Simulation of Asphaltene Aggregation through Molecular Dynamics: Insights and Limitations

T. F. Headen, † E. S. Boek, † G. Jackson, † T. S. Totton, ‡ and E. A. Müller*∥§

†Department of Chemical Engineering, Imperial College London, London SW7 2AZ, U.K.
‡ISIS Neutron and Muon Source, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Campus, Didcot OX11 0QX, U.K.
∥Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, U.K.
§BP Exploration Operating Co. Ltd., Sunbury-on-Thames TW16 7LN, U.K.

ABSTRACT: We report classical atomistic molecular dynamics simulations of four structurally diverse model asphaltenes, a model resin, and their respective mixtures in toluene or heptane under ambient conditions. Relatively large systems (∼50,000 atoms) and long time scales (>80 ns) are explored. Wherever possible, comparisons are made to available experimental observations asserting the validity of the models. When the asphaltenes are dissolved in toluene, a continuous distribution of cluster sizes is observed with average aggregation number ranging between 3.6 and 5.6, monomers and dimers being the predominant species. As expected for mixtures in heptane, the asphaltene molecules tend to aggregate to form a segregated phase. There is no evidence of the distinct formation of nanoaggregates, and the distribution of clusters is found to be continuous in character. Analysis of the shape of the clusters of asphaltenes suggests that they are generally spherical in character, with the archipelago models favoring longer prolate structures and the continental model tending toward oblate structures. The aggregates are seen to be diffuse in nature, containing at least 50% solvent on average, being denser in heptane than in toluene. Mixtures of asphaltenes with different architecture are found to have cluster properties that are intermediate between those of the individual components. The presence of resins in the mixture does not appear to alter the shape of the asphaltene aggregates or their size or density when toluene is the solvent; on the other hand, the resins lead to an increase in the density of the resulting aggregates in heptane. Quantification of these observations is made from histograms of the cluster distributions, potential of mean force calculations, and an analysis of the shape factors. We illustrate how the time scales for complete aggregation of molecules in heptane are longer than the longest of the simulations reported in the open literature and as an example report a long simulation (0.5 μs) that fails to reach an equilibrium state, suggesting that acceleration techniques (e.g., using coarse-grained models) are needed to appropriately explore these phenomena.

1. INTRODUCTION

Asphaltenes are a solubility fraction of crude oils, defined as being the part of the crude insoluble in n-alkanes (usually heptane) and soluble in aromatic solvents (usually toluene). Asphaltenes are among the most polar and heaviest (in terms of their molecular size) components of mixtures of carbonaceous fluids and hence are prone to aggregation, flocculation, and deposition. Aggregation is an extremely complex process and may be triggered by reductions in pressure and temperature and/or by blending with other incompatible oils and/or by the incorporation of gases such as CO₂. The study of the phase behavior and aggregate structures of carbonaceous deposits is an issue of continuous importance throughout the oil industry, particularly in the area of flow assurance.1

The complexity of the crude oil mixture in terms of the number of possible individual components makes the experimental analysis an ongoing tour de force that, although enlightening, has not yet provided the community with unequivocal answers as to the aggregation mechanisms and physical behavior of asphaltenes in solution. Models for asphaltene aggregation are plentiful and conflicting and span decades of research. For example, the “Yen–Mullins” model is a well-cited interpretation of asphaltene aggregation predicting that asphaltenes form small and dense nanoaggregates, which in turn aggregate loosely into clusters. While the hierarchical structure of two aggregate sizes seems to explain, for example, the discrepancy between some experimental results4,5 that indicate an aggregation number of ∼4 to 8 and X-ray and neutron scattering results6,7 that show larger, more diffuse clusters, the model has not been universally accepted or confirmed. The effect of concentration on the phase behavior is also nontrivial. Asphaltenes may be present at parts per million levels in the live unstable crudes, generating troublesome organic deposits; however, Boscan crudes are known to contain 17% asphaltenes and are stable with no deposits formed at any stage of production, while highly paraffinic crudes with asphaltene contents of 0.1% or less present serious production issues. It is surprising that in spite of more than half a decade of experimental and theoretical modeling, questions still remain in terms of the description and interpretation of the most
fundamental details of asphaltene precipitation. Molecular simulation provides a potential capability to expand our knowledge of the aggregation and deposition mechanisms of asphaltenes through an increased understanding of aggregate structure and energetics and is the topic of our current contribution.

To be able to successfully simulate systems containing asphaltenes, it is vital to have an accurate description of the molecular structure of asphaltenes. This is a challenging step, as asphaltenes represent a whole class of molecules rather than an individual species. In general, an assumption is made that a molecule, or a small group of molecules, that have chemical characteristics similar to those of asphaltenes, can represent the continuum of asphaltene molecular structures. The crux is to obtain molecular structures that faithfully represent the entire chemical family. There are very few systematic methodologies to propose sensible asphaltene structures, and the folklore is based on experimental information that is only marginal and in many instances inconclusive and contradictory. Special mention is made of the quantitative molecular representation (QMR) methods that have been used to propose representative structures of asphaltenes in order to remove some of the empiricism surrounding this aspect of the problem.

We first review the state of the art of asphaltene simulation and the types of molecular structures that have been used to date. Atomistic simulation of asphaltene aggregation is then presented at a size and time scale that is at the limit of what can be reasonably achieved using modern multicores processors with GPU acceleration of a parallel molecular dynamics (MD) code. Finally, we discuss the insight that simulation can give regarding the aggregation mechanisms for asphaltene in good and bad solvents, the limitations of this atomistic approach, and potential methods for increasing the size and length scales accessible through simulation.

2. BACKGROUND

2.1. Atomistically Detailed Simulation Studies. In recent years, as high-performance computing has become more prevalent, there has been an increasing effort to understand the structure of asphaltene aggregation at the molecular level using molecular simulation. The earliest simulation papers were constrained by the computational resources available and as a consequence were limited to considering single model asphaltene molecules or asphaltenes without explicit solvent molecules in what today would be considered rudimentary systems, not because of the scientific content but rather as result of the inexorable advance of Moore’s law. In addition, only relatively short simulation times (up to ~500 ps) were achievable, and therefore, dynamic properties such as diffusion or the observation of spontaneous aggregation were not possible. Instead, the early studies tended to use a combination of energy minimization and “manual docking” to explore different conformations for asphaltene aggregation and examined data during a short MD run in this conformation.

One of the earliest studies, by Rogel, investigated asphaltene aggregation through molecular dynamics simulations of two model asphaltene structures. Asphaltenes aggregates were formed through a combination of molecular mechanics and dynamics to produce reasonable low-energy structures. Solubility parameters where calculated from monomers and aggregates of the model asphaltenes, showing a reduction in the solubility parameter with increasing aggregate size. Some simulations were also carried out with a small number (45) of toluene or heptane solvating molecules. The interaction energies between asphaltene molecules were calculated as −160 and −40 kJ/mol in heptane and toluene, respectively, showing qualitative agreement with the solubility definition but suggesting a rather large difference (>100 kJ/mol) in the solvation energies of the respective solvents. A later study considered a larger diversity of asphaltene model molecules in order to estimate asphaltene densities. The simulations tended to overestimate the densities compared with experiment. It should be noted that all of the asphaltene structures used had a relatively large number (8–22) of condensed aromatic rings.

In another pioneering study, Murgich et al. conducted molecular mechanics calculations using relatively large “continental” asphaltene structures (with 24 aromatic rings) in conjunction with resins and showed that aggregation mainly occurred as a result of the stacking of the polyaromatic cores of the asphaltene. A later study using a smaller model asphaltene molecule with seven fused rings in its polyaromatic core led to similar findings. The structures used in both of these studies were built to be consistent with measured spectroscopic data for asphaltenes from Venezuelan crude and Athabasca bitumen.

Pacheco-Sánchez et al. used MD to show that asphaltene aggregation occurs spontaneously even for smaller asphaltene molecules forming dimers, trimers, and tetraters during a short 100 ps simulation. The structures of the aggregates formed displayed no overriding structure type: face-to-face, offset-stacked, and T-shaped aggregates were observed. An asphaltene structure in the solid phase was also examined, comparing the simulated structures to experimental scattering data. Four different asphaltene structures were chosen from the literature: models were extracted from the work of Groenzin and Mullins, Murgich et al., Zajac et al., and Speight, each formulated from analytical study of asphaltenes. All four simulation studies replicated the experimentally determined structure factor to reasonable accuracy. In a further study, the Groenzin–Mullins asphaltene model was simulated in a box of explicit solvent, and the effect of pressure on aggregation was investigated. Finally, Vicente et al. adopted the same simulation approach to provide the enthalpy of mixing and cohesive energies for the Groenzin–Mullins model molecule in order to determine the Hildebrand solubility parameter, closely matching experimentally determined results.

