Replica-Exchange Molecular Dynamics Simulation of the Natural Evolution of a Model Type I Kerogen

Jean-Marc Leyssale, Pierre-Louis Valdenaire, Kevin Potier, Roland J.-M. Pellenq

Replica exchange molecular dynamics simulation of the natural evolution of a model type I kerogen

Jean-Marc Leyssale,*,†,‡ Pierre-Louis Valdenaire,‡ Kevin Potier,† and Roland J.-M. Pellenq‡,¶,§

†*Institut des Sciences Moléculaires, Univ. Bordeaux, CNRS UMR 5255, 351 Cours de la Libération, 33405 Talence, France*

‡*CNRS/MIT/Aix Marseille Univ. joint lab "MultiScale Materials Science for Energy and Environment", IRL 3466, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

¶*EpiDaPo Lab - CNRS / George Washington University, Children's National Medical Center, Children's Research Institute, 111 Michigan Ave. NW, Washington, DC 20010, USA*

§*Institut Europeen des Membranes, CNRS and University of Montpellier, 100 Av. du Prof. Jeanbrau, 34090 Montpellier, France*

E-mail: jean-marc.leyssale@u-bordeaux.fr

Abstract

Kerogen evolution in organic-rich sedimentary rocks, by its timescale and complexity, is a particularly challenging process for atomistic modeling and simulations. Yet, such modeling approaches, provided that appropriate simulation techniques are used and reasonable simplifications are made, can unravel the most significant physico-chemical mechanisms underlying oil and gas generation and the evolution of the kerogen

structure and properties. Here we combine reactive molecular dynamics and the replica exchange molecular dynamics simulation technique to simulate a simple model of type I organic matter, described as a mixture of five unsaturated fatty acids as a simplified model of algae, all along its maturation pathway, from diagenesis to late metagenesis. We describe fluid production, from the oil to gas windows, and the kerogen evolution in terms of structure and properties. We show that the formation of permanent microporosity within the kerogen matrix is intimately correlated to both the fluid production and the rigidification of the kerogen skeleton occuring in the gas window, i. e., at low O/C and H/C atomic ratios. We show that because of their pristine molecular structures with a more or less long tail made of alternating single and double carbon-carbon bonds, fatty acid molecules do decompose into alkanes of various length, up to octane in our simulations, by contrast to lignin and cellulose that only produces short alkanes, mostly methane

Introduction

Kerogen, which can both serve as a hydrocarbon reservoir[1,2] and as a possible host site for the geological sequestration of carbon dioxide,[3] as proposed in recent enhanced gas recovery (EGR) technologies, is thus key to important energy and environmental issues.[4,5] As shale hydrocarbons can be tightly adsorbed in kerogen micropores, and mesopores, efficient recovery in the EGR process strongly depends on the thermomechanical properties of the kerogen/gas system – competitive adsorption between CO_2 and hydrocarbons, adsorption induced-deformations, etc. – and the transport properties of the involved gas species.

A particularly sensitive aspect is the nature and properties of kerogen. Indeed, the term kerogen, defined as the insoluble organic matter (OM) present in sedimentary rocks,[6] actually covers a wide family of carbonaceous materials with broad ranges of chemical compositions and microstructures.[7] Therefore, kerogen is often described according to its type, related to its origin, and maturity which defines the degree of evolution of the OM on the path between

its parent form, and pure carbon. Three kerogen types are mostly considered as sources of shale oil and gas. Type I kerogen originates from lacustrine environment and is characterized by a large H/C atomic ratio and a low O/C ratio in the early stage of its evolution path. Type II is from marine environments and has intermediate values of the H/C and O/C ratios while type III kerogen, from terrestrial origin, has the largest O/C and lowest H/C.

Three stages characterize the natural evolution, or maturation, of kerogen:[7] diagenesis during which kerogen looses most of its oxygen by releasing water and other inorganic gases while establishing an insