

Modeling of bulk kerogen porosity: Methods for control and characterization

Manolis Vasileiadis¹, Loukas D. Peristeras^{1,}, Konstantinos D. Papavasileiou¹, Ioannis G. Economou^{1,2,*}*

¹Molecular Thermodynamics and Modeling of Materials Laboratory, Institute of Nanoscience and Nanotechnology, National Center for Scientific Research “Demokritos”, GR-15310 Aghia Paraskevi Attikis, Greece

²Chemical Engineering Program, Texas A&M University at Qatar, P.O. Box 23874, Doha, Qatar

Abstract

Shale gas is an unconventional source of energy, which has attracted a lot of attention during the last years. Kerogen is a prime constituent of shale formations and plays a crucial role in shale gas technology. Significant experimental effort in the study of shales and kerogen has produced a broad diversity of experimentally determined structural and thermodynamic properties even for samples of the same well. Moreover, proposed methods reported in the literature for constructing realistic bulk kerogen configurations have not been thoroughly investigated. One of the most important characteristics of kerogens is their porosity due to its direct connection with their transport properties and its potential as discriminating and classifying metric between samples. In this study, Molecular Dynamics (MD) simulations are used to study the porosity of model

1 kerogens. The porosity is controlled effectively with systematic variation of the number and the
2 size of dummy LJ particles that are used during the construction of system's configuration. The
3 porosity of each sample is characterized with a newly proposed algorithm for analyzing the free
4 space of amorphous materials. It is found that with moderately sized configurations, it is possible
5 to construct percolated pores of interest in shale gas industry.

6 **1. Introduction**

7
8 The continuously increasing energy demands require new alternative sources of energy.¹ Shale
9 gas is an unconventional energy source, showing great potential to play an eminent role in the
10 energy market of the immediate future. Oil shales are mineral formations whose primary
11 constituent is clays that host sedimentary organic matter.² Kerogen is usually defined as the
12 insoluble in common organic solvents part of the organic matter found in shale formations capable
13 of generating oil or gas. Based on the hydrogen/carbon (H/C) and oxygen/carbon (O/C) ratios of
14 their elemental analysis, kerogens are classified in three types, namely I, II and III. A complete
15 definition of kerogen, its types and a variety of other relevant information can be found in the
16 review of Vandembroucke.³ Briefly, each type of kerogen is represented by specific regions in van
17 Krevelen diagrams, where the H/C is plotted against the O/C ratio. The type of kerogen found in
18 each shale formation is a consequence of many contributing factors such as the presence or not of
19 humic precursors, the level of maturation and the environment, just to name a few. For example,
20 type I kerogen is usually obtained in lacustrine environments, while marine environments usually
21 result in type II kerogen.

22 Although its existence is known for many years, shale gas attracted a lot of attention over the
23 last decades after the successful efforts to commercially exploit the Barnett shale in Texas, USA.⁴

1 ⁶ The commercial production of Barnett shale was made possible primarily by the development of
2 two new revolutionary technologies, the extended horizontal drilling⁵ and the hydraulic
3 fracturing.^{5,7} Despite the optimism about shale gas, there are a number of poorly understood open
4 issues that hinder a widespread production. The well productivity usually decays very fast^{8,9}
5 requiring re-stimulation for maximum yield, a process which is not always straightforward and
6 effective.^{4,10,11} In addition, a shale play may not always be productive with current technologies
7 available, despite the fact that it is known to host gas.^{12,13}

8 The environmental impact of the shale gas is also still under investigation.¹⁴ Although the
9 combustion of the produced shale gas has lower emissions than other fossil fuels,¹⁵ there are
10 numerous issues related to the techniques that are used for its extraction. Only a proportion of the
11 hydraulic fracturing fluid is recovered from the wells and its final destination is not known.^{8,9}
12 Furthermore, there are reports of high CH₄ concentration in water at regions close to shale fields,¹⁶
13 while commercial shale gas extraction activities have also been related to increased concentrations
14 of NO_x and volatile organic compounds (VOC)¹⁷ in nearby regions, as well as with the possibility
15 of pollution by radioactive materials.¹⁸ Finally, investigations have also been conducted with
16 regard to the relation of hydraulic fracturing and increased seismic activity around the shale
17 fields.¹⁹

18 It is evident that shale gas production spread is inherently linked with addressing at least a part
19 of the open questions associated with its environmental impact. For this reason, there is significant
20 industrial and academic research effort aiming in settling the various controversies about shale gas
21 technology. Composition, density, porosity, total organic carbon (TOC) and transport properties
22 are only a few of the properties examined in several experimental studies of various shale and
23 kerogen samples of different fields. One of the most distinctive characteristics of shale samples

1 and kerogen, in particular, is their complex porous network. Various imaging, scattering and
2 adsorption techniques^{20,21} have revealed that pores of different sizes, ranging from micro (<20Å)
3 to macro pores, are found in kerogen and clays of shale samples. Adsorption experiments
4 performed to shale samples, dried or not, have showed that in many cases their maximum CH₄
5 sorption capacity is linearly correlated to their TOC content.²²⁻²⁴ This is a strong indication that
6 organic matter dispersed in shale has the most significant contribution to gas adsorption.

7 Apart from the experimental research on kerogens, a continuously increasing interest in
8 modeling work is also seen in the literature. To this end, various efforts have been made to study
9 kerogen and other porous carbonaceous materials by means of molecular modeling. A review by
10 Palmer and Gubbins²⁵ resulted to the categorization of the methods originally been used for the
11 study of disordered nanocarbons (DNCs) and their porous network. Although these materials
12 have significantly different chemistry compared to kerogen, the methods are not limited to DNCs
13 and can potentially be used for the study of kerogen. This categorization discriminated the methods
14 into i) constitutive, ii) mimetic or iii) reconstruction methods. Constitutive methods include
15 studies on slit pores of graphite,²⁵ using coarse grained^{26,27} and other approximate models.^{25,28}
16 Mimetic and reconstruction methods utilize many body potentials like reactive force fields such
17 as REBO²⁹ and ReaxFF.³⁰ Mimetic techniques include methods such as the quenched molecular
18 dynamics (QMD), while reconstruction methods refer to techniques such as the reverse Monte
19 Carlo method (RMC)³¹ and its variations such as the latest advance hybrid reverse Monte Carlo
20 method (HRMC)³² and molecular dynamics hybrid reverse Monte Carlo method (MD-HRMC).³³
21 RMC based methods usually employ a reactive force field and aim at constructing a realistic model
22 of the material under study by modifying a configuration of the system via a MC scheme, in a way
23 that experimental information (usually the radial distribution function, $g(r)$ ^{25,33,34}) is reproduced.

1 Bousige et al.³⁴ used a MD-HRMC method with the REBO²⁹ force field in the study of four
2 different kerogen samples. They constructed kerogen models consisting of C, H and O, which they
3 validated against structural experimental data (e.g. the radial distribution function, pore size
4 distribution) and elastic properties (e.g. Young modulus). This approach produced realistic
5 structures despite the fact that a small proportion of unphysical features such as three membered
6 rings or cumulene chains were also present. Similar HRMC based models have also been used for
7 the study of alkane adsorption,³⁵ mixture diffusion in kerogen and the validation of a free volume
8 theory.³⁶ A major drawback of the reactive force fields is the high computational cost compared
9 to their classical counterparts. In addition, tedious effort is required for their parameterization in
10 order to describe accurately the specific chemistry of the target system.

11 A fourth approach refers to the representative molecular models method. It involves the
12 construction of a model molecule of kerogen based on available experimental data (such as
13 elemental analysis or scattering data even of a specific sample) and prior knowledge on molecular
14 fragments that likely exist in kerogen. To the best of our knowledge, the first effort in this direction
15 was made by Behar and Vandenbrouke³⁷ who presented representative 2D molecular diagrams of
16 the three types of kerogen at three stages of their maturation. Maturation was reflected in the
17 decreasing size of the molecules of each type, whose initial size (less mature) was in the order of
18 15,000 carbon atoms. This approach was later extended to 3D structures.³⁸ The models of Behar
19 and Vandenbrouke have been also used in the study of the volumetric properties of the bulk Green
20 River immature type I kerogen.³⁹ Another 2D model for the Green river kerogen was proposed by
21 Siskin et al.,⁴⁰ with its major difference from the previously mentioned models lying in the kerogen
22 representation, which is performed by six molecules. 3D configurations of the Siskin's model
23 generated by *ab-initio* calculations have also been published.⁴¹

1 Recently, Ungerer et al.⁴² published a number of kerogen molecules of smaller size, representing
2 different kerogen types at different maturation stage constructed in a similar framework based on
3 experimental data⁴³ and prior knowledge about molecular fragments. The geometry of these model
4 molecules was subsequently optimized using semi-empirical quantum chemistry methods, before
5 using the molecules for the calculation of different thermodynamic and volumetric properties via
6 different computational methods. In this approach, the porosity can be imposed using LJ dummy
7 particles,⁴⁴ or by including directly the organic mixture to be studied during the bulk kerogen
8 construction stage.⁴⁵ This type of kerogen models are constructed using equilibrium Molecular
9 Dynamics (MD). Various structural, thermodynamic and transport properties, such as adsorption
10 and/or desorption of small molecule and the diffusivity of pure fluids and fluid mixtures were
11 calculated.^{42,44-49}

12 Between the four aforementioned broad categories of methods i.e. i) constitutive, ii) memetic
13 iii) reconstructive and iv) representative molecular models, we adopted the latter in this study. In
14 view of the broad diversity of experimentally determined physical properties such as TOC, density,
15 porosity, permeability etc. not only for different shale fields of similar type kerogen but also for
16 different samples of the same well,^{21,24,50-52} it seems appropriate to investigate ways to adapt the
17 methodology of Ungerer et al.⁴² and Collell et al.⁴⁴ so that the resulting porosity has characteristics
18 tailored to the specific sample under study. In this respect, we systematically investigate how
19 several factors of the methodology affect the porosity of the resulted models. We have used the
20 published overmature type II model molecule (type IID) of Ungerer et al.⁴² to construct systems
21 of various sizes employing GAFF⁵³ and Dreiding⁵⁴ force fields. We observe the effect of system
22 size on the accessible space characteristics of the bulk. Porosity is controlled by choosing the

1 number and size of LJ dummy particles and the effect of these choices on the free and accessible
2 volume are assessed.

3 A number of algorithms for analyzing void space are available in the literature. Greenfield and
4 Theodorou⁵⁵ studied the diffusion pathways in atactic polypropylene using a Delaunay tessellation
5 based scheme. By joining adjacent accessible tetrahedra they traced the void clusters in the system,
6 while repeating the same cluster analysis for different probe radii they were able to identify how
7 these clusters are connected. A more "exotic" analysis by Alinchenko et al.⁵⁶ was used to study the
8 morphology of the void space distributed in a hydrated DMPC bilayer, based on the generalized
9 Voronoi S-tessellation. In this case, the specific features of the underlining S-tessellation
10 complicates and limits the applicability of the method. Finally, based on Voronoi tessellation and
11 by reviewing and implementing a set of algorithms for performing void space geometry-based
12 analysis of crystalline porous materials, Willems et al. developed the zeo++ tool.⁵⁷ In the current
13 study, two methodologies are used for the estimation and characterization of the free and accessible
14 volume: a) a grid based as implemented in Poreblazer⁵⁸ software and b) a Voronoi tessellation
15 based method developed in house and implemented as a python script.

16

17 **2. Methodology for constructing bulk kerogen models**

18

19 Bulk kerogen configurations are constructed using MD simulations. All the calculations
20 presented were performed using the II-D molecular model proposed by Ungerer et al.⁴² (Figure
21 1a), thereby work focused on overmature type II kerogen, which is representative of many
22 productive shale fields such as the Duvernay and Barnett.^{42,43} This is an approximately planar
23 molecule of almost disk-like shape, consisting of seven big fragments. Each of the constituent

1 fragments is planar, constructed primarily from fused benzene rings, although other five and six
2 member rings containing oxygen, sulfur and nitrogen atoms are present. The molecular formula of
3 II-D is $C_{175}H_{102}N_4O_9S_2$ and its molecular weight is $2468.8 \text{ g mol}^{-1}$. Although it is well understood
4 that kerogens are in reality a polydisperse multi-component mixture, the single unit approximation
5 is adopted here. The construction of the initial configuration is based on random placement of
6 kerogen molecules in a cubic box at very low density. Then following Ungerer et al.,⁴² the system
7 is cooled from a high temperature to the standard ambient temperature of 298.15 K. During the
8 cooling procedure, the pressure was kept constant at 1 atm, although in the original scheme a
9 higher pressure (100 bar) was used during the first stages of the procedure.

10 Calculations were performed using two force fields, namely, the Dreiding force field⁵⁴ with
11 Gasteiger atomic point charges⁵⁹ and the General AMBER force field (GAFF).⁵³ Dreiding was
12 used as a simple, widely available and easy to use force field. GAFF is a force field compatible
13 with the models usually employed for modeling molecules hosted by kerogen, e.g. small organic
14 molecules. Testing on a wide array of molecules showed that GAFF produces minimized structures
15 closer to their crystallographic counterparts, (i.e. lower root mean square displacement) compared
16 to Dreiding.⁵³ In order to derive GAFF parameters, we used the published conformation of Ungerer
17 et al.⁴² as the initial structure, which was then optimized at the B3LYP/6-31G* level of theory,⁶⁰⁻
18 ⁶² by means of the Gaussian 09⁶³ suite of programs. Partial atomic charges were obtained from the
19 optimized geometry at the HF/6-31G* level of theory, according to the Mertz-Kollman population
20 analysis scheme.^{64,65} Then, assuming no significant conformational changes from this geometry
21 during the dynamical simulations, partial atomic charges were derived according to the Restrained
22 Electrostatic Potential (RESP) protocol,⁶⁶ utilizing the ANTECHAMBER⁶⁷ module of the

1 AMBER12⁶⁸ suite of programs. Atom type definitions for the GAFF force field are shown in
2 Figure 1b.

3 As an inevitable consequence of the nature of the system, i.e. molecules of considerable size and
4 reduced flexibility, it is anticipated that the ergodic hypothesis⁶⁹ is not necessarily valid. In order
5 to address this issue, and to study the effects of a) the system size, b) the force field used and c) the
6 presence of the LJ particles in the resulted porosity, a number of initial configurations were created
7 whose details are presented in Table 1. The Amorphous Builder⁷⁰ software integrated in
8 Scienomics MAPS platform⁷¹ was used to place a number of type II-D molecules inside the
9 simulation box with density of 0.01 g cm⁻³.

10 In order to impose porosity on the constructed models, LJ dummy particles were used in a
11 manner similar to Collell et al.⁴⁴ The numbers of configurations created for each combination of
12 system size and LJ dummy particle sets for both force fields used are detailed in Table 1. The
13 number of LJ dummy particles varied from one to three, with their diameter varying from 15 to 40
14 Å. In all cases, the LJ particles had an interaction energy parameter⁴⁹ equal to 0.20 kcal mol⁻¹ and
15 mass equal to that of Argon (MW=39.948 g mol⁻¹). In this way, significant parameters of the
16 simulation that affect the characteristics of the resulting porosity (e.g. force field, number and size
17 of dummy particle etc.) were studied, allowing ultimately for the desired porosity to be achieved.

18 Each of the generated initial configurations was then relaxed using a staged cooling protocol by
19 MD simulations at the isothermal isobaric statistical ensemble (NPT).⁶⁹ The MD simulations were
20 performed with the LAMMPS⁷² MD simulation package. Throughout the NPT simulations the
21 pressure was kept constant at 1 atm. For the 15 molecule system, the simulations started at 2000
22 K with one NPT equilibration run of 1 ns. The system was then cooled using a stepwise procedure
23 of repeated simulation cycles. Each cycle was composed of two stages, a stage where temperature

1 was reduced linearly from an initial to a final temperature, followed by a small NPT equilibration
2 stage, as shown in Figure 2. In order to mitigate the computational cost, the 50 and 100 molecule
3 simulations started at 1100 K cooled down to 300 K by following the same two stage simulation
4 cycles approach as in the case of 15 molecule systems. However, the cooling stage of each
5 simulation cycle in these 50 and 100 molecule systems was 0.05 ns long with a cooling rate of
6 2000 K ns⁻¹, while the NPT equilibration stage lasted for 0.8 ns. The same cooling protocols were
7 used for systems with the corresponding size containing LJ dummy particles. In this case, an extra
8 step of NPT equilibration of 2 ns at 298.15 K and 398.15 K after LJ dummy particle removal was
9 applied. For all the MD simulations, the temperature and pressure were kept constant using Nose-
10 Hoover thermostat and barostat respectively.^{73,74} The integration time step was set to 1 fs for all
11 calculations. The traditional Ewald summation technique^{69,75} was used for the calculation of
12 electrostatic interactions in all systems. A simple cut-off at 12 Å was applied for all non-bonded
13 interactions.

14 Due to the elevated temperature and pressure conditions that exist in a shale gas reservoir,⁷⁶ the
15 effect of pressure on the density was investigated by sequential pressurization of the ambient
16 pressure structures obtained at 298.15 K and 398.15 K to 100 and 250 atm. The obtained kerogen
17 structures were analyzed with tools described in the next section.

18

19 **3. Calculation and characterization of porosity**

20

21 The porosity was studied using two approaches; a grid based algorithm as implemented in
22 Poreblazer open source software⁵⁸ and a Voronoi tessellation based methodology developed here
23 that does not rely on a grid. Each one used for the calculation of different porosity characteristics.

1 Information obtained by the analysis of the Voronoi network includes: (a) the CH₄ accessible
2 volume and surface, (b) the limiting and the maximum pore diameters, and (c) the identification
3 and the analysis of the diffusion paths spanning the configuration, while Poreblazer was used only
4 for the calculation of the pore size distribution (PSD). The limiting pore diameter (LPD) or largest
5 free sphere diameter (D_f),⁷⁷ is defined from the critical spherical penetrant diameter, above which,
6 diffusion would not be possible due to the lack of percolated pores, while the maximum pore
7 diameter (MPD) or larger included sphere diameter (D_i),⁷⁷ is the maximum penetrant diameter that
8 can fit into the structure. All the aforementioned properties were obtained as averages over time
9 and for different initial configuration of similar characteristics (force field, system size, number
10 and size of dummy LJ particles).

11 Poreblazer was used with the default settings (12.8 Å cut off distance, diameters of 3.314 Å and
12 2.58 Å for N₂ and He respectively and for all other elements values for the σ and ϵ parameters of
13 the Universal force field, UFF⁷⁸) except for the grid spacing. The cubelet size was set to 0.7 Å for
14 the 100 molecule simulations and 0.5 Å for the other simulations, in order to reduce the required
15 computational cost. More details for the grid based algorithm and functionality of Poreblazer can
16 be found in relevant literature by Sarkisov and Harrison.⁵⁸

17 The scheme developed for the detailed characterization of the void space in periodic crystalline
18 or amorphous molecular systems combines the Voronoi tessellation with concepts and algorithms
19 from graph theory. Furthermore, the Dodd and Theodorou⁷⁹ algorithm was used for the analytical
20 calculation of the free/accessible volume. The proposed approach was implemented as a python
21 script with parts of the functionality implemented in FORTRAN for efficiency reasons. The
22 calculation of Voronoi polyhedra was performed by means of voro++ software⁸⁰ written in c++.
23 In the next section, the constituent steps of the new algorithm are described thoroughly.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

3.1. Voronoi tessellation based analysis of the free space

The workflow of the scheme developed for the accessible space analysis is pictorially shown on the flowchart of Figure 3. In the first step of our scheme, the radical Voronoi tessellation is performed using Voropp⁸¹ and the corresponding polyhedra (hereafter referred as Voronoi cells or just cells) are calculated for each atom (Figure 4a). This is a weighted tessellation where the separation plane between two spheres depends on their radius i.e. is the locus of points with equal tangent relative to the two spheres.⁸² This property, in the case where the system constituent atoms are of different sizes, ensures that the center of each atom lies inside the corresponding Voronoi polyhedron resulting in a more realistic per atom decomposition of the space required in our analysis.⁸³ Using the obtained information (i.e. vertices, edges and faces defining the Voronoi cells), we then proceed with the analysis of the void space and the calculation of the quantities mentioned in the previous section. In order to define the accessible volume, a spherical probe is assumed and subsequently the radius of all atoms is increased by the radius of the probe. Here it is worth noting that the radius of the kerogen atoms can be defined according to any force field. In this study the diameters of the UFF force field have been used in all cases, for reasons of compatibility with the calculations performed using Poreblazer. The probe radius assumed is 1.865 Å, which is the radius of CH₄ in TraPPE force field.⁸⁴ A description of the way that the aforementioned quantities of interest are calculated follows.

(a) Calculation of the accessible surface and volume

The total accessible volume of the system of N atoms, V_{tot}^{acc} , is the sum of the individual cell contributions:

$$V_{tot}^{acc} = \sum_{i=1}^N V_i^{acc} \quad (1)$$

1 where, V_i^{acc} is the accessible volume of the i^{th} cell. V_i^{acc} is calculated as the total volume of the
 2 Voronoi cell, decreased by the volume of the fraction of the atom that is within the cell. The latter
 3 is obtained by evaluating the remaining sphere (atom) volume, intersected by a number of planes
 4 using the Dodd and Theodorou algorithm.⁷⁹ In this case, the atom defines the sphere, while the
 5 faces of the corresponding Voronoi polyhedron define the cutting planes. The same calculation
 6 provides also the cell contribution, S_i^{acc} , to the total accessible surface S_{tot}^{acc} :

$$S_{tot}^{acc} = \sum_{i=1}^N S_i^{acc} \quad (2)$$

7

8 (b) *Identification of accessible vertices and edges*

9 The remaining analysis requires the identification of the vertices and edges of the constructed
 10 polyhedra that do not overlap with systems' atoms (Figure 4b). These lie in the void space of the
 11 material and host the probing molecule. For each accessible vertex, we calculate the maximum
 12 probe that it may host; for each accessible edge, we calculate the maximum probe possible to
 13 traverse it. These quantities will be used for the identification of diffusion paths along percolated
 14 pores, and several characteristic pore diameters (e.g. maximum or limiting pore diameters).
 15 Therefore the accuracy by which they are going to be calculated will ultimately determine the
 16 accuracy of the subsequent calculations.

17 Each vertex is checked against its nearest atoms, i.e. the atoms that correspond to the Voronoi
 18 cells by which it is shared. Subsequently, for every non-overlapping vertex j , the minimum
 19 distance, d_j , from the surface of the four nearest atoms is calculated analytically. An edge is
 20 considered as overlapping if at least one of its vertices is either overlapped or intersected by a

1 nearby sphere. The latter condition is verified by solving analytically the line-sphere intersection
2 problem for the atoms (spheres) of the cells where the edge (line) participates in. In our approach,
3 every non-overlapping edge (k) is characterized by the passing probe diameter (d_k^{PPD}), which is
4 the maximum diameter of the spherical probe that is able to pass from one constituent vertex to
5 the other. If the edge intersects the triangle defined by its three atom participating cell (Figure 5a),
6 the d_k^{PPD} is calculated analytically as the diameter of the sphere tangent to surfaces of the atoms,
7 with its center lying on the triangle's plane. Note that the center of this sphere is equidistant from
8 the atoms' surface. If the edge does not intersect the aforementioned triangle (Figure 5b), a number
9 of equally distributed edge points (including the vertices) are selected and the distance of each
10 point from the surface of the closest atom is calculated. In this case, d_k^{PPD} is specified as the
11 minimum of these distances. In this case, the passing pore diameters calculation is not analytical
12 and the accuracy is affected by the number of points selected along the edge. In this study, a large
13 number of ten points along the edges of this type are chosen (this is almost every 0.1 Å on average).

14 The non-overlapping vertices and edges form a network spanning the void space of the periodic
15 simulation box (Figure 4c). Due to the periodic boundary conditions (PBC) imposed in our models,
16 a non-overlapping vertex v_i may have one or more periodic images $v_{i'}$. At this point, the periodic
17 indices $\mathbf{i} = (i_x, i_y, i_z)$ of each image $v_{i'}$ are identified.

18 (c) *Identification of pores*

19 The analysis of the resulted network leads to the identification of the pores existing in the
20 structure. In the context of graph theory, this network can be seen as a bidirectional, disconnected
21 (in the general case) graph G , while the clusters of interconnected vertices are the constituent
22 connected sub-graphs, G_i .

1 Topologically, the G_i clusters may be connected due to the periodic boundary conditions
2 imposed in our models (Figure 6). In this sense, a pore is a set of connected clusters. Two clusters,
3 C_1 and C_2 , are considered to be connected if at least one vertex $v_{i'}^{C_2}$ of the second is the periodic
4 image of a vertex $v_i^{C_1}$ belonging to the first or *vice versa*. The pores that exist in the system are
5 reconstructed by merging together their constituent connected clusters. Two clusters (C_1, C_2) are
6 merged by translating the vertices of C_2 by the minimum vector separating a shared pair of parent-
7 image vertices $v_i^{C_1}, v_{i'}^{C_2}$ (Figure 7). In order to keep track of the relevant cluster positions, the
8 periodic indices of the image vertices that either themselves or their parents belong to C_2 , are
9 updated according to the translation performed. Finally, the image vertices of the resulting cluster
10 ($C_{12} = C_1 \cap C_2$) coinciding with their parents (i.e. their periodic indexes after the update are zero)
11 are fused with their parents i.e. they are removed and the connectivity of their parents is updated
12 properly taking care that no duplicated edges should exist.

13 The free/accessible volume of the located pores, percolated or not, is calculated based on eq. (1)
14 from the contributions of pores' constituent cells. It is worth noticing that if a Voronoi cell is shared
15 between two or more pores, only the part of the cell belonging to the pore under consideration is
16 taken into account.

17 (d) *Percolation analysis*

18 The resulted pores are then examined for percolation. If at least one pair of parent-image vertices
19 $v_i, v_{i'}$ exist in the pore, then it is considered percolated with the periodic indices (i_x, i_y, i_z) of the
20 image $v_{i'}$ indicating the direction of percolation. The set of unique periodic indices found for all
21 parent-image pairs reveal the directions of the percolation channels that exist in the pore (Figure
22 6b). The alternative approach for assessing percolation is to check the spatial characteristics of the

1 pore. In this case, if a dimension of the pore's bounding box is greater than the corresponding
 2 dimension of the simulation box then the system is considered to be percolated along this direction.

3 (e) *Identification of diffusion paths and their limiting pore diameters*

4 To systematically trace the paths traversing a percolated pore it is useful to define the notion of
 5 an *island*. The image vertices $v_{i'}$ of the pore are grouped based on their periodic indices, $\mathbf{i} =$
 6 (i_x, i_y, i_z) . For each group the aggregations of interconnected vertices are identified. We refer to
 7 such aggregations as *islands* and the m^{th} island of group \mathbf{i} is denoted as $I_{\mathbf{i}}^m$, with $m = 1, 2, \dots, N_{\mathbf{i}}^I$
 8 and $N_{\mathbf{i}}^I$ being the number of *islands* of group \mathbf{i} (Figure 8). The parents of the images belonging to
 9 island $I_{\mathbf{i}}^m$ are also interconnected and define the parent *island*, hereby denoted as $I_{\mathbf{i}}^m$. Each pair of
 10 parent-image *islands* defines a percolation channel which is characterized by a maximum size of
 11 the probe able to traverse it. Then, all possible pairs of islands $I_{\mathbf{i}}^m, I_{\mathbf{i}}^{m'}$ with $m, m' = 1, 2, \dots, N_{\mathbf{i}}^I$
 12 are considered and for every pair a path connecting the two *islands* is found.

13 According to graph theory, a path is a sequence of connected edges and the length of the path is
 14 equal to the number of its edges. In a percolated pore, at least a path of length ℓ connects a pair of
 15 parent-image vertices, $v_i, v_{i'}$ (or equivalently a pair of parent-image islands $I_{\mathbf{i}}^m, I_{\mathbf{i}}^{m'}$) of a
 16 percolated pore and is denoted as $P_{v_i, v_{i'}}^{\ell}$. The shortest percolated path is the one of the smaller
 17 length:

$$P_{v_i, v_{i'}}^{\min} = \min_{\ell} P_{v_i, v_{i'}}^{\ell} \quad (3)$$

18 In the context of amorphous materials, we are interested in identifying the path that has the
 19 potential to allow faster and easier diffusion in the material. This is the shortest path composed of
 20 edges of the larger possible passing pore diameter, d_k^{PPD} . In order to have a more complete
 21 exploration of the diffusion possibilities in the material, a path of these characteristics is being
 22 found for every pair $I_{\mathbf{i}}^m, I_{\mathbf{i}}^{m'}$ that has earlier been determined. This task is performed using the

1 Dijkstra algorithm,⁸⁵ as implemented in networkX python library.⁸⁶ The unbiased Dijkstra
 2 algorithm identifies the path of minimum length connecting two vertices. If weights are assigned
 3 to the edges, the algorithm traces the path with the minimum sum of weights. Therefore, in order
 4 to find the path that meets our criteria for the pair $I_i^m, I_i^{m'}$, we choose the parent-image pair with
 5 the larger distance from atoms' surface as the two vertices and we assign appropriate weights, w_k ,
 6 to the edges of the pore:

$$w_k = (M - d_k^{PPD} + m) a \quad (4)$$

7 where, a is a scale factor used to increase the separation between the weights that correspond to
 8 the smaller, m , and the larger, M , values of d_k^{PPD} values met in the pore:

$$M = \max_k \{d_k^{PPD} : k \in \text{pore edges}\}, \text{ and} \quad (5)$$

9

$$m = \min_k \{d_k^{PPD} : k \in \text{pore edges}\}. \quad (6)$$

10 It should be noted that according to (5) the larger the d_k^{PPD} , the smaller the corresponding weight
 11 w_k . Therefore, the traced paths meet the criteria set and the maximum probe that can traverse the
 12 path corresponds to the edge with the largest w_k .

13 Due to the nature of the systems we are interested in, it is possible that the path traced for the
 14 pair $I_i^m, I_i^{m'}$ does not include the edge corresponding to the real maximum passing probe diameter.
 15 This is due to the fact that a path of much smaller length may result to a smaller sum of weights
 16 even if the weights of the edges are relatively high. To overcome this problem, the edge with the
 17 maximum weight, w_k , i.e. the minimum passing probe diameter, located from the Dijkstra
 18 algorithm is eliminated from the pore's graph and the path connecting the *islands* is recalculated.
 19 This procedure is repeated until the graph becomes disjointed (i.e. there is no path connecting these
 20 vertices). The free particle diameter corresponds to the examined islands pair, $d_{I_i^m, I_i^{m'}}^{PPD}$ is the

1 maximum value of the calculated minimum passing pore diameters of all the traced paths. In turn,
2 the maximum value of the free pore diameters of all the identified channels corresponds to the
3 limiting pore diameter d_{pore}^{PPD} of a percolated pore.

4 (f) *Determination of the maximum and limiting pore diameters*

5 In the final step, the maximum included sphere, D_I , that can be hosted in the structure and the
6 limiting pore size or free sphere diameter, D_f , are calculated. D_I , is obtained straightforwardly as
7 the maximum value of the distances d_j that were determined earlier for each accessible edge:

$$D_I = \max(\{d_j\}) \quad (7)$$

8 D_f is the critical penetrant diameter above which diffusion would not be possible due to the lack
9 of accessible percolated pores and therefore, by definition, is the maximum value of the calculated
10 limiting pore diameters:

$$D_f = \max\{d_{pore_i}^{PPD}\} \quad (8)$$

11 given that percolated pores exist in the system for the probe diameter in use. Thus, when the
12 accessible volume comprises no percolated pores, the assumed probe radius is reduced gradually
13 until it is sufficiently small in order to be able to freely diffuse in the structure. Then, D_f , is
14 calculated accurately.

15

16 **4. Results**

17

18 All the thermodynamic and structural properties presented here are calculated as averages over
19 time and different initial configuration of the same characteristics, i.e. force field, size and number
20 of LJ particle(s). In this way the effect of thermal motion of the kerogen molecules is reflected on
21 the calculated distributions. Since we are interested for systems relevant to shale gas, the Voronoi

1 tessellation methodology was applied using a probe of radius of 1.865 Å which is the radius of the
2 united atom representation of CH₄ in the TraPPE force field⁸⁴.

3 The density as a function of temperature for the different configurations created without the aid
4 of LJ dummy particles is shown in Figure 9 for Dreiding and GAFF force fields, respectively. In
5 both cases, the average value of the density decreases with system size reaching a limiting value
6 already for the 50 molecules system. As expected, the statistical uncertainty becomes narrower
7 with increasing systems size. The GAFF force field produces denser structures for every system
8 size and temperature by a factor of 6.9% on average and a maximum of 16%. The system size
9 effect is evident for both force fields. The obtained densities are expected to remain unaffected
10 upon re-heating of the low temperature structure at least up to 700 K, according to the findings of
11 Ungerer et al.⁴². This is partly verified by heating the 15 II-D structures obtained using GAFF at
12 298.15 K, to 900 K by a protocol identical to the one in Figure 2. The obtained densities (not
13 presented here) by this procedure are identical to the densities of Figure 9b.

14 The key characteristics of the void space are summarized in Table 2; namely, the accessible
15 surface (S_{acc}) and accessible volume (V_{acc}), the % porosity, the limiting and maximum pore
16 diameters (LPD and MPD, respectively), the percentage of percolated structures and the volume
17 and surface of the larger pore in the structure, for the different structures constructed. The pore
18 size distributions calculated for the systems constructed without the presence of LJ dummy
19 particle(s) at 298 K are shown in Figure 10 for the two force fields considered. For the system of
20 15 molecules, the PSD is calculated also at 600 K. In all cases, the smallest pore diameter is
21 approximately 2.5 Å. The maximum pore diameter observed using Dreiding force field is roughly
22 12 Å for all sizes, while for structures constructed with GAFF and 15 or 50 II-D molecules
23 approach the value of 10 Å and jumps to 14 Å for the 100 II-D molecules systems. For both force

1 fields, temperature is found to have negligible effect on the pore sizes of the constructed kerogen
2 models, a finding suggesting limited mobility of the bulky II-D molecules in a congested
3 environment, even at temperatures as high as 600 K. As the system size increases, larger pores
4 appear in the configurations, while only for the GAFF 100 II-D molecule system this is combined
5 with an increased maximum pore diameter from 12 to 14 Å. This indicates that the PSD is system
6 size dependent and suggest that it is necessary to move to relatively large system sizes consisting
7 of at least several tens of molecules when simulating kerogen with this type of models.

8 Even though the pore size distribution describes the type of pores of the material, it is not
9 sufficient to provide a complete picture of the pathways available for the diffusion of gases in
10 kerogens. In this respect, the limiting pore diameter or free sphere radius is necessary as a rough
11 estimate for the size of the molecules possible to traverse the percolated paths spanning the
12 structure. Normalized LPD distributions (LPDD) defined equivalently to PSD and calculated
13 using standard binning procedure are shown in Figure 10c and 10d for the various system sizes for
14 both force fields at 298 K and for the 15 II-D system also at 600 K. In all cases the LPD value does
15 not exceed the diameter of TraPPE CH₄ i.e. 3.73 Å which indicates that even the diffusion of small
16 molecules such as CH₄ is difficult to occur especially in a static structure. The effect of system
17 size in the LPDD is in general considered minor.

18 Considering the diversity of experimental results, a number of kerogen models were constructed
19 with the aid of LJ dummy particles aiming to impose in a controlled manner the major
20 characteristics of the porosity. This effort focused to 50 II-D structures using GAFF since the bulk
21 density reached a limiting value for this system size and GAFF is more accurate in the structural
22 reproduction of small organic molecules. Figure 11 shows the effect of the LJ dummy particles
23 and pressure on the calculated density for these configurations at 298 K. As expected, the structures

1 generated with the aid of LJ dummy particles are less dense with decreasing density as the size of
2 the particle(s) increase. This behavior is not observed for the structures constructed with one 30 Å
3 and one 40 Å LJ particle where the structure constructed with 40 Å LJ particle is denser. The
4 majority of the models are denser than 1.1 g cm⁻³ with the smallest observed density of 1.04 g cm⁻³
5 obtained when two 30 Å LJ particles are used. With the exception of the case of the one LJ
6 particle with diameter 20 Å, the increase of the pressure from 1 to 250 atm results in an increase
7 of the density around 0.1% indicating the small effect of the pressure on the density.

8 The PSD and LPDD of the 50 molecules GAFF structures constructed with the aid of different
9 number of 15 Å LJ dummy particles are shown in Figure 12, together with the distributions
10 (defined and calculated equivalently to LPDD) of the maximum pore diameter (MPDD) and the
11 % porosity (PD). The % porosity is calculated as the fraction of CH₄ accessible volume with
12 respect to the total volume of the system. The use of LJ dummy particle(s) increases the accessible
13 volume in the structure as it is evident from PD (Figure 12a). This extra space is distributed in
14 larger pores, as it is indicated by the broadened PSD (Figure 12b) with maximum pore diameters
15 approaching 20 Å. The presence of bigger pores is also reflected in the MPD distribution (Figure
16 12d). In general, inclusion of more LJ particles allows the construction of bigger pores in addition
17 to the small pores (~2-5Å) that remain in the structure. On the other hand, LPD distribution (Figure
18 12c) is not significantly affected by the increase in the number of LJ particles.

19 The effect of the variation of the LJ dummy particle diameter on the characteristics of the
20 accessible volume is shown in Figure 13. Porosity can also be increased by increasing the LJ
21 dummy particle size as it is intuitively expected. While this is true in principle, there is an upper
22 limit in the porosity that can be achieved. When increasing the size from 30Å to 40Å, no significant
23 increase in the induced porosity is observed (Figure 13a). In every case, no more than 10% porosity

1 is being obtained. Comparing to increasing the number of LJ particles, variation of their size is
2 more effective in inducing porosity in the sense that equal or bigger porosity can be achieved with
3 a small increase of the size of one particle. Furthermore, increasing the LJ dummy particle diameter
4 from 15 to 30 Å results in larger pores (with bigger change from 15 to 20 Å), as it is indicated by
5 the PSDs of Figure 13b. The broadening trend of the PSD is not observed when the diameter of
6 the LJ particle is increased from 30 to 40 Å. This suggests an upper limit in the pore sizes that can
7 be constructed using one LJ particles with approximate value of 23 Å. The effect of increasing the
8 LJ particle size on pore diameters shown by PSDs is also reflected on MPDD, which is translated
9 towards bigger pores. In agreement with previous findings this trend ends when LJ particle size
10 increases from 30 to 40 Å (Figure 13d). The cease of the LJ particle size effect on pore diameter
11 is a consequence of the size of the II-D molecule, whose diameter is approximately 28 Å in its gas
12 phase conformation⁴² and can be even smaller in the bulk. Overall, the LJ particle diameter
13 variation has a more pronounced effect on LPD distribution (Figure 13c) compared to changing
14 the number of smaller size LJ particles. Limiting pore distributions become broader and less sharp,
15 reaching values greater than the diameter of CH₄ (something which couldn't be achieved with
16 variation of then number of LJ dummy particles) up to 5 Å. It is evident that the increase of LJ
17 dummy particle size is a much more effective in controlling the major characteristics of porosity,
18 being the only way that allows the creation of diffusion channels in the bulk (i.e. limiting pore
19 diameters greater than the diameter of CH₄).

20 The influence of particles of 20 and 30 Å size on the limiting pore is shown on Figure 14. It is
21 clear that an increased number of larger particles can result in well-shaped, wide percolated pores
22 that allow large size molecule diffusion. A considerable number of pores with limiting diameters
23 higher than the diameter of CH₄ and up to 10 Å are created.

1 Overall no single model shares all types of micropores that are known to exist in kerogen.
2 Despite this, the various combinations of LJ dummy particle numbers and sizes have made it
3 possible to create structures of different porosity characteristics that are all present in kerogen.
4 These characteristics are closely linked to phenomena that are of interest in modeling systems
5 relevant to shale gas.

6 Although direct quantitative comparisons between the calculated quantities presented here and
7 experimental results is not possible, all the models that have been generated are in general within
8 or close to experimental range. The majority of the models give densities between 1.1 and 1.3
9 g cm^{-3} , values very close the commonly observed densities for type II kerogen⁵². PSDs are in
10 accordance with the experimental pores sizes of micropores, having diameters of a few Å.^{20,34}
11 Ultimately, to the best of our knowledge, the only information on limiting pores is related to
12 Barnett shale, where experiments have proven the existence of pore throats of less than 5 Å,^{4,7} a
13 finding that corroborates our calculations.

14 Percolation analysis reveals that CH₄ accessible pores can be percolated only when a LJ dummy
15 particle is used. Other simulation parameters such as the system size or the force field seem to
16 have minor effect on the more important characteristics of the porosity. The percentage of
17 configurations comprising percolated pores is significant only in the cases of 3x20 Å, 2x30 Å and
18 1x40 Å LJ dummy particle sets (Table 2). When probing the accessible volume with a 2 Å particle
19 diameter, all approaches with both force fields result in a large number of percolated
20 configurations, exhibiting only a single pore in the vast majority of the cases (more than 99%). For
21 this reason, analysis of the diffusion paths is performed with probe diameters equal to 2 Å, so that
22 pore characteristics of all procedures are acquired. Figure 15a shows the % of each GAFF atom
23 type (see Figure 1b) located at the larger percolated pore surface, as a fraction of the total number

1 of atoms of this type found in the configuration. There is clear tendency of some specific atom
2 types to lie predominantly at the pore surface instead of the bulk. For example, more than half of
3 the pyrrolic hydrogens (atom type hn) in each configuration are found at the surface of pores. This
4 finding may have important implications when studying the diffusion of polar molecules. The total
5 height of the bars on Figure 15a, is related to the total number of atoms by which the pore surface
6 is constructed. Therefore, a correlation between the total height of bars for each modeling approach
7 and the corresponding larger pore surface area (see penultimate column of Table 2) is anticipated.
8 Such a correlation between pore surface and number of pore surface atoms is observed for all cases
9 except for the case of one 40 Å dummy particle. The bars for the structures constructed with a 40
10 Å LJ particle are shorter than the bars of the a 30 Å LJ particle structures despite the fact that pore
11 surface is higher for the 40 Å LJ particle. This is a consequence of the parallel alignment of the
12 molecules in the void surface (unlike the perpendicular alignment of the previous cases), as it is
13 visually verified.

14 The fraction of the atoms of the pore that are found at limiting diameter regions is shown in
15 Figure 15b. We define as limiting regions the pore segments which are accessible by probes up to
16 20 % larger than CH₄ i.e. of 4.476 Å diameter. The reduction of bars in Figure 15a to the bars in
17 Figure 15b does not follow a specific trend. Finally, Figure 15c shows the fraction of each atom
18 type in the limiting region, with respect to the total number of atoms in the limiting region. There
19 is a measurable fraction of pyrrolic groups, indicated by the presence of hn atoms, while the
20 limiting region environment is dominated by aromatic carbon and hydrogen ca and ha respectively.

21

22 **5. Conclusions**

23

1 Dreiding and GAFF force fields were used for the construction of bulk kerogen models using a
2 model kerogen molecule, namely II-D, available in the literature. A staged cooling technique was
3 used for the construction of systems of varying size up to 100 II-D molecules. Different sets of LJ
4 particles were used during the cooling stage for imposing porosity in a systematic and controlled
5 manner. Density was calculated and porosity was analyzed using two algorithms. The force field
6 was found to have a minor effect on the calculated properties, which is consistent with calculated
7 densities reported in the literature for other force fields. Systems size is found to have some effect
8 on densities and the created pores. While decreasing the system size increases the computational
9 efficiency, it seems that it is necessary to have a considerable system size of at least 50 II-D
10 molecules in order to have a representative PSD and LPD distributions. The method of dummy LJ
11 particles is found to be appropriate in the effort to have a representative porous space in the bulk.
12 While in every case the ultralow size pores of below 10 Å are present, including LJ dummy
13 particles allows the modeling of the remaining spectrum of micro-pores (up to 20 Å). Increasing
14 the number of LJ particles has a broadening effect on the PSD, but does not seem very effective in
15 altering the LPD distribution. On the other hand, the increasing size of a single LJ particle seems
16 the appropriate way to affect in a controlled way both PSD and LPD distribution. Certain
17 combinations of size and number of particles allow the construction of pores large enough to study
18 the diffusion of several molecules of considerable size.

19 Elemental analysis of the pore surface indicates a variation in the preference of certain atom
20 types to occupy the surface of the pore instead of the bulk. Overall, important parameters in the
21 construction method for bulk kerogen models were investigated on mature type II kerogen and
22 effective ways that allow one to construct models of tailored porosity in accordance with the

1 diverse experimental results were proposed. In the future, the diffusion of small molecules in the
2 kerogen structure will be examined.

3

4 AUTHOR INFORMATION

5 **Corresponding Authors**

6 *Emails: i.economou@inn.demokritos.gr, l.peristeras@inn.demokritos.gr.

7 **Author Contributions**

8 The manuscript was written through contributions of all authors. All authors have given approval
9 to the final version of the manuscript.

10 **Funding Sources**

11 All the results reported here were performed under ShaleXenvironment, H2020 project, funded
12 by the European Commission Horizon 2020 Framework Programme for Research and Innovation
13 (Grant Agreement no 640979).

14

15 **ACKNOWLEDGMENT**

16 We are grateful to European Commission for funding our research under Horizon 2020
17 Programme for Research and Innovation and to Professor Andrey G. Kalinichev and Professor
18 Martin Lisal for fruitful discussions on modeling kerogens. This work was supported by
19 computational time granted from the Greek Research & Technology Network (GRNET) in the
20 National High Performance Computing facility - ARIS - under project ID002043. We are also
21 thankful to the High Performance Computing Center of Texas A&M University at Qatar for
22 generous resource allocation. Scienomics SARL is acknowledged for providing MAPS software

1 used to generate initial structures of the systems simulated. The code for the Voronoi based
2 characterization of the accessible space of an amorphous material is available upon request.

3

4

5

1
2 **Table 1:** Details of the number of configurations constructed, categorized based on (a) the size of
3 the system, (b) the size and the number of the LJ particles and (c) the applied force field. The bar
4 (-) represents not attempted calculations.

# II-D molecules	LJ particle		# configurations	
	σ (Å)	# particles	Dreiding	GAFF
15	None		16	16
	15	1	16	-
50	None		5	6
	15	1	-	6
		2	-	6
		3	-	6
	20	1	-	6
		2	-	6
		3	-	6
	30	1	-	6
		2	-	6
	40	1	-	6
100	None		5	6

16
17
18

Table 2. Summary of the average values of the major characteristics of the constructed bulk kerogen models. Numbers in parentheses are the corresponding standard deviations. The Dreiding force field is abbreviated with D and the GAFF force field with G.

# II-D.	LJ particles		S_{acc} ($m^2 g^{-1}$)		V_{acc} ($10^{-10}m^3 g^{-1}$)		% porosity		LPD (\AA)		MPD (\AA)		% percolated configurations [†]		Larger pore [†] (GAFF)	
	σ (\AA)	#	D	G	D	G	D	G	D	G	D	G	D	G	S_{acc} ($m^2 g^{-1}$)	V_{acc} ($10^{-10}m^3 g^{-1}$)
15	None		188 (59)	145 (53)	88 (53)	74 (50)	1.0 (0.6)	0.9 (0.6)	2.3 (0.3)	2.1 (0.3)	7.4 (1.3)	7.3 (1.2)	0	0	76 (41)	49 (38)
	15.0	1	339 (59)	-	312 (86)	-	3.5 (0.9)	-	2.8 (0.6)	-	11.6 (1.2)	-	11	-	-	-
50	None		257 (10)	212 (20)	187 (9)	119 (19)	2.1 (0.1)	1.5 (0.2)	2.6 (0.3)	2.2 (0.2)	11.7 (0.3)	9.1 (1.0)	0	0	50 (17)	41 (21)
	15.0	1	-	270 (33)	-	222 (48)	-	2.7 (0.6)	-	2.4 (0.5)	-	11.4 (1.7)	-	0	127 (43)	137 (52)
		2	-	253 (35)	-	228 (50)	-	2.7 (0.6)	-	2.5 (0.3)	-	12.7 (0.8)	-	0	103 (36)	128 (54)
		3	-	313 (30)	-	335 (95)	-	3.9 (1.0)	-	2.4 (0.3)	-	13.9 (2.2)	-	0	160 (68)	232 (132)
	20.0	1	-	289 (46)	-	328 (96)	-	3.9 (1.0)	-	2.5 (0.4)	-	16.5 (2.4)	-	1	175 (52)	276 (98)
		2	-	381 (55)	-	616 (123)	-	6.9 (1.3)	-	3.0 (1.0)	-	18.1 (1.2)	-	17	297 (81)	572 (143)
		3	-	436 (39)	-	813 (77)	-	8.9 (0.7)	-	4.7 (2.3)	-	19.6 (2.0)	-	48	372 (62)	776 (112)
	30.0	1	-	379 (48)	-	564 (740)	-	6.4 (0.8)	-	2.8 (0.4)	-	19.9 (1.2)	-	1.5	267 (64)	497 (92)
		2	-	498 (35)	-	1196 (231)	-	12.4 (2.1)	-	4.9 (1.5)	-	(23.7) (2.5)	-	67	445 (50)	1167 (246)
	40.0	1	-	350 (42)	-	561 (103)	-	6.4 (1.0)	-	3.0 (0.8)	-	19.8 (2.6)	-	25	274 (65)	518 (131)
	100	None		271 (6)	232 (270)	168 (4)	165 (41)	2.0 (0.1)	2.0 (0.5)	2.5 (0.1)	2.2 (0.2)	10.4 (0.4)	11.5 (1.5)	0	0	65 (23)

[†]CH₄ based

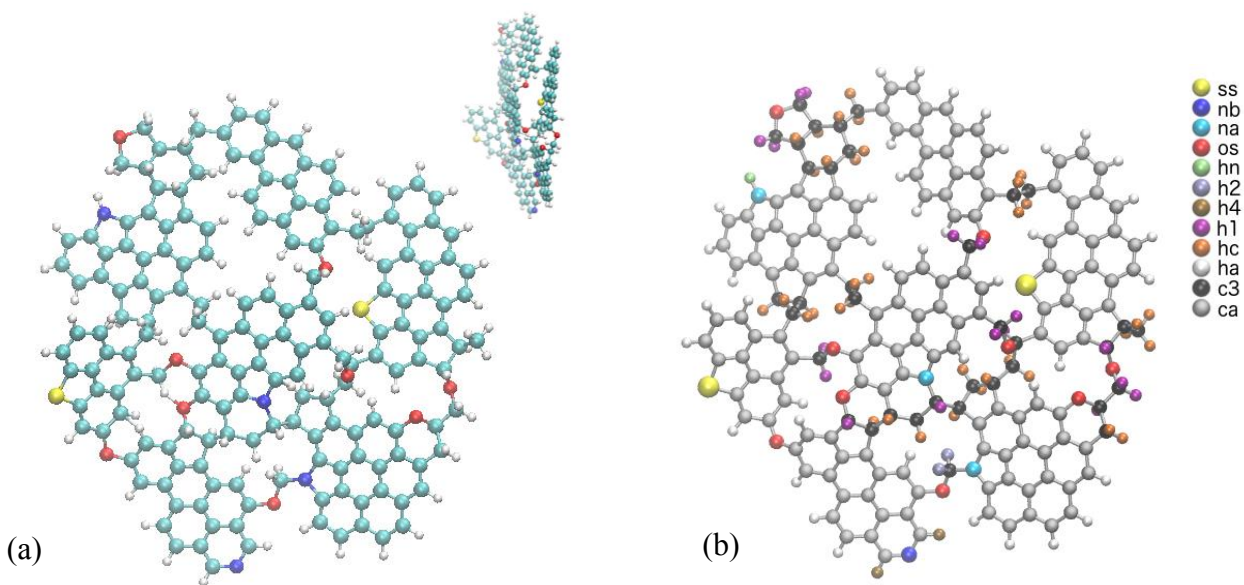


Figure 1: (a) Top and side (inset upper right) views of the molecular model for the II-D type kerogen. The molecular formula of II-D model is $C_{175}H_{102}N_4O_9S_2$ and its molecular weight is $2468.8 \text{ g mol}^{-1}$. It consists of five and six member fused rings assembled in a disk-like shape. Color code: green (C), white (H), blue (N), red (O) and yellow (S). (b) Model view where the GAFF force field atom types are displayed: black (c3), grey (ca), purple (h1), light blue (h2), ochre (h4), white (ha), light orange (hc), lime (hn), blue (nb), red (os), yellow (ss).

