

Introducing a force-matched united atom force field to explore larger spatiotemporal domains in molecular dynamics simulations of bitumen

Assaf, Eli I.; Liu, Xueyan; Lin, Peng; Erkens, Sandra

DOI

[10.1016/j.matdes.2024.112831](https://doi.org/10.1016/j.matdes.2024.112831)

Publication date

2024

Document Version

Final published version

Published in

Materials and Design

Citation (APA)

Assaf, E. I., Liu, X., Lin, P., & Erkens, S. (2024). Introducing a force-matched united atom force field to explore larger spatiotemporal domains in molecular dynamics simulations of bitumen. *Materials and Design*, 240, Article 112831. <https://doi.org/10.1016/j.matdes.2024.112831>

Important note

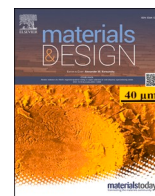
To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Introducing a force-matched united atom force field to explore larger spatiotemporal domains in molecular dynamics simulations of bitumen

Eli I. Assaf^{a,*}, Xueyan Liu^a, Peng Lin^{a,b}, Sandra Erkens^{a,b}

^a Delft University of Technology, Delft, The Netherlands

^b Ministry of Infrastructure and Water Management (Rijkswaterstaat), The Netherlands

ARTICLE INFO

Keywords:

Coarse-graining
Force field
Force-matching
Molecular dynamics
Morphology
Bitumen
Phase separation
Segregation

ABSTRACT

This paper presents a United Atom (UA) force field for simulating hydrocarbon molecules in bituminous materials, integrating explicit hydrogens into beads with their parent atom. This method simplifies all-atom molecular models, significantly accelerating Molecular Dynamics (MD) simulations of bitumen by 10 to 100 times. Key advantages include halving the particle count, eliminating complex hydrogen interactions, and decreasing the degrees of freedom of the molecules. Developed by mapping forces from an all-atom model to the centers of mass of UA model beads, the force field ensures accurate replication of energies, forces, and molecular conformations, mirroring properties like pressure and density. It features 17 bead types and 287 interaction types, encompassing various hydrocarbon molecules. The UA force field's stability, surpassing all-atom models, is a notable achievement. This stability, stemming from smoother potential energy surfaces, leads to consistent property measurements and improved stress tensor accuracy. It enables the extension of MD simulations to larger spatiotemporal scales, crucial for understanding complex phenomena such as phase separation in bituminous materials. This foundational work sets the stage for future developments, including refining parameters and introducing new bead types, to enhance the modeling capabilities of the force field, thereby advancing the application and understanding of bituminous materials.

1. Introduction

Atomistic simulations play a crucial role in understanding the properties of materials at the atomic level, which is essential for designing new materials. These simulations, however, are computationally demanding, limiting their application to systems comprised of a few hundred molecules at most, hardly surpassing nanoscale domains. This constraint becomes particularly problematic when studying bituminous materials, which exhibit complex intermolecular interactions and morphologies that span well into the microscale. To fully capture the mechanisms that dictate the material's mechanical and rheological responses, it is necessary to analyze bitumen across various scales.

Coarse Graining (CG) techniques, which reduce the complexity of atomistic models by representing groups of atoms with simplified particles or "beads," offer a way to improve computational efficiency. These techniques are gaining traction in the study of bituminous materials as they allow for the exploration of larger scales, especially the microscale. Such scale-up makes it possible to fundamentally observe the evolution of important phenomena like phase separation, crystallization, and the

development of complex intermolecular features, crucial in establishing a more comprehensive model of bitumen [1]. Despite the existence of various CG approaches, there is a notable absence of methods tailored specifically for capturing the unique behaviors of bituminous materials. The inherent limitation of CG force fields lies in their simplification, making them most effective for systems analogous to the reference systems utilized in their development.

Force-matched United Atom (UA) CG models offer an attractive approach for simulating hydrocarbon mixtures by grouping hydrogen atoms with their respective parent atoms. Designed to replicate the forces present in all-atom simulations, these models manage to retain a significant portion of atomic detail within the CG models [2]. This strategy is viewed as a prudent progression in the CG modeling of bituminous materials, facilitating intuitive interpretation of results and comparisons with all-atom models [3]. Nonetheless, current UA force fields encounter difficulties in accurately capturing the characteristics of bituminous molecules, particularly those with large molecular masses, planar and aromatic structures, and a propensity for dense molecular stacking influenced by electron delocalization effects.

* Corresponding author.

E-mail address: e.i.assaf@tudelft.nl (E.I. Assaf).

<https://doi.org/10.1016/j.matdes.2024.112831>

Received 18 December 2023; Received in revised form 27 February 2024; Accepted 8 March 2024

Available online 11 March 2024

0264-1275/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

This manuscript aims to advance Molecular Dynamics (MD) simulations of bitumens and similar complex hydrocarbon mixtures by extending the temporal and spatial scales of these simulations from the nanoscale to the microscale, to capture critical phenomena affecting their mechanical and rheological responses. It introduces a UA-CG force field specifically designed for high molecular weight, planar resinous, and asphaltenic molecules in bitumens. This force field, comprised of 17 bead types and over 287 tabulated potentials, enables modeling of most molecules in bitumen, including polycyclic aromatics and non-aromatics, and phenolic, pyrrolic, pyridinic, thiophenic, and sulfoxidic functional groups, while increasing computational performance by approximately 100 times. It seeks to accurately reflect molecular conformations and thermodynamic properties, such as pressure and density, and to tighten the gap of all-atom studies with those studying bitumen at Engineering scales.

The paper begins with an overview of previous research on bitumen, emphasizing both experimental and computational strategies for its characterization and the application of all-atom and CG MD methods. The methodology section details the development and validation of the new UA-CG force field, including molecule selection, parameterization, bead definition, mathematical modeling, and a thorough list of validation and benchmarking procedures.

In the results and discussion, the paper evaluates the new force field's performance in modeling bitumen, focusing on its applicability to various hydrocarbon molecules, computational efficiency, and its ability to provide insights into phase behavior and the rearrangement of bitumen molecules to form different morphologies. It also establishes benchmarking criteria to aid in the comparison to all-atom models and the UA force field developed by Weiner et al. [4]. The conclusion suggests potential uses for the force field, anticipated benefits, and areas for further refinement, underscoring the impact of this development on the modeling of bituminous materials and similarly complex hydrocarbon mixtures, like in the case of rejuvenators, additives, and polymers.

2. Literature review

Bitumen, a crucial asphalt component, is a viscoelastic substance from crude oil distillation, comprising complex high molecular weight hydrocarbons [5]. Its global production surpasses 100 million tons annually, underlining its industrial importance. Bitumen's chemical composition varies with the crude oil source and refining methods, leading to diverse properties. Its organic molecules differ in aromaticity, saturation, and polarity, exhibiting phase behavior and unique intermolecular and microstructural features that heavily impact its mechanical and rheological responses. Characterizing bitumen is challenging due to the diverse scalability of these features [6].

Chemical characterization involves separating bitumen molecules into fractions based on solubility. Initially, bitumen was categorized into asphaltenes and maltenes. However, this was inadequate due to maltenes' molecular complexity. Corbett [7] refined this by separating maltenes into saturates, aromatics, and resins using elution-adsorption liquid chromatography. This established the SARA (Saturates, Aromatics, Resins, and Asphaltenes) fractions as a comprehensive method for chemical elucidation of bituminous samples [8,9].

Bitumen's chemical structure's understanding evolved with studies on how these fractions interact to form a seemingly homogeneous mixture. Initial models suggested a colloidal dispersion with asphaltenes suspended in maltenes [10,11]. It was proposed that asphaltenes rearrange to form agglomerates several micrometers in size, affecting the behavior of bitumens. Pfeiffer and Saal introduced sol and gel bitumen concepts, differentiated by their composition and microstructural arrangements. Most bitumen samples have characteristics of both sol and gel microstructures [12].

Despite advances in understanding bitumen's microstructure, debates persist about the nature of its nanostructural and microstructural features. Techniques like X-ray diffraction and Atomic Force Microscopy

have revealed distinct microstructures, including *bee* and *worm* formations, which correspond to the agglomerates of similar SARA fractions [13]. However, their experimental interpretation remains unvalidated.

Experimental methods in nanostructural analysis of bituminous materials face limitations, leading to the use of Quantum Mechanics (QM) and MD simulations for detailed modeling. These simulations, crucial for predicting large-scale properties from atomic and molecular interactions, are constrained by computational resources and the need for detailed molecular structure data.

Works by Hansen [3], Qu [14], and Greenfield [15] have been vital in developing bitumen models for simulations that consider all SARA components. Nonetheless, these models often overlook potential phase behavior and molecular rearrangement to form different morphologies as seen in Fig. 1. Efforts continue to create more accurate bitumen models, focusing on integrating larger and longer simulations for better nanostructural and microstructural understanding.

Recent research on the spatial distribution of SARA fractions in standard MD bitumen models shows promising, albeit preliminary results. Scientists have expanded model sizes and simulation times to nearly microscale spatiotemporal domains. These extended simulations, although computationally intractable, reveal that SARA fractions rearrange [16,17], with asphaltenes forming aggregates surrounded by resins, while saturates and aromatics fill other spaces, indicating phase separation and more heterogeneous morphologies being more stable than homogenous counterparts.

To bypass long MD simulations for heterogeneous molecular formations, scientists have resorted to Monte Carlo methods to recreate intermolecular features observed experimentally. This technique, starting with pre-heterogenized models, allows for shorter simulations and has been key in developing stable bitumen models with inherent heterogeneity. These models demonstrate that intermolecular feature formation leads to energetically favorable structures, proving that highly heterogeneous morphologies are expected to occur naturally.

The challenge in manually creating highly heterogeneous structures involves predicting morphology and arranging molecules strategically [18]. While the Monte Carlo method has offered insights into energetically favorable morphologies, it raises questions about the dynamics and kinetic feasibility of these formations in real bituminous materials; there is growing interest in allowing system dynamics to naturally reach stable morphological configurations. This would require scientists to move from Monte Carlo back to MD, requiring state-of-the-art methods to allow these simulations to become computationally feasible.

To address the computational challenges in all-atom MD simulations, researchers have increasingly turned to CG techniques [19–22]. These aim to simplify molecular systems by representing multiple atoms as singular, larger entities known as beads. This approach facilitates the simulation of extensive numbers of molecules across broader spatiotemporal scales. Although CG techniques abstract away from detailed atomic interactions, they have proven instrumental in investigating phenomena such as phase separation and the emergence of microstructural features in bituminous materials—tasks that are unfeasible with all-atom models due to their high computational requirements [23].

Additionally, force-matching CG methods are grounded in Classical Newtonian principles, reflecting the foundational dynamics characteristic of all-atom MD simulations. This grounding facilitates a more straightforward comparison with traditional MD methods, in contrast to alternatives like Dissipative Particle Dynamics (DPD) or other CG methods. Although DPD simulations also rely on Newton's second law of motion, they incorporate additional forces—dissipative and random forces—beyond those typically accounted for in all-atom MD simulations. These forces introduce complexities in accurately capturing the microscale phenomena, making DPD distinct and somewhat less directly comparable to the detailed molecular interactions described by all-atom methods [24]. Consequently, employing force-matched CG force fields emerges as a more logical selection over methods necessitating extensive

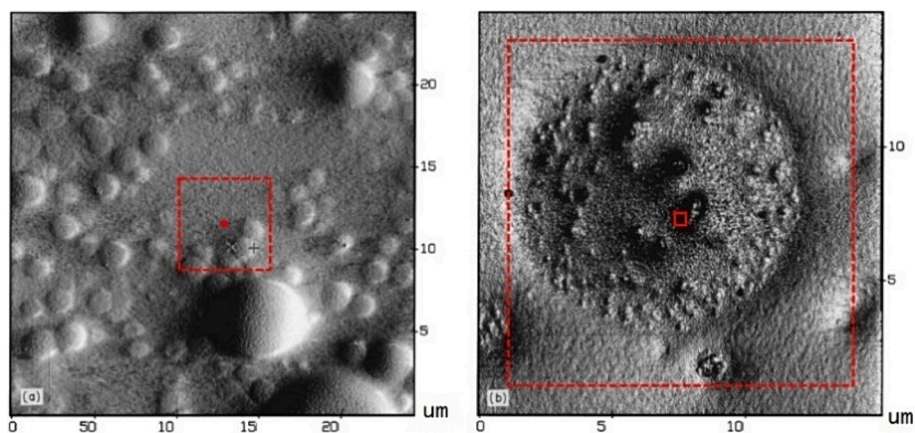


Fig. 1. SARA fractions within a bitumen sample as seen under an Atomic Force Microscope (AFM). This image captures a gel form of bitumen, highlighting the presence of round asphaltene agglomerates that reach micrometer sizes. The smaller red square indicates the scope of all-atom MD simulations, whereas the larger red square shows the scope attainable by UA-CG simulations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

comparability techniques.

The strategy for mapping atoms to beads in CG models varies, with the level of detail being a discretionary choice made by the force field designers. The objective is to find an optimal balance between the required level of detail and the anticipated computational efficiency. Strategies range from UA models, which incorporate hydrogen atoms into their parent atoms while retaining significant detail, to more simplified CG models. An example of the latter is the Martini force field [25], which groups four atoms into beads for the construction of organic molecules, offering a more moderate level of detail. At the other end of the spectrum are ultra-CG models, which represent entire molecules or groups of molecules as single beads, sacrificing detail for the ability to study micro- and millimeter scale phenomena such as those in biological systems or fluid dynamics. The application of force-matching UA techniques is particularly pertinent in the study of bituminous materials, where accurately capturing thermodynamic and kinetic properties is crucial for understanding the materials' mechanical and rheological responses. The cautious application of UA force fields in this domain is essential due to the limited prior application of CG techniques to bituminous material research and the critical importance of maintaining geometric, thermodynamic, and kinetic fidelity.

