Petroleum Science 23 (2026) 1280–1292

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

A novel application of a stepwise extraction protocol: resolving the challenge of biomarker detection in high-maturity shale systems

Jing-Ya Zhang [a,b,*], Bo-Wen Meng [a,b], Ru-Kai Zhu [a,b], Hua-Sen Zeng [b,c], Chang Liu [a,b]

[a] Research Institute of Petroleum Exploration and Development, PetroChina, Beijing, 100083, China
[b] State Key Laboratory of Continental Shale Oil, Daqing, 163712, Heilongjiang, China
[c] Research Institute of Petroleum Exploration and Development of Daqing Oilfield Company Ltd., Daqing, 163712, Heilongjiang, China

ARTICLE INFO

Article history:
Received 5 March 2025
Received in revised form
19 November 2025
Accepted 17 December 2025
Available online 5 January 2026

Edited by Min Li

Keywords:
Source rocks
Biomarker identification
High maturity
Stepwise extraction
Gas chromatography-mass spectrometry (GC-MS)

ABSTRACT

Biomarkers in source rocks are crucial in uncovering biological source and depositional environment. However, the detection of critical biomarkers such as steranes and terpanes in high-maturity source rocks is often challenging. On one hand, biomarkers in high-maturity source rocks are inherently present at low concentrations; on the other hand, the detection of these low-concentration biomarkers can be hampered by the high concentrations of interfering compounds, such as *n*-alkanes and other branched/cyclic alkanes. In this study, we developed a sample pretreatment protocol of stepwise extraction, using the combination of solvent in varying polarities (i.e., *n*-hexane and dichloromethane) and samples of different sizes (i.e., blocks >3 cm, 2.5–5 mesh, 10–20 mesh). The specific procedures are as follows: (1) removing the high amounts of interfering compounds during the first-step extraction to increase the relative proportion of biomarkers in saturated fraction by extracting coarse samples with relatively weak-polar solvents; (2) analyzing the soluble organic matter extracted in the second step by Gas chromatography-mass spectrometry (GC-MS) for biomarker (i.e., steranes and terpanes) detection by extracting samples of 100–200 mesh with solvents of stronger polarity. This method was demonstrated using organic-rich shales with high maturity (R_o = 1.42%) from the Qingshankou Formation (Fm.) in the Gulong Sag, and it was validated using the medium maturity (R_o = 1.0%) shales from the Qingshankou Fm., Sanzhao Sag, Songliao Basin. The results show that under routine Soxhlet extraction combining GC-MS analysis, sterane and terpane biomarkers in the Qingshankou shales are unidentifiable, regardless of whether urea adduction is applied or not. In contrast, these biomarkers can be successfully identified utilizing stepwise extraction coupled with GC-MS analysis. The effectiveness of biomarker identification is influenced by the interplay of both the shale particle size and organic solvent polarity. The integration of biomarker indexes was then employed to interpret the biological source and the depositional environment of the organic matter, providing detailed insights of the hydrocarbon generation potential of shales. This information can further provide additional guidance for selecting favorable resource zones for shale oil exploration in the Gulong Sag, Songliao Basin.

1. Introduction

The concept of biomarker, which refers to the molecular fossils broadly existing in sediments, rocks and crude oil, could date back to as early as 1934 (Treibs, 1934; Peters and Moldowan, 1993). Since the mid-1970s, the identification of these biomarkers has been significantly enhanced by the commercial availability of gas chromatography-mass spectrometry (GC-MS) (Hunt et al., 2002). To date, the routine pretreatment involves solvent extraction followed by solid phase column chromatography prior to the GC-MS analysis (Philp and Lewis, 1987; Hwang, 1990). However, such a simple process is not always sufficient for detecting biomarkers. For example, Li et al. (2006) observed an absence of steranes and terpanes in the high wax oils from an Ordovician buried hill when investigating the oil-source correlation. They suggested that the high content of waxy alkanes (>40%) and the significant proportion of saturated hydrocarbons

* Corresponding author.
E-mail address: zhangjingya@petrochina.com.cn (J.-Y. Zhang).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2025.12.030

(88%) hampered the identification of the cyclic biomarkers. Therefore, molecular sieving or urea adduction have been proposed to remove the *n*-alkanes and thereby concentrate branched/cyclic alkanes, on the basis that molecular diameter of *n*-alkanes is smaller enough to be adsorbed in pores within molecular sieve or urea crystal (Michalczyk, 1985; Peters et al., 2005; Xu and Sun, 2005; Grice et al., 2008; Zhang et al., 2020). Using these methods, steranes and terpanes were successfully detected in extremely waxy samples in Bohai Bay Basin in China (Li et al., 2006; Lü et al., 2008) and Tarinaki Basin in New Zealand, etc. (Czochanska et al., 1988).

However, we notice that, for the high-mature shales from the Cretaceous Qingshankou Formation (Fm.) in the Gulong sag, Songliao Basin, biomarkers in source rocks remain unidentifiable even after urea adduction process. The Qingshankou Fm. which extensively distributes throughout the basin is regarded as the major source rock in the Songliao Basin (Feng et al., 2010). Previous studies showed that the Qingshankou Fm. is characterized by high total organic carbon (TOC) content and good organic matter types (oil-prone Type I and II kerogens), which is mainly contributed by prokaryotes and eukaryotic algae input (Zhang P. L. et al., 2021; Liu et al., 2022; Xiao et al., 2024). Such reliable information for immature to mature shales ($0.8\% < R_o \leq 1.3\%$) could be easily provided by maceral composition and biomarkers. However, as the thermal maturity of kerogen evolves ($R_o > 1.3\%$), the fluorescence weakened due to strengthened absorption (Lin and Davis, 1988; Lin and Ritz, 1993; Hackley et al., 2017). Meanwhile, the higher stage of thermal evolution accelerates the degradation of biomarkers, including steranes and terpanes, resulting in inherently very low concentrations of biomarker (Farrimond et al., 1998; Fang et al., 2019). Additionally, high thermal maturity facilitates the cracking of long-chain hydrocarbons, forming a large amount of aliphatic component of medium chains and ring structures (Kikuchi et al., 2010; Tissot and Welte, 2013; Vinnichenko et al., 2021), further interfering the detection of the biomarkers at low concentrations (Huang et al., 1992; Philp, 2000).

In this study, the target layer Qingshankou Fm. in the Gulong Sag bears very high maturity (R_o in the range of 1.15%–1.45%) (Sun et al., 2021; Zhang et al., 2023), significantly challenging the detection of sterane and terpane at low concentrations (Feng et al., 2011; Bi et al., 2021). Yet, little is known about the characteristics of biomarkers in the high-maturity shales in the Gulong Sag, and hence limited geological and geochemical information is available for this region. To this end, we designed a series of stepwise extraction experiments using organic solvents of varying polarity and rock samples of different sizes to determine the optimal conditions for the biomarker detection. Compared with the routine Soxhlet extraction either with urea adduction or not, the stepwise extraction combined with GC-MS is highly effective in reducing the impact of excessive *n*-alkanes and interfering cyclic/branched compounds, thereby resolving biomarkers (i.e., sterane and terpane) in source rocks. This approach for biomarker detection yields satisfactory results in the high-maturity shales of the Qingshankou Fm., Gulong Sag. The biomarker indexes obtained by this method were then utilized to interpret the origin, type and depositional environment of organic matter, providing an in-depth insight into the hydrocarbon generation potential of the high-maturity Qingshankou shales. This information shed light on the resource potential and exploration prospect of the shale oil in the Gulong Sag.

2. Analytical methods

2.1. Samples

The highly mature core samples of organic-rich shale in the Qingshankou Fm. were retrieved from Well GY8HC, Gulong Sag, Songliao Basin (Fig. 1). The shale oil produced from this well in the Qingshankou Fm. is characterized by high wax content (28.0%) and low crude oil density (0.8053 cm^3/g). The shale samples are in high maturity, with the vitrinite reflectance (R_o) of 1.42% and T_{max} values of 460–467 °C. The shales in Gulong Sag are rich in organic matter, characterized by the TOC of 2.42–3.15 wt%. The content of free hydrocarbon (S_1) and pyrolysis hydrocarbon (S_2) for Sample A is 3.65 and 7.29 mg/g respectively; and it is 2.69 and 4.83 mg/g respectively for Sample B (Table 1). In the following Section 3, Sample A was selected as a representative case to illustrate the results obtained under different experimental conditions for the purpose of determining the optimal conditions for biomarker identification. Subsequently, the optimal experimental conditions were then applied to both Samples A and B to obtain biomarker indexes for further geological insights of the shales, as discussed in Section 4.

The core sample of lower maturity was retrieved from the Qingshankou Fm. in Well ZY1, located in the Sanzhao Sag, Songliao Basin, to validate the designed experiment. The sample shows maturity with an R_o value of 1.0%, and it is rich in organic matter, exhibiting TOC of 1.76%, S_1 of 1.86 mg/g and S_2 of 9.29 mg/g (see Table 1). The experimental process was conducted using both routine Soxhlet extraction and stepwise extraction, based on the optimal experimental conditions determined from the analysis of the GY8HC sample.

2.2. Routine Soxhlet extraction

Core sample of ~30 g (according to the content of organic matter) was powdered into 100–200 mesh after the outer layer was removed to avoid the interference of drilling fluid. The powder sample was then extracted by dichloromethane (DCM) in a Soxhlet apparatus at 60 °C for 72 h to ensure complete extraction of soluble organic matter. Activated copper was added in the flask of the Soxhlet extractor to remove sulfur from the organic extract.

2.3. Stepwise extraction

The proposed experiments designed in batches are based on the combination of varying solvent polarity and samples sizes. Fig. 2 shows the sketch of the technical process. In detail, high-maturity rock samples of ~200 g, depending on the content of the extractable organic matter (EOM) were prepared after removing the outer layer of rocks, in order to avoid the interference of drilling fluids during core sampling. The sample was evenly separated by weight and crushed into three different sample sizes, namely blocks >3 cm, 2.5–5 mesh, and 10–20 mesh, respectively (each sample weighs approximately 20–30 g), while samples of the same size were further divided evenly by weight into two halves and were extracted by different solvent combination.

Of all 6 batches of samples, 3 batches of each size were extracted continuously in a Soxhlet apparatus at 60 °C for 24 h to get soluble extract using *n*-hexane of 150 mL, while the rest 3 batches were extracted continuously for 24 h using DCM of 150 mL. The extracts were labeled as EOM A1–EOM A6, in the order of descending sample size and increasing solvent polarity (Fig. 2).

