Simple Statistical Model for Branched Aggregates: Application to Cooee Bitumen

Claire A. Lemarchand*,[†] and Jesper S. Hansen[†]

[†]DNRF Centre "Glass and Time", IMFUFA, Department of Sciences, Roskilde University, Universitetsvej 1, Postbox 260, DK-4000 Roskilde, Denmark

ABSTRACT: We propose a statistical model that can reproduce the size distribution of any branched aggregate, including amylopectin, dendrimers, molecular clusters of monoalcohols, and asphaltene nanoaggregates. It is based on the conditional probability for one molecule to form a new bond with a molecule, given that it already has bonds with others. The model is applied here to asphaltene nano-aggregates observed in molecular dynamics simulations of Cooee bitumen. The variation with temperature of the probabilities deduced from this model is discussed in terms of statistical mechanics arguments. The relevance of the statistical model in the case of asphaltene nanoaggregates is checked by comparing the predicted value of the probability for one molecule to have exactly i bonds with the same probability directly measured in the molecular dynamics simulations. The agreement is satisfactory.
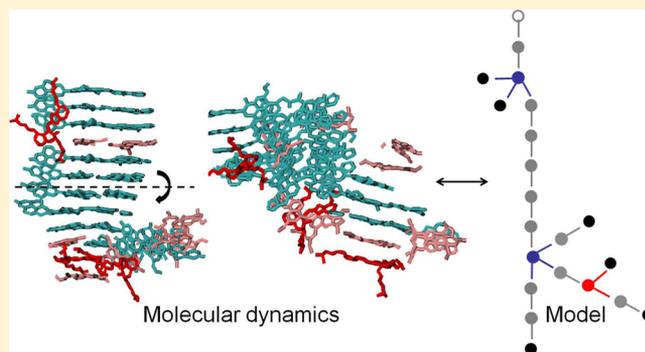
1. INTRODUCTION

Many chemical systems used in industrial applications have a molecular or supramolecular structure that is branched. These systems include dendrimers, used for drug delivery and having a hyperbranched and highly symmetric structure,[1,2] amylopectin, which is a branched polysaccharide present in cereal starch,[3,4] asphaltene nanoaggregates, present in bitumen and that can be branched depending on the asphaltene molecular structure,[5] and molecular clusters of monoalcohols.[6] The topology as well as the polydispersity in size of these branched structures can impact the macroscopic properties of the system. It is therefore important to understand and control the size distribution of these structures.

Theoretical models already exist that connect the macroscopic properties of the system to the ability of each molecule to form one or more bonds. "Bond" is to be understood here in a very general sense and can represent a covalent bond as well as much weaker connections such as bonds holding together molecular clusters. The Wertheim theory,[7−10] in particular, predicts the temperature and pressure dependence of the bond forming probability and the equation of state of associating molecular fluids with multiple binding sites.[11] However, the size distribution of the branched clusters is not directly addressed in this theory. Alternatively, the Flory–Stockmayer model[12−14] proposes a prediction for the size distribution of branched polymers at equilibrium, but this model assumes that all functional groups have the same reactivity. For different reasons, including steric hindrance, it is likely that forming a first bond occurs at a different probability than forming a second bond given that a first one has already been formed. This fact is pointed out by Sillren et al.[6] in the case of

monoalcohol clusters. They consequently extend the Flory model[15] using one parameter for the probability of one molecular unit to form one bond by introducing a second parameter corresponding to the conditional probability for one molecular unit to form a second bond given that a first one is already present. The aim of the present paper is to generalize the formalism introduced by Sillren et al. to molecular units that can form an arbitrarily large number of bonds.

To check for the relevance of the generalized model, we apply it to the case of asphaltene nanoaggregates in bitumen. The current experimental knowledge on bitumen molecular structure is usually summarized in the Yen–Mullins model,[16−18] which introduces nanoaggregates and clusters. Nanoaggregates are typically composed of less than 10 asphaltene molecules aligned due to π-stacking interactions. Clusters, generally around 5 nm wide, are composed of nanoaggregates and could be created by interactions between asphaltene alkyl side chains.[19] Another model developed by Gray et al.[20] considers more loosely defined structures where many different possible interactions between the asphaltene molecules can be involved. Molecular simulations have also helped to get a picture of the asphaltene nanoaggregates and clusters at the atomic and molecular levels. Many simulation studies were conducted on solutions of asphaltene molecules in a given solvent[5,21−26] and some in more complex bitumen mixtures.[27−32] These studies have shown that the structure and the size of the aggregates depend on the solvent and on the

concentration and structure of the asphaltene molecules themselves. In particular, an asphaltene of the archipelago type, i.e., containing aromatic parts linked by alkyl chains, is known to lead to branched structures.[5] Moreover, long alkyl chains on asphaltene molecules have been seen to reduce the nanoaggregate size through steric hindrance.[5,23] The present work examines molecular dynamics simulations of the Cooee bitumen.[29,33] This model bitumen contains saturated hydrocarbons, and three types of aromatic molecules: resinous oil, resin, and asphaltene. It is able to reproduce generic rheological properties[34] and formation of branched nanoaggregates.[35] Consequently, it constitutes a good model to appreciate the applicability of the generalized statistical model.

The paper is organized as follows. Section 2 describes the structure of the nanoaggregates as seen in the simulations and provides simulations details. In section 3 the generalized statistical model is presented. In section 4 the model is applied to MD results on Cooee bitumen. Finally, section 5 contains a summary and a conclusion.

2. STRUCTURE OF NANOAGGREGATES IN SIMULATIONS OF COOEE BITUMEN

2.1. Nanoaggregate Structure. The simulated system contains four types of molecules, chosen to resemble the SARA classification:[36] a Saturated hydrocarbon, a resinous oil molecule (which is denoted Aromatic in the SARA scheme), a Resin molecule, and an Asphaltene molecule. The molecular structures chosen are shown in Figure 1. Asphaltene, resin, and
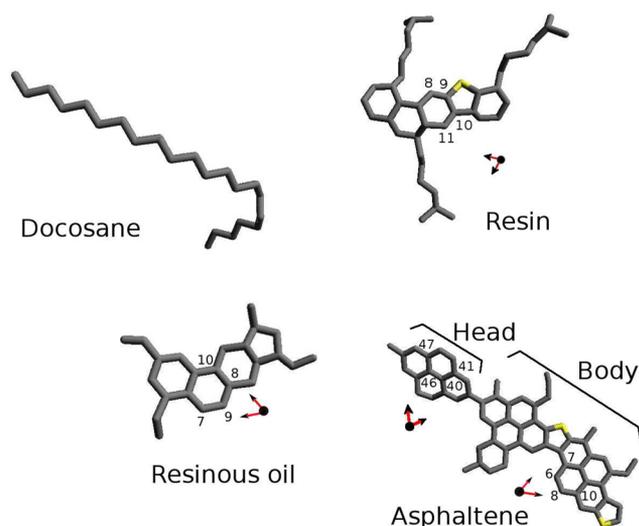

