A Catalogue of Plausible Molecular Models for the Molecular Dynamics of Asphaltenes and Resins obtained from Quantitative Molecular Representation

Jason C. Law[1], Thomas F. Headen[1,‡], Guadalupe Jiménez-Serratos[1], Edo S. Boek[1,†], Juan Murgich[2] and Erich A. Müller[1,*]

[1] Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom

[2] Calle Hermosilla 75, 5-20, 28001 Madrid, Spain

Keywords: asphaltenes, resins, molecular dynamics (MD) simulation, molecular structures

‡ Present address: ISIS Neutron and Muon Source, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Campus, Didcot OX11 0QX, U.K.

† Present address: Division of Chemical Engineering & Renewable Energy, School of Engineering and Materials Science, Queen Mary University of London, London E1 4NS, U.K.

* Corresponding author: e.muller@imperial.ac.uk

ABSTRACT

Computer simulation studies aimed at elucidating the phase behavior of crude oils inevitably require atomistically-detailed models of representative molecules. For the lighter fractions of crudes, such molecules are readily available, as the chemical composition can be resolved experimentally. Heavier fractions pose a challenge, on one hand due to their polydispersity and on the other due to poor description of the morphology of the molecules involved. The Quantitative Molecular Representation (QMR) approach is used here to generate a catalogue of 100 plausible asphaltene and resin structures based on elemental analysis and ^{1}H – ^{13}C NMR spectroscopy experimental data. The computer-generated models are compared in the context of a review of previously proposed literature structures and categorized by employing their molecular weights, double bond equivalents (DBE) and hydrogen to carbon (H/C) ratios. Sample atomistic molecular dynamics simulations were carried out for two of the proposed asphaltene structures with contrasting morphologies, one island-type and one archipelago-type, at 7 wt% in either toluene or heptane. Both asphaltene models, which shared many characteristics in terms of average molecular weight, chemical composition and solubility parameters showed marked differences in their aggregation behavior. The example showcases the importance of considering diversity and polydispersity when considering molecular models of heavy fractions.

INTRODUCTION

The molecular modelling of petroleum fluids imposes particular challenges that do not surface when studying other complex fluids. Crude oils are mixtures of many hundred thousand or more constituents with an extraordinary polydispersity in terms of molecular weight and morphology, albeit some remarkable similarities in terms of chemical nature. While relatively routine techniques such as gas chromatography provide information on the lighter ends, which can be uniquely identified, the intermediate molecular weight fractions become increasingly complex to isolate and identify. Through thermal fractionation, even more information can be garnered about some of the intermediate molecular weight fractions, and a rather fine "lumping" of some of the presumed components can be made into pseudo-components. However, even up to this intermediate molecular weight, the complexity of the saturate and the aromatic fractions becomes evident.[1,2] For the heavier fractions (i.e. those characterized by 30 or more heavy atoms) it is customary to "give up" in the attempt to describe the individual constituents of the mixture and the characterization is made in terms of a practical (but very unsatisfactory from a molecular viewpoint) fractionation into resins and asphaltenes.

Asphaltenes are the heaviest of the polar constituents in crude oil, defined in terms of their insolubility in nonpolar solvents n-heptane while being soluble in aromatic solvents, such as toluene and benzene.[3] They are of particular concern in the oil industry as a fraction of the asphaltenes of certain crudes may precipitate spontaneously upon decompression during production. Their behavior has puzzled researchers for decades: many crudes with a high asphaltene content are produced for decades without any precipitation of solids (i.e. Boscan crudes, which may have up to 17 wt% of asphaltenes), while some light Arabian crudes with only a few percent of asphaltenes are particularly prone to solid deposition. Solid organic deposits are formed either by the crystallization of paraffins and/or the precipitation of certain types of asphaltenes, which can plug the porous media of the reservoir and/or pipes, resulting in increased efforts in oil extraction, decreased production rates and negative financial effects.[4]

The complexity of the asphaltene fraction, the fact that its geological history and origin are determinant and the insistence of attempting to assign a "universal" model molecule to what is in essence a solubility class, have all conspired against obtaining clear insights on the phase behavior

of these fractions. Discussions focusing on quantifying apparently simple general properties, such as the average molecular weight of an asphaltene, have become open, and sometimes bitter debates.[5,6,7] The efforts to experimentally determine the nature of the "prototypical" asphaltene molecule have been metaphorized[8] by using the Hindu tale of a group blind men attempting to agree on the description of an elephant after each of them only had the experience of each touching a different part of the animal; proclaiming that the elephant is like a tree (by whom had touched its legs), or like a snake (by whom had touched its nose), a wall (by touching its side), and so forth. In analogy, the only possible way to describe asphaltene fractions (or molecules) is to take into account a large amount of inferred experimental information to build a picture of the most likely scenario. Despite decades of extensive research involving both experimental work and theoretical modelling, many aspects of the physical behavior and aggregation mechanisms of asphaltenes remain inconclusive, such as the discrepancy in aggregation number and cluster size between some experimental X-ray and neutron scattering results.[9] Noteworthy, high resolution atomic force microscopy (AFM) studies,[10,11] have provided a unique vision (in the literal sense of the word) of the heavy fractions of crude oils molecules, suggesting some of them present cores composed primarily of polynuclear aromatic hydrocarbons (PAH) with side aliphatic chains.[12] This model is usually referred to as an "island" morphology, as opposed to the "archipelago" structures where a larger number of smaller aromatic moieties are linked together by aliphatic chains. The core structures "seen" with AFM have been used as the basis for proposing asphaltene models.[13] The most recent comprehensive characterizations performed using extrographic fractionation and ultrahigh-resolution mass spectrometry seem to indicate that, as one would expect, asphaltenes comprise a structural continuum of both continental (island) and archipelago motifs.[14, 15, 16]

In this scenario, not surprisingly, the theoretical descriptions lag behind. Any equation-of-state modelling of asphaltene phase behavior and/or deposition process will invariably suffer from the lack of an appropriate description of what, in essence, is a continuous molecular distribution, and as such, current models struggle to provide quantitative predictions.[17] Molecular modelling can come to aid in the understanding of the physical behavior of asphaltenes in solutions. The first step to simulating a crude oil, however, is to construct accurate molecular structures of asphaltenes from our limited anecdotal evidence. This contribution is aimed at providing a discrete set of plausible molecular models that can be used directly in the molecular simulation of crude oils.

MOLECULAR MODELS FOR ASPHALTENES

A Brief Historical Perspective

The first suggestions

A variety of model asphaltenes have been published in the literature representing a wide range of molecular weights and structural types, all based either on particular experimental data and/or heuristic and plausibility arguments. Here we do not attempt to make an abridged review of the literature on the subject nor a judgmental analysis, but rather focus on the (sometimes circular) progress that has been made in the past decades in the elucidation of asphaltene and resin molecular structures. More extensive reviews of molecular modelling of asphaltenes can be found in references 18, 19 and others included in this manuscript. The inclusion (or exclusion) of a particular model in this discussion is not a prejudice for or against any particular model and/or research group. It is worth noting, however, that the PAHs in multiple reprinted drawings do not follow Clar's rule,[20,21] show atypical numbers of bonds with the neighboring aromatic rings, and/or are not Kekulé resonance structures; figures with such PAHs are marked with asterisks (*). Models structures marked in such a way should not be employed as presented and should be modified appropriately. Separately, some proposed models do not contain heteroatoms and lack polarity; figures with such models are marked with daggers (†).

In 1974, Yen[22] proposed arguably the first reported "drawing" of an asphaltene model, a hypothetical structure from a Lagunillas oil shown in Figure 1. It is a polymeric structure where the monomer unit has an empirical formula of $C_{74}H_{87}NS_2O$ and a molecular weight of 1071 Da; the positions of all substituents are arbitrary, and are pieced together by analyzing five major hydrogen types detected using nuclear magnetic resonance (NMR) – aromatic, α-substituted, naphthenic, methylene and saturated methyl types. It has historically been misinterpreted as a single molecule, when in fact the author describes asphaltenes as having at least three said monomers. A similarly related model, proposed by Speight and Moschopedis[23] is shown in Figure 2. With a molecular formula of $(C_{79}H_{92}N_2S_2O)_3$ and a molecular weight of 3449 Da, the model was constructed with the assumption that asphaltenes, viewed structurally, were condensed polynuclear aromatic ring systems bearing alkyl side-chains. The use of NMR data along with direct observation with scanning tunneling microscopy led to hypothetical asphaltene structures from a Mayan crude oil, known as the Zajac model.[24] A widely cited version of the Zajac model

(Figure 3) consists of a molecular formula of $C_{54}H_{57}NS$ and a molecular weight of 752 Da, with nine aromatic rings, two alicyclic rings in the fused aromatic region, a single aliphatic chain containing an isopropyl group, and nitrogen and sulfur heteroatoms in pyridinic and thiophenic forms, respectively.[25] The original authors noted that small structures were found in the presence of much larger structures, but could not determine whether the large structures were aggregates of smaller units or large single structures connected by aliphatic linkages. The latter option would eventually be known as the archipelago structural type, which was not widely recognized at the time. Figure 4 shows a molecular representation of a Ratawi asphaltene with a molecular formula of $C_{70}H_{94}$ and molecular weight of 944 Da drawn based on liquid- and solid-state NMR, proposed by Storm et al.[26] The model relied on using of the chemical relationship between the amounts of bridgehead carbons and the size of certain polynuclear aromatics, and suggested the use of smaller molecular sizes and shorter aliphatic groups.

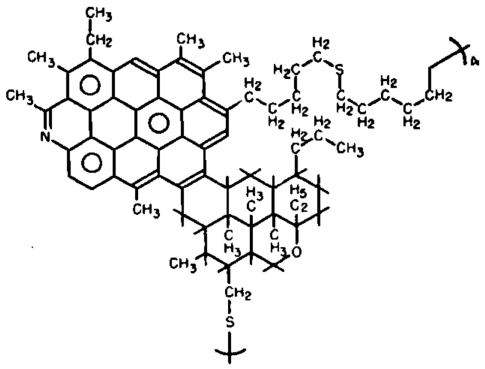


Figure 1. A monomer unit from one of the first depictions of asphaltenes corresponding to the formula $C_{74}H_{8}7NS_2O$, from a Lagunillas oil proposed by Yen.[22]

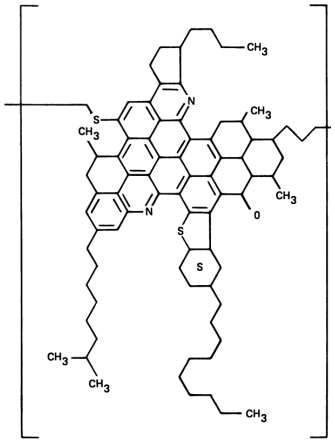


Figure 2. Hypothetical structure of a segment (monomer) of a petroleum asphaltene, $(C_{79}H_{92}N_2S_2O)_3$, proposed by Speight and Moschopedis.[23] Reprinted with permission from ref. 23. Copyright 1982 American Chemical Society.

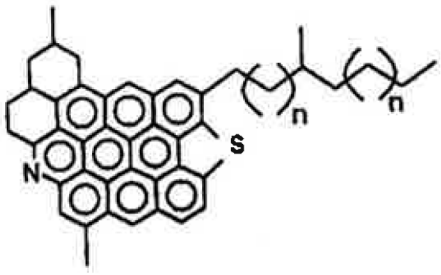


Figure 3. (*) Hypothetical asphaltene molecule structure proposed by Zajac et al.[24.] n=7 aliphatic carbon atoms. Reprinted with permission from the author.

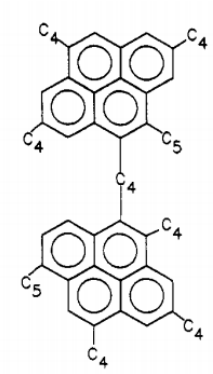


Figure 4. (†) Molecular representations of a Ratawi asphaltene based on liquid- and solid-state NMR, proposed by Storm et al.[26] Reprinted with permission from ref. 26 Copyright 1994 American Chemical Society.

