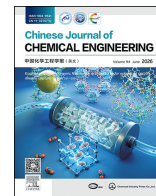




Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Full Length Article

High-resolution mass spectrometry-informed molecular reconstruction of vacuum gas oil

Dong Guan [#], Xiangyu Zhao [#], Shuofan Li, Zhengyu Chen ^{*}, Quan Shi, Chunming Xu, Linzhou Zhang ^{*}

State Key Laboratory of Heavy Oil Processing, Petroleum Molecular Engineering Center, China University of Petroleum, Beijing 102249, China

ARTICLE INFO

Article history:

Received 27 September 2025

Received in revised form

18 December 2025

Accepted 19 December 2025

Available online 13 February 2026

Keywords:

Molecular reconstruction

High-resolution mass spectrometry

Vacuum gas oil

Composition similarity

ABSTRACT

Accurate quantification of the molecular composition is essential for precisely modeling the petroleum refining process. Conventional molecular reconstruction methods, which primarily rely on bulk properties, fail to accurately predict molecular distributions due to the lack of incorporating molecular characterization information. This study presents a novel methodology for reconstructing the molecular composition of vacuum gas oil using high-resolution mass spectrometry (HRMS) data as a reference. A dual-objective optimization strategy was developed to simultaneously align bulk properties and composition similarities. The results demonstrate that integrating HRMS data significantly enhances the accuracy of molecular distribution prediction. The predicted distributions of double bond equivalents and carbon numbers for different classes of species are in good agreement with experimental observations. The method was also applied to various vacuum gas oils, and the robustness and reliability of this method are validated.

© 2026 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

Vacuum gas oil (VGO) is involved in various refining processes, such as crude oil distillation, fluid catalytic cracking (FCC), hydrocracking, etc. Process simulation is of vital importance for the design and optimization of the VGO upgrading processes. Traditional process models, such as lumped kinetic models, lack the ability to predict molecular distributions and conversion pathways, leading to a low precision of product quality prediction. Molecules are the basic units of reaction and separation, and molecular-level modeling is a promising technique for VGO refining process optimization [1].

The molecular-level process model primarily consists of molecular compositional models, molecular-level kinetic models, and molecular-level separation models. The molecular compositional model serves as the foundation for molecular-level process model development, providing input containing feedstock molecular type, structure, and quantity information. In the early stages, it was challenging to obtain the molecular content of complex

molecular systems due to the limitations of analytical techniques. The averaged molecular structures were frequently employed to represent petroleum molecules [2,3]. Subsequently, the petroleum molecular reconstruction method was proposed to calculate the molecular composition of petroleum fractions based on bulk properties. In this field, Klein's research group has done a lot of pioneering work. For example, Neurock and Klein *et al.* [4–7] proposed a molecular compositional reconstruction method utilizing elemental and ¹H NMR data as input. In their study, the bond-electron matrix (BEM) and probability density functions (PDF) were employed to represent molecular structures and distributions, respectively. Deniz *et al.* [8,9] introduced artificial intelligence methods, such as artificial neural networks, in the construction of the compositional model, which effectively promoted the improvement of model calculation efficiency. Simultaneously, Quann and Jaffe *et al.* [10–12] proposed structure-oriented lumping (SOL) to represent molecular structure, and the petroleum fractions were reconstructed at the structural group level. The SOL method has been widely applied to develop molecular-level process models for petroleum refining [13–19]. Additionally, Peng [20] developed the molecular type-homologous series (MTHS) method to represent petroleum composition in terms of a pseudo-component matrix. Hudebine *et al.* [21–23] developed the maximum information entropy method to fine-

^{*} Corresponding authors.

E-mail addresses: zychen@cup.edu.cn (Z. Chen), Lzz@cup.edu.cn (L. Zhang).

[#] These authors contributed equally to this work.

tune compositional distribution in the absence of detailed molecular-level characterization data. Recently, Zhang's research group [24] proposed a hybrid structural unit and bond electron matrix framework (SU-BEM), offering a flexible approach to representing digital petroleum molecules. Several compositional models have been developed for different petroleum fractions based on the SU-BEM framework [25,26]. To date, various molecular compositional reconstruction methods have been proposed. However, most of the reported molecular compositional models were built predominantly depending on bulk properties. This built compositional model commonly fails to reflect the actual molecular distribution for the complex hydrocarbon mixtures since detailed molecular compositional information was not provided.

With the progress of analytical technology, not only the bulk property data, but also molecular characterization data can be used for molecular reconstruction. The rapid development of molecular characterization technologies provides more detailed molecular information. For example, the molecular composition measured from gas chromatography coupled with a flame ionization detector (GC-FID) technology was incorporated into the compositional model of gasoline fractions [27,28]. The models can calculate both the bulk properties and molecular distribution accurately. Comprehensive two-dimensional gas chromatography (GC $\times$ GC) can provide the detailed molecular composition of diesel and VGO [29–31]. For the heavy petroleum fraction, high-resolution mass spectrometry (HRMS) and noncontact atomic force microscope (AFM) can be used to obtain the molecular composition and structure [32–35]. HRMS (e.g., FT-ICR MS and orbitrap MS) offer qualitative compositional information of heavy oil. Qian *et al.* [36,37] made pioneering work in the development of molecular characterization methods for heavy oil based on HRMS. Marshall and Rodgers *et al.* [38–43] proposed the concept of 'Petroleomic', which studies the relationship between the molecular composition with property and reaction activity. Subsequently, a series of methods for molecular characterization of heavy fractions based on HRMS were developed [44–46]. Among them, Shi's research group [47–50] achieved a breakthrough in the semi-quantitative characterization of heavy oil. Furthermore, in terms of structural analysis, Zhang *et al.* [34,51–57] directly observed molecular structures using the AFM.

When more and more heavy oil molecules were recognized, compositional models combining molecular characterization information have gradually attracted widespread research interest. Zhang *et al.* [58] developed a molecular compositional model for

heavy oil and compared it with HRMS data, demonstrating that the molecular distribution was consistent with the spectrometry data in terms of carbon number and double bond equivalents (DBE) range. Alvarez-Majmutov *et al.* [59] constructed the molecular compositional model of residue, and the model results conform to the molecular distribution of the HRMS data. Guan *et al.* [60] built a compositional model of heavy oil using HRMS data as validation, and the accuracy of molecular fraction distributions was improved. Although there have been some research advancements in combining HRMS data with molecular composition models, the molecular reconstruction method of heavy oil directly using the HRMS data as input has not been reported. It means that the accuracy of the molecular content in the model cannot be strictly guaranteed.

In our previous work, a novel molecular reconstruction method was developed, named petroleum molecular reconstruction based on property and composition similarities [61]. This method was applied to the diesel fraction, ensuring the simultaneous accuracy of bulk properties and molecular fraction distribution. However, compared to diesel, VGO exhibits a more complex molecular composition. HRMS can obtain the molecular-level characterization information of VGO. Thus, this work proposed an HRMS-informed molecular reconstruction method, and the molecular composition of four VGOs was reconstructed precisely. Simultaneously, the issue of multiple solutions in the traditional compositional model has also been resolved. The results demonstrate that the HRMS-informed molecular reconstruction method can improve both the accuracy and repeatability of the compositional models.

2. Molecular Reconstruction Modeling Approach

2.1. Method overview

To incorporate qualitative or semi-quantitative molecular characterization data into the molecular reconstruction method, we proposed the concept of composition similarity and developed a novel molecular reconstruction approach in our previous work [61]. According to the novel method, the qualitative HRMS data of VGO can also be used, as shown in Fig. 1. In this method, HRMS data were employed as input for composition similarity calculation in the molecular reconstruction process. At the same time, the property similarity can be calculated, which is similar to the previous studies [25]. Finally, both property and composition similarities were combined as optimization objectives in a dual-

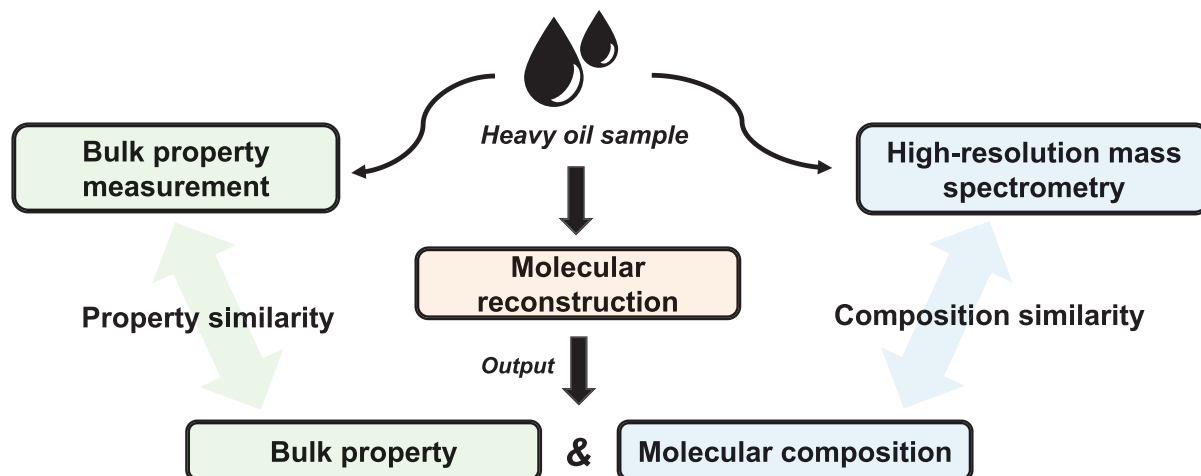


Fig. 1. Schematic diagram of molecular reconstruction based on HRMS.