Simulations of asphaltene dimers with explicit solvent molecules conducted over 100 ps by Carauta et al. have shown that the asphaltene dimers bind face-to-face at distances of 3.6 Å in heptane and 5 Å in toluene, matching the solubility trend. Furthermore, they show that the effect of increasing the temperature (from 323 to 573 K) is to decrease the distance in the asphaltene dimer. A further study used molecular mechanics followed by quantum-mechanical calculations to study the structure of asphaltene dimers in solution using smaller model compounds (with approximately six aromatic rings) containing oxygen heteroatoms.

In general, these early studies are limited by either the lack of explicit solvent or the short time scales over which structural and thermodynamic data were collected. However, the results are qualitatively consistent with the solubility behavior of asphaltenes and indicate that face-to-face stacking of the aromatic cores is a primary structural characteristic, as shown through X-ray diffraction from solid asphaltene samples. In more recent years, increased computing power has led to the study of larger model systems being simulated for longer
periods of time, bringing the simulations a step closer to the experimental systems.

Greenfield and co-workers have written a series of papers on the simulation of model asphalt described as a ternary mixture containing model asphaltenes, resins, and maltenes. The initial studies used model asphaltenes from Groenzin and Mullins and from a study by Artok et al. employing pyrolysis–GC–MS, NMR spectroscopy, gel-permeation chromatography, and MALDI-TOF-MS to study asphaltene structure and derive model asphaltenes. The other components of the asphalt were modeled as dimethylnaphthalene and the alkane n-C_{20} for the aromatic and saturate fractions of the crude oil, respectively. The simulations were used to estimate the dynamical properties of the model asphalt such as bulk viscosity and diffusion of the molecular components. This relatively simplistic simulation was able to predict experimentally determined viscosities to within an order of magnitude. A detailed analysis of the orientational structure within the mixture was obtained by considering the angle between the asphaltene aromatic planes as a function of the radial distribution function. The addition of a polymer modifier (polystyrene) was also examined. More recent studies have used a modified model system. Adapted versions of asphaltene molecules from Mullins were used as the asphaltene component—the positions of the alkyl chains on the asphaltene core were changed to reduce the high internal energies caused by steric repulsion. Polar molecules were also used to more realistically model the aromatic component of the asphalt. A similar analysis of the dynamical properties of the asphalt was conducted with this improved model system.

Given that there is a continuing level of uncertainty in the general molecular structural characteristics of asphaltenes, i.e., the recognition that asphaltenes have a wide variety of different structures, it makes sense to try and understand the effects that different structural characteristics have on the physical properties. Along this vein, the separate studies by Kuznicki et al. and Ungerer et al. focused on investigating the differences among these geometries. Kuznicki et al. used three different molecular structures of asphaltenes, one archipelago structure with two aromatic cores and two similar island molecules, one of them with a carboxylate group (COO−) terminating a side chain. Simulations of 12 continental asphaltenes with 12 archipelago asphaltenes were then conducted in three different solvents (toluene, heptane, and water). The study also considered a two-phase system of water and toluene and compared the differing aggregation at the interface with and without the anionic carboxyl group. In the single-solvent simulations, similar structural features for aggregation were seen in all of the solvents. However, the strength of the aggregation varied, as indicated by identical peak positions in the radial distribution function with varying peak heights in the order toluene < heptane < water. The simulations in the mixed toluene/water system suggested that the presence of ionic terminal groups on the aliphatic chains of specific asphaltene fractions can dramatically enhance the interfacial activity of the entire asphaltene population. Ungerer et al. used three different asphaltene structures that were consistent with elemental analysis data for asphaltenes from an Arabian crude: an “island” structure with eight aromatic rings, a larger “continental” structure with 15 aromatic rings, and an “archipelago” structure with three conjoined polyaromatic cores. Simulations of nine asphaltene molecules and 600 solvent (toluene or heptane) molecules were conducted. The results showed that limited aggregation occurred for the archipelago model, whereas strong and irreversible aggregation occurred for the continental model in both toluene and heptane. The island model formed some small aggregates of two or three molecules in both toluene and heptane, with larger and more stable aggregates occurring in heptane, qualitatively matching experimental observations.

For the simulations discussed thus far, the asphaltene structures were essentially constructed “by hand” to be consistent with a particular set of experimental data. There is no guarantee that the model molecules are the most representative structure for any particular data set. Methods for automatically generating structures and then selecting a subset that best matches experimental data are termed quantitative molecular representation (QMR) methods. One starts the procedure with a database of smaller structural units, which are then connected—following defined chemical rules—to form a large number of potential asphaltene structures. This large set is then refined to a smaller subset by minimizing the difference between experimentally determined analytical data and simulated data from the asphaltene structures, as defined through an empirical objective function. It has been found that a group of three to five molecular structures is sufficient to accurately reflect the analytical data. The initial studies of Sheremata et al. allowed connection of the molecular building blocks only through alkyl chains, generating large archipelago-type structures. This approach was extended by Boek et al. to include other potential linkages as long as they were chemically consistent, in so doing generating island structures with larger aromatic cores comprising four to seven aromatic moieties. Following this approach, a set of three structures generated from elemental analysis and 1H/13C NMR data of Athabasca bitumen with a QMR method were simulated by Headen et al. The molecular weight used is a variable factor in the QMR method and was chosen, for that study, as the rather small value of 750 Da. Simulations of six asphaltene molecules were conducted in toluene and heptane for each of the structures. Although both the island and archipelago asphaltenes exhibited aggregation in toluene and heptane, the aggregation “events” were seen to be short-lived, with asphaltene–asphaltene contacts being broken and reformed several times over the course of a 20 ns simulation. Simulation of the resin molecules showed considerably less aggregation. Additionally a constraint-force method was used to obtain an estimate of the potential of mean force (PMF) between two asphaltene molecules, yielding a free energy of dimer formation of ~10 kJ/mol. Studies based on the QMR-generated molecules have been extended to studies of asphaltene aggregation in supercritical carbon dioxide and preliminary studies of asphaltene aggregation on a calcite surface and of asphaltene molecules at the oil–water interface.

The asphaltene molecules used by Headen et al. have been employed by other groups as reference asphaltene compounds in their own simulation studies. Sedghi et al. used eight different variations of the “asphaltene C” structure of Headen et al. in order to study the effects of structural changes. The work also represented an increase in scale: the authors kept the same asphaltene concentration of 7 wt % but simulated 36 molecules for 90 ns, claiming that the formation of asphaltene nanoaggregates followed later by the formation of clusters/flocs in accordance with the modified Yen model of asphaltene aggregation. An umbrella sampling method was used to
calculate the free energy of dimer formation for each of the different asphaltene structures. It was shown that the size of the aromatic core has the largest effect on the aggregation energy; for example, increasing the number of aromatic rings from eight to 11 led to an increase in the free energy of aggregation from $-14$ to $-19.4$ kJ/mol. The length of the peripheral alkyl chains had little effect on the aggregation energy. The influence of the inclusion of heteroatoms was less conclusive. No clear link between the aggregation energy and the polarity of the molecule was found, but the addition of heteroatoms into the aromatic core did increase the energy of aggregation.

QMR is not without issues. It requires as an input library of fragments to generate the larger structures, so unless there is certainty in the choice of the fragments, the validity of the resulting structures is open to question. Moreover, the sets of molecules that match the experimental data are generally not unique, so a large number structures are obtained and a level or arbitrariness is introduced when deciding which one of these is the most representative of the asphaltene fraction. A similar approach to QMR for generating asphaltenes was employed by Strausz et al.47 These models were constructed from a detailed aggregation of their models. A total of nine structures were chosen for the study, with molecular weights ranging from 324 to 1625 Da. Simulations of aggregation in a number of different solvents (tetrahydrofuran, toluene, heptane, and pentane) were conducted for two of the larger asphaltene models (MW = 1519 and 1467 Da). Face-to-face aggregation of the aromatic planes was seen in each of the model solvents. The study also simulated four asphaltenes as pure components to obtain values of the Hildebrand solubility parameter between 17.7 and 23.4 MPa$^{1/2}$, which are close to the values for aromatic solvents such as toluene (18.9 MPa$^{1/2}$). However, as in QMR, the result of the Frigerio and Molinari model is biased by the number and morphology of the building blocks employed: a large proportion of their building blocks had an unexpectedly large proportion of aromatic moieties and no saturated rings. The resulting molecules thus exhibited large pericondensed domains, which explains the observation that aromatic stacking would be the predominant mode of aggregation of their models.

