

# Development of Digital Oil for Heavy Crude Oil: Molecular Model and Molecular Dynamics Simulations

*Motoaki Iwase<sup>1,2</sup>, Shumpei Sugiyama<sup>3</sup>, Yunfeng Liang<sup>2</sup>, Yoshihiro Masuda<sup>1,2</sup>, Masato Morimoto<sup>4</sup>,  
Toshifumi Matsuoka<sup>5</sup>, Edo S. Boek<sup>6</sup>, Ryo Ueda<sup>7</sup>, and Kazunori Nakagawa<sup>7</sup>*

*<sup>1</sup>Department of Systems Innovation, The University of Tokyo, Tokyo 113-8656, Japan*

*<sup>2</sup>Center for Engineering, Research into Artifacts (RACE), The University of Tokyo, Chiba 277-8568,  
Japan*

*<sup>3</sup>Environment and Resource System Engineering, Kyoto University, Kyoto 615-8540, Japan*

*<sup>4</sup>National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki 305-8569, Japan*

*<sup>5</sup>Fukada Geological Institute, Tokyo 113-0021, Japan*

*<sup>6</sup>School of Engineering and Materials Science, Queen Mary University of London, London E1 4NS,  
United Kingdom*

*<sup>7</sup>Japan Petroleum Exploration Co., Ltd. (JAPEX), Chiba 261-0025, Japan*

\*Corresponding author: Yunfeng Liang

Postal address: The University of Tokyo  
Room 566, Kashiwa Research Complex 5F, Kashiwa Campus  
5-1-5 Kashiwanoha, Kashiwa  
Chiba 277-8568, Japan

Tel: +81-3-7136-4271

Email: [liang@race.u-tokyo.ac.jp](mailto:liang@race.u-tokyo.ac.jp)

\*Corresponding author: Yoshihiro Masuda

Email: [masuda@race.u-tokyo.ac.jp](mailto:masuda@race.u-tokyo.ac.jp)

## Abstract

We constructed a molecular model (digital oil model) for heavy crude oil based on analytical data and our newly developed method. Crude oil was separated into four fractions: saturates, aromatics, resins, and asphaltenes (SARA). Although it is classified as a heavy crude oil, the asphaltenes turned out to be at very low weight concentration (~0.4 wt. %), and were ignored in our study. The digital oil was constructed as a mixture of representative molecules of four fractions: saturates, aromatics, resins, and lost components (which resulted from our SARA analysis). Representative molecules were generated by quantitative molecular representation (QMR), a technique that provides a set of molecules consistent with analytical data, such as elemental composition, average molecular mass, and the proportions of structural types of hydrogen and carbon atoms, as revealed by  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance. To enable the QMR method to be applicable to saturates, we made two developments: the first was the generation of non-aromatic molecules by a new algorithm that can generate a more branched structure by separating the chain bonding into main and subsidiary processes; the second was that the molecular mass distribution of the model could be fitted to that obtained from experiments. To validate the digital oil thus obtained, we first confirmed the validity of the model for each fraction in terms of plots of double-bond equivalent as a function of carbon number. We then calculated its density and viscosity by molecular dynamics simulations. The calculated density was in good agreement with experimental data for crude oil. The calculated viscosity was higher than experimental values; however, the error appeared systematic, being a factor of ~1.5 higher than that of experiments. Moreover, the calculated viscosity as a function of temperature was well described by the Vogel–Fulcher–Tammann equation. Digital oil will be a powerful tool to analyze both macroscopic properties and microscopic phenomena of crude oil under any thermodynamic conditions.

## 1. Introduction

Heavy oil, which includes extra heavy oil and bitumen, is attractive due to its huge resources. Conventional oil makes up only about 30 % of the world's total oil resources, with the remainder occurring as heavy oil, extra heavy oil, and bitumen.<sup>1,2</sup> The recovery and refinery of heavy oil are, however, more difficult and costly than those of conventional oil because of its high viscosity and heavier molecular compositions.<sup>1-4</sup> Unlike conventional light oil, enhanced oil recovery (EOR) is an essential technique for heavy oil recovery even at early development stage. In solvent-injection EOR, solvents reduce the viscosity of heavy oil, but solvents that are effective vary depending on the oil field.<sup>1,2</sup> It is therefore necessary to understand the oil components and their property changes on adding solvents to enable the effectiveness of different solvents to be evaluated.

Owing to the development of computational technology, molecular-scale simulations are now widely used in petroleum engineering.<sup>5-15</sup> The advantage of those simulations is to provide insights that cannot be extracted from experimental data alone. For molecular-scale simulations, it is important to construct an accurate molecular model of crude oil. Early studies reported many average structures of various fractions, such as saturates, aromatics, resins, and asphaltenes;<sup>12,16-19</sup> however, a full molecular model of crude oil containing gas, light components, heavy components, and asphaltenes has not been reported. Theoretically, a molecular model of crude oil must contain all the components of that oil, but it is impossible to detect all of these. One approach to characterize a complex mixture is to use averaged data based on simple measurements of the whole sample.<sup>3</sup> We have therefore utilized a digital oil technique. Digital oil is a molecular model of crude oil, which is represented as a mixture of representative molecules of each fraction. Because the representative molecules are generated based on the average information for that fraction, we need not detect all the components (when the separation is unworkable). Once we construct a digital oil, we can analyze both its macroscopic properties and microscopic phenomena under any thermodynamic conditions using molecular dynamics (MD) simulations. Eventually, we may use the digital oil to make predictions when the relevant experimental data are not easily accessible.

In previous work, we constructed a digital oil model for light oil.<sup>15</sup> The gaseous and light fractions

were detected by gas chromatography, and hundreds of molecular species, including n-alkanes, isoalkanes, naphthenes, alkylbenzenes, and polyaromatics, were modeled. The heavy fraction and asphaltenes were represented as molecular mixtures by quantitative molecular representation (QMR)<sup>10</sup>, which is a technique that can generate representative molecules consistent with analytical data, such as elemental composition, average molecular mass, and **the proportions** of structural types of hydrogen and carbon atoms obtained from <sup>1</sup>H and <sup>13</sup>C nuclear magnetic resonance (NMR) experiments. The digital oil was constructed as a mixture of the molecular models of all fractions. It could successfully reproduce the properties of light crude oil, such as density and viscosity.

When we applied this digital oil method to heavy oils, the major difficulty was that we could not characterize the light fraction by gas chromatography, even when we separated this fraction into saturates and aromatics. In addition, when we applied QMR to the light fraction, especially saturates, another problem arose: **the current QMR could not generate non-aromatic molecules; that is, we could not construct a model for saturates** due to the software restrictions. In this study, we therefore developed the QMR to be applicable to saturates and expanded the digital oil **technique** to describe heavy crude oil.

## **2. Methodology**

In this study, we constructed a digital oil model for a heavy crude oil (in North Japan, API gravity: 11–17°). First, we separated a crude oil sample into the four fractions determined by SARA (saturates, aromatics, resins, and asphaltenes) analysis. Second, we experimentally analyzed each fraction and obtained average structural information. Third, we generated representative molecules of each fraction by QMR. **Fourth**, the digital oil was constructed as a mixture of representative molecules of all fractions. **Finally, the crude oil properties (density and viscosity) were calculated using MD simulations with the digital oil model. As part of investigation of the crude oil properties, the density and viscosity of the crude oil were experimentally measured at a few temperature and pressure conditions. They are reported in this paper in order to validate our digital model. In the following Sections 2.1 to 2.3, we will make a brief summary on our experiments. In Sections 2.4 to 2.6, we will**

discuss the model and simulation details.

## 2.1. Separation of Crude Oil

We separated the crude oil sample into four fractions: saturates, aromatics, resins, and asphaltenes, based on SARA analysis.<sup>3</sup> These class designations are indirect labels in terms of chemical structure and composition.<sup>3</sup> Ideally, saturates contain only aliphatic compounds; however, the saturate fraction resulting from this analysis often contains a significant amount of mono-aromatic molecules.<sup>16</sup> The aromatic component contains a variety of aromatic compounds with saturated groups attached. Resins contain a greater number of heteroatoms and have a higher concentration of aromatic carbon, while asphaltenes are the highest molecular mass fraction and contain most of the polar compounds.<sup>3</sup> The distinctions in molecular structure between aromatics and resins, and between resins and asphaltenes, therefore, are not clear.<sup>3</sup> In the SARA analysis, the crude oil sample was first refluxed with heptane and separated into soluble maltenes and insoluble asphaltenes, after which the saturates, aromatics, and resins were extracted from the maltenes in an alumina column using heptane, toluene, and methanol–toluene solvents, respectively.

## 2.2. Analysis of Crude Oil Fractions

To generate representative molecules using QMR, we required the analytical data for each fraction, including the elemental composition, average molecular mass, and **the proportions** of structural types of hydrogen and carbon atoms as obtained from <sup>1</sup>H- and <sup>13</sup>C-NMR experiments.

**Elemental composition.** The elemental compositions of carbon, hydrogen, nitrogen, and sulfur were directly measured by CHNS elemental analysis using a FLASH2000 analyzer (Thermo Fisher Scientific, USA), while the amount of oxygen was calculated by difference.

**Average molecular mass.** The average molecular mass was measured by gel permeation chromatography (GPC) and gas chromatography distillation (GCD). GPC, also called size-exclusion chromatography, is a technique to determine the average molecular mass based on molecular size by column chromatography. The permeation and exclusion limits of the column were 200 and 70 000 u,

respectively. For GPC, we used PU-980 (JASCO, Japan), AS-2055 (JASCO), CO-965 (JASCO), KF-403HQ (Showa Denko, Japan), and an Infinity ELSD (Agilent Technologies, USA) as the high-performance liquid chromatography (HPLC) pump, auto injector, thermostatic bath, column, and evaporative light-scattering detector, respectively. Tetrahydrofuran (THF) was used as the eluent. We dissolved each sample (2 mg) in THF (1 mL) and injected it into the HPLC system using the auto injector. We used four kinds of polystyrenes as standards to draw a calibration curve correlating retention time with molecular mass.<sup>20</sup> In this study, GPC was used for the resins.

GCD is a technique for separating and analyzing components of a mixture based on their boiling points. In the same way that we can obtain a distillation curve by using a calibration curve correlating retention time and boiling point, we used a calibration curve correlating retention time and molecular mass to obtain the average and distribution of molecular mass. In this study, GCD was used for saturates and aromatics. We used gas chromatographs (Hewlett Packard/Agilent, USA and A.C. Analytical Controls, Netherlands) with a flame-ionization detector (FID). N-alkanes and alkylbenzenes were used as the standards for the saturates and aromatics, respectively.