Existing UA force fields enable the simulation of diverse organic mixtures in MD studies [4,26,27]. For instance, the work by Weiner and Jorgensen [28] facilitates the representation of hydrogens bonded to parent atoms, providing various bead types for constructing a wide array of organic molecules such as aliphatic chains, aromatics, phenols, thiophenes, and oxanes, all prevalent in bitumen. These UA force fields, often derived from their all-atom force field counterparts, permit the creation of CG models for hydrocarbon mixtures with relative ease. However, the application of these force fields to bituminous systems faces challenges. These challenges arise mainly due to the lower transferability of CG models [2] and the large size of molecules modeling bitumen (e.g., asphaltenes with more than 100 carbon atoms and a mass of 2000 g/mol), their highly aromatic and planar structures, and the presence of strong electron delocalization effects, which result in unique conformations of functionalized groups [29]. This issue is compounded in such force fields, where the majority of parameterizations and mappings from all-atom to UA-CG models are based on reference molecules that typically do not exceed 12 carbon atoms and 200 g/mol in molecular mass. Additionally, the mathematical representations of two-body, three-body, and four-body interactions in these UA force fields often assume a single equilibrium position [30], suggesting a preference for one molecular conformation. This simplification may not realistically capture the forces and energies associated with the aggregation of asphaltenes and the π - π interactions between their aromatic rings,

leading to inaccuracies in modeling critical properties such as pressure, density, or diffusion coefficients [31,32]. There exists a need to develop a force-matched UA force field that accurately covers the molecules used to model bitumen. This force field must accurately represent large organic molecules, enable microscale phenomena exploration within reasonable computational efforts, and adhere closely to the thermodynamical and kinetics-related fundamentals observed in all-atom MD simulations of bitumen.

3. Methodology

This section details the methodologies, assumptions, inputs, and processes essential for developing the CG force field, organized systematically into twelve steps. The steps include defining bead types, mapping rules, interaction types, the core algorithms used to perform the force-matching process, and a thorough list of validation and benchmarking checks to make sure the CG force field generated captures the dynamics of a wide variety of all-atom hydrocarbon systems.

3.1. Simplification strategy and bead types

This study adopts a full UA approach to establish the mapping scheme for the CG force field, following a simplification strategy that mirrors the one used by Weiner et al. This approach simplifies the simulation by excluding detailed hydrogen interactions, thereby reducing the number of particles required. This simplification smooths out the forces and motion of the beads, decreases the degrees of freedom for particle movement, and accelerates kinetics-related phenomena. Despite these simplifications, the method aims to maintain the overall molecular geometries and closely approximate conformational and thermodynamic properties such as pressure and density, making it an attractive, but conservative approach for the modelling of bituminous molecules.

The force field employs seven criteria for bead differentiation: the chemical element of the parent atom, number of attached hydrogens, hybridization state, degree, ring presence, participation in an aromatic system, and the list of atomic elements bonded to the parent atom. Three combination tests are conducted with varying criteria: Test 1 using four criteria (parent atom element, hybridization, ring presence, and aromatic participation), Test 2 using six criteria (excluding the list of atomic elements bonded to the parent atom), and Test 3 considering all seven. These tests aim to ensure unique representation for each bead type while striking a balance [23]. An excessive number of bead types can risk statistical invalidity due to limited all-atom MD reference data, while too few bead types may lead to force-matching incongruities [33]. An

example application of how bead types are identified in Tests 1, 2, and 3 in a phenolic asphaltene pseudomolecule is seen in Fig. 2.

3.2. Reference molecules

The molecular set used in this study is derived from the work of Shisong et al. [34], building upon Greenfield's original set with additional modifications tailored to align with the SARA fractions and functional analysis of three 70/100 penetration grade bitumen samples from TotalEnergies, Norbit, and FHWA (Federal Highway Administration). The molecules, along with their chemical formula, mass, and estimated density, are presented in Table 1. This selection is designed to encompass a broad spectrum of hydrocarbon classes commonly present in bituminous materials such as alkanes, alkenes, polycyclic aromatics, and non-aromatics, along with functional groups such as phenols, oxanes, pyridines, thiophenes, and sulfoxides. This diverse composition makes these molecules particularly suitable for creating a CG force field capable of mimicking the behavior of complex hydrocarbon mixtures [35], especially comprised of molecules of high molecular mass and aromaticity, ubiquitous in bitumens. These are also selected to cover a wide array of molecules that cannot be accurately modelled using current UA force fields while addressing key applications in the field of bitumen studies.

3.3. Selecting the interaction types

The forces that govern the motion of particles, usually of atoms (in all-atom models) or beads (in CG models), are often represented as potential energy functions as the relationship between the force acting on a system of atoms and their energy [36] is expressed in Equation (1), as

$$F = -\nabla E \quad (1)$$

where F is the force field, and E is the potential energy function. In these force fields, multiple terms are included to account for the force contributions from various types of interactions, such as 2-, 3-, and 4-body interactions. This study's CG force field includes three interaction types: 2-body non-bonded pairwise and pairwise bonded interactions, and 3-body angular interactions. The system's total energy, E , is the sum of these interactions, given by Equation (2), as

$$E = \sum E^{nb} + E^b + E^a \quad (2)$$

where E^{nb} is energy from non-bonded interactions, E^b from 2-body bonded interactions, and E^a from 3-body bonded interactions. For non-bonded interactions, energy is calculated considering the number of beads (N) and interparticle energy E_{ij}^{nb} between particles i and j , with 153 unique pairwise interactions tabulated for LAMMPS. Bonded interactions use the energy E_{ij}^b between bonds ij , incorporating only 36 bonds based on their realistic occurrence in bitumen simulations. For 3-

body angular interactions, energy E_{ijk}^a between angles ijk formed by three bonded particles is calculated, covering 98 angular potentials. Therefore, the total energy E is given by Equation (3), as

$$E = \sum_{i=1}^{N_a} E_{ijk}^a + \sum_{i=1}^{N_b} E_{ij}^b + \sum_{i=1}^N \sum_{j=i+1}^N E_{ij}^{nb} \quad (3)$$

where N_b , and N_a are the total number of bonded and angular interactions in the system. Dihedral interactions are omitted due to the balance between accuracy and computational effort, and their vast potential types given that the potentials for different beads are not averaged or estimated using combination expressions [37]. This omission also aids in the force field's flexibility and transferability to different molecules in LAMMPS. Overall, the force field comprises 287 tabulated potentials, producing net energies and forces in a system whose molecules are analogous to those described in Table 1.

3.4. Building the reference all-atom MD models for use in the CG process

Reference all-atom systems are constructed according to their capacity to yield a wide range of equilibrium trajectories. This approach aims to maximize both the quantity and quality of energy and force data available for each bead and interaction type, which are essential for creating the tabulated force fields. The all-atom reference models are constructed using LAMMPS loaded with the Polymer Consistent Force Field (PCFF) [38]. The total energy of the all-atom system estimated using the PCFF, E_{PCFF} , is given by Equation (4), as

$$E_{PCFF} = \sum E^b + E^{ia} + E^{oa} + E^t + E^{bb} + E^{ba} + E^{bt} + E^{aa} + E^{at} + E^{tt} + E^{VDW} + E^{coul} \quad (4)$$

where the detailed description of the interaction terms can be found in the Supplementary Information (SI), Table S1. Folder `/0.pcff_sim` in the SI contains the input scripts necessary to run a sample bitumen simulation with the PCFF parameters loaded. The LAMMPS routines used to initialize these systems involve eight sequential steps, listed as follows:

1. Obtaining the SMILES notation of the molecules to be used and using the Rdkit Python module to initialize them into a stable 3D conformer.
2. Evenly distributing the molecules in a low-density box (0.20 g/cm³), positioning them away from z-direction walls to facilitate uniaxial compression to equilibrium density.
3. Loading systems into LAMMPS, assigning force field types and charges and minimizing the atomic positions using a conjugate gradients method until the energies and the forces converge to magnitudes below 1.0.
4. Compressing systems to target density under NVT control, over 5 ns, a true strain rate of 1 %, and deformations occurring every 1000 steps.

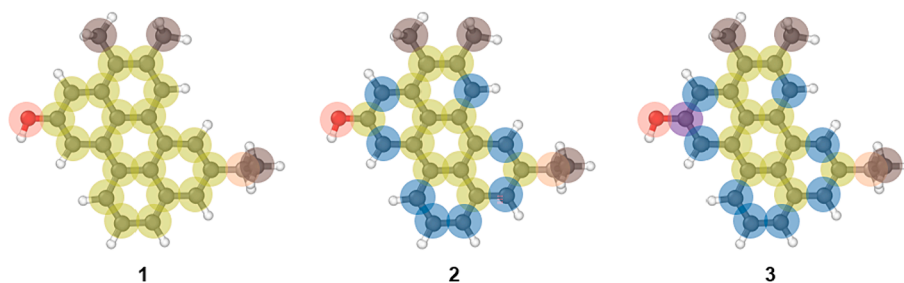


Fig. 2. Transformation of an all-atom phenolic asphaltene, with a molecular formula of $C_{24}H_{20}O$ consisting of 45 atoms, into its UA analog. The UA version simplifies the molecule to 25 beads, with each bead representing a parent atom along with its associated bonded hydrogens. Tests 1, 2, and 3 use different combinations of bead identification criteria (parent element, bonded hydrogens, hybridization, degree, ring participation, aromaticity, and neighboring elements) to generate 4, 6, and 7 bead types respectively.

Table 1

Illustration of the chemical structures of the molecules utilized in this investigation. Within the Saturates category, (1) squalane and (2) hopane are depicted. The Aromatics group includes (3) dioctylcyclohexane naphthalene and (4) perhydrophenanthrene naphthalene. In the Resins category, (5) quinolinohopane, (6) thioisorenieratane, (7) benzobisbenzothiophene, (8) pyridinohopane, and (9) trimethylbenzeneoxane are shown. Lastly, the Asphaltenes category comprises (10) phenolic asphaltene, (11) pyrrolic asphaltene, and (12) thiophenic asphaltene.

 1 C ₃₀ H ₆₂ 422.8 0.803	 2 C ₃₅ H ₆₂ 482.9 0.913	 3 C ₃₀ H ₆₆ 406.7 0.916	 4 C ₃₅ H ₆₄ 464.7 1.030	 5 C ₄₀ H ₅₀ N 553.9 1.007	 6 C ₄₀ H ₆₀ OS 589.0 1.010
 7 C ₃₈ H ₁₀ OS ₂ 306.4 1.540	 8 C ₃₀ H ₂₇ N 503.9 0.977	 9 C ₂₉ H ₅₀ O 414.7 0.893	 10 C ₄₂ H ₅₁ O 574.9 1.049	 11 C ₆₀ H ₆₁ N 888.4 1.104	 12 C ₅₁ H ₆₂ S 707.1 1.100

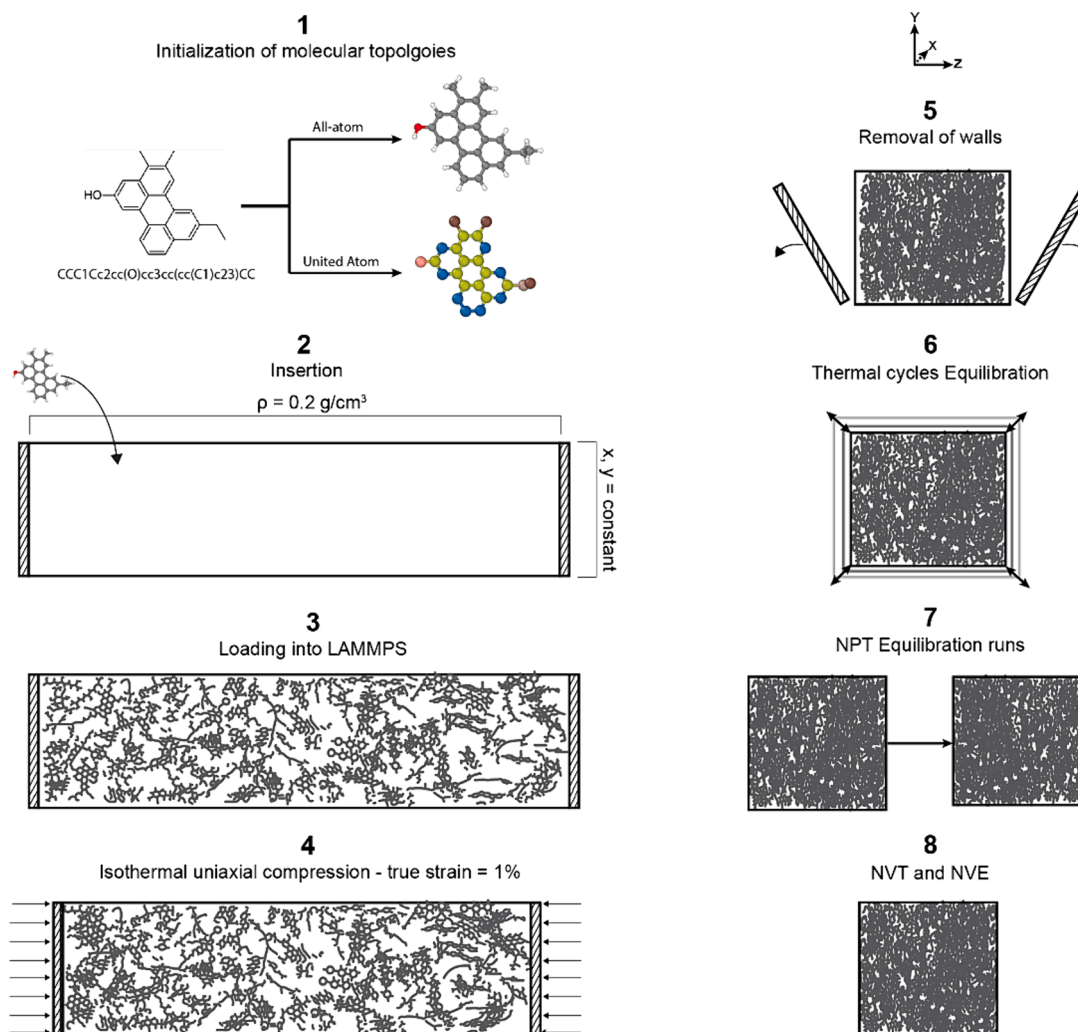


Fig. 3. Visual representation of the steps performed in this study to initialize molecular systems in both all-atom and CG models.

5. Removing compression walls, applying full periodic boundary conditions.
6. Conducting 50 NPT heating cycles for system equilibration. The temperature of the system is varied sinusoidally with an amplitude of $\pm 25\%$ of the original equilibrium temperature and a frequency of 1 ns during these cycles.
7. Stabilizing the density of the system at equilibrium temperature and pressure by using two consecutive 50 ns-long NPT stages. The final density is obtained from the average density of the second NPT stage.
8. Conducting dynamics stages under NVT and NVE conditions, each for 50 ns, to obtain atomic trajectories that are uninterrupted by the use of a barostat or a thermostat.

Fig. 3 depicts the visual representation of the steps employed for initializing molecular systems to their final, condensed density, both in all-atom and CG models. Modifications, omissions, or extensions to these steps may occur to enhance efficiency or meet particular requirements, which will be clearly indicated when such adjustments are made.