The mega-asphaltene proposals

One of the first and possibly most long-standing debates in the asphaltene literature surrounds the specification of an average molecular weight for prototypical asphaltenes. Reported molecular weights of petroleum asphaltenes range anywhere from 400 Da to 10 000 Da, although most current estimates are closer to the 800 - 1500 Da average value.[27,28,29,30,31,32] In context, the larger values contributed to the proposal of molecular pictures of large asphaltenes, such as those provided by INTEVEP-Petróleos de Venezuela S. A. and investigated by researchers from the Venezuelan Institute of Scientific Research (IVIC). These were models of a "continental" morphology (similar to "island" motifs, but with a much larger PAH core). An example is given in Figure 5, where two-dimensional and space-filling drawings of an asphaltene studied by Murgich et al.[33] are shown. These are among the earliest models of asphaltenes to be employed directly in molecular modelling, which constituted a "heroic" computational effort at the time. The asphaltene structure in Figure 5 has a molecular formula of $C_{149}H_{175}N_3O_2S_2$ and a molecular weight of 2104 Da. The space-filling drawing of the asphaltene structure demonstrates the complexity and stability of the conformation, and the models were reported to have maintained similar appearances while interacting with toluene, cyclohexane and *n*-octane solvents. Another example is given in Figure 6, where one of the several structures for the asphaltene samples from Venezuelan crudes proposed by Rogel and Carbognani [34] is shown. Molecular dynamics simulations were carried out to estimate the densities of the structures, which were found to show the correct tendencies, but were often underestimated compared to experimental values. The authors reported that asphaltenes from unstable crude oils and deposits exhibited higher densities, higher aromaticities, and lower hydrogen-to-carbon ratios than asphaltenes from stable crude oils.

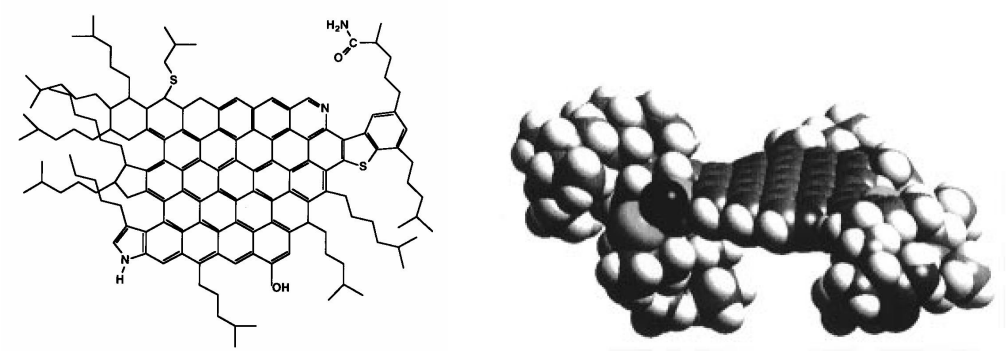


Figure 5. Model asphaltene structure, $C_{149}H_{175}N_3O_2S_2$, studied by Murgich et al.[33] This was one of the first models used in molecular mechanics simulations. Reprinted with permission from ref. 33. Copyright 1996 American Chemical Society.

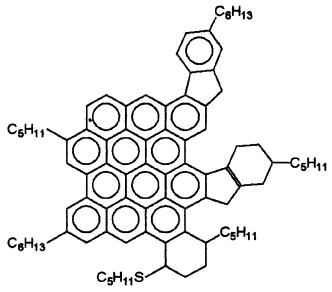


Figure 6. (*†) Model MG27, one of several average structures for the asphaltene samples proposed by Rogel and Carbognani [34] with a formula of $C_{87}H_{92}S_1$ and a MW of 1170 Da. Reprinted with permission from ref. 34. Copyright 2003 American Chemical Society.

Strausz et al.[31] championed the use of the ruthenium ion catalyzed oxidation (RICO) reaction on asphaltenes in an attempt to obtain information about the sizes of aromatic regions and alkyl side chains, about the naphthenic rings attached to aromatic fragments, and about the bridges between aromatic and naphthenic fragments. Along with the results from NMR and pyrolysis studies, they constructed a hypothetical asphaltene model based on Alberta asphaltenes, as shown in Figure 7. The model presented a paradigm change in the way asphaltenes are conceptualized, it consisted of smaller aromatic and naphthenic units linked by aliphatic bridges (instead of a single condensed aromatic system with a large number of rings). The model has an elemental formula of $C_{420}H_{496}N_6S_{14}O_4V$, an H/C atomic ratio of 1.18 and a molecular weight of 6191 Da. Further examples of these macro-archipelago models are shown in Figure 8, again inferred from information from analytical methods and elemental properties, with a combined molecular formula

of $C_{490}H_{528}N_4O_5S_{15}$, and a molecular weight of 7034 Da.[35] The type and distribution of aromatic main units, aliphatic chain structures, functionalities, and naphthenic structures were deduced from and based on RICO, pyrolysis gas chromatography (py/GC) and NMR data. Notably, the organic phase from the RICO reaction of the asphaltene contained sulfur and nitrogen heteroatoms, and because the X-ray photoelectron spectroscopy (XPS) study revealed that the sulfur structures were sulfide, thiophenic and sulfoxide, the existence of sulfides and nitrogen-containing saturated structures in asphaltene structures was proven. Other examples of large archipelago structures have been proposed by Sheremata et al.[36] (see Figure 17 and discussion in the next section) and those suggested by Ali, Ghaloum and Hauser[37] (Figure 9) who subjected two asphaltene samples from Kuwaiti residual oils to preparative gel-permeation chromatography (GPC), before analyzing them using NMR and XRD studies to investigate the relationship between the molecular size of the asphaltene fraction and its structural features.

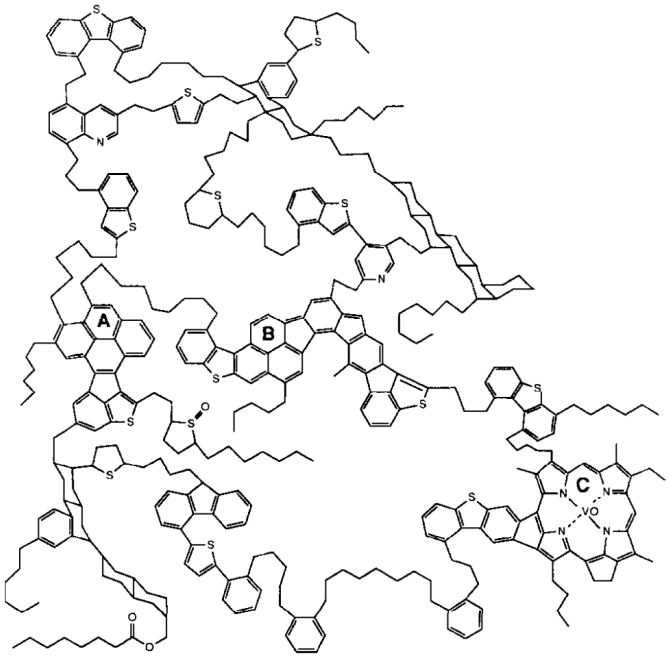


Figure 7. Hypothetical mega-archipelago asphaltene molecule structure, $C_{420}H_{496}N_6S_{14}O_4V$, proposed by Strausz et al. [31] Reprinted with permission from ref. 31. Copyright 1992 Elsevier.

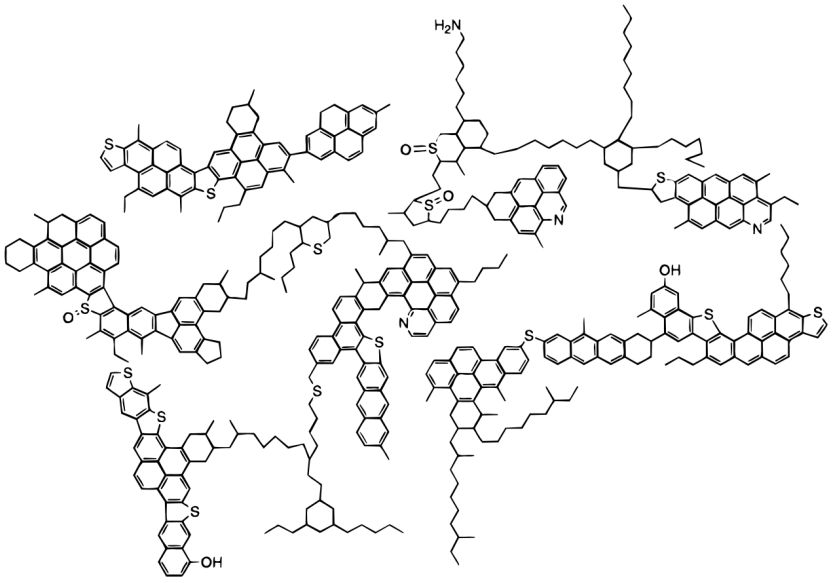


Figure 8. Model structures of asphaltene with a combined molecular formula of $C_{490}H_{528}N_4O_5S_{15}$ studied by Artok et al.[35] Note that there are four distinct archipelago models in the figure. Reprinted with permission from ref. 35. Copyright 1999 American Chemical Society.

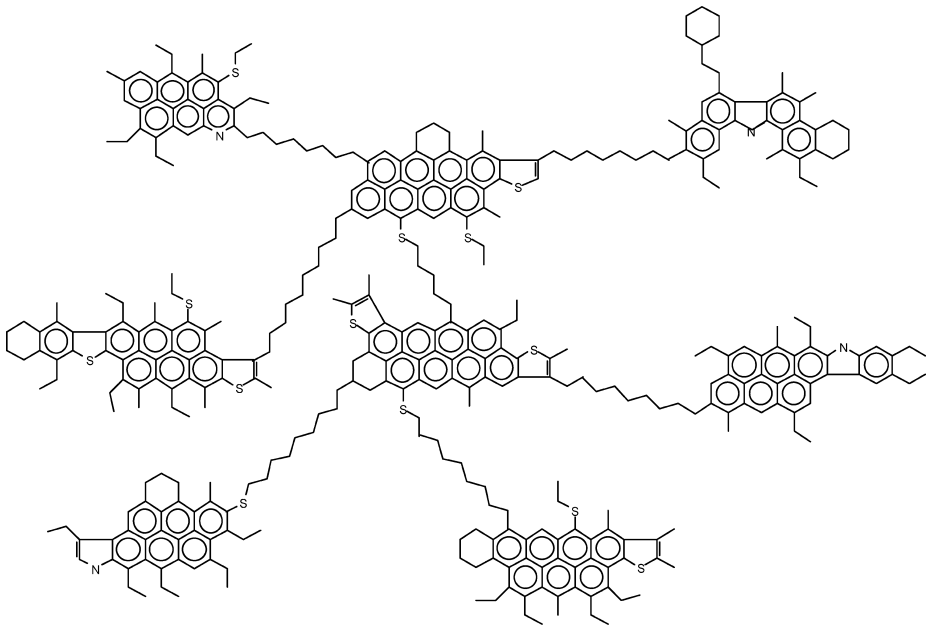


Figure 9. (*) Structure representation ($C_{367}H_{418}S_{12}N_3$) of asphaltene GPC Fractions derived from Kuwaiti residual oils.[37] The lower right-side fragment does not have a valid Kekulé structure. Reprinted with permission from ref. 37. Copyright 2006 American Chemical Society.

Recent models

At the start of the century, Mullins et al.[38] championed a series of asphaltene prototypes based heavily on results from fluorescence depolarization measurements arguing that mean molecular weights should be roughly 750 Da with a range of 500-1000 Da. Mullins et al. proposed that the typical asphaltene molecules were characterized by a single dominating polycyclic aromatic hydrocarbon (PAH) core consisting of 4-10 fused rings per asphaltene molecule. Four of the original examples are shown in Figure 10 with molecular formulas of $C_{51}H_{62}S$ (top left), $C_{66}H_{81}N$ (top right), $C_{56}H_{71}N$ (bottom left) and $C_{42}H_{54}O$ (bottom right) and molecular weights of 707 Da, 888 Da, 758 Da and 575 Da, respectively. The Mullins molecules (also referenced in the context of a proposal of an aggregation model called the Yen-Mullins model[39,40]) have had an overarching influence in the way the community depicted asphaltene molecules and many subsequent works have either employed the models directly[41,42,43] or have been inspired by them, sometimes correcting some of the irregularities in the original models. [44, 45]

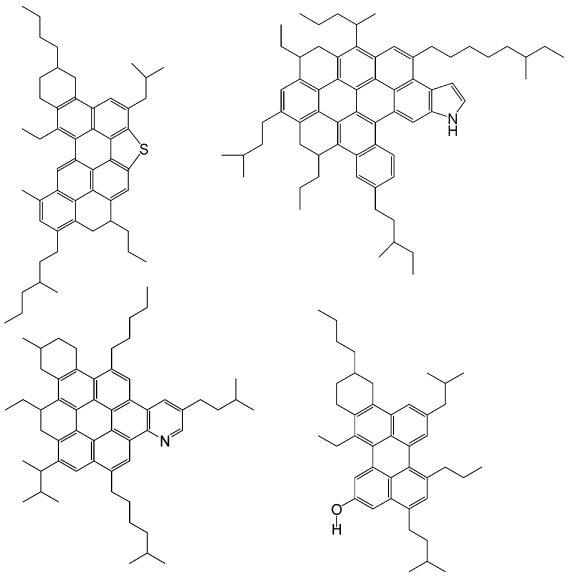


Figure 10. The asphaltene structures (modified Yen model) , $C_{51}H_{62}S$ (top left), $C_{66}H_{81}N$ (top right), $C_{56}H_{71}N$ (bottom left) and $C_{42}H_{54}O$ (bottom right), proposed by Mullins.[39] Reprinted with permission from ref. 39. Copyright 2010 American Chemical Society.