**Structural types.** We conducted <sup>1</sup>H- and <sup>13</sup>C-NMR spectroscopic measurements using a Lambda 500 spectrometer (Japan Electron Optics Laboratory (JEOL), Japan). Details can be found in our previous study.<sup>15</sup> The structural types of hydrogen and carbon atoms were classified based on their chemical shifts. **The proportions** of these structural types could then be estimated on the basis of <sup>1</sup>H- and <sup>13</sup>C-NMR spectroscopies.

### **2.3. Measurement of Crude Oil Properties**

We measured the physical properties (density and viscosity) of the crude oil to validate our digital oil model. The density was measured at atmospheric pressure (0.1 MPa) and at 288 K using specific gravity bottle (pycnometer), Wadon (JIS K 2249 and JIS R 3503, Hokkai, Japan). We first filled the sample (50 mL) to the pycnometer and measured its weight. We then calibrated the density using the water equivalent of the pycnometer.

The kinematic viscosity was measured at atmospheric pressure (0.1 MPa) and at 303, 310.8, and

323 K using Ubbelohde viscometer (JIS Z 8803 and JIS Z 8809, Sibata Scientific Technology, Japan). Ubbelohde viscometer can measure a wide range of kinematic viscosity (0.3 to 100 000 mm<sup>2</sup>/s). We measured the time the sample (4.0 mL) flowed out through the capillary (inner diameter: 1.03 mm), and then, calculated the kinematic viscosity based on the Hagen-Poiseuille law.

## 2.4. Generation of Representative Molecules

Based on the results of the above experiments, we generated representative molecules using QMR. The method consisted of two processes: generation and optimization. In the generation process, each molecule was described in terms of following attributes: (1) number of unit sheets, (2) number of aromatic rings per unit sheet, (3) number of naphthenic rings per unit sheet, (4) number of alkyl chains attached to each unit sheet, and (5) length of each alkyl chain. These parameters were randomly sampled from each probability distribution function (PDF) by the Monte Carlo algorithm.<sup>10</sup> Note that the average values were estimated by the average structure analysis,<sup>21</sup> and the minimum and maximum values were determined based on previous work.<sup>15</sup> Molecules were then constructed by assembling building blocks according to their sampling parameters.<sup>10</sup> Here, building blocks are considered as a set of basic structures of molecules, and classified into two groups: aromatic sheets and aliphatic chains. The building blocks contained not only hydrocarbon, but also heteroatoms, such as nitrogen, sulfur, and oxygen. These building blocks were prepared based on the earlier studies.<sup>3,15,22-26</sup> We did not use carboxyl functional group (>C=O and -COOH) as one of the oxygen-containing building blocks because the <sup>13</sup>C-NMR results showed that our sample did not contain the carboxyl carbon.

In the optimization process, the objective function,  $F_1$ , given by Eq. (1),<sup>10</sup> was calculated for all combinations of the candidate molecules:

$$F_1 = \sum_{i=1}^7 \left( \frac{\mu'_i - \mu_i}{\sigma_i} \right)^2 + \frac{1}{11} \sum_{i=8}^{18} \left( \frac{\mu'_i - \mu_i}{\sigma_i} \right)^2, \quad (1)$$

where  $\mu'_i$  and  $\mu_i$  are the calculated and experimental values, respectively, and  $\sigma_i$  is the experimental error, which was taken from previous work.<sup>27</sup> All eighteen parameters ( $i = 1-18$ ) are listed in **Table 1**. The first seven parameters ( $i = 1-7$ ) provide elemental composition and number-average molecular mass, and the other parameters ( $i = 8-18$ ) provide the structural parameters determined from NMR spectroscopies. Among them,  $Q_1$  is the alkyl-substituted aromatic quaternary carbon,  $Q_2$  is the bridgehead aromatic quaternary carbon,  $C_1$  is the aromatic CH beside  $Q_2$ , and  $C_2$  is the aromatic CH except  $C_1$ . The details on the other parameters can be found in previous works.<sup>10,15,27</sup> The accuracy of structural parameters is relatively lower than that of elemental composition and molecular mass, so the contribution of those parameters to the objective function was weighted by a factor of 1/11.<sup>10</sup> The best combination that provided the lowest deviation was selected as the representative molecules.

## 2.5. Development of QMR

Because the QMR method was originally designed for asphaltenes and heavy fractions,<sup>10</sup> it could not provide non-aromatic molecules like saturates. In the case of the previously studied light oil, we applied QMR only to the heavy fraction and asphaltenes, but for heavy oil, we needed to apply QMR to all SARA fractions, including saturates. We therefore developed an appropriate QMR method in this study.

**Non-aromatic molecules.** We extended the QMR method to construct a model for the saturate fraction, which consists of monoaromatic and non-aromatic compounds. **Figure 1** shows a schematic of generation process, where one molecule is composed by three attributes: naphthenic rings and/or an aromatic ring (unit sheet), long side chains (main), and short branches of the main chains (sub). First, the numbers of aromatic and naphthenic rings were determined, and a unit sheet was selected among candidates prepared in advance (**Figure 1a**). Next, the number and the length of main chains were determined, which were attached on the unit sheet. After that, the number and the length of sub chains were determined, and all the sub chains were attached to the main chains. In the case of paraffins, no unit sheet was formed (**Figure 1b**), so only the main chain and sub chains were

configured. All the parameters, such as the number, the length, and the position of chains, were determined by the Monte Carlo algorithm with given minimum and maximum values for each parameter. The optimization process of the extended QMR was the same as the previous one mentioned above (Eq. (1)).

Using this method, we can select whether a particular unit sheet is present and whether it contains aromatic rings. Moreover, we can generate many branched structures by separating the chain bonding into main and subsidiary (“sub”) processes. As a result, it is now possible to generate non-aromatic molecules, such as paraffins and naphthenes, with more complicated chain structures than previous studies.

**Molecular mass distribution.** Our extended QMR method enables us to generate representative molecules for all fractions, including saturates. Although these representative molecules reproduced the number-average molecular masses, they could not reproduce the molecular mass distributions. To improve the accuracy of the models, especially for high proportions of saturates, we fitted the molecular mass distribution of the models to that obtained experimentally. First, we prepared thirty sets of model molecules, all of which provided sufficiently small  $F_1$  deviations. We then uniformly divided the molecular mass range into  $S$  grids and calculated the deviation between the model and the experimental values for each grid. Finally, we calculated the total deviations,  $F_2$ , and selected the best combination with the lowest deviation. Deviation  $F_2$  is expressed as follows:

$$F_2 = \sum_{i=1}^S (w'_i - w_i)^2, \quad (2)$$

where  $w'_i$  and  $w_i$  are the mole fractions of the model and experiments, respectively.

## 2.6. Molecular Dynamics Simulations

We calculated the physical properties of the digital oil by MD simulations. The GROMACS package (version 4.6.7) was used.<sup>28</sup> The CHARMM general force field (CGenFF) was used for describing the

molecular models.<sup>29-32</sup> It was shown in previous studies that the CGenFF (and the CHARMM force field) could reproduce the densities for organic liquids and the interfacial tension of oil–water interfaces very well with an uncertainty less than 1 % and 3 %, respectively.<sup>11,29,33</sup> The cutoff distances were 1.2 nm for both the Lennard–Jones and electrostatic potentials. The particle mesh Ewald summation method was used for long-range electrostatic interactions.<sup>34</sup> To remove excess potential energy prior to the calculation of density, we conducted an energy minimization step using the steepest descent method, after which we conducted an NPT (isothermal–isobaric) ensemble using the velocity-rescaling thermostat<sup>35</sup> and Berendsen barostat<sup>36</sup> until the system reached equilibrium. The density was then calculated using an NPT simulation with the Nosé–Hoover thermostat<sup>37,38</sup> and Parrinello–Rahman barostat<sup>39</sup> for 3 ns. The shear viscosity was calculated under NVT simulation (canonical ensemble) with the Nosé–Hoover thermostat for 5–20 ns. The temperature ranged from 403 K to 303 K, and the equilibrium volumes and coordinates were obtained from the NPT simulations for density calculation.

### 3. Results and Discussion

#### 3.1. Analytical Experiments

**Table 1** shows the weight percentages of crude oil fractions, **elemental composition**, number-average molecular mass, and **the proportions** of structural types of hydrogen and carbon atoms of each fraction. Asphaltenes were ignored in the digital oil because they comprised only 0.4 wt. %, and their yield was too low for their analytical determination. Because detailed characterization for a similar grade of heavy oil was not available, we **compared** our analytical results with those of bitumen.<sup>16</sup> The H/C ratios of saturates, aromatics, and resins were 1.63, 1.11, and 1.15, respectively, while those in the earlier study were 1.75–1.89, 1.40–1.46, and 1.38–1.47.<sup>16</sup> We thus found that the H/C ratios of our samples were lower than those of the samples used in the earlier study. In addition, the aromaticities of the saturates, aromatics, and resins were 14.7, **57.1**, and **48.4 %**, respectively, while values of 7.1–9.0, 35.9–37.7, and 36.2–43.0 % were previously reported.<sup>16</sup> As these values suggested, the aromaticities of our samples were higher than those of the samples used in the earlier study. These

results show that our crude oil sample is rich in cyclic hydrocarbons and aromatic carbons. The number-average molecular masses of the saturates, aromatics, and resins were 315, 473, and 515, respectively, while those of the earlier study were 324–380, 465–530, and 739–1010.<sup>16</sup> Although the prior study focused on bitumen, these values showed good agreement, especially for saturates and aromatics.

**Table 1** also indicates that 39.4 wt. % of all components were lost during SARA analysis. It is known that components with low boiling points can vaporize during this analysis, but it is impossible to analyze these directly. **We therefore** estimated the average molecular mass of the lost components from the molecular mass distribution obtained by GCD. **Figure 2** shows the molecular mass distributions of the saturates, aromatics, the crude oil sample, and the lost components. The molecular mass distribution of the lost components was calculated as that of the crude oil sample minus the saturates and aromatics. There were no aromatic molecules distributed at low molecule mass (< 260), so we assumed that the lost components were all saturates. **In fact, the weight percentages of carbon and hydrogen of the lost components determined by mass balance were 86.8 and 12.6 wt. %, which were almost the same as those of the saturates (87.9 and 12.0 wt. %).** **In this paper,** we assumed that the elemental composition and the structural parameters of the lost components were the same as those of the saturates. We thereby **inferred** analytical data for these lost components.