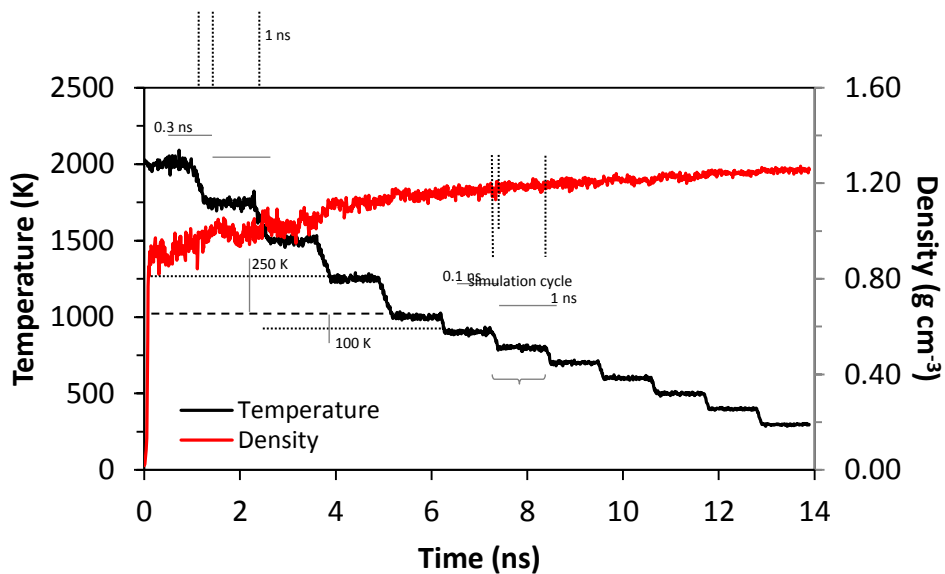


Figure 2. Staged cooling protocol for the construction of 15 II-D bulk kerogen models demonstrated for a run using GAFF. The system is cooled by repeated simulation cycles each one composed by two stages: a linear cooling stage followed by a NPT equilibration. As temperature is reduced to 1000K, each simulation cycle comprises a 0.3 ns cooling stage, with a rate equal to 833 K ns^{-1} . As we move from 1000 to 298.15 K, each cooling stage is 0.1 ns long and has a cooling rate of 1000 K ns^{-1} . Throughout this procedure, the NPT equilibration stages are 1 ns long. The density evolution along the run is also shown for comparison.

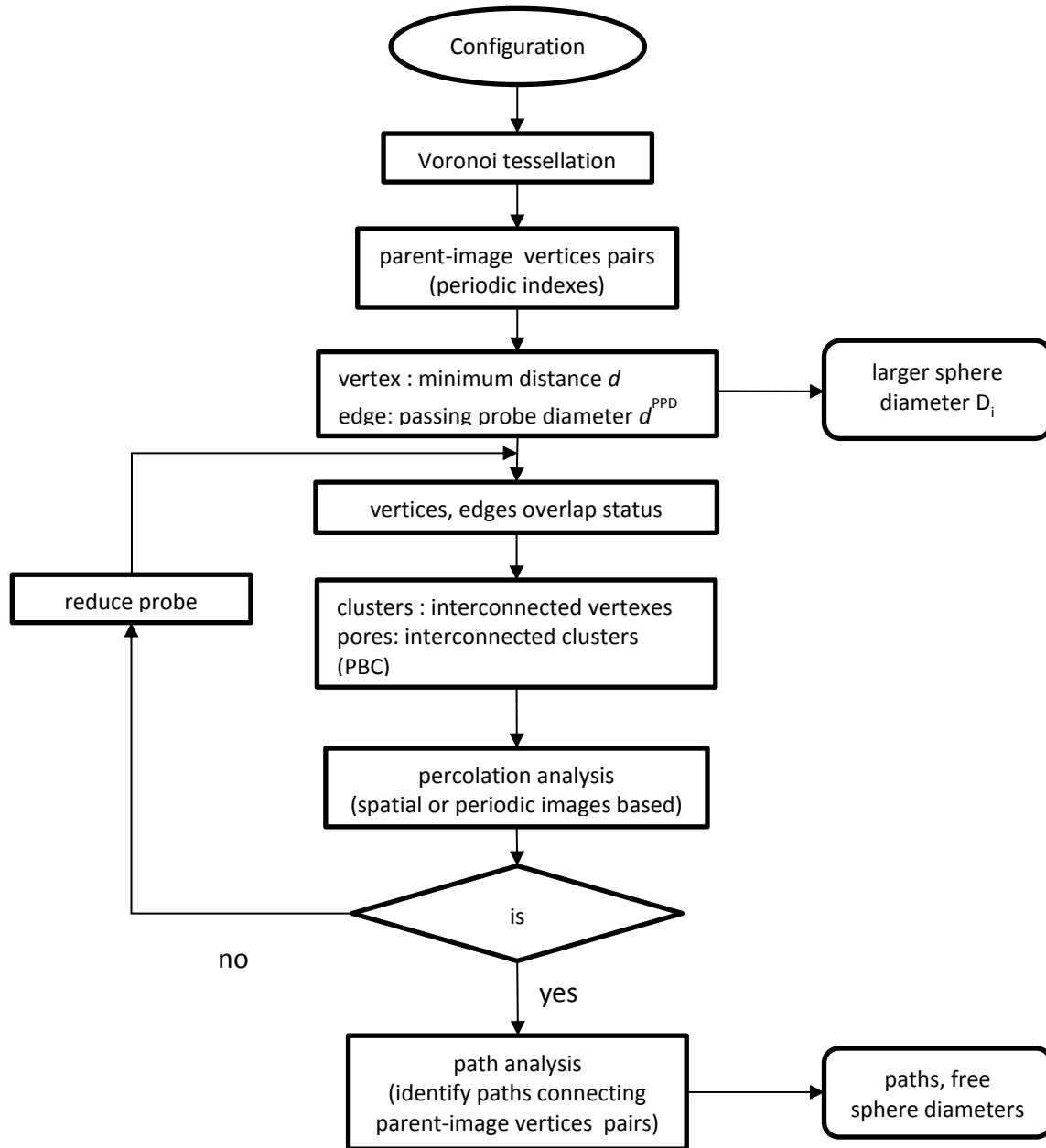


Figure 3: Flowchart of the methodology developed for the estimation and characterization of the accessible space in an amorphous material.

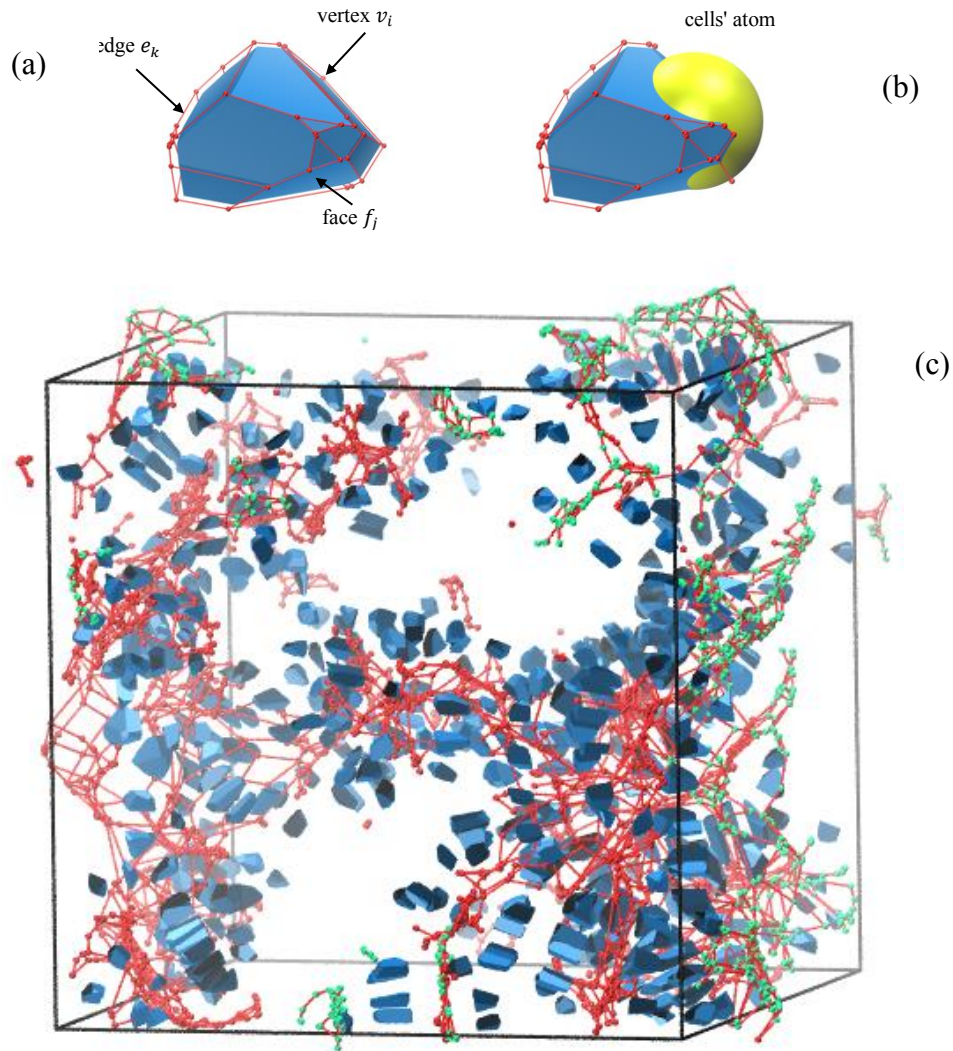


Figure 4: (a) An indicative polyhedron resulted from the Voronoi tessellation of the system. (b) The vertices, v_i , and the edges, e_k , of each cell that do not overlap with the atom that corresponds to the Voronoi cell are identified. The radius of the atom has been previously increased by the radius of the spherical probe considered. (c) The set of non-overlapping vertices and edges forms a graph/network, which topologically span the simulation box. Their corresponding Voronoi polyhedra are also drawn in blue (scaled down by an affine transformation with respect to their centers). Due to the periodic boundary conditions applied, pairs of parent-image vertices exist in the system (image vertices colored green here).

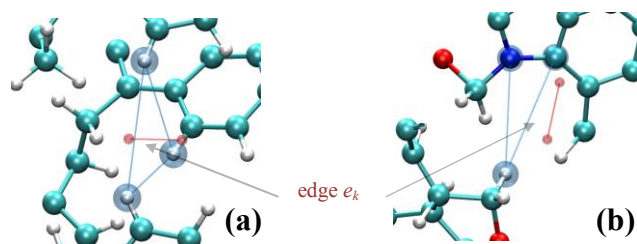


Figure 5: Each non-overlapping edge, e_k , (red line), is assigned with the passing probe diameter i.e. the maximum diameter of the spherical probe which is possible to pass from the one constituent vertex to the other depending on the relative position of edge with the atoms by which it is shared: (a) e_k intersects with the triangle formed by the atoms by which e_k is shared and (b) e_k does not intersect with the triangle created by the atoms whose cells share e_k .

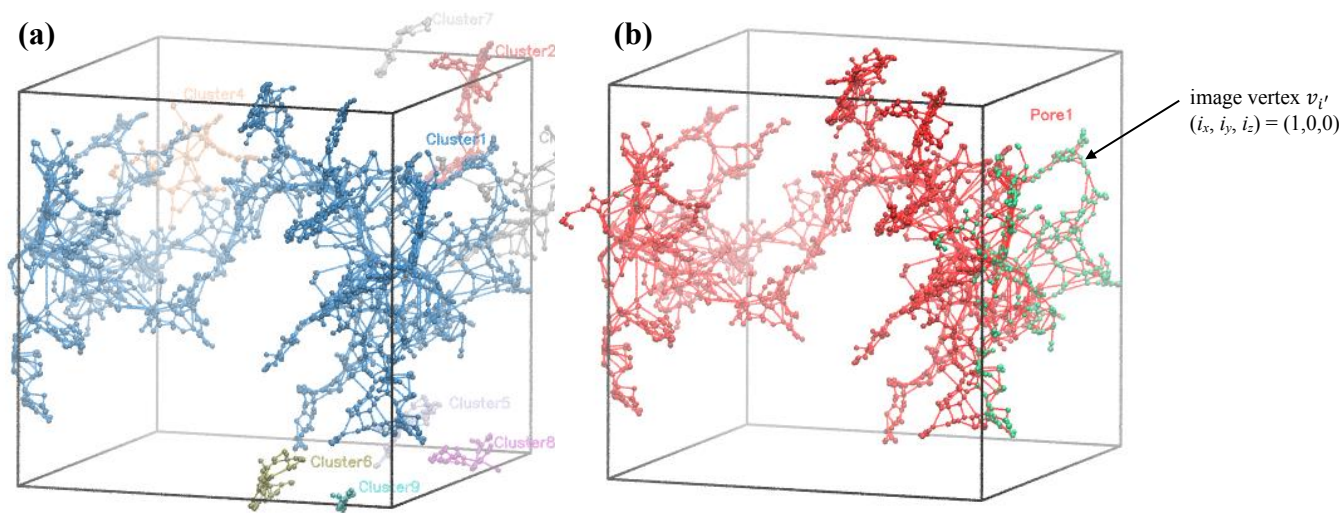


Figure 6: (a) A pore is a set of clusters (sub-graphs) connected through the periodic boundary condition applied in our system. (b) The constituent clusters are merged to reconstruct the corresponding pore. The resulted pore is percolated if image vertices (marked with green color) still exist while their periodic image indicates the direction of the channel.