The "island" nature of the Mullins' models influenced, but did not dominate, the propositions and models for years to come. The asphaltene model shown in Figure 11, built based on samples characterized by elemental, molecular weight and NMR (^{1}H and ^{13}C) analyses to assist in the investigation of the toluene-insoluble and toluene-soluble asphaltene fractions, is an example[46].

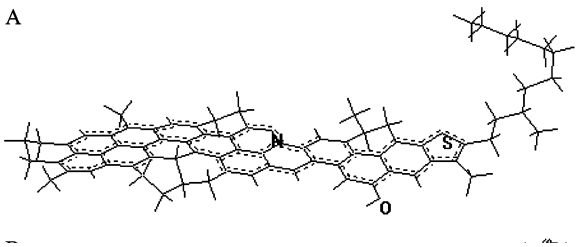


Figure 11. Model structure for an asphaltene of $C_{76}H_{73}NSO$ proposed by Acevedo et al.[46] Reprinted with permission from ref. 46 . Copyright 2004 American Chemical Society.

However, one encounters in the recent literature many other propositions, which somehow interpolate between the characteristics of the island and archipelago motifs. Figure 12 shows a further four examples of medium-sized asphaltenes where the island, continental and archipelago motifs are evidenced. These models were based on establishing the number of saturated and unsaturated rings using their molecular formulas and the percentage of aromatic carbon obtained mainly by analysis of NMR data, which were then used to adjust the structures and estimate the size and position of alkyl chains. It is worth reiterating that some of the models shown fail to be valid structures and/or fail to have the expected heteroatoms and/or polarity and have not stood the test of time.

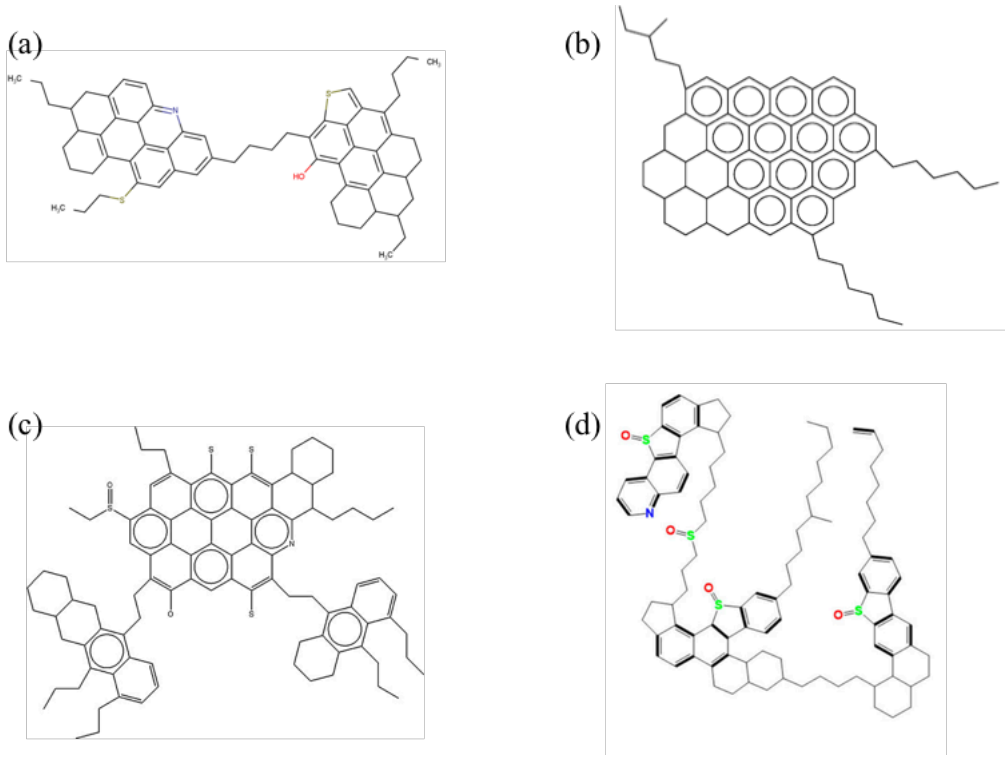


Figure 12. (a) Example of a model structure representing asphaltene samples from Khafji oil proposed by Takanohashi et al.[47] (b) (*†) Model structures for Asphaltene A, ($C_{69}H_{73}N_2O$), proposed by Carauta et al.[48] (c) Asphaltene models representing Kuh-e-Mond oils ($C_{95}H_{102}O_2NS_4$) proposed by Sayyad Amin et al.[49] (d) Molecular representations of interfacially active asphaltene (IAAs), ($C_{97}H_{117}NO_4S_4$) proposed by Yang et al.[50] Reprinted with permission from ref. 47. Copyright 2004 Taylor & Francis. Reprinted with permission from ref. 48. Copyright 2005 American Chemical Society. Reprinted with permission from ref. 49 Copyright 2011 Elsevier. Reprinted with permission from ref. 50. Copyright 2015 American Chemical Society.

There is now a much more modern view that has been also supported by extensive experimental evidence that there are asphaltene molecules with PAH cores with heteroatoms, alkyl side chains and some have alkyl bridges joining PAH cores. These bridges may also contain heteroatoms such as S and O. Molecular weight seems to be in the 1000 +/- 300 Da region. This restricted portrayal is insufficient to fully characterize the complexity present in this class of molecules. The number, size, shape, distribution and location of aromatic cores and aliphatic chains, as well as the presence of heteroatoms, all contribute delicately to the physical behavior of the molecule and affect its

aggregation, flocculation and deposition mechanisms.[51] Unsurprisingly, there is no universal agreement on the subject. There is experimental evidence[14] that two main structural motifs exist, called popularly archipelago-type and continental-type asphaltenes, hence from a modelling and/or theoretical perspective, it is useful to have a spectrum (or at the very least a pair) of models that span these two extremes. An example are the asphaltene models proposed by Ungerer et al.[52] with H/C ratios of 1.2 and molecular weights between 1300 Da and 1360 Da. The archipelago-type and continental-type asphaltenes have molecular formulas of $C_{93}H_{116}O_2S_2$ and $C_{94}H_{112}O_2S_2$, respectively. As illustrated in Figure 13, archipelago asphaltenes consist of small aromatic PAH domains connected by aliphatic chains while island asphaltenes feature a single large aromatic core, decorated by aliphatic chains connected to the edge of the core and are, by virtue of the sheer shape and relative size of their PAH cores, more prone to aggregation[52,53]. Another example of pairs of proposed models where an island is assessed side-by-side to an archipelago one of roughly the same size is given by Headen et al.[19] (c.f. Figure 19) and by Kuznicki et al[54, 55].

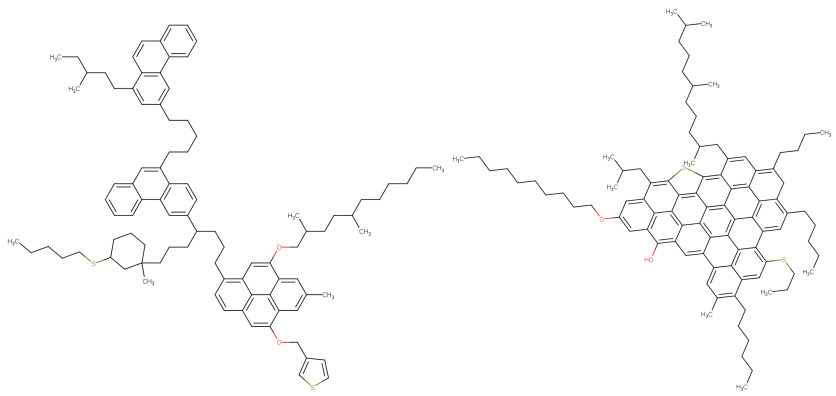


Figure 13. Pair of Archipelago-type asphaltene (left), $C_{93}H_{116}O_2S_2$, and continental-type asphaltene (right), $C_{94}H_{112}O_2S_2$, models studied by Ungerer et al. [52]

Resins

Resin structures are much less discussed in the literature. The fact that they have lower molecular weights make them inherently less prone to precipitating. From the scarce available data, it is known that the resins are highly polydisperse and may be classified into three broad groups: saturated, aromatics and mixed, depending on the largest fragment present in the molecules. The

aromatic resins are likely to interact with similar aromatic regions in the asphaltenes and form molecular aggregates with them. It makes sense, then, to select resins of the aromatic group as part of the molecular aggregates (clusters) to be formed. Two examples are provided in Figure 15, others are given in ref. 56.

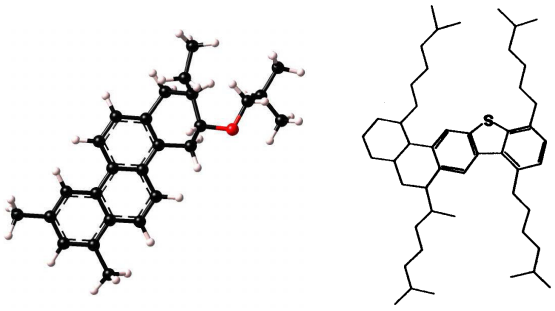


Figure 14. Model structures of resins, (left) Resin structure with a molecular formula of $C_{26}H_{32}O$ and MW of 360.5 Da proposed by Collell et al.[57] Reprinted with permission from ref. 57. Copyright 2014 American Chemical Society. (right) resin with molecular formula $C_{49}H_{70}S_1$, studied by Murgich et al.[33] Reprinted with permission from ref. 33. Copyright 1996 American Chemical Society.

Model chemical compounds

In the same way one can derive models for molecules from heuristic arguments in order to test their performance *in silico*, a similar exercise can be performed by synthesizing molecules (or by employing existing chemicals), to test the performance of asphaltene-like molecules under experimental conditions.[58,59,60] The reader is referred to a recent review by Sjöblom et al. [61] for a comprehensive account of the state of the art. Clearly, this approach is hampered by the difficulties associated with the practical aspects of organic chemical synthesis and is limited to structures based on the availability of chemically reactive building blocks. In some cases, these synthesized molecules have also been studied by computer simulation. The advantage here is that both the experiments and the simulations are probing the same molecular system and any disparity in results can be traced to deficiencies in either experimental or simulation protocols.

QUANTITATIVE MOLECULAR REPRESENTATION (QMR)

Previous Deployments

From the above discussion, it becomes apparent that the literature is riddled with models for asphaltene molecules with an extremely wide variability in terms of morphology and size. Even upon narrowing down parameter space in terms of molecular weight, chemical composition, degree of aromaticity, etc., most references present at the most a handful of plausible chemical descriptions of molecules. It is anticipated that the large diversity of naturally-occurring molecules will not be appropriately modelled with a small discrete set of representative models, which diminishes the (most likely to be key) effects of polydispersity and mixture behavior in computer simulation studies.

Crude oil is the result of an extraordinarily complex set of chemical reactions. Its generation starts with the decomposition of biopolymers into lighter molecular fragments that are later bonded back together in a rather random way to form a geopolymer (kerogen). Kerogen contains also smaller molecules that are trapped in the voids present in its structure. Under proper conditions of pressure and temperature (the oil window, 80 to 120 °C), the geopolymer breaks down into smaller molecular fragments that form, with the many trapped molecules, crude oil and gas. The low temperatures involved (<120 °C) in the transformations of the geopolymers are such that not all possible chemical reactions are allowed between the elements contained in the biopolymers. Then, not all possible combination of the atoms forming the original biopolymer will, then, be present in the crude oils. The number of different molecules is enormous in crude oils but not as astronomical as predicted by simple combinatorics. There is a general consensus between petroleum geochemists that all crude oils are formed by the same type of compounds. The main differences are in the amount of each of the compounds found in a particular crude. Figure 15 shows the ^{13}C - NMR spectra of the purified asphaltenes of two crudes of very different origins and contrasting characteristics. From the similarity found between the spectra, it is clear that the asphaltenes share significant commonalities and only subtle variations are found between those obtained from different crudes.