### 3.2. Construction of Digital Oil

**Based on the above results,** we generated representative molecules for each fraction using our extended QMR method. **In our study, a simple random distribution was chosen for saturates and a  $\gamma$  distribution<sup>10,15</sup> was chosen for PDFs of aromatics and resins.** The sampling parameters for the saturates, aromatics, and resins used in the generation process are listed in **Table 2** and **Table 3**. **For the lost components, we used the same sampling parameters as those of the saturates.** **Figure 3** shows the relationships between **the objective** deviation  $F_1$  and the number of optimized molecules,  $M$ , in the optimization process. All fractions achieved convergence when  $M$  was 5 or higher. **It was found that the deviation of resins was comparable to that of asphaltenes and heavy fractions in the earlier**

studies, with a range of 20 to 60.<sup>10,15</sup> Interestingly, we obtained much lower deviations for light fractions, which was presumably because the molecular structures of light fractions were less complicated than those of heavy fractions, resins, and asphaltenes. The optimized parameters of the model for each fraction are listed in **Table 1**, in comparison with those obtained from experiments.

**Figure 4** shows comparisons of the experimental molecular mass distributions with one and two or three QMR sets for saturates and aromatics. In the case of saturates, the deviation  $F_2$  of a single QMR set was at least 0.2; however, when we used three QMR sets, the deviation became 0.005. Similarly in the case of aromatics, the deviation  $F_2$  of two QMR sets was 0.005, while that of a single QMR set was at least 0.03. Finally, we constructed the digital oil as a mixture of these representative molecules, which contained 36 kinds of molecules and 917 molecules in total. Representative molecules of each fraction are discussed below.

**Saturates.** **Figure 5** shows the saturates model. Two naphthenic hydrocarbons and nine monoaromatics were selected as representative. Similar structures, such as naphthenes with alkyl chains and monoaromatics with naphthenic rings, were found in the earlier study.<sup>16</sup> The average number of unit sheets was 1.0, that of aromatic rings per unit sheet was 0.8, and the average chain length was 6.9. The aromaticity of the model was 21.8 %, while the experimental value was 14.7 %. Three QMR sets were selected for the distribution fitting. The  $F_1$  deviations of the sets were 5.9, 7.2, and 8.9.

**Lost components.** **Figure 6** shows the model of the lost components. Two isoalkanes, one naphthenic hydrocarbon, and five monoaromatics were selected as the representative of this fraction. It is noteworthy that similar structures were found in the database of our light oil model.<sup>15</sup> The average number of unit sheets was 0.9, that of aromatic rings per unit sheet was 0.7, and the average chain length was 5.1. The aromaticity of the model was 26.4 %. Two QMR sets were selected for the distribution fitting, having  $F_1$  deviations of 5.1 and 10.1.

**Aromatics.** **Figure 7** shows the aromatics model. One monoaromatic and eleven polyaromatics were selected as representative. The molecules contained a few heteroatoms, such as nitrogen, sulfur, and oxygen. Similar structures, such as polyaromatics with naphthenic rings and thiophenic sulfur, were

found in the earlier study.<sup>16,17</sup> The average number of unit sheets was 1.7, that of aromatic rings per unit sheet was 3.0, and the average chain length was 2.4. The aromaticity of the model was 58.1 %, while that of experiments was 57.1 %. Two QMR sets were selected for the distribution fitting, with  $F_1$  deviations of 3.1 and 4.1.

**Resins.** Figure 8 shows the resins model. Five polyaromatics were selected as these representative molecules. They contained more heteroatoms than those of the aromatic fraction. Similar structures, such as phenolic hydroxyl group and thiophenic sulfur, were found in the earlier studies.<sup>12,16,17,19</sup> In addition, we found that the heaviest two molecules resembled archipelago model for asphaltenes, while no molecules were similar to island model. The average number of unit sheets was 1.3, that of aromatic rings per unit sheet was 2.6, and the average chain length was 4.5. The aromaticity of the model was 54.9 %, compared with an experimental value of 48.4 %. Because of the low proportion of this class in the SARA fractions, we did not apply a distribution fitting. The  $F_1$  deviation of this model was 49.2.

**Plots of double-bond equivalent against carbon number.** Planar limits, defined as the lines generated by connecting maximum double-bond equivalent (DBE) values at given carbon numbers, have been proposed as a means of predicting and understanding the molecular structure of the compounds in crude oil.<sup>40</sup> For validation of each model, we plotted DBE as a function of carbon number and compared these with the planar limits. DBE represents the degree of unsaturation, which, for elemental formulae of  $C_cH_hN_nO_oS_s$ , is equal to the number of rings and double bonds involving carbon.<sup>25,40</sup>

$$\text{DBE} = c - h/2 + n/2 + 1 \quad , \quad (3)$$

Figure 9 shows the plots of DBE against carbon number for each model. The planar limits of the saturates (PSL), aromatics (PAL), and resins (PRL)<sup>40</sup> are also shown. All plots located below the planar limit of each fraction, which implied that the molecular model of each fraction was reasonable in terms of molecular structure.

### 3.3. Investigation of Crude Oil Properties

**Calculation of density.** First, we calculated the density of the digital oil over a wide range of temperature. The pressure was controlled at 0.1 MPa, and the temperature ranged from 403 K to 288 K. **Figure 10** shows the calculated density as a function of temperature. **The experimental datum (of the same oil sample) at 288 K is also shown. The calculated and measured values were in good agreement with each other,** having the values of 942.4 and 954.7 kg/m<sup>3</sup>, respectively. In addition, we found that there was no glass transition in this temperature range. Because the density as a function of temperature was close to linear, we compared the slope of the digital oil with that of other related chemicals for further validation. The density slope of the digital oil was  $-0.76 \text{ kg/m}^3 \text{ K}$ , while that of an asphalt model was  $-0.68 \text{ kg/m}^3 \text{ K}$ ,<sup>14</sup> that of an experimental asphalt value was  $-0.60 \text{ kg/m}^3 \text{ K}$ ,<sup>41</sup> and that of toluene was  $-0.98 \text{ kg/m}^3 \text{ K}$ .<sup>42</sup> Considering that crude oil is a mixture of kinds of **asphalts and light fractions**, the digital oil **was** reasonable for crude oil compounds in terms of its density slope.

**Calculation of viscosity.** **We then** calculated the shear viscosity of the digital oil by equilibrium MD (EMD) simulations.<sup>43–45</sup> Although non-equilibrium MD (NEMD) **method has** become increasingly popular for calculating the viscosities of liquids, it was confirmed that EMD **method** enabled us to obtain the viscosity with comparable accuracy and reliability to NEMD **method**.<sup>45</sup> In contrast to NEMD **method**, EMD **method does** not require additional adjustments of shear rate.<sup>45</sup> **Figure 11** shows the temporal-evolution profiles of shear viscosity for correlation time at different temperatures. The first plateau in the profile showed the value of shear viscosity used in the EMD simulations.<sup>43–45</sup> Longer calculation time was required for lower temperatures, because it was difficult to obtain plateaus under these conditions. At the lowest temperature (288 K), the profile **did not** show the plateau in our simulations, **so** we judged that we were not able to calculate the viscosity at this temperature using EMD method.

**We experimentally measured the kinematic viscosity of the crude oil at 303, 310.8, and 323 K to validate the calculated viscosity of the digital oil.** The kinematic viscosity is the ratio of the dynamic viscosity to the density,<sup>46</sup> as given by Eq. (4):

$$v = \frac{\mu}{\rho}, \quad (4)$$

where  $v$  is the kinematic viscosity,  $\mu$  is the dynamic viscosity, and  $\rho$  is the density. The calculated kinematic viscosity values at 303, 310.8, and 323 K were 71.8, 51.9, and 27.8 mm<sup>2</sup>/s, respectively, which were higher than experimental values (50.0, 32.5, and 18.8 mm<sup>2</sup>/s). However, the error appeared systematic, being a factor of ~1.5 higher than that of experiments. We considered several reasons why the calculated values were overestimated (as discussed below).

To further validate the viscosity, we tried fitting the calculated viscosity to equations that express the relationship between viscosity and temperature of liquids. Two well-known empirical expressions (with their respect theoretical bases) are the Arrhenius and Vogel–Fulcher–Tammann (VFT) equations.<sup>47</sup> The Arrhenius equation is a two-parameter correlation, given by Eq. (5):

$$\eta = A_s \exp\left(\frac{E_a}{RT}\right), \quad (5)$$

where  $\eta$  is the dynamic viscosity,  $T$  is the absolute temperature,  $A_s$  is the substance-dependent constants,  $E_a$  is the activation energy, and  $R$  is the universal gas constant.<sup>47</sup> It is considered that Arrhenius equation works well at low temperatures, but fails in the high-temperature range,<sup>47</sup> primarily because it is expressed by only two parameters. The VFT equation is therefore commonly used for analysis of experimental data of temperature dependence of viscosity. This is a three-parameter correlation, given by Eq. (6).<sup>47,48</sup>

$$\eta = \eta_0 \exp\left(\frac{DT_0}{T - T_0}\right), \quad (6)$$

where  $\eta$  is the dynamic viscosity,  $T$  is the absolute temperature,  $\eta_0$  is the pre-exponential factor (viscosity at infinite temperature),  $D$  is the fragility parameter, and  $T_0$  is the Vogel temperature

(ideal glass transition temperature).<sup>47,48</sup> The last three parameters are experimentally determined substance-dependent constants. The VFT equation can be applied for wide ranges of temperature, types of chemical compounds, and fields of research.<sup>47,48</sup> Thus, we used the VFT equation for correlation in this study. **Figure 12** shows a comparison of calculated viscosity with the VFT equation. **The estimated experimental shear viscosity values based on the experimental kinematic viscosity and the calculated density are also shown.** We used the least-squares method for fitting the parameters. The fitting parameter  $\eta_0$  was **0.41 cP**,  $D$  was **1.81**, and  $T_0$  was **224 K**. The parameter  $D$  relates to the liquid fragility, which reflects how quickly the viscosity increases with decreasing temperature. For strong liquids (*i.e.*, those that easily form glass, like SiO<sub>2</sub>), it takes a large value (typically  $D \geq 100$ ), while for fragile liquids, it takes a small value ( $D < 10$ ).<sup>48</sup> Thus, the value of  $D$  showed that our crude oil sample **was not a strong liquid**. In addition, the value of  $T_0$  showed that the glass transition temperature was far below 288 K. As mentioned above, the calculated density also showed that the glass transition temperature was outside the range of 288–403 K. Therefore, the calculated viscosity showed the consistency with the calculated density **in terms of glass transition temperature**.