Table 1

Analysis method of molecular composition for VGO fraction.

Molecular type	Chemical derivatization methods	Ionization methods
Saturated hydrocarbons	Redox	–ESI
Aromatic hydrocarbons	Sulfonation	–ESI
Sulfidic compounds/thiophenic compounds	Methylation/demethylation	+ESI
Basic nitrogen compounds	–	+ESI
Non-basic nitrogen compounds	–	–ESI
Acidic oxygen compounds	–	–ESI

objective optimization algorithm. The incorporation of HRMS data will effectively constrain molecular distribution, thereby enhancing the reliability and repeatability of VGO compositional models. Thus, the method was named HRMS-informed molecular reconstruction.

2.2. Experimental method

The VGO molecular composition data for various compound types were obtained through HRMS combined with different ionization and chemical derivatization methods. In our previous work, Li *et al.* [62] established a semi-quantitative analysis strategy for the molecular characterization of VGO fractions by an Orbitrap Fusion MS (Thermo Scientific, USA). In Li's work, the different types of compounds in VGO were characterized by electrospray ionization (ESI). Among them, polar compounds (basic nitrogen compounds, non-basic nitrogen compounds, phenols, and carboxylic acids) were injected directly into HRMS for characterization, while nonpolar/weakly polar compounds (saturated hydrocarbons, aromatic hydrocarbons, sulfides, and thiophenes) were pre-treated through chemical derivatization, as shown in Table 1. In this work, the VGO was also characterized by the same method, and details for sample preparation, ionization conditions, and characterization methods were summarized in the Supplementary Material. After the molecular characterization was completed, the obtained data would be employed for constructing a preset molecular library and optimizing molecular fractions. Since HRMS can only obtain the DBE and molecular formula, the

molecular structure was determined by the molecular structure image obtained via AFM.

In addition to the molecular characterization experiment, a series of bulk properties was also measured. Carbon, hydrogen, and oxygen contents were measured using a Vario EL Cube elemental analyzer (Elementar, Germany) following ASTM D5291-2002 (C and H) and ASTM D5622-1995 (O). Sulfur and nitrogen contents were determined using an EA 3100 elemental analyzer (Analytik Jena, Germany) in accordance with ASTM D5762-2002 and ASTM D5453-2004, respectively. The distillation curve was obtained by an Agilent 6890 N GC system (Agilent, USA) using the ASTM D2887EXT method.

2.3. Preset molecular library

The preset molecular library is a series of molecular structures defined according to molecular characterization data, providing specific molecular properties for bulk property calculations. It is a challenging task to preset a suitable molecular library. The molecular structures of VGO are extremely complex, with a wide distribution of carbon numbers and a vast number of isomers.

To address the challenge, this work proposed a method for constructing a preset molecular library of VGO based on HRMS data, as shown in Fig. 2 [24,25]. Generally, HRMS can detect approximately 8000–10000 molecules in VGO, and the molecular structure can be inferred by the obtained DBE and chemical formulas information. Thus, the DBE range and chemical formulas detected by HRMS were initially used to determine core molecular structures. Then, different side chains were added to the molecular core to generate a homologous series, followed by obtaining the preset molecular library. Compared to traditional preset molecular libraries based on virtual molecules [20,23], our proposed molecular library covered almost all detectable molecules. Moreover, representative isomer structures were added to the library, such as *n*-paraffins and *iso*-paraffins, and there are 12138 molecules in the VGO molecular library. The molecular types, carbon numbers, and DBE ranges of the VGO preset molecular library are depicted in Fig. 2. Detailed molecular information of the preset molecular library can be found in Supplementary Material 1.

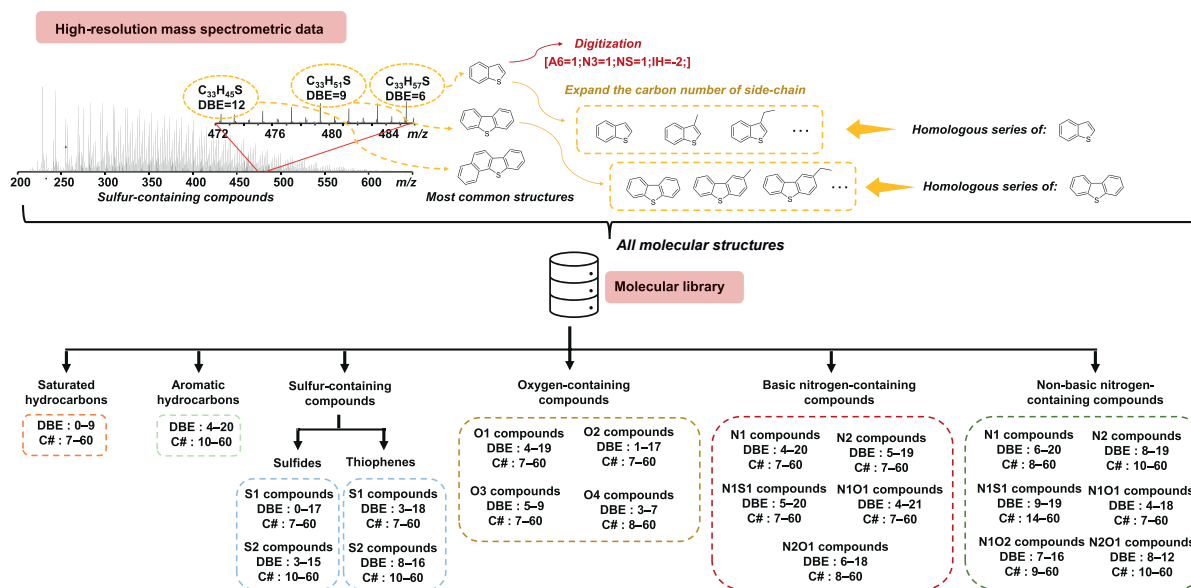


Fig. 2. VGO preset molecular library construction based on the HRMS.

2.4. Probability density function

The preset molecular library provides the molecular structures, while the molecular fractions are determined through the calculation of the probability density function (PDF) of different molecular attributes. The PDF parameters were tuned during the optimization step. The commonly used PDF functions are Gamma and Histogram PDFs, as presented in Eqs. (1) and (2), respectively.

$$f(x) = \frac{(x - \gamma)^{\alpha-1} e^{-(x-\gamma)/\beta}}{\Gamma(\alpha)\beta^\alpha} \quad (1)$$

$$f(x) = X_i \quad (2)$$

The Gamma PDF has three adjustable parameters (α , β , γ) for continuous property constraints such as carbon number and boiling point. In the Histogram PDF, the value of X_i represents the content of a specific molecular class, such as hydrocarbon, sulfur-containing compounds, nitrogen-containing compounds, and oxygen-containing compounds. The PDF combination form used in this work is depicted in Fig. 3. The PDF combination form is consistent with the molecular types detected by the HRMS to maximize the optimization efficiency. The Gamma PDFs of boiling point and DBE were set to ensure that the carbon number and DBE distribution of the compositional model agree well with the HRMS data. For example, as for basic N1S1 compounds, the total content is first constrained by Histogram PDFs (molecule type, nitrogen-containing compounds, basic nitrogen, N1S1, and AN/NN). The total content of basic N1S1 compounds is obtained by multiplying the function values between them. And then, the carbon number and DBE distribution in basic N1S1 compounds are constrained by Gamma PDFs, and the content of each compound is finally obtained. It is worth noting that C₄₀ series biomarker molecules exhibit unique content distribution patterns in hydrocarbons and sulfur-containing compounds in VGOs [63–66]. Therefore, a specific Histogram PDF was added to calculate the content of biomarker molecules in the compositional model. Among them, the PDFs contain 157 optimization parameters. The boundary conditions of different types of PDF functions are inconsistent. The

boundary conditions of the Gamma PDF and Histogram PDF are 380–980 and 0–1, respectively. The details of PDFs and boundary conditions of parameters were summarized in Table S2 of the Supplementary Material.