We also highlight the archipelago molecules generated by Jian et al.49 These models were constructed from a detailed work to identify the most abundant fragments using mild pyrolysis and several very selective chemical decomposition schemes of asphaltene from Athabasca. This work contains valuable quantitative information about (a) the distribution of the sizes of the side chains and the type of bridges (pure –CH$_2$– groups and those with –S—, –O—, and ester links) existing between the ring systems, (b) the size of the aromatic ring systems, and (c) biomarkers and porphyrins covalently attached to some asphaltene molecules.

An alternative approach to the ad hoc proposal of asphaltene structures for simulation is to model existing compounds or asphaltene surrogates (e.g., dyes and chromophores) that are presumably similar in structure to asphaltenes.48 As a proof of concept, Jian et al.49 simulated derivatives of violanthrone-78 as a model for asphaltene. Another recent example is the study of alkyl-substituted hexabenzocoronene molecules.50 The potential advantage of these studies is that it may be possible to obtain experimental data for the pure compound that can be readily compared to simulation results, and therefore, they serve as an important validation of the classical force fields in widespread use in the community.

### 2.2. Comparison of Results from Simulation Studies

An important aim of simulations of asphaltenes is to be able to predict properties inaccessible through experiments and by this to gain physical insight into the molecular-level processes that drive macroscopically observed behavior. A key issue, however, is to assess the reliability of the asphaltene models, as the results are clearly contingent on this choice. This apparently simple point is a very contentious one, as there is an inherent difficulty in mapping out a “characteristic asphaltene”. The ultimate behavior of a molecular model of an asphaltene is dictated by the details of the molecular geometry and chemical constituents, i.e., the morphology of the molecule. Even the calculation of the most basic of properties, such as the average or representative molecular weight, is the subject of a long-standing debate with values that vary by orders of magnitude even for laboratories separated by two floors within the same building.51–53 The futility of attempting to compare directly with small-scale experiments then becomes apparent. Hence, the simulation studies generally involve comparisons with previous results. A particularly interesting breakthrough on the elucidation of what an asphaltene molecule “looks like” has recently been provided by Schuler et al.54 In that paper, a combination of atomic-resolution imaging of over 100 asphaltene molecules using atomic force microscopy and molecular orbital imaging by scanning tunneling microscopy is presented. The work confirms that some real asphaltene molecules consist of a central aromatic core with peripheral alkane chains. In some cases, this central core is divided into several distinct interconnected polycyclic aromatic cores (i.e., “archipelago-type molecules”), while in others a single aromatic core with peripheral alkanes is the dominant asphaltene molecular architecture. A clear conclusion is that real cuts will exhibit a complex mixture of these two general motifs. This is a particularly reassuring result as it suggests that most of the recent literature has focused on suitable types of molecules. The other key issue in simulations is the use of appropriate force fields. In the case of the atomistic models described in this article, the simulations are based on intermolecular potentials historically validated for liquid-phase simulation of well-known and well-characterized small molecules. These potentials are the workhorse of computational biochemistry, are assumed to be transferable to other molecules, and in general seem to be in quantitative agreement in all but very extreme scenarios.55

The computer modeling studies to date have focused on the energetics of aggregation and the diffusion of asphaltene molecules. The asphaltene–asphaltene interaction energy is the sum of the nonbonded interactions between a pair of asphaltene molecules. In general, the overall magnitudes of the asphaltene–asphaltene interaction energies are similar among the various models, especially when molecular weight is taken into account—the largest molecules have the largest interaction energies. The results reported in the literature span values from $-89$ kJ/mol (ref 56) to as high as $-372$ kJ/mol (ref 9). This calculation does not include the change in energy due to the asphaltene–solvent interactions or the entropy change in the process of aggregation. For this one would have to consider the free energy of aggregation $\Delta G_{agg}$ in the given solvent (as opposed to a calculation of the docking energy of molecules in vacuum). Only two sets of authors have reported this type of
calculation: Headen et al.\textsuperscript{40} calculated the PMF by conducting a series of simulations in which two asphaltene molecules were constrained by the SHAKE algorithm to be a fixed distance apart. The PMF (or equivalently the free energy of aggregation) can be calculated from the force required to keep the asphaltene molecules at the set constrained distance. Sedghi et al.\textsuperscript{44} used a similar umbrella sampling method\textsuperscript{57} to assess the force-distance curve for asphaltene–asphaltene interactions: a series of simulations were conducted in which harmonic-spring bias potentials were used to keep the asphaltene molecules within a range of distances from each other. From this a more complete force-distance curve could be built. The results from the two calculations gave very similar results. Sedghi et al.\textsuperscript{44} observed a slightly higher aggregation energy in heptane than in toluene, a result that meets the experimental definition of asphaltenes.

Diffusion coefficients can be calculated directly in MD simulations by tracking the mean-square displacement of the center of mass of asphaltene molecules as a function of time.\textsuperscript{58} The literature reports a relatively wide variation in the simulation results from $0.01 \times 10^{-10}$ to $8 \times 10^{-10}$ m$^2$/s.\textsuperscript{25,32,59} These values are consistent with diffusion constants from NMR measurements, which suggest values between $1 \times 10^{-10}$ and $3 \times 10^{-10}$ m$^2$/s depending on concentration\textsuperscript{60} and structure.\textsuperscript{61}

To date, the longest and largest atomistic simulations reported have been conducted for 80 ns on systems containing 36 asphaltene molecules in explicit heptane at 7 wt %.\textsuperscript{44} Complete aggregation of the system is seen to occur on this time scale, specifically as a precipitous event at approximately 60 ns. Even in this poor solvent, the aggregation into clusters of five to 10 molecules takes over 20 ns. These time scales suggest some minimum simulation times that should be used to understand asphaltene aggregation. Hence, we examine large system sizes (27 asphaltene molecules) and, most particularly, long simulation times (80–500 ns) in an attempt to reach equilibrium conditions.

3. METHODS

3.1. Asphaltene Structures. A common underlying assumption in atomistic simulations is that a unique asphaltene molecule may be representative of the full solubility class. It is known that asphaltenes have a wide multimodal distribution of sizes, so it is unlikely that only one average molecule can be a valid descriptor for such a complex system. For practical reasons, however, any study is bound to be limited to a small number of prototypical molecules, mostly as a consequence of the rather limited computational power available coupled with the uncertainty and experimental challenge of the description of asphaltenes. We chose five structures from the literature that are consistent with available analytical data for asphaltenes and that among them represent different aspects of the asphaltene structure (e.g., island vs archipelago). These hypothetical structures are presented in Figure 1.

The first three structures, asphaltenes A and C and resin B, are the same as those employed in a previous study of MD simulation of asphaltenes\textsuperscript{40} and can be traced to the output of a QMR analysis.\textsuperscript{8} We denote structure “B” as a resin because it has a low molecular weight, a small aromatic core, and a reduced number of polar functional groups, consistent with the structural characteristics of a resin. We note that the structures depicted in Figure 1 slightly differ from those shown in the original publication, with additional CH$_2$/3 groups on asphaltenes A and C. This corrects a mistake in the drawing of the structures used in the original paper. Asphaltenes D and E are structures empirically suggested by Ungerer et al.\textsuperscript{35} These are larger structures than A–C and help inform which structural characteristics of asphaltenes affect aggregation. Asphaltenes D and E are large continental asphaltene with a single aromatic core, while asphaltene E is an archipelago structure.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{asphaltene Structures.png}
\caption{Asphaltene molecular structures used in this study. A–C are from Headen et al.\textsuperscript{40} and D and E from Ungerer et al.\textsuperscript{35}}
\end{figure}
intermolecular potentials of the OPLS-AA force-temperature and pressure were maintained using the Nose basic nitrogen sites. Selective methylation of asphaltenes has been a limitation of these models is the absence of carboxylic acid groups or cases (e.g., single molecules in toluene) with recently published results. The rotation of suitably rigid using the LINCS algorithm. A cubic periodic boundary of 1 fs was used for all of the simulations, and bond lengths were kept bond format (and is therefore electron-deficient). To resolve this, one of the peripheral aromatic carbon atoms has been switched to an sp³ CH₃ carbon.

Clearly, the choice of the model molecules in Figure 1 is arbitrary and incomplete, but it allows one to compare some of the limiting cases (e.g., single molecules in toluene) with recently published results. A limitation of these models is the absence of carboxylic acid groups or basic nitrogen sites. Selective methylation of asphaltenes has been shown to drastically reduce the resulting MW, indicating that hydrogen bonding may be relevant to the aggregation process.

3.2. Simulation Details. We used well-defined classical intra- and intermolecular potentials of the OPLS-AA force-field family to describe interatomic forces. In view of the uncertainties surrounding larger issues, such as the particular morphology, concentration, and composition of asphaltic systems, the choice of this particular force field over any other modern and well-validated one is not of prime importance. Notwithstanding, the OPLS force field has been shown to work well for aromatic liquids in reproducing experimental data. It is important to note that improper dihedrals must be used with aromatic structures in order to keep the aromatic ring flat.

Simulation cells were constructed by random placement and rotation of asphaltene molecules followed by random placement and rotation of sufficient solvent molecules to give the required concentration. This procedure was carried out at low density to reduce the probability of molecular overlap. After an initial energy minimization step to remove any high-energy structures, the system was allowed to reach equilibrium density by running an MD simulation in the isobaric–isothermal (NPT) ensemble. The temperature and pressure were maintained using the Nose–Hoover and Parinello–Rahman algorithms, respectively. A pressure of 1 bar and temperature of 300 K were used throughout our study.

The GROMACS MD simulation code (version 4.6) was employed for all of the simulations using the leapfrog MD algorithm and the Verlet pair list scheme for neighbor searching. This allowed the use of GPU acceleration of the MD code, increasing the speed of execution on a workstation by a factor of approximately 4. A time step of 1 fs was used for all of the simulations, and bond lengths were kept rigid using the LINCS algorithm. Cubic periodic boundary conditions were employed to approximate infinite bulk behavior. The cutoff of the nonbonded interactions was set at 1 nm, with a standard correction for the energy and pressure employed to account for long-range dispersion interactions. Long-range electrostatic interactions were handled using the particle mesh Ewald procedure. The atom positions for the asphaltene molecules were recorded every 1000 time steps (1 ps) to allow for comprehensive analysis of the asphaltene clusters.

Bulk fluid simulations, in which asphaltene molecules of interest were simulated in solvent at 7 wt %, form the core of the results. The concentration is representative of the upper limit of concentrations expected, typical of heavier crudes such as those of Venezuelan or Mexican origin. This rather high concentration allows for simulations containing a minimum number of solvent molecules and represents the cases where precipitation is enhanced. The simulations were run for at least 80 million time steps (80 ns), and the minimum number of asphaltene molecules considered was 27. In the case of the systems with resin molecules, they replaced solvent molecules to give an approximate composition of 7 wt % asphaltene, 7 wt % resin, and 86 wt % solvent. The total weight percent of the asphaltene fraction was again capped at 7% in the case of the mixtures considered. Table S1 in the Supporting Information gives details of the individual simulations.

To complement the results, a second type of simulation was employed to calculate the asphaltene–asphaltene PMF curves using umbrella sampling. A detailed description of the method can be found elsewhere; in particular, an application to asphaltene simulations is described by Sedghi et al. The method relies on performing a series of simulations in which a biasing potential is used to keep the asphaltene centers of mass at an average fixed distance that is varied discretely between a large separation and a close distance. The results from these simulations are then recombined and the effect of the biasing potential is removed using a weighted histogram analysis method. This is implemented in GROMACS using the g_wham utility. The spring constant (bias) was chosen to ensure sufficient overlap between each umbrella window simulation and complete sampling of the reaction coordinate. For all of the simulations presented below, a spring constant of 1000 kJ/mol·nm⁻² was deemed satisfactory. Each umbrella window simulation was run with sufficient solvent to give an asphaltene concentration of 1 wt % for 10 million time steps in the NVT ensemble after initial equilibration of 500 ps in the NPT ensemble to obtain the initial equilibrium density. Table S2 gives the details of the PMF simulations. For PMF calculations of this type, where the molecules are free to move in three dimensions, a correction is applied to remove the increase in entropy as the distance between the molecules, r, is increased: this correction was performed by adding k₆T ln(4πr³/3).²⁷

3.3. Cluster Analysis Methods. The aggregation number, i.e., the number of molecules that constitute a cluster over the course of the simulations, was used as a guide to monitor both the approach to equilibrium and the expected average asphaltene cluster size. The GROMACS utility g_clustersize was employed to calculate the average cluster size and a histogram of cluster sizes from the simulation trajectory file. Molecules were arbitrarily considered to be clustered if the minimum distance between molecules (the distance of closest approach) was less than 3.5 Å. The tracking of the minimum distance between two randomly chosen molecules of asphaltene C in heptane is displayed in Figure 2. The events where the molecules were clustered are clearly visible from the collection of state points where the minimum approach distance was below 3.5 Å. The average cluster size,

---

**Figure 2.** Plot of the minimum distance between two individual asphaltene molecules during a 5 ns time window of an 80 ns simulation of 27 molecules of asphaltene C in heptane. The red horizontal dashed line marks the cutoff employed to discriminate whether molecules form part of a cluster.
number of clusters, and cluster size distribution were recorded from each frame of the simulation trajectory. The average cluster size, $N_{ave}$, was calculated from

$$N_{ave} = \frac{\sum_{i} N_i}{\sum N_i}$$  \hspace{1cm} (1)

where $N_i$ is the number of clusters comprising $i$ molecules. The sum in eq 1 runs from $i = 2$, i.e., it does not account for monomers. In section 5 we discuss the effect of including the monomers in the calculations. The histogram of cluster sizes was calculated by averaging the number of clusters of $n$ molecules over the full length of the simulation.

$$S = \frac{1}{N} \begin{pmatrix} \sum_i (x_i - x_{cm})^2 & \sum_i (x_i - x_{cm})(y_i - y_{cm}) & \sum_i (x_i - x_{cm})(z_i - z_{cm}) \\ \sum_i (y_i - y_{cm})(x_i - x_{cm}) & \sum_i (y_i - y_{cm})(z_i - z_{cm}) \\ \sum_i (z_i - z_{cm})(x_i - x_{cm}) & \sum_i (z_i - z_{cm})(y_i - y_{cm}) & \sum_i (z_i - z_{cm})^2 \end{pmatrix}$$  \hspace{1cm} (3)

where the subscript $i$ represents atom $i$ of the cluster and the subscript “cm” represents the center of mass of the cluster. Transformation of the gyration tensor to the principal axis involves the diagonalization of $S$:

$$S = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$$  \hspace{1cm} (4)

The eigenvalues of $S$ (the principal moments) are commonly sorted in the order $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and the square of the radius of gyration is recovered from the sum of the eigenvalues: $R_g^2 = \lambda_1 + \lambda_2 + \lambda_3$. A measure of the dimensionality and symmetry of a cluster is given by the relative shape anisotropy, $\kappa^2$, defined as

$$\kappa^2 = 1 - 3 \left( \frac{\lambda_1 \lambda_2 \lambda_3 + \lambda_1 \lambda_3 + \lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_1 \lambda_3 + \lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_1 \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \right)^2$$  \hspace{1cm} (5)

$\kappa^2$ can take values between 0 and 1, where a value of 0 corresponds to a perfectly spherical cluster and a value of 1 corresponds to a linear chain.

Another key factor in understanding the structure of asphaltene aggregates is the level of solvent penetration into the aggregate. What is needed is an estimate of the density of the asphaltene aggregate in the simulations. To obtain this estimate, we assume that the volume of each cluster can be approximated by calculating the volume of an “effective” ellipsoid that would give the same principal moments of the gyration tensor ($\lambda_1$, $\lambda_2$, and $\lambda_3$). This equivalent ellipsoid would have axes $a$, $b$, and $c$ equal to $\sqrt{\frac{M_i}{n}}$ ($n = 1–3$), and therefore, we can approximate the “encompassing” volume of the cluster as

$$V_{\text{cluster}} = \frac{4}{3}\pi \sqrt[3]{\lambda_1 \lambda_2 \lambda_3}$$  \hspace{1cm} (6)

Further discussion of the analysis is given in the Supporting Information. The mass of the cluster can then be readily calculated from the mass of each molecule and the number of molecules in the cluster. The estimate of the density of the cluster is therefore obtained from

$$\rho_{\text{cluster}} = \frac{\sum m_i}{V_{\text{cluster}}}$$  \hspace{1cm} (7)

where $m_i$ is the mass of the $i$th molecule in the cluster. We note that the cluster calculations were performed by considering only asphaltene molecules. This means that the solvent molecules, even if they were within a cluster, were not part of the shape analysis.