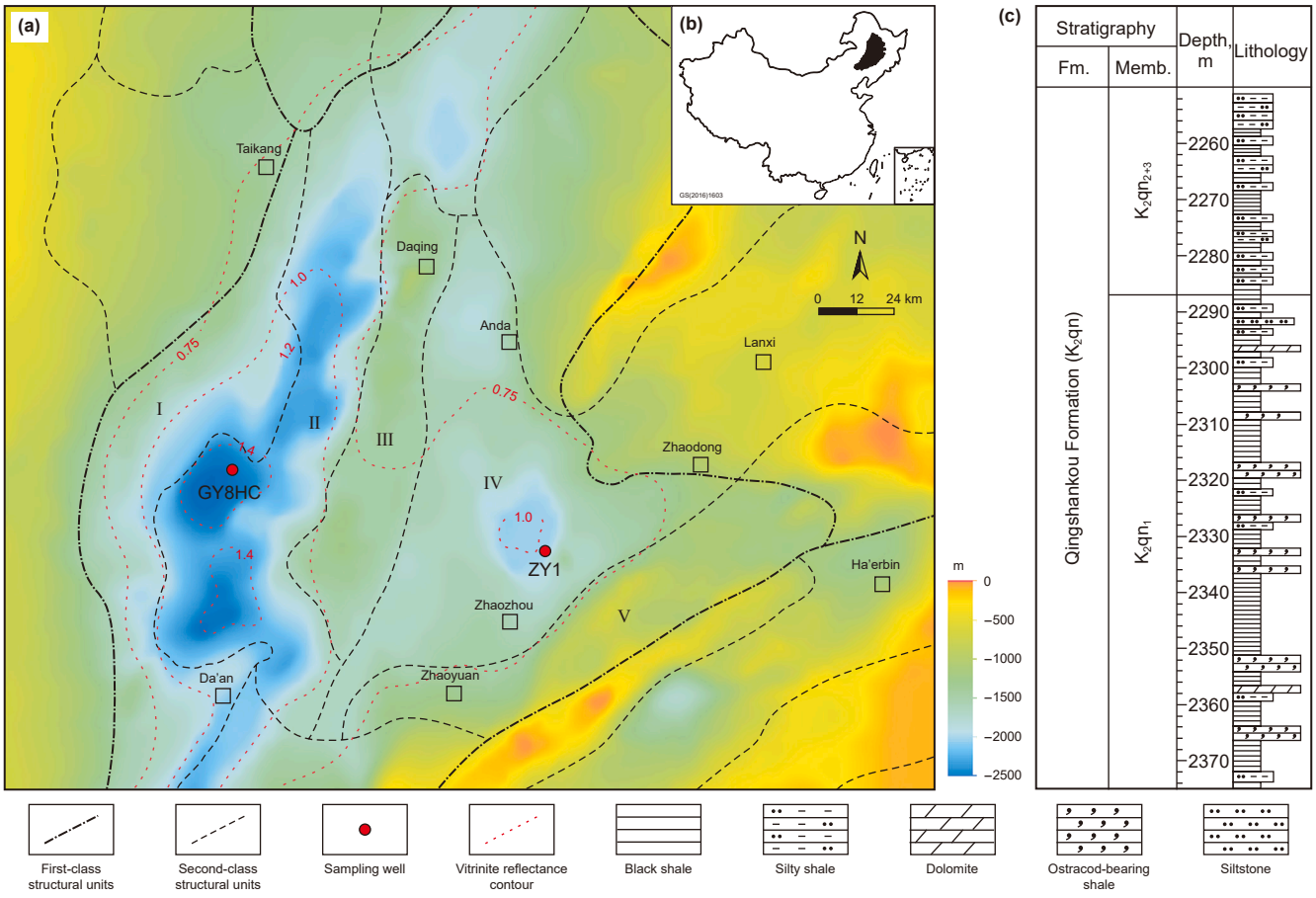

Fig. 1. Location map of the Gulong Sag, Songliao Basin **(a)**, with inset showing the location of Songliao Basin in black color **(b)** and general stratigraphic column of the Qingshankou Fm **(c)**. I: Longhupao-da'an Terrace; II: Qijia-Gulong Sag; III: Daqing placanticline; IV: Sanzhao Sag; V: Chaoyanggou Terrance. The color ramp shows the altitudes of the bottom of the Qingshankou Formation.

Table 1
Bulk organic geochemical data for the shale samples from the Qingshankou Fm. in Gulong and Sanzhao Sag, Songliao Basin.

Well	Sample	TOC, %	S_1, mg/g	S_2, mg/g	T_{max}, °C
GY8HC	A	3.15	3.65	7.29	467
GY8HC	B	2.42	2.69	4.83	460
ZY1	C	1.76	1.86	9.29	445

Then, all 6 batches of residual shale samples after the 1st step extraction were powered into 100–200 mesh and extracted by 150 mL DCM at 60 °C for 48 h continuously. The extracts were noted as EOM B1–EOM B6 (Fig. 2). All extracts of EOM A1–EOM A6 and EOM B1–EOM B6 were collected and analyzed to quantify chloroform bitumen "A" content.

2.4. SARA (saturate, aromatic, resin, and asphaltene) fractionation

The separation procedure of EOM from both the routine extraction and designed stepwise extraction is described as the following. Asphaltenes were precipitated applying the method described by Theuerkorn et al. (2008). The soluble extract were afterward eluted on silica gel/Al_2O_3 column by *n*-hexane, DCM and *n*-hexane (2:1, *v:v*) and chloroform accordingly into saturated fraction, aromatic fraction and resin respectively. Quantification of each fraction was accomplished by weighing the residue after solvent was evaporated and the percentage of each fraction was determined accordingly.

2.5. GC-MS analysis

The saturated fraction was analyzed using Agilent 7890 GC coupled with an Agilent 5975C MSD. An HP-5 MS capillary column (60 m × 250 μm × 0.25 μm) was equipped, using He as carrier gas at a flow rate of 1.0 mL/min. The temperature program was set as 50 °C (held for 1 min) to 120 °C at 20 °C/min, followed by 4 °C/min to 250 °C and finally 3 °C/min to 310 °C (held for 30 min). The MS was operated in electron impact (EI) mode, with ionizing energy of 70 eV, ion source of 230 °C and scan range of *m/z* 50–600. Quantification of *n*-alkanes, steranes and terpanes in saturated hydrocarbon fraction were accomplished using D_{50}-nC_{24} as internal standard.

2.6. Urea adduction

n-Hexane of 1 mL was added to the vial of saturated fraction to fully dissolve the hydrocarbons. Urea-methanol solution was then added by drop until the floc stopped precipitation. After the vial was sealed and placed under 4 °C for 12 h, *n*-hexane was used to extract the branched/cyclic alkanes. The inclusion compound was then dissolved using ultrapure water and were extracted by *n*-

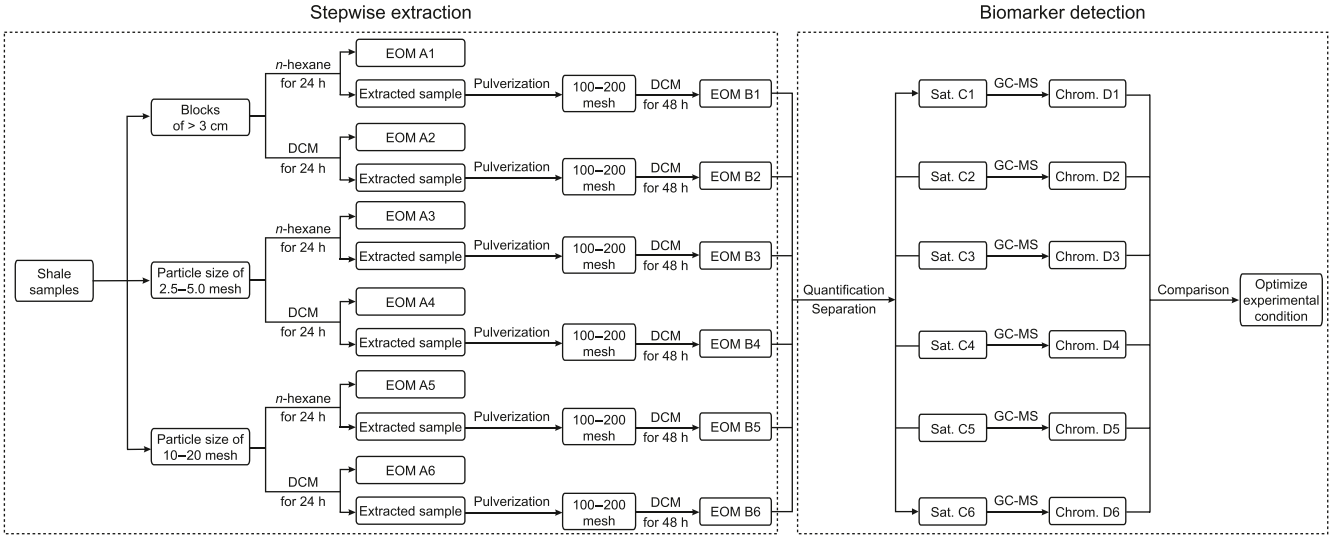

Fig. 2. Flow chart of the proposed stepwise extraction and biomarker detection approach. EOM: extractable organic matter; Sat.: saturated hydrocarbon fraction; Chrom.: ion chromatogram.

hexane. Both branched/cyclic alkane and *n*-alkane fraction were further eluted with *n*-hexane on a silica gel column to remove the remained urea.

3. Result and discussion

3.1. Routine Soxhlet extraction and GC-MS analysis

The integration of Soxhlet extraction, SARA fractionation, and GC-MS analysis of saturated hydrocarbon remains the current routine analytical method for detecting biomarkers. The result shows that the shale sample in the Qingshankou Fm. contains chloroform bitumen "A" content of 8.1 mg/g, which is composed of saturated hydrocarbons, aromatic hydrocarbons, resins, and asphaltenes of 79.8%, 4.5%, 5.4% and 5.5% respectively, demonstrating that saturated hydrocarbon is the dominant fraction (Table 2). The saturated fraction is mainly characterized by *n*-alkanes from nC_{10}–nC_{34}, with concentration of 4527.9 μg/g rock, showing an asymmetric distribution without odd to even carbon preference (carbon preference index, CPI = 1.11), with the main peak carbon as nC_{17} (Fig. 3(a)). The significantly high content of *n*-alkanes and other branched/cyclic alkanes entirely masks the signals of the biomarker, leading to the series of sterane and terpane unidentifiable and thus not quantifiable by the routine analysis of GC-MS (Fig. 3(b) and (c)).

3.2. Urea adduction and GC-MS analysis

To remove the remarkably high content of *n*-alkanes, we utilized urea adduction to separate the saturated hydrocarbon into *n*-alkane and branched/cyclic alkane fractions. The result suggested that *n*-alkanes are largely removed from saturated hydrocarbon fractions by urea adduction (Fig. 4(a)). However, even after urea adduction, the biomarkers including the tricyclic terpane,

Table 2
Comparison of routine and stepwise extraction methods on extract yields and SARA fractions in high-maturity shales of the Qingshankou Fm., Songliao Basin.