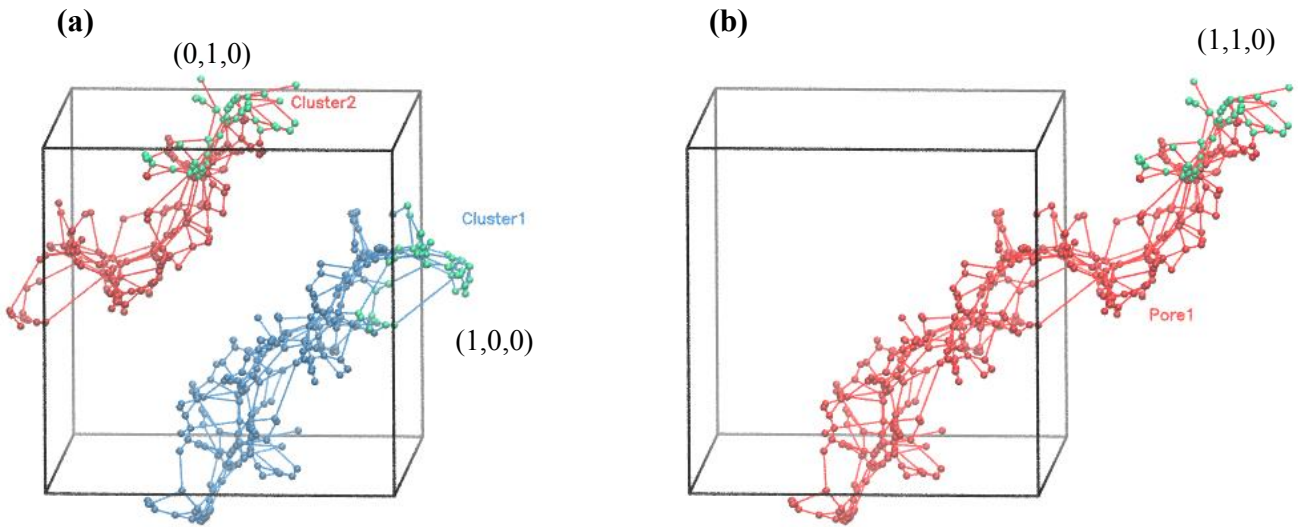


Figure 7: **(a)** Two clusters, Cluster1 (C_1) and Cluster2 (C_2) of accessible vertices in a fully periodic system. The image vertices of parent-image pairs are shown in green. Since the system is fully periodic, cluster C_1 is connected (forming a pore) with the periodic image of C_2 . This can be realized as an action of merging the two clusters (into a pore) after translation of C_2 by the vector separating the image vertices of C_1 by their parents in C_2 **(b)**. After merging the clusters, the periodic indices of the surviving image vertices are properly updated indicate the direction of the percolated channel.

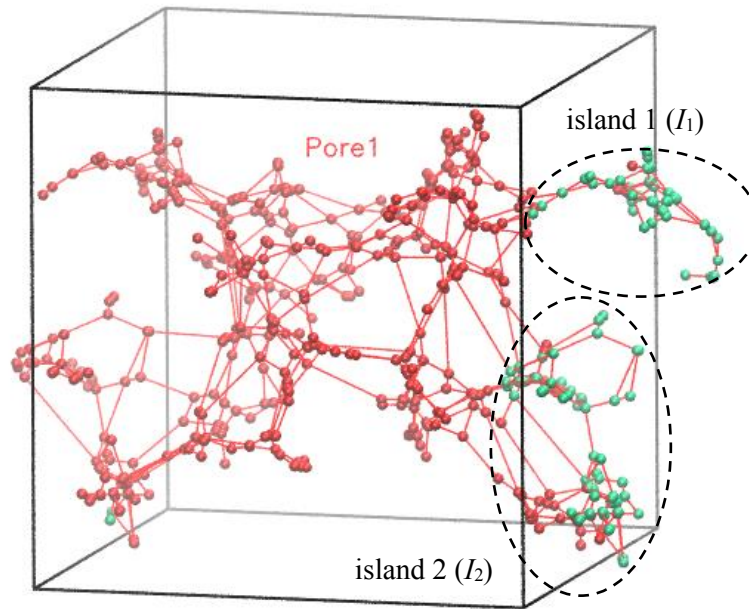


Figure 8: Image vertices (highlighted in green) of the pore are grouped based on the periodic indices identifying existing channels. Then, interconnected vertices of the same channel are aggregated and form islands.

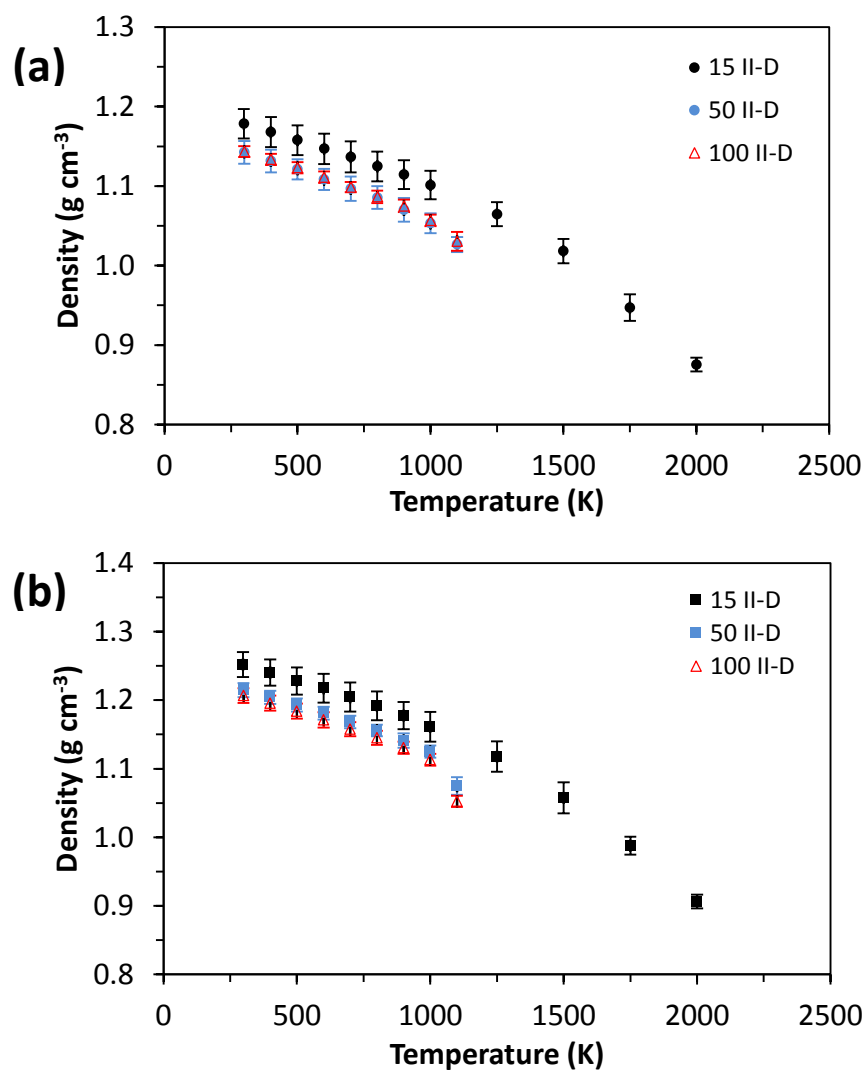


Figure 9. Density of kerogen structure as a function of temperature along the cooling procedure for (a) Dreiding and (b) GAFF force fields.

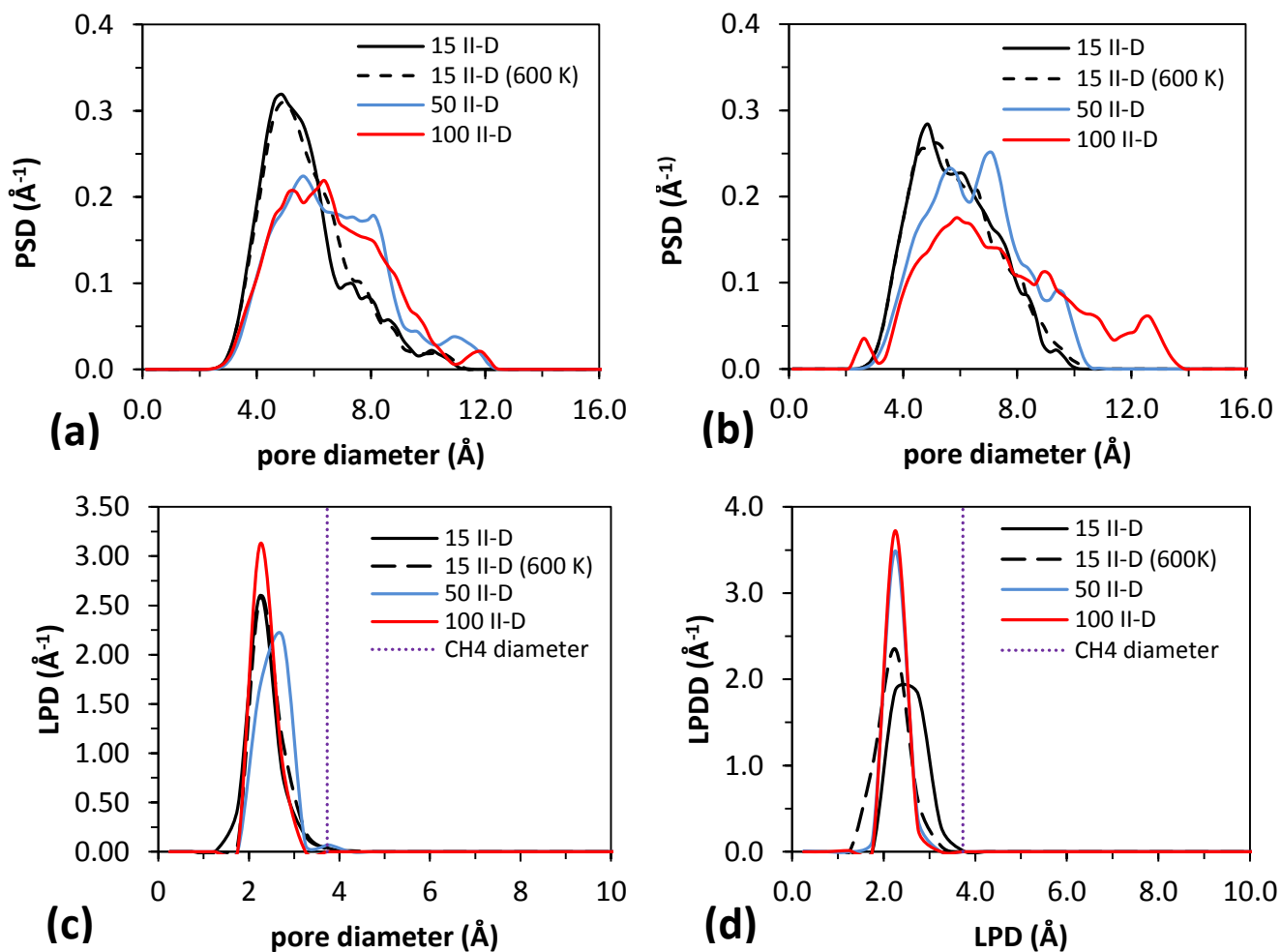


Figure 10 Average pore size (a and b) and limiting pore diameter distributions (c and d) for structures of different size (15 to 100 II-D molecules) at 298 K constructed using Dreiding (a, c) and GAFF (b, d) force fields. The PSD of the 15 II-D system at 600 K is also presented.

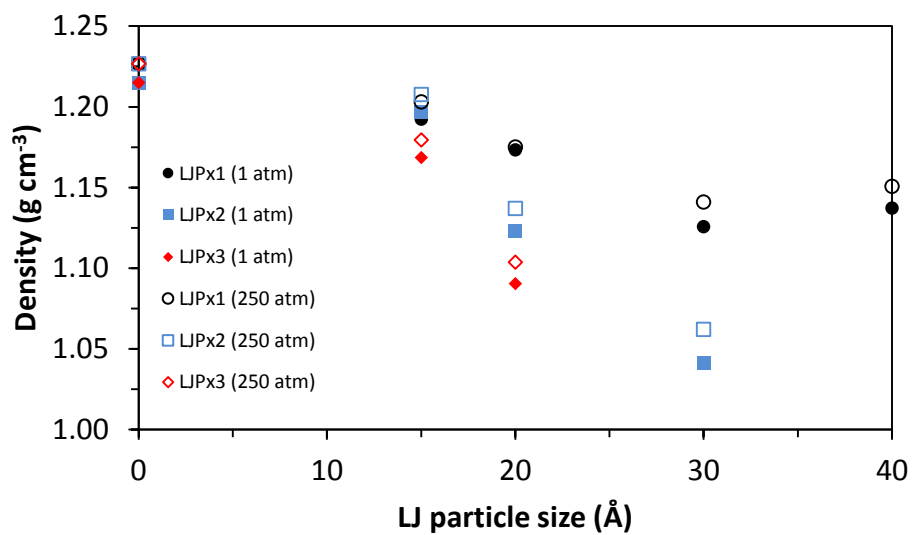


Figure 11. Density as function of number and size of LJ dummy particles (LJP) for runs with 50 II-D molecules using GAFF force field.