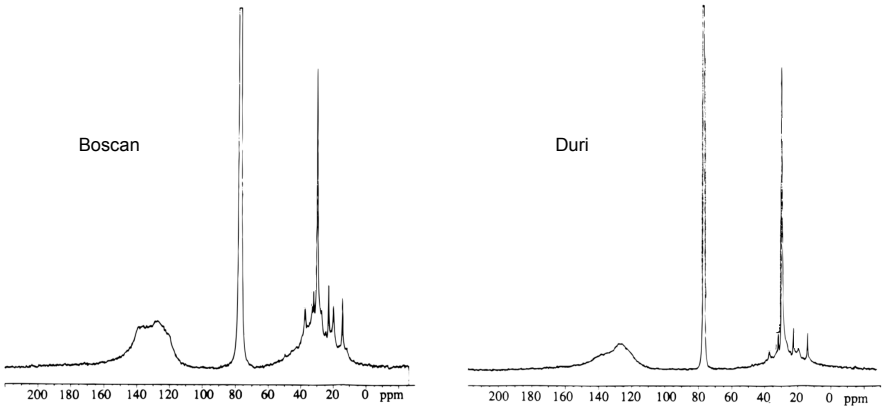


Figure 15. NMR spectra of asphaltenes extracted from Boscan (a heavy aromatic crude from Venezuela) and Duri (a light paraffinic crude from Indonesia) crudes. Adapted from ref. 62.

It is known that one can relate components of the saturated fraction with the aromatic part: the same structure of saturated polycyclics is found in molecules of the aromatic fraction. What follows from this argument is that if one selects some plausible building blocks, one can then form plausible molecules from a rearrangement of the chemical moieties, subject to a number of known experimentally accessible boundary conditions.

Most proposed asphaltene models in the literature were built upon heuristic arguments, and those that relied on experimental data still required decisions to be manually made on certain qualitative levels. It is in this space that this manuscript is positioned: we use a systematic approach, the quantitative molecular representation (QMR) methodology[36, 63, 64] to generate populations of molecular structures based on experimental data, in an attempt to introduce diversity and arguments that are more quantitatively measurable and less focused on one single average generalized structure.

Klein and co-workers[63,65] are the first to report the construction of asphaltene molecules using a Monte Carlo method by stochastically assembling 10 000 different molecules based on an off-shore Californian asphaltene sample, a Kern River heavy oil and sour import heavy gas oil. Important structural attributes including the number of aromatic rings, naphthenic rings and different lengths of aliphatic chains were adjusted according to their quantitative probability

density functions. Other reports from the group[66] described resin molecule attributes in terms of quantitative probability density functions, and optimized the properties of Monte Carlo-constructed molecules against experimental data from ^{1}H NMR spectroscopy, elemental analysis, and vapor-pressure osmometry (VPO) studies. Further to this they reported[67] a hybrid Monte Carlo-quadrature method, which analyzed molecules as collections of molecular attribute building blocks, each represented by a probability distribution function that was sampled via a Monte Carlo simulation yielding a large ensemble of representative molecules, designed such that the properties of the ensemble were restrained by experimentally measured analytical data. A quadrature method was then used to select an optimal small set (∼10–20) of molecules, with mole fractions optimized to match experimental data with key analytical properties of a complex petroleum feedstock. Recently, the method was further refined and applied to Turkish asphaltenes.[68] A pair of the generated molecules is shown in Figure 16.

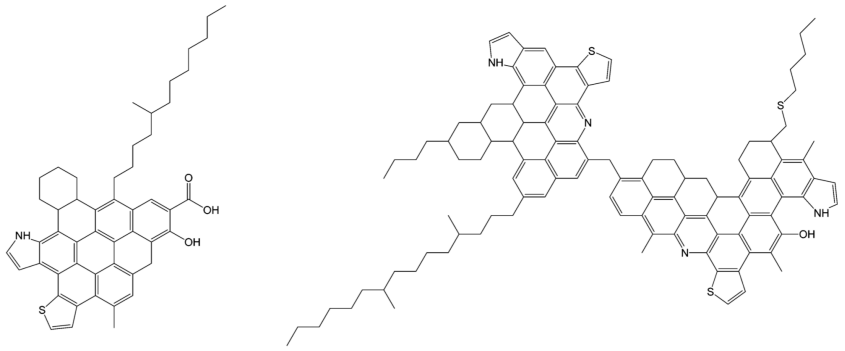


Figure 16. Island (left) and archipelago (right) models of asphaltenes generated by stochastic molecular reconstruction proposed by structure generated through stochastic assembly proposed by Denitz et al.[68]

Sheremata et al.[36] further developed the QMR generation model to incorporate structural data from ^{13}C NMR spectroscopy, a distribution of molecular building blocks following the archipelago-type structural framework, as well as both thioether and alkyl bridges. The authors reported a series of QMR-generated archipelago-type asphaltenes: $C_{283}H_{337}N_3O_4S_9$, $C_{230}H_{302}N_4O_2S_{10}$ and

$C_{318}H_{395}N_6O_6S_8V$, with molecular weights of 4133 Da, 3476 Da and 4705 Da, respectively, of which the final molecule is shown in Figure 17. The non-rigid asphaltene structures were constructed segment by segment via a Monte Carlo methodology, and nonlinearly optimized based on elemental and NMR spectroscopy (both ^{13}C and ^{1}H) analyses for Athabasca asphaltenes. The resulting structure is reminiscent of the mega-asphaltenes, most likely due to the imposition of the boundary condition of having a MW close to 4k Da.

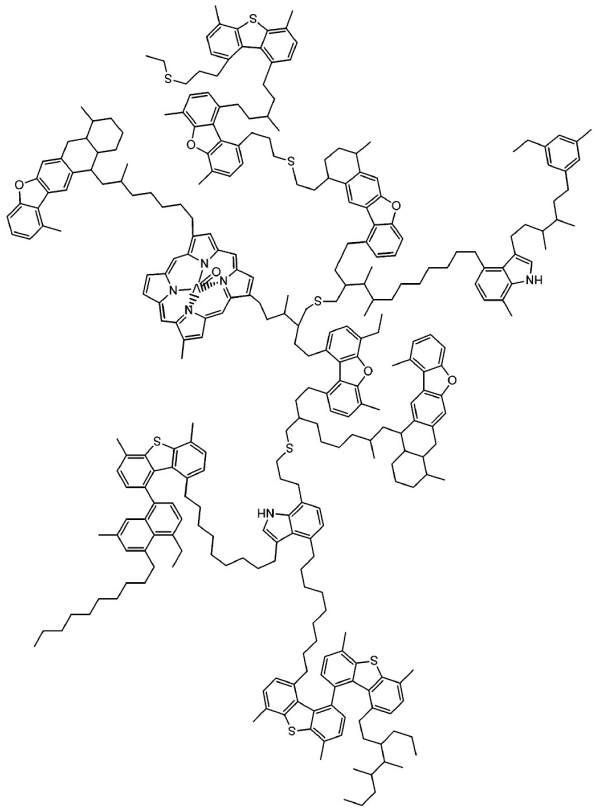


Figure 17. Archipelago-type asphaltene, $C_{318}H_{395}N_6O_6S_8V$, generated by Sheremata et al.[36] using a QMR algorithm. Reprinted with permission from ref. 36. Copyright 2004 American Chemical Society.

Figure 18 shows one of the asphaltene models generated by Frigerio and Molinari[69] using the Scienomics Materials Processes and Simulation (MAPS) platform.[70] Similar to the QMR approach, asphaltene molecules were virtually assembled from molecular building blocks which were selected from structures and functional groups that were detected in asphaltenes by

experimental characterization, such as chemolysis and pyrolysis. Seventeen molecules were chosen and published, with molecular weights ranging from 347 Da to 1625 Da, from a larger generated population by selecting the best fits to the chemical characterization data, which consisted of elemental analysis, GPC and $^{1}H - {}^{13}C$ NMR results. Rueda-Velázquez and Gray[71] also offer a proposition of building distributions of asphaltenes by Monte Carlo constructions attempting to map to molecular weight distributions, although no explicit structures were published.

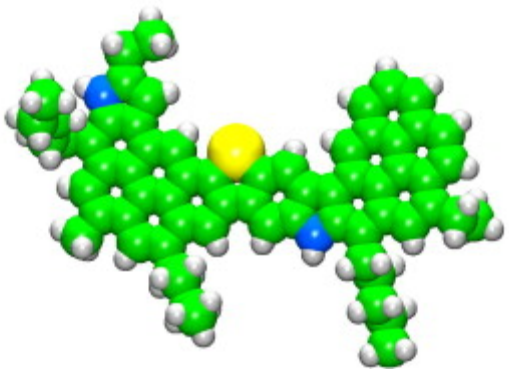


Figure 18. (*) A representative asphaltene model from of a population generated using the Scienomics Materials Processes and Simulation (MAPS) platform[70] reported by Frigerio and Molinari[69] Reprinted with permission from ref. [69]. Copyright 2011 Elsevier.

The procedures of Sheremata et al.[36] generated large archipelago-type structures by connecting building blocks only through alkyl chains. Boek et al. [64, 72, 73] extended the approach by also including other linkages, generating continental structures with larger aromatic cores. The set of asphaltene and resin structures (examples provided in Figure 19) were simulated in heptane and toluene. Asphaltene aggregation was reported in both heptane and toluene systems, but the aggregates persisted for longer in heptane than in toluene. On the other hand, the resin structures formed no aggregates in toluene, with some aggregation occurring in heptane. The Headen models[19] have been employed in extraordinarily long simulations (over 0.5 μs) revealing how aggregation is a "slow" process as compared to the solvent dynamics. These models and modified versions have been used in a number of simulation studies. [73,1974,75,76,77] So called "digital oils", including an *in silico* production of whole crudes [78, 79], of petroleum fractions[80], and vacuum

residue fractions[81] have also been produced in a similar fashion. A related procedure, labelled "structure-oriented lumping" [82] has been also extended to vacuum residue fractions.[83]

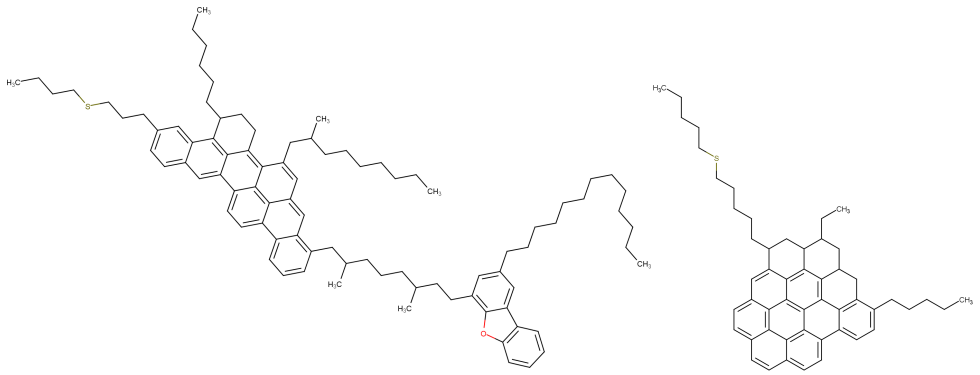


Figure 19. QMR-generated archipelago-type (left) "Asphaltene A" and continental-type (right) "Asphaltene C" proposed by Headen et al.[73] Reprinted with permission from ref. 73. Copyright 2009 American Chemical Society.

Methodology

The QMR approach developed and described in detail by Boek et al[64] is used to generate asphaltene and resin structures based on elemental analysis, ^{1}H – ^{13}C NMR data of an Athabasca bitumen (data set 'S' reported by Sheremata et al.[36]) and two new asphaltene samples (data sets 'C' and 'G'), as presented in the Supporting Information.