2.5. Objective function

The definition of the objective function is crucial for the accurate optimization of molecular fractions in the compositional model. In the traditional molecular reconstruction method, the only objective function is the property similarity, as shown in Eq. (3).

$$\text{Prop sim} = -\text{Prop error} = -\text{mean}(\delta |P_{\text{exp}} - P_{\text{pred}}|) \quad (3)$$

where P_{exp} is the experimental data of the bulk property, P_{pred} is the predictive value of the bulk property, and δ is the weight factor for different bulk properties.

In this work, composition similarity is introduced as a secondary objective. Due to variations in chemical derivatization and ionization methods, HRMS data across different compound classes cannot be compared directly [47]. To effectively utilize HRMS data, we proposed a method for calculating composition similarity between the compositional model and the HRMS data, as shown in Fig. 4. When calculating the composition similarity, the similarity for different compound types is first determined. Next, the average composition similarity of different compound types is calculated, and it will be used as the composition similarity between the compositional model and the HRMS data, as shown in Eq. (4).

$$\text{Comp sim} = \frac{\sum_{i=1}^k \text{CS}_i}{k} \quad (4)$$

where Comp sim represents the composition similarity between the compositional model and the HRMS data, k represents the number of different types of compounds, CS_i represents the

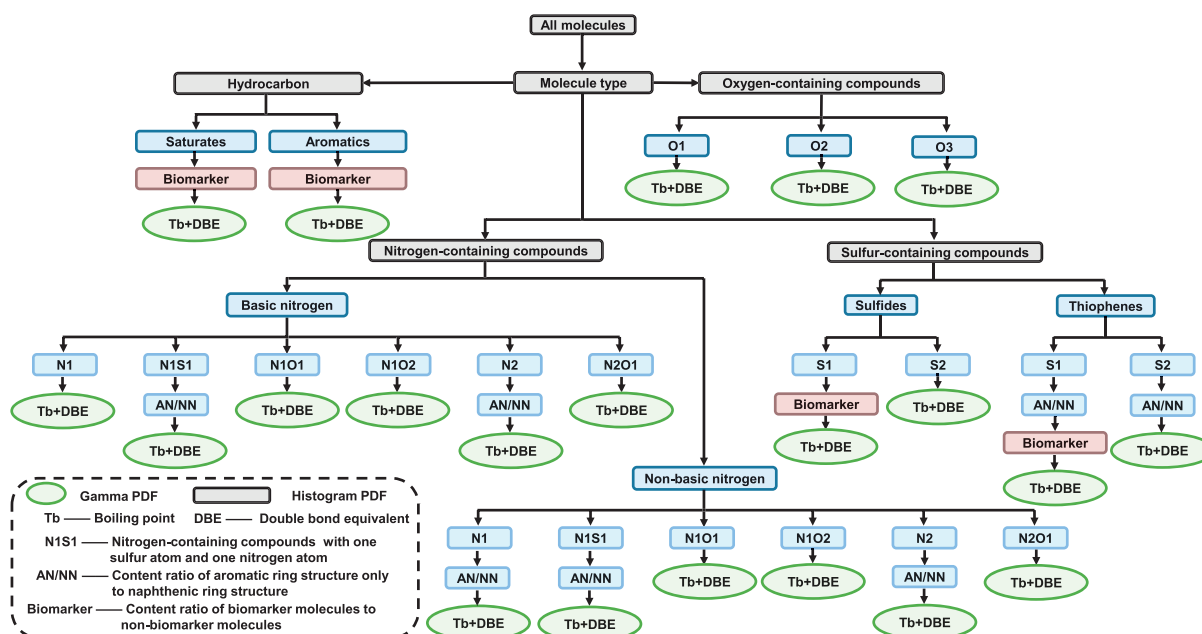


Fig. 3. PDF combination form used in the molecular reconstruction of VGO.

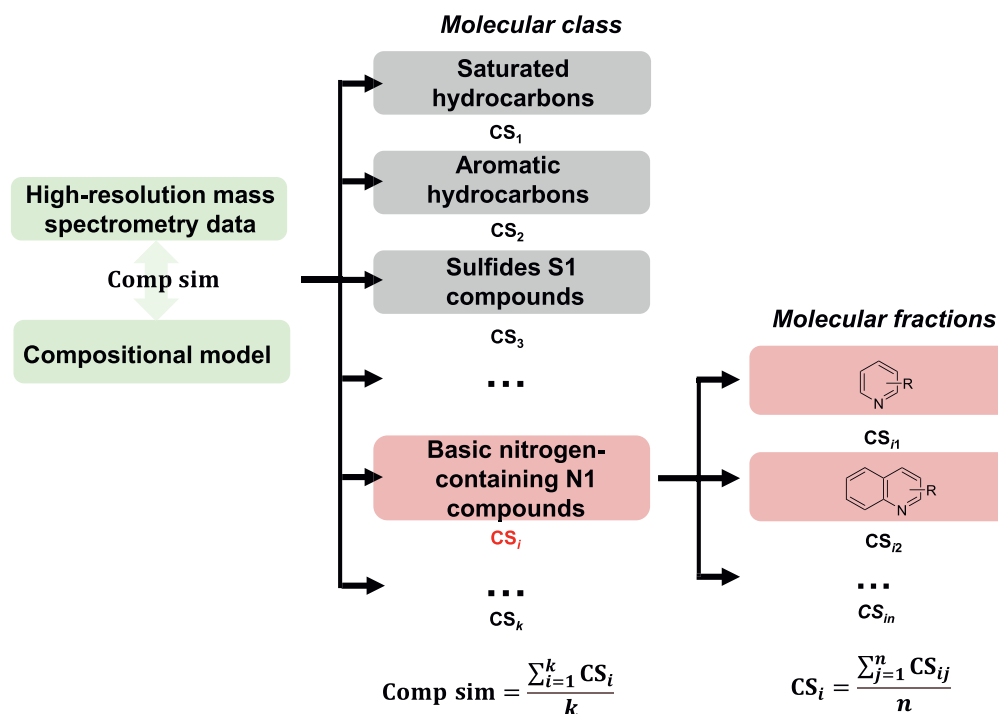


Fig. 4. The calculation method of the composition similarity between the compositional model and the HRMS data.

compositional similarity of compounds class i . When calculating the composition similarity of different compound types, the composition similarity of homologous species in this type of

compound should be calculated first. The average value is then used as the composition similarity for that type of compound, as shown in Eq. (5).

$$\text{CS}_i = \frac{\sum_{j=1}^n \text{CS}_{ij}}{n} \quad (5)$$

Herein, n is the number of different homologous species, CS_{ij} represents the composition similarity of j -type homologs species in i -type compounds. Based on the previous work [61], the cosine similarity formula is suitable to be used to calculate the composition similarity, as shown in Eq. (6).

$$\text{CS}_{ij} = \frac{\sum m_{\text{HRMS},ij} \times m_{\text{model},ij}}{\sqrt{\sum m_{\text{HRMS},ij}^2} \times \sqrt{\sum m_{\text{model},ij}^2}} \quad (6)$$

where, $m_{\text{model},ij}$ is the molecular fraction of the compositional model, $m_{\text{HRMS},ij}$ is the molecular fraction of HRMS data.

2.6. Optimization algorithm

Based on the above discussion, a mathematical model of molecular reconstruction method based on HRMS data can be proposed, and the simple addition form was used to transform the dual-objective optimization problem into a single-objective optimization problem, as shown in Eq. (7).

$$\begin{aligned} \min f(\mathbf{x}) &= \text{Prop error} - \omega \times \text{Comp sim} \\ &= \frac{\sum_{L=1}^N \delta_L |P_{\text{exp},L} - P_{\text{pred},L}|}{n} - \omega \frac{\sum_{i=1}^k \text{CS}_i}{k} \\ \text{s.t. } &x_{\text{low}} \leq x \leq x_{\text{up}} \end{aligned} \quad (7)$$

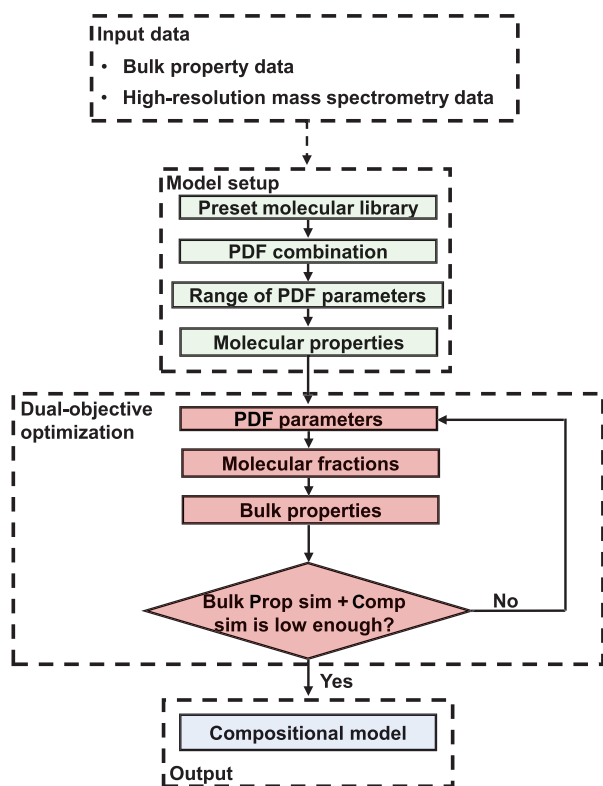


Fig. 5. The algorithm of molecular reconstruction based on HRMS.