Figure 1. Structure of the four molecules in the "Cooee bitumen" model. Gray edges represent the carbon groups CH$_3$, CH$_2$, and CH and yellow edges represent sulfur atoms. The "head" and "body" of the asphaltene molecule are shown. Numbers and arrows indicate bond-vectors used to quantify the nanoaggregate structure. Reprinted with permission from C. A. Lemarchand et al., *J. Chem. Phys.* **2013**, *139*, 124506. Copyright 2013 American Institute of Physics.

resinous oil molecules are aromatic molecules and are flat to a good approximation. They consequently tend to align with respect to each other.[30] They are found to align at a distance of around 4.0 Å, close to the minimum of the Lennard-Jones potential between the molecules, and in agreement with the experimental literature.[17] The Lennard-Jones potential used in the molecular dynamics (MD) simulations mimics the π-

stacking interaction between aromatic molecules. In the Cooee bitumen model, the alignment of the aromatic molecules is the basis of the nanoaggregate formation. An example of a nanoaggregate found in the simulations is given in Figure 2.
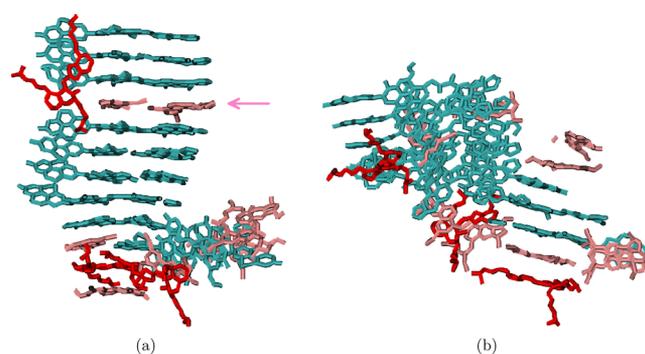

Figure 2. Two snapshots of the same branched nanoaggregate obtained in molecular dynamics. Asphaltene molecules are in blue, resin molecules are in red, and resinous oil molecules are in pink. The pink arrow indicates two resinous oil molecules aligned on the same side of an asphaltene body. The right view is obtained by a 60° rotation of the left view about a horizontal axis.

The definition of a nanoaggregate in the MD simulations is based on the following rule: two aromatic molecules are nearest neighbors in the same nanoaggregate if they are "well-aligned" and "close enough".[30] A nanoaggregate is composed of all the aromatic molecules connected by this rule. Please note here that a nanoaggregate is composed not only of asphaltene molecules but also of resin and resinous oil molecules. More precisely, the neighboring rule is based on three thresholds.[30] The first threshold imposes boundaries to the angle θ between the normal vectors to the aromatic planes of two different molecules. We choose $0° \leq \theta \leq 34°$ and $149° \leq \theta \leq 180°$.[30] The second threshold imposes a maximum value to the distance d_1 between the aromatic planes of two different molecules. We impose $d_1 \leq 6$ Å.[30] The third threshold imposes a maximum value to the distance d_2 between the center of mass of the first molecule and the projection of the center of mass of the second molecule on the plane of the first molecule. We fix $d_2 \leq 0.7d_A$ Å,[30] where d_A is the typical length of an asphaltene molecule and equals to $d_A = 13.1$ Å in the Cooee model.

Parts a−c of Figure 3 show three examples of conformations of two molecules considered as nearest neighbors according to this rule. Conversely, parts d and e of Figure 3 show two conformations where the two molecules are not linked.

Furthermore, the asphaltene molecule chosen in this model has two parts, a flat head and a flat body, which are not aligned. The flat head and the flat body are both considered as aromatic planes subjected to the neighboring rule but counted as a single molecule. The aromatic molecules can align in the direction of an asphaltene body or in the direction of an asphaltene head, thus creating branched nanoaggregates. This branching phenomena could be one of the mechanisms, which cause the formation of the clusters of nanoaggregates. In this paper, the branching phenomena observed in the Cooee bitumen creates a good example to test the generalized statistical model. The statistical model is believed to be general enough to be adapted to other types of asphaltene molecules and other formation mechanisms of clusters. Two views of the same branched nanoaggregate obtained from the MD simulations are shown in Figure 2. The fact that the asphaltene molecule
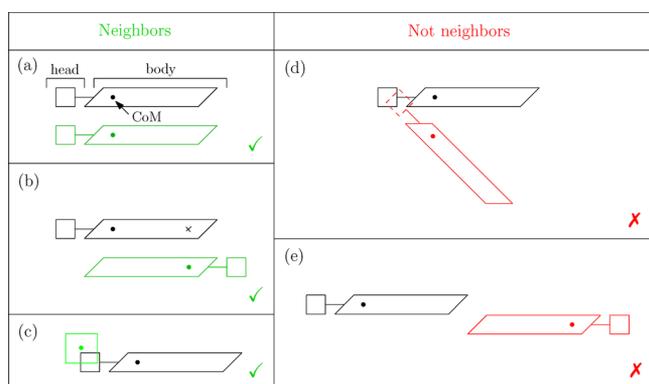
Figure 3. (a)−(c) Scheme of limiting cases for which a molecule (in green) is the nearest neighbor of an asphaltene molecule (in black). (d) and (e) Scheme of limiting cases for which a molecule (in red) is not the nearest neighbor of an asphaltene molecule (in black). The dot represents the center of mass of the molecule. (a) Head-to-head conformation. (b) Head-to-tail conformation. (c) Molecule aligned to the head of the asphaltene molecule. (d) Nonaligned molecules. (e) Molecules far from each other and aligned. Modified with permission from C. A. Lemarchand et al., *J. Chem. Phys.* **2013**, *139*, 124506. Copyright 2013 American Institute of Physics.

chosen here has two flat parts implies that an asphaltene molecule can have at least four nearest neighbors, in contrast to only two nearest neighbors in purely linear nanoaggregates. Moreover, resin and resinous oil molecules have a smaller aromatic structure than asphaltene molecules. Consequently, two resin or resinous oil molecules can align on the same side of an asphaltene body, as illustrated in Figure 2. The maximum number of nearest neighbors of an asphaltene molecule can thus be six. For resin and resinous oil molecules, the maximum number of nearest neighbors is two.

For the sake of brevity, the terms "aggregates" and "nanoaggregates" will be used indistinctly in this paper. The term aggregate should not be confused here with the macroscale rocks glued together by bitumen and constituting the road pavement.