The radius of gyration, relative shape anisotropy, and density estimate were calculated for each cluster in trajectory frames every 10 000 time steps (10 ps). The average of each of these quantities was calculated in each trajectory frame, and the distributions of the properties were calculated over the length of the whole simulation. On rare occasions a cluster was larger than the size of the simulation periodic cell. In order to avoid spurious results, such clusters were ignored in the calculation of averages and distributions of cluster properties.

4. SIMULATION RESULTS AND ANALYSIS

4.1. Simulations of 27 Identical Asphaltenes in Toluene or Heptane. In Figure 3 we show equilibrium snapshots of simulations of 27 molecules of asphaltene C in toluene and heptane at an asphaltene concentration of 7 wt %. Solvent molecules have been removed for clarity. The cyan, white, and yellow spheres represent the carbon, hydrogen, and sulfur atomic centers, respectively. The blue boxes outline the simulation cells. Periodic boundary conditions were applied in all Cartesian directions.

Figure 3. Representative snapshots from simulations of 27 molecules of asphaltene C in toluene and heptane at an asphaltene concentration of 7 wt %. Solvent molecules have been removed for clarity. The cyan, white, and yellow spheres represent the carbon, hydrogen, and sulfur atomic centers, respectively. The blue boxes outline the simulation cells. Periodic boundary conditions were applied in all Cartesian directions.

A standard metric for the size of polymers or other macromolecules in solution is the radius of gyration ($R_g$), defined as

$$R_g = \frac{1}{N} \sum_{i=1}^{N} (\hat{r}_i - \hat{r}_{cm})^2$$  \hspace{1cm} (2)

where $\hat{r}_i$ is the position vector of atom $k$ and $\hat{r}_{cm}$ is the position vector of the center of mass of the cluster. Although the radius of gyration provides a useful, shape-independent measure of cluster size, it does not give information about the shape of the cluster. A more appropriate related property is the gyration tensor ($S$), given by

$$S = \begin{pmatrix} \sum_i (x_i - x_{cm})^2 & \sum_i (x_i - x_{cm})(y_i - y_{cm}) & \sum_i (x_i - x_{cm})(z_i - z_{cm}) \\ \sum_i (y_i - y_{cm})(x_i - x_{cm}) & \sum_i (y_i - y_{cm})(z_i - z_{cm}) \\ \sum_i (z_i - z_{cm})(x_i - x_{cm}) & \sum_i (z_i - z_{cm})(y_i - y_{cm}) & \sum_i (z_i - z_{cm})^2 \end{pmatrix}$$  \hspace{1cm} (3)

where $\hat{r}_i$ is the position vector of atom $k$ and $\hat{r}_{cm}$ is the position vector of the center of mass of the cluster. Although the radius of gyration provides a useful, shape-independent measure of cluster size, it does not give information about the shape of the cluster. A more appropriate related property is the gyration tensor ($S$), given by

$$S = \begin{pmatrix} \sum_i (x_i - x_{cm})^2 & \sum_i (x_i - x_{cm})(y_i - y_{cm}) & \sum_i (x_i - x_{cm})(z_i - z_{cm}) \\ \sum_i (y_i - y_{cm})(x_i - x_{cm}) & \sum_i (y_i - y_{cm})(z_i - z_{cm}) \\ \sum_i (z_i - z_{cm})(x_i - x_{cm}) & \sum_i (z_i - z_{cm})(y_i - y_{cm}) & \sum_i (z_i - z_{cm})^2 \end{pmatrix}$$  \hspace{1cm} (3)

where $\hat{r}_i$ is the position vector of atom $k$ and $\hat{r}_{cm}$ is the position vector of the center of mass of the cluster. Although the radius of gyration provides a useful, shape-independent measure of cluster size, it does not give information about the shape of the cluster. A more appropriate related property is the gyration tensor ($S$), given by

$$S = \begin{pmatrix} \sum_i (x_i - x_{cm})^2 & \sum_i (x_i - x_{cm})(y_i - y_{cm}) & \sum_i (x_i - x_{cm})(z_i - z_{cm}) \\ \sum_i (y_i - y_{cm})(x_i - x_{cm}) & \sum_i (y_i - y_{cm})(z_i - z_{cm}) \\ \sum_i (z_i - z_{cm})(x_i - x_{cm}) & \sum_i (z_i - z_{cm})(y_i - y_{cm}) & \sum_i (z_i - z_{cm})^2 \end{pmatrix}$$  \hspace{1cm} (3)
further aggregation is expected (see the discussion in section 6.2, where we expand on this comment). The average cluster size in heptane approaches an upper limit that is commensurate with the total number of molecules, and therefore, it is presumed that the cluster size reported is limited by finite-size effects, as the trend suggests complete phase segregation. The capped cluster size seen in toluene, however, implies that there is a degree of clustering and aggregation that is limited not by mass but rather by the balance between the energetically favorable cohesion and the entropically unfavorable clustering.

For the case of the archipelago-type asphaltene A, the difference between the clustering behaviors in toluene and heptane is less distinct. A plot of aggregate sizes over 80 ns simulations is shown in Figure 5. There is an indication that asphaltene A is less soluble in toluene at these concentrations, tending to form larger clusters and separating out. Lighter crude oils (e.g., Arabian crudes) presumably have asphaltenes with a smaller proportion of pericondensed rings and are particularly sensitive to asphaltene precipitation at very low concentrations, much below the concentration studied here. While not conclusive, there is a suggestion that the more open asphaltene structures may have a lower solubility in toluene. The key differences between the island-type asphaltene C and the archipelago-type asphaltene A are morphological, involving both the aggregate shape and density. A comparison of the distributions of cluster radius of gyration, relative shape anisotropy, and density is given in Figure 6. As asphaltene A is a larger molecule, it is unsurprising that its cluster radius of gyration is generally larger than that for asphaltene C (Figure 6a). It is worth noting that for asphaltene C in heptane (blue dashed curve) there is a bimodal distribution in the cluster size, indicating larger clusters ($R_g \sim 15$ Å) in coexistence with smaller clusters ($R_g \sim 8$ Å); however, this might be an artifact of the finite size of the cell. For this particular system/solvent, the final equilibrium state is most likely complete phase separation, with a (relatively pure) asphaltene phase in
coexistence with a supernatant solvent with some (very limited) asphaltene solubility. Attaining this equilibrium state will require simulations with orders of magnitude larger length- and time-scales, and the results shown here correspond to the initial stages of this process.

The distributions of relative shape anisotropy (eq 5) are given in Figure 6b. The clusters of asphaltene A are less spherical than those of asphaltene C. The extended archipelago structure is likely to produce more extended asymmetric clusters, analogous to what is commonly seen for elongated polymers in solution. The estimated densities of the clusters also show differences (Figure 6c). Clusters of asphaltene C are consistently denser than those formed by asphaltene A, and asphaltenes in heptane show denser aggregates than those in toluene. In general, the results are in agreement with the measurements of Gawrys et al., who suggested a range of solvent entrainment from 30 to 50% (v/v). For asphaltene C in toluene, the average density suggests that half of the volume assigned to the cluster is filled with entrained solvent. It is reassuring to note that this observation is consistent with observations from small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) experiments and previous simulations.

The large continental asphaltene D, with its large aromatic core, is expected to show high levels of clustering and aggregation. However, the corresponding analysis in terms of cluster aggregation numbers suggests that the cluster sizes in heptane are lower than for the other asphaltenes and that the difference in aggregation number for simulations in toluene and heptane is small. (The average cluster size over the simulation and the average cluster size distribution are reproduced in the Supporting Information.) At first glance, the observation of lower aggregate size is anomalous, but the distribution of cluster sizes indicates that aggregation is indeed strong, as there are very few monomers in solution. A likely explanation is that in the early stages of the simulation the molecules quickly form dimers, trimers, or other small clusters. There is little disaggregation and reformation of asphaltene contacts for asphaltene D compared with the smaller asphaltenes—the only way for clusters to grow is therefore for these small clusters to coalesce. Diffusion of the larger clusters is particularly slow, and therefore, coalescence is likely to take considerably longer, beyond the time scales reasonably available to atomistic molecular dynamics simulations.