The reported asphaltenes have been precipitated and purified using different procedures. It is expected that the final composition of the resulting asphaltenes will be different. It is quite difficult to estimate the impact of these variations on the structure of the resulting average molecules. There is no work, to the best of our knowledge, that deals with the impact of these variations on the resulting structure. However, we have found that the C-13 NMR spectra of the asphaltene of the Boscan heavy crude differed only slightly for samples obtained in laboratories using different procedures. This suggest that the impact of the purification method, at least on this crude is relatively small. Nevertheless, further work is required to determine how much the precipitation and purification methods influence the resulting average structure of the asphaltenes.

The QMR procedure starts with the generation of a large population of 5000 to 10 000 potential asphaltene and resin molecules, each constructed by connecting smaller molecular segments following defined chemical rules. Two types of building blocks are used – aromatic sheets and aliphatic chains. The aromatic sheets were extended with naphthenic rings. The aliphatic chains were then attached to the aromatic cores, forming a single unit. The aliphatic chains consisted of between 1 to 50 carbons, and included n-alkanes, branched alkanes and chains that comprised heteroatoms. A non-linear optimization procedure (see Supporting Information) was used to select small subsets of molecules from the initial population that gave the best match with experimental data by considering the input and output parameters for each of 18 properties: molecular weight, elemental compositions from elemental analyses, and ^{1}H – ^{13}C NMR data. A target average MW of ~1000 Da for the optimized subsets was chosen in line with recent estimates.[84,85] For the generation of continental-type and archipelago-type asphaltenes, the numbers of aromatic unit sheets per molecule were limited to one and eight, with maximum numbers of 20 and 10 aromatic rings per unit sheet, respectively.

CATALOGUE OF ASPHALTENE AND RESIN MOLECULAR MODELS

A catalogue of 100 QMR-generated molecular models of asphaltenes and resins was produced, which is summarized in the Supporting Information. For each structure, the molecular formula, MW, H/C ratio, DBE, elemental weight percentages (wt%), the total number of aromatic rings and those in the largest aromatic core are reported. In the Supporting Information we include a 2D and 3D visualization for each molecule. The 2D structures were built and displayed using MarvinSketch (version 15.3.30.0).[86] The output files were then read by Avogadro (version 1.2.0),[87] an open-source molecular builder and visualization tool, which generated the 3D display structure. The catalogue is by no means comprehensive nor represents an ultimate collection of structures. Heavy metals, e.g. nickel and vanadium porphyrins, present only in ppm quantities and oxygenated and/or carboxylic moieties, which are relevant in adsorption on surfaces, all contribute to the collective behavior of asphaltenic systems, but were not considered in the ensemble. It is clearly inferred from experiments that these chemical groups will play an important role in the precipitation of asphaltenes but with the finite size of accessible atomistic simulations limited to thousands of molecules, it is still difficult to include them in appropriate and meaningful quantities.

Molecules A1 to A69 were generated with the continental-type structural configuration, of which A1 to A40 were based on data set 'S', and A41 to A69 were based on data set 'G'. Molecules A70 to A100 were generated with the archipelago-type structural configuration based on data set 'C'. Aromatics, resins, and asphaltenes form a continuum[88,89] of increasing molar mass, polarity, and heteroatom content, the catalogue also presents a distribution of molecules, where molecules that could either be resins or asphaltenes are present. In a recent study,[90] Arabian heavy oil (ARH) samples were analyzed by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) based on saturates, aromatics, resins, and asphaltenes (SARA) fractionations. The average carbon number and double bond equivalent (DBE) of the resin fraction were reported to be 37.9 and 15.9, respectively. For the asphaltene fraction, they were reported to be 44.7 and 19.9 respectively. To reflect these results, molecules with 7 or less aromatic rings in total (models A90 to A101) have been categorized as resins, such that the maximum carbon number, DBE and MW reach 36, 16 and 523 Da respectively. The classification is by no means definitive.

Distribution histograms of the MWs, carbon numbers, H/C ratios, DBEs and largest aromatic core sizes of the QMR-generated molecules are shown in Figure 20. The MWs of the QMR-generated island-type asphaltene structures average at 872 Da and range from 572 Da to 1147 Da. For the MWs of the archipelago-type asphaltene and resin structures, the averages are at 946 Da and 375 Da, and range from 748 Da to 1273 Da and 268 Da to 523 Da, respectively. The MW distribution is shown to peak around the 1000 Da region. The average DBEs of the continental-type asphaltenes, archipelago-type asphaltenes and resins are 35, 29 and 12 respectively. A mapping of DBE vs. carbon numbers of the QMR-generated structures along with a few of the proposed structures reviewed in the earlier section 'A Brief Historical Perspective' are presented in figure 21. These representations have been used recently by Rodgers et al. to present experimental data from direct mass spectral detection which help quantify the general structure of the asphaltenes.[91] In this representation, for example, the continental molecules are characterized by generally higher DBE values. Regions for the island-type and archipelago-type structures were roughly drafted based on observation of the conformations in the catalogue. Asphaltene and resin regions suggested by Cho et al.[90] are shown to agree well with the catalogue molecules. Also in the plot, we mark the upper and lower boundaries of compositional space as suggested by Podgorski et al. [92] In all of these boundaries, the molecules suggested in the catalogue seem to

straddle the general experimental observations and are similar to other representations in the literature.

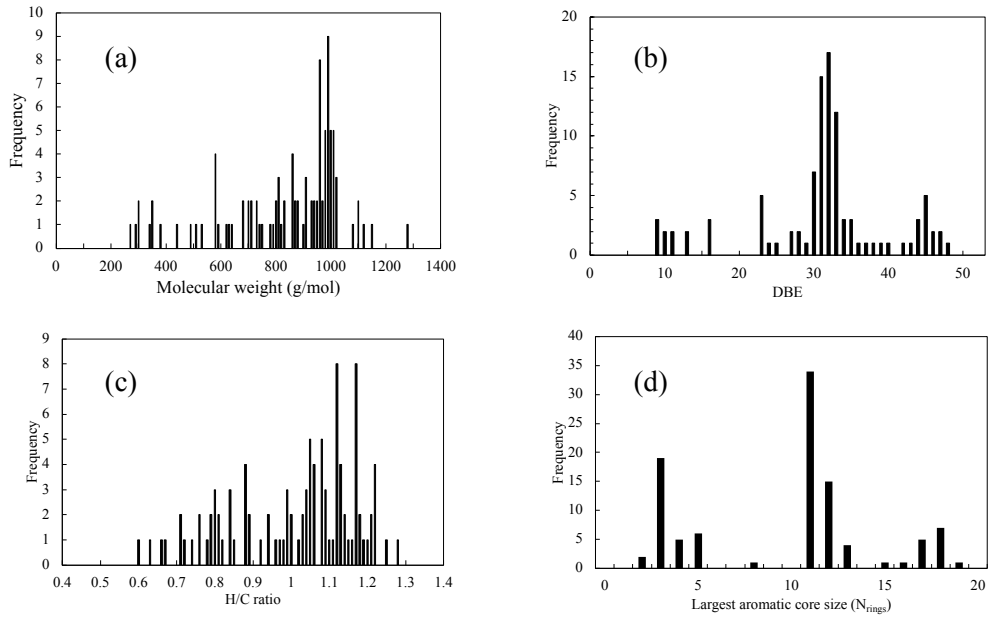


Figure 20. Distribution histograms of the (a) MW, (b) DBE, (c) H/C ratios, (d) and largest aromatic core sizes of the QMR-generated molecules.

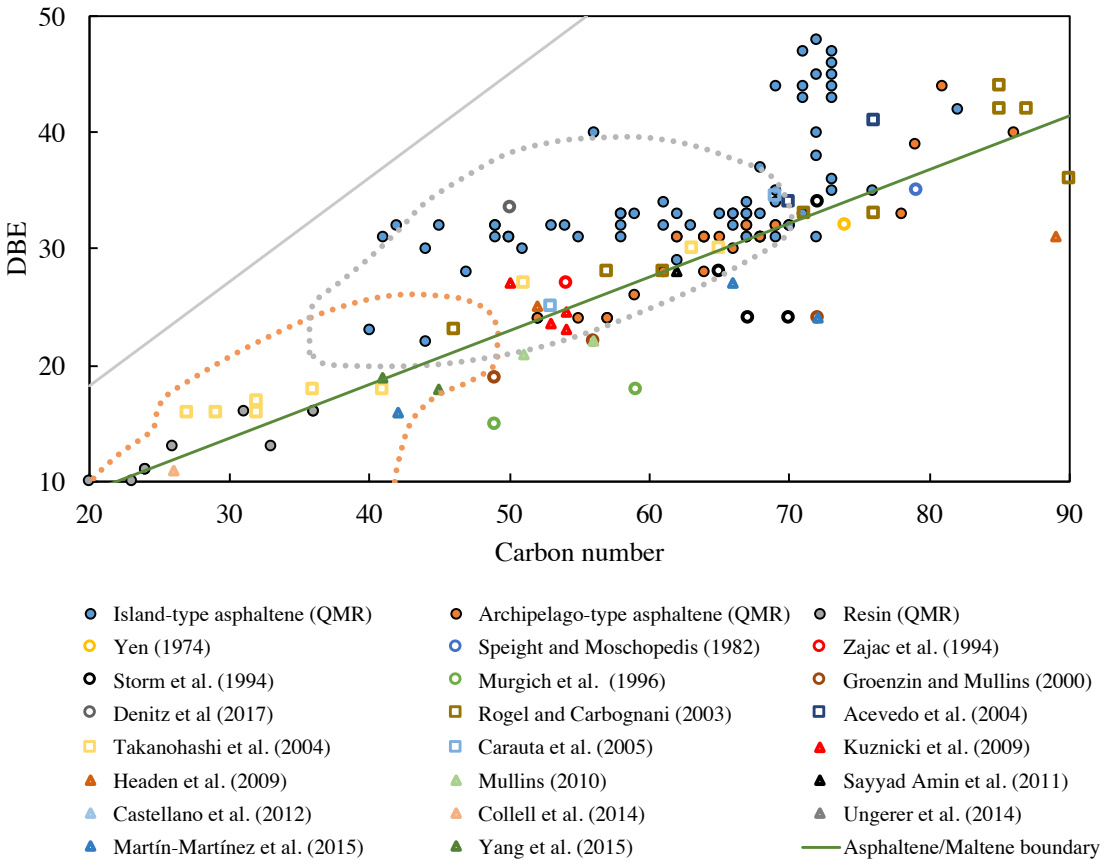


Figure 21. Double bond equivalents (DBE) as a function of carbon numbers for some of the molecules described in the text and the catalogue molecules (QMR). Dotted regions correspond to the experimental asphaltene (gray) and resin (orange) regions reported by Cho et al.[Error! Bookmark not defined.] Grey line is the theoretical upper bound for DBE, green line is the asphaltene/maltene boundary as proposed by Podgorski et al. [92]

SIMULATION OF ASPHALTENE AGGREGATION THROUGH MOLECULAR DYNAMICS

The physical behavior of asphaltenes depend on an exquisite balance between the intermolecular forces of all the molecules present in the system and these in turn depend on not only the molecular structural properties, but also the number and nature of the elements (particularly heteroatoms)

present in the systems. While the catalogue has a large variety of molecular models, we have chosen, without prejudice to the other members of the ensemble, two representative examples of similar molecular weight but very different morphology to present as an example: molecule A18 (Figure 22) an island-type asphaltene model featuring a large aromatic core at the center, and molecule A81 (Figure 23), an archipelago-type asphaltene model featuring smaller aromatic islands connected and extended by chains. Both molecules are almost isomeric, with similar chemical structure, MW (968 and 984 Da respectively) and solubility parameter (18.8 $MPa^{0.5}$ and 18.7 $MPa^{0.5}$ respectively).

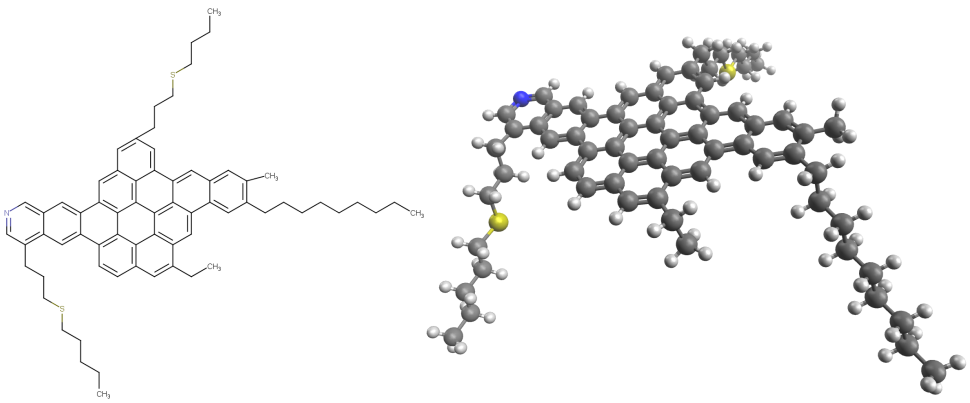


Figure 22. Molecule A18, $C_{68}H_{73}NS_2$, an island-type asphaltene structure, in 2D (left) and 3D (right).