Well	No.	Sample size	Solvent (1st step)	Solvent (2nd step)	The 1st step extraction					The 2nd step extraction					Proportion of the 1st step extract, %	Proportion of the 2nd step extract, %
					"A", mg/g	Sat., wt%	Aro., wt%	Res., wt%	Asp., wt%	"A", mg/g	Sat., wt%	Aro., wt%	Res., wt%	Asp., wt%		
GY8HC	1	100–200 mesh	DCM	n.a.	8.10	79.8	4.5	5.4	5.5	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	2	blocks >3 cm	*n*-hexane	DCM	2.81	87.5	3.2	3.8	1.4	4.65	74.5	6.8	10.4	5.2	37.6	62.4
	3	2.5–5 mesh	*n*-hexane	DCM	6.33	82.0	5.7	3.9	2.0	1.16	37.7	14.8	36.9	5.1	84.6	15.4
	4	10–20 mesh	*n*-hexane	DCM	7.08	82.8	4.8	4.4	1.6	0.33	40.7	20.4	24.1	9.3	95.6	4.4
	avg.	–	*n*-hexane	DCM	5.41	84.1	4.6	4.0	1.7	2.05	51.0	14.0	23.8	6.5	72.6	27.4
	5	>3 cm blocks	DCM	DCM	3.90	76.9	8.4	9.0	2.8	3.69	64.6	9.1	11.0	8.4	51.4	48.6
	6	2.5–5 mesh	DCM	DCM	7.15	74.5	6.7	9.5	3.8	1.00	57.4	10.3	21.7	8.5	87.8	12.2
	7	10–20 mesh	DCM	DCM	7.76	73.1	7.5	9.0	5.1	0.31	30.4	25.3	26.6	12.7	96.2	3.8
	avg.	–	DCM	DCM	6.27	74.8	7.5	9.2	3.9	1.67	50.8	14.9	19.8	9.9	78.5	21.5
ZY1	8	100–200 mesh	DCM	n.a.	4.11	67.0	8.4	12.4	6.9	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	9	10–20 mesh	*n*-hexane	DCM	2.81	79.7	3.4	2.5	2.5	1.84	67.6	8.8	14.7	6.5	60.4	39.6

n.a.: not available; avg.: average value; "A": chloroform bitumen "A"; Sat.: saturate; Aro.: aromatic; Res.: resin; Asp.: asphaltene.

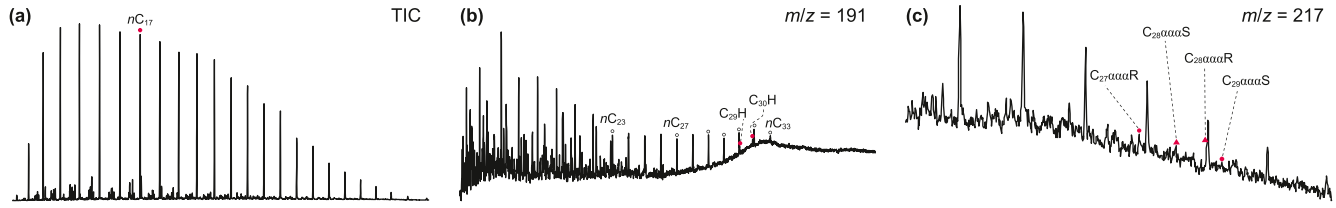

Fig. 3. Chromatograms of TIC **(a)**, *m/z* 191 **(b)** and *m/z* 217 **(c)** of saturated hydrocarbons in the shale of the Qingshankou Formation, Gulong Sag. $C_{29}H$: 17α, 21β-30-norhopane; $C_{30}H$: 17α, 21β-hopane; C_{27}αααR: 20R-ααα-cholestane; C_{28}αααS: 20S-24-methyl-ααα-cholestane; C_{28}αααR: 20R-24-methyl-ααα-cholestane; C_{29}αααS: 20S-24-ethyl-ααα-cholestane.

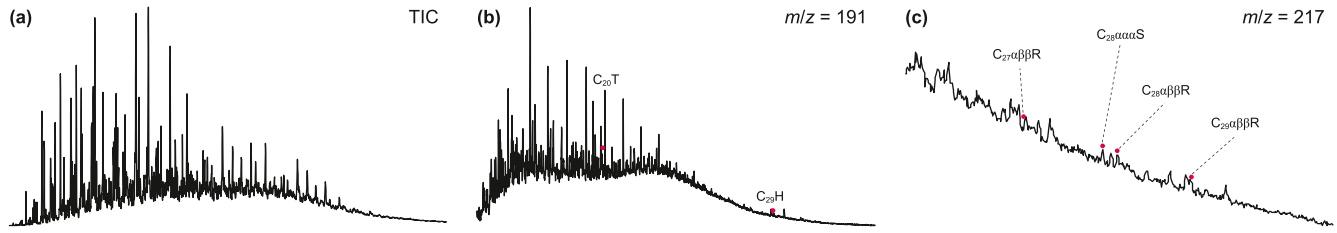

Fig. 4. Chromatograms of TIC **(a)**, *m/z* 191 **(b)** and *m/z* 217 **(c)** of saturated hydrocarbons after urea adduction process in the shale sample from the Qingshankou Formation, Gulong Sag. $C_{20}T$: C_{20} tricyclic terpane; $C_{29}H$: 17α,21β-30-norhopane; C_{27}αββR: 20R-αββ-cholestane; C_{28}αααS: 20S-24-methyl-ααα-cholestane; C_{28}αββR: 20R-24-methyl-αββ-cholestane; C_{29}αββR: 20R-24-ethyl-αββ-cholestane.

pentacyclic terpane and the regular sterane series remain unidentifiable in the shales of the Qingshankou Fm., Gulong Sag (Fig. 4(b) and (c)). Moreover, it may also concurrently lead to a significant loss of pentacyclic triterpenoid and steranes with relatively higher molecular mass (Han et al., 2017). Therefore, the distribution of the mass chromatogram such as *m/z* 191 and *m/z* 217 is distorted (Fig. 4(b) and (c)).

3.3. Stepwise extraction and GC-MS analysis

3.3.1. Experimental procedure and basic principles

Stepwise extraction has been applied in reservoir geochemistry research, primarily focusing on hydrocarbon migration and accumulation in multi-stages, and the characteristics of microscopic oil occurrence in pore spaces (Wilhelms et al., 1996; Gorynski et al., 2019; Zhang et al., 2019, 2023; Liu et al., 2020). Each step of the extraction allows the removal of soluble organic matter with distinct properties, which presumably represents hydrocarbon recharging in different stages or distinguishes free oil from adsorbed oil in pore spaces (Sajgó et al., 1983; Price and Clayton, 1992; Liu et al., 2020). Here, we apply the approach innovatively to detect biomarker effectively, by designing a series of stepwise extraction experiment. As such, the impact of the interfering compounds including high concentration of *n*-alkane and other branched/cyclic alkanes is decreased, thereby enhancing the relative abundance of biomarker and improving their detection efficacy.

The procedure of stepwise extraction adopted in this study is detailed in Section 2.3 (Fig. 2). To determine the optimal experimental condition for high maturity shales of the Qingshankou Fm., 6 sets of experiments were designed for comparison, with organic solvents of varying polarity and rock samples of different grain size. The 1st step extraction on coarse samples (noted as EOM A1–EOM A6 in Fig. 2) mainly aims at removing the majority of *n*-alkanes and other branched/cyclic alkanes from the saturated hydrocarbon compounds. The 2nd step involves extracting on powdered samples (the obtained extracts are noted as EOM B1–EOM B6 in Fig. 2) for the analysis and identification of biomarkers such as steranes and terpanes. The EOM B1–EOM B6 was then quantified and separated into SARA fractions. Then the saturated hydrocarbon fractions (noted as Sat. C1–Sat. C6 in Fig. 2) were analyzed using GC-MS. Subsequently, the ion chromatograms of the 6 batches of saturated hydrocarbon fractions (noted as Chrom. D1–Chrom. D6 in Fig. 2) were compared to determine the optimal solvent combination and sample size for the biomarker detection. The fundamental principle is that provided biomarkers such as steranes and terpanes can be clearly identified in the chromatograms, the content of key biomarkers should be maximized.

3.3.2. Content of EOM and SARA fractions

The content of EOM and SARA fractions for 6 batches of stepwise extraction is listed in Table 2. For each set of experiment, the sum of extracted chloroform bitumen "A" by the two steps of extraction (ranging from 7.41 to 8.15 mg/g) is approximate to that of routine extraction (8.1 mg/g). For shale samples of the same particle size, the combination of DCM (1st step) and DCM (2nd step) (noted as DCM + DCM) extracted more chloroform bitumen "A" than the combination of *n*-hexane (1st step) and DCM (2nd step) (noted as Hex + DCM), whereas particle size had minimal impact on the total content of chloroform bitumen "A" extracted.

The proportion of extracts in each step is influenced by both of particle size and organic solvent combination. For the 1st step extraction, the content of chloroform bitumen "A" using *n*-hexane increases from 2.81, 6.33–7.08 mg/g as the particle size of the samples decrease from blocks >3 cm, 2.5–5 mesh to 10–20 mesh (Table 2). Whereas for the 2nd step extraction, the content of extracts decreases from 4.65, 1.16 to 0.33 mg/g as the particle size becomes finer. Similarly, the shale samples extracted by DCM + DCM show a consistent pattern with extracts by Hex + DCM. Such results imply that finer particle size allows for a larger contact area between the samples and the solvent, thus increasing the efficiency of the 1st step extraction and facilitating the extraction of more soluble organic matter. The organic solvent polarity affects the content and proportion of extracts for each step as well. The 1st step extraction using DCM could extract additional 0.7 mg/g (for samples of 10–20 mesh), 0.8 mg/g (for samples of 2.5–5 mesh) and 1.1 mg/g (for samples of blocks >3 cm) of chloroform bitumen "A" than using *n*-hexane in the 1st step, due to the lower polarity of *n*-hexane compared with that of DCM.

Meanwhile, polarity of organic solvent also exert influence on the percentage of saturated hydrocarbon. When *n*-hexane is used

in the 1st step extraction, the percentage of saturated hydrocarbon (average = 84.1 wt%) is higher than that of the routine extraction (79.8 wt%). In contrast, the extraction with DCM solvent displays a lower percentage of saturated hydrocarbon, with an average of 74.8 wt% (Table 2). This is due to the higher polarity of DCM, which extracts a relatively larger amount of organic matter of polar compounds, lowering the proportion of saturated fraction. For the 2nd step extraction, the percentage of saturated hydrocarbon is lower than that of the 1st step, with a more pronounced decrease observed as the particle size becomes finer, regardless of with *n*-hexane or DCM.