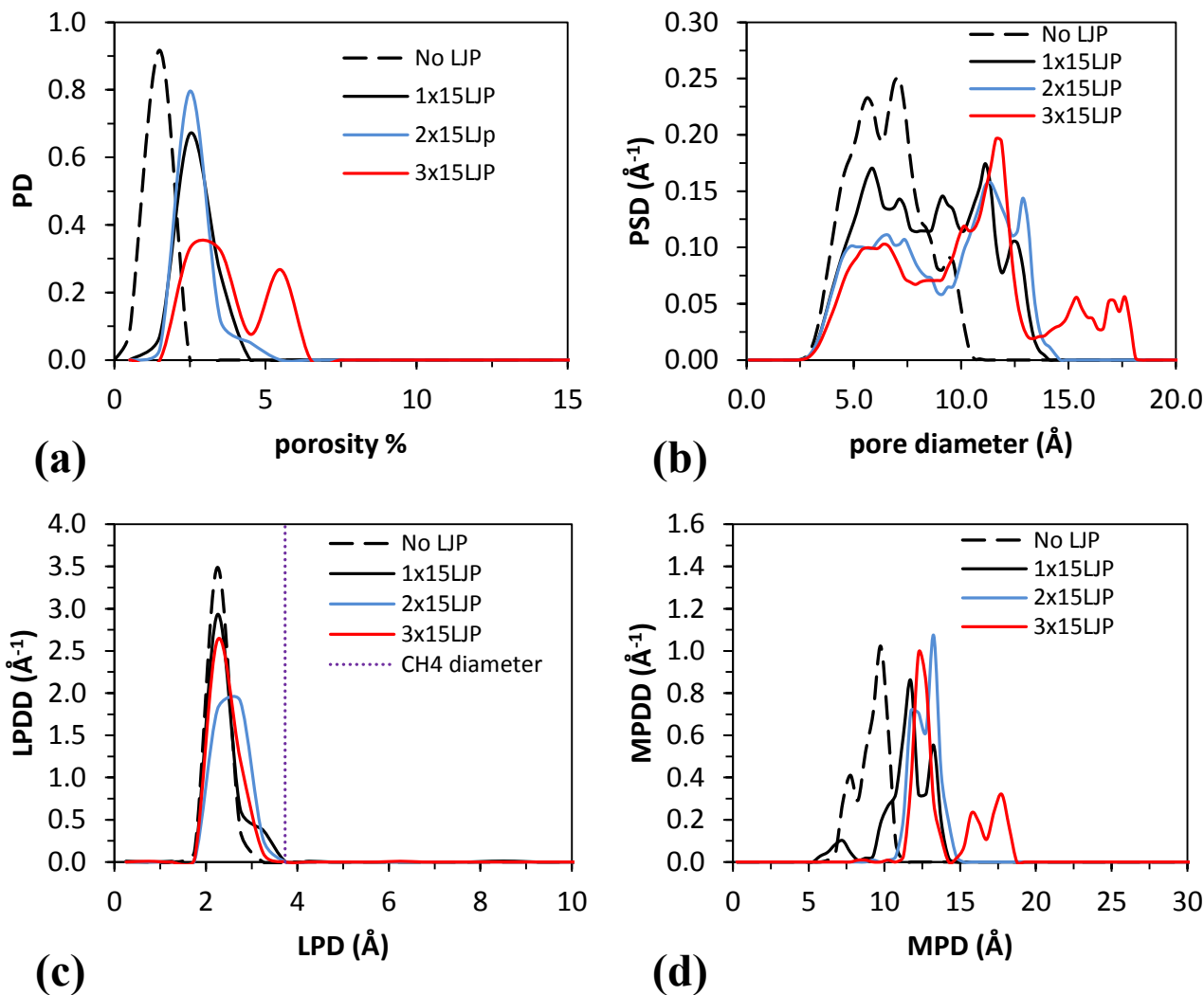


Figure 12. Accessible volume characteristics for 50 II-D bulk kerogen systems modeled using GAFF and constructed with the aid of different number of 15 Å LJ dummy particles. (a) PD, (b) PSD (c) LPDD and (d) MPDD.

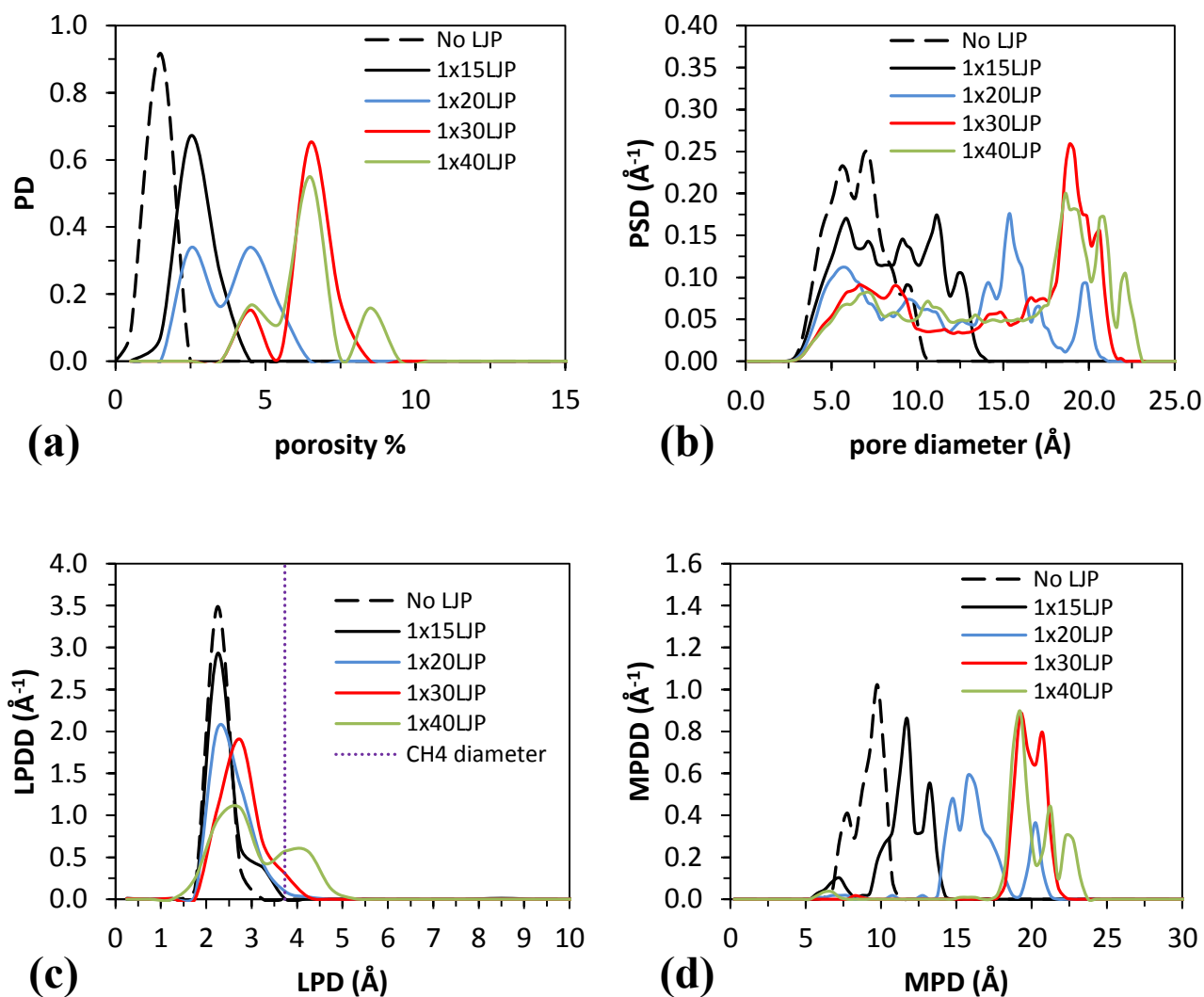


Figure 13. Accessible volume characteristics for 50 II-D bulk kerogen systems modeled using GAFF and constructed with the aid of varying size (15 to 40 Å) of a single LJ dummy particle. (a) PD, (b) PSD, (c) LPDD and (d) MPDD.

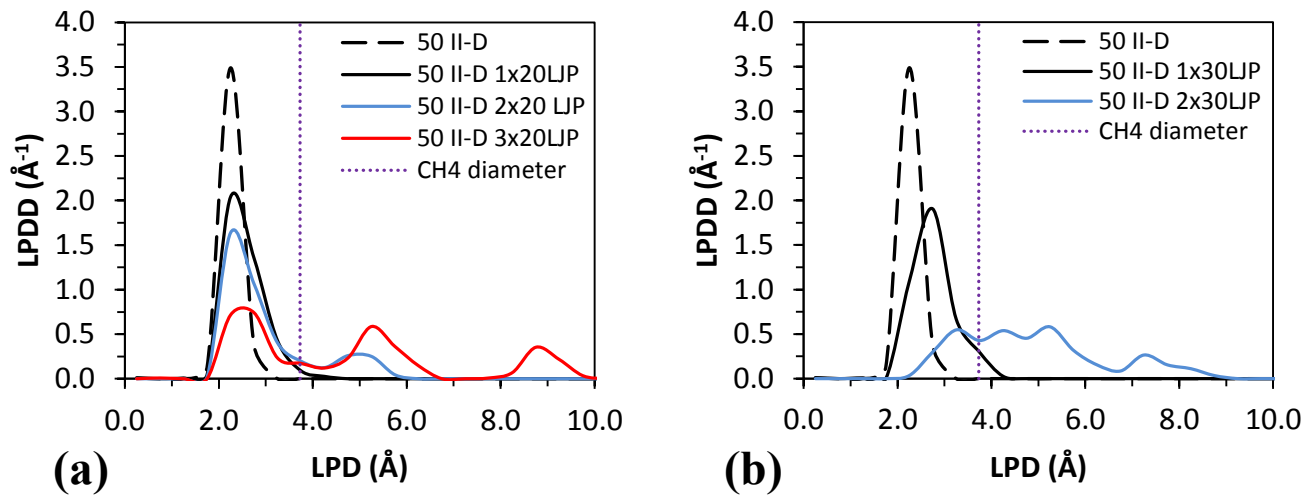


Figure 14. LPDD for different number of LJ particles of size (a) 20 and (b) 30 \AA .

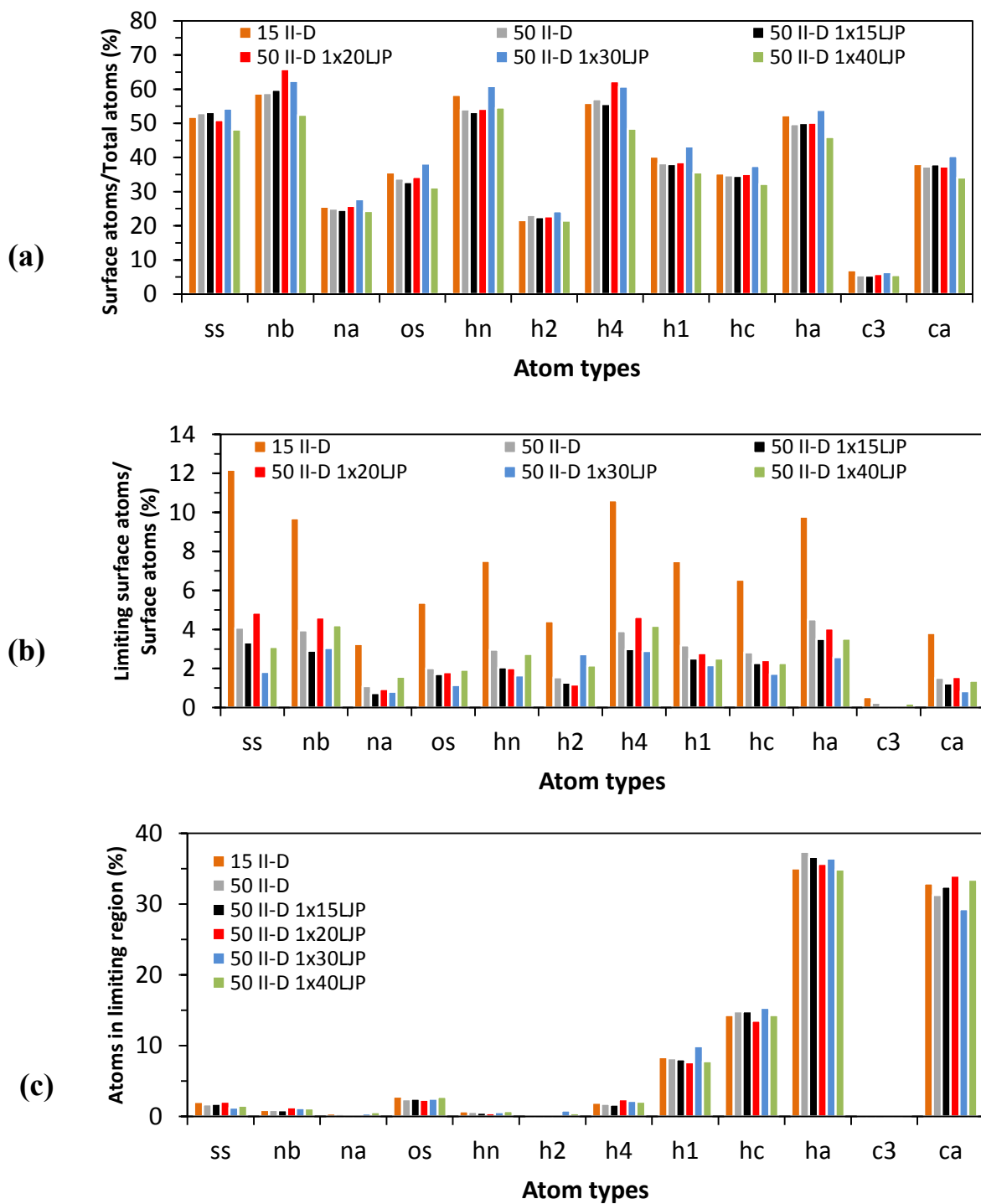


Figure 15. GAFF atom type analysis of the surface of the pores. (a) Percentage of atoms of each type that form the surface with respect to the total number of atoms of this type in the whole structure, (b) Percentage of the surface atoms that are found at the limiting regions of the pore (limiting regions are defined as a radius of $3.73 \cdot 1.2 = 4.476 \text{ \AA}$ around the pore). (c) Percentage of the atoms of the limiting regions that are of specific type.