The thermostat and barostat involve the use of a modified Nose-Hoover [39,40] integration algorithm that incorporates a drag factor to reduce oscillatory effects, with a pressure and temperature damping factors equal to 100 steps, and a particle velocity drag coefficient equal to 1.0 unless explicitly noted. Each LAMMPS subroutine is set to generate 1000 trajectories and instantaneous properties at each point, unless noted otherwise.

The reference models fed to the CG algorithm are derived from the final NVE simulation, encompassing 2000 trajectories detailing particle positions (unscaled and unwrapped), velocities, forces, and masses in the LAMMPS trajectory file. The emphasis was on using systems run under NVE conditions to minimize artificial perturbations to particle motion and avoid unrealistic values of energies and forces in the tabulated potentials.

Additionally, bitumen systems, with their diverse, high molecular hydrocarbons, present a unique challenge when creating reference all-atom models to feed to the CG algorithm. Ideally, an all-atom MD model should encompass all molecules to be parametrized, but this approach imposes an inherent imbalance in bead type occurrences especially when using the molecules of Table 1, leading to gaps in interaction distances and parametrization quality. To overcome these challenges, a systematic approach is used to develop an appropriate force field for each bead and interaction type, starting with reference systems that contain all 12 molecule types. If bead types fail to meet validation standards, they are excluded initially and addressed in a secondary attempt by building reference systems with altered molecule numbers containing a favorable count of unparametrized bead types.

For bead types that remain unparametrized, reference systems containing just one molecule type are built. This third attempt allows for the creation of single-component mixtures where the number of beads that specifically require more attention is maxed out. If tabulated potentials for certain bead types remain elusive, a final attempt is made by using pseudomolecules [38]. These chemically analogous, but structurally different molecules fill gaps in the force field's tabulated data, but their use may affect the force field's accuracy [41]. The comprehensive list of molecules and pseudomolecules used as reference in this study can be found in Table S2 and S3 in the SI. The list of systems of molecules used as reference can be found in Tables S5 and S6 in the SI.

3.5. Thermodynamic conditions

Defining thermodynamic conditions is a key step in CG force field development, as these force fields detail energy and force interactions at varying interparticle distances (r_{ij}). Key properties such as temperature, pressure, and density are integral to these systems and must be accurately represented within the force field's range [42]. High-pressure scenarios, for instance, require calibration for shorter interparticle distances.

Bitumens experience diverse conditions throughout their lifecycle, including processing, manufacturing, and other Civil Engineering applications. While this study's CG force field does not simulate production phases (e.g., refinery conditions), it must represent conditions encountered in applications like asphalt mixing, road usage, and laboratory settings (e.g., aging bitumen samples at 80 °C and 2000 atm).

Therefore, the CG force field is designed to cover temperatures from 0 to 250 °C and pressures from 0.5 to 2500 atm, where energies and forces within all-atom reference systems are parameterized across these conditions. Therefore, the reference systems generated in Section 3.4 are built under three specific equilibrium scenarios: at 25 °C and 1 atm (typical ambient conditions), at 250 °C and 1 atm (high-temperature scenarios), and at 250 °C and 2500 atm (extreme conditions in lab conditioning settings).

3.6. Force-matching algorithm

A force-matching approach is utilized to derive the force field of this study, using the OpenMCSG package developed by Voth et al. [43]. The initial phase involves the computation of the COM for each CG bead, derived from the all-atom model [43] and the mapping rules defined in Section 3.1. The COM for a bead, comprising atoms with positions r_i and masses m_i , is calculated using Equation (5),

$$\text{COM}_i = \frac{1}{M} \sum_j m_j r_j \quad (5)$$

Here, M represents the total mass of the atoms in bead i . The forces F_i acting on each COM are obtained from the all-atom reference trajectories and aggregated for all atoms within a CG bead, resulting in the net force on a CG bead as the sum of forces on its constituent atoms. The force on a CG bead F^{CG} is then determined in Equation (6),

$$F^{\text{CG}} = \sum_{i \in \text{bead}} F_i \quad (6)$$

The force-matching algorithm tunes the parameters θ of the CG potential to ensure the forces $F_i^{\text{CG}}(\theta)$ it predicts closely resemble those from the all-atom model [33,43]. This objective is formulated as a minimization-optimization problem, as given by Equation 7,

$$\min_{\theta} \sum_{i=1}^N \|F_i^{\text{AA}} - F_i^{\text{CG}}(\theta)\|^2$$

In this formulation, θ symbolizes the CG potential parameters, N is the number of CG beads, F_i^{AA} is the force on the i -th CG bead from the all-atom model, and $F_i^{\text{CG}}(\theta)$ is the predicted force on the CG bead. The algorithm commences with initial guesses for θ and iteratively refines them to minimize the objective function. The optimization continues until the objective function converges to 0.01 % of its magnitude using a conjugate gradient method [44]. The parameter θ can encompass various quantities, including potential energy functional forms, force constants, and equilibrium values like bond lengths, angles, and dihedral angles, alongside more complex thermodynamic properties [45]. However, only the forces are mirrored in the force field generated in this study. This is because additional constraints can improve CG force field fidelity but may limit the solution space or lead to local minima convergence, reducing its transferability to other systems beyond those used as reference [23].

Tabulated energy and force values are generated for various bead and interaction types. These tables contain 1000 points, obtained using sixth-degree B-spline interpolations. The module exports 2-body non-bonded and bonded interactions in terms of interparticle distance in Angstroms, and 3-body angular interactions in degrees, ready to be used in LAMMPS simulations involving *Real* units. The ranges of these tables, and thus the ranges of the force fields, are derived from the Radial Distribution Functions (RDFs) obtained from the system's interparticle distances, bond lengths, and angles. These RDFs are extracted from the

reference all-atom CG trajectories.

Points positioned within 10 % of the tables' boundaries are excluded due to their association with rare geometrical conformations, which tend to yield statistically unreliable force and energy values. The thermodynamic scenarios detailed in Section 3.5, namely 250 °C and 1 atm, in addition to 250 °C and 2500 atm, generate tabulated ranges extending beyond those observed in systems stabilized at standard conditions, which are 25 °C and 1 atm. The incorporation of energy and force values from these extended ranges into the tables' boundaries serves to broaden the range of the original tables, capturing conformations that would only occur at higher pressure or temperature conditions, enhancing the CG force field's ability to model high density systems and improving Initialization and Equilibration routines in MD simulations.

The OpenMSCG package, in its developmental phase, requires adaptations for effective, efficient, and automated processing of complex hydrocarbon mixtures. These adaptations include automation of input generation, sanity checks, post-processing, and seamless LAMMPS integration in high-performance computing environments. A Python script is developed for handling operations, accommodating hundreds of interaction types.

3.7. Postprocessing energies and forces

Data generated for tabulated potentials through force matching algorithms can sometimes struggle to accurately represent the forces and energies in all-atom MD simulations. This issue is especially prominent at the extremes of the data range, where there are fewer data points [46]. To address these issues, padding techniques are applied at the potential's distance range limits. This ensures reasonable behavior at extreme distances, enhancing the simulation's physical realism at distances poorly covered (if at all) by the tables [43].

Control over these extensions is achieved through a custom algorithm, enabling iterative adjustments of various parameters for a smooth and controlled extension. The algorithm extends data points along a linear trend derived from the last values of the existing dataset, maintaining equal spacing between the new points and those already present. For 2-body non-bonded pairwise interactions and bonded interactions, and 3-body angular interactions, the extensions are applied to cover interatomic distances of 0.5 to 20 Å, 0.5 to 5 Å, and angular ranges from 0 to 180 degrees, respectively.

Parameters for this linear extension vary, including the number of points used from each extreme side and the level of smoothing applied to the plot. Once the extended tabulated potentials are generated, they undergo a validation check to ascertain whether the data points adhere to physically realistic trends before they are tested in simulations. If the data points do not conform to realistic standards, refinement is applied. For non-bonded pairwise interactions, both energies and forces undergo checks to ensure repulsive behavior at shorter distances and tending to zero at longer distances, with equilibrium points reflecting Lennard-Jones/Coulomb potential characteristics. Bonded pairwise and 3-body angular interactions are predominantly characterized by harmonic features, implying quadratic energy profiles and nearly linear force trends. The presence of multiple equilibrium points is expected and is not considered a factor to reject the validity of a tabulated potential.

Data points before extension should demonstrate smoothness, tested using a Savitzky-Golay filter. Postprocessed potentials are then scored based on the number of refinement iterations required, indicating the accuracy and extent of modification. Scores range from 1, indicating minimal modification, to higher values signifying increased modifications or the need for recreation of the table for specific interaction types. Therefore, a score of 1 representing only one iteration is considered ideal. Scores exceeding the maximum iteration limit indicate the need to recreate the table using other, more relevant reference systems. Intermediate scores, though not ideal, could be acceptable if attaining a score of 1 is highly challenging for the specific interaction type being handled.

3.8. Validation against all-atom reference models

This section conducts six validation checks to ensure that the CG models correctly capture the stability, conformational accuracy, and the thermodynamic and kinetic characteristics of the all-atom reference models [35,47]. If a check fails, the CG force field undergoes optimization before proceeding with further validation checks. The CG systems are created by using a Python algorithm designed to transform all-atom models into their CG counterparts. The algorithm maps the positions of beads from an all-atom trajectory file and assigns appropriate CG force field types, thereby allowing CG simulation routines to be run after a specific all-atom trajectory. For Validations 1, 2, and 3, the methodology employs the final trajectory from the second NPT simulation described in Section 3.4. In contrast, Validations 4, 5, and 6 utilize the initial trajectories from the all-atom simulations, which are then subject to LAMMPS routines akin to those described in Section 3.4. All the all-atom systems used as reference trajectories to build the CG force field of this paper are subjected to these validation checks. These validation checks are listed and described as follows.

Validation 1 evaluates the force field's numerical stability in LAMMPS, beginning with a minimization stage using a conjugate gradients method until the energies and the forces converge to magnitudes below 1.0, followed by an NPT stage with a 0.01 fs integration step and temperature and pressure adjustments every 10 steps. The simulations are run for 100 ps. This highly detailed simulation routine helps in identifying the source (i.e., which bead and interaction type) of possible issues caused by instabilities, such as lost particles, significant bond or angle deviations from the reference systems, or failure to reach equilibrium, like thermodynamic properties not stabilizing (e.g., 298 ± 182 K).

Validation 2 tests the force field's stability under different control conditions through the use of NPT, NVT, and NVE simulations. This includes five subsequent NPT simulations of 500 picoseconds each (after that of Validation 1), a time step of 1 fs, each with control frequencies varying from every 25, 100, and 500 steps, and back. This assesses the force field's response to more normal simulation parameters and changing thermostat and barostat levels. This is followed by five NVT stages, mirroring the NPT pattern, but just with a thermostat. A single NVE simulation 5 ns long without external controls completes this check, where temperature and pressure are monitored to stay within 10 % of their initial average values, verifying the force field's ability to perform without any artificial dynamics corrections to stabilize the simulation properties.

Validation 3 further tests the force field's stability under diverse conditions by replicating the NPT, NVT, and NVE stages from Validation 2, but at 250 °C and 1 atm, and then 250 °C and 2500 atm. This process assesses the force field's stability under more extreme conditions, especially during the NVE stage involving no external control.

Validation 4 ensures observed properties are not artifacts of MD algorithm corrections. It firstly assesses if the particles' kinetic energy follows a Maxwell-Boltzmann distribution, using the probability density function of Equation (8), given as follows,

$$f(E) = \frac{2}{kT} \left(\frac{E}{kT} \right)^{0.5} e^{-\frac{E}{kT}} \quad (8)$$

where $f(E)$ is the probability density function of kinetic energy, k is the Boltzmann constant, T is the temperature of the system in Kelvin, and E is the kinetic energy of the particles. This ensures early identification of unrealistic energy distributions among the particles.

A secondary check examines particle distribution in the simulation box, identifying any significant empty spaces or uneven distributions. This is crucial for verifying that average pressure or density accurately reflects overall particle interactions. Density profiles are calculated in the x , y , and z dimensions, covering 5 % of the box's length, with a criterion that their densities should not vary by more than 20 % from the

box's average density.

Validation 5 verifies the alignment of conformations and geometries between the CG and all-atom models. It involves calculating and comparing RDFs for all 287 interaction types in both CG and all-atom systems. For example, a bond distance of $1.45 \pm 0.12 \text{ \AA}$ between CH2 and CH beads in the all-atom model should closely match in the CG model, confirming similar geometries. Accurate replication of molecular geometries, especially for structures like planar polycyclic aromatic groups in asphaltene molecules, is crucial for realistically representing bitumen's molecular characteristics. The RDF is calculated using Equation (9), as

$$g(r) = \frac{V}{N^2} \frac{dN(r)}{4\pi r^2 dr} \quad (9)$$

where $g(r)$ is the radial distribution function at a distance r , V is the volume of the system, N is the total number of interaction types being analyzed (e.g., number of bond types), and $dN(r)$ is the average number of particles found within a spherical shell of radius r and thickness dr surrounding a reference particle. This validation check is performed on the trajectories produced by the NVE routine of Section 3.4. This ensures the CG model replicates the all-atom model's geometric and spatial properties in simulations without artificial control interruptions.

Validation 6 assesses the force field's precision in mirroring key thermodynamic properties like potential energy, temperature, pressure, and density, comparing mean values between CG and all-atom MD simulations and, where possible, against experimental data. The validation includes statistical analysis to ensure measurements fall within a 95 % confidence interval and match the all-atom model's standard deviations.

3.9. Testing multiple molecular systems

In this section, the transferability and flexibility of the force field are tested with molecules that are not part of the reference set of molecules used to build the CG force field [48]. The construction of these molecular systems is done following the steps detailed in Section 3.4, employing both the PCFF all-atom force field and the CG force field to create two parallel models. Validation checks 3 through 6 from Section 3.8 are applied to these systems. The systems created include various molecules used in bitumens, rejuvenators, and other related substances, covering over 100 different molecules and mixtures. This includes the 12-molecule bitumen samples from TotalEnergies, Norbit, and FHWA (presented in Table 1), and a wide range of additional hydrocarbons and functionalized organic compounds, sorted by molecular weight from 50 g/mol to over 2000 g/mol, and ensuring all are liquid at 298 K and 1 atm (presented in Table 2). The molecules in Table 2 involve adapting bitumen molecules from Table 1 to align with those commonly used in the literature, focusing on how molecular structure changes impact properties in the CG model. Smaller molecules are purposely tested to evaluate the force field's expected inability to capture molecules that resemble those used to parametrize other UA force fields. The SMILES notations of the molecules tested, along with the number of molecules used in each system are available in Table S4 and S7 in the SI.