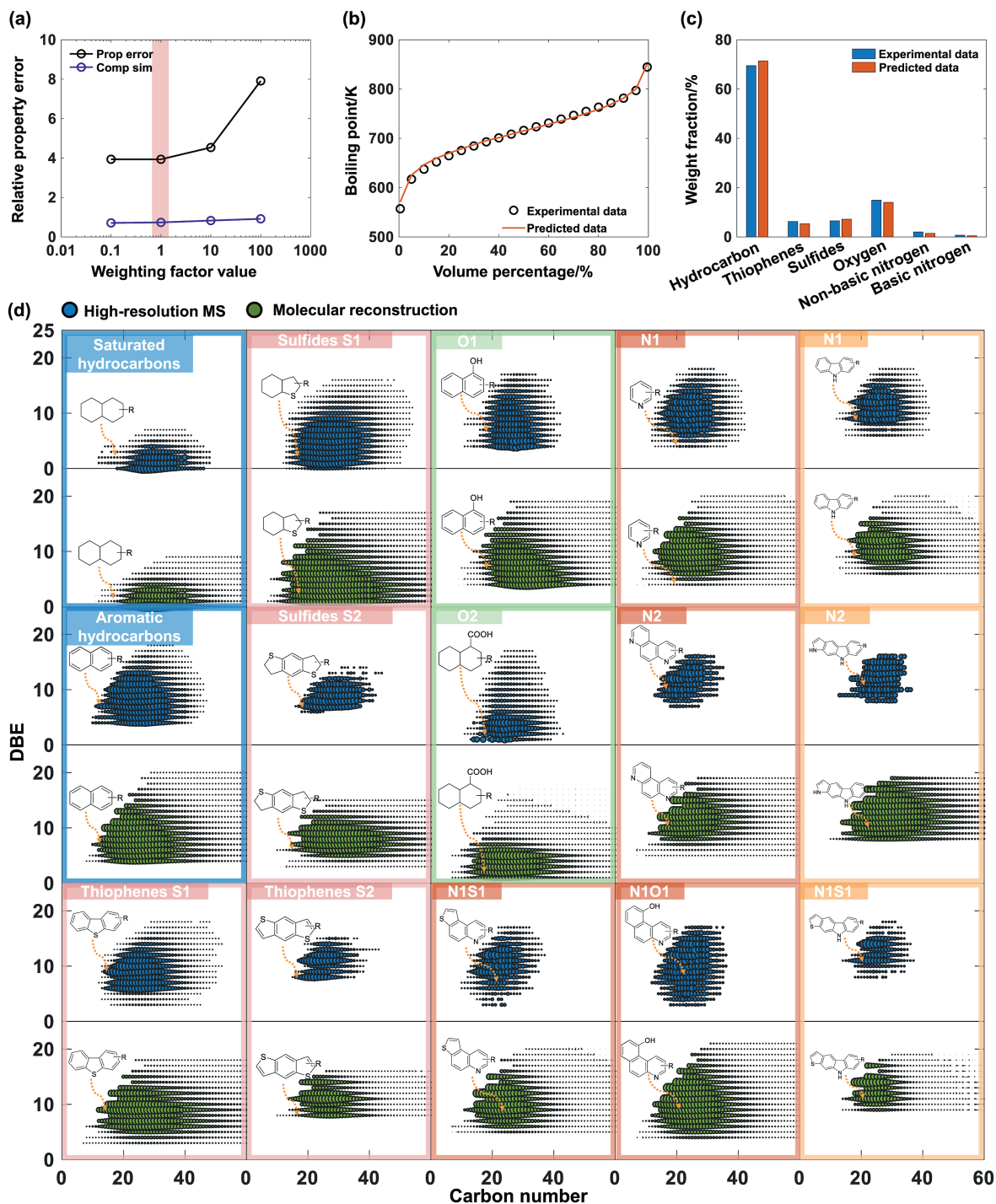


Fig. 6. The typical VGO compositional model results: (a) weight factor determination, (b) comparison of distillation curve between model and experimental data, (c) comparison of the content of different types of compounds between model and experimental data, (d) comparison of the molecular fraction distribution of different compound types between model and HRMS data.

where x is the optimization variable, x_{low} and x_{up} is the minimum and maximum value of the optimization variables. ω is the weight factor between the two objective functions. In Eq. (7), composition similarity is used to constrain molecular distribution, while bulk properties are employed to correct potential mass shifts in HRMS data.

Based on the above mathematical model, the algorithm of the molecular reconstruction method based on HRMS data is shown in

Fig. 5. Firstly, according to the HRMS data, we set the molecular library and PDF combination, determined the value range of optimization parameters, and generated the initial guesses of optimization variables. On this basis, the genetic algorithm (GA) was used to optimize the molecular fractions. When the bulk properties and molecular fractions of the compositional model are consistent with the experimental data and HRMS data, the final solution has been obtained.

3. Results and Discussion

3.1. Model result of a typical VGO

To obtain the optimal molecular composition, we need to first determine the weight factor between the property similarity and the composition similarity in the objective function. The results for weight factor determination are shown in Fig. 6(a). It can be seen from the result that the relative property error and relative composition similarity increase as the weight factor increases. Since the HRMS data is qualitative, the bulk property similarity is more accurate and should be guaranteed first. As shown in Fig. 6(a), when the weight factor is greater than 1, the bulk

property error increases sharply. Therefore, it is appropriate to set the weight factor to 1.

When the weight factor was set to 1.0, a typical VGO compositional model was constructed, as shown in Fig. 6(b)–(d). The results show that the distillation curve of the compositional model agrees well with the experimental data, and the content of different compound types in the compositional model is consistent with the experimental results published in the literature [62]. The comparison of the molecular distribution of different compounds between the model and the HRMS data is shown in Fig. 6(d). In addition to a slight tailing phenomenon in the calculated value, the molecular distribution of hydrocarbon compounds, sulfur-containing compounds, oxygen-containing