2.2. Simulation Details. The main system studied in this paper contains 410 saturated hydrocarbons, 50 resinous oil molecules, 50 resin molecules, and 50 asphaltene molecules. The methyl (CH_3), methylene (CH_2), and methine (CH) groups are represented by the same united atom unit of molar mass 13.3 g·mol^{-1} and the sulfur atoms are represented by a united atom unit with a different molar mass, 32 g·mol^{-1}. The potential between the united atom units contains four terms: an intermolecular potential, corresponding to a Lennard-Jones potential with parameters $\sigma = 3.75$ Å and $\epsilon/k_B = 75.4$ K, where k_B is the Boltzmann constant, and three terms for the intramolecular potential. These three terms describe the bond length between two connected particles, the angle between three consecutive particles, and the dihedral angle between four consecutive particles. The parametrization of the intramolecular potential is described in detail in ref 29. One bond type is used to represent all bonds. Two angle types are used to represent the angles in an aromatic ring and around a tetrahedral carbon. Three dihedral types are used. They are all proper and correspond to the cis, trans, and anti or gauche conformations. In particular, the dihedral angles at the junction of an aromatic ring are treated with the potential corresponding to anti and gauche conformations. The simulations are performed in the canonical ensemble (*NVT*) at a constant density and six

different temperatures in the range 452 K ≤ T ≤ 641 K. The density ρ = 0.964 kg·L^{-1} is chosen to obtain an average pressure around the atmospheric pressure at T = 452 K. A Nosé−Hoover thermostat is used. The time step is $\Delta t = 0.86$ fs, which corresponds to 0.0005 in reduced units of time. A temperature ramp beginning at the highest considered temperature is carried out to minimize the equilibration time. Each simulation is equilibrated for a period of 17 ns and lasts thereafter 86 ns. Eight independent simulations are performed at the same state point. The molecular dynamics package RUMD[37] is used to perform the calculation.

3. STATISTICAL MODEL FOR BRANCHED AGGREGATES

The statistical model proposed in this work can reproduce the size distribution of branched molecular aggregates containing molecules with an arbitrarily large number of bonds. This model is a generalization of the model designed by Sillren et al.[6] to characterize the equilibrium size distribution of branched monoalcohol clusters. The model of Sillren et al. considers that a molecular unit can have 0, 1, or 2 bonds. In this work, we consider that a molecular unit can have up to m bonds, with $m > 1$.

In the framework of the statistical model, the topology of a given molecular cluster or aggregate is described in the following way. A molecular unit is said to be of type M_i if it has exactly i bonds, with $i = 0, ..., m$. In the model, a bond is oriented and attributed to a single molecular unit. An aggregate starts with a molecular unit called root. If the root is of type M_i, i branches are initiated from the root. The branches can grow linearly by the addition of units of type M_1. They can themselves branch if units of type M_i, with $i \geq 2$, are added. The branches always terminate with a unit of type M_0 with 0 bonds. This last requirement is equivalent to the assumption that the aggregate does not contain any loop; i.e., two branches cannot merge. An example of an aggregate containing molecular units with up to 5 bonds is illustrated in Figure 4.
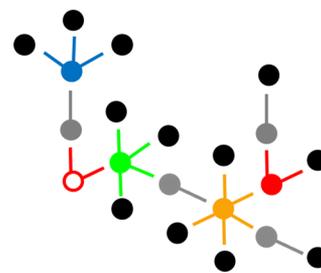

Figure 4. Example of an aggregate with six types of molecular units, which can have from 0 to 5 bonds. The open sphere is the root, here of type M_2. Black spheres represent molecules of type M_0 with zero bonds, gray spheres are molecules of type M_1 with one bond, red spheres are molecules of type M_2 with two bonds, blue spheres are molecules of type M_3 with three bonds, green spheres are molecules of type M_4 with four bonds, and orange spheres are molecules of type M_5 with five bonds.

We denote by n_i, with $i = 0, ..., m$, the number of molecular units of type M_i in a given aggregate and by n the total number of molecular units in the aggregate. By definition, the conservation relation

$$n = \sum_{i=0}^{m} n_i \tag{1}$$

holds. Moreover, as a branch is always terminated by a molecular unit of type M_0, the number n_0 of M_0 unit is

$$n_0 = 1 + \sum_{i=2}^{m} (i-1)n_i \tag{2}$$

The number 1 corresponds to the end of the branch directly opposing the root. Each time a unit of type M_i is added, $i-1$ new branches are initiated and need to be terminated. An aggregate described by the $m + 2$ numbers $(n, n_0, n_1, n_2, ..., n_m)$ can consequently be described by m numbers: $(n, n_2, ..., n_m)$.

The statistical model makes use of two sets of probabilities. The conditional probabilities p_i, with $1 \leq i \leq m$, are the probabilities for a molecular unit to form an ith bond, given that it already has $i-1$ bonds. The probabilities q_i, with $0 \leq i \leq m$, are the probabilities for a molecular unit to have exactly i bonds. In other words, the probabilities q_i are the probabilities for a molecule to have a first, a second, ..., and an ith bond but not an $(i+1)$th. The probabilities q_i and p_i are consequently related by

$$q_0 = 1 - p_1 \tag{3a}$$

$$q_i = (1 - p_{i+1}) \prod_{j=1}^{i} p_j \qquad \text{for } i = 1, ..., m-1 \tag{3b}$$

$$q_m = \prod_{j=1}^{m} p_j \tag{3c}$$

The statistical model assumes that the event defined as "a randomly chosen molecular unit is of type M_i," is independent of the event "a randomly chosen molecular unit is of type M_j" for $i \neq j$. The probability to find an aggregate with n molecular units, among which n_2 are of type M_2, ..., and n_m are of type M_m is then given by a multinomial distribution:

$$P(n, n_2, ..., n_m) = \frac{(n)!}{n} \prod_{i=0}^{m} \frac{q_i^{n_i}}{n_i!} \tag{4}$$

where the q_i's are given in eqs 3a−3c and the dependence of the right-hand side on n_0 and n_1 is suppressed by using eqs 1 and 2. The factor $1/n$ arises from indiscernibility. As a consequence, the probability to find an aggregate with n molecular units is given by

$$P(n) = \sum_{n_2=0}^{n_2^{\max}} ... \sum_{n_m=0}^{n_m^{\max}} P(n, n_2, ..., n_m) \tag{5}$$

where $n_2^{\max} = \left\lfloor \frac{n-1}{2} \right\rfloor$ and $n_i^{\max} = \left\lfloor \frac{n - 1 - \sum_{j=2}^{i-1} j n_j}{i} \right\rfloor$, for $3 \leq i \leq m$. The symbol $\lfloor x \rfloor$ denotes the closest integer lower than x. It is also called the integer part of x and is given by the floor function. The formula given in eq 5 is the main result of the statistical model and can describe any size distribution of branched molecular aggregates, provided that the aggregates do not contain any loop and the molecular types are independent. The average and moments of this distribution are given in the Appendix.