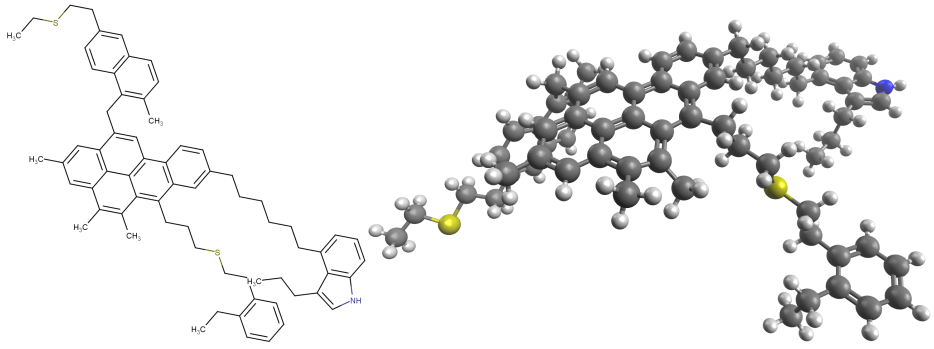


Figure 23. Molecule A81, $C_{69}H_{77}NS_2$, an archipelago -type asphaltene structure, in 2D (left) and 3D (right)

Classical atomistic MD simulations of the asphaltene models A18 and A81, separately in toluene or heptane, were carried out utilizing intra- and intermolecular potentials of the OPLS-AA force-field[93] to describe interatomic forces, which had been shown to work well for aromatic liquids in reproducing experimental data.[94] 27 asphaltene molecules and the required solvent molecules to provide for a 7 wt% asphaltene concentration were initially placed in random positions and equilibrated before running the system in the isobaric−isothermal (NPT) ensemble. Temperatures and pressures were maintained at 298 K and 1 bar using the Nosé−Hoover and Parinello−Rahman algorithms, respectively.[95] All simulations were carried out using the GROMACS MD simulation code (version 4.6).[96] The leapfrog MD algorithm and the Verlet scheme was used with cubic simulation cells with periodic boundary conditions. Bond lengths were regulated using the LINCS algorithm.[97] The time step is set at 1 fs and the cutoff of non-bonded interactions is set at 1 nm, while long-range electrostatic interactions are accounted for by the particle-mesh Ewald (PME) implementation.[95] The simulations were run for at least 120 million time steps, or 120 ns, and the system sizes were ~10 nm in each dimension. The GROMACS utility *g_clustsize* was used to calculate the average (excluding monomers) and maximum asphaltene cluster sizes and numbers of clusters over time from the simulation trajectory file. Molecules that were less than 3.5 Å from each other were considered to be clustered (although different criteria may lead to slightly different profiles).[53]

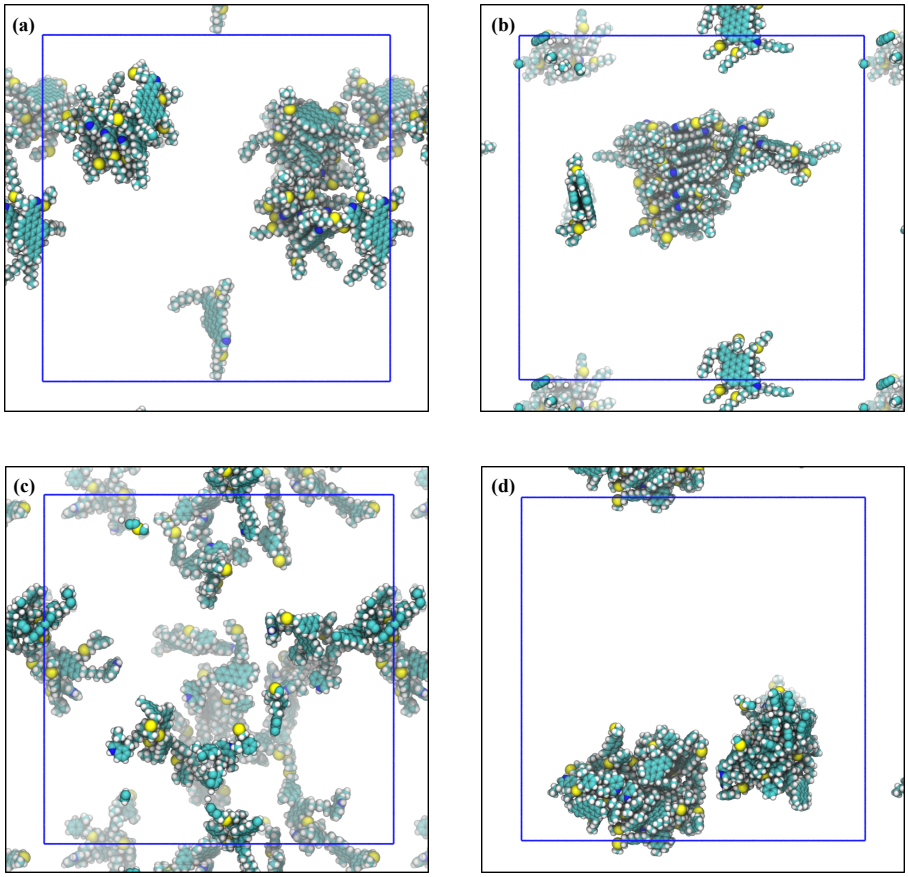


Figure 24. Snapshots of equilibrium configurations at 1 bar and 298 K of a 7 wt% asphaltene in pure solvent after 120 ns of simulations. (a) Molecule A18 (island) in toluene (b) Molecule A18 (island) in heptane (c) Molecule A81 (archipelago) in toluene (d) Molecule A81 (archipelago) in heptane. Cyan, white, blue and yellow spheres represent the carbon, hydrogen, nitrogen and sulfur atomic centers, respectively. Solvent molecules are not shown for clarity. The simulation cell boundaries are shown in blue.

As expected, asphaltene aggregation levels are shown to be higher in heptane than in toluene for both asphaltene models. For the simulations in heptane, the maximum cluster sizes approached the total numbers of asphaltenes in the system, 27, suggesting that finite-size effects were in place and the system was essentially phase-separating into an asphaltene rich phase and a supernatant solvent. For the simulations in toluene, however, the development of smaller aggregates demonstrated the balance between the energetically favorable cohesion and the entropically unfavorable aggregation.

In spite of the similarities, the morphology of the asphaltenes plays a crucial role[13]: there is a marked difference between the aggregation behavior of the archipelago-type asphaltene when compared to the more compact island-type. Figure 24 shows snapshots of typical equilibrated configurations.[98] In toluene, the more compact (island) A18 molecules shows some appreciable aggregation (Figure 24a). The system size is small and it is not possible to conclude whether the system is fully phase separated or not. For this particular asphaltene, a concentration of 7% might be above its solubility limit. The A18 molecules shows a noticeable aggregation on heptane (Figure 24b). The stacking of the PAH cores is apparent in the middle of the sub-figure. Given a longer simulation time, it is plausible that the system undergoes a complete phase separation[19]. The twin molecule A81, on the other hand exhibits a very different behavior, it is seen to be very soluble in toluene (Figure 23c) , with only a small number of dimers, trimers, etc. present. In heptane, A81 phase separates clearly. Again, the presence of more than one cluster in the cell is most likely an artifact of the (relatively) small simulation times rather than the presence of a recognizable stable "nanoggregate".[53] In the Supporting Information we provide for time-resolved graphs which track the average and maximum size clusters seen in the simulations along with the total number of cluster seen in the simulation cells. The classification of asphaltenes into "good" and "bad" or equivalently "soluble" and "insoluble"[99] seems to be intricately linked to the archipelago or continental nature of the model, although clearly many other factors apart from the morphology of the molecules are presumably behind these differences e.g. hydrogen bonds, polarity, acid-base interactions, presence of nickel and vanadium porphyrins, etc.

CONCLUSIONS

A catalogue of 100 QMR-generated continental-type asphaltene, archipelago-type asphaltene and resin molecular structures is reported, with molecular weights ranging from 268 Da to 1273 Da. The catalogue is by no means comprehensive or unique and in some sense, it might be argued that the effort is futile given that there are tens of thousands of plausible elemental compositions for asphaltenes.[100] The aggregation mechanisms that classify asphaltenes as "good" or "bad" crucially depend on the molecular structural properties and in this sense the catalogue may be used as a starting point for the study of more complex asphaltene mixture behavior. The collection presented in the catalogue, is of course very dependent on the, arguably arbitrary, selection of

molecular weight distributions, maximum and minimum number of rings present in the molecules and the number and type of heteroatoms.

Atomistic molecular dynamics simulations of ~100 000 atoms and 120 ns were carried out to showcase the physical aggregation behavior of two examples from the catalogue; a continental-type asphaltene A18 and archipelago-type asphaltene A81 in explicit solvents of either heptane or toluene, under ambient conditions. At this level of detail, for a small sample of 27 identical asphaltene molecules and reasonably high concentrations (hence less solvent molecules) the system sizes were ~100 Å in each dimension. Providing that the radius of gyration of asphaltenes (or asphaltene clusters) in toluene is ~50-100 Å, as seen by small-angle scattering,[101] the system sizes in this study of ~100 Å in each dimension inevitably suffer from finite-size system effects. A recent direct comparison between MD simulation and scattering, for more concentrated solutions, shows that simulation boxes of a similar size fail to replicate the observed SANS scattering.[76] As such, in order to fully observe asphaltene aggregation mechanisms, understand the effect of polydispersion and to attempt to monitor the kinetics of cluster formations, amongst others, simulations with length and time scales of several orders of magnitude larger are necessary. However, these simulations shown herein are already large by the standards of current computational power, and represent the practical upper length and time scale limits for atomistic simulations. One way forward is to coarse-grain (CG) the intermolecular potentials[102,103,53,104] in such a way that groups of atoms are represented by larger but fewer beads, increasing the particle sizes and dramatically reducing the number of inter-particle calculations, hence enabling longer time steps and giving access to length and time scales of higher orders.

Evidently, asphaltenes and resins are very polydisperse at the molecular level, much in the same way all snowflakes in a snow storm are unique. Following the analogy, in spite of the impossibility of obtaining "the" asphaltene molecule, one could expect to find some repeating patterns and characteristics which would allow an ensemble of prototypical or surrogate molecules to be used as appropriate models for the discrete molecular modelling of these families of fluids. While current computational power casts a limitation on the practical maximum number of molecules in a simulation, the remarkable advances in computational technologies suggest that in a matter of

time we will have the possibility of modelling fully polydisperse mixtures of heavy crudes and other complex fluids. Our catalogue of plausible molecules provides a self-consistent starting point, which along with other similar collections provide the starting elements of a whole new age of *in silico* studies.

ASSOCIATED CONTENT

Supporting Information.

The Supporting Information is available free of charge on the ACS Publications website at DOI: XX.XXXX/pubs.acs.org and includes links to simulation files, clustering analysis of models A18 and A81 in toluene and heptane, discussions on molecular morphologies which have been ruled out and the raw NMR data.

A catalogue of 100 QMR-generated asphaltene and resin molecular structures is provided as part of the supporting information.

AUTHOR INFORMATION

Corresponding author

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

ORCID

Erich A. Müller: 0000-0002-1513-6686

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been partially funded by EPSRC through research grants to the Molecular Systems Engineering group (Grant Nos. EP/E016340, EP/J014958 and EP/R013152). J.C.L. has

been partially supported by Shell Global Solutions International B.V. through project PT45895. Contributing discussions with Dr. Majeed Shaik (Shell) and Prof. George Jackson (Imperial College London) are gratefully acknowledged. Part of the NMR data employed in this project was provided by Shell Global Solutions International. Computations were performed using the computational resources provided by the UK Materials and Molecular Modelling Hub, which is partially funded by EPSRC (EP/P020194/1) and by the Imperial College Research Computing Service, DOI: 10.14469/hpc/2232. Rendering of simulation snapshots was made using VMD[105].