Overall, the crude oil properties of density and viscosity were well reflected by the digital oil. In future, we will consider several possibilities to improve our digital oil model. The first issue is about the force field used in **MD simulations**. **As we mentioned in Section 2.6, we used CGenFF in this study because it well reproduced the experimental density in comparison with the other force fields, such as general AMBER force field (GAFF) and optimized potentials for liquid simulations (OPLS).**<sup>29,33</sup> However, it is not known *a priori* regarding the viscosity. Thus, we need to consider **using various force fields, and if necessary, tuning parameters for viscosity calculation**. The second issue is the optimization process where large errors of **the structural parameters** were used. **As we mentioned in Section 2.4, the contribution of the structural parameters to the objective function  $F_1$  was weighted by a factor of 1/11 because of the relatively low accuracy (Eq. (1)).** Therefore, even though the deviation was small, these structural parameters can differ considerably from experimental data. In fact, these differences caused the errors in aromaticity, especially for saturates. If the aromaticity of the saturates model approaches the experimental value, from 21.8 % to 14.7 %, more

paraffinic and naphthenic hydrocarbons will be generated as the representative molecules, which could lead to a decrease in the viscosity of the digital oil. The third issue is the simplified fitting of the molecular mass distributions. In this fitting, the molecular mass range was divided into uniform grids and the deviation was calculated for each grid; however, to improve the accuracy of the distribution, a finer classification is required in the higher mole fraction range. If we use denser grids, the accuracy of the distribution can be improved, which could lead to a decrease in viscosity.

In addition, we need to enrich the kinds of molecules that can be generated by QMR. The most essential point is to enable a variety of isomers of representative molecules to be considered in the digital oil. It is well known that viscosities may differ greatly between isomers even though these densities are close. In some cases, they may not be distinguishable from each other in structural parameters determined by NMR spectroscopies. Therefore, if we consider the isomers as alternatives for the representative molecules, the viscosity of the digital oil could change while keeping the density and the structural parameters. Moreover, considering isomers can also be helpful in terms of heteroatom compounds structures. In the case of nitrogen compounds, for example, pyrrole indicates an acidic or neutral property, while pyridine indicates a basic property; that is, indole and cyclopentapyridine have the same elemental compositions, the same molecular masses, and the same structural parameters, but different  $pK_a$  values because of the position of nitrogen atom. Regarding our digital oil, nitrogen existed only in five-membered aromatic ring like pyrrole and indole, which showed acidic or neutral properties. Thus, if we consider the isomers of nitrogen compounds, we can freely change  $pK_a$  values of these molecules, which could enable a wider variety of molecules to be considered in the digital oil.

#### 4. Conclusion

In this study, we expanded the digital oil technique to model heavy crude oil. In our expanded method, QMR was developed to be applicable to saturates. The first development was a new generation algorithm that allows our QMR method to generate non-aromatic molecules like saturates. In addition, it can also generate more branched structures by separating the chain bonding into main and

subsidiary processes. The second development was that we fitted the molecular mass distribution of the model to that obtained from experiments. As a result, we successfully constructed a digital oil for heavy crude oil comprising a mixture of 36 types of representative molecules.

To validate the digital oil, we first confirmed the validity of the model for each fraction in terms of the plots of DBE against carbon number. In addition, we calculated the density and viscosity of the digital oil by MD simulations, and confirmed good agreement with experimental data for crude oil. Furthermore, we compared the calculated viscosity with the VFT equation. The fitting parameter  $D$  showed that our crude oil sample **was not a strong liquid**. The fitting parameter  $T_0$  showed that the glass transition temperature was far below 288 K. The calculated density also showed that the glass transition temperature was outside the range of 288–403 K, thereby confirming consistency between the calculated **viscosity and density**. Consequently, we were able to confirm the validity of the digital oil in terms of molecular structure and physical properties. To further improve the digital oil, we could increase the accuracy of the aromaticity and the molecular mass distribution of the models in optimization process. In addition, we could consider a variety of isomers **to refine our models, which will lead to further investigation of crude oil properties**. Moreover, it will be very interesting to study **the digital oil with respect to fractions upon distillation,<sup>49,50</sup> and understand the nature of a super mixture behavior of a heavy crude oil; that is, why we could not characterize the individual molecules even for light fractions**.

## **Acknowledgement**

The authors thank the Japan Society for the Promotion of Science (JSPS) for a Grant-in-Aid for Scientific Research A (No. 24246148) and Grant-in-Aid for Scientific Research C (Nos. 16K06925 and 17K06988). We further acknowledge funding from Japan Petroleum Exploration Co., Ltd. (JAPEX). We also wish to acknowledge Dr. Shinya Sato of AIST, and Mrs. Hironori Imazato, Masanori Nakano, Hideyuki Nakashima, and Tomoya Ohata of JAPEX for technical support and insightful comments.

## References

- (1) Curtis C. *et al.* Heavy-Oil Reservoirs. *Oilfield Review* **2002**, *14*, 30–51.
- (2) Alboudwarej, H. *et al.* Highlighting Heavy Oil. *Oilfield Review* **2006**, *18*, 34–53.
- (3) Gray, M. R. *Upgrading Petroleum Residues and Heavy Oils*; Marcel Dekker: New York, 1994.
- (4) Mullins, O. C. *Petroleomics and Structure-Function Relations of Crude Oils and Asphaltenes. Asphaltenes, Heavy Oils, and Petroleomics*. Edited by Mullins, O.C.; Sheu, E.Y.; Hammami, A.; Marshall, A.G. Springer, **2007**, pp.1–16.
- (5) van Buuren, A.; Marrink, S.; Berendsen, H. A Molecular-Dynamics Study of the Decane Water Interface. *J. Phys. Chem.* **1993**, *97*, 9206–9212.
- (6) Jang, S. S.; Lin, S.-T.; Maiti, P. K.; Blanco, M.; Goddard, W. A., III. Molecular Dynamics Study of a Surfactant-Mediated Decane-Water Interface: Effect of Molecular Architecture of Alkyl Benzene Sulfonate. *J. Phys. Chem. B.* **2004**, *108*, 12130–12140.
- (7) Zhang, L.; Greenfield, M. L. Molecular Orientation in Model Asphalts Using Molecular Simulation. *Energy Fuels* **2007**, *21*, 1102–1111.
- (8) Zhang, L.; Greenfield, M. L. Analyzing Properties of Model Asphalts Using Molecular Simulation. *Energy Fuels* **2007**, *21*, 1712–1716.
- (9) Bresme, F.; Chacon, E.; Tarazona, P.; Tay, K. Intrinsic Structure of Hydrophobic Surfaces: The Oil-Water Interfaces. *Phys. Rev. Lett.* **2008**, *101*, 056102.
- (10) Boek, E. S.; Yakovlev, D. S.; Headen, T. F. Quantitative Molecular Representation of Asphaltenes and Molecular Dynamics Simulation of Their Aggregation. *Energy Fuels* **2009**, *23*, 1209–1219.
- (11) Kunieda, M.; Nakaoka, K.; Liang, Y.; Miranda, C. R.; Ueda, A.; Takahashi, S.; Okabe, H.; Matsuoka, T. Self-Accumulation of Aromatics at the Oil-Water Interface through Weak Hydrogen Bonding. *J. Am. Chem. Soc.* **2010**, *132*, 18281–18286.
- (12) Castellano, O.; Gimon, R.; Canelon, C.; Aray, Y.; Soscun, H. Molecular Interactions between Orinoco Belt Resins. *Energy Fuels* **2012**, *26*, 2711–2720.
- (13) Mikami, Y.; Liang, Y.; Matsuoka, T.; Boek, E. S. Molecular Dynamics Simulations of

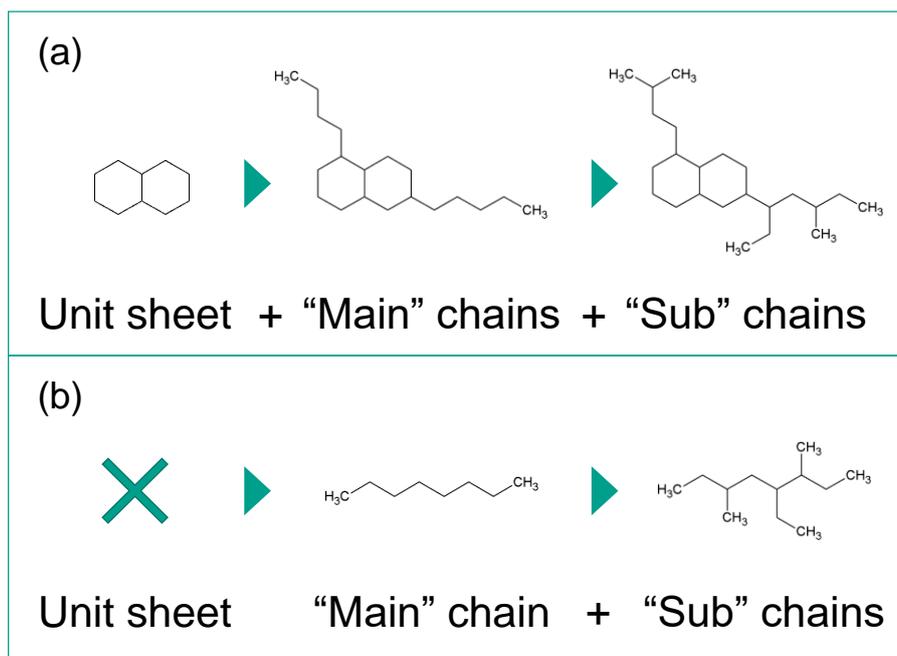
- Asphaltenes at the Oil-Water Interface: From Nanoaggregation to Thin-Film Formation. *Energy Fuels* **2013**, *27*, 1838–1845.
- (14) Li, D. D.; Greenfield, M. L. Chemical Compositions of Improved Model Asphalt Systems for Molecular Simulations. *Fuel* **2014**, *115*, 347–356.
- (15) Sugiyama, S.; Liang, Y.; Murata, S.; Matsuoka, T.; Morimoto, M.; Ohata, T.; Nakano, M.; Boek, E. S. Construction of Digital Oil for Investigation of Crude Oil Properties at Different Thermodynamic Conditions. *International Petroleum Technology Conference* **2016**, IPTC-18779-MS; *SPE Journal* **2017**, SPE-189465-PA.
- (16) Suzuki, T.; Itoh, M.; Takegami, Y.; Watanabe, Y. Chemical Structure of Tar-sand Bitumens by  $^{13}\text{C}$  and  $^1\text{H}$  N.M.R Spectroscopic Methods. *Fuel* **1982**, *61*, 402–410.
- (17) Jennings, P. W.; Pribanic, J. A.; Desando, M. A.; Raub, M. F.; Stewart, F.; Hoberg, J.; Moats, R.; Smith, J. A.; Mendes, T. M.; McGrane, M.; Fanconi, B.; VanderHart, D. L.; Manders, W. F. Binder Characterization and Evaluation by Nuclear Magnetic Resonance Spectroscopy. *SHRP-A-335*, Strategic Highway Research Program, Washington DC. **1993**.
- (18) Jones, D. R. SHRP Materials Reference Library: Asphalt Cements: A Concise Data Compilation. *SHRP-A-645*, Strategic Highway Research Program, Washington DC. **1993**.
- (19) Murgich, J.; Abanero, J. A. Molecular Recognition in Aggregates Formed by Asphaltene and Resin Molecules from the Athabasca Oil Sand. *Energy Fuels* **1999**, *13*, 278–286.
- (20) Sato, S.; Takanohashi, T.; Tanaka, R. Molecular Weight Calibration of Asphaltenes Using Gel Permeation Chromatography/mass Spectrometry. *Energy Fuels* **2005**, *19*, 1991–1994.
- (21) Sato, S. Fundamental of Structural Analysis for Heavy Oil: Method to Construct the Average Molecular Structure. *PetroTech* **2010**, *33*, 581–588.
- (22) Yamamoto, M.; Taguchi, K.; Sasaki, K. Basic Nitrogen Compounds in Bitumen and Crude Oils. *Chemical Geology* **1991**, *93*, 193–206.
- (23) Green, J. B.; Yu, S. K. T.; Pearson, C. D.; Reynolds, J. W. Analysis of Sulfur Compound Types in Asphalt. *Energy Fuels* **1993**, *7*, 119–126.
- (24) Gray, M. R.; Ayasse, A. R. Kinetics of Hydrodesulfurization of Thiophenic and Sulfide Sulfur

- in Athabasca Bitumen. *Energy Fuels* **1995**, *9*, 500–506.
- (25) Marshall, A. G.; Rodgers, R. P. *Petroleomics: Chemistry of the Underworld. Proc. Nat. Acad. Sci. U. S. A.* **2008**, *105*, 18090–18095.
- (26) Speight, J. G. *The Chemistry and Technology of Petroleum, Fifth Edition*; CRC Press: 2014.
- (27) Sheremata, J. M.; Gray, M. R.; Dettman, H. D.; McCaffrey, W. C. Quantitative Molecular Representation and Sequential Optimization of Athabasca Asphaltenes. *Energy Fuels* **2004**, *18*, 1377–1384.
- (28) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (29) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (30) Vanommeslaeghe, K.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) 1: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.
- (31) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) 2: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.
- (32) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D., Jr. Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulation. *J. Comput. Chem.* **2012**, *33*, 2451–2468.
- (33) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van del Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.
- (34) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth

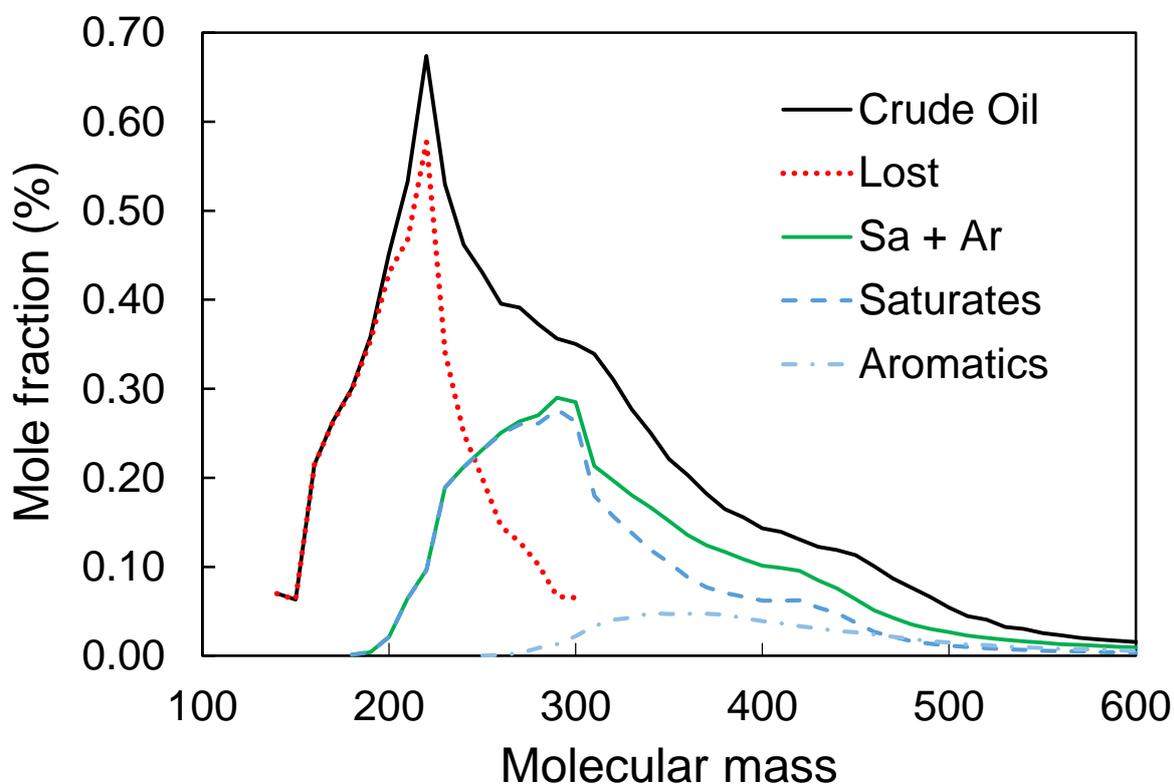
- Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (35) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Phys. Chem.* **1984**, *81*, 3684–3690.
- (37) Nosé, S. A molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- (38) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695–1697.
- (39) Parrinello, M.; Rahman, A. Strain Fluctuations and Elastic Constants. *J. Chem. Phys.* **1983**, *76*, 2662–2666.
- (40) Cho, Y.; Kim, Y. H.; Kim, S. Planar Limit-Assisted Structural Interpretation of Saturates/Aromatics/Resins/Asphaltenes Fractionated Crude Oil Compounds Observed by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 6068–6073.
- (41) Robertson, R. E. *et al.* *Fundamental Properties of Asphalts and Modified Asphalts. Vol. 1: Interpretive Report.* FHWA-RD-99-212. Federal Highway Administration, Washington DC. **1993**.
- (42) Lemmon, E. W.; McLinden, M. O.; Friend, D. G. "Thermophysical Properties of Fluid Systems" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Linstrom, P. J.; Mallard, W. G. *National Institute of Standards and Technology, Gaithersburg MD, 20899, doi:10.18434/T4D303, (retrieved December 15, 2017).*
- (43) Hess, B. Determining the Shear Viscosity of Model Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 209–217.
- (44) Zhang, L.; Greenfield, M. L. Relaxation Time, Diffusion, and Viscosity Analysis of Model Asphalt Systems Using Molecular Simulation. *J. Chem. Phys.* **2007**, *127*, 194502.
- (45) Chen, T.; Smit, B.; Bell, A. T. Are Pressure Fluctuation-Based Equilibrium Methods Really

Worse than Nonequilibrium Methods for Calculating Viscosities? *J. Chem. Phys.* **2009**, *131*, 246101.

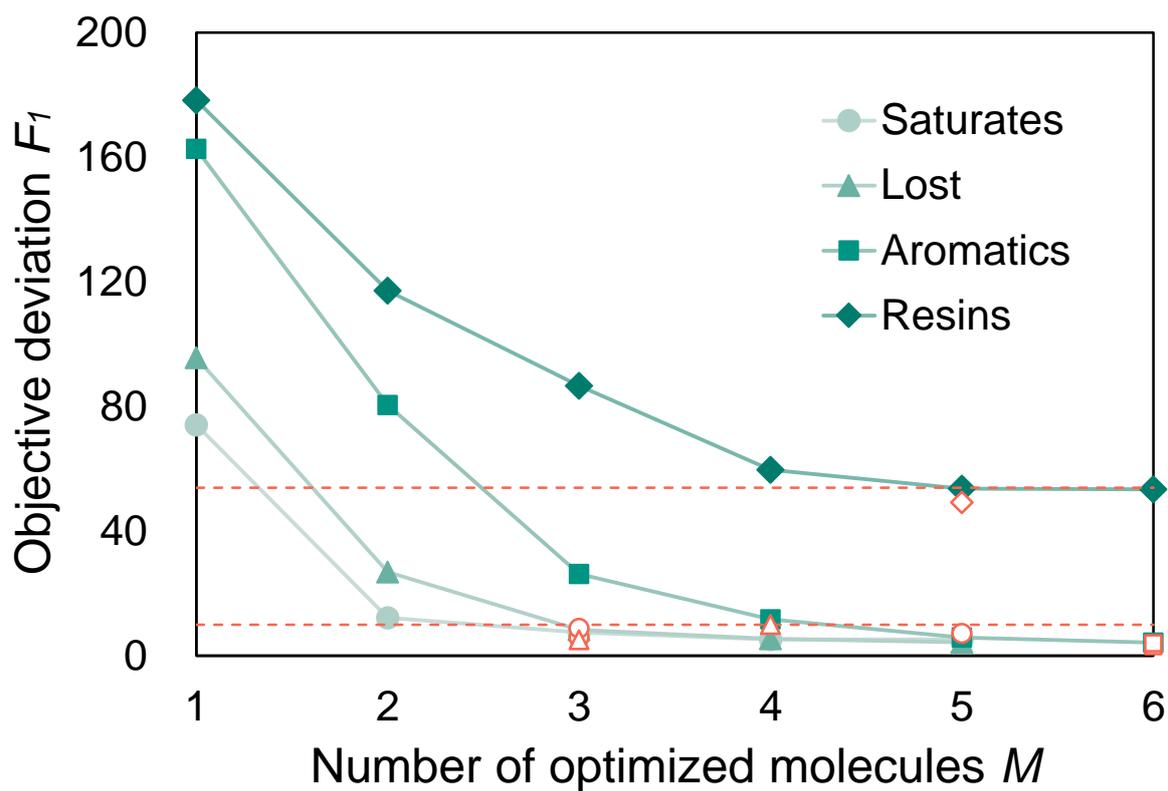
- (46) Pedersen, K. S.; Christensen, P. L. *Phase Behavior of Petroleum Reservoir Fluids*; CRC Press: 2006.
- (47) Tian, J.; Zhang, L.; Zheng, M. A New Correlation in Predicting Temperature-Dependent Viscosity of Saturated Liquids. *Mod. Phys. Lett. B* **2017**, *31*, 1750014.
- (48) Ikeda, M.; Aniya, M. Understanding the Vogel-Fulcher-Tammann Law in Terms of the Bond Strength-Coordination Number Fluctuation Model. *J. Non-Cryst. Solids* **2013**, *371-372*, 53–57.
- (49) Hudebine, D.; Verstraete, J.; Chapus, T. Statistical Reconstruction of Gas Oil Cuts. *Oil Gas Sci. Technol.* **2011**, *66*, 461–477.
- (50) Alvarez-Majmutov, A.; Gieleciak, R.; Chen, J. Deriving the Molecular Composition of Vacuum Distillates by Integrating Statistical Modeling and Detailed Hydrocarbon Characterization. *Energy Fuels* **2015**, *29*, 7931–7940.