4. APPLICATION TO COOEE BITUMEN

4.1. Results. The size distributions of nanoaggregates obtained in MD simulations are shown in Figure 5 for two
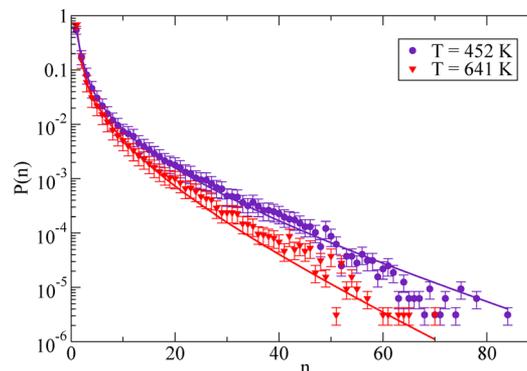

Figure 5. Probability distribution $P(n)$ of obtaining aggregates of n molecular units in molecular dynamics simulations for two temperatures $T = 452$ K and $T = 641$ K. The errors are evaluated by the ratio of the standard deviation and the square root of the number of independent results. The lines are obtained by setting $m = 5$ and using p_i for $i = 1, ..., 5$ as fitting parameters in eq 5. The results obtained for intermediate temperatures are omitted for the sake of clarity.

different temperatures. In the MD simulations, periodic boundary conditions are applied and it can happen that large nanoaggregates span the entire simulation box. However, finite size effects due to these large nanoaggregates are negligible when it comes to the size distributions shown in Figure 5. Two independent arguments can be invoked to support this claim. First, we checked that in a system that is 5 times smaller, the size distributions $P(n)$ of branched nanoaggregates are the same within error bars as those presented in Figure 5 up to $n = 40$. Second, the total number of nanoaggregates spanning the entire box is lower than 1% for all temperatures. Nanoaggregates with a chance to span the entire box are defined by the fact that the distance between the centers of mass of two of their molecules is larger than the box size minus the cutoff distance $d_1 = 6$ Å. These nanoaggregates can be of various sizes ranging from 17 to 82 molecules with an average size of 34 or 36 depending on the temperature.

To fit the probability $P(n)$ given in eq 5 to the MD results, the maximum number m of bonds that a molecular unit can have is chosen to be 5, because in the MD simulations, an asphaltene molecule can have up to 6 neighbors, among which one relates the molecule to the root and the 5 others correspond to 5 oriented bonds attributed to the asphaltene molecule. The fit to the MD results is shown in Figure 5 for two different temperatures. The probabilities p_i for a molecular unit to form an ith bond given that it already has $i-1$ bonds are used as fitting parameters. Each p_i value is deduced from the probability $P(n = i)$ to have an aggregate of i molecular units in the following, recursive way: p_1 is deduced from $P(n = 1) = 1 - p_1$ and p_i, with $i = 2, ..., 5$, is deduced from $P(n = i)$ and the p_j's for $j = 1, ..., i - 1$. Thus, only the first five points of the curves are used for the fit, the rest of the fitting curve follows. Moreover, the fitted p_i values do not depend on the choice of the maximum number of bonds m. Increasing m amounts to increasing the number of fitting parameters and to matching the tail of the distribution more closely, but it does not change the p_i values for $i \leq 5$. The agreement between the model and the MD results with $m = 5$ is good.

Figure 6 displays the temperature variation of the conditional probabilities p_i for i ranging from 1 to 3. We observe that p_1
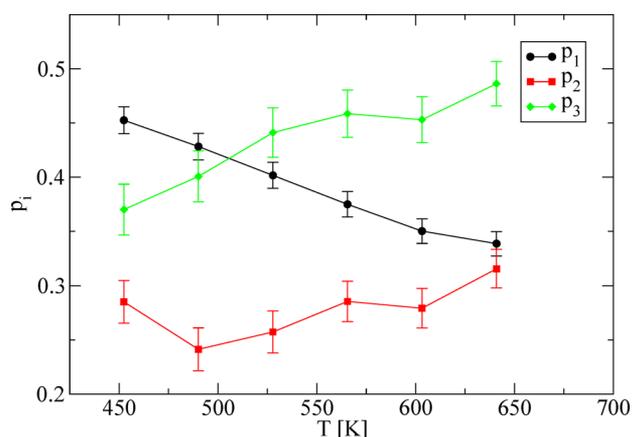

Figure 6. Temperature variation of the conditional probabilities p_i for a molecular unit to form an ith bond given that it has already $i - 1$ bonds, for $i = 1, ..., m$, deduced from molecular dynamics simulations when setting $m = 5$ and using the p_i's as fitting parameters in eq 5. The errors are evaluated by the ratio of the standard deviation and the square root of the number of independent results.

decreases with increasing temperature, which shows that the number of free molecules increases with increasing temperature. This result agrees with the expected behavior according to which thermal noise outcompetes the energy of interaction between molecular units as temperature increases. It is also in agreement with previous molecular dynamics results.[22] The probability p_2 of forming two bonds is rather constant, whereas the probability p_3 increases with increasing temperature. This quite surprising result can be explained in the following way. An aggregate involves several types of molecules of which the asphaltene molecules are the largest. We have proved in a previous work[34] that the fraction of asphaltene molecules in large aggregates increases with increasing temperature, whereas the fraction of resin and resinous oil molecules in large aggregates decreases with increasing temperature. In other words, when temperature increases, only the molecules with the largest aromatic structure and probably the largest interaction energy stay in large aggregates. The smaller molecules are more easily detached from the aggregates by thermal perturbation. As asphaltene molecules are the only considered molecules able to

branch, the higher proportion of asphaltene molecules in large aggregates as temperature increases can be correlated with the increase of the probability p_3 to form three bonds as temperature increases. The values of p_4 and p_5, deduced from the small values of $P(n=4)$ and $P(n=5)$ are too noisy for a conclusion to be drawn.

4.2. Relevance and Limitation of the Model. To evaluate the relevance of the branching model with up to $m = 5$ bonds per molecular unit for the Cooee bitumen, we compare the probabilities q_i deduced from the fit of $P(n)$ with the average fractions $\langle N_i/N \rangle$ of molecular units with i bonds directly determined from the MD configurations. The average is a time and ensemble average over eight different initial configurations. The MD simulations provide the number of links of each molecular unit, which differs from the definition of bonds used to count the different types of molecular unit. Links are shared by two molecules whereas bonds are oriented and attributed to a single molecular unit, as already mentioned. To convert the number of links into the number of bonds per molecular unit, it is necessary to define a root for each aggregate identified in the simulations. The idea of a root comes from the analogy with alcohol clusters, for which the orientation of the hydrogen bonds are obvious from a chemical point of view. Hydrogen bonds can be oriented from O to H for example. In that case, the root is the only molecule in the cluster with a free H. This molecule can have an arbitrary number of molecules linked to its O atom. By analogy with the alcohol clusters, the root is chosen randomly in the MD simulations between all the possible roots, i.e., the molecules that do not have 6 links. Indeed, a molecule with 6 links can only be described as a molecular unit with 5 bonds and the 6th link is a bond attributed to another molecule. This destroys the ability of the molecules with 6 links to be the root. For the root, the number of bonds is equal to the number of links. For the other molecular units, the number of bonds is given by the number of links minus one, the link which relates them through their branch to the root. As shown in Figure 7, the agreement between q_i and $\langle N_i/N \rangle$ is quite good and does not depend on temperature. In particular, the agreement remains acceptable for high values of i, despite the small values of q_4 and q_5 and the large error bars. The very small values of $\langle N_6/N \rangle$ obtained reveal that molecules with 6 bonds are very improbable, which validates the hypothesis of the model stating that a molecular unit can have at most five bonds.
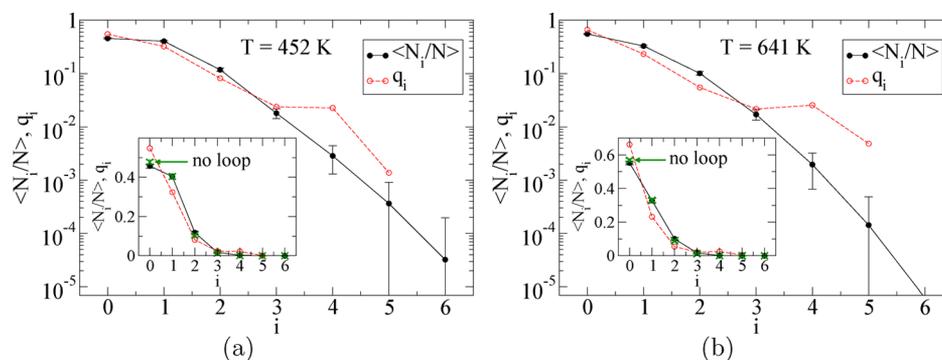