3.3.3. GC-MS of saturated hydrocarbons

The GC-MS results of EOM A1–EOM A6 show that *n*-alkanes are the predominant compounds in the saturated fractions, although varied owing to the solvent used and sample size (Fig. 5). Chromatograms of *m/z* 191 and *m/z* 217 demonstrate that terpanes and steranes are not able to be detected at this experimental step regardless of solvent polarity or sample particle size employed (Fig. 5). The results of GC-MS analysis of the saturated fractions of EOM B1–EOM B6 are presented in Fig. 6. For the samples of blocks >3 cm, the *n*-alkane distribution exhibits unimodal pattern with the major peak of nC_{17} and nC_{20}, respectively (see TIC in Fig. 6(a) and (b)). The content of *n*-alkane is 2430.6 μg/g rock by Hex + DCM and 1543.9 μg/g rock by DCM + DCM solvent combination respectively (Table 3). However, the mass chromatograms of *m/z* 191 show that terpane biomarkers were mostly unidentifiable, except C_{21} tricyclic terpane ($C_{21}T$), C_{23} tricyclic terpane ($C_{23}T$) and C_{30} 17α-hopane ($C_{30}H$), either by extraction with Hex + DCM or DCM + DCM solvent combination (shown as *m/z* 191 in Fig. 6(a)

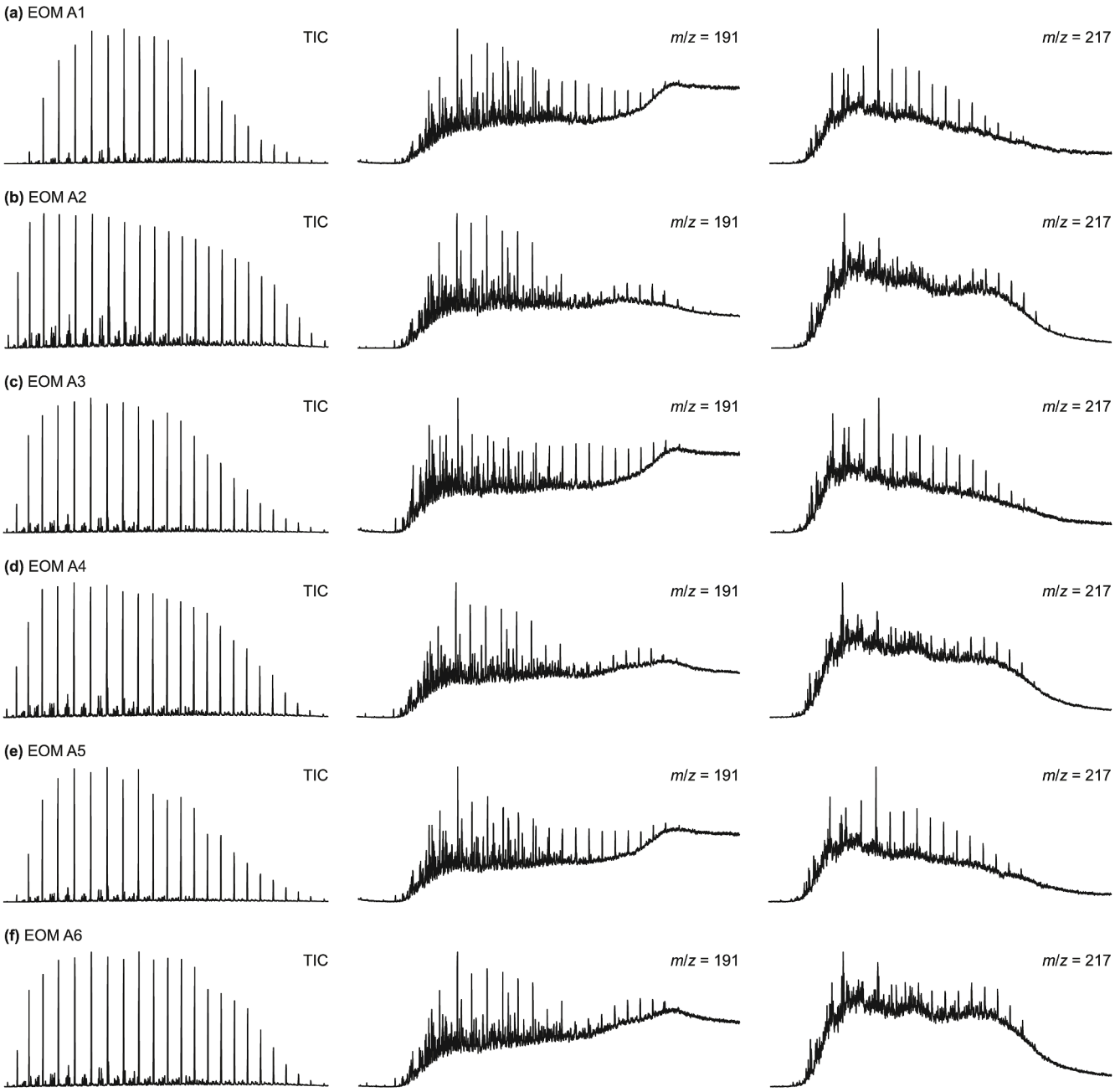

Fig. 5. Chromatograms (from left to right: TIC, *m/z* 191 and *m/z* 217) of saturated fractions of the 1st step extracts (EOM A1–A6) from high-maturity shales in the Qingshankou Formation using stepwise extraction.

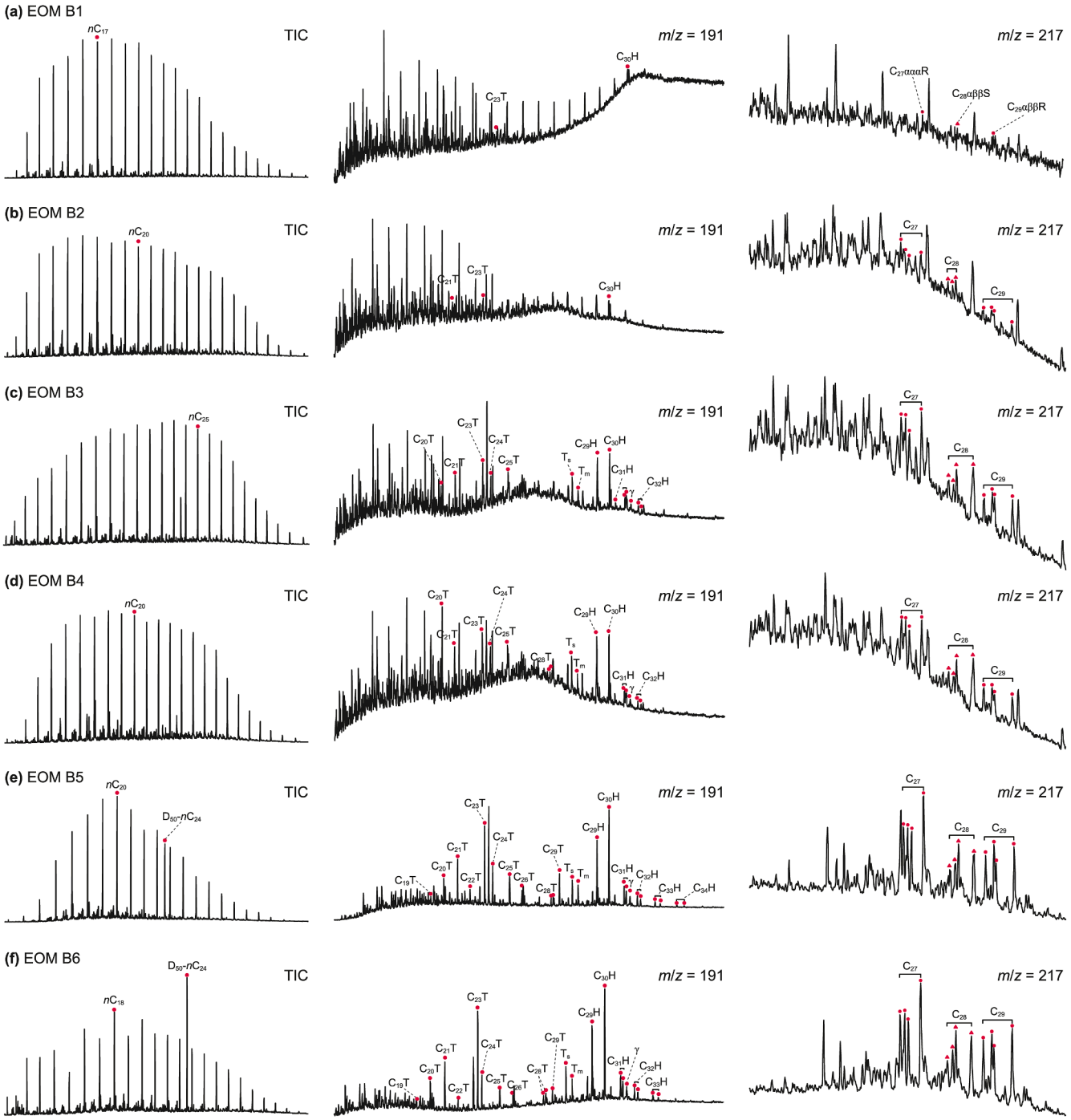

Fig. 6. Chromatograms (from left to right: TIC, *m/z* 191 and *m/z* 217) of saturated fractions of the 2nd step extracts (EOM B1–B6) from high-maturity shales in the Qingshankou Formation using stepwise extraction. Pr: Pristane; Ph: Phytane; T: tricyclic terpane; H: hopane; γ: gammacerane; C_{27}: C_{27} regular sterane; C_{28}: C_{28} regular sterane; C_{29}: C_{29} regular sterane.

and (b)), which is similar to the result of routine Soxhlet extraction (Fig. 3). Similar issue also arises during the identification of steranes, where peaks of regular sterane are still suppressed even after stepwise extraction (shown as *m/z* 217 in Fig. 6(a) and (b)). Therefore, only a few molecular configurations of regular steranes have been identified vaguely (e.g., C_{27} ααα20R, C_{28} ααα20S and C_{29} αββ20R regular sterane, etc.). In comparison, the sample of blocks >3 cm extracted by DCM + DCM display more clearly identifiable peaks of regular sterane in its *m/z* 217 chromatogram compared with that extracted by Hex + DCM.

For shale samples of 2.5–5 mesh, the *n*-alkane distribution is unimodal with the major peak of nC_{25} and nC_{20} (see TIC in Fig. 6(c) and (d)). The content of *n*-alkane is 285.5 μg/g rock with Hex + DCM, and 378.3 μg/g rock with DCM + DCM respectively (Table 3). Although *n*-alkanes and branched alkanes still interfere with biomarker identification, most of the regular steranes, tricyclic terpanes and hopanes could be identified in both saturated fractions of EOM B3 and EOM B4 (Fig. 6(c) and (d)). However, the identification of short-chain tricyclic terpanes (e.g., C_{19} to C_{22} tricyclic terpanes) is still hindered by some branched alkanes co-

Table 3
The content of *n*-alkane, sterane and terpane in high-maturity shales from the Qingshankou Fm. by stepwise extraction using different organic solvent and sample size.

No.	Sample size	Solvent (1st step)	Solvent (2nd step)	The 2nd step extraction			
				Content of *n*-alkane, μg/g rock	Content of terpanes, μg/g rock	Content of steranes, μg/g rock	Sum of steranes and terpanes, μg/g rock
1	Blocks >3 cm	*n*-Hexane	DCM	2430.6	n.d.	n.d.	n.d.
2	2.5–5 mesh	*n*-Hexane	DCM	285.5	0.225	0.059	0.284
3	10–20 mesh	*n*-Hexane	DCM	79.8	0.373	0.089	0.462
4	Blocks >3 cm	DCM	DCM	1543.9	n.d.	n.d.	n.d.
5	2.5–5 mesh	DCM	DCM	378.3	0.389	0.097	0.485
6	10–20 mesh	DCM	DCM	45.2	0.178	0.048	0.226

n.d.: not detected.

eluted. Additionally, chromatograms of both *m/z* 191 and *m/z* 217 show a hump in baseline, indicating a need for further concentration of biomarkers and additional separation of interfering compounds (Fig. 6(c) and (d)).

TIC chromatograms of 10–20 mesh reveal a major peak of *n*-alkane peak at nC_{20} by Hex + DCM extraction and at nC_{18} by DCM + DCM extraction, respectively (Fig. 6(e) and (f)). Notably, the *n*-alkane for samples of 10–20 mesh is 79.8 μg/g rock with Hex + DCM extraction and 45.2 μg/g rock with DCM + DCM extraction, which is much less than that of >3 cm and 2.5–5 mesh, as seen in Table 3. This possibly reduces the interference on the detection of biomarkers. The mass chromatograms of *m/z* 191 and *m/z* 217 demonstrate that steranes and terpanes are fully identifiable regardless of the solvent combination (Fig. 6(e) and (f)). In comparison, Hex + DCM extraction yields superior baselines in the mass chromatograms of *m/z* 191 and *m/z* 217 (Fig. 6(e)). More importantly, the content of biomarkers including terpanes and steranes using Hex + DCM extraction is 0.462 μg/g rock, double the content obtained with DCM + DCM (0.226 μg/g rock) (Table 3).

We therefore suggest that for shales in the Qingshankou Fm., Gulong Sag, sample size of 10–20 mesh extracted with Hex + DCM represents the most suitable experimental condition. The results indicate that the content of terpane and sterane extracted is 0.373 and 0.089 μg/g rock respectively, and hence the content of key biomarkers in total is 0.462 μg/g rock for the 2nd step extraction (Table 3). Considering the majority of chloroform bitumen "A" (7.08 mg/g) removed in the 1st step extraction, we therefore estimate the key biomarker content for the original shale sample to be no less than 0.021 μg/g rock, as some biomarkers were also removed during the 1st step extraction.

3.4. Validation of designed experiment

3.4.1. Samples and experimental procedure

The medium thermal maturity shale sample (Sample C, R_o = 1.0%, see Table 1) from Well ZY1 in the Qingshankou Fm., Songliao Basin was analyzed to validate the proposed stepwise extraction method for biomarker detection. Initially, both routine Soxhlet extraction and stepwise extraction were performed. The specific experimental procedures for the stepwise extraction adhered to the optimal conditions described in Section 3.3.3. This process involved the use of shale samples with a particle size of 10–20 mesh, with *n*-hexane employed for the 1st step extraction and DCM for the 2nd step. Subsequently, SARA fractionation and GC-MS analysis were carried out on both the routine Soxhlet extracts and stepwise extracts. The ratios of biomarker parameters obtained from the stepwise extraction were compared to those from the routine Soxhlet extraction to assess the reliability of the experiment designed.

3.4.2. Content of EOM and SARA fractions

The results concerning the contents of soluble organic matter and the characteristics of SARA fraction are presented in Table 2. It indicates that the shale sample subjected to routine Soxhlet extraction contains chloroform bitumen "A" content of 4.11 mg/g, comprising saturated hydrocarbons (67.0%), aromatic hydrocarbons (8.4%), resins (12.4%), and asphaltenes (6.9%). The outcomes of the stepwise extraction reveal that the chloroform bitumen "A" content for the 1st step extraction is 2.81 mg/g, which constitutes 60.4% of the total; while the 2nd step extraction yields a content of 1.84 mg/g, accounting for 39.6%. The ultimate extract (the 2nd step) exhibits SARA fractions of 67.6%, 8.8%, 14.7% and 6.5%, which closely correspond to those obtained through routine Soxhlet extraction.

3.4.3. GC-MS results and biomarker parameters

(1) *n*-alkanes and isoprenoids

The results of GC-MS analysis of the saturated fraction obtained from routine Soxhlet extraction are presented in Fig. 7(a). The extract demonstrates a carbon number distribution that spans the entire range of *n*-alkanes and isoprenoids from C_{11} to C_{36}, characterized by a unimodal distribution, with the predominant constituents of nC_{20}. The CPI value for the extracts is 1.12, while the pristane to phytane (Pr/Ph) ratio is 0.95 (Tables 4 and 5).

Similarly, the final results of the stepwise extraction exhibit comparable unimodal characteristics of *n*-alkanes. The carbon number in the 2nd step extracts also ranges from C_{11} to C_{36}, with nC_{20} as the dominant peak (Fig. 7(c)). The CPI value for these extracts is 1.1, and the Pr/Ph ratio remains at 0.95 (Tables 4 and 5).

(2) Terpanes

The distributions of tricyclic and pentacyclic terpanes from routine Soxhlet extraction and stepwise extraction are illustrated in the ion chromatograms of *m/z* 191 (Fig. 7), with the relative abundances and relevant parameters presented in Tables 4 and 5 In both extracts, tricyclic terpanes are predominantly composed of C_{21} tricyclic terpanes, followed closely by C_{23} tricyclic terpanes (Fig. 7(a) and (c)). The ratios of short-chain tricyclic and tetracyclic terpanes for both extracts are consistent. For example, the ratios of $(C_{19}+C_{20})/C_{23}$ tricyclic terpane, $C_{19}/(C_{19}+C_{23})$ tricyclic terpane and C_{24} tetracyclic terpane/(C_{24} tetracyclic terpane + C_{23} tricyclic

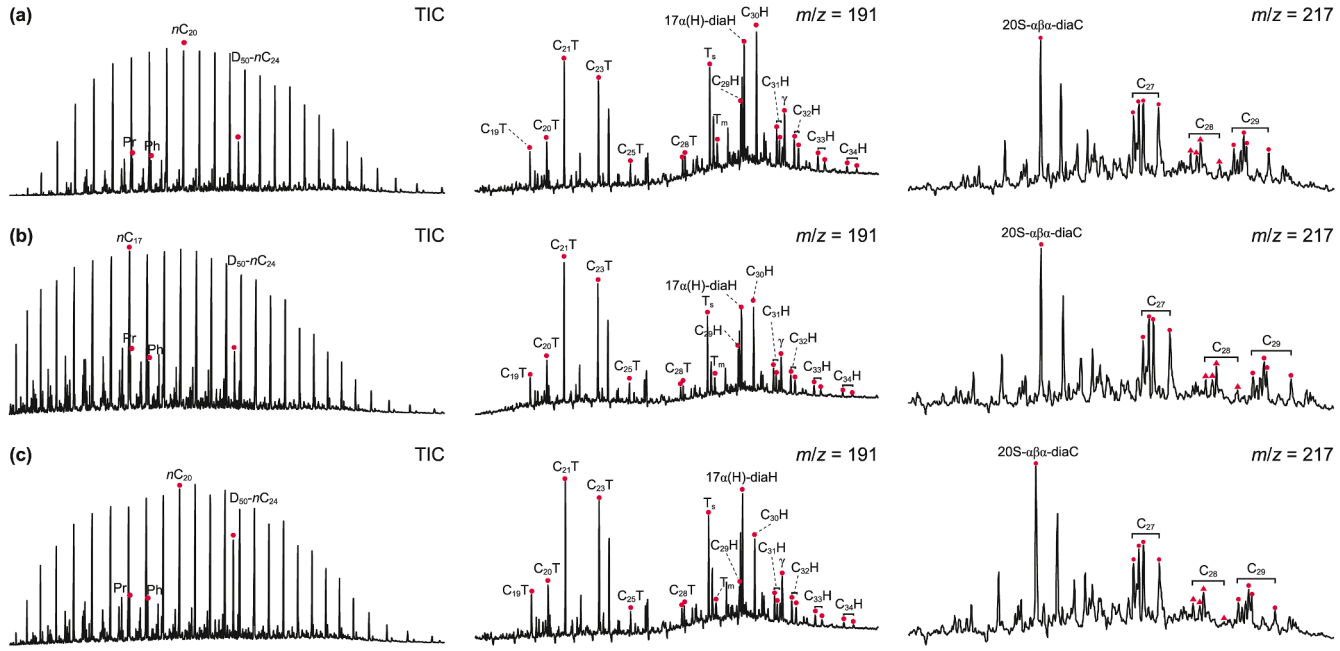

Fig. 7. Chromatograms (from left to right: TIC, *m/z* 191 and *m/z* 217) of saturated fractions from medium-maturity shales in the Qingshankou Formation obtained by routine Soxhlet extraction and stepwise extraction. **(a)** routine Soxhlet extraction, samples of 100–200 mesh, extracted by DCM; **(b)** the 1st step extracts (EOM A5) of stepwise extraction; **(c)** the 2nd extracts (EOM B5) of stepwise extraction. Pr: Pristane; Ph: Phytane; T: tricyclic terpane; H: hopane; γ: gammacerane; 17α(H)-diaH: 17α(H)-diahopane; C_{27}: C_{27} regular sterane; C_{28}: C_{28} regular sterane; C_{29}: C_{29} regular sterane; 20S-αβα-diaC: 20S-αβα-diacholestane.

Table 4
Comparison of biomarker parameters in medium-maturity shales from the Qingshankou Fm., Songliao Basin: routine Soxhlet extraction vs. stepwise extraction.

Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Routine Soxhlet extraction	1.12	0.20	0.21	0.95	65.8	8.8	25.4	0.52	0.44	0.59	0.55	2.24	1.01	0.29	0.21	10.12	0.77	0.27
Stepwise extraction (1st step)	1.14	0.13	0.16	0.45	58.4	10.8	30.8	0.53	0.43	0.62	0.56	3.04	0.84	0.23	0.22	9.46	0.71	0.28
Stepwise extraction (2nd step)	1.11	0.21	0.22	0.95	68.4	3.3	28.3	0.52	0.42	0.73	0.55	2.38	0.92	0.28	0.20	11.11	0.76	0.28

Table 5
Definitions of the biomarker parameters in Table 4.

Number	Parameters	Definitions
1	CPI	1/2[(C_{25}+C_{27}+C_{29}+C_{31}+C_{33})/(C_{24}+C_{26}+C_{28}+C_{30}+C_{32})+(C_{25}+C_{27}+C_{29}+C_{31}+C_{33})/(C_{26}+C_{28}+C_{30}+C_{32}+C_{34})]
2	Pr/nC_{17}	Pristane/nC_{17}
3	Ph/nC_{18}	Phytane/nC_{18}
4	Pr/Ph	Pristane/Phytane
5	%C_{27}steranes	C_{27}/(C_{27}+C_{28}+C_{29})*100
6	%C_{28}steranes	C_{28}/(C_{27}+C_{28}+C_{29})*100
7	%C_{29}steranes	C_{29}/(C_{27}+C_{28}+C_{29})*100
8	C_{29} sterane ββ/(αα+ββ)	C_{29} sterane αββ/(C_{29} sterane αββ+C_{29} sterane ααα)
9	C_{29} sterane 20S/(20S + 20R)	C_{29} sterane 20S/(C_{29} sterane 20S + C_{29} sterane 20R)
10	G/C_{30}	Gammaceranes/C_{30} hopane
11	C_{31} homohopane 22S/(22S + 22R)	C_{31} homohopane 22S/(C_{31} homohopane 22S + C_{31} homohopane 22R)
12	T_s/T_m	C_{27} 18α-22, 29, 30-Trisnorneohopane/C_{27} 17α-22, 29, 30-Trisnorhopane
13	(C_{19}T + C_{20}T)/C_{23}T	(C_{19} tricyclic terpane + C_{20} tricyclic terpane)/C_{23} tricyclic terpane
14	C_{19}T/(C_{19}T + C_{23}T)	C_{19} tricyclic terpane/(C_{19} tricyclic terpane + C_{23} tricyclic terpane)
15	C_{24}Te/(C_{24}Te + C_{23}T)	C_{24} tetracyclic terpane/(C_{24} tetracyclic terpane + C_{23} tricyclic terpane)
16	C_{26}T/C_{25}T	C_{26} tricyclic terpane/C_{25} tricyclic terpane
17	C_{24}T/C_{23}T	C_{24} tricyclic terpane/C_{23} tricyclic terpane
18	C_{22}T/C_{21}T	C_{22} tricyclic terpane/C_{21} tricyclic terpane

terpane) for routine Soxhlet extracts are 1.01, 0.29 and 0.21, respectively, while the corresponding values for the stepwise extracts are 0.92, 0.28 and 0.20. Additionally, other parameters such as C_{26}/C_{25} tricyclic terpane, C_{24}/C_{23} tricyclic terpane and C_{22}/C_{21} tricyclic terpane are comparable across both extraction methods as well (Tables 4 and 5).

In the hopane series, C_{30}H is the most abundant compound in the routine Soxhlet extracts, followed by 17α(H)-diahopane (Fig. 7(a)). In constrast, the stepwise extracts are predominantly composed of 17α(H)-diahopane, which is present in higher quantities than C_{30}H (Fig. 7(c)). Notably, in both routine Soxhlet and stepwise extracts, the content of homohopanes decreases

progressively from C_{31} to C_{34}. The biomarker parameters used for maturity evaluation obtained from both extraction methods are also consistent. For instance, the C_{31} homohopane 22S/(22S + 22R) values for both methods are 0.55, while the T_s/T_m ratios are 2.24 and 2.38, respectively. The ratios of gammacerane to $C_{30}H$ (gammacerane index, $\gamma/C_{30}H$) obtained through the two methods exhibit slight variation, with values of 0.59 for routine Soxhlet extraction and 0.73 for stepwise extraction (Tables 4 and 5).

(3) Steranes

The ion chromatograms of *m/z* 217, obtained from both routine Soxhlet extraction and stepwise extraction, display remarkably similar sterane distributions. Notably, 20S-αβα-diacholestane and 20R-αβα-diacholestane are the most abundant components (Fig. 7). The regular sterane assemblages in both extracts show a clear predominance of C_{27} components, following a distinct $C_{27} > C_{29} > C_{28}$ trend (Fig. 7). Quantitatively, the average percentages of C_{27}, C_{28}, and C_{29} regular steranes determined by the routine Soxhlet extraction method are 65.8%, 8.8% and 25.4%, respectively, while the corresponding values from stepwise extraction are 68.4%, 3.3%, and 28.3% (Tables 4 and 5). Furthermore, the sterane isomerization index from both extraction methods are highly consistent. The C_{29} sterane ββ/(αα+ββ) ratios are 0.52 for both extractions, while the C_{29} sterane 20S/(20S + 20R) ratios are 0.44 for routine Soxhlet extracts and 0.42 for stepwise extraction, indicating comparable thermal maturity signals (Tables 4 and 5).

Through the aforementioned analysis and comparison, it can be demonstrated that the SARA fractions of shales obtained via stepwise extraction exhibit similarities to those derived from routine Soxhlet extraction. The GC-MS chromatograms of both extraction methods reveal a high degree of concordance, and the calculated ratios of biomarker parameters are remarkably consistent. Consequently, the proposed stepwise extraction method (utilizing samples sized 10–20 mesh, extracted by Hex + DCM) is considered validated for biomarker detection and holds potential for extending its application to highly mature shales. However, it is noteworthy that the biomarkers obtained by stepwise extraction protocol are likely predominantly originate from organic matter adsorbed on both the kerogen skeleton and mineral surfaces (Qian et al., 2017; Li et al., 2021). Compared with free-phase biomarkers, these adsorbed-phase biomarkers exhibit a lower chemical maturity signal (Schwark et al., 1997). However, they still provide indispensable critical evidence in high-maturity systems where free-phase biomarkers are substantially depleted.

4. Implication

4.1. Origin of organic matter

The distribution of sterane and terpane biomarkers, along with the derived molecular geochemical proxies obtained through stepwise extraction from high-maturity shales in the Qingshankou Fm., Gulong Sag, are presented in Fig. 6 and Table 6. The shale samples contain high abundance of tricyclic terpanes, with homologues detected from C_{19} to at least C_{29} (Fig. 6). The low C_{22}/C_{21} tricyclic terpane ratio (0.22) and low to medium $(C_{19}+C_{20})/C_{23}$ tricyclic terpane ratio (0.49) imply a major contribution of aquatic organic matter and minimal terrestrial input, which is consistent with the relatively low C_{24} tetracyclic terpane/(C_{24} tetracyclic terpane + C_{23} tricyclic terpane) ratio (Table 6). This is because short-chain tricyclic terpanes (C_{19} and C_{20}) are typically derived from higher plant fragments, particularly diterpenoids formed in vascular plants, whereas C_{23} tricyclic terpane is usually associated with aquatic organisms (Peters et al., 2005; Volk et al., 2005; Zhang J. et al., 2021). Furthermore, relative abundance of regular steranes is another useful indicator for biological origins of organic matter. The ternary diagram of ααα20RC_{27}-ααα20RC_{28}-ααα20RC_{29} steranes shows a higher abundance of C_{27} steranes (43%), followed by C_{29} (30%) and C_{28} (27%) regular steranes (Fig. 8), which is interpreted as a dominant contribution from planktonic algae and bacterial organic matter, with a low contribution of terrestrial organic matter in the shales in the Gulong Sag (Seifert and Moldowan, 1978; Zhang J. et al., 2021).

It should be noted that most biomarkers used in petroleum geochemistry, including aromatic steroid hydrocarbons, are strongly influenced by thermal maturity (Huang et al., 2022). As a result, some source-related biomarker parameters may become unreliable or misleading in highly mature samples. Therefore, provenance input and sedimentary facies were considered as supporting evidence. We further compared regular sterane data of the Qingshankou shales in the Gulong Sag with equivalent data from other structural units in northern Songliao Basin. As shown in

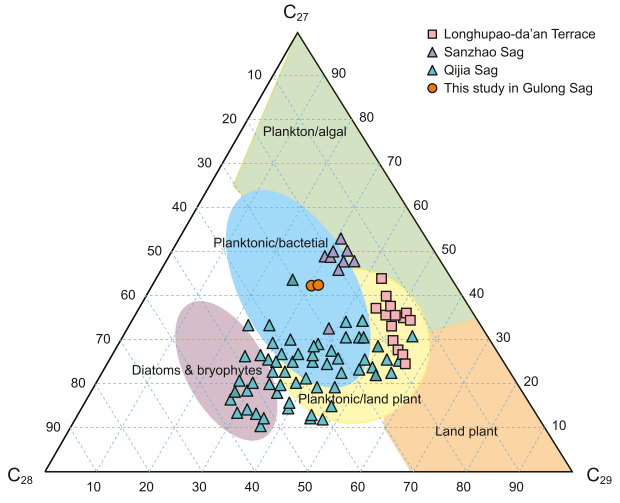

Fig. 8. Ternary diagram of regular steranes (C_{27}, C_{28}, and C_{29}) indicating organic matter sources of the shales in the Qingshankou Fm. from different structural units, northern Songliao Basin.

Table 6
Biomarker ratios of saturated hydrocarbons in high-maturity shales from the Qingshankou Fm., Gulong Sag, Songliao Basin.

Sample	$C_{22}T/C_{21}T$	$(C_{19}T + C_{20}T)/C_{23}T$	$C_{24}T/C_{23}T$	$C_{24}Te/(C_{24}Te + C_{23}T)$	C_{27} Ster.%	C_{28} Ster.%	C_{29} Ster.%	Pr/Ph	$G/C_{30}H$
A	0.22	0.49	0.54	0.31	0.43	0.27	0.30	0.64	0.16
B	0.20	0.53	0.55	0.31	0.43	0.26	0.31	0.73	0.17

T: tricyclic terpane; Te: tetracyclic terpane; Ster.: regular sterane; G: gammacerane; H: hopane; Pr: pristine; Ph: phytane.

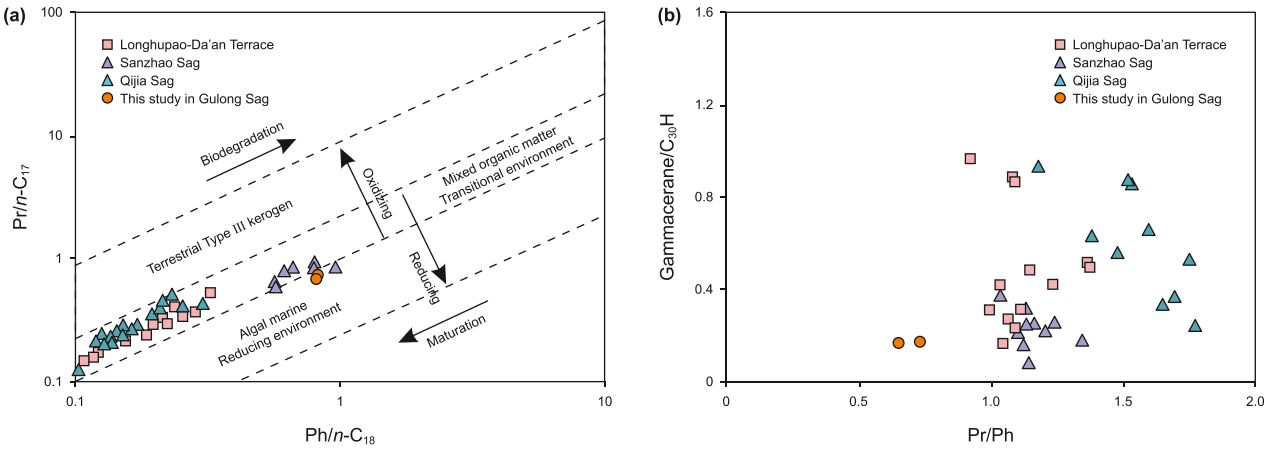

Fig. 9. Indicators of depositional environment showing the water properties of the Qingshankou shales in different structural units, northern Songliao Basin. **(a)** Phytane to n-C_{18} alkane (Ph/n-C_{18}) vs. pristane to n-C_{17} alkane (Pr/n-C_{17}) (modified from Shalaby et al. (2012)); **(b)** Pristane to phytane (Pr/Ph) vs. gammacerane/C_{30} hopane.