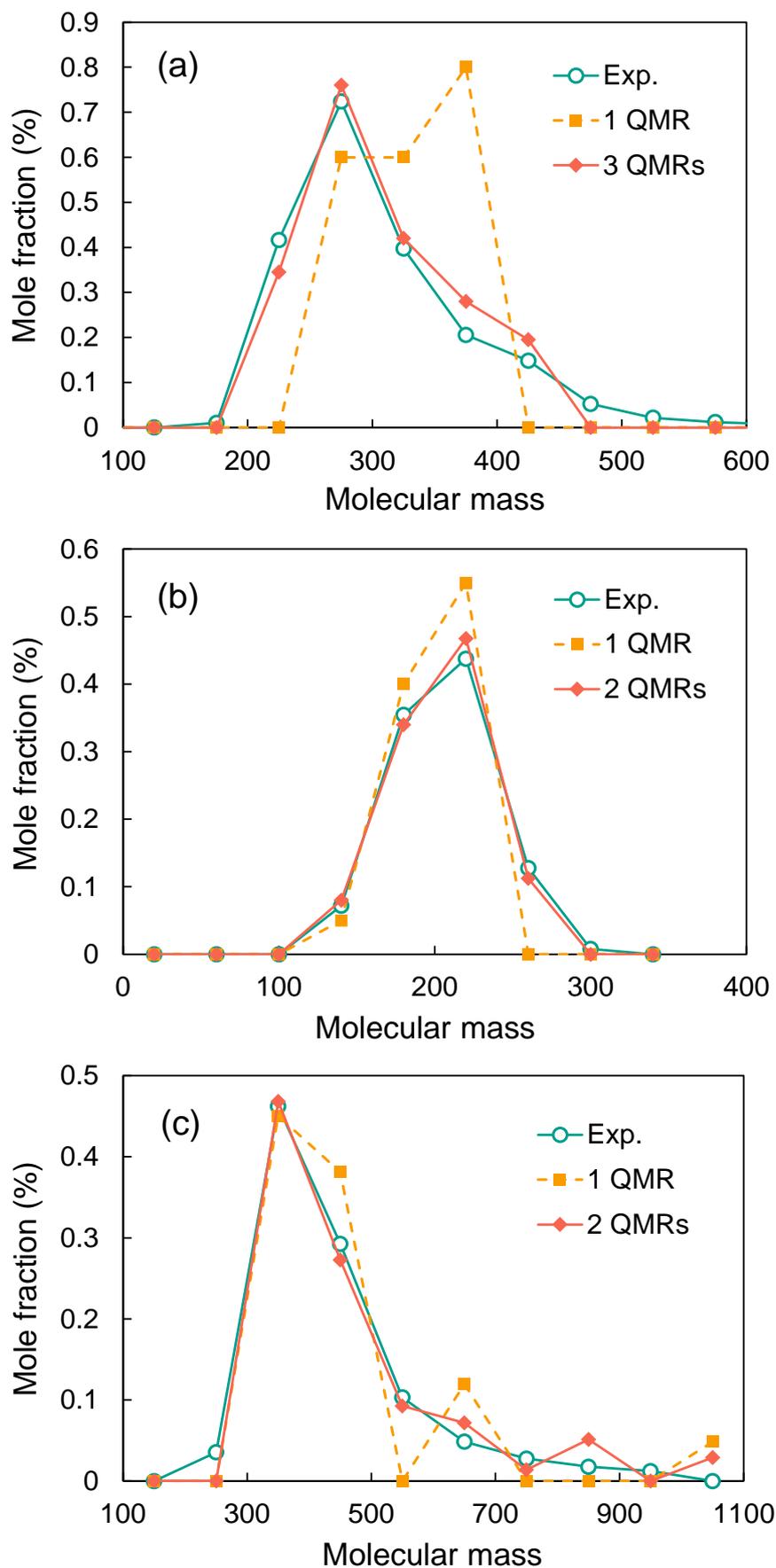
**Figure 1.** Schematic of generation algorithm in our extended QMR method. (a) Unit sheet is present and there are no aromatic rings; (b) unit sheet is not present.



**Figure 2.** Molecular mass distributions of crude oil sample, saturates, aromatics, and lost components obtained by gas chromatography distillation.

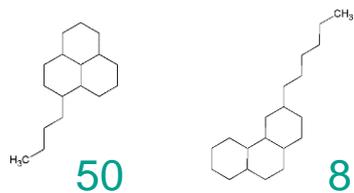


**Figure 3.** Relationships between **objective deviation** and number of optimized molecules. The dashed lines show thresholds: the upper threshold was set to 54 for resins; the lower was set to 10 for saturates, lost components, and aromatics. Open symbols show the values of our models.

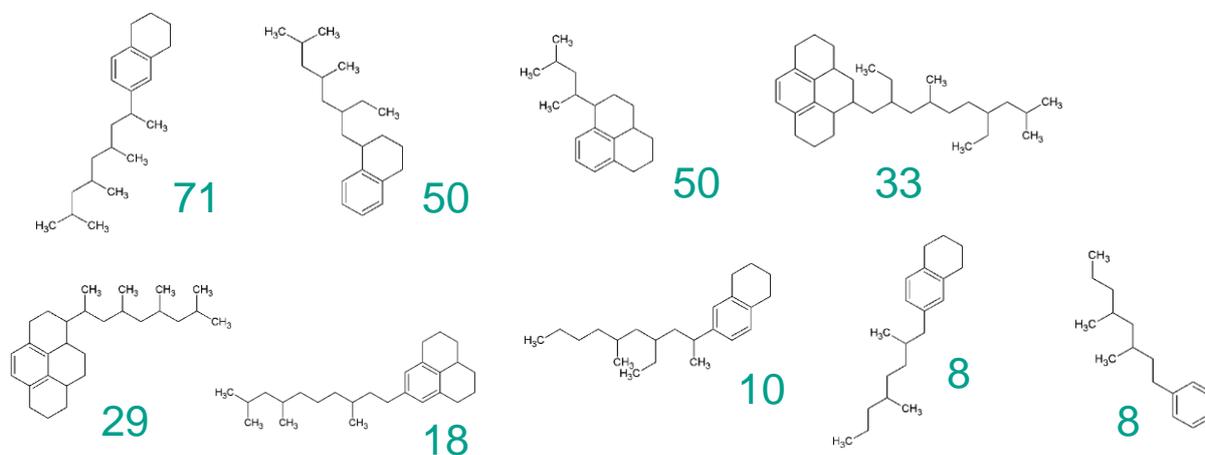


**Figure 4.** Comparisons of experimental molecular mass distributions with those determined using one QMR set and two or three QMR sets: (a) Saturates; (b) lost components; (c) aromatics.

(a) Naphthenic

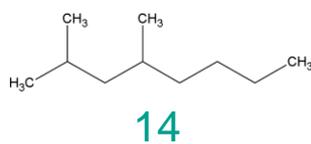
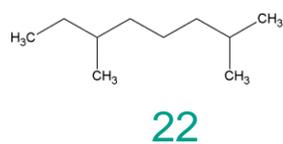


(b) Monoaromatic

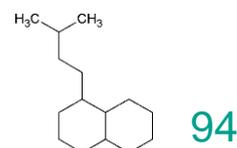


**Figure 5.** Representative molecules of saturates. The number beside each molecule shows the number of that molecule contained in the digital oil.

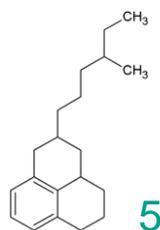
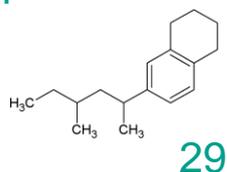
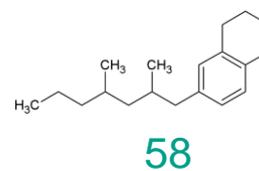
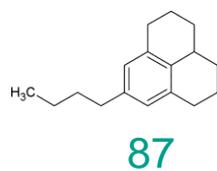
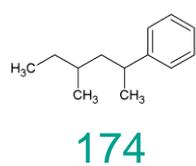
(a) Isoalkane