Figure 7. Fraction $\langle N_i/N \rangle$ (filled circles, black solid line) of molecules with i bonds versus i in the MD simulations and probability q_i (open circles, red dashed line) for a molecule to have exactly i bonds deduced from eqs 3a−3c for the p_i values associated with the probability $P(n=i)$ of obtaining aggregates of i molecular units. Temperature is fixed at (a) $T = 452$ K and (b) $T = 641$ K. Insets: same data on a linear scale. The green crosses correspond to MD data when all loops are broken.
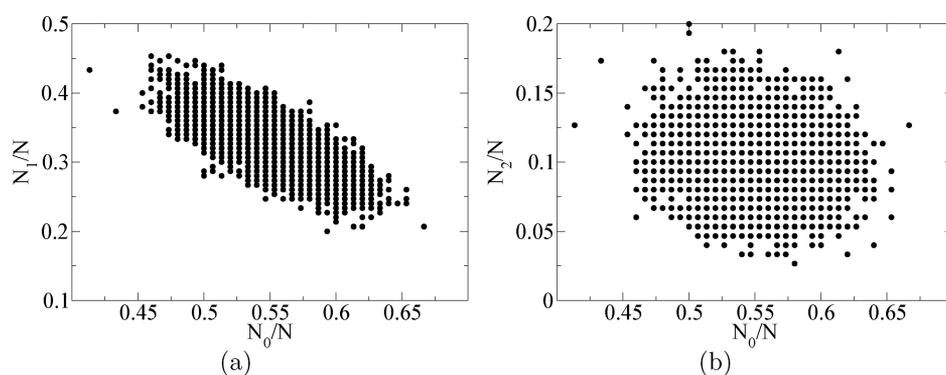
Figure 8. (a) Fraction N_1/N of molecules with 1 bond versus fraction N_0/N of molecules with 0 bonds in the MD simulations. (b) Fraction N_2/N versus fraction N_0/N. Temperature is fixed at $T = 641$ K. The quantities N_0, N_1, N_2, and N are integers, so that the fractions N_0/N, N_1/N, and N_2/N appear to be on a lattice, with a lattice constant equals to $1/N = 1/500$.

The small discrepancy between q_i and $\langle N_i/N \rangle$ for $i \leq 3$ is emphasized in the insets of Figure 7. It cannot be ignored because the relative errors are very small for small i's. Different hypotheses of the model can be reconsidered to account for these small deviations. We first examine whether the architecture of the simulated aggregates differ from the assumed shape with a root and $\left(1 + \sum_{i=2}^{5}(i-1)n_i\right)$ branches. In particular, some aggregates may contain loops formed by merged branches and, consequently, have a number of branches smaller than $\left(1 + \sum_{i=2}^{5}(i-1)n_i\right)$. For every temperature, the number of loops is quantified as the number of links detected in MD between two molecules that are already counted as part of the same nanoaggregate. The fraction of molecules closing a loop is defined as the number of loops found in one configuration over the total number N of molecules. It is found to range from 1.7% to 2.4% depending on temperature. It can account for part of the disagreement between the $\langle N_i/N \rangle$ and q_i but not for all of it. To test this, the ratios $\langle N_i/N \rangle$ are recalculated from the MD simulations in the case where all loops are artificially broken. The results are plotted in the insets of Figure 7. It is clear from this figure that breaking the loops is not enough to recover an agreement between $\langle N_i/N \rangle$ and q_i.

Second, we suspect that, in the closed system considered, the ability of molecular units to create new bonds depends on the other already formed aggregates. More precisely, we reconsider the relevance of the independence of the two following events: "a randomly chosen molecular unit has i bonds" and "a randomly chosen molecular unit has j bonds". Figure 8 reveals that, depending on i and j, some events are correlated. Specifically, we compute all the correlation coefficients

$$C_{ij} = \frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\sqrt{\langle N_i^2 \rangle - \langle N_i \rangle^2} \sqrt{\langle N_j^2 \rangle - \langle N_j \rangle^2}} \quad (6)$$

between the numbers N_i and N_j of molecular units with i and j bonds, respectively. The results are shown in Figure 9. Only C_{01} and C_{12} differ significantly from zero. The negative value of C_{01} is related to the elliptical shape heading down of the scattered plot in Figure 8a . The correlation coefficients are nearly unchanged in the entire temperature range. To check for finite size effects, the correlation coefficients are also calculated at a smaller system size and at one temperature. The results are essentially the same for the smaller system size: only the correlation coefficients C_{01} and C_{12} are nonvanishing. As already mentioned, we also checked that the size distribution $P(n)$ of branched nanoaggregates is the same within error bars
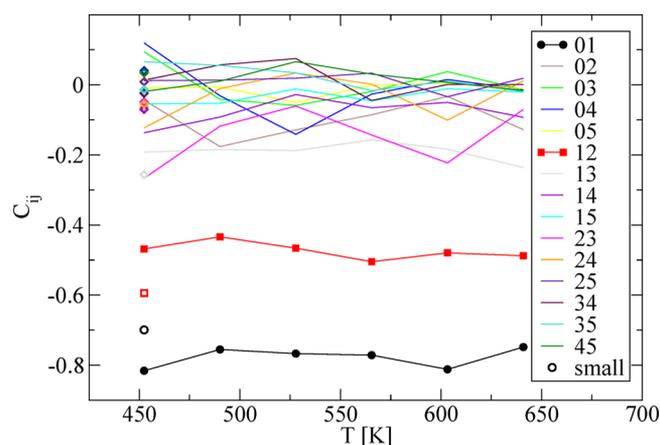