The force field's ability to simulate established hydrocarbon behaviors is evaluated given the broad availability of molecular systems. If the force field contradicts these known trends, it is deemed inadequate. These trends include: 1) A decrease in density with the addition of more branches to a molecule, 2) A rise in density with an increased number of aromatic rings in the compound, 3) An increase in density due to the incorporation of polar functional groups, often through heteroatoms or reduced carbons, and 4) A density increase correlating with a rise in the molecular weight of the compound [49,50].

3.10. Equivalency in all-atom and CG timescales

Section 3.10 in CG force field development focuses on measuring the timescale equivalency between CG and all-atom MD models. Establishing a conversion factor for time translation between CG and all-atom simulations is crucial, as the timescales in CG simulations correspond to longer periods in all-atom MD simulations [51].

To establish a conversion factor for these differing timescales, a Python algorithm is employed to analyze both the all-atom and CG trajectories. This algorithm determines the time required for molecules to be displaced over a certain distance, utilizing their mean-squared displacement (MSD) and self-diffusivity coefficients. Additionally, it focuses on calculating vibratory frequencies (and corresponding times) in bonding and angular interactions present in the system. From these computations, an overall conversion factor for the timescales in both models is derived. This conversion factor facilitates the comparison of timescales between the CG and all-atom models, given in Equation (10),

$$t_{\text{net}} = \frac{t_{\text{CG}}}{t_{\text{aa}}} \quad (10)$$

where t_{CG} and t_{aa} are the elapsed times from the CG and the all-atom simulations. The Python algorithm operates by processing the trajectories from either the all-atom or CG systems. It executes three sub-routines in parallel. The first subroutine computes the MSD and diffusivities of the COMs of the molecules in the system, focusing on their net displacement. MSD is calculated for entire molecules rather than individual particles or beads. This approach was chosen to exclude the influence of bonding and angular vibrations on the computed displacement. The MSD is calculated using Equation (11),

$$D_n = \frac{1}{2} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle (r_n(t) - r_{n,0})^2 \rangle, \quad (11)$$

where D_n is the self-diffusivity constant in the n^{th} dimension, $r_n(t)$ is the position vector of the COM of the molecules at time t , $r_{n,0}$ is the position of the COM of the molecules at time $t = 0$, and the brackets represent the average over the molecules. The second and third sub-routines are designed to handle harmonic motions in 2-body bonded and 3-body angular interactions. They calculate the interparticle distance (or angle) over an extended duration, encompassing multiple oscillation cycles. The average time required to complete a full cycle is then considered the time constant for both the all-atom and CG systems.

3.11. Density benchmark: CG and Weiner's

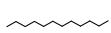

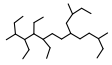

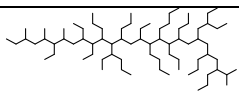



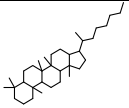
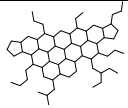



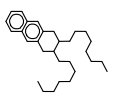
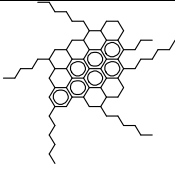



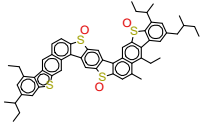
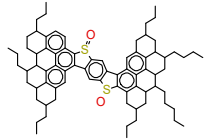

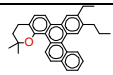
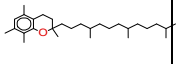
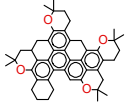
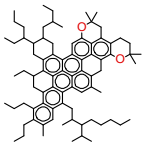
Section 3.11 focuses on comparing the capacity of the developed CG force field to that of Weiner's et Al. in accurately representing the densities of the CG models. Correctly estimating the density in an MD model is a decent indicative of the force field realistically representing the energies and forces of the particles in the simulation, given the already comprehensive list of validation checks performed in Section 3.8. This is done by estimating the density computed using Weiner's force field (ρ_{Weiner}). This aims to establish an idea on how well applicable both force fields are to estimate the density of a hydrocarbon system, especially in relation to the variability of the molecules presented in Table 2. The construction of molecular systems employing Weiner's force field follows a methodology akin to that outlined in Section 3.4, replicating the procedures delineated therein.

3.12. Performance benchmarking

Section 3.12 involves benchmarking the computational efficiency of the CG force field against the all-atom model, crucial for assessing practical benefits in computation time and resources. Two benchmarks measure performance:


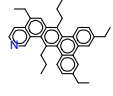
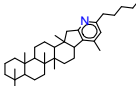
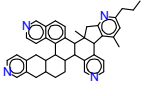
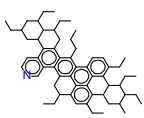

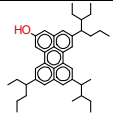
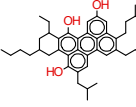
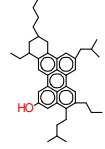
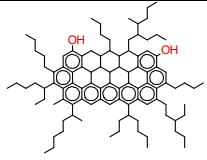


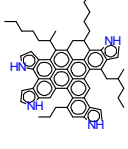
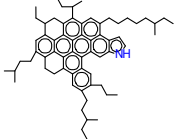
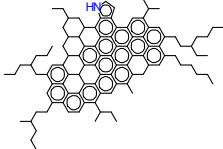


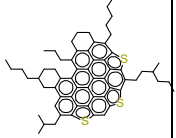
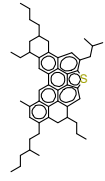
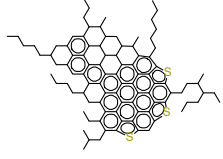
Table 2

Key molecules employed for evaluating the effectiveness of the CG force field on systems not used as reference. The numeric classification (1–9) identifies different molecule families, each distinguished by a distinct functional group. The character classification (A–E) represents increasingly complicated iterations of the principal molecule. The caption X | Y | Z comprises the chemical formula, molecular mass (g/mol), and the estimated density (g/cm³) of the molecule.

	A	B	C	D	E
1	 C ₁₂ H ₂₆ 170.3 0.751	 C ₁₃ H ₂₈ 184.4 0.755	 C ₂₇ H ₅₆ 380.7 0.798	 C ₃₀ H ₆₂ 422.8 0.803	 C ₇₇ H ₁₅₆ 1082.1 0.827
2	 C ₁₅ H ₂₆ 206.4 0.893	 C ₂₁ H ₃₄ 286.5 0.965	 C ₃₀ H ₅₂ 412.7 0.889	 C ₃₅ H ₆₂ 482.9 0.913	 C ₆₇ H ₁₁₀ 915.6 0.951
3	 C ₁₂ H ₁₆ 160.2 0.909	 C ₁₄ H ₁₄ 182.3 1.069	 C ₂₂ H ₃₀ 294.5 0.961	 C ₃₀ H ₄₆ 406.7 0.916	 C ₇₅ H ₉₈ 999.6 1.084
4	 C ₁₂ H ₈ OS 200.25 1.410	 C ₁₈ H ₁₀ S ₂ 290.4 1.417	 C ₁₈ H ₁₀ S ₂ 306.4 1.540	 C ₅₆ H ₅₄ O ₂ S ₄ 903.3 1.390	 C ₇₇ H ₉₄ O ₂ S ₂ 1115.7 1.240
5	 C ₁₁ H ₁₄ O 162.2 0.978	 C ₃₂ H ₃₂ O 432.6 1.125	 C ₂₉ H ₅₀ O 414.7 0.893	 C ₄₇ H ₅₀ O ₄ 678.9 1.234	 C ₇₇ H ₉₄ O ₂ S ₂ 1115.7 1.240

(continued on next page)

Table 2 (continued)

					$C_{61}H_{106}O_2$ 1111.7 1.066
6	 $C_9H_{13}N$ 135.2 0.910	 $C_{37}H_{39}N$ 497.7 1.105	 $C_{36}H_{57}N$ 503.8 0.977	 $C_{42}H_{42}N_4$ 602.8 1.197	 $C_{61}H_{77}N$ 824.3 1.047
7	 $C_{10}H_8O$ 144.2 1.181	 $C_{41}H_{54}O$ 562.9 1.035	 $C_{42}H_{54}O_3$ 606.9 1.114	 $C_{42}H_{54}O$ 574.9 1.049	 $C_{102}H_{136}O_2$ 1394.2 1.082
8	 C_8H_7N 117.1 1.149	 $C_{22}H_{13}N$ 291.3 1.397	 $C_{61}H_{64}N_4$ 853.2 1.251	 $C_{66}H_{81}N$ 888.4 1.104	 $C_{108}H_{113}N$ 1425.0 1.237
9	 C_8H_6S 134.2 1.187	 $C_{24}H_{12}S$ 332.4 1.483	 $C_{64}H_{66}S_3$ 931.4 1.295	 $C_{51}H_{62}S$ 707.1 1.100	 $C_{98}H_{102}S_3$ 1376.0 1.282

- Marking 1: Compares the time to complete an identical number of simulation steps in both models.
- Marking 2: Considers timescale equivalency and the time required to effectively simulate 50 ns, factoring in the CG and all-atom timescale conversion.

Simulations involve initialized and equilibrated bitumen models as per Table 1 and Table 2, testing five system sizes with molecule counts scaled by factors of 1, 2, 4, 8, and 16. The molecular composition of the systems (with scale factor equal to 1) can be found on Table S7. Performance data is normalized against the all-atom simulations for comparative analysis, with constant computational resources across simulations to focus solely on the CG model's inherent efficiency. The systems used in this section are built using steps akin to those of Section 3.4. The performance assessment is conducted using brief NPT simulations of 100 picoseconds at equilibrium density, considering the substantial scale of the systems and the absence of a requirement to model extended simulations for evaluating algorithm performance.

3.13. Testing for phase separation

In this testing phase, the CG force field's capability to replicate intermolecular structure formation, as highlighted by Assaf et al. [18] and others [19,52–54], is assessed. These intermolecular formations are pivotal for defining phase-separated mixture morphologies and microstructural characteristics in bituminous materials, emerging from the agglomeration of molecules of similar size, shape, and polarity.

Moreover, the agglomeration of particles holding different interaction potentials is not indicative of correct redistribution of phases. Inaccurate force field definitions can lead to the formation of unrealistic formations (if they form at all), hence the importance on measuring the force field's capacity to replicate this phenomenon correctly through the evaluation of the evolution of morphological features over simulation time.

Six hydrocarbon mixtures are built and tested to evaluate this phenomenon, deliberately using molecules of Table 2 that exhibit distinct polarities, in both all-atom and CG models. These systems ranged from simple binary mixtures with significantly differing molecular polarities to more complex mixtures resembling bitumen, featuring slight polarity variations among multiple molecule types. The construction method for these mixtures parallels that described in Section 3.4, with an alteration in the simulation process where the second NPT stage (step 4) is prolonged to a certain number of steps until phase separation metrics (described below) stabilize within 2 % of their final 1 million steps average, or until the total number of steps reaches $5e9$, foregoing the NVT and NVE steps. This is done to account for the impact of t_{net} in the CG model which is expected to have faster kinetics (see Section 3.10) and thus exhibit phase separation over fewer simulation steps. The simulations are performed at 298 K and 1 atm, with an integration step of 1 fs in both all-atom and CG models. The molecule quantities in these mixtures are adjusted to ensure a comparable number of beads per phase, aiming to enhance phase separation effects and inhibit the development of minor aggregates that could elude precise detection by the metrics mentioned below. In each all-atom mixture, the total atom count is set to 30,000, leading to CG models that have a bead count

Table 3

Compilation of mixtures containing molecules from different SARA fractions. Molecules (and their numbers) are selectively chosen to enhance phase separation phenomena. The structure of these molecules is sourced from Table 2, ensuring uniform bead counts across phases, with an objective to maintain a total particle count of 30,000 in the all-atom models for each mixture.

Mixture	Molecular composition				Particle count	
	Saturates	Aromatics	Resins	Asphaltenes	All-atom	CG
1	1A: 395	–	–	8C: 116	30	12
					000	295
2	–	3D: 197	–	8C: 116	30	13
					000	479
3	–	–	4C: 484	8C: 116	30	17
					000	719
4	1A: 263	–	4C: 323	8C: 78	30	14
					000	971
5	1A: 197	3D: 99	4C: 242	8C: 58	30	14
					000	189
6	1A: 941B: 94	3B: 1073C: 107	4B: 1254C: 125	8C: 288D: 28	30	14
					000	906

reduced by approximately half. The molecular composition and particle count of the mixtures built are presented in Table 3.

To quantitatively analyze the distribution of these phases within the mixtures over time, four metrics are consistently monitored throughout the simulations in both the all-atom and CG models. These metrics aim to reflect the agglomeration phenomena observed in all-atom simulations within the CG models, providing a comparative benchmark to assess the CG force field capacity to replicate these morphologies when compared to the all-atom models.

Metric 1. The total potential energy of the system is evaluated, represented by E_{PCFF} in the all-atom model and E_{CG} in the CG model. A decline in total potential energy, accompanied by the emergence of more heterogeneous morphologies, suggests the system's thermodynamic inclination towards phase separation.

Metric 2. This metric is designed to measure the degree to which particles are neighbored by others belonging to the same phase in comparison to those from different phases. The metric is defined as $P_{Y|X}$, formulated in Equation (12) as

$$P_{Y|X} = 100 \frac{N_{Y|X}}{N_{total|X}} \quad (12)$$

In this expression, $P_{Y|X}$ represents the proportion of phase Y particles in the proximity of phase X particles. $N_{Y|X}$ refers to the count of phase Y particles neighboring phase X particles, while $N_{total|X}$ is the total count of particles neighboring phase X particles, encompassing those from all phases. A higher value of $P_{X|X}$ (where X is the same phase) indicates a more pronounced aggregation, pointing to a stronger inclination towards aggregation and heterogeneous formations. The estimation of neighboring particles is based on phase-specific RDFs, $g_X(r)$, derived through the integration of Equation (9). Particles are considered neighbors if their interparticle distance is below 10 Å. There are four possible types of fractions in a mixture, each represented by a specific number: Saturates (1), Aromatics (2), Resins (3), and Asphaltenes (4). For instance, the notation $P_{3|4}$ signifies the ratio of asphaltenic particles neighboring resinous particles.