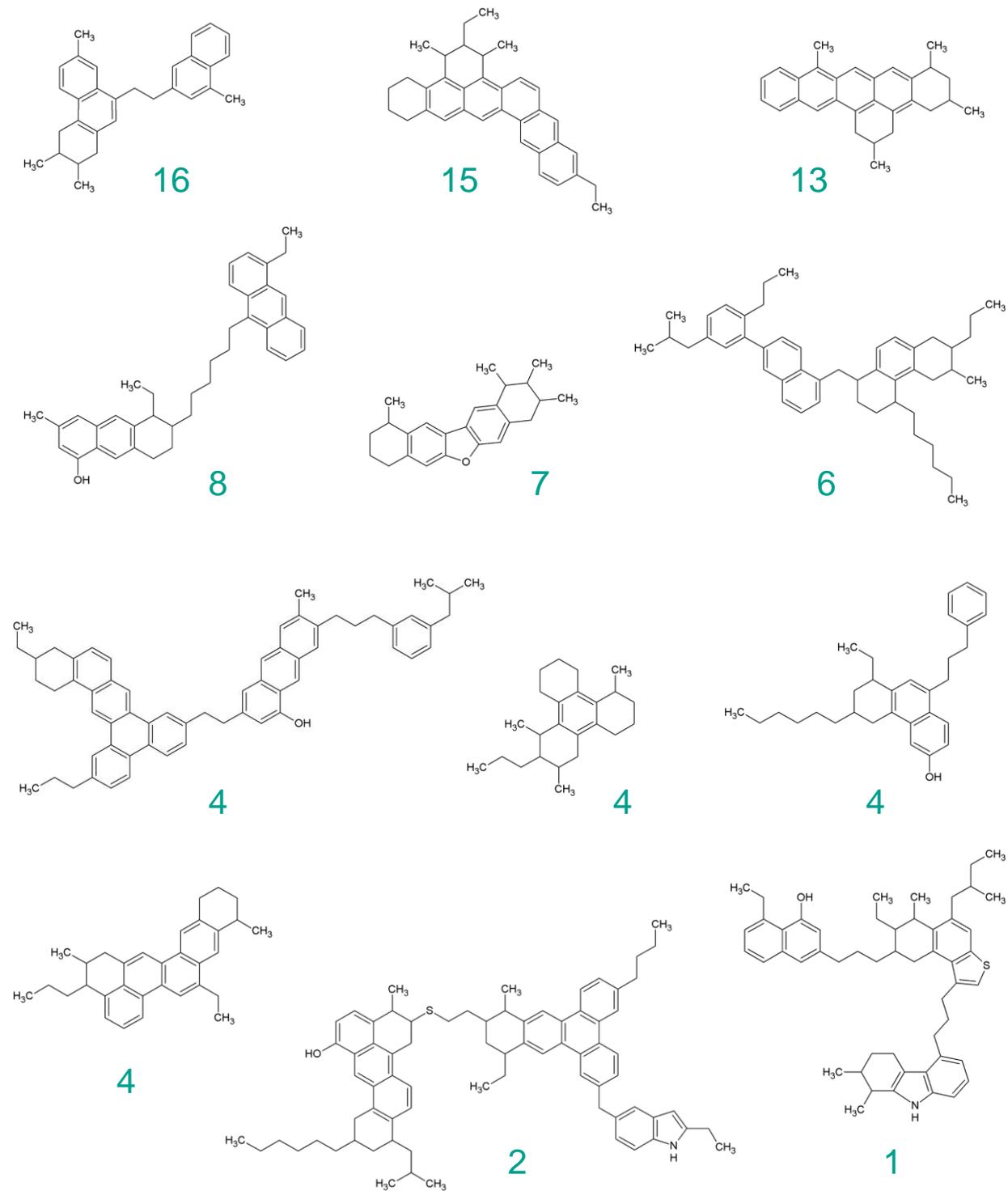
(b) Naphthenic



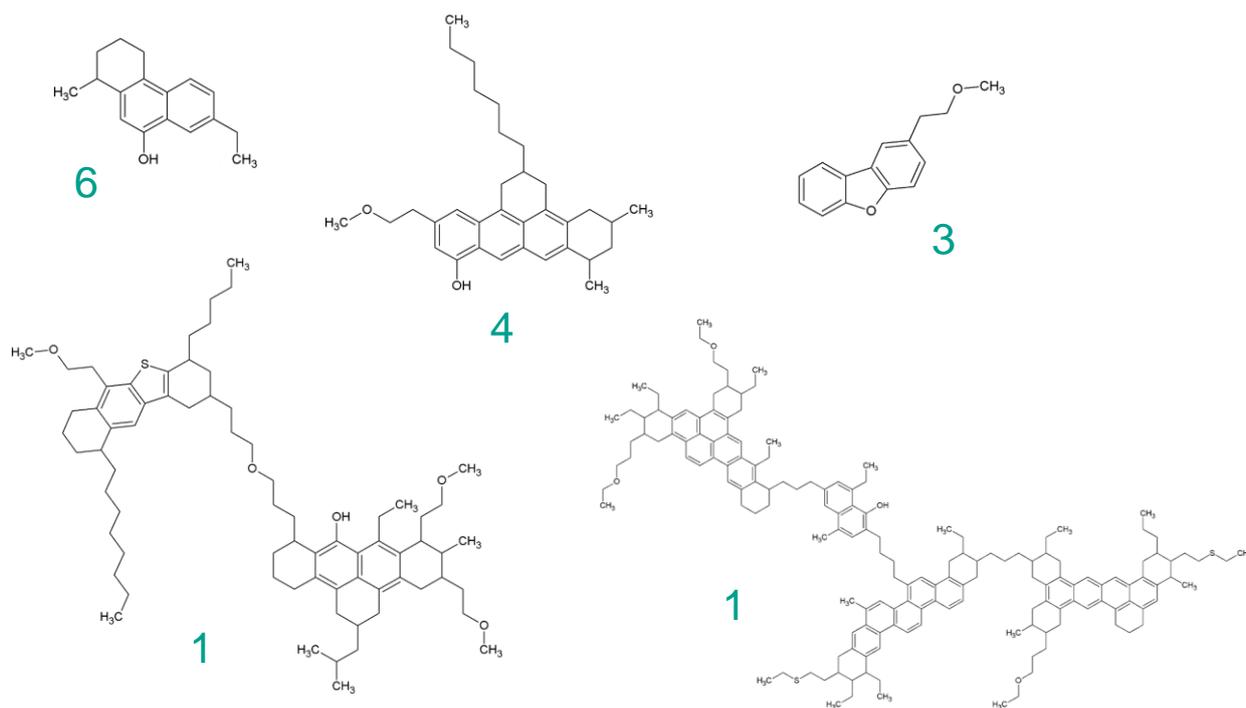
(c) Monoaromatic



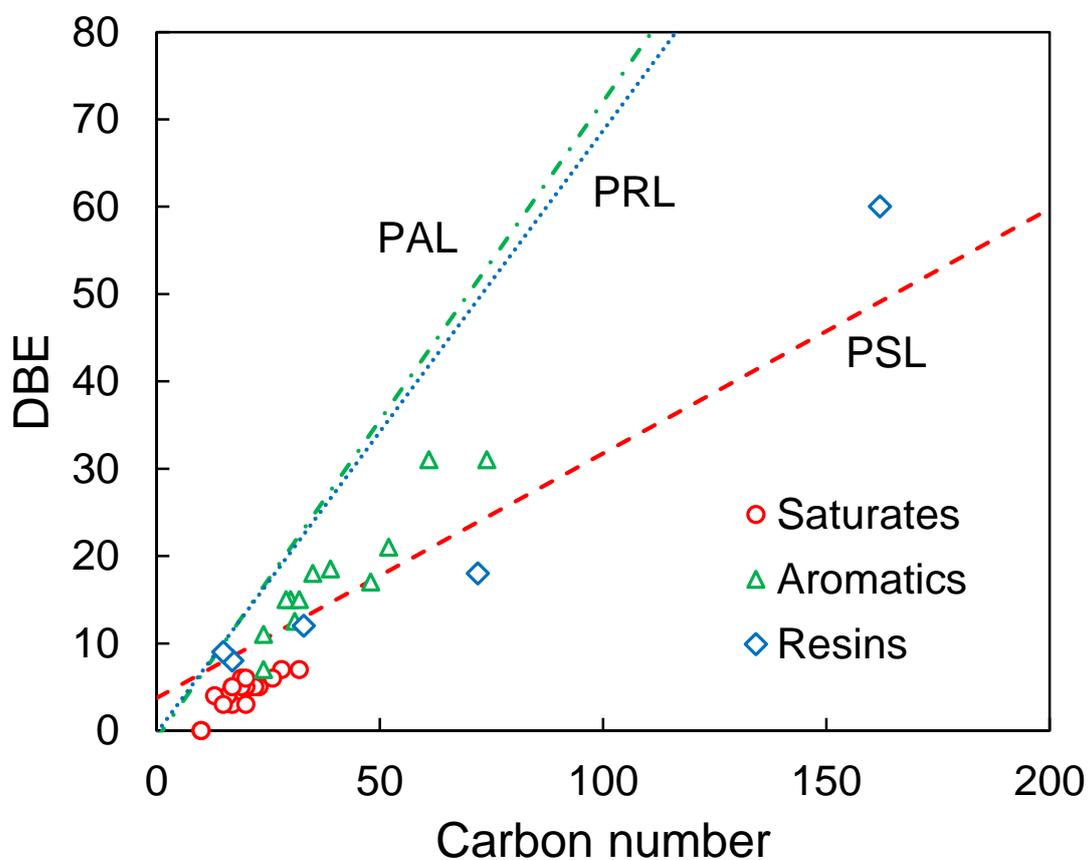
**Figure 6.** Representative molecules of lost components. The number beside each molecule shows the number of that molecule contained in the digital oil.



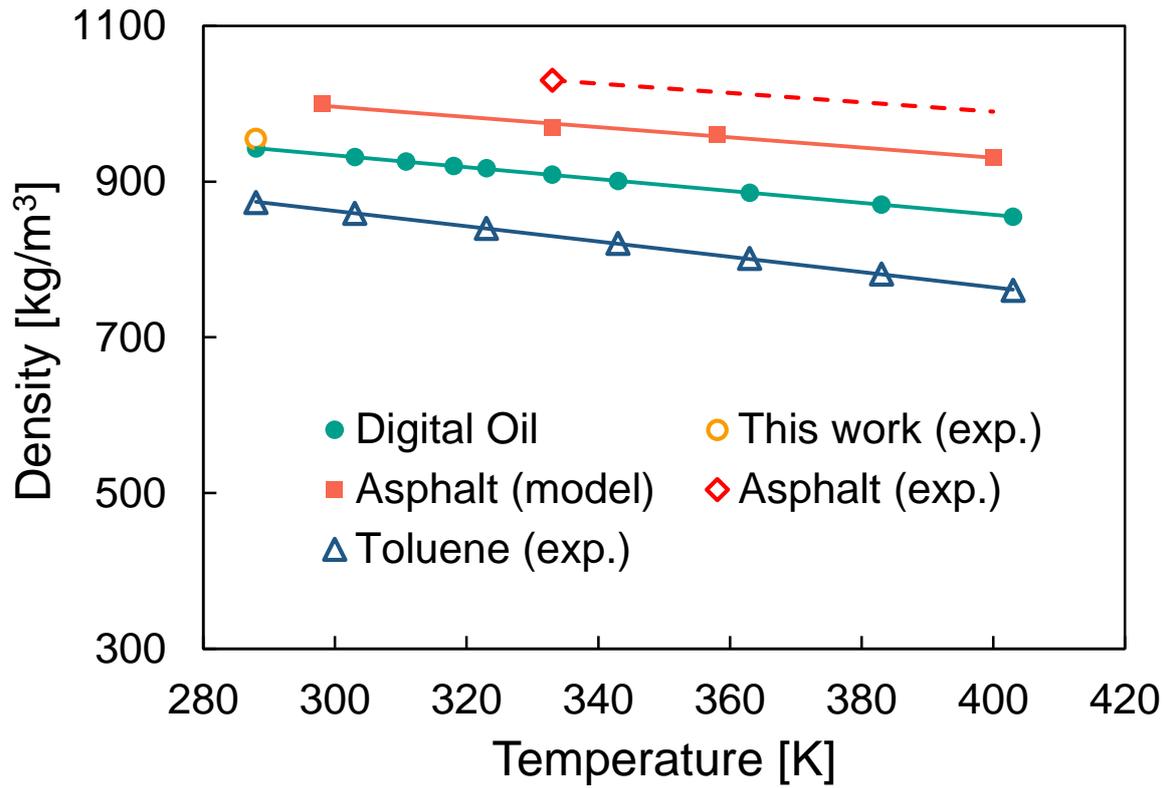
**Figure 7.** Representative molecules of aromatics. The number beside each molecule shows the number of that molecule contained in the digital oil.