Figure 9. Correlation coefficients C_{ij} between the numbers N_i and N_j of molecules with i and j bonds, respectively, versus temperature in the MD simulations. The lines correspond to the results obtained in a big system of 15570 atoms. The solid symbols refer to C_{01} and C_{12} showing the strongest anticorrelation effect. The open symbols give the results in a small system of 3114 atoms for $T = 452$ K.

up to $n = 40$ for both system sizes. This indicates that the nonvanishing values of the correlation coefficients are probably not due to a finite size effect. The fact that the system is closed, i.e., with a constant total number of molecules, can explain why some correlation coefficients do not vanish. Indeed, the behavior of the correlation coefficient C_{01} can be related to the simultaneous decrease of N_1 and increase of N_0, due to the breaking of linear chains into single molecules. The negative value of C_{12} highlights that the breaking of branched aggregates at the level of a unit M_2, which decreases N_2, mainly results in the formation of linear chains, which increases N_1, rather than in the formation of single molecules M_0. Indeed, this last type of events would have led to a negative value of C_{02}, which is not observed. The vanishing values of C_{ij} for $(i, j) \neq (0, 1)$ and $(1, 2)$ reveals that molecular unit of type M_i with $i > 2$ can change into any other type M_j of molecular units without significant preference.

To conclude on the limitations of the model, one can say that even if the assumption of independent events is not strictly true in the case of a closed system of asphaltene nanoaggregates as seen in MD simulations of Cooee bitumen, it does not seem to call into question the functional form given in eq 5. It shifts slightly the values of the probabilities q_i from their expected values. However, the correlations themselves can be used to

gain insights about the formation mechanism of the branched nanoaggregates.

5. SUMMARY AND CONCLUSION

We proposed a statistical model describing the size distribution of branched molecular aggregates with molecular units having up to m bonds, with $m > 1$. The model has m parameters, namely the probabilities p_i, which correspond to the conditional probabilities for a molecule to form an ith bond given that it already has $i - 1$ bonds. The model is based on two main assumptions: (i) the molecular aggregates do not contain any loop; (ii) the different molecular types are independent. This statistical model can be used *a priori* to describe the size distribution of any branched aggregates. It is used in this paper to reproduce the equilibrium size distribution of branched nanoaggregates as observed in molecular dynamics (MD) simulations of Cooee bitumen. The MD results satisfactorily compare with the model, even though the assumption of independent events is not strictly verified in the MD simulations. The model allows us to recognize the conditional probabilities p_i as relevant parameters from a statistical mechanics point of view. The probability p_1 to form a first bond decreases with increasing temperature, because of thermal noise, whereas the probability p_3 to form a third bond increases with increasing temperature, as the proportion of asphaltene molecules in the aggregate increases. A study of the variation of the conditional probabilities p_i with the chemical composition of the mixture would be of interest. The addition of asphaltene molecules in the system is expected to increase p_3, because of the branching abilities of these molecules. On the contrary, the addition of resin and resinous oil molecules is expected to both reduce p_3, because of a lack of branching abilities of these molecules, and reduce p_1, because their free energy of self-assembly is smaller than that of asphaltene molecules. The statistical model also motivates a deeper study of the correlations between different types of molecular units, which in turn can lead to new insights on the mechanism of formation of the branched aggregates.

The statistical model can *a priori* be used to explain experimental data on bitumen or solutions of asphaltene molecules. Much progress has been done lately to quantify the distribution in molecular weight of asphaltene nanoaggregates in solutions of asphaltene molecules using small angle neutron and X-ray scattering,[38] laser desorption/ionization mass spectrometry,[39] and field desorption mass spectrometry.[40] However, the question of how the distribution obtained experimentally exactly relates to the probability of having a nanoaggregate of a given size is unresolved because it can depend on the experimental technique. Moreover, most of the experimental literature focuses on the nanoaggregate size in solutions of asphaltene molecules. The nanoaggregate size in bitumen mixtures can be quite different.

Furthermore, the conditional probabilities p_i for a molecule to form an ith bond can be extended to probabilities of transition between types of molecular units in the framework of a master equation approach.[41,42] The master equation associated with interacting molecules would provide a mesoscopic description of the dynamics of the system. In particular, it could reproduce the equilibrium size distribution as well as the equilibrium correlations between some types of molecular units observed in the MD simulations of a closed system. In addition, the master equation approach could recreate the dynamics of relaxation of the system toward the equilibrium size distribution.

Once the fundamental reactions between molecule types are uncovered by the master equation framework, a complementary thermodynamics approach based on the same mechanism can be introduced. This thermodynamics approach could be on the line of that developed in refs 43 and 44 but based on the precise mechanism deduced from the MD simulations. Furthermore, the free energies of interaction necessary to the thermodynamics approach could be independently checked by MD simulations using for example umbrella sampling, as was carefully done for some asphaltene molecules and some geometries of interaction in ref 45.

■ MOMENTS OF THE DISTRIBUTION OF THE BRANCHED AGGREGATE SIZE

To obtain the moments of the distribution defined in eqs 4 and 5, we use an approach also suggested by Sillren et al.[6] It consists in differentiating the probability $P(n)$ with respect to the m independent parameters $q_1, ..., q_m$, before performing the average of the equation thus obtained. The procedure is detailed for the average of the distribution and summarized for the other moments in the following.

According to eqs 4 and 5, the probability $P(n)$ can be written as

$$P(n) = \sum_{n_2=0}^{n_2^{max}} \cdots \sum_{n_m=0}^{n_m^{max}} \frac{(n)!}{n} \prod_{i=0}^{m} \frac{q_i^{n_i}}{n_i!} \tag{7}$$

Differentiating with respect to q_j, with $j \geq 1$, yields

$$\frac{\partial P(n)}{\partial q_j} = \sum_{n_2=0}^{n_2^{max}} \cdots \sum_{n_m=0}^{n_m^{max}} P(n, n_2, ..., n_m) \left(\frac{n_j}{q_j} - \frac{n_0}{q_0} \right) \tag{8}$$

where the term n_0/q_0 comes from the use of $q_0 = 1 - \sum_{i=1}^{m} q_i$. Summing over n from 1 to ∞ gives

$$\frac{\langle n_j \rangle}{q_j} - \frac{\langle n_0 \rangle}{q_0} = 0 \tag{9}$$

where the symbol $\langle \cdot \rangle$ denotes the average over the distribution. We have used the fact that $\sum_{n=1}^{\infty} \partial P(n)/\partial q_j = \partial 1/\partial q_j = 0$. We also know that the following conservation relations are true for every aggregate and also on average:

$$\langle n_0 \rangle = 1 + \sum_{i=2}^{m} (i - 1)\langle n_i \rangle \tag{10}$$

$$\langle n_1 \rangle = \langle n \rangle - 1 - \sum_{i=2}^{m} i\langle n_i \rangle \tag{11}$$