Fig. 8, the shales in the Qingshankou Formation, Gulong Sag contain higher contents of planktonic and bacterial organic matter along with lower terrestrial plant input compared with those from the Qijia Sag and Longhupao-Da'an Terrace. Such results align with the sediment provenance distribution in the northern Songliao Basin. The Gulong Sag, located closer to the lake center, is more distant from terrestrial source areas, whereas the Qijia Sag and Longhupao-Da'an Terrace received more terrestrial input from the northern and western provenances, respectively (Wang et al., 2023).

4.2. Depositional environment

The Pr/Ph ratio serves as a redox indicator of paleoenvironmental conditions during organic matter deposition (Didyk et al., 1978; Ten Haven et al., 1987, 1988). Reducing conditions in sediments promote cleavage of the phytyl side chain to yield phytol and ultimately phytane, whereas oxic conditions favor the conversion of phytol to pristine on the contrary (Peters et al., 2005). Therefore, Didyk et al. (1978) interpreted that Pr/Ph < 1 suggests anoxic source rock deposition, while Pr/Ph > 1 indicates oxic deposition. The Pr/Ph ratio of the Qingshankou shales in Gulong Sag ranges from 0.64 to 0.73 (Table 6), notably lower compared to the averages in Sanzhao Sag (1.16), Qijia Sag (1.46) and Longhupao-Da'an Terrace (1.08) (Fig. 9(a) and (b)). This suggests a reducing depositional environment with algal Type I kerogen enrichment in the Qingshankou shales of the Gulong Sag.

Gammacerane, a biomarker indicative of hypersaline and stratified water columns (Moldowan et al., 1985; Sinninghe Damsté et al., 1995; Holba et al., 2003), is found in low abundance in the shales of the Qingshankou Fm., Gulong Sag. The gammacerane index of the shales in the Gulong Sag ranges from 0.16 to 0.17 (Table 6), significantly lower than the averages observed in Qijia Sag (0.59) and Longhupao-Da'an Terrace (0.51), indicating a fresh to brackish water depositional environment with unstable water column stratification.

5. Conclusion

To resolve the challenges of biomarker identification in high-maturity shales, a stepwise extraction approach combined with GC-MS analysis was developed, employing organic solvents of varying polarity and rock samples of different size. This pretreatment enabled the successful detection of sterane and terpane biomarkers in shales of the Qingshankou Fm., Gulong Sag, Songliao Basin. Furthermore, biomarker parameters are applied to infer the source of organic matter and depositional environment of the organic-rich shales, providing insights into their geological significance. The key conclusions are as follows:

(1) In the high-maturity (R_o = 1.42%), organic-rich shales of the Qingshankou Fm. in the Gulong Sag, Songliao Basin, routine Soxhlet extraction combined with GC-MS analysis failed to identify sterane and terpane biomarkers, even after urea adduction process. In contrast, the stepwise extraction effectively removed interfering compounds from saturated hydrocarbons, enabling clear biomarker identification. Optimal results were achieved using samples of 10–20 mesh, sequentially extracted with n-hexane followed by DCM.

(2) The proposed stepwise extraction method was validated using a medium-maturity shale (R_o = 1.0%) from the Qingshankou Fm. in the Sanzhao Sag, Songliao Basin. The results showed a high degree of concordance in SARA fractions, GC-MS chromatograms, and biomarker parameters between the stepwise and routine Soxhlet extraction. This agreement confirms the reliability of the proposed approach for biomarker detection and holds its potential for further application to highly mature shales.

(3) The diagnostic biomarker parameters obtained through our stepwise extraction method provide key geological insights. They reveal that the organic matter in the shales of the Qingshankou Fm., Gulong Sag was primarily derived from aquatic sources (plankton, algae and bacteria), with limited terrestrial input. Furthermore, the data suggest a depositional environment characterized by fresh to brackish water, reducing conditions and an unstable water column stratification.

CRediT authorship contribution statement

Jing-Ya Zhang: Writing – review & editing, Writing – original draft, Investigation, Funding acquisition, Formal analysis. **Bo-Wen Meng:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis. **Ru-Kai Zhu:** Investigation, Funding acquisition, Conceptualization. **Hua-Sen Zeng:** Investigation, Formal analysis. **Chang Liu:** Funding acquisition, Formal analysis.

Declaration of competing interest

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.

Acknowledgement

This study is funded by the National Natural Science Foundation of China (Project No. 42202152, U22B6004 and 42302158), Young Elite Scientists Sponsorship Program by China Association for Science and Technology (CAST) (Project No. 2023QNRC001) and Heilongjiang Province open competition projects "Research on the diagenetic dynamic evolution process and its coupling relationship with pores and fractures" (Project No. 2021ZXJ01A09). We sincerely thank Daqing Oilfield Academician Workstation and State Key Laboratory of Continental Shale Oil for providing the core samples.

References

Bi, H., Li, P., Jiang, Y., Fan, J.J., Chen, X.Y., 2021. Effective source rock selection and oil–source correlation in the Western slope of the northern Songliao Basin, China. Pet. Sci. 18, 398–415. https://doi.org/10.1007/s12182-021-00554-1.

Czochanska, Z., Gilbert, T.D., Philp, R.P., Sheppard, C.M., Weston, R.J., Wood, T.A., Woolhouse, A.D., 1988. Geochemical application of sterane and triterpane biomarkers to a description of oils from the Taranaki Basin in New Zealand. Org. Geochem. 12, 123–135. https://doi.org/10.1016/0146-6380(88)90249-5.

Didyk, B.M., Simoneit, B.R.T., Brassell, S.C., Eglinton, G., 1978. Organic geochemical indicators of palaeoenvironmental conditions of sedimentation. Nature 272, 216–222. https://doi.org/10.1038/272216a0.

Fang, R.H., Littke, R., Zieger, L., Baniasad, A., Li, M.J., Schwarzbauer, J., 2019. Changes of composition and content of tricyclic terpane, hopane, sterane, and aromatic biomarkers throughout the oil window: a detailed study on maturity parameters of Lower Toarcian Posidonia Shale of the Hils Syncline, NW Germany. Org. Geochem. 138, 103928. https://doi.org/10.1016/j.orggeochem.2019.103928.

Farrimond, P., Taylor, A., Telns, N., 1998. Biomarker maturity parameters: the role of generation and thermal degradation. Org. Geochem. 29 (5–7), 1181–1197. https://doi.org/10.1016/S0146-6380(98)00079-5.

Feng, Z.H., Fang, W., Li, Z.G., Wang, X., Huo, Q.L., Huang, C.Y., Zhang, J.H., Zeng, H.S., 2011. Depositional environment of terrestrial petroleum source rocks and geochemical indicators in the Songliao Basin. Sci. China (Earth Sciences) 54, 1304–1317. https://doi.org/10.1007/s11430-011-4268-0.

Feng, Z.Q., Jia, C.Z., Xie, X.N., Zhang, S., Feng, Z.H., Cross, T.A., 2010. Tectonostratigraphic units and stratigraphic sequences of the nonmarine Songliao basin, northeast China. Basin Res. 22, 79–95. https://doi.org/10.1111/j.1365-2117.2009.00445.x.

Gorynski, K.E., Newhart, R.E., Tobey, M.H., Smagala, Thomas M., Enriquez, D.A., Dreger, J.L., 2019. Quantification and characterization of hydrocarbon-filled porosity in oil-rich shales using integrated thermal extraction, pyrolysis, and solvent extraction. AAPG Bull. 103, 723–744. https://doi.org/10.1306/08161817214.

Grice, K., de Mesmay, R., Glucina, A., Wang, S., 2008. An improved and rapid 5A molecular sieve method for gas chromatography isotope ratio mass spectrometry of *n*-alkanes (C_8–C_{30+}). Org. Geochem. 39, 284–288. https://doi.org/10.1016/j.orggeochem.2007.12.009.

Hackley, P.C., Walters, C.C., Kelemen, S.R., Mastalerz, M., Lowers, H.A., 2017. Organic petrology and micro-spectroscopy of Tasmanites microfossils: Applications to kerogen transformations in the early oil window. Org. Geochem. 114, 23–44. https://doi.org/10.1016/j.orggeochem.2017.09.002.

Han, Y., Horsfield, B., Curry, D.J., 2017. Control of facies, maturation and primary migration on biomarkers in the Barnett Shale sequence in the Marathon 1 Mesquite well, Texas. Mar. Petrol. Geol. 85, 106–116. https://doi.org/10.1016/j.marpetgeo.2017.04.018.

Holba, A.G., Dzou, L.I., Wood, G.D., Ellis, L., Adam, P., Schaeffer, P., Albrecht, P., Greene, T., Hughes, W.B., 2003. Application of tetracyclic polyprenoids as indicators of input from fresh–brackish water environments. Org. Geochem. 34, 441–469. https://doi.org/10.1016/S0146-6380(02)00193-6.

Huang, H.P., di Primio, R., Pedersen, J.H., Silva, R., Algeer, R., Ma, J.P., Larter, S., 2022. On the determination of oil charge history and the practical application of molecular maturity markers. Mar. Petrol. Geol. 139, 105586. https://doi.org/10.1016/j.marpetgeo.2022.105586.

Huang, Y.S., Geng, A.S., Fu, J.M., Sheng, G.Y., Zhao, B.Q., Cheng, Y.X., Li, M.F., 1992. The investigation of characteristics of biomarker assemblages and their precursors in Damintun ultra-high wax oils and related source rocks. Org. Geochem. 19, 29–39. https://doi.org/10.1016/0146-6380(92)90025-S.

Hwang, R.J., 1990. Biomarker analysis using GC-MSD. J. Chromatogr. Sci. 28, 109–113. https://doi.org/10.1093/chromsci/28.3.109.

Hunt, J.M., Philp, R.P., Kvenvolden, K.A., 2002. Early developments in petroleum geochemistry. Org. Geochem. 33 (9), 1025–1052. https://doi.org/10.1016/S0146-6380(02)00056-6.

Kikuchi, T., Suzuki, N., Saito, H., 2010. Change in hydrogen isotope composition of *n*-alkanes, pristane, phytane, and aromatic hydrocarbons in Miocene siliceous mudstones with increasing maturity. Org. Geochem. 41, 940–946. https://doi.org/10.1016/j.orggeochem.2010.05.004.