References

- (1) Dresselhaus, M. S.; Thomas, I. L. *Nature* **2001**, *414* (6861), 332–337.
- (2) Bussell, I. S. *Oil shale developments*; Nova Science Publishers: New York, 2009.
- (3) Vandenbroucke, M. *Oil Gas Sci. Technol.* **2003**, *58* (2), 243–269.
- (4) Montgomery, S. L.; Jarvie, D. M.; Bowker, K. A.; Pollastro, R. M. *AAPG Bull.* **2005**, *89* (2), 155–175.
- (5) Alexander, T.; Baihly, J.; Boyer, C.; Clark, B.; Waters, G.; Jochen, V.; Le Calvez, J.; Lewis, R.; Miller, C. K.; Thaeler, J.; Toelle, B. E. *Oilfield Rev.* **2011**, *23* (3), 40–55.
- (6) Boyer, C.; Clark, B.; Jochen, V.; Lewis, R.; Miller, C. K. *Oilfield Rev.* **2011**, *23* (3), 28–39.
- (7) Bowker, K. A. *AAPG Bull.* **2007**, *91* (4), 523–533.
- (8) King, G. E. *J. Pet. Technol.* **2012**, *64* (04), 34–42.
- (9) King, G. E. Proceedings of the SPE annual technical conference and exhibition, Florence, Italy, 19-22 Spetember 2010; SPE paper ID: SPE-133456-MS.
- (10) Reeves, S. R.; Hill, D. G.; Hopkins, C. W.; Conway, M. W.; Tiner, R. L.; Mohaghegh, S. Proceedings of the SPE annual technical conference and exhibition, Houston, TX, 3-6 October 1999; SPE paper ID: SPE-56482-MS.
- (11) Eshkalak, M. O.; Aybar, U.; Sepehrnoori, K. *Pet. Sci.* **2015**, *12* (3), 553–559.

(12) SourceOECD (Online service); International Energy Agency; Organisation for Economic Co-operation and Development. *World energy outlook 2008*; OECD/IEA: Paris, 2008.

(13) U.S. Energy Information Administration (EIA)
<http://www.eia.gov/analysis/studies/worldshalegas/> (accessed Feb 10, 2017).

(14) Gadhamshetty, V.; Shrestha, N.; Chilkoor, G.; Bathi, J. R. In *ACS Symposium Series*; Drogos, D. L., Ed.; American Chemical Society: Washington, DC, 2015; Vol. 1216, pp 151–180.

(15) Bayramoglu, T.; Ari, Y. O. *Glob. Econ. Obs.* **2015**, 3 (2), 4.

(16) Jackson, R. B.; Vengosh, A.; Darrah, T. H.; Warner, N. R.; Down, A.; Poreda, R. J.; Osborn, S. G.; Zhao, K.; Karr, J. D. *Proc. Natl. Acad. Sci.* **2013**, 110 (28), 11250–11255.

(17) Olaguer, E. P.; Erickson, M.; Wijesinghe, A.; Neish, B.; Williams, J.; Colvin, J. *J. Air Waste Manag. Assoc.* **2016**, 66 (2), 173–183.

(18) Nelson, A. W.; Knight, A. W.; May, D.; Eitheim, E. S.; Schultz, M. K. In *ACS Symposium Series*; Drogos, D. L., Ed.; American Chemical Society: Washington, DC, 2015; Vol. 1216, pp 89–128.

(19) Frohlich, C.; Brunt, M. *Earth Planet. Sci. Lett.* **2013**, 379, 56–63.

(20) Clarkson, C. R.; Solano, N.; Bustin, R. M.; Bustin, A. M. M.; Chalmers, G. R. L.; He, L.; Melnichenko, Y. B.; Radliński, A. P.; Blach, T. P. *Fuel* **2013**, 103, 606–616.

(21) Gu, X.; Mildner, D. F. R.; Cole, D. R.; Rother, G.; Slingerland, R.; Brantley, S. L. *Energy Fuels* **2016**, 30 (6), 4438–4449.

- (22) Zhang, T.; Ellis, G. S.; Ruppel, S. C.; Milliken, K.; Yang, R. *Org. Geochem.* **2012**, *47*, 120–131.
- (23) Ross, D. J. K.; Marc Bustin, R. *Mar. Pet. Geol.* **2009**, *26* (6), 916–927.
- (24) Weniger, P.; Kalkreuth, W.; Busch, A.; Krooss, B. M. *Int. J. Coal Geol.* **2010**, *84* (3–4), 190–205.
- (25) Palmer, J. C.; Gubbins, K. E. *Microporous Mesoporous Mater.* **2012**, *154*, 24–37.
- (26) Segarra, E. I.; Glandt, E. D. *Chem. Eng. Sci.* **1994**, *49* (17), 2953–2965.
- (27) Liu, J.-C.; Monson, P. A. *Adsorption* **2005**, *11* (1), 5–13.
- (28) Biggs, M.; Agarwal, P. *Phys. Rev. A* **1992**, *46* (6), 3312–3318.
- (29) Ni, B.; Lee, K.-H.; Sinnott, S. B. *J. Phys. Condens. Matter* **2004**, *16* (41), 7261–7275.
- (30) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *J. Phys. Chem. A* **2001**, *105* (41), 9396–9409.
- (31) McGreevy, R. L.; Pusztai, L. *Mol. Simul.* **1988**, *1* (6), 359–367.
- (32) Opletal, G.; Petersen, T.; O'Malley, B.; Snook, I.; McCulloch, D. G.; Marks, N. A.; Yarovsky, I. *Mol. Simul.* **2002**, *28* (10–11), 927–938.
- (33) Bousige, C.; Boţan, A.; Ulm, F.-J.; Pellenq, R. J.-M.; Coasne, B. *J. Chem. Phys.* **2015**, *142* (11), 114112.

- (34) Bousige, C.; Ghimbeu, C. M.; Vix-Guterl, C.; Pomerantz, A. E.; Suleimenova, A.; Vaughan, G.; Garbarino, G.; Feygenson, M.; Wildgruber, C.; Ulm, F.-J.; Pellenq, R. J.-M.; Coasne, B. *Nat. Mater.* **2016**, *15* (5), 576–582.
- (35) Falk, K.; Pellenq, R.; Ulm, F. J.; Coasne, B. *Energy Fuels* **2015**, *29* (12), 7889–7896.
- (36) Obliger, A.; Pellenq, R.; Ulm, F.-J.; Coasne, B. *J. Phys. Chem. Lett.* **2016**, 3712–3717.
- (37) Behar, F.; Vandenbroucke, M. *Org. Geochem.* **1987**, *11* (1), 15–24.
- (38) Faulon, J. L.; Vandenbroucke, M.; Drappier, J. M.; Behar, F.; Romero, M. *Org. Geochem.* **1990**, *16* (4–6), 981–993.
- (39) Zhang, L.; LeBoeuf, E. J. *Org. Geochem.* **2009**, *40* (11), 1132–1142.
- (40) Siskin, M.; Scouten, C. G.; Rose, K. D.; Aczel, T.; Colgrove, S. G.; Pabst, R. E. In *Composition, Geochemistry and Conversion of Oil Shales*; Snape, C., Ed.; Springer Netherlands: Dordrecht, 1995; pp 143–158.
- (41) Orendt, A. M.; Pimienta, I. S. O.; Badu, S. R.; Solum, M. S.; Pugmire, R. J.; Facelli, J. C.; Locke, D. R.; Chapman, K. W.; Chupas, P. J.; Winans, R. E. *Energy Fuels* **2013**, *27* (2), 702–710.
- (42) Ungerer, P.; Collell, J.; Yiannourakou, M. *Energy Fuels* **2015**, *29* (1), 91–105.
- (43) Kelemen, S. R.; Afeworki, M.; Gorbaty, M. L.; Sansone, M.; Kwiatek, P. J.; Walters, C. C.; Freund, H.; Siskin, M.; Bence, A. E.; Curry, D. J.; Solum, M.; Pugmire, R. J.; Vandenbroucke, M.; Leblond, M.; Behar, F. *Energy Fuels* **2007**, *21* (3), 1548–1561.
- (44) Collell, J.; Galliero, G.; Gouth, F.; Montel, F.; Pujol, M.; Ungerer, P.; Yiannourakou, M. *Microporous Mesoporous Mater.* **2014**, *197*, 194–203.

- (45) Collell, J.; Ungerer, P.; Galliero, G.; Yiannourakou, M.; Montel, F.; Pujol, M. *Energy Fuels* **2014**, *28* (12), 7457–7466.
- (46) Yiannourakou, M.; Ungerer, P.; Leblanc, B.; Rozanska, X.; Saxe, P.; Vidal-Gilbert, S.; Gouth, F.; Montel, F. *Oil Gas Sci. Technol. – Rev. D'IFP Energ. Nouv.* **2013**, *68* (6), 977–994.
- (47) Collell, J.; Galliero, G.; Vermorel, R.; Ungerer, P.; Yiannourakou, M.; Montel, F.; Pujol, M. *J. Phys. Chem. C* **2015**, *119* (39), 22587–22595.
- (48) Ho, T. A.; Criscenti, L. J.; Wang, Y. *Sci. Rep.* **2016**, *6*, 28053.
- (49) Michalec, L.; Lísal, M. *Mol. Phys.* **2016**, 1–18.
- (50) Anovitz, L. M.; Cole, D. R.; Sheets, J. M.; Swift, A.; Elston, H. W.; Welch, S.; Chipera, S. J.; Littrell, K. C.; Mildner, D. F. R.; Wasbrough, M. J. *Interpretation* **2015**, *3* (3), SU59-SU70.
- (51) Emmanuel, S.; Eliyahu, M.; Day-Stirrat, R. J.; Hofmann, R.; Macaulay, C. I. *Mar. Pet. Geol.* **2016**, *70*, 175–184.
- (52) Okiongbo, K. S.; Aplin, A. C.; Larter, S. R. *Energy Fuels* **2005**, *19* (6), 2495–2499.
- (53) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
- (54) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94* (26), 8897–8909.
- (55) Greenfield, M. L.; Theodorou, D. N. *Macromolecules* **1993**, *26* (20), 5461–5472.
- (56) Alinchenko, M. G.; Anikeenko, A. V.; Medvedev, N. N.; Voloshin, V. P.; Mezei, M.; Jedlovsky, P. *J. Phys. Chem. B* **2004**, *108* (49), 19056–19067.

(57) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141.

(58) Sarkisov, L.; Harrison, A. *Mol. Simul.* **2011**, *37* (15), 1248–1257.

(59) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36* (22), 3219–3228.

(60) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.

(61) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627.

(62) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56* (5), 2257–2261.

(63) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr., J. E. P.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *GAUSSIAN 09, Revision D.01*; Gaussian, Inc, Wallingford, CT, 2009.

(64) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J Comput Chem* **1990**, *11* (4), 431–439.

- (65) Singh, U. C.; Kollman, P. A. *J Comput Chem* **1984**, *5* (2), 129–145.
- (66) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280.
- (67) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260.
- (68) Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Swails, J. .; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2016*; University of California, San Francisco, 2016.
- (69) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*, 2nd ed.; Computational science series; Academic Press: San Diego, 2002.
- (70) Ramos, J.; Peristeras, L. D.; Theodorou, D. N. *Macromolecules* **2007**, *40* (26), 9640–9650.
- (71) MAPS platform. Current version is 4.0 | Scienomics
<http://www.scienomics.com/products/molecular-modeling-platform> (accessed Sep 22, 2016).
- (72) Plimpton, S. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (73) Hoover, W. G. *Phys Rev A* **1985**, *31* (3), 1695–1697.
- (74) Shinoda, W.; Shiga, M.; Mikami, M. *Phys. Rev. B* **2004**, *69* (13).

- (75) Ewald, P. P. *Ann. Phys.* **1921**, 369 (3), 253–287.
- (76) Rexer, T. F.; Mathia, E. J.; Aplin, A. C.; Thomas, K. M. *Energy Fuels* **2014**, 28 (5), 2886–2901.
- (77) Foster, M. D.; Rivin, I.; Treacy, M. M. J.; Delgado Friedrichs, O. *Microporous Mesoporous Mater.* **2006**, 90 (1–3), 32–38.
- (78) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, 114 (25), 10024–10035.
- (79) Dodd, L. R.; Theodorou, D. N. *Mol. Phys.* **1991**, 72 (6), 1313–1345.
- (80) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, 149 (1), 134–141.
- (81) Rycroft, C. H. *Chaos Interdiscip. J. Nonlinear Sci.* **2009**, 19 (4), 041111.
- (82) Richard, P.; Oger, L.; Troadec, J. P.; Gervois, A. *Eur. Phys. J. E* **2001**, 6 (4), 295–303.
- (83) Gellatly, B. J.; Finney, J. L. *J. Non-Cryst. Solids* **1982**, 50 (3), 313–329.
- (84) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, 102 (14), 2569–2577.
- (85) Dijkstra, E. W. *Numer. Math.* **1959**, 1 (1), 269–271.
- (86) NetworkX <https://networkx.github.io/> (accessed Nov 11, 2016).