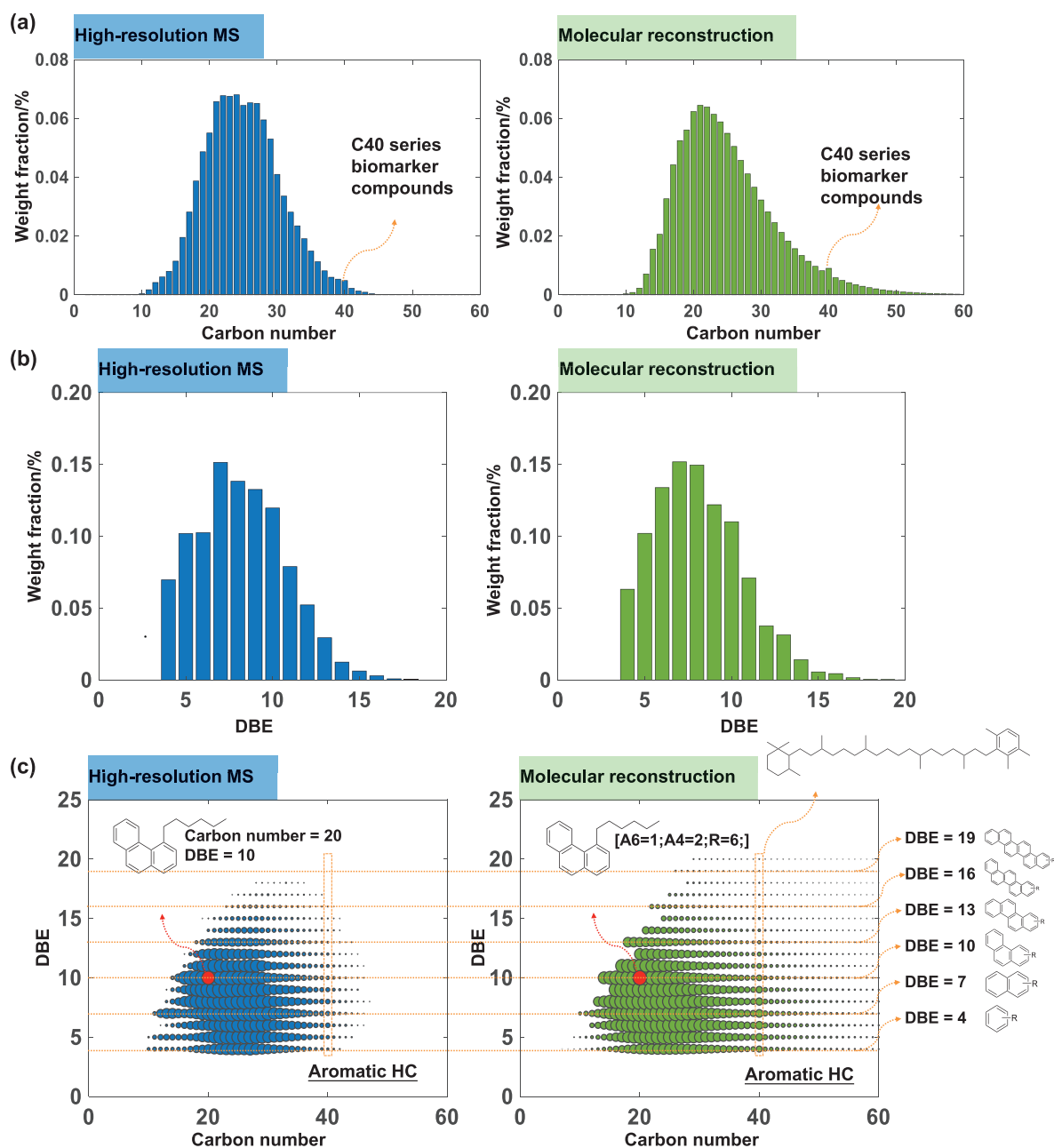


Fig. 7. The comparison of aromatic molecular fraction distributions of the typical VGO between compositional model and HRMS data: (a) comparison of carbon number distribution, (b) comparison of DBE distribution, (c) comparison of carbon number vs. DBE distribution.

compounds, basic nitrogen-containing compounds, and non-basic nitrogen-containing compounds of the model has a good agreement with the HRMS data. The trailing phenomenon is primarily due to mass discrimination occurring at the high-mass end of HRMS. Thus, this work used the boiling point distribution data to correct for the content of molecules with high molecular weight. Overall, the calculated results show that the molecular reconstruction method based on HRMS is effective, and the bulk properties and molecular fraction distributions of the VGO compositional models are accurate. According to the results, we repeatedly constructed the compositional model of the typical VGO. The results prove that the method has good repeatability. Moreover, the structural information of each VGO molecule can

be found in Supplementary Material 1, and the molecular content information can be found in Supplementary Material 2.

To further evaluate the accuracy of the novel method, we compared the calculated carbon number and DBE distribution of aromatic molecules with the HRMS data, as shown in Fig. 7. In the carbon number and DBE distribution diagram, each point represents a series of compounds with the same carbon number and DBE, and the point size represents its relative content. The points with the same DBE and different carbon numbers represent a series of homologs. For example, species with DBE = 4 representing alkyl benzene compounds and species with DBE = 10 and carbon number = 20 should be hexyl-phenanthrene. In Fig. 7(c), the content of C_{40} series biomarkers is higher than that of other

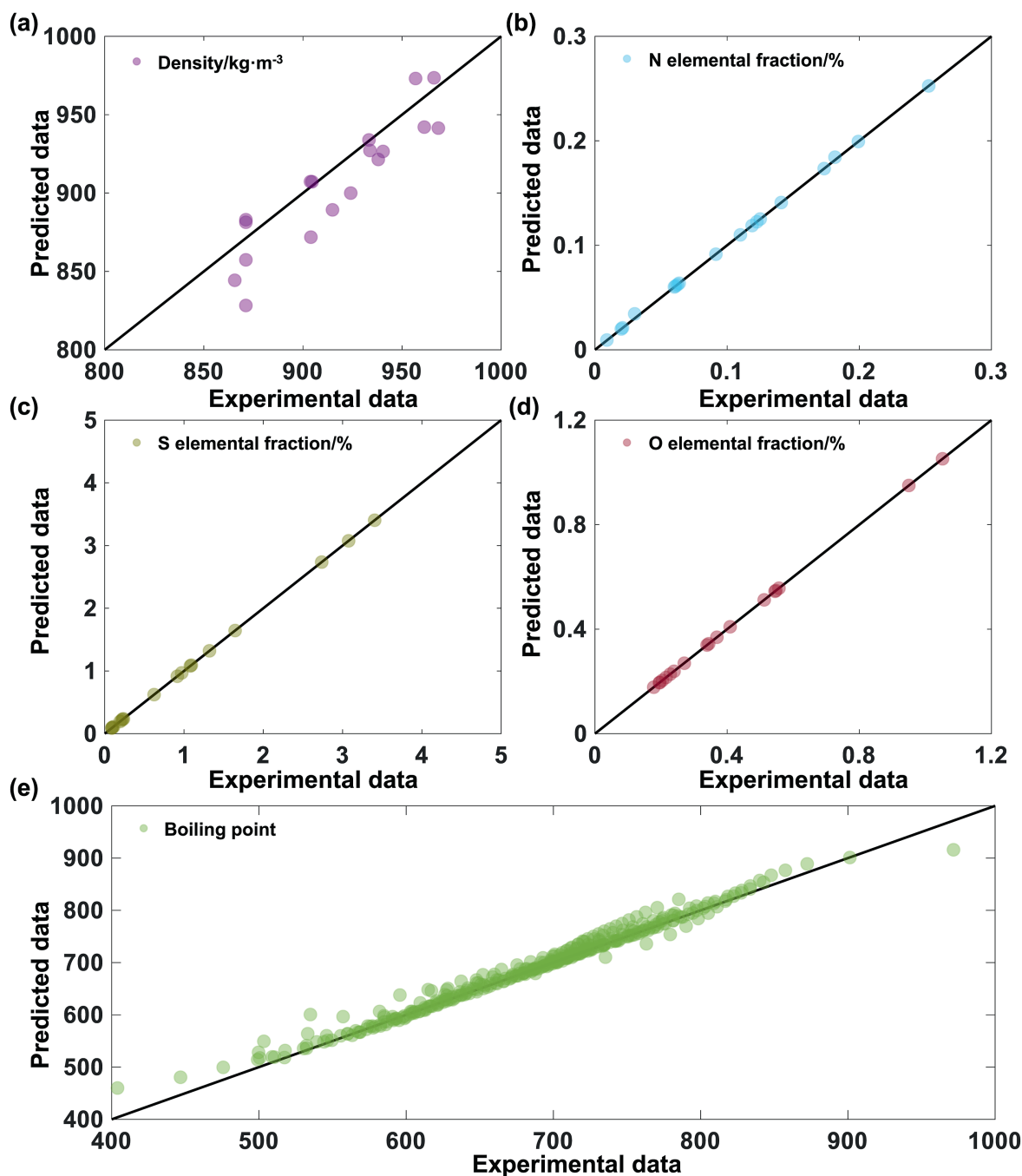


Fig. 8. Statistical bulk property accuracy of VGO compositional models from different origins: (a) density, (b)–(d) elemental fraction, (e) boiling point.

molecules, and the compositional model is consistent with the HRMS data. It means the model captures homologous series patterns.

3.2. Applications of VGOs from different origins

To validate the generalization of the molecular reconstruction based on the HRMS method, this work developed a series of VGO compositional models from different sources. The developed model reflects the enrichment of N- and S-containing species in higher boiling fractions. The statistical bulk property accuracy of the models is shown in Fig. 8. Overall, the calculated bulk property has a good agreement with the experimental data. However, compared with elemental content and boiling point, the density has a relatively larger error. It is primarily because density experiments usually have greater uncertainty compared to elemental analysis and simulated distillation. Thus, the weight factor of density is relatively low, and there is a relatively high calculated error. This work also calculated the mean absolute error (MAE) of density, which is $0.016 \text{ g} \cdot \text{cm}^{-3}$, and this result is acceptable for the VGO fraction.

The comparison of molecular fraction distributions between the compositional models and the HRMS data is shown in Fig. 9. The content of various compound types in the compositional model agrees well with the experimental data obtained from the literature [62]. As can be seen from the figure, the molecular compositions of typical compounds in VGO1-VGO3 are notably different. For example, VGO2 is rich in biomarker molecules. The results show that developed compositional models effectively capture these molecular distribution patterns, and it also validates that the method has great generalization.