Metric 3. This metric builds upon Metric 2 by assessing the proportion of particles within two distinct regions of a system: *interface* and *bulk* regions. Particles are classified as part of interface regions if they are mainly surrounded by particles from a differing phase, or as part of the bulk regions if the majority of their neighboring particles are from the same phase. This proportion, P_{bulk} , is given by Equation (13)

$$P_{bulk} = 100 \frac{N_{bulk}}{N_{bulk} + N_{interface}} \quad (13)$$

where $N_{interface}$ is the number of particles in interfacial regions, and N_{bulk} is the number of particles in bulk regions. The method for determining the presence of particles in these regions is consistent with that used to compute Metric 2, where particles in interface regions are those with the highest percentage of $P_{X|Y}$ for Y representing a phase different from their own, and particles in bulk regions are those with the highest percentage of $P_{X|X}$ for X corresponding to their own phase. In systems where phases are more uniformly distributed, P_{bulk} is anticipated to be lower compared to systems with molecules in more clustered formations.

Metric 4. Metric 4 is designed to quantify the number of asphaltene clusters present in mixtures at any given time, given that all six mixtures contain an asphaltenic phase. This is achieved by employing a molecular cluster counting algorithm during each step of the simulations. According to the algorithm, an asphaltene molecule is part of a cluster if the carbon atoms in one of its aromatic rings are within a neighboring distance to a carbon atom in the aromatic ring of another asphaltene molecule. The neighboring distance is set to 5 Angstroms, which corresponds to the observed average distance between stacked aromatic rings in asphaltene clusters within bitumen [55]. Isolated molecules are considered individual clusters if no other molecules are in proximity. This implies that in an ideal dispersion of asphaltenes, where each molecule is isolated, the count of asphaltene clusters would match the total number of asphaltene molecules present in the system. In other words, a higher number of asphaltene clusters denotes a less agglomerated system.

The evolution of these metrics in the CG model differs from that in the all-atom models due to the distinct time scales governing each model (see Section 3.10). Despite these differences, it is expected that the phase-specific metrics will stabilize to comparable magnitudes across both models. Similarly, the observed trends in potential energy measurements are expected to converge, indicating that, despite the inherent timescale differences, both models should reflect consistent thermodynamic and structural behaviors regarding the morphological distribution of its SARA fractions. Significant discrepancies between all-atom and CG metrics would indicate that the developed CG force field does not realistically capture close-range energies and forces.

4. Results

The section commences with an introduction to the nature of the force field beads and their interactions. It then proceeds to discuss the stability and the ability of the CG force field to replicate the geometries of all-atom molecules. Following this, the results comparing the thermodynamic properties, timing benchmarks, and the capabilities of the force field in modeling large-scale phenomena, such as phase separation, are presented.

4.1. Bead types

Test 1 (see Section 3.1), employing four differentiation criteria based on the parent atom's chemical element, hybridization, presence in a ring, and participation in an aromatic system, yielded nine bead types. This number was proved insufficient as it often necessitated using the same bead for distinct chemical structures, particularly when parent atoms had varying numbers of adjacent parent atoms. Such an approach led to instability in simulations, where systems failed to achieve numerical equilibrium or did so only through stringent equilibration algorithms, enforcing non-realistic spatial conformations.

Test 3, expanding the criteria to seven, including the parent atom's chemical element, number of bonded hydrogens, hybridization, degree, ring presence, aromatic system participation, and the chemical element of neighboring atoms for each bead, resulted in 26 bead types. This increase, though substantial for the range of cases in the force field and for similar beads requiring no differentiation, demanded extensive

computational resources and a broader range of all-atom reference systems. Gathering statistical data to fill tabulated potentials for closely related bead types (e.g., differing in only one criterion) often led to noisy, unrealistic, and low-quality potentials that failed to meet early validation checks of Section 3.8.

Test 2, adopting six criteria – the parent atom’s chemical element, the number of bonded hydrogens, hybridization, degree, ring presence, and aromatic system inclusion – struck a better balance. This approach did not account for the explicit chemical elements of neighboring parent atoms but used the parent atom’s hybridization and degree for effective differentiation. This method passed all checks of Section 3.8 without issues. Consequently, the resulting CG force field includes 17 bead types covering various atom-hydrogen groups in the molecules listed in Table 1 and Table 2, including sulfoxides, ketones, phenols, thiophenes, pyrroles, oxanes, quinolines, and sulfoxides. These beads still facilitate constructing CG models of other hydrocarbons frequently mentioned in bitumen literature, like saturates, aromatics, resins, asphaltenes, some polymers, rejuvenators, additives, bio-binders, and common solvents. The bead types, labeled from A to Q, have been assigned specific colors, with detailed descriptions presented in Table 4. Table 5 shows the molecules of Table 1 converted into their CG analog, depicting all 17 bead types being utilized to form a variety of bitumen molecules.

4.2. Mapping quality and distribution

The molecules selected to create reference systems, as detailed in Table 1, display an uneven distribution of bead types. For these systems, the skewness in bead type distribution is significant. For every bead of types E, G, M, O, and P, there are approximately 100 times more beads of types Q, J, K, L, A, C, and D. This imbalance is anticipated, with the former group representing unique functional groups appearing once per molecule (e.g., oxane oxygens of type G), and the latter forming the core

structure of every hydrocarbon molecule (beads depicting CH combinations in branches and cycloalkanes, of type J). This imbalance becomes more pronounced when deriving tabulated potentials for angular interactions. Combinations in these scenarios are rarer in the MD models, and their skewed distribution becomes even more noticeable (e.g., obtaining datapoints for angular interactions with infrequent bead combinations like A-E-A).

Out of the 287 required potentials, 244 passed the validation checks in Section 3.8. Notably, these accepted potentials mainly involved the most frequent bead types, despite some reference systems incorporating single-molecule specialized mixtures. The remaining 43 interaction types, predominantly rare and intermolecular (e.g., A-E non-bonded interaction), necessitated more tailored reference systems. Of these, 37 were validated after employing multiple molecule mixtures, leaving six unresolved. The final six interaction types, crucial yet rare, particularly in bead types that correspond to less than 0.1 % of the total number of beads, could not generate acceptable potentials through conventional means. Consequently, pseudomolecules were used as reference systems to generate higher quality data points for these interaction types.

Folder */1_cg_force_field* in the SI encompasses all 287 tabulated potentials generated for the CG force field of this study, prepared for integration with LAMMPS. Additionally, Folder */2_sample_cg_sim* in the SI comprises the essential files for conducting a sample CG simulation, which includes an NPT simulation lasting 50 picoseconds of a mixture containing 1024 phenolic asphaltene molecules, depicted in Table 1.

4.3. Spatial fidelity

All the reference systems tested successfully passed validation Check 4, which compared the geometries of specific interaction types in CG simulations with their all-atom equivalents. However, the confidence intervals varied across these interactions. Angular and bonded

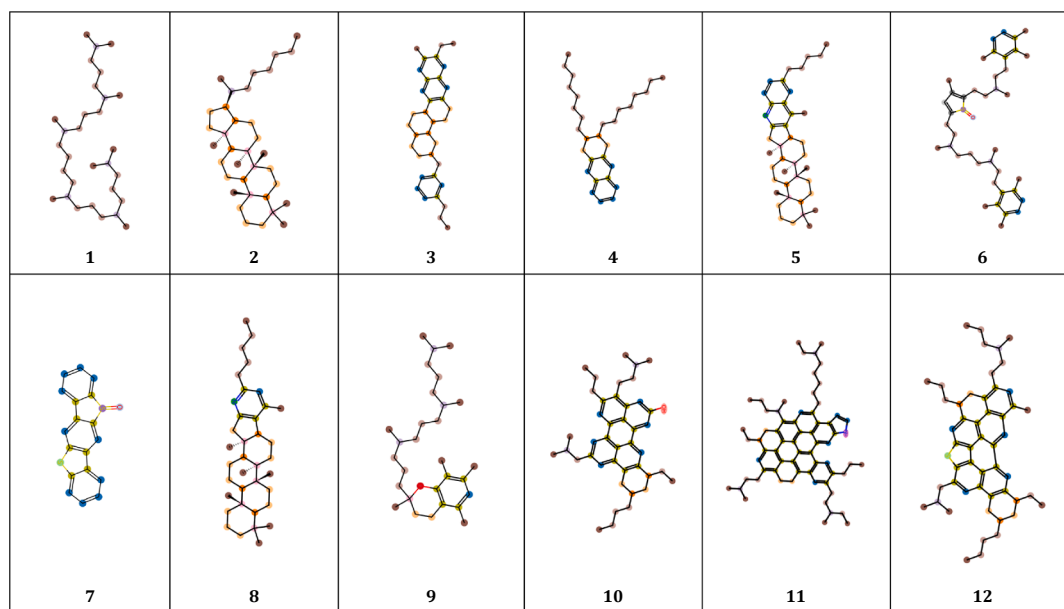
Table 4

Depiction of the 17 distinct bead types that constitute the newly formulated CG force field. These bead types are categorized according to six attributes, including the elemental identity of the parent atom they represent, the number of attached hydrogen atoms, hybridization, bond degree, the presence in rings, and inclusion within aromatic systems.

Type	Color	Parent atom	H bond	Hybridization	Degree	In Ring?	Aromatic?	Description
Q		C	0	SP2	3	Yes	Yes	Aromatic quaternary carbon
O		C	0	SP2	3	Yes	No	Thiophenic quaternary carbon
N		C	0	SP3	4	Yes	No	Quaternary carbon
A		C	1	SP2	2	Yes	Yes	Aromatic tertiary carbon
P		C	1	SP2	2	Yes	No	Thiophenic tertiary carbon
C		C	1	SP3	3	Yes	No	Cycloalkane tertiary carbon
J		C	1	SP3	3	No	No	Tertiary carbon
D		C	2	SP3	2	Yes	No	Cycloalkane secondary carbon
L		C	2	SP3	2	No	No	Secondary carbon
K		C	3	SP3	1	No	No	Primary carbon
E		N	0	SP2	2	Yes	Yes	Pyridinic nitrogen
M		N	1	SP2	2	Yes	Yes	Amine nitrogen
B		O	0	SP2	1	No	No	Sulfoxide oxygen
G		O	0	SP2	2	Yes	No	Oxane oxygen
H		O	1	SP2	1	No	No	Phenolic oxygen
F		S	0	SP2	2	Yes	Yes	Thiophenic sulfur
I		S	0	SP3	3	Yes	No	Sulfoxide sulfur

Table 5

Molecules from Table 1 depicted in their UA-CG representation, featuring all bead types (and corresponding colors) as listed in Table 4.



interaction types consistently achieved a 95 % confidence interval. This holds true even for combinations involving less common bead types. The RDFs for non-bonded pairwise interactions align within 90 % confidence intervals compared to their all-atom counterparts, with most incongruences occurring on infrequently occurring beads type combinations. Table 6 in the manuscript illustrates the RDFs for the most common, moderately common, and rare bead combination types, providing examples for encompassing non-bonded pairwise, bonded, and angular interactions.

4.4. Thermodynamics and performance benchmark

Table 7 presents the thermodynamic properties obtained from real bituminous samples using the all-atom and the CG force field. These properties are derived from the reference systems that were used in the development of the CG force field parameters. Conversely, Table 8 outlines the thermodynamic properties derived from the mixtures that include molecules not initially used as references (as elaborated in Section 3.9 and Table 2), in addition to those obtained from their all-atom analogs. For clarity, Table 8 restricts its focus to three molecule types (A, D, and E, for small, regular, and very large molecules respectively). The comprehensive version of this table is available in the SI, under Table S8. Additionally, both Table 7 and Table 8 report the benchmarking values of t_{net} and ρ_{Weiner} , which were found to differ across the tested systems. Both tables display mean values accompanied by their respective 95% confidence intervals, which were derived from the second NPT routine described in Section 3.4.

4.5. Microstructural features and phase separation

Table 9 presents the count of steps necessary to achieve equilibrium states, encompassing potential energy, neighbor composition, the ratio of particles in the bulk to interface, and the number of asphaltene clusters. It details the initial and final values of these metrics for all six mixtures in their all-atom configurations, as described in Section 3.12. In contrast, Table 10 provides these metrics for the CG models. The simulations reveal significant molecular rearrangement within the systems, leading to the formation of phase-rich regions – reflected by consistently higher values of $P_{X|X}$ in Table 10. Additionally, the increase in P_{bulk}

values indicate the aggregation of phases into larger bulk regions with comparatively smaller interface areas. This molecular reorganization results in a decrease in the systems' overall potential energy, suggesting a shift towards a thermodynamically more stable structure. The similarity in all-atom and CG metric values, both at the start and end of the simulations, indicates that the morphological characteristics of both systems are comparable. Fig. 4 visually depicts the rearrangement of phases in the simulation involving Mixture 4 (composed of a ternary mixture of saturates, resins, and asphaltenes), illustrating the dynamic evolution towards more heterogeneous, but more stable morphologies. Reaching highly heterogeneous morphologies in all-atom models takes a considerable number of steps, reaching well into microsecond durations, making them computationally intractable for regular use. To match results with CG models, all-atom simulations need about 60 times more steps on average.

5. Discussion

This section analyzes the previously presented results of the CG force field. Following its comprehensive validation, the emphasis is on identifying the force field's strengths and weaknesses. The analysis covers the CG force field's stability, its efficacy in replicating spatial conformations, its ability to reproduce thermodynamic and large-scale properties, and concludes with a discussion on its performance benefits.