The archipelago-type asphaltene E shows cluster properties similar to those of archipelago-type asphaltene A, and for succinctness, the average cluster properties and plots of average cluster size and distributions of cluster properties for this asphaltene are given in the Supporting Information.

4.2. Simulation of Mixtures of Asphaltene Structures.

4.2.1. Effect of Resins. To evaluate the effect of the presence of resin molecules, 7 wt % of the solvent (toluene or heptane) was replaced with an equivalent mass of resin B (Figure 1). For simulations of 27 molecules of the island-type asphaltene C,
there is little to no difference in the cluster size or properties when 7 wt % of toluene is replaced by resin. In the case that this fraction of heptane is replaced by resins, there are indications of an improvement in the solubility of the asphaltene: there is a clear reduction in the density of the asphaltene clusters, and the radius of gyration of the clusters becomes smaller (Figure 7). It should be noted that only molecules of asphaltene C were included in the calculation of the cluster properties—for these calculations the resin was considered part of the solvent.

The changes observed in the heptane-based systems upon addition of resin can be rationalized by exploring the behaviors of the pure resin (27 molecules; 7 wt %) in toluene and heptane (without any asphaltenes). As expected, the resin is soluble in both cases, but inspection of the resin–resin radial distribution function (RDF) $g(r)$ (Figure 8) shows no peak in the RDF in toluene, whereas a peak is present in heptane, indicating some level of incompatibility with the solvent. When asphaltenes are also present, the formation of resin–asphaltene aggregates is likely to be favorable in heptane, thus reducing the density of the asphaltene aggregate as a result of entrainment of the resin. Our simulations show no evidence of resin molecules solvating or solubilizing the asphaltene clusters (as in a micellar picture of asphaltene solubility$^{81}$); the distribution of resin molecules within a cluster is therefore homogeneous and not limited to the periphery (Figure 9). These observations are consistent with the absence of peaks in $g(r)$ for the aromatic—
4.2.2. Mixtures of Asphaltenes. We performed simulations of a mixture of two morphologically different asphaltene structures. There is an expectation that the presence of polydispersity in asphaltenes may hinder aggregation by virtue of steric hindrance due to the increased incompatibility in the shapes of the different molecules. A mixture (50:50 on a weight basis) of island-type asphaltene C with archipelago-type asphaltene A was studied, with the total combined weight fraction of the asphaltene fraction kept at 7 wt %, as in the pure-component cases. The results (see the Supporting Information) indicate that for all the cluster properties—aggregation number, radius of gyration, shape anisotropy, and density—the values of the properties obtained for the simulation of mixed species lie between the values obtained for the pure components, i.e., the mixture behaves in an ideal fashion in terms of the asphaltene composition dependence. There is no obvious reduction in cluster size upon mixing of the different asphaltene structures, nor is there a significant change in the distribution of the cluster properties. As displayed in Figure 10, the distributions of the cluster radius of gyration and density lie between the distributions for the single (pure) asphaltene simulations. A similar result is also seen for simulations of a mixture of the two archipelago-type asphaltenes A and E (see the Supporting Information). These observations are consistent with recent report on simulations of mixtures of asphaltenes by Santos Silva et al.,22 where the behavior of a mixture of models was found to be consistent with the behaviors of the individual asphaltenes.

In a similar fashion, we probed the effect of the addition of the resin to the solvent on a mixed asphaltene system. Here we ran the latter asphaltene mixture (A + C) with a proportion of the solvent (∼7 wt %) replaced by resin B. The calculated distributions of cluster sizes and cluster densities are given in Figure 11. The effect on the cluster radius of gyration is limited, although in both toluene and heptane the introduction of resin does appear to produce a bimodal distribution, increasing the number of smaller clusters and favoring the solubility of the asphaltenes. The differences in cluster density in the simulations are more marked. Unlike the simulation where only asphaltene C is present, where the addition of resin has a negligible effect on the cluster properties, there is a clear decrease in cluster density for the mixture of asphaltenes A and C and an increase in the overall solubility with the introduction of resin. Clearly, more cases with larger degrees of polydispersity in the asphaltene, resin and solvent fractions need to be considered in order to reach more general and meaningful conclusions.

4.3. Potential of Mean Force between Pairs of Asphaltene Molecules. The potential of mean force (PMF) between the centers of mass of two asphaltene molecules provides a quantitative description of the strength of aggregation between asphaltene molecules.81 The difference in energy between the minimum of the PMF and the value of the PMF at long distances can be interpreted as the free energy of dimer formation, ΔG_{dimer}. Values of the free energy of dimer formation for asphaltene models are given in Table S2 and are seen to exhibit a wide spread of over an order of magnitude, from 8.4 kJ/mol for asphaltene C in toluene to 66.4 kJ/mol for asphaltene D in heptane.82 This is partially due to the inherent differences in close packing of the molecules, where the more continental morphologies can form highly energetic face-to-face interactions, and also partially due to the difference in flexibility of the molecules, as the springs fix the center of mass but the different molecules will bend and flip in significantly different fashions depending on their general morphology. As an example, the PMF curves for the island-type asphaltene C and continental-type asphaltene D are given in Figure 12. The equilibrium separation between the centers of mass of the molecules forming a dimer is ∼6 Å. The free energies of dimer formation for the large continental asphaltene D in both toluene and heptane (Figure 12) are considerably higher than those for asphaltene C, as the large aromatic core of asphaltene D makes aromatic stacking highly favorable. Similarly, the free energy of dimer formation is considerably higher in heptane than in toluene, indicating the propensity for clustering in
heptane. What is notable is the small difference in the solvation energies between solvents, suggesting that the precipitation of asphaltenes is crucially driven by extremely small changes in the physicochemical characteristics of the mixture (solvent).

We conducted umbrella sampling simulations involving a single asphaltene molecule and an asphaltene dimer to explore the energy cost of “growing” clusters from dimers. In these simulations, two umbrella sampling “springs” were used. One “spring” was used to maintain a preformed dimer at a fixed separation—in this case the equilibrium separation was the same for all of the umbrella windows and was set at 6 Å. The second spring was fixed between the center of mass of a third molecule and the center of mass of one of the molecules in the preformed dimer. The equilibrium distance for the second “spring” of the reference molecule to the third molecule was then varied between 3 and 20 Å. The resulting PMF curve for asphaltene C trimer formation in toluene is shown in Figure 13 as a function of the distance between the centers of mass of the third molecule and the reference molecule. There is a small reduction in the free energy of aggregation compared with that of dimer formation, which indicates that as the aggregate increases in size further aggregation becomes less favored.

4.4. Dimensional Map of Asphaltene Aggregates. It is instructive to represent the shape of a cluster in terms of ratios of the principal moments of the gyration tensor ($\lambda_1$, $\lambda_2$, and $\lambda_3$). We can define the following ratios, remembering that by definition $\lambda_1 > \lambda_2 > \lambda_3$:

$$ r_1 = \sqrt{\frac{\lambda_1}{\lambda_2}}, \quad r_2 = \sqrt[3]{\frac{\lambda_2}{\lambda_3}} $$

Figure 10. Distributions of asphaltene cluster radius of gyration and (inset) estimate of cluster density from simulations of asphaltene C, asphaltene A, and a 50:50 (by weight) mixture in toluene (left) and heptane (right).

Figure 11. Distributions of cluster radius of gyration (left) and estimate of cluster density (right) for simulations of mixtures of asphaltenes A and C in toluene and heptane with and without resin B.

Figure 12. Potential of mean force between two molecules of asphaltene C in toluene (purple) and heptane (green) and between two molecules of asphaltene D in toluene (cyan) and heptane (gold).
where \( r_1 = 1 \) for an oblate shape \( (\lambda_1 = \lambda_2 > \lambda_3) \) and \( r_2 = 1 \) for a prolate shape \( (\lambda_1 > \lambda_2 = \lambda_3) \). A plot of \( r_1 \) against \( r_2 \) can graphically represent the shape of the cluster; the closer the point for the cluster is to the \( x \) or \( y \) axis, the more it can be thought of as prolate or oblate, respectively. It should be noted that when \( r_1 = r_2 \) we obtain \( \lambda_2 = \sqrt{\lambda_1 \lambda_3} \), so that the cluster is neither prolate or oblate, while a sphere would be the special case where \( r_1 = r_2 = 1 \) (corresponding to the origin of the plot).