3.3. Application on VGO subfractions

Furthermore, this study developed a range of VGO compositional models obtained from various distillates of the vacuum column, and the results are displayed in Fig. 10. The results showed that the content of heteroatom compounds in the compositional models increases with the boiling point increase of the distillates, which conforms to the experimental regulation. As the distillate becomes heavier, the molecular fraction distributions shift towards regions of higher carbon number and DBE. The models are in good agreement with the experimental data. Overall, the

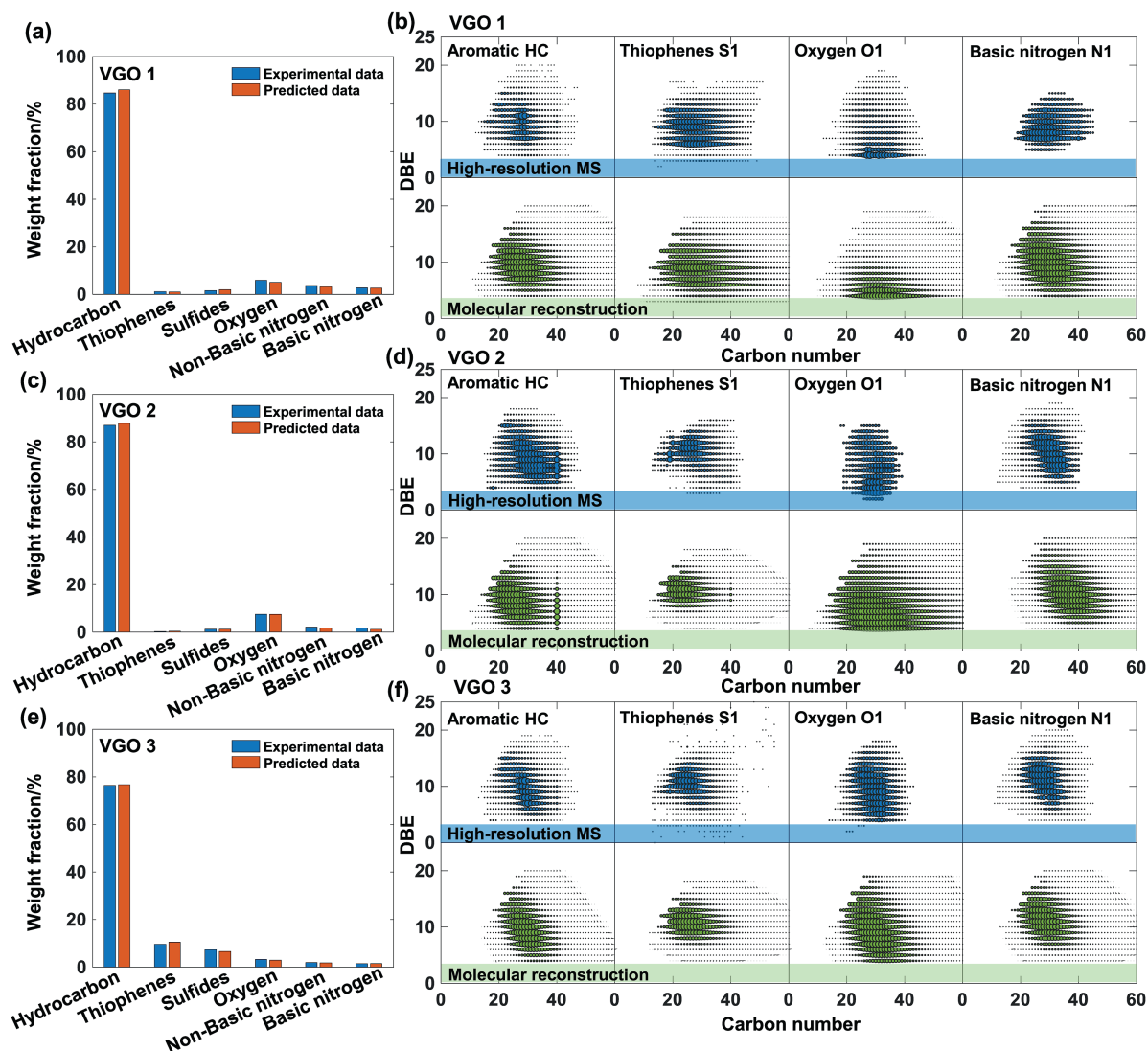


Fig. 9. The results of VGO compositional models from different origins: (a), (c) and (e) the comparison of contents for different types of compounds between compositional models and HRMS data; (b), (d) and (f) the comparison of fractions for different types of compounds between compositional models and HRMS data.

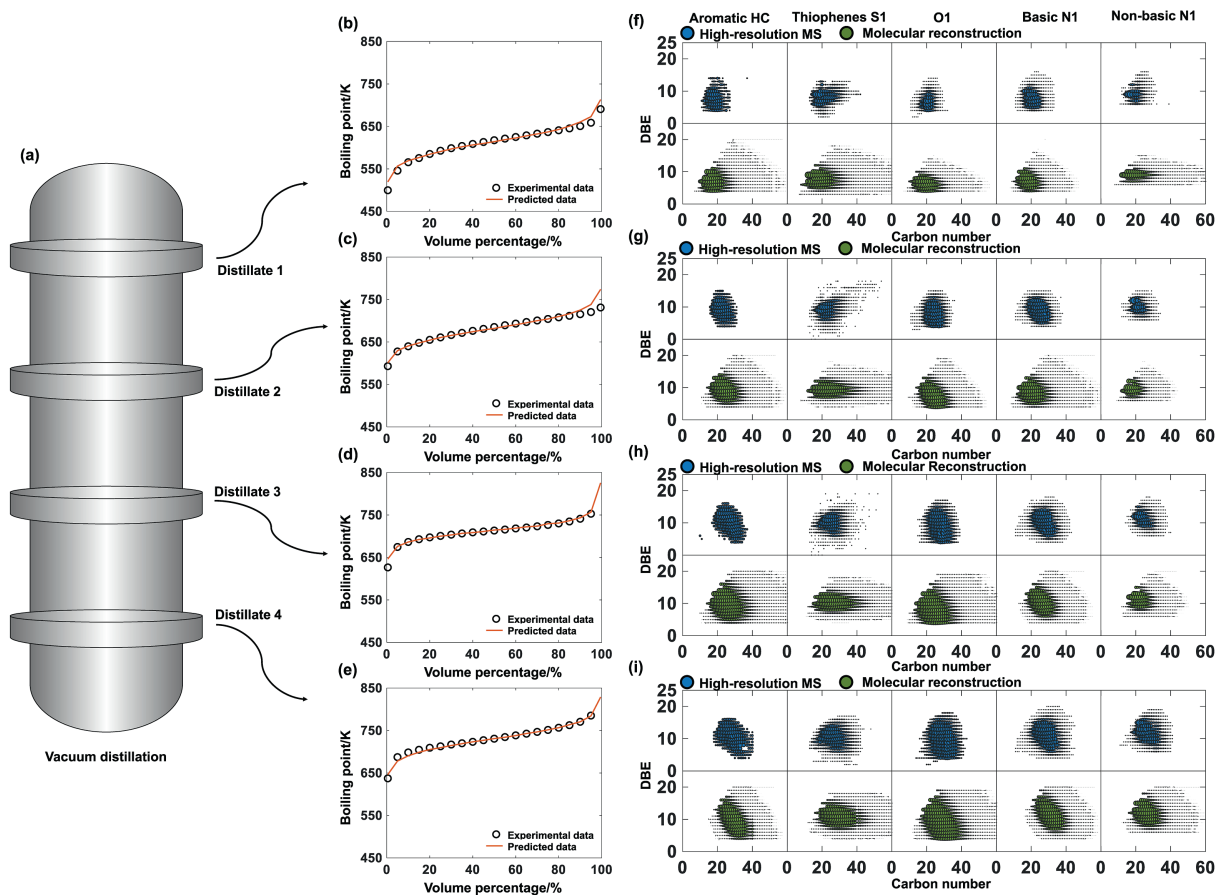


Fig. 10. The results of VGO compositional models of different distillates: (a) vacuum tower, (b)–(e) the comparison of the distillation curve between the model and experimental data of different distillates, (f)–(i) the comparison of fractions of different types of compounds between compositional models and HRMS data.

molecular reconstruction based on the HRMS method has good accuracy and generalization.

4. Conclusions

In this work, an HRMS-informed molecular reconstruction method was developed and applied in the VGO fraction. Unlike traditional models that only fit bulk property data, we incorporate HRMS-derived composition as a direct constraint in the objective function to better screen feedstocks. According to the molecular-level characterization information of VGO, a suitable PDF combination and composition similarity calculation method was established. As for the special molecular distribution of biomarker molecules, the corresponding PDF function for the content of biomarker molecules was set. The results from multiple VGO compositional models demonstrate that the calculated bulk properties have a good agreement with experimental data, and the molecular distributions of various compound types align with HRMS data. The molecular reconstruction method proposed in this study exhibits good repeatability with HRMS data and is expected to be applied for heavy oil processing and geochemistry studies.

CRediT Authorship Contribution Statement

Dong Guan: Writing – review & editing, Writing – original draft, Validation, Methodology. Xiangyu Zhao: Writing – review & editing, Data curation. Shuofan Li: Writing – review & editing,

Data curation. Zhengyu Chen: Writing – review & editing, Validation, Funding acquisition. Quan Shi: Writing – review & editing, Supervision, Resources. Chunming Xu: Supervision, Resources, Project administration. Linzhou Zhang: Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgements

This work was supported by the Coal-Major Project (2024ZD1700401-05), National Key Research and Development Program of China (2021YFA1501201), the National Natural Science Foundation of China (22222815, 22508414, 22021004), and the Science Foundation of China University of Petroleum, Beijing (2462023QNXZ004, 2462025BJRC007).