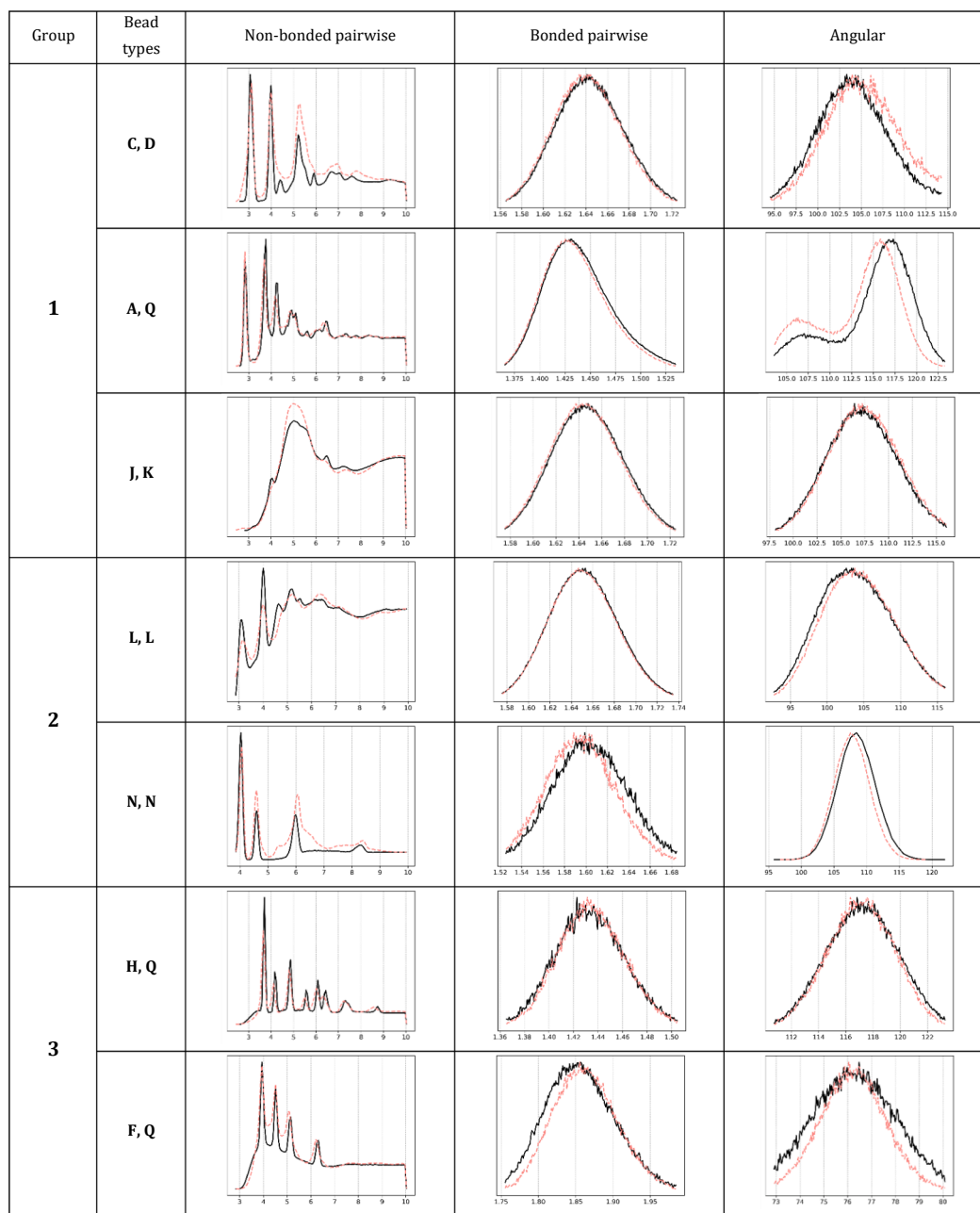
5.1. Bead selection and parametrization

The majority of the force field interactions met the accuracy criteria set out in the validation checks described in Section 3.8, and notably, less than 10 % of the tabulated potentials necessitated specialized postprocessing for data point quality enhancement. This indicates a high level of initial accuracy in the tabulated potentials. This also indicates that the custom all-atom systems created for this study prove useful in producing statistically relevant data points for the CG force field.

As expected, bead types that occur less frequently generally yielded lower accuracy scores. This necessitated the use of several reference systems to generate robust data sets for these bead types. Interestingly, some of the more common bead types also required meticulous tuning. Notable among these are bead types J, K, and L, which represent

Table 6

Radial distribution functions for bead pairings in a bitumen model akin to Table 1 and Table 2. Interactions are divided into three categories: Group (1) (most common), Group (2) (moderately common), and Group (3) (rarely common). Black graphs denote all-atom baseline interactions, while red graphs show CG interactions. Values on the y-axis are normalized.

**Table 7**

Thermodynamic properties of the three bitumen types produced by the CG force field compared to those obtained from the all-atom analogs. These are part of the reference systems used to generate the tabulated potentials of the force field.

Bitumen type	All-atom				CG				Benchmark	
	E[kcal/mol]	T[K]	P[atm]	ρ [kg/m ³]	E[kcal/mol]	T[K]	P[atm]	ρ [kg/m ³]	t_{net} [s]	ρ_{Weiner} [s]
1 TotalEnergies	47850 ±	298.150 ±	12.7 ±	1.012 ±	-850	298.150 ±	0.988 ±	1.020 ±	82.3	0.914 ±
	98.5	0.038	121	0.023	±22.5	0.028	0.104	0.010		
2 Norbit	49522 ±	298.151 ±	-11 ±	1.021 ±	-788	298.154 ±	1.092 ±	1.031 ±	79.9	0.927 ±
	78.6	0.045	98.2	0.013	±52.6	0.009	0.113	0.009		
3 FHWA-1	51466 ±	298.149 ±	-9.7 ±	1.019 ±	-514	298.148 ±	0.999 ±	1.017 ±	81.1	0.919 ±
	211	0.012	78.1	0.017	±25.2	0.021	0.371	0.028		

Table 8

Thermodynamic properties derived from the CG force field in comparison to their counterparts obtained from the corresponding all-atom models. Specifically, it focuses on Molecules A, D, and E. For a comprehensive overview, including additional molecules, refer to Table S8 in the SI.

Molecule ID		All-atom				CG				Benchmark	
		E[kcal/mol]	T[K]	P[atm]	ρ [kg/m ³]	E[kcal/mol]	T[K]	P[atm]	ρ [kg/m ³]	t_{net} [:]	ρ_{Weiner} [:]
1	A	-1223 ± 541	298.152 ± 0.112	12.51 ± 35.25	0.751 ± 0.004	-1001 ± 15	298.164 ± 0.674	0.985 ± 0.174	0.899 ± 0.086	25.1	0.754 ± 0.004
	D	-1342 ± 856	298.148 ± 0.007	5.65 ± 21.11	0.803 ± 0.003	-1442 ± 29	298.160 ± 0.132	1.101 ± 0.178	0.804 ± 0.005	100.9	0.699 ± 0.029
	E	-3045 ± 421	298.150 ± 0.004	4.99 ± 19.84	0.827 ± 0.004	-2817 ± 30	298.274 ± 0.287	0.922 ± 0.452	0.824 ± 0.003	110.8	0.655 ± 0.052
2	A	-1085 ± 197	298.153 ± 0.039	0.963 ± 36.12	0.893 ± 0.005	-1020 ± 97	298.111 ± 0.342	0.903 ± 0.125	0.698 ± 0.054	34.2	0.894 ± 0.006
	D	-977 ± 215	298.147 ± 0.001	0.988 ± 9.95	0.913 ± 0.009	-977 ± 215	298.230 ± 0.008	1.002 ± 0.036	0.910 ± 0.008	87.9	0.810 ± 0.026
	E	189 ± 117	298.740 ± 0.045	0.393 ± 77.95	0.951 ± 0.003	120.1 ± 115	297.934 ± 1.325	0.393 ± 4.737	0.949 ± 0.009	100.6	0.784 ± 0.029
3	A	-668 ± 185	298.149 ± 0.025	1.000 ± 49.32	0.909 ± 0.004	-543.9 ± 40	298.054 ± 0.071	1.009 ± 0.178	0.799 ± 0.071	19.8	0.912 ± 0.007
	D	298 ± 125	298.150 ± 0.987	0.12 ± 15.54	0.916 ± 0.003	237.6 ± 42	298.102 ± 2.107	0.520 ± 0.559	0.915 ± 0.008	87.0	0.840 ± 0.021
	E	-1399 ± 69	298.148 ± 0.087	1.54 ± 16.87	1.084 ± 0.004	-1465 ± 11	298.35 ± 0.251	0.742 ± 0.617	1.090 ± 0.004	99.3	0.741 ± 0.028
4	A	-1255 ± 99	298.158 ± 0.004	1.10 ± 18.55	1.410 ± 0.005	-1075 ± 80.1	298.177 ± 0.040	0.926 ± 0.135	0.689 ± 0.046	27.4	1.397 ± 0.004
	D	-1355 ± 442	298.158 ± 0.077	1.11 ± 20.33	1.390 ± 0.046	-1296 ± 15.0	298.141 ± 0.020	0.392 ± 5.016	1.382 ± 0.030	87.5	1.010 ± 0.019
	E	-1785 ± 336	298.152 ± 0.036	1.23 ± 27.65	1.240 ± 0.004	-1486 ± 23.8	298.152 ± 0.091	1.005 ± 0.113	1.22 ± 0.005	90.1	0.874 ± 0.024
5	A	-1223 ± 88	298.141 ± 0.006	1.16 ± 21.55	1.410 ± 0.004	-91.2 ± 85.6	298.154 ± 0.043	0.955 ± 0.156	0.980 ± 0.076	27.4	1.398 ± 0.005
	D	-1477 ± 551	298.151 ± 0.087	1.15 ± 28.35	1.390 ± 0.058	-1496 ± 16.0	298.149 ± 0.022	0.477 ± 0.216	1.392 ± 0.029	87.5	1.017 ± 0.021
	E	-2101 ± 399	298.149 ± 0.036	1.36 ± 24.76	1.240 ± 0.005	-1577 ± 34.8	298.151 ± 0.082	1.014 ± 0.145	1.254 ± 0.004	90.1	0.874 ± 0.027
6	A	-999.8 ± 115	298.150 ± 0.002	0.944 ± 18.66	0.910 ± 0.004	-1254 ± 12.5	298.15 ± 0.045	1.045 ± 0.185	0.814 ± 0.045	24.1	0.911 ± 0.007
	D	-4215 ± 1024	298.157 ± 0.078	1.215 ± 51.44	1.197 ± 0.057	-4576 ± 190	298.150 ± 0.07	1.065 ± 0.081	1.201 ± 0.030	78.5	0.945 ± 0.021
	E	-3895 ± 215	298.152 ± 0.010	1.050 ± 33.98	1.047 ± 0.006	-3903 ± 52.9	298.143 ± 0.026	1.048 ± 0.144	1.048 ± 0.004	99.6	0.884 ± 0.031
7	A	-311 ± 546	298.147 ± 1.987	-2.51 ± 45.62	1.181 ± 0.009	-311.8 ± 75.8	298.146 ± 2.816	0.778 ± 4.894	0.879 ± 0.065	19.8	1.180 ± 0.005
	D	-1281 ± 178	298.152 ± 0.007	1.00 ± 28.88	1.049 ± 0.003	-1281 ± 12.9	298.147 ± 0.014	1.021 ± 0.044	1.048 ± 0.011	78.4	0.945 ± 0.017
	E	-1984 ± 984	298.151 ± 0.002	2.900 ± 7.44	1.082 ± 0.006	-1984 ± 95.7	298.149 ± 0.005	1.005 ± 0.011	1.082 ± 0.006	99.9	0.887 ± 0.031
8	A	-2366 ± 588	298.745 ± 0.171	0.861 ± 35.65	1.149 ± 0.003	-2297 ± 39.7	298.052 ± 0.276	0.845 ± 7.405	0.985 ± 0.025	27.8	1.145 ± 0.002
	D	-573 ± 788	298.151 ± 0.002	1.010 ± 4.223	1.104 ± 0.006	-460.8 ± 30.3	298.151 ± 0.028	1.021 ± 0.095	1.099 ± 0.009	68.5	1.001 ± 0.032
	E	1699 ± 1522	298.151 ± 0.111	0.49 ± 89.55	1.237 ± 0.003	1623 ± 45.6	298.051 ± 2.764	0.798 ± 7.418	1.234 ± 0.004	70.2	0.886 ± 0.034
9	A	855 ± 560	298.156 ± 0.001	0.994 ± 31.25	1.187 ± 0.007	663.1 ± 123	298.130 ± 0.266	0.978 ± 0.676	1.005 ± 0.027	21.5	1.187 ± 0.009
	D	-1223 ± 541	298.152 ± 0.112	12.51 ± 35.25	1.100 ± 0.004	-1001 ± 15	298.164 ± 0.674	0.985 ± 0.174	1.103 ± 0.010	68.7	1.007 ± 0.023
	E	-1187 ± 542	298.156 ± 0.071	15.64 ± 24.51	1.282 ± 0.005	-1137 ± 25	298.207 ± 0.712	0.991 ± 0.283	1.281 ± 0.005	89.5	0.998 ± 0.035

Table 9

Simulation duration, potential energy, neighbor composition, interface to bulk proportion, and number of asphaltene clusters metrics at the start and end of the simulations for all-atom mixtures, as detailed in Section 3.12.

Mix ID	Duration	All-atom Start	End													
			E _{PCFF} [kcal/mol]	P _{1 1}	P _{2 2}	P _{3 3}	P _{4 4}	P _{bulk}	N _C	E _{CG} [kcal/mol]	P _{1 1}	P _{2 2}	P _{3 3}	P _{4 4}	P _{bulk}	N _C
1	2880	58,475	51.0	-	-	-	49.9	51.2	109	112.6	65.7	-	-	65.1	65.8	107
2	2490	57,451	-	50.7	-	-	49.9	51.4	108	-422.5	-	63.0	-	62.4	62.0	100
3	3549	54,556	-	-	50.9	49.7	53.6	104	-648.5	-	-	65.2	64.4	64.2	99	
4	3827	52,097	33.3	-	33.3	33.3	34.5	70	-888.9	49.1	-	50.2	56.7	71.1	67	
5	4855	51,112	26.1	25.4	25.9	24.9	24.9	54	-1255.6	48.5	47.2	44.1	51.6	70.1	52	
6	5000	49,785	25.4	25.9	25.6	24.9	24.8	54	-1388.1	41.5	42.2	44.1	45.6	70.0	53	

Table 10

Simulation duration, potential energy, neighbor composition, interface to bulk proportion, and number of asphaltene clusters metrics at the start and end of the simulations for CG mixtures, as detailed in Section 3.12.