For each of the simulations of 27 identical asphaltenes, a two-dimensional histogram was constructed by calculating the number of clusters formed during the simulation that fell within bounds \( \Delta r_1, \Delta r_2 \). The corresponding plots are shown in Figure 14 for the island-type asphaltene C and the large continental-type asphaltene D (dimensional maps for other simulations are available in the Supporting Information). There are clear differences in the shapes of the clusters for the different asphaltenes. The smaller asphaltene C forms predominantly slightly prolate-shaped clusters. For the large continental-type asphaltene D, on the other hand, there is a clear preference for oblate shapes. This is likely to be due to the underlying shape of the continental asphaltene: aromatic stacking will bias columnar-like stacking when few molecules make up the clusters.

In order to directly compare the cluster shapes for clusters of different asphaltenes, in Figure 15 we have plotted the average
values of \( r_1 \) and \( r_2 \) for each simulation. Clusters in heptane are on average more spherical, mostly by a reduction in \( r_1 \) (a reduction in the longest dimension of the cluster). This analysis agrees with the fact that heptane is a poor solvent and the asphaltene molecules will attempt to form the densest clusters with the least available exposed area. The archipelago-type asphaltenes A and E are on average more prolate than oblate, which is perhaps no surprise given their molecular shapes and is consistent with the general behavior of oligomers in good solvents.

5. MODELS OF ASPHALTENE AGGREGATION IN TOLUENE

The aggregation structure of asphaltenes in good solvents and stable crudes has received considerable attention in recent years.\(^2\) Here we do not attempt to review all of the literature on the subject; a brief overview is that there are currently two models considered for asphaltene association at the nanoscale. The first is the Yen—Mullins model,\(^2\) which postulates two levels of aggregation: the formation of nanoaggregates—aggregates containing <10 asphaltene molecules—predominantly held together by stacking of aromatic cores and the posterior appearance of nanoaggregates that consist of more loosely bound clusters with a size of approximately 10 nm. The second model, proposed by Gray et al.,\(^8\) is a "supramolecular assembly model", whereby a broad range of intermolecular interactions (\( \pi-\pi \) stacking) lead to diffuse asphaltene aggregation.

What insights can the simulations described above give as to the nature of asphaltene aggregation at the nanoscale? The first thing to note is that the distributions of aggregate sizes throughout the simulations exhibit a large number of monomers and dimers: for all but the simulations of the large continental-type asphaltene D (which tends to be an anomalous molecule), the monomer is the most common "aggregate". We believe that this is not just an effect on the short time scales involved in the simulation (see the next section).

The fact that we have a highly skewed distribution of aggregate sizes means that we need to be careful about how the "average" aggregation number is defined. In fact, depending on the experimental technique employed, the measure of aggregation may change: if the method is sensitive only to aggregation "events", it will not include monomers in the calculation, but if the method essentially calculates the molecular weight of the aggregate, then monomers will be included. For example, small-angle scattering provides a weight-averaged molecular weight,\(^6\) which is best compared to a weight-averaged cluster size, \( \bar{N}_w \), where the index \( i \) runs from \( i = 1 \):

\[
\bar{N}_w = \frac{\sum i^2 \cdot N_i}{\sum i \cdot N_i}
\]

The average cluster sizes given in Table S1 are an average (defined by eq 1) for clusters containing at least two molecules. If one considers evaluating eq 1 including the monomers (the index \( i = 1, 2, 3, \ldots \)), for simulations of asphaltene C in toluene the number- and weight-averaged cluster sizes are 2.4 and 4.4 respectively. The weight-averaged aggregation number approaches the values stated from NMR, and DC conductivity measurements,\(^5\) approximately 8 and 6, respectively. A recent analysis of small-angle X-ray scattering from low concentrations of asphaltenes in toluene using a two-state aggregation model gave an aggregation number of 3.3 ± 1.3,\(^6\) in good agreement with our current findings. It is important to note that with a weight-averaged aggregation number of 4.4, the most common "aggregate" is still an asphaltene monomer. These results therefore suggest that the nanoaggregate model should be interpreted in terms of a preferential distribution of asphaltene cluster sizes where monomers, dimers, and trimers are present in significant numbers, or better yet should be discarded in favor of the recognition of a continuous distribution of \( n \)-mers.

The second clear result from the analysis of the asphaltene clusters is that they can be diffuse, containing a significant level of solvent—50% or more. This compares well with results from absolute intensities of small-angle scattering experiments.\(^7\) The nanoaggregate model implies that the nanoaggregates themselves are dense—formed predominantly through \( \pi-\pi \) stacking—and the diffuse nature of the clusters measured through SAXS/SANS is due to fractal aggregation of these dense nanoaggregates. The simulations described above run counter to the hypothesis that dense nanoaggregates are formed—they tend to show that the density of the aggregates formed is similar regardless of their size. There is no evidence of a model where two steps are involved. However, it is quite clear that much larger simulations will be needed to show this with confidence.

Analysis of the principal dimensions of all of the asphaltene aggregates present during the simulation suggests that prolate-type structures are more common than oblate ones. The exception is the continental-type asphaltene D, which has a very large aromatic core and a strongsteric propensity for \( \pi-\pi \) stacking, leading to oblate cylindrical structures. The size and shape of asphaltene aggregates have been extensively studied by small-angle X-ray and neutron scattering. The interpretation of the scattering data to specific form factors is difficult because of the polydispersity of the system. A particular distribution (e.g., Schultz, log-normal, etc.) has to be assumed from the outset, and often many parameters have to be estimated from the data.

One of the most thorough small-angle scattering studies of asphaltenes in solution was conducted by Eyssautier et al.,\(^3\) who used both SAXS and SANS with solvent contrast variation to obtain a comprehensive data set. The results were interpreted as fractal clusters of disk-shaped nanoaggregates with a height of 6.7 Å and a diameter of 32 Å. On the dimensional maps plotted in Figure 15, this would be represented at \( (r_1 = 1, r_2 = 2.2) \), which is considerably different from the values seen for clusters in the simulations above. Notwithstanding, the present simulations are too small and too short to observe clusters at the length scale seen by small-angle scattering \( (R_g \approx 50 \text{ Å}) \). Larger and longer simulations are required to determine if there are indeed two aggregation length scales (i.e., nanoaggregates forming clusters) or if these loose aggregates continue to increase in size as larger system sizes are used.

Are the time and length scales too short/small for all of the simulations presented in the current work? We believe that for the simulations in toluene the simulation time was long enough (but possibly a larger system would improve the statistics). To support this assertion, we ran a "cluster breaking" simulation—essentially taking the asphaltene cluster structure from the long simulation of asphaltene C in heptane, where all of the molecules are in a cluster, removing the heptane, and replacing the solvent with toluene (keeping the asphaltene atoms "frozen" during the process). We observed (see Figure S12)
that in less than 20 ns the system evolved to a similar average cluster size as for the simulation with toluene (simulation 1), suggesting that (a) we sampled for a long enough time in toluene and (b) the aggregation observed in our simulations is reversible in nature.