Supplementary Material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cjche.2025.12.038>.

References

- [1] W.H. Green, Moving from postdictive to predictive kinetics in reaction engineering, *AIChE J.* 66 (11) (2020) e17059.
- [2] J.K. Brown, W.R. Ladner, A study of the hydrogen distribution in coal-like materials by high-resolution nuclear magnetic resonance spectroscopy. 2. A comparison with infra-red measurement and the conversion to carbon structure, *Fuel* 39 (1) (1960) 87–96.
- [3] W.J. Lyu, L.Z. Zhang, K.Y. Li, G. Wang, Q. Shi, S.Q. Zhao, C.M. Xu, Average molecule construction of petroleum fractions based on ¹H-NMR, *AIChE J.* 65 (1) (2019) 270–280.
- [4] M. Neurock, C. Libanati, A. Nigam, M.T. Klein, Monte Carlo simulation of complex reaction systems: molecular structure and reactivity in modelling heavy oils, *Chem. Eng. Sci.* 45 (8) (1990) 2083–2088.
- [5] M. Neurock, A. Nigam, D. Trauth, M.T. Klein, Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms, *Chem. Eng. Sci.* 49 (24) (1994) 4153–4177.
- [6] D.M. Campbell, M.T. Klein, Construction of a molecular representation of a complex feedstock by Monte Carlo and quadrature methods, *Appl. Catal. Gen.* 160 (1) (1997) 41–54.
- [7] D.M. Trauth, S.M. Stark, T.F. Petti, M. Neurock, M.T. Klein, Representation of the molecular structure of petroleum resid through characterization and Monte Carlo modeling, *Energy Fuel* 8 (3) (1994) 576–580.
- [8] C.U. Deniz, M. Yasar, M.T. Klein, Stochastic reconstruction of complex heavy oil molecules using an artificial neural network, *Energy Fuel* 31 (11) (2017) 11932–11938.
- [9] C.U. Deniz, Artificial neural network modeling of the heavy petroleum fractions at molecular level, Ph.D. Thesis, Istanbul University, Turkey, 2017.
- [10] R.J. Quann, S.B. Jaffe, Structure-oriented lumping: describing the chemistry of complex hydrocarbon mixtures, *Ind. Eng. Chem. Res.* 31 (11) (1992) 2483–2497.
- [11] R.J. Quann, S.B. Jaffe, Building useful models of complex reaction systems in petroleum refining, *Chem. Eng. Sci.* 51 (10) (1996) 1615–1635.
- [12] S.B. Jaffe, H. Freund, W.N. Olmstead, Extension of structure-oriented lumping to vacuum residua, *Ind. Eng. Chem. Res.* 44 (26) (2005) 9840–9852.
- [13] L.D. Tian, J.M. Wang, B.X. Shen, J.C. Liu, Building a kinetic model for steam cracking by the method of structure-oriented lumping, *Energy Fuel* 24 (8) (2010) 4380–4386.
- [14] L.D. Tian, B.X. Shen, J.C. Liu, A delayed coking model built using the structure-oriented lumping method, *Energy Fuel* 26 (3) (2012) 1715–1724.
- [15] L.D. Tian, B.X. Shen, J.C. Liu, Building and application of delayed coking structure-oriented lumping model, *Ind. Eng. Chem. Res.* 51 (10) (2012) 3923–3931.
- [16] J.C. Liu, H. Chen, Z.P. Pi, Y.F. Liu, H. Sun, B.X. Shen, Molecular-level-process model with feedback of the heat effects on a complex reaction network in a fluidized catalytic cracking process, *Ind. Eng. Chem. Res.* 56 (13) (2017) 3568–3577.
- [17] X.L. Qin, J.C. Liu, C. Wang, L. Ye, B. Xing, W.X. Yu, J.Q. Xie, H.Z. Wang, Y. Ji, D.N. Lu, Molecular level analysis on performance of diameter expanding reactor to improve gasoline quality in FCC process, *Fuel* 290 (2021) 119978.
- [18] B. Xing, L. Ye, J.C. Liu, X.L. Qin, W.X. Yu, J.Q. Xie, L.X. Hou, H.Z. Wang, Y. Ji, D.N. Lu, J.G. Zhao, H. Sun, H. Ling, Reaction network of sulfur compounds in delayed coking process, *Chem. Eng. J.* 422 (2021) 129903.
- [19] J.C. Chen, Z. Fang, T. Qiu, Molecular reconstruction model based on structure oriented lumping and group contribution methods, *Chin. J. Chem. Eng.* 26 (8) (2018) 1677–1683.
- [20] B. Peng, Molecular modelling of petroleum processes, Ph.D Thesis, The University of Manchester, UK, 1999.
- [21] D. Hudebine, J.J. Verstraete, Molecular reconstruction of LCO gasoils from overall petroleum analyses, *Chem. Eng. Sci.* 59 (22–23) (2004) 4755–4763.
- [22] J.J. Verstraete, N. Revellin, H. Dulot, D. Hudebine, Molecular reconstruction of vacuum gasoils, *Prep. Am. Chem. Soc. Div. Fuel Chem.* 49 (1) (2004) 20–21.
- [23] J.J. Verstraete, P. Schnongs, H. Dulot, D. Hudebine, Molecular reconstruction of heavy petroleum residue fractions, *Chem. Eng. Sci.* 65 (1) (2010) 304–312.
- [24] S. Feng, C. Cui, K.Y. Li, L.Z. Zhang, Q. Shi, S.Q. Zhao, C.M. Xu, Molecular composition modelling of petroleum fractions based on a hybrid structural unit and bond-electron matrix (SU-BEM) framework, *Chem. Eng. Sci.* 201 (2019) 145–156.
- [25] D. Guan, L.Z. Zhang, Initial guess estimation and fast solving of petroleum complex molecular reconstruction model, *AIChE J.* 68 (10) (2022) e17782.
- [26] Y.M. Guan, D. Guan, C. Zhang, S.H. Yuan, G.Q. Cai, L.Z. Zhang, Diesel molecular composition and blending modeling based on SU-BEM framework, *Pet. Sci.* 19 (2) (2022) 839–847.
- [27] C. Cui, L.Z. Zhang, Y.J. Ma, T. Billa, Z. Hou, Q. Shi, S.Q. Zhao, C.M. Xu, M.T. Klein, Computer-aided gasoline compositional model development based on GC-FID analysis, *Energy Fuel* 32 (8) (2018) 8366–8373.
- [28] G.Q. Cai, Z.F. Liu, L.Z. Zhang, Q. Shi, S.Q. Zhao, C.M. Xu, Systematic performance evaluation of gasoline molecules based on quantitative structure-property relationship models, *Chem. Eng. Sci.* 229 (2021) 116077.
- [29] J.B. Phillips, J. Beens, Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions, *J. Chromatogr. A* 856 (1–2) (1999) 331–347.
- [30] A. Alvarez-Majmutov, J.W. Chen, R. Gieleciak, D. Hager, N. Heshka, S. Salmon, Deriving the molecular composition of middle distillates by integrating statistical modeling with advanced hydrocarbon characterization, *Energy Fuel* 28 (12) (2014) 7385–7393.
- [31] A. Alvarez-Majmutov, R. Gieleciak, J.W. Chen, Deriving the molecular composition of vacuum distillates by integrating statistical modeling and detailed hydrocarbon characterization, *Energy Fuel* 29 (12) (2015) 7931–7940.
- [32] A.G. Marshall, R.P. Rodgers, Petroleomics: the next grand challenge for chemical analysis, *Acc. Chem. Res.* 37 (1) (2004) 53–59.
- [33] R.P. Rodgers, T.M. Schaub, A.G. Marshall, Petroleomics: MS returns to its roots, *Anal. Chem.* 77 (1) (2005) 20A–27A.
- [34] B. Schuler, S. Fatayer, G. Meyer, E. Rogel, M. Moir, Y.L. Zhang, M.R. Harper, A. E. Pomerantz, K.D. Bake, M. Witt, D. Peña, J.D. Kushnerick, O.C. Mullins, C. Ovalles, F.G.A. van den Berg, L. Gross, Heavy oil based mixtures of different origins and treatments studied by atomic force microscopy, *Energy Fuel* 31 (7) (2017) 6856–6861.
- [35] B. Schuler, G. Meyer, D. Peña, O.C. Mullins, L. Gross, Unraveling the molecular structures of asphaltene by atomic force microscopy, *J. Am. Chem. Soc.* 137 (31) (2015) 9870–9876.
- [36] K.N. Qian, K.E. Edwards, M. Siskin, Characterization of large nonvolatile polyaromatic molecules by a combination of in-source pyrolysis and field desorption mass spectrometry, *Energy Fuel* 15 (4) (2001) 949–954.
- [37] K.N. Qian, W.K. Robbins, C.A. Hughey, H.J. Cooper, R.P. Rodgers, A.G. Marshall, Resolution and identification of elemental compositions for more than 3000 crude acids in heavy petroleum by negative-ion microelectrospray high-field Fourier transform ion cyclotron resonance mass spectrometry, *Energy Fuel* 15 (6) (2001) 1505–1511.
- [38] A.M. McKenna, G.T. Blakney, F. Xian, P.B. Glaser, R.P. Rodgers, A.G. Marshall, Heavy petroleum composition. 2. Progression of the Boduszynski model to the limit of distillation by ultrahigh-resolution FT-ICR mass spectrometry, *Energy Fuel* 24 (5) (2010) 2939–2946.
- [39] A.M. McKenna, J.M. Purcell, R.P. Rodgers, A.G. Marshall, Heavy petroleum composition. 1. Exhaustive compositional analysis of Athabasca bitumen HVGO distillates by Fourier transform ion cyclotron resonance mass spectrometry: a definitive test of the Boduszynski model, *Energy Fuel* 24 (5) (2010) 2929–2938.
- [40] A.M. McKenna, L.J. Donald, J.E. Fitzsimmons, P. Juyal, V. Spicer, K.G. Standing, A.G. Marshall, R.P. Rodgers, Heavy petroleum composition. 3. Asphaltene aggregation, *Energy Fuel* 27 (3) (2013) 1246–1256.
- [41] A.M. McKenna, A.G. Marshall, R.P. Rodgers, Heavy petroleum composition. 4. Asphaltene compositional space, *Energy Fuel* 27 (3) (2013) 1257–1267.
- [42] D.C. Podgorski, Y.E. Corilo, L. Nyadong, V.V. Lobodin, B.J. Bythell, W.K. Robbins, A.M. McKenna, A.G. Marshall, R.P. Rodgers, Heavy petroleum composition. 5. Compositional and structural continuum of petroleum revealed, *Energy Fuel* 27 (3) (2013) 1268–1276.
- [43] B.D. Marshall, C.P. Bokis, A PC-SAFT model for hydrocarbons I: Mapping aromatic π - π interactions onto a dipolar free energy, *Fluid Phase Equilib.* 489 (2019) 83–89.
- [44] X. Chen, H.D. Li, L.Z. Zhang, Q. Shi, S.Q. Zhao, C.M. Xu, CYHPO oxidation followed by methylation for selective characterization of thiophenic and sulfidic compounds in petroleum via ESI FT-ICR MS, *Fuel* 265 (2020) 116907.
- [45] Y. Zhang, X. Chen, L.Z. Zhang, Q. Shi, S.Q. Zhao, C.M. Xu, Specification of the nitrogen functional group in a hydrotreated petroleum molecule using hydrogen/deuterium exchange electrospray ionization high-resolution mass spectrometry, *Analyst* 145 (13) (2020) 4442–4451.
- [46] A. Bojkovic, F.H. Vermeire, M. Kuzmanović, H. Dao Thi, K.M. van Geem, Analytics driving kinetics: advanced mass spectrometric characterization of petroleum products, *Energy Fuel* 36 (1) (2022) 6–59.
- [47] H.D. Li, Y.H. Zhang, C.M. Xu, S.Q. Zhao, K.H. Chung, Q. Shi, Quantitative molecular composition of heavy petroleum fractions: a case study of fluid catalytic cracking decant oil, *Energy Fuel* 34 (5) (2020) 5307–5316.
- [48] Q. Shi, J.X. Wu, Review on sulfur compounds in petroleum and its products: state-of-the-art and perspectives, *Energy Fuel* 35 (18) (2021) 14445–14461.
- [49] Q. Shi, Y.H. Zhang, K.H. Chung, S.Q. Zhao, C.M. Xu, Molecular characterization of fossil and alternative fuels using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry: recent advances and perspectives, *Energy Fuel* 35 (22) (2021) 18019–18055.
- [50] H.D. Li, S.F. Li, J.X. Wu, L.Y. Xie, Y.M. Liang, Y.H. Zhang, S.Q. Zhao, C.M. Xu, Q. Shi, Molecular characterization of aromatics in petroleum fractions by combining silica sulfuric acid sulfonation with electrospray ionization high-resolution mass spectrometry, *Fuel* 317 (2022) 123463.
- [51] Y.L. Zhang, Nonalternant aromaticity and partial double bond in petroleum molecules revealed: theoretical understanding of polycyclic aromatic hydrocarbons obtained by noncontact atomic force microscopy, *Energy Fuel* 33 (5) (2019) 3816–3820.
- [52] P. Zahl, Y.L. Zhang, Guide for atomic force microscopy image analysis to discriminate heteroatoms in aromatic molecules, *Energy Fuel* 33 (6) (2019) 4775–4780.
- [53] Y.L. Zhang, F. Schulz, B.M. Rytting, C.C. Walters, K. Kaiser, J.N. Metz, M.R. Harper, S.S. Merchant, A.S. Mennito, K.N. Qian, J.D. Kushnerick, P.K. Kilpatrick, L. Gross, Elucidating the geometric substitution of tetraporphyrins by spectroscopic analysis and atomic force microscopy molecular imaging, *Energy Fuel* 33 (7) (2019) 6088–6097.
- [54] B. Schuler, Y.L. Zhang, F. Liu, A.E. Pomerantz, A.B. Andrews, L. Gross, V. Pauchard, S. Banerjee, O.C. Mullins, Overview of asphaltene nanostructures