Li, S.M., Pang, X.Q., Liu, K.Y., Jin, Z.J., 2006. Origin of the high waxy oils in Bohai Bay Basin, east China: Insight from geochemical and fluid inclusion analyses. J. Geochem. Explor. 89, 218–221. https://doi.org/10.1016/j.gexplo.2005.11.045.

Li, S.T., Li, S.X., Guo, R.L., Zhou, X.P., Wang, Y., Chen, J.L., Zhang, J.Q., Hao, L.W., Ma, X.F., Qiu, J.L., 2021. Occurrence state of soluble organic matter in shale oil reservoirs from the Upper Triassic Yanchang formation in the ordos Basin, China: insights from multipolarity sequential extraction. Nat. Resour. Res. 30, 4379–4402. https://doi.org/10.1007/s11053-021-09959-6.

Lin, R., Davis, A., 1988. A fluorogeochemical model for coal macerals. Org. Geochem. 12, 363–374. https://doi.org/10.1016/0146-6380(88)90010-1.

Lin, R., Ritz, G.P., 1993. Studying individual macerals using i.r. microspectrometry, and implications on oil versus gas/condensate proneness and "low-rank" generation. Org. Geochem. 20, 695–706. https://doi.org/10.1016/0146-6380(93)90055-G.

Liu, B., Wang, Y., Tian, S., Guo, Y., Wang, L., Yasin, Q., Yang, J., 2022. Impact of thermal maturity on the diagenesis and porosity of lacustrine oil-prone shales: insights from natural shale samples with thermal maturation in the oil generation window. Int. J. Coal Geol. 261, 104079. https://doi.org/10.1016/j.coal.2022.104079.

Liu, C., Xu, X., Liu, K., 2020. Pore-scale oil distribution in shales of the Qingshankou formation in the Changling Sag, Songliao Basin, NE China. Mar. Petrol. Geol. 120, 104553. https://doi.org/10.1016/j.marpetgeo.2020.104553.

Lü, H., Chen, Z.L., Wang, Z., Tang, H.S., 2008. Distribution of high molecular weight hydrocarbons and genesis of high wax content oil from the south slope zone of the Dongying Sag, the Jiyang Depression. Oil Gas Geol. 29, 355–360 (in Chinese).

Michalczyk, G., 1985. Determination of n- and iso-paraffins in hydrocarbon waxes – comparative results of analyses by gas chromatography, urea adduction, and molecular sieve adsorption. Fette Seifen Anstrichm. 87, 481–486.

Moldowan, J.M., Seifert, W.K., Gallegos, E.J., 1985. Relationship between petroleum composition and depositional environment of petroleum source rocks. AAPG Bull. 69, 1255–1268. https://doi.org/10.1306/AD462BC8-16F7-11D7-8645000102C1865D.

Peters, K.E., Moldowan, J.M., 1993. The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments. Prentice Hall, Englewood Cliffs, New Jersey, USA.

Peters, K.E., Walters, C.C., Moldowan, J.M., 2005. The Biomarker Guide: Biomarkers and Isotopes in Petroleum Systems and Earth History, second ed., vol. 2. Cambridge University Press, Cambridge.

Philp, R.P., Lewis, C.A., 1987. Organic geochemistry of biomarkers. Annu. Rev. Earth Planet Sci. 15, 363–395. https://doi.org/10.1146/annurev.ea.15.050187.002051.

Philp, R.P., 2000. Geochemical analysis: gas chromatography and GC-MS. In: Wilson, I.D. (Ed.), Encyclopedia of Separation Science. Academic Press, Oxford.

Price, L.C., Clayton, J.L., 1992. Extraction of whole versus ground source rocks: Fundamental petroleum geochemical implications including oil-source rock correlation. Geochem. Cosmochim. Acta 56, 1213–1222. https://doi.org/10.1016/0016-7037(92)90057-P.

Qian, M.H., Jiang, Q.G., Li, M.W., Li, Z.M., Liu, P., Ma, Y.Y., Cao, T.T., 2017. Quantitative characterization of extractable organic matter in lacustrine shale with different occurrences. Petrol. Geol. Exp. 39 (2), 278–286. https://doi.org/10.11781/sysydz201702278 (in Chinese).

Sajgó, C.S., Maxwell, J.R., Mackenzie, A.S., 1983. Evaluation of fractionation effects during the early stages of primary migration. Org. Geochem. 5, 65–73. https://doi.org/10.1016/0146-6380(83)90004-9.

Schwark, L., Stoddart, D., Keuser, C., Spitthoff, B., Leythaeuser, D., 1997. A novel sequential extraction system for whole core plug extraction in a solvent flow-through cell — application to extraction of residual petroleum from an intact pore-system in secondary migration studies. Org. Geochem. 26 (1–2), 19–31. https://doi.org/10.1016/S0146-6380(96)00163-5.

Seifert, W.K., Moldowan, M.J., 1978. Applications of steranes, terpanes and monoaromatics to the maturation, migration and source of crude oils. Geochem. Cosmochim. Acta 42, 77–95. https://doi.org/10.1016/0016-7037(78)90219-3.

Shalaby, M.R., Hakimi, M.H., Abdullah, W.H., 2012. Organic geochemical characteristics and interpreted depositional environment of the Khatatba Formation, northern Western Desert, Egypt. AAPG Bull. 96, 2019–2036. https://doi.org/10.1306/04181211178.

Sinninghe Damsté, J.S., Kenig, F., Koopmans, M.P., Koster, J., Schouten, S., Hayes, J.M., De Leeuw, J.W., 1995. Evidence for gammacerane as an indicator of water column stratification. Geochem. Cosmochim. Acta 59, 1895–1900. https://doi.org/10.1016/0016-7037(95)00073-9.

Sun, L.D., Liu, H., He, W.Y., Zhang, S.C., Zhu, R.K., Jin, X., Meng, S.W., Jiang, H., 2021. An analysis of major scientific problems and research paths of Gulong shale oil in Daqing Oilfield, NE China. Petrol. Explor. Dev. 48 (3), 527–540. https://doi.org/10.1016/S1876-3804(21)60043-5.

Ten Haven, H.L., De Leeuw, J.W., Rullkötter, J., Damsté, J.S., 1987. Restricted utility of the pristane/phytane ratio as a palaeoenvironmental indicator. Nature 330, 641–643. https://doi.org/10.1038/330641a0.

Ten Haven, H.L., Rullkötter, J., De Leeuw, J.W., Damsté, J.S., 1988. Pristane/phytane ratio as environmental indicator. Nature 333, 604. https://doi.org/10.1038/333604b0.

Theuerkorn, K., Horsfield, B., Wilkes, H., di Primio, R., Lehne, E., 2008. A reproducible and linear method for separating asphaltenes from crude oil. Org. Geochem. 39 (8), 929–934. https://doi.org/10.1016/j.orggeochem.2008.02.009.

Tissot, B.P., Welte, D.H., 2013. Petroleum Formation and Occurrence, second ed. Springer Science & Business Media.

Treibs, A., 1934. The occurrence of chlorophyll derivatives in an oil shale of the upper Triassic. Justus Liebigs Ann. Chem. 517, 103–114.

Vinnichenko, G., Jarrett, A.J., van Maldegem, L.M., Brocks, J.J., 2021. Substantial maturity influence on carbon and hydrogen isotopic composition of *n*-alkanes in sedimentary rocks. Org. Geochem. 152, 104171. https://doi.org/10.1016/j.orggeochem.2020.104171.

Volk, H., George, S.C., Middleton, H., Schofield, S., 2005. Geochemical comparison of fluid inclusion and present-day oil accumulations in the Papuan Foreland - evidence for previously unrecognized petroleum source rocks. Org. Geochem. 36, 29–51. https://doi.org/10.1016/j.orggeochem.2004.07.018.

Wang, X.J., Bai, X.F., Lu, J.M., Jin, Z.J., Wang, G.W., Kuang, L.C., Li, J.H., Li, J.S., Zhang, J.Y., Sun, L.D., Wu, J.F., Sun, H.L., 2023. New fields, new types and resource potentials of oil-gas exploration in northern Songliao Basin. Acta Pet. Sin. 44, 2091–2178 (in Chinese).

Wilhelms, A., Horstad, I., Karlsen, D.A., 1996. Sequential extraction—a useful tool for reservoir geochemistry? Org. Geochem. 24, 1157–1172. https://doi.org/10.1016/S0146-6380(96)00100-3.

Xiao, F., Yang, J.G., Li, S.C., Yao, Y.L., Huang, Y.M., Gao, X.Y., 2024. Enrichment and movability of lacustrine tight shale oil for the first member of the Upper Cretaceous Qingshankou Formation in the Sanzhao Sag, Songliao Basin, NE China: Insights from saturated hydrocarbon molecules. Fuel 368, 131615. https://doi.org/10.1016/j.fuel.2024.131615.

Xu, S., Sun, Y., 2005. An improved method for the micro-separation of straight chain and branched/cyclic alkanes: urea inclusion paper layer chromatography. Org. Geochem. 36 (9), 1334–1338. https://doi.org/10.1016/j.orggeochem.2005.04.003.

Zhang, H., Huang, H.P., Li, Z., Liu, M., 2019. Oil physical status in lacustrine shale reservoirs – a case study on Eocene Shahejie Formation shales, Dongying depression, East China. Fuel 257, 116591. https://doi.org/10.1016/j.fuel.2019.116027.

Zhang, J.Y., Sun, M.L., Liu, G.D., Cao, Z., Kong, Y.H., 2021. Geochemical characteristics, hydrocarbon potential, and depositional environment evolution of fine-grained mixed source rocks in the Permian Lucaogou Formation, Jimusaer Sag, Junggar Basin. Energy & Fuels 35, 264–282. https://doi.org/10.1021/acs.energyfuels.0c02500.

Zhang, J.Y., Zhu, R.K., Wu, S.T., Jiang, X.H., Liu, C., Cai, Y., Zhang, S.R., 2023. Microscopic oil occurrence in high-maturity lacustrine shales: Qingshankou Formation, Gulong Sag, Songliao Basin. Pet. Sci. 20, 2726–2746. https://doi.org/10.1016/j.petsci.2023.08.026.

Zhang, P.L., Misch, D., Hu, F., Kostoglou, N., Sachsenhofer, R.F., Liu, Z.J., Meng, Q.T., Bechtel, A., 2021. Porosity evolution in organic matter-rich shales (Qingshankou Fm.; Songliao Basin, NE China): Implications for shale oil retention. Mar. Petrol. Geol. 130, 105139. https://doi.org/10.1016/j.marpetgeo.2021.105139.

Zhang, Y.D., Sun, Y.G., Chen, J.P., 2020. Stable carbon isotope evidence for the origin of C_{28} steranes in lacustrine source rocks from the Qikou Sag, Bohai Bay Basin, Eastern China. Org. Geochem. 145, 104028. https://doi.org/10.1016/j.orggeochem.2020.104028.