Equations 9−11 constitute a set of $m + 2$ linear equations with $m + 2$ unknowns, namely $\langle n \rangle$, $\langle n_0 \rangle$, ..., $\langle n_m \rangle$. The following procedure can be used to solve it. Using eqs 10 and 9 yields

$$\langle n_0 \rangle = \frac{q_0}{q_0 - \sum_{i=2}^{m} (i - 1)q_i} \tag{12}$$

Moreover, using eq 9 in the case $j = 1$ and eq 11 gives

$$\frac{\langle n \rangle}{q_1} - \frac{1}{q_1} - \sum_{i=2}^{m} \frac{i\langle n_i \rangle}{q_1} - \frac{\langle n_0 \rangle}{q_0} = 0 \tag{13}$$

Inserting the expression of $\langle n_j \rangle$ in terms of $\langle n_0 \rangle$ given in eq 9 yields

$$\langle n \rangle = 1 + \sum_{i=1}^{m} i \frac{q_i}{q_0} \langle n_0 \rangle \tag{14}$$

Finally, using the expression of $\langle n_0 \rangle$ given in eq 12 yields the result for the average $\langle n \rangle$ of the distribution:

$$\langle n \rangle = \frac{1}{1 - \sum_{i=1}^{m} j q_j} \tag{15}$$

To obtain the kth moment of the distribution in terms of the $(k-1)$th moment, we start again from eq 8. Multiplying by n^{k-1}, $k \geq 1$, and summing over n from 1 to ∞ yields

$$\frac{\partial \langle n^{k-1} \rangle}{\partial q_j} = \frac{\langle n^{k-1} n_j \rangle}{q_j} - \frac{\langle n^{k-1} n_0 \rangle}{q_0} \qquad j \geq 1 \tag{16}$$

If $\partial \langle n^{k-1} \rangle / \partial q_j$ is known, eq 16 and the conservation relations eqs 10 and 11 constitute a set of $m+2$ linear equations with $m+2$ unknowns, namely, $\langle n^k \rangle$, $\langle n^{k-1} n_0 \rangle$, ..., $\langle n^{k-1} n_m \rangle$. It can be solved using a procedure very similar to that used to obtain the average $\langle n \rangle$. The result is a recursive formula for the kth moment:

$$\langle n^k \rangle = q_1 \partial_{q_1} \langle n^{k-1} \rangle + \langle n^{k-1} \rangle + \sum_{i=2}^{m} i q_i \partial_{q_i} \langle n^{k-1} \rangle$$
$$+ \sum_{i=1}^{m} i q_i \frac{\langle n^{k-1} \rangle + \sum_{j=1}^{m} (j-1) q_j \partial_{q_j} \langle n^{k-1} \rangle}{1 - \sum_{j=1}^{m} j q_j} \tag{17}$$

where the symbol ∂_{q_i} denotes the partial derivative with respect to q_i. This equation can be used to obtain for example the second moment:

$$\langle n^2 \rangle = \langle n \rangle \left(3 \langle n \rangle - 1 + \langle n \rangle^2 \left(-1 + \sum_{i=1}^{m} i^2 q_i \right) \right) \tag{18}$$

■ AUTHOR INFORMATION

Corresponding Author
*C. A. Lemarchand. E-mail: clairel@ruc.dk.
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Hummelen, J. C.; van Dongen, J. L. J.; Meijer, E. W. Electrospray Mass Spectrometry of Poly(propylene imine) Dendrimers-The Issue of Dendritic Purity or Polydispersity. *Chem. - Eur. J.* **1997**, *3*, 1489–1493.

(2) Yu, H.; Schlüter, A. D.; Zhang, B. Synthesis of High Generation Dendronized Polymers and Quantification of Their Structure Perfection. *Macromolecules* **2014**, *47*, 4127–4135.

(3) Buléon, A.; Colonna, P.; Planchot, V.; Ball, S. Starch Granules: Structure and Biosynthesis. *Int. J. Biol. Macromol.* **1998**, *23*, 85–112.

(4) Thompson, D. B. On the Non-Random Nature of Amylopectin Branching. *Carbohydr. Polym.* **2000**, *43*, 223–239.

(5) Murgich, J. Molecular Simulation and the Aggregation of the Heavy Fractions in Crude Oils. *Mol. Simul.* **2003**, *29*, 451–461.

(6) Sillren, P.; Bielecki, J.; Mattson, J.; Börjesson, L.; Matic, A. A Statistical Model of Hydrogen Bond Networks in Liquid Alcohols. *J. Chem. Phys.* **2012**, *136*, 094514.

(7) Wertheim, M. S. Fluids with Highly Directional Attractive Forces. I. Statistical Thermodynamics. *J. Stat. Phys.* **1984**, *35*, 19–34.

(8) Wertheim, M. S. Fluids with Highly Directional Attractive Forces. II. Thermodynamic Perturbation Theory and Integral Equations. *J. Stat. Phys.* **1984**, *35*, 35–47.

(9) Wertheim, M. S. Fluids with Highly Directional Attractive Forces. III. Multiple Attraction Sites. *J. Stat. Phys.* **1986**, *42*, 459–476.

(10) Wertheim, M. S. Fluids with Highly Directional Attractive Forces. IV. Equilibrium Polymerization. *J. Stat. Phys.* **1986**, *42*, 477–492.

(11) Gubbins, K. E. The theory of Non-Electrolyte Solutions: an Historical Review. *Mol. Phys.* **2013**, *111*, 3666–3697.

(12) Stockmayer, W. H. Theory of Molecular Size Distribution and Gel Formation in Branched-Chain Polymers. *J. Chem. Phys.* **1943**, *11*, 45–55.

(13) Flory, P. J. Molecular Size Distribution in Three Dimensional Polymers. VI. Branched Polymers Containing A-R-Bf-1 Type Units. *J. Am. Chem. Soc.* **1952**, *74*, 2718–2723.

(14) Donoghue, E.; Gibbs, J. H. Mean Molecular Size Distributions and the Sol-Gel Transition in Finite, Polycondensing Systems. *J. Chem. Phys.* **1979**, *70*, 2346.

(15) Flory, P. J. Molecular Size Distribution in Linear Condensation Polymers. *J. Am. Chem. Soc.* **1936**, *58*, 1877–1885.

(16) Yen, T. F.; Erdman, J. G.; Pollack, S. S. Investigation of the Structure of Petroleum Asphaltenes by X-Ray Diffraction. *Anal. Chem.* **1961**, *33*, 1587–1594.

(17) Mullins, O. C.; Sabbah, H.; Eyssautier, J.; Pomerantz, A. E.; Barré, L.; Andrews, A. B.; Ruiz-Morales, Y.; Mostowfi, F.; McFarlane, R.; Goual, L.; et al. Advances in Asphaltene Science and the Yen-Mullins Model. *Energy Fuels* **2012**, *26*, 3986–4003.

(18) Hosseini-Dastgerdi, Z.; Tabatabaei-Nejad, S. A. R.; Khodapanah, E.; Sahraei, E. A Comprehensive Study on Mechanism of Formation and Techniques to Diagnose Asphaltene Structure; Molecular and Aggregates: a Review. *Asia-Pac. J. Chem. Eng.* **2015**, *10*, 1–14.