- and thermodynamic applications, *Energy Fuel*. 34 (12) (2020) 15082–15105.
- [55] Y.L. Zhang, M. Siskin, M.R. Gray, C.C. Walters, R.P. Rodgers, Mechanisms of asphaltene aggregation: puzzles and a new hypothesis, *Energy Fuel*. 34 (8) (2020) 9094–9107.
- [56] Y.L. Zhang, Applications of noncontact atomic force microscopy in petroleum characterization: opportunities and challenges, *Energy Fuel*. 35 (18) (2021) 14422–14444.
- [57] P.C. Chen, J.N. Metz, A.S. Mennito, S. Merchant, S.E. Smith, M. Siskin, S.P. Rucker, D.C. Dankworth, J.D. Kushnerick, N. Yao, Y.L. Zhang, Petroleum pitch: exploring a 50-year structure puzzle with real-space molecular imaging, *Carbon* 161 (2020) 456–465.
- [58] L.Z. Zhang, Z. Hou, S.R. Horton, M.T. Klein, Q. Shi, S.Q. Zhao, C.M. Xu, Molecular representation of petroleum vacuum resid, *Energy Fuel*. 28 (3) (2014) 1736–1749.
- [59] A. Alvarez-Majmutov, R. Gieleciak, J.W. Chen, Modeling the molecular composition of vacuum residue from oil sand bitumen, *Fuel* 241 (2019) 744–752.
- [60] D. Guan, Z.Y. Chen, X. Chen, Y. Zhang, Q.Y. Qi, Q. Shi, S.Q. Zhao, C.M. Xu, L.Z. Zhang, Molecular-level heavy petroleum hydrotreating modeling and comparison with high-resolution mass spectrometry, *Fuel* 297 (2021) 120792.
- [61] D. Guan, G.Q. Cai, L.Z. Zhang, Dual-objective optimization for petroleum molecular reconstruction based on property and composition similarities, *AIChE J.* 69 (8) (2023) e18108.
- [62] S.F. Li, J.X. Wu, Y.D. Wang, Y.Y. Li, W.L. Zhang, Y.H. Zhang, K.Y. He, C.X. Cai, G. Y. Bian, H.Z. Wang, Y. Ji, Q. Shi, Semi-quantitative analysis of molecular composition for petroleum fractions using electrospray ionization high-resolution mass spectrometry, *Fuel* 335 (2023) 127049.
- [63] J.X. Wu, W.L. Zhang, C. Ma, L.M. Ren, S.M. Li, Y.H. Zhang, Q. Shi, Separation and characterization of squalene and carotenoids derived sulfides in a low mature crude oil, *Fuel* 270 (2020) 117536.
- [64] W.M. Liu, Y.H. Liao, Q. Shi, C.S. Hsu, B. Jiang, P.A. Peng, Origin of polar organic sulfur compounds in immature crude oils revealed by ESI FT-ICR MS, *Org. Geochem.* 121 (2018) 36–47.
- [65] Y. Hebling, P. Schaeffer, A. Behrens, P. Adam, G. Schmitt, P. Schneckeburger, S. M. Bernasconi, P. Albrecht, Biomarker evidence for a major preservation pathway of sedimentary organic carbon, *Science* 312 (5780) (2006) 1627–1631.
- [66] S.F. Zhang, J. Huo, X.G. Sun, F.X. Yang, P. Wang, J.X. Wu, Y.H. Zhang, Q. Shi, Molecular composition reveals unique rheological property of Karamay heavy crude oil, *Energy Fuel*. 35 (1) (2021) 473–478.