References

[1] Strausz, O. P.; Morales-Izquierdo, A.; Kazmi, N.; Montgomery, D. S.; Payzant, J. D.; Safarik, I.; Murgich, J. Chemical Composition of Athabasca Bitumen: The Saturate Fraction. *Energy Fuels* **2010**, 24 (9), 5053-5072.

[2] Strausz, O. P.; Lown, E. M.; Morales-Izquierdo, A.; Kazmi, N.; Montgomery, D. S.; Payzant, J. D.; Murgich, J. Chemical Composition of Athabasca Bitumen: The Distillable Aromatic Fraction. *Energy Fuels* **2011**, 25 (10), 4552-4579.

[3] A very illustrative dissertation on the history of asphaltenes and the problems associated with their definition as a solubility class is provided in M. E. Moir "Asphaltenes, What Art Thou?", Chapter 1 of *The Boduszynski Continuum: Contributions to the understanding of the molecular composition of Petroleum"* C. Ovalles and M.E. Moir Eds. ACS Symposium series 1282, ACS, Washington (2018)

[4] Papadimitriou, N. I.; Romanos, G. E.; Charalambopoulou, G. C.; Kainourgiakis, M. E.; Katsaros, F. K.; Stubos, A. K. Experimental investigation of asphaltene deposition mechanism during oil flow in core samples. *J. Pet. Sci. Eng.* **2007**, 57 (3−4), 281−293.

[5] Badre S.; Goncalves C.C.; Norinaga K., Gustavson G., Mullins O.C. Molecular size and weight of asphaltene and asphaltene solubility fractions from coals, crude oils and bitumen. *Fuel* **2006**, 85, 1–11 and the letter to the editor by Herod, A. A.; Kandiyoti R.; Bartle K.D. *Fuel* **2006**, 85, 1950–1951 and the subsequent rebuttal by Mullins O.C. *Fuel* **2007**, 86, 309–312.

[6] Behrouzi, M.; Luckham, P. F. Limitations of Size-Exclusion Chromatography in Analyzing Petroleum Asphaltenes: A Proof by Atomic Force Microscopy. *Energy Fuels* **2008**, 22, 1792−1798 and the comment by Herod, A. A.; Kandiyoti, R. *Energy Fuels* **2008**, 22, 4307−4309.

[7] Mullins, O. C.; Martínez-Haya, B.; Marshall, A. G. Contrasting Perspective on Asphaltene Molecular Weight. This Comment vs the Overview of A. A. Herod, K. D. Bartle, and R. Kandiyoti. *Energy Fuels* **2008**, 22, 1765−1773.

[8] Coletti, F.; Crittenden, B. D.; Haslam, A. J.; Hewitt, G. F.; Jackson, G.; Jimenez-Serratos, G.; Macchietto, S.; Matar, O. K.; Müller, E. A.; Sileri, D.; Yang, J. Chapter 5 - *Modelling of Fouling from Molecular to Plant Scale. In: Crude Oil Fouling: Deposit Characterization, Measurements, and Modeling*; Coletti, F., Hewitt, G. F., Eds.; Gulf Professional Publishing: Boston, **2014**; pp 179−320.

[9] Hoepfner, M. P.; Fogler, H. S. Multiscale Scattering Investigations of Asphaltene Cluster Breakup, Nanoaggregate Dissociation, and Molecular Ordering. *Langmuir* **2013**, 29 (49), 15423–15432.
[10] Schuler, B.; Meyer, G.; Peña, D.; Mullins, O. C.; Gross, L. Unraveling the Molecular Structures of Asphaltenes by Atomic Force Microscopy. *J. Am. Chem. Soc.* **2015**, 137 (31), 9870–9876.
[11] B. Schuler et al. Heavy Oil Based Mixtures of Different Origins and Treatments Studied by Atomic Force Microscopy. *Energy Fuels* **2017**, 31 (7), 6856–6861.
[12] It is important to note the limitations behind the impressive feat of visualizing a molecule with AFM. Clearly there are significant discrepancies between the structures determined by AFM and a long list of experimental results; e.g. a) A mild pyrolysis of the observed molecules of asphaltenes will not generate the well-known range of products found for this process. b) The number of methyl and methylene groups in these asphaltene molecules cannot explain the observed 1-H and 13-C NMR spectra, c) These molecules cannot explain the latest results obtained via FT-ICR Mass spectroscopy, where the coexistence of both archipelago and much smaller island asphaltenes has been confirmed. Until these discrepancies are fully explained, we cannot assume that they do represent the majority of the asphaltene molecules.
[13] Silva, H. S.; Alfarra, A.; Vallverdu, G.; Bégué, D.; Bouyssiere, B.; Baraille, I. Asphaltene aggregation studied by molecular dynamics simulations: role of the molecular architecture and solvents on the supramolecular or colloidal behavior, *Pet. Sci.* **2019** https://doi.org/10.1007/s12182-019-0321-y
[14] Chacoń-Patiño, M. L.; Rowland, S. M.; Rodgers, R. P. Advances in Asphaltene Petroleomics. Part 1: Asphaltenes Are Composed of Abundant Island and Archipelago Structural Motifs. *Energy Fuels* **2017**, 31 (12), 13509−13518.
[15] Chacoń-Patiño, M. L.; Rowland, S. M.; Rodgers, R. P. Advances in Asphaltene Petroleomics. Part 2: Selective Separation Method That Reveals Fractions Enriched in Island and Archipelago Structural Motifs by Mass Spectrometry. *Energy Fuels* **2018**, 32 (1), 314−328.
[16] Chacoń-Patiño, M. L.; Rowland, S. M.; Rodgers, R. P. Advances in Asphaltene Petroleomics. Part 3. Dominance of Island or Archipelago Structural Motif Is Sample Dependent. *Energy Fuels* **2018**, 32, 9106.
[17] Rueda-Velásquez, R. I.; Gray, M. R. Monte Carlo Simulation of Asphaltenes and Products from Thermal Cracking. *Energy Fuels* **2014**, 28, 2352−2364.
[18] Li, D. D.; Greenfield, M. L. High Internal Energies of Proposed Asphaltene Structures. *Energy Fuels* **2011**, 25, 3698−3705.
[19] Headen, T. F.; Boek, E. S.; Jackson, G.; Totton, T. S.; Muller, E. A. Simulation of Asphaltene Aggregation through Molecular Dynamics: Insights and Limitations. *Energy Fuels*, **2017**, 31(2), 1108–1125.
[20] Clar, E. *The Aromatic Sextet*. Wiley, New York (1972).
[21] Solà, M. Forty years of Clar's aromatic π-sextet rule. *Front. Chem*. **2013**, 1, 22.
[22] Yen, T. F. Structure of Petroleum Asphaltene and Its Significance. *Energy Sources* **1974**, 1 (4), 447−463.
[23] Speight, J. G.; Moschopedis, S. E. On the Molecular Nature of Petroleum Asphaltenes . In " Chemistry of asphaltenes" ed. Bunger, J.W. and Li, N.C. Chapter 1 *Adv. Chem. Ser.* **1982**, 195, 1−15.
[24] Zajac, G. W.; Sethi, N. K.; Joseph, J. T. Molecular imaging of petroleum asphaltenes by scanning-tunneling-microscopy-verification of structure from C-13 and proton nuclear-magnetic-resonance data. *Scanning Microsc.* **1994**, 8, 463−470.

[25] Pacheco-Sánchez, J. H.; Álvarez-Ramírez, F.; Martínez-Magadán, J. M. Morphology of Aggregated Asphaltene Structural Models. *Energy Fuels* **2004**, 18(6), 1676–1686.
[26] Storm, D. A.; Edwards, J. C.; DeCanio, S. J.; Sheu, E. Y. Molecular representations of Ratawi and Alaska north slope asphaltenes based on liquid-and solid-state NMR. *Energy Fuels* **1994**, 8, 561-566.
[27] Andersen, S. I.; Speight, J. G. Thermodynamic models for asphaltene solubility and precipitation. *J. Petroleum Sci. Eng.* **1999**, 22(1), 53–66.
[28] Ferris, S. W.; Black, E. P.; Clelland, J. B. Aromatic Structure in Asphalt Fractions. *Ind. Eng. Chem. Prod. Res. Dev.* **1967**, 6, 127–132.
[29] Miller, J. T.; Fisher, R. B.; Thiyagarajan, P.; Winans, R. E.; Hunt, J. E. Subfractionation and Characterization of Mayan Asphaltene. *Energy Fuels* **1998**, 12, 1290–1298
[30] Boduszynski, M. M. Subfractionation and characterization of Mayan asphaltene. *Energy Fuels* **1988**, 2, 597–613
[31] Storm, D. A.; DeCanio, S. J.; DeTar, M. M.; Nero, V. P. Upper bound on number average molecular weight of asphaltenes. *Fuel* **1990**, 69, 735–738
[32] Strausz, O. P.; Mojelsky, T. W.; Lown, E. M. The molecular structure of asphaltene: an unfolding story. *Fuel* **1992**, 71, 1355–1363.
[33] Murgich, J.; Rodríguez, M. J.; Aray, Y. Molecular recognition and molecular mechanics of micelles of some model asphaltenes and resins. *Energy Fuels* **1996**, 10, 68–76.
[34] Rogel, E.; Carbognani, L. Density estimation of asphaltenes using molecular dynamics simulations. *Energy Fuels* **2003**, 17, 378–386.
[35] Artok, L.; Su, Y.; Hirose, Y.; Hosokawa, M.; Murata, S.; Nomura, M. Structure and reactivity of petroleum-derived asphaltene. *Energy Fuels* **1999**, 13, 287–296.
[36] Sheremata, J. M.; Gray, M. R.; Dettman, H. D.; McCaffrey, W. C. Quantitative molecular representation and sequential optimization of Athabasca asphaltenes. *Energy Fuels* **2004**, 18, 1377−1384.
[37] Ali, F. A.; Ghaloum, N.; Hauser, A. Structure representation of asphaltene GPC fractions derived from Kuwaiti residual oils. *Energy Fuels* **2006**, 20, 231-238.
[38] Groenzin, H.; Mullins, O. C. Molecular size and structure of asphaltenes from various sources. *Energy Fuels* **2000**, 14, 677–684.
[39] Mullins, O. C. The modified Yen model. *Energy Fuels* **2010**, 24, 2179−2207.
[40] O. C. Mullins, et al. Advances in Asphaltene Science and the Yen–Mullins Model. *Energy Fuels* **2012**, 26 (7), 3986–4003.
[41] Zhang, L. Q.; Greenfield, M. L. Relaxation time, diffusion, and viscosity analysis of model asphalt systems using molecular simulation. *J. Chem. Phys.* **2007**, 127, 194502.
[42] Li, D. D.; Greenfield, M. L. Viscosity, relaxation time, and dynamics within a model asphalt of larger molecules. *J. Chem. Phys.* **2014**, 140, 034507.
[43] Tirjoo, A.; Bayati, B.; Rezaei, H.; Rahmati, M. Molecular dynamics simulation of the effect of ions in water on the asphaltene aggregation. *J. Molecular Liq.* **2019**, 277, 40–48.
[44] Martín-Martínez, F. J.; Fini, E. H.; Buehler, M. Molecular asphaltene models based on Clar sextet theory. *RSC Adv.* **2015**, 5, 753−759.
[45] De Leon, J.; Velásquez, A. M.; Hoyos, B. A. A stochastic method for asphaltene structure formulation from experimental data: avoidance of implausible structures. *Phys. Chem. Chem. Phys.* **2017**, 19, 9934-9944.