Mix ID	Duration Steps[1e6]	CG Start							End						
		E_{PCFF} [kcal/mol]	$P_{1 1}$	$P_{2 2}$	$P_{3 3}$	$P_{4 4}$	P_{bulk}	N_C	E_{CG} [kcal/mol]	$P_{1 1}$	$P_{2 2}$	$P_{3 3}$	$P_{4 4}$	P_{bulk}	N_C
1	48.0	56,486	54.5	–	–	50.4	52.4	1	–70.135	66.2	–	–	66.0	65.2	1
2	49.8	54,503	–	54.3	–	51.3	52.2	1	–535.66	–	64.0	–	63.4	63.6	1
3	50.7	52,466	–	–	53.1	52.8	55.6	1	–788.94	–	–	66.7	64.5	63.6	1
4	59.8	49,884	35.7	–	35.4	34.4	36.3	1	–1023.55	50.5	–	51.1	57.6	72.5	1
5	77.5	49,746	27.0	26.1	27.6	27.6	26.0	3	–1415.8	49.4	49.2	46.9	52.4	71.2	3
6	80.1	47,998	26.8	27.4	27.9	27.9	26.1	4	–1599.7	42.3	43.8	45.9	46.7	71.0	4

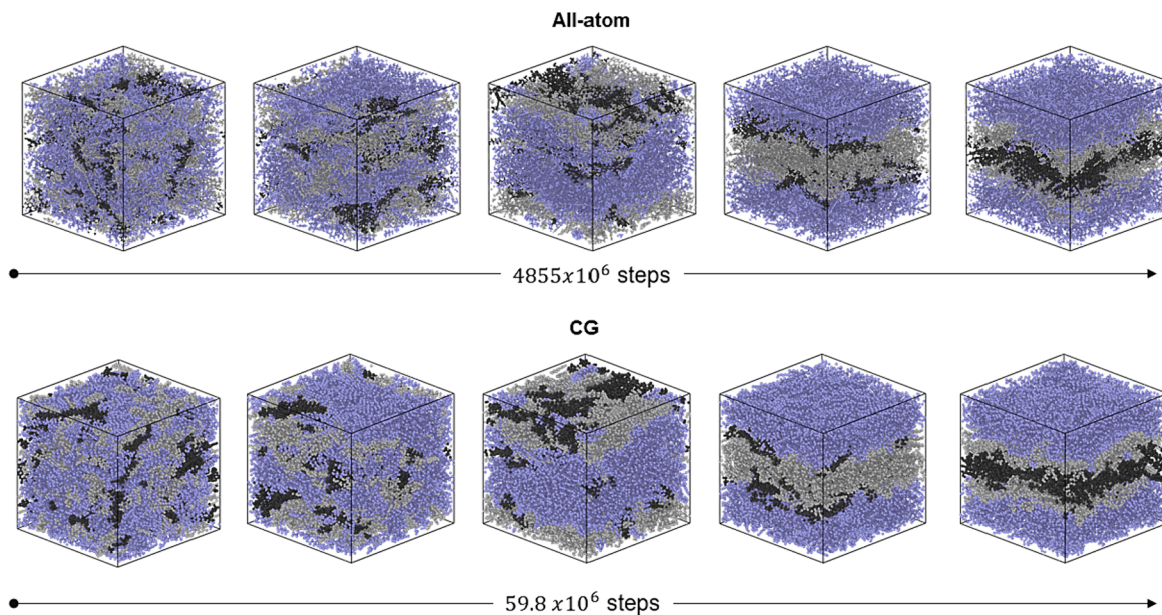


Fig. 4. Mixture 4 undergoing phase separation using both the all-atom and the CG force field. The saturates fraction of this ternary mixture is represented in purple, the resins in grey, and the asphaltenes in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

aliphatic beads, as well as types C and D, associated with aromatic rings. This raised a particular case in the context of asphaltenes and resins. These beads did not accurately represent the same types in smaller molecules, and this incongruity was evident in both directions. This reflects the emphasis on ensuring the CG force field's effectiveness in capturing the complex behaviors of larger, structurally complex hydrocarbons, such as asphaltenes and resins, which are often complex and of greater interest in bituminous materials research.

The simulations employing the CG force field demonstrated numerical stability in every interaction type and combination, demonstrated by the successful completion of Validation checks 1, 2, and 3 in all the systems. However, it was observed that these simulations required a longer duration to reach equilibrium compared to those using all-atom models. Although CG simulations achieve equilibrium in fewer steps due to the smoother potential energy surfaces, considering the total effective time (t_{net}), they necessitate approximately five times longer to reach equilibrium. This prolongation in equilibration time can be attributed to multiple factors. The MD algorithms used are generally optimized for scenarios involving all-atom models, suggesting they may not be fully adapted to the distinct dynamics encountered in CG simulations. Additionally, the current limited understanding of CG-specific dynamics exacerbates this issue, as it impedes accurate predictions and efficient management of the simulations' progression towards equilibrium.

5.2. Spatial fidelity and thermodynamics

Table 6 presents favorable outcomes in evaluating the CG force field's ability to simulate mixtures of 12-molecule bituminous mixtures. However, it is of greater significance to analyze the data from Table 7. In this context, the distinct advantages and limitations of the CG force field become more apparent, particularly in systems where a singular chemical attribute predominates over others. For instance, the CG force field exhibits limitations in accurately modeling molecules with molecular masses below 150 g/mol. This discrepancy is particularly evident in Type A and B molecules of Table 2, where the accuracy observed in larger molecules is not achieved. Such deviations were anticipated, given the force field's calibration against reference molecules of considerably higher molecular masses and shapes. This limitation is manifest not only in unrealistic thermodynamic properties but also in diminished geometric fidelity compared to their all-atom counterparts. The divergence is especially pronounced in small alkanes and molecules containing only one ring, such as dodecane or benzene, suggesting that the CG force field is less suitable for modeling such molecular structures.

Furthermore, the CG force field encounters difficulties in precisely modeling some molecules categorized under C and D, specifically those that are highly functionalized (e.g., molecules 7C, 8C, 9C, 5D, and 4D). This difficulty aligns with the instability observed in their all-atom model equivalents, likely due to the concentrated aggregation of highly polar groups. Such clustering may lead to instabilities in

conformation, energy, and forces within these molecules, particularly since they are artificial representations of compounds that might not be present in real bituminous samples. Moreover, highly polar compounds are known to often necessitate custom force field parameters, both in all-atom and in CG modeling techniques. Consequently, this suggests a potential requirement for the development of new force field parameters to accurately represent highly polar molecules, albeit a scenario not typically encountered in bituminous materials.

The performance of the CG force field also exhibits certain limitations when applied to small, highly planar, functionalized hydrocarbons. This is particularly evident in molecules akin to pseudo-asphaltenes or pseudo-resins (like in the case of small benzobisbenzothiophene derivatives), which often exhibit densities above 1.3 g/cm^3 . Fortunately, bituminous molecules tend to be larger and have branching structures, which the force field accurately represents.

Table 8 illustrates that Weiner's UA force field exhibits optimal performance in simulating molecules of types A, B, and C, which are situated at the smaller end of the molecular spectrum examined in this research. This observation is supported by the precise representation of system densities, which show remarkable correlation with the densities obtained from all-atom models. Conversely, for molecules of types D and E, which are the most prevalent in MD studies of bitumen, Weiner's force field demonstrates diminished applicability. This limitation is evident from the observed densities in these systems, which are slightly lower than those predicted by all-atom models, particularly for larger molecules of type E. The force field's failure to replicate fully condensed densities suggests potential discrepancies in molecular conformation, energy, and force interactions. Despite densities being relatively high (approximately 0.7 g/cm^3), the presence of additional space may lead to inaccuracies in kinetic behaviors and compromise the understanding of phase separation phenomena, which is heavily dependent on close range interactions, thus necessitating accurate high-density measurements. In contrast, the CG force field developed in this study exhibits an inverse trend. For smaller molecules of types A, B, and C, particularly type A, it struggles to match the density values of all-atom models. However, as molecular size and complexity increase, the CG force field achieves density values closely aligned with those of all-atom models. This highlights the capability of the newly developed force field to accurately simulate the behavior of complex bituminous molecules, particularly of the resinous or asphaltenic varieties comprised of high molecular masses.

Consequently, the optimal applicability of the CG force field is observed with molecules that closely resemble those listed in Table 1. This encompasses molecules with a substantial molecular mass range of 500–2000 g/mol. The structure of these molecules typically features a combination of multiple aromatic and non-aromatic rings, with at least two to three of each category being present. Additionally, a certain level of branching within their structures contributes to their compatibility with the force field. These molecules usually display densities within the range of 0.9 to 1.1 g/cm^3 , a parameter within which the CG force field demonstrates its strengths in accurate modeling and simulation.

Despite certain limitations in accurately capturing the properties and geometries of some molecules, such as small alkanes or small cyclic compounds (e.g., benzene, toluene, or cyclopentane) the trends outlined in Section 3.8 remain consistent. It is observed that several factors influence the density of the mixture: increasing molecular weight, introducing branching, adding polar functional groups, and enhancing planarity through the addition of aromatic rings. This aspect is crucial, as the CG force field is primarily intended for capturing general trends in organic chemistry. While there may be deviations in the average properties of certain molecules, the overall trends align well with established chemical principles.

For instance, in the case of Group (1) molecules (alkanes), there is an observed discrepancy where the CG model predicts densities approximately 10 to 15 % higher than their all-atom counterparts. However, as we progress from types A through E, representing a transition to larger

and more complex alkanes, the increment in density mirrors the trends observed in the all-atom models. This consistency in trend replication is a significant aspect of the CG model's utility.

Another notable trend is the impact of branching on molecular density. The introduction of branches typically results in a decrease in density. This effect is particularly evident when comparing molecules with no branches (type B) to those with branching (types C or D) or molecules of type 1. The CG force field accurately captures this trend, underscoring its effectiveness in modeling dynamic changes in molecular structure and their resultant impact on physical properties.

In the realm of all-atom models, it is observed that the standard deviations for certain properties, notably potential energy and pressure, are significantly large, often surpassing the magnitude of the average values themselves. This phenomenon is not unexpected, given the complex potential energy surfaces employed in these all-atom models. These surfaces account for highly detailed interactions (see E_{PCFF} in Equation (4) of Section 3.4), such as highly vibratory C-H bonds, which can induce considerable fluctuations in energies, forces, and consequently, the stress tensors and pressures within the simulation box.

Conversely, the CG force field demonstrates a marked advantage in this context. By simplifying the potential energy surfaces—achieved through the elimination of highly vibrational interactions and the aggregation of several atoms into beads—this approach results in smoother and more uniform particle motion. Consequently, the standard deviations in properties like potential energy and pressure are substantially reduced, often by an average factor of 10 compared to all-atom models. This attribute of the CG force field is particularly beneficial for researchers focusing on properties that depend heavily on the stability of energies, forces, and stress tensors. Such properties include viscoelastic-dependent characteristics like viscosity or complex moduli, where the minimized fluctuations in the CG model can provide more reliable and consistent data over a wide range of stress regimes. This is particularly relevant in the study of bitumens, where the driving forces behind observed phenomena are frequently subtle and may not be readily detectable in short, all-atom simulations that are known to produce noisy stress tensor-related properties.

5.3. Phase separation

The proficiency of the CG force field in faithfully mirroring phenomena inherent to bituminous materials stands out. It accurately simulates the clustering of asphaltene molecules and the phase differentiation among molecules of varying chemical properties, such as those in non-polar and polar phases. This functionality is crucial, especially for its proposed use in simulations aimed at exploring these phenomena within bituminous materials.

In the study, all six mixtures presented in Table 3 exhibited at least some level of phase reorganization from their initial uniform distribution, as observed in both all-atom and CG simulations. The initial discrepancy from an initially even phase distribution (for instance, a $P_{X|X}$ or $P_{X|Y}$ equal to 50 in a binary system) may result from the model construction stage or the operation of neighbor-detection algorithms discussed in Section 12.

Throughout the simulations, there was a consistent increase in the values of $P_{X|X}$ and P_{bulk} , pointing to the development of phase-rich regions within the systems. The observed reduction in the number of asphaltene clusters over the duration of the simulation aligns with the operational principles of the cluster detection algorithm, which tracks the coalescence of small, molecule-sized clusters into fewer, larger aggregates. For Mixtures 1 through 4, the culmination of the simulation process is characterized by the presence of a singular asphaltene cluster, suggesting the formation of a large, unified agglomerate with minimal likelihood of subsequent rearrangement. Conversely, in the case of simulations associated with Mixture 5 and 6, the cluster counts were observed to be 2 and 4, respectively. This variation hints at the potential

for future structural rearrangements within these mixtures. Such a differentiation in behavior can be attributed to the quaternary nature of Mixtures 5 and 6, alongside a lesser variance in molecular polarity compared to the other, more simplistic mixtures under investigation.

The magnitude of these values was comparable between all-atom and CG models, suggesting that the morphologies derived from both modeling approaches are similar and can be compared. As expected, phase separation occurred more rapidly in simpler mixtures, characterized by smaller molecules with greater differences in polarity. In the case of Mixture 6, a quaternary mixture with minor polarity differences, equilibrium was not achieved within the number of simulation steps run. Yet, the emergence of a more heterogeneous morphology was evident and thus sufficient to demonstrate the CG model's capability to replicate phase segregation akin to that observed in the all-atom model.

A critical observation was made regarding the role of t_{net} in assessing the progression of these simulations across both all-atom and CG models. Given an equivalent number of steps, CG models exhibited faster phase separation, attributed to the time-related scalability differences of the CG models. To achieve similar morphological states, all-atom simulations required, on average, 60 times more steps than CG models, aligning with the t_{net} values presented in Table 8 on the lower end. The lower value can be explained by other factors not accounted for when computing t_{net} , such as the difference in particle count or force field formulation complexity.

5.4. Performance

The comparison of simulation time efficiency between CG models and all-atom models reveals significant variances, with CG simulations typically completing 5 to 50 times faster. However, providing a precise efficiency ratio is challenging due to several factors influencing the speed of LAMMPS simulations. These factors include the number of particles, the diversity of particle types, the complexity of interactions, the average number of neighboring particles for each modeled particle, and the efficiency of the algorithms to compute neighbor lists.

5.4.1. Marking 1: Time scale without τ_{net}

The benchmarks conducted on condensed models, with densities ranging from 0.9 to 1.2 g/cm³ under NVE conditions and using identical integration steps and elapsed times, underscored the impact of particle count, bead types, and interaction complexities on the speed advantages of CG models. When considering solely the number of particles—assuming constant complexities and bead types, which are known to affect both modeling approaches linearly—the CG technique shows its greatest advantages in larger systems, particularly those with over 25,000 particles. In systems where the scale factor is only x1 or x2, the time gains are approximately 5 times faster. However, in very large systems (scale factors of x8 and x16), especially those with a singular molecule type and a limited number of bead types (four or fewer), the efficiency gains are considerably more pronounced, reaching up to 50 times faster.