(19) Majumdar, R. D.; Gerken, M.; Hazendonk, P. Solid-State 1H and 13C Nuclear Magnetic Resonance Spectroscopy of Athabasca Oil Sands Asphaltenes: Evidence for Interlocking Pi-Stacked Nano-aggregates with Intercalated Alkyl Side Chains. *Energy Fuels* **2015**, *29*, 2790–2800.

(20) Gray, M. R.; Tykwinski, R. R.; Stryker, J. M.; Tan, X. Supramolecular Assembly Model for Aggregation of Petroleum Asphaltenes. *Energy Fuels* **2011**, *25*, 3125–3134.

(21) Rogel, E. Studies on Asphaltene Aggregation via Computational Chemistry. *Colloids Surf., A* **1995**, *104*, 85–93.

(22) Pacheco-Sánchez, J. H.; Zaragoza, I. P.; Martínez-Magadán, J. M. Asphaltene Aggregation under Vacuum at Different Temperatures by Molecular Dynamics. *Energy Fuels* **2003**, *17*, 1346–1355.

(23) Teklebrhan, R. B.; Ge, L.; Bhattacharjee, S.; Xu, Z.; Sjöblom, J. Probing Structure-Nanoaggregation Relations of Polyaromatic Surfactants: A Molecular Dynamics Simulation and Dynamic Light Scattering Study. *J. Phys. Chem. B* **2012**, *116*, 5907–5918.

(24) Ungerer, P.; Rigby, D.; Leblanc, B.; Yiannourakou, M. Sensitivity of the Aggregation Behaviour of Asphaltenes to Molecular Weight and Structure Using Molecular Dynamics. *Mol. Simul.* **2014**, *40*, 115–122.

(25) Wang, S.; Xu, J.; Wen, H. The Aggregation and Diffusion of Asphaltenes Studied by GPU-Accelerated Dissipative Particle Dynamics. *Comput. Phys. Commun.* **2014**, *185*, 3069−3078.

(26) Ruiz-Morales, Y.; Mullins, O. C. Coarse-Grained Molecular Simulations to Investigate Asphaltenes at the Oil-Water Interface. *Energy Fuels* **2015**, *29*, 1597−1609.

(27) Zhang, L.; Greenfield, M. L. Molecular Orientation in Model Asphalts Using Molecular Simulation. *Energy Fuels* **2007**, *21*, 1102−1111.

(28) Greenfield, M. L. Molecular Modelling and Simulation of Asphaltenes and Bituminous Materials. *Int. J. Pavement Eng.* **2011**, *12*, 325−341.

(29) Hansen, J. S.; Lemarchand, C. A.; Nielsen, E.; Dyre, J. C.; Schrøder, T. B. Four-Component United-Atom Model of Bitumen. *J. Chem. Phys.* **2013**, *138*, 094508.

(30) Lemarchand, C. A.; Schrøder, T. B.; Dyre, J. C.; Hansen, J. S. Cooee Bitumen: Chemical Aging. *J. Chem. Phys.* **2013**, *139*, 124506.

(31) Li, D. D.; Greenfield, M. L. Viscosity, Relaxation Time, and Dynamics within a Model Asphalt of Larger Molecules. *J. Chem. Phys.* **2014**, *140*, 034507.

(32) Li, D. D.; Greenfield, M. L. Chemical Compositions of Improved Model Asphalt Systems for Molecular Simulations. *Fuel* **2014**, *115*, 347−356.

(33) CO_2 Emission Reduction by Exploitation of Rolling Resistance Modelling of Pavements, http://www.cooee-co2.dk/.

(34) Lemarchand, C. A.; Bailey, N. P.; Todd, B. D.; Daivis, P. J.; Hansen, J. S. Non-Newtonian Behavior and Molecular Structure of Cooee Bitumen under Shear Flow: A Non-Equilibrium Molecular Dynamics Study. *J. Chem. Phys.* **2015**, *142*, 244501.

(35) Lemarchand, C. A.; Schrøder, T. B.; Dyre, J. C.; Hansen, J. S. Cooee Bitumen. II. Stability of Linear Asphaltene Nanoaggregates. *J. Chem. Phys.* **2014**, *141*, 144308.

(36) ASTM. *Annual Book of Standards*; American Society for Testing and Materials: Philadelphia, PA, 1995; method D-2007.

(37) Bailey, N. P.; Ingebrigtsen, T. S.; Hansen, J. S.; Veldhorst, A. A.; Bøhling, L.; Lemarchand, C. A.; Olsen, A. E.; Bacher, A. K.; Larsen, H.; Dyre, J. C.; et al. RUMD: A General Purpose Molecular Dynamics Package Optimized to Utilize GPU Hardware Down to a few Thousand Particles. (submitted for publication to *J. Comput. Chem.* (see also http://rumd.org).

(38) Eyssautier, J.; Levitz, P.; Espinat, D.; Jestin, J.; Gummel, J.; Grillo, I.; Barré, L. Insight into Asphaltene Nanoaggregate Structure Inferred by Small Angle Neutron and X-ray Scattering. *J. Phys. Chem. B* **2011**, *115*, 6827−6837.

(39) Martínez-Haya, B.; Hortal, A. R.; Hurtado, P.; Lobato, M. D.; Pedrosa, J. M. Laser Desorption/Ionization Determination of Molecular Weight Distributions of Polyaromatic Carbonaceous Compounds and their Aggregates. *J. Mass Spectrom.* **2007**, *42*, 701−713.

(40) Qian, K.; Edwards, K. E.; Siskin, M.; Olmstead, W. N.; Mennito, A. S.; Dechert, G. J.; Hoosain, N. E. Desorption and Ionization of Heavy Petroleum Molecules and Measurement of Molecular Weight Distributions. *Energy Fuels* **2007**, *21*, 1042−1047.

(41) Gardiner, C. W. *Handbook Of Stochastic Methods For Physics, Chemistry, And The Natural Sciences*; Springer-Verlag: Berlin, Heidelberg, Germany, 1985.

(42) Heuett, W. J.; Qian, H. Grand Canonical Markov Model: A Stochastic Theory for Open Nonequilibrium Biochemical Networks. *J. Chem. Phys.* **2006**, *124*, 044110.

(43) Aguilera-Mercado, B.; Herdes, C.; Murgich, J.; Müller, E. A. Mesoscopic Simulation of Aggregation of Asphaltene and Resin Molecules in Crude Oils. *Energy Fuels* **2006**, *20*, 327−338.

(44) Acevedo, S.; Caetano, M.; Ranaudo, M. A.; Jaimes, B. Simulation of Asphaltene Aggregation and Related Properties Using an Equilibrium-Based Mathematical Model. *Energy Fuels* **2011**, *25*, 3544−3551.

(45) Sedghi, M.; Goual, L.; Welch, W.; Kubelka, J. Effect of Asphaltene Structure on Association and Aggregation Using Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117*, 5765−5776.