6. CONCLUSIONS

6.1. General Observations. We have employed atomistic molecular dynamics simulations of large systems (≈50,000 atoms) and long time scales (80 ns) focused on four different asphaltene structures, a model resin molecule, and their mixtures. Asphaltenes were simulated in an explicit solvent, either toluene or heptane, under ambient conditions. These conditions are not the same as experienced in either an upstream or downstream oil production environment, where asphaltene aggregation and deposition are likely to be a problem. By definition, however, they do represent the solubility extremes for asphaltenes and as such are vital model systems for comparison to experiment. A case is also to be made toward the degree to which the models represent real systems with respect to morphology, number and location of heteroatoms, average molecular weight, etc. While all of these aspects could be contentious, some observations of asphaltene aggregation behavior are common to all asphaltene structural types and are presumed to be universal. These include the following:

- In a good solvent there is a distribution of cluster sizes where monomers and dimers constitute the most common “cluster” size.
- The average aggregation number of clusters ranges between 3.6 and 5.6 molecules (as defined by eq 1). This is lower than seen by some indirect experimental methods (six to eight molecules), but the number is dependent on the type of average that is measured. The results closely match the estimate of aggregation numbers obtained from small-angle X-ray scattering data.
- Analysis of the principal axes of the gyration tensor indicates that on average the clusters are relatively spherical with $\sqrt{\lambda_1/\lambda_2} < 1.9$ and $\sqrt{\lambda_2/\lambda_3} < 1.6$. There is no evidence of thin disk-shaped nanoaggregates as suggested by small-angle scattering results. Differences in the average shape of the clusters are due to the underlying structure of the molecule, with archipelago-type asphaltene favoring longer prolate structures and large continental-type asphaltene favoring oblate shapes. This is in agreement with previous observations from both experiment and simulation.
- Estimates of the density of the aggregates from simulation show that they are diffuse in nature, containing on average at least 50% solvent, in excellent agreement with experimental observations.
- Simulations of mixtures of different types of asphaltene structure show cluster properties that are intermediate between those of the individual components. There does not appear to be an overall reduction in the asphaltene cluster size when two asphaltene structures are used.
- The replacement of a portion of the solvent molecules (7 wt %) with a model resin does not alter the asphaltene aggregate shape, size, or density when toluene is the solvent. However, in heptane the effect of resin is to decrease the density of the asphaltene aggregates, most possibly by intrusion. These results indicate that the resins do not play a solvation (or surfactant-like) role in the formation of asphaltene clusters. The results in heptane match the experimental findings of the coprecipitation of more soluble molecules with asphaltenes.
- Potential of mean force calculations using umbrella sampling and weighted histogram analysis show that the free energy of dimerization of asphaltenes is lower in toluene than in heptane. For asphaltene C, the free energy of forming a trimer from a dimer and a monomer in toluene is lower than the free energy of dimerization. If this trend repeats itself for larger clusters, it would explain the fact that in toluene the aggregation of asphaltenes is limited to a small number of molecules—i.e., reductions in free energy of aggregation as more molecules are added to the cluster eventually limit the cluster growth. In heptane, complete aggregation should
molecules in clustered states. Here there is clear evidence that the probability of finding a molecule in a cluster of a given size. Such a plot gives an average snapshot of the distribution of aggregation numbers for the first and last 100 ns of 500 ns simulations of 27 molecules of asphaltene C in toluene and heptane at 7 wt %.

Figure 17. Distributions of cluster sizes for the first (red) and last (green) 100 ns of 500 ns simulations of 27 molecules of asphaltene C in toluene and heptane.

occur by definition, and this would be in line with an increase in free energy as the cluster size increases.

- We find no evidence to support the claim that there is a fundamental difference between the aggregates of a few asphaltene molecules (so-called nanoaggregates) and those of larger clusters. In fact, the asphaltene cluster distributions are seen to be continuous, suggesting a distributed model is more accurate than the Yen–Mullins model, though the latter is conceptually appealing and widely accepted.

6.2. Outlook. As computational power has increased, the time scales achieved for molecular dynamic simulation of asphaltenes have also increased, from O(10−100 ps) up to O(10 ns). However, even these time scales are most likely small (cf. Figure 5), where the relaxation times for asphaltene A are clearly large, as only after ∼40 ns one can observe differences in behavior between toluene and heptane. We have run uniquely long 500 ns (0.5 μs) simulations of the island-type asphaltene C in toluene and heptane, continuing from the end of the 80 ns simulations (simulations 1 and 2). The average cluster sizes over the course of these simulations are depicted in Figure 16. There is no obvious drift in the average aggregation number for the simulation in toluene, with the cluster size oscillating between 3 and 5. However, in heptane there is a discernible upward trend in the aggregation number, although this is only clear over such a long time scale (by the standards of atomistic MD).

The differences in aggregation at the beginning and end of the 500 ns simulations are clearer if we look at the distributions of aggregation numbers for the first and last 100 ns (Figure 17). In Figure 17 we have plotted the number of clusters of each size, N(cluster size), multiplied by the cluster size for each particular instance. The ordinate is therefore proportional to the probability of finding a molecule in a cluster of a given size. Such a plot gives an average snapshot of the distribution of molecules in clustered states. Here there is clear difference between the behavior in toluene and that in heptane. In heptane, there are few monomers and dimers, and the majority of asphaltene form part of large clusters (or a single cluster including all of the molecules); in toluene, the most likely “cluster” is in fact an asphaltene monomer—there are observable numbers of dimers and trimers, but as the cluster size increases, the likelihood of the formation of a multi-
molecular cluster decays rapidly. However, in toluene there is little difference in the distribution from the beginning to the end of the 500 ns simulation, with monomers or small aggregates being most common, whereas in heptane by the last 100 ns of the simulation there is a clear preference for larger clusters of 18 to 25 molecules, in sharp contrast to the results inferred from looking only at the first 100 ns of the simulation. Here we are clearly encountering finite size effects, as the number of molecules in the box is 27. If a larger system were simulated, one may assume that the cluster size would continue to grow.

These simulations are the largest and longest atomistic molecular dynamics simulations conducted on asphaltene in solution at the present time. In spite of this, it is clear that these simulations fall short of being able to fully replicate the aggregation behavior even for these modest-sized systems in good solvents. The radius of gyration of asphaltene in toluene as seen by small-angle scattering is on the order of ∼50 Å. To accurately represent this, a simulation box would need to be larger, say at least 200 Å in each dimension, compared to ∼80 Å for the simulations reported here. To put this into context, simulations would take on the order of 15 times longer to run. In addition to the simulation box size, our 500 ns simulations indicate that in poor solvents (e.g., heptane), changes can occur over time scales that are very long compared with those associated with atomistic MD.

A way to access these longer time and length scales is to use coarse-graining (CG) of the force field, representing groups of atoms by a single “bead”. The simplest form of CG for organic systems is to lump together the hydrogen with the carbon in “united atom” models, as used in some simulations of asphaltene. This is usually implemented solely for the aliphatic carbons. It reduces the total number of atoms (quite considerably for highly saturated organics) and also removes the highest-frequency molecular vibrations, the C–H bonds, therefore increasing the time step that can be used. However, more aggressive CG is needed to explore the problem in an appropriate fashion.

To date there have been relatively few attempts to study asphaltene over larger length and time scales using coarse-grained models with molecular mechanics, standard molecular dynamics, or dissipative particle dynamics. However, most of the aforementioned models lack a traceable
link to the atomistic model they attempt to represent. The beads (or superatoms) have a distant relation to the underlying atomistic force fields. A further consideration is that the loss of structural detail in the CG approach does require careful benchmarking against experimental and/or more detailed simulation results. In need of particular care is the modeling of the polynuclear aromatic hydrocarbon core of the models, for which the current models, based on the properties of benzene, provide a poor representation and an unphysical overprediction of the face-to-face aggregation.94 Provided that for which the current models, based on the properties of one has faith in both the representability of the proposed large-scale modeling of petroleum compared with fully atomistic MD and are the way forward for overprediction of the face-to-face aggregation.94 Notwithstanding the hurdles, CG methods have the potential to increase the size or time scale available by a factor of 10−100 compared with fully atomistic MD and are the way forward for large-scale modeling of petroleum fluids.95,96

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.energyfuels.6b02161.

Tables with simulation details of all runs conducted and numerical results presented (S1); additional results for asphaltene E, asphaltene D, and a mixture of asphaltene A and asphaltene E in toluene and heptane (S2−S4); cluster dimensional maps for asphaltenes A and E (S5); plot of the numbers of monomers during 500 ns simulations of asphaltene C in toluene and heptane (S6); discussion of the reversibility of the clustering process (S7); and further discussion of eq 6 (S8) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: e.muller@imperial.ac.uk.

ORCID

E. A. Müller: 0000-0002-1513-6686

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the funding and technical support from BP through the BP International Centre for Advanced Materials (BP-ICAM), which made this research possible. Enlightening discussions with Prof. Juan Murghie are gratefully acknowledged. The simulations described herein were performed using the facilities of the Imperial College High Performance Computing Service and the BP High Performance Computer Centre. Additional support of the Molecular Systems Engineering Group from the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom (Grants EP/E016340 and EP/J014958) is also gratefully acknowledged.

REFERENCES

(9) Murghie, J.; Rodríguez, J.; Aray, Y. Energy Fuels 1996, 10, 68−76.
(40) Heiden, T. F.; Boek, E. S.; Skipper, N. T. Energy Fuels 2009, 23, 1220−1229.
(56) Rogel, E. Energy Fuels 2000, 14, 566–574.