5.4.2. Marking 2: Time scale with τ_{net}

When evaluating the impact of τ_{net} , from using CG simulations as opposed to all-atom models, it is observed that the gains typically range from 5 to 50 times faster. This efficiency gain is more pronounced when simulating smaller, less branched molecules compared to larger, more complex hydrocarbons like resinous or asphaltenic molecules. Smaller molecules, particularly those without significant steric hindrance constraints such as interlocking, have greater freedom of movement, leading to higher net displacement. This increased mobility generally translates to a more pronounced efficiency gain in simulations using the CG model. In contrast, larger molecules that resemble resins or asphaltenes exhibit behaviors such as interlocking and stacking into stable aggregates. These behaviors, while reducing the general displacement on a per-molecule basis, can result in significant net displacement of the

entire molecular conglomerate.

Interestingly, despite the lower per-molecule MSDs observed in larger molecules—which might suggest lower efficiency gains—the complexity of these large molecules is also mirrored in their all-atom counterparts. Consequently, the simulation of complex, branched, and aromatic molecules yield higher τ_{net} gains when using the CG approach. This is due to the significant computational demands of accurately modeling the atomic structures and interactions in all-atom simulations, demands that are substantially reduced in CG models while still capturing the essential dynamics and interactions of these complex molecules. Therefore, a simulation performed using the CG force field could theoretically run 1000–2000 times faster, but this is seldom the case. The overall compounded gain is often in the range of 50–200 times faster than all-atom simulations, particularly for systems involving bituminous molecules structures and numbers akin to those in Table 1. τ_{net} varies across each molecular system, as evidenced by the data in Table 6 and Table 7. Comparing CG to all-atom systems is essential to understand both the physical characteristics and performance enhancements for every specific system studied.

6. Conclusion

In this research, we introduced a United Atom Force-Matched force field tailored for coarse-grained modeling of bituminous molecules, effectively incorporating explicit hydrogen atoms into their bonded parent atoms. This method substantially decreases computational requirements, halving the number of particles involved, eliminating computationally expensive hydrogen interactions, and decreasing degrees of freedom in the motion of the particles. As a direct consequence, the performance of MD simulations of bitumen is enhanced by an average factor of 100, enabling the exploration of microscale phenomena that were previously beyond reach due to computational constraints. Specifically, this includes the possibility to study phase separation, aggregation, and crystallization within bitumen's SARA fractions, areas where traditional all-atom models faltered.

The force field's accuracy is achieved through meticulous force mapping from all-atom to CG models, utilizing the OpenMSCG package developed by Voth et Al. This ensures that the CG force field replicates the molecular energies, forces, and conformations with high fidelity. It mirrors the geometric and thermodynamic properties observed in bitumen closely, including pressure and density values. The CG force field includes 17 bead types, distinguished by six criteria, and 287 interaction types, enabling it to simulate the complex makeup of bituminous materials effectively, particularly compounds prevalent in asphaltenes and resins, characterized by high molecular mass (above 500 g/mol), high aromaticity, planarity, and densities between 0.95 and 1.1 g/cm³.

The comparative advantage of this force field lies in its specialization for large hydrocarbon molecules, displaying superior numerical stability, spatial fidelity, and accurate representation of thermodynamic and kinetics-related properties for molecules within the specified structural characteristics. While it exhibits limitations in modeling smaller or highly polar molecules, it outperforms other United Atom force fields in replicating the properties of larger hydrocarbon compounds. The force field is also proven capable of closely replicating phase behavior and morphological developments in multi-phase hydrocarbon mixtures, akin to those observed in all-atom counterparts, thereby promising a valuable tool for investigating larger spatiotemporal phenomena in bituminous materials, essential to realistically capture its mechanical and rheological responses.

7. Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study. However, the [Supplementary Information](#) includes the necessary

information to reproduce most of the simulations performed in this study (all-atom and CG).

8. Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used OpenAI's ChatGPT4 to shorten the length of certain sections. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRedit authorship contribution statement

Eli I. Assaf: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xueyan Liu:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization. **Peng Lin:** Writing – review & editing, Validation, Supervision, Conceptualization. **Sandra Erkens:** Validation, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We developed a force field for use in LAMMPS. We will make the force field available upon publication of our manuscript.

Acknowledgments

This paper/article is created under the research program Knowledge-based Pavement Engineering (KPE). KPE is a cooperation between the Ministry of Infrastructure and Water Management (Rijkswaterstaat), TNO, and TU Delft in which scientific and applied knowledge is gained about asphalt pavements and which contributes to the aim of Rijkswaterstaat to be completely climate neutral and to work according to the circular principle by 2030. The opinions expressed in these papers are solely from the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matdes.2024.112831>.

References

- C. Giavarini, D. Mastrofini, M. Scarsella, L. Barré, D. Espinat, Macrostructure and rheological properties of chemically modified residues and bitumens, *Energy Fuel* 14 (2) (2000) 495–502.
- S. Riniker, J.R. Allison, W.F. van Gunsteren, On developing coarse-grained models for biomolecular simulation: a review, *PCCP* 14 (36) (2012) 12423–12430.
- J.S. Hansen, C.A. Lemarchand, E. Nielsen, J.C. Dyre, T. Schröder, Four-component united-atom model of bitumen, *J. Chem. Phys.* 138 (9) (2013).
- S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, *J. Am. Chem. Soc.* 106 (3) (1984) 765–784.
- J. Read, D. Whiteoak, The shell bitumen handbook, Thomas Telford (2003).
- D. Lesueur, The colloidal structure of bitumen: consequences on the rheology and on the mechanisms of bitumen modification, *Adv. Colloid Interface Sci.* 145 (1–2) (2009) 42–82.
- L.W. Corbett, Composition of asphalt based on generic fractionation, using solvent deasphalting, elution-adsorption chromatography, and densimetric characterization, *Anal. Chem.* 41 (4) (1969) 576–579.
- A. D-01, Standard test methods for separation of asphalt into four fractions, ASTM International (2001).
- A. Ecker, The application of iastrocan-technique for analysis of bitumen, *Pet. Coal* 43 (1) (2001) 51–53.
- A. Rosinger, Beiträge zur Kolloidchemie des Asphalts, *Kolloid-Zeitschrift* 15 (5) (1914) 177–179.
- F. Nellensteyn, The constitution of asphalt, *J. Inst. Pet. Technol* 10 (1924) 311–323.
- A.T. Pauli, Asphalt compatibility testing using the automated Heithaus titration test, *Preprints of Papers-American Chemical Society Division Fuel Chemistry* 41 (1996) 1276–1281.
- J.F. Masson, V. Leblond, J. Margeson, Bitumen morphologies by phase-detection atomic force microscopy, *J. Microsc.* 221 (1) (2006) 17–29.
- X. Qu, Q. Liu, M. Guo, D. Wang, M. Oeser, Study on the effect of aging on physical properties of asphalt binder from a microscale perspective, *Constr. Build. Mater.* 187 (2018) 718–729.
- D.D. Li, M.L. Greenfield, Chemical compositions of improved model asphalt systems for molecular simulations, *Fuel* 115 (2014) 347–356.
- D.D. Li, M.L. Greenfield, Viscosity, relaxation time, and dynamics within a model asphalt of larger molecules, *J. Chem. Phys.* 140 (3) (2014).
- T. Headen, E. Boek, G. Jackson, T. Totton, E. Müller, Simulation of asphaltene aggregation through molecular dynamics: insights and limitations, *Energy Fuel* 31 (2) (2017) 1108–1125.
- E.I. Assaf, X. Liu, P. Lin, S. Erkens, S. Nahar, L.I. Mensink, Studying the impact of phase behavior in the morphology of molecular dynamics models of bitumen, *Mater. Des.* 230 (2023) 111943.
- G. Li, M. Han, Y. Tan, A. Meng, J. Li, S. Li, Research on bitumen molecule aggregation based on coarse-grained molecular dynamics, *Constr. Build. Mater.* 263 (2020) 120933.
- J. Wong-Ekkabut, S. Baoukina, W. Triampo, I.-M. Tang, D.P. Tieleman, L. Monticelli, Computer simulation study of fullerene translocation through lipid membranes, *Nat. Nanotechnol.* 3 (6) (2008) 363–368.
- S.Y. Joshi, S.A. Deshmukh, A review of advancements in coarse-grained molecular dynamics simulations, *Mol. Simul.* 47 (10–11) (2021) 786–803.
- P.J. Bond, J. Holyoake, A. Ivetac, S. Khalid, M.S. Sansom, Coarse-grained molecular dynamics simulations of membrane proteins and peptides, *J. Struct. Biol.* 157 (3) (2007) 593–605.
- W.G. Noid, J.-W. Chu, G.S. Ayton, V. Krishna, S. Izvekov, G.A. Voth, A. Das, H. C. Andersen, The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models, *J. Chem. Phys.* 128 (24) (2008).
- M. Liu, G. Liu, L. Zhou, J. Chang, Dissipative particle dynamics (DPD): an overview and recent developments, *Arch. Comput. Meth. Eng.* 22 (2015) 529–556.
- S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. De Vries, The MARTINI force field: coarse grained model for biomolecular simulations, *J. Phys. Chem. B* 111 (27) (2007) 7812–7824.
- W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* 118 (45) (1996) 11225–11236.
- L. Yang, C.-H. Tan, M.-J. Hsieh, J. Wang, Y. Duan, P. Cieplak, J. Caldwell, P. A. Kollman, R. Luo, New-generation Amber united-atom force field, *J. Phys. Chem. B* 110 (26) (2006) 13166–13176.
- W.L. Jorgensen, J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *J. Am. Chem. Soc.* 110 (6) (1988) 1657–1666.
- N. Golmohammadi, M. Boland-Hemmat, S. Barahmand, H. Eslami, Coarse-grained molecular dynamics simulations of poly(ethylene terephthalate), *J. Chem. Phys.* 152 (11) (2020) 114901.
- G.C.Q. da Silva, G.M. Silva, F.W. Tavares, F.P. Fleming, B.A.C. Horta, Are all-atom any better than united-atom force fields for the description of liquid properties of alkanes? *Journal of Molecular Modeling* 26 (11) (2020) 296.
- J. Wu, D. Mukherji, Comparison of all atom and united atom models for thermal transport calculations of amorphous polyethylene, *Comput. Mater. Sci* 211 (2022) 111539.
- J.S. Hansen, C.A. Lemarchand, E. Nielsen, J.C. Dyre, T. Schröder, Four-component united-atom model of bitumen, *J. Chem. Phys.* 138 (9) (2013) 094508.
- Y. Wang, W.G. Noid, P. Liu, G.A. Voth, Effective force coarse-graining, *PCCP* 11 (12) (2009) 2002–2015.
- S. Ren, X. Liu, P. Lin, S. Erkens, Y. Xiao, Chemo-physical characterization and molecular dynamics simulation of long-term aging behaviors of bitumen, *Constr. Build. Mater.* 302 (2021) 124437.
- J. Jin, A.J. Pak, A.E. Durumeric, T.D. Loose, G.A. Voth, Bottom-up coarse-graining: principles and perspectives, *J. Chem. Theory Comput.* 18 (10) (2022) 5759–5791.
- D.C. Rapaport, The art of molecular dynamics simulation, Cambridge university press 2004.
- S.C. Glotzer, W. Paul, Molecular and mesoscale simulation methods for polymer materials, *Annu. Rev. Mat. Res.* 32 (1) (2002) 401–436.
- H. Sun, S.J. Mumby, J.R. Maple, A.T. Hagler, An ab initio CFF93 all-atom force field for Polycarbonates, *J. Am. Chem. Soc.* 116 (7) (1994) 2978–2987.
- S.J.M.p. Nosé, A molecular dynamics method for simulations in the canonical ensemble, *52(2)* (1984) 255–268.
- A. Guy, A.J.J.o.C. Bowling, N. Dynamics, Modification of Nosé–Hoover Thermostat to Improve Temperature Response in Molecular Simulations, *12(3)* (2017).
- F.-Y. Lin, A.D. MacKerell, Force fields for small molecules, *Biomolecular simulations: Methods and protocols* (2019) 21–54.
- R. Pastor, A. MacKerell Jr, Development of the CHARMM force field for lipids, *J. Phys. Chem. Lett.* 2 (13) (2011) 1526–1532.
- Y. Peng, A.J. Pak, A.E. Durumeric, P.G. Sahrman, S. Mani, J. Jin, T.D. Loose, J. Beiter, G.A. Voth, OpenMSCG: A Software tool for bottom-up Coarse-graining, *J. Phys. Chem. B* 127 (40) (2023) 8537–8550.

- [44] S.-I. Amari, Backpropagation and stochastic gradient descent method, *Neurocomputing* 5 (4–5) (1993) 185–196.
- [45] W. Noid, P. Liu, Y. Wang, J.-W. Chu, G.S. Ayton, S. Izvekov, H.C. Andersen, G. A. Voth, The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models, *J. Chem. Phys.* 128 (24) (2008).
- [46] F. Ercolessi, A molecular dynamics primer, spring college in computational physics, ICTP, Trieste 19 (1997).
- [47] H. Wang, C. Junghans, K. Kremer, Comparative atomistic and coarse-grained study of water: what do we lose by coarse-graining? *Eur. Phys. J. E* 28 (2009) 221–229.
- [48] H. Hernandez, Standard Maxwell-Boltzmann distribution: definition and properties, *ForsChem Res. Rep.* 2 (2017) 2017.
- [49] A.K. Rappé, C.J. Casewit, K. Colwell, W.A. Goddard III, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *J. Am. Chem. Soc.* 114 (25) (1992) 10024–10035.
- [50] M.G. Martin, J.I. Siepmann, Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes, *J. Phys. Chem. B* 102 (14) (1998) 2569–2577.
- [51] K.P. Santo, A.V. Neimark, Dissipative particle dynamics simulations in colloid and Interface science: a review, *Adv. Colloid Interface Sci.* 298 (2021) 102545.
- [52] E.S. Boek, D.S. Yakovlev, T.F. Headen, Quantitative molecular representation of asphaltenes and molecular dynamics simulation of their aggregation, *Energy Fuel* 23 (3) (2009) 1209–1219.
- [53] J. Murgich, Molecular simulation and the aggregation of the heavy fractions in crude oils, *Mol. Simul.* 29 (6–7) (2003) 451–461.
- [54] J. Tang, H. Wang, Coarse grained modeling of nanostructure and asphaltene aggregation in asphalt binder using dissipative particle dynamics, *Constr. Build. Mater.* 314 (2022) 125605.
- [55] L.M. da Costa, S.R. Stoyanov, S. Gusarov, X. Tan, M.R. Gray, J.M. Stryker, R. Tykwinski, J.W.de M. Carneiro, P.R. Seidl, A. Kovalenko, Density functional theory investigation of the contributions of π - π stacking and hydrogen-bonding interactions to the aggregation of model asphaltene compounds, *Energy Fuel* 26 (5) (2012) 2727–2735.