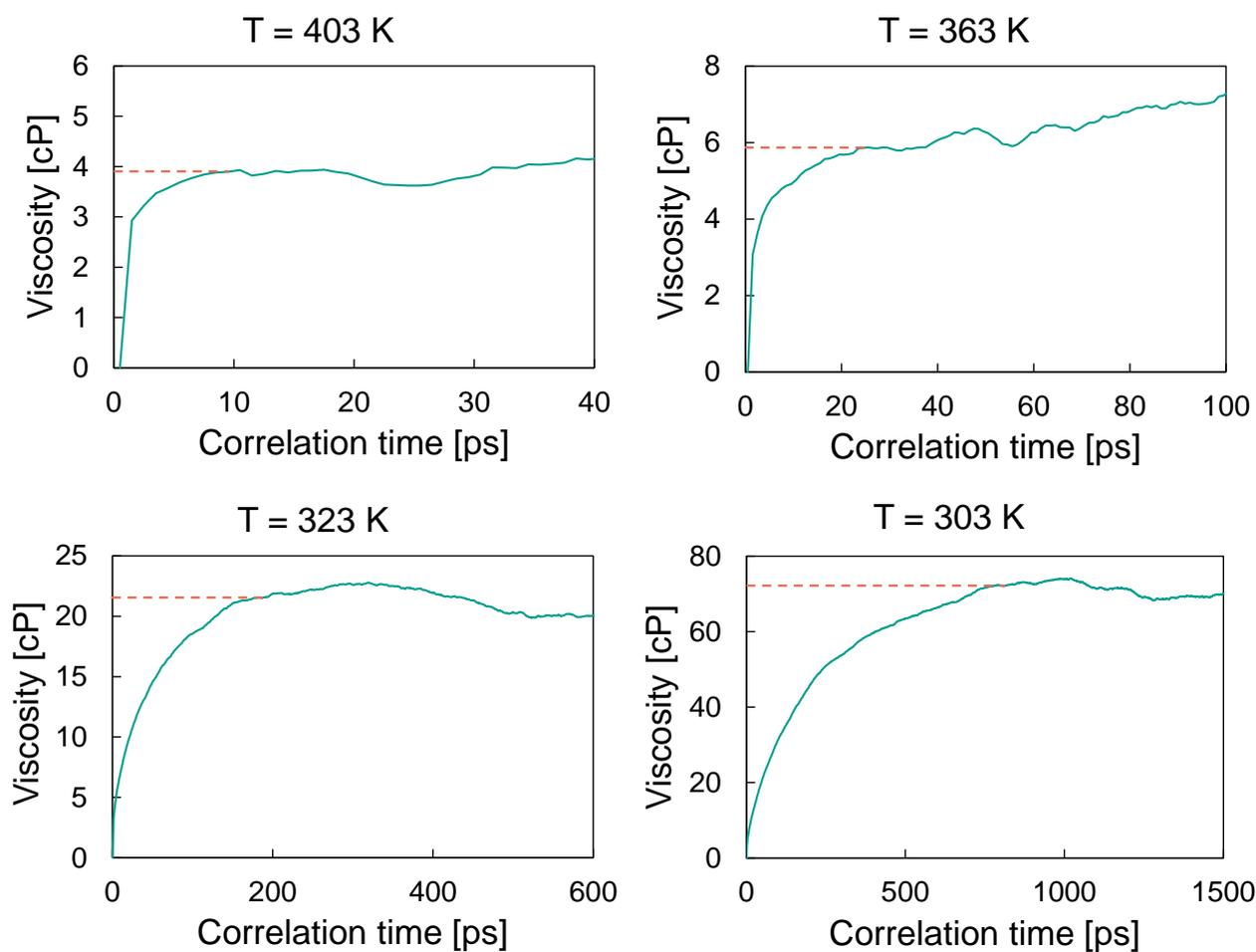
**Figure 8.** Representative molecules of resins. The number beside each molecule shows the number of that molecule contained in the digital oil.



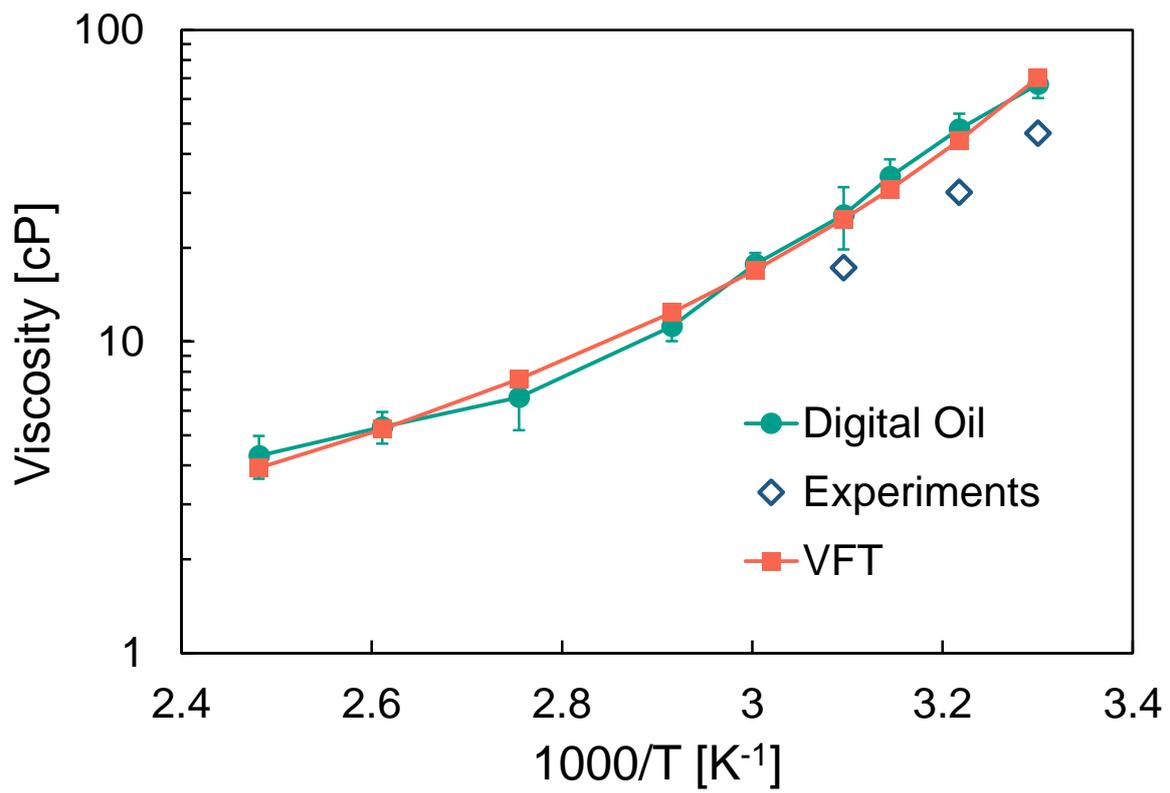
**Figure 9.** Plots of double-bond equivalent (DBE) as a function of carbon number of saturates, aromatics, and resins. The planar limit for each is also shown as a line.



**Figure 10.** Calculated density as a function of temperature. An experimental datum of our crude, calculated values of an asphalt model, and experimental values for asphalt and toluene are also shown for comparison.



**Figure 11.** Temporal-evolution profiles of viscosity for correlation time in Equilibrium MD simulations. The dashed lines show the first plateaus, at which the values of viscosity were taken.



**Figure 12.** Calculated shear viscosity as a function of temperature. Available experimental data and the Vogel-Fulcher-Tammann equation are also shown for comparison.

**Table 1.** Comparison of experimental and model parameters for crude oil fractions. According to NMR spectroscopies, olefinic carbons were 0.0 % for saturates, 0.3 % for aromatics, and 0.5 % for resins, respectively. They were ignored in our model.

No.	Fraction	Saturates		Lost components		Aromatics		Resins	
	wt. %	41.1		39.4		16.1		3.0	
	Property	Exp.	Model	Exp.	Model	Exp.	Model	Exp.	Model
1	C (wt. %)	87.9	88.0	-	88.1	90.3	90.3	80.0	83.4
2	H (wt. %)	12.0	12.0	-	11.9	8.4	8.4	7.7	8.2
3	N (wt. %)	0.0	0.0	-	0.0	0.1	0.1	0.7	0.0
4	S (wt. %)	0.0	0.0	-	0.0	0.2	0.2	0.4	0.4
5	O (wt. %)	0.1	0.0	-	0.0	1.0	1.1	11.3	8.0
6	V (wt. %)	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0
7	Mn	315	304	213	203	473	472	515	500
	Mw	345	316	219	209	595	523	1174	1053
8	Q <sub>1</sub> (%)	6.3	10.7	-	10.1	16.1	17.1	12.4	17.4
9	Q <sub>2</sub> (%)	1.2	0.0	-	0.0	8.7	15.9	12.8	14.7
10	C <sub>1</sub> (%)	1.3	0.0	-	0.0	20.2	14.9	12.4	13.4
11	C <sub>2</sub> (%)	5.9	11.1	-	16.3	12.1	10.2	10.8	9.4
	Carbon aromaticity (%)	14.7	21.8	-	26.4	57.1	58.1	48.4	54.9
12	Other aliphatic CH <sub>2</sub> (%)	23.1	35.3	-	31.9	10.6	17.2	11.7	22.0
13	Total $\alpha$ -CH <sub>3</sub> (%)	13.2	10.8	-	12.1	9.5	8.8	8.0	4.3
14	Total $\beta$ -CH <sub>3</sub> (%)	3.6	2.2	-	3.2	1.8	2.0	2.4	2.6
15	Chain CH <sub>2</sub> (%)	0.5	0.5	-	0.5	1.2	1.2	4.1	3.1
16	Aliphatic CH (%)	34.7	20.0	-	15.4	15.1	8.8	18.7	6.6
17	Naphthenic CH <sub>2</sub> (%)	7.3	7.2	-	8.0	2.9	2.5	4.1	2.6
18	$\gamma$ -CH <sub>3</sub> (%)	2.9	2.2	-	2.5	1.5	1.4	2.1	3.9

**Table 2.** Sampling parameters for the saturates.

	Saturates			
	Non-aromatics		Monoaromatics	
	Min	Max	Min	Max
Number of naphthenic rings	0	3	0	3
Number of main chains	1	4	1	1
Length of main chains	4	10	4	10
Number of sub chains	0	10	0	10
Length of sub chains	1	2	1	2

**Table 3.** Sampling parameters for the aromatics and resins. Note: three parameters (average, minimum, and maximum values) are necessary to generate a  $\gamma$  distribution.

	Aromatics			Resins		
	Ave.	Min	Max	Ave.	Min	Max
Unit sheets (US)	1.6	1	4	1.3	1	4
Aromatic rings / US	2.1	1	6	2.9	1	8
Naphthenic rings / US	1.3	0	6	2.3	0	8
Alkyl chain length	2.3	1	8	2.6	1	10
Substitution of aromatic atoms with alkyl chains (%)	11.9	0	34	14.2	0	34
Substitution of naphthenic atoms with alkyl chains (%)	43.5	0	84	43.2	0	84