[46] Acevedo, S.; Escobar, O; Echevarría, L.; Gutierrez, L. B.; Méndez, B. Structural analysis of soluble and insoluble fractions of asphaltenes isolated using the PNP method. Relation between asphaltene structure and solubility. *Energy Fuels* **2004**, 18, 305–311.
[47] Takanohashi, T.; Sato, S.; Tanaka, R. Structural relaxation behaviors of three different asphaltenes using MD calculations. *Pet. Sci. Technol.* **2004**, 22, 901–914.
[48] Carauta, A. N. M.; Seidl, P. R.; Chrisman, E. C. A. N.; Correia, J. C. G.; de O.Menechini, P.; Silva, D. M.; Leal, K. Z.; deMenezes, S. M. C.; de Souza, W. F.; Teixeira, M. A. G. Modeling solvent effects on asphaltene dimers. *Energy Fuels* **2005**, 19, 1245–1251.
[49] Sayyad Amin, J.; Nikooee, E.; Ghatee, M. H.; Ayatollahi, Sh.; Alamdari, A.; Sedghamiz, T. Investigating the effect of different asphaltene structures on surface topography and wettability alteration. *Appl. Surf. Sci.* **2011**, 257, 8341−8349.
[50] Yang, F.; Tchoukov, P.; Dettman, H.; Teklebrhan, R. B.; Liu, L.; Dabros, T.; Czarnecki, J.; Masliyah, J.; Xu, Z. Asphaltene subfractions responsible for stabilizing water-in-crude oil emulsions. Part 2: Molecular representations and molecular dynamics simulations. *Energy Fuels* **2015**, 29, 4783−4794.
[51] Murgich, J. Molecular simulation and the aggregation of the heavy fractions in crude oils. *Mol. Simul.* **2003**, 29, 451–461.
[52] 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.
[53] Jiménez-Serratos, G.; Totton, T. S.; Jackson, G.; Müller, E. A. Aggregation Behavior of Model Asphaltenes Revealed from Large-Scale Coarse-Grained Molecular Simulations. *J. Phys. Chem. B* **2019**, 123(10), 2380–2396.
[54] Kuznicki, T.; Masliyah, J. H.; Bhattacharjee, S. Molecular dynamics study of model molecules resembling asphaltene-like structures in aqueous organic solvent systems. *Energy Fuels* **2008**, 22, 2379–2389
[55] Kuznicki, T.; Masliyah, J. H.; Bhattacharjee, S. Aggregation and partitioning of model asphaltenes at toluene− water interfaces: Molecular dynamics simulations. *Energy Fuels* **2009**, 23, 5027–5035.
[56] Castellano, O.; Gimon, R.; Canelon, C.; Aray, Y.; Soscun, H. Molecular interactions between Orinoco belt resins. *Energy Fuels* **2012**, 26 (5), 2711–2720.
[57] Collell, J.; Ungerer, P.; Galliero, G.; Yiannourakou, M.; Montel, F.; Pujol, M. Molecular simulation of bulk organic matter in type II shales in the middle of the oil formation window. *Energy Fuels* **2014**, 28(12), 7457-7466.
[58] Rakotondradany, F.; Fenniri, H.; Rahimi, P.; Gawrys, K. L.;Kilpatrick, P. K.; Gray, M. R. Hexabenzocoronene model compounds for asphaltene fractions: synthesis & characterization. *Energy Fuels* **2006**, 20, 2439–2447.
[59] Tan, X.; Fenniri, H.; Gray, M. R. Pyrene derivatives of 2, 2'-bipyridine as models for asphaltenes: synthesis, characterization, and supramolecular organization. *Energy Fuels* **2008**, 22, 715–720.
[60] He, L.; Wang, P.; He, L.; Qu, Z.; Luo, J.; Peng, B.; Tang, X.; Pei, Y. Molecular dynamics simulations of the self-organization of side-chain decorated polyaromatic conjugation molecules: phase separated lamellar and columnar structures and dispersion behaviors in toluene solvent. *RSC Adv.* **2018**, 8, 11134-11144.
[61] Sjöblom, J.; Simon, S.; Xu, Z. Model molecules mimicking asphaltenes. *Adv. Colloid Interface Sci.* **2015**, 218, 1−48.

[62] Strausz, O. P.; Mojelsky, T.W.; Lown, E. M.; Kowalewski, I. ; Behar, F. Structural Features of Boscan and Duri Asphaltenes, *Energy Fuels*, **1999**, 13(2), 228–247.
[63] Neurock, M.; Nigam, A.; Trauth, D.; Klein, M. T. Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms. *Chem. Eng. Sci.* **1994**, 49, 4153−4177.
[64] Boek, E. S.; Yakovlev, D. S.; Headen, T. F. Quantitative molecular representation of asphaltenes and molecular dynamics simulation of their aggregation. *Energy Fuels* **2009**, 23, 1209-1219.
[65] Neurock, M.; Libanati, C.; Nigam, A.; Klein, M. T. Monte Carlo simulation of complex reaction systems: Molecular structure and reactivity in modelling heavy oils. *Chem. Eng. Sci.* **1990**, 45, 2083-2088.
[66] Trauth, D. M.; Stark, S. M.; Petti, T. F.; Neurock, M.; Klein, M. T. Representation of the molecular structure of petroleum resid through characterization and Monte Carlo modeling. *Energy Fuels* **1994**, 8, 576-580.
[67] Campbell, D. M.; Klein, M. T. Construction of a molecular representation of a complex feedstock by Monte Carlo and quadrature methods. *Appl. Catal. A* **1997**, 160, 41-54.
[68] Deniz, C. U.; Yasar, M.; Klein, M. T. A New Extended Structural Parameter Set for Stochastic Molecular Reconstruction: Application to Asphaltenes. *Energy Fuels* **2017**, 31 (8), 7919–7931.
[69] Frigerio, F.; Molinari, D. A multiscale approach to the simulation of asphaltenes. *Comput. Theor. Chem.* **2011**, 975, 76−82.
[70] Scienomics, Paris, France. http://www.scienomics.com.
[71] Rueda-Velásquez, R. I.; Gray, M. R. Monte Carlo simulation of asphaltenes and products from thermal cracking. *Energy Fuels* **2014**, 28, 2352−2364.
[72] Al Halwachi, H. K.; Yakovlev; D. S.; Boek, E. S. Systematic optimization of asphaltene molecular structure and molecular weight using the quantitative molecular representation approach. *Energy Fuels* **2012**, 26, 6177-6185.
[73] Headen, T. F.; Boek, E. S.; Skipper, N. T. Evidence for asphaltene nanoaggregation in toluene and heptane from molecular dynamics simulations. *Energy Fuels* **2009**, 23, 1220−1229.
[74] Sedghi, M.; Goual, L.; Welch, W.; Kubelka, J. *J. Phys. Chem. B* **2013**, 117, 5765.
[75] Mikami, Y.; Liang, Y.; Matsuoka, T.; Boek, E.S. *Energy Fuels* **2013**, 27, 1838.
[76] Headen, T.F.; Hoepfner, M.P. *Energy Fuels* **2019**, 33, 3787-3795.
[77] Mehana, M., Fahes, M., Huang L. Asphaltene Aggregation in Oil and Gas Mixtures: Insights from Molecular Simulation *Energy Fuels* **2019**, 33(6) 4721-4730.
[78] Iwase, M.; Sugiyama, S.; Liang, Y.; Masuda, Y.; Morimoto, M.; Matsuoka, T.; Boek, E. S.; Ueda, R.; Nakagawa, K. Development of Digital Oil for Heavy Crude Oil: Molecular Model and Molecular Dynamics Simulations. *Energy Fuels* **2018**, 32 (3) 2781-2792.
[79] Sugiyama, S.; Liang, Y.; Murata, S.; Matsuoka, T.; Morimoto, M.; Ohata, T.; Nakano, M.; Boek, E. S. Construction, validation, and application of digital oil: Investigation of asphaltene association toward asphaltene-precipitation prediction. *SPE J.* **2018**, 23 (3), 952-968.
[80] Lyu, W.; Zhang, L.; Li, K.; Wang, G.; Shi, Q.; Zhao, S.; Xu, C. Average Molecule Construction of Petroleum Fractions Based on 1H-NMR. *AIChE J.* **2018**, 65 (1) 270–280.
[81] Verstraete, J. J.; Schnongs, P.; Dulot, H.; Hudebine, D. Molecular reconstruction of heavy petroleum residue fractions. *Chem. Eng. Sci.* **2010**, 65(1), 304–312.
[82] Quann, R. J.; Jaffe, S. B. Structure-oriented lumping: describing the chemistry of complex hydrocarbon mixtures. *Ind. Eng. Chem. Res.* **1992**, 31(11), 2483–2497.
[83] Jaffe, S. B.; Freund, H.; Olmstead, W. N. Extension of Structure-Oriented Lumping to Vacuum Residua. *Ind. Eng. Chem. Res.* **2005**, 44(26), 9840–9852.

[84] 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.
[85] McKenna, A. M.; Donald, L. J.; Fitzsimmons, J. E.; Juyal, P.; Spicer, V.; Standing, K. G.; Marshall, A. G.; Rodgers, R. P. Heavy Petroleum Composition. 3. Asphaltene Aggregation. *Energy Fuels* **2013**, 27(3) 1246–1256.
[86] MarvinSketch (version 15.3.30.0), calculation module developed by ChemAxon, https://chemaxon.com/products/marvin
[87] Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchinson, G. R. J. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Cheminf.* **2012**, 4, 17.
[88] Alboudwarej, H.; Beck, J.; Svrcek, W. Y.; Yarranton, H. W. Sensitivity of asphaltene properties to separation techniques. *Energy Fuels* **2002**, 16 (2), 462–469.
[89] Shi, Q.; Hou, D.; Chung, K. H.; Xu, C.; Zhao, S.; Zhang, Y. Characterization of Heteroatom Compounds in a Crude Oil and Its Saturates, Aromatics, Resins, and Asphaltenes (SARA) and Non-basic Nitrogen Fractions Analyzed by Negative-Ion Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2010**, 24, 2545–2553.
[90] Cho, Y. J.; Na, J.-G.; Nho, N.-S.; Kim, S. H.; Kim, S. Application of Saturates, Aromatics, Resins, and Asphaltenes Crude Oil Fractionation for Detailed Chemical Characterization of Heavy Crude Oils by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Equipped with Atmospheric Pressure Photoionization. *Energy Fuels* **2012**, 26 (5), 2558−2565.
[91] McKenna, A. M.; Purcell, J. M.; Rodgers, R. P.; Marshall, A. G. Heavy Petroleum Composition. 1. Exhaustive Compositional Analysis of Athabasca Bitumen HVGO Distillates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Definitive Test of the Boduszynski Model. *Energy Fuels* **2010**, 24 (5), 2929−2938.
[92] Podgorski, D. C.; Corilo, Y. E.; Nyadong, L.; Lobodin, V. V.; Bythell, B. J.; Robbins, W. K.; McKenna, A. M.; Marshall, A. G.; Rodgers, R. P. Heavy petroleum composition. 5. Compositional and structural continuum of petroleum revealed. *Energy Fuels* **2013**, 27 (3), 1268−1276.
[93] Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, 118, 11225−11236.
[94] Fu, C.-F.; Tian, S. X. A comparative study for molecular dynamics simulations of liquid benzene. *J. Chem. Theory Comput*. **2011**, 7, 2240−2252.
[95] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*. 2nd Ed. Oxford (2017).
[96] Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, 29, 845−54.
[97] Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, 4, 116−122.
[98] We assume at that after 120 ns the systems have equilibrated. This might not be true for system in heptane that phase separate, where extremely long simulations are required.
[99] Vilas Bôas Fávero, C.; Maqbool, T.; Hoepfner, M.; Haji-Akbari, N.; Fogler, H. S. Revisiting the flocculation kinetics of destabilized asphaltenes. *Adv. Colloid Interface Sci.* **2017**, 244, 267−280.
[100] McKenna, A. M.; Marshall, A. G.; Rodgers, R. P. Heavy Petroleum Composition .4. Asphaltene Compositional Space. *Energy Fuels* **2013**, 27 (3), 1257−1267.

[101] Gawrys, K. L.; Blankenship, G. A.; Kilpatrick, P. K. Solvent entrainment in and flocculation of asphaltenic aggregates probed by small-angle neutron scattering. *Langmuir* **2006**, 22, 4487−4497.
[102] B. 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.
[103] Herdes, C.; Totton, T. S.; Müller, E. A. Coarse grained force field for the molecular simulation of natural gases and condensates. *Fluid Phase Equilib*. **2015**, 406, 91−100.
[104] Kaimaki, D. M.; Haire, B. T.; Ryan, H. P.; Jiménez-Serratos, G.; Alloway, R. M.; Little, M.; Morrison, J.; Salama, I. E.; Tillotson, M. J.; Smith, B. E.; Moorhouse, S. J.; Totton, T. S. ; Hodges, M.; Yeates, S. G.; Quayle, P.; Clarke, S. M.; Müller, E. A.; Durkan, C. Multiscale Approach Linking Self-Aggregation and Surface Interactions of Synthesized Foulants to Fouling Mitigation Strategies. *Energy Fuels*, **2019**, 33, 7216–7224.
[105] Humphrey, W., Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, 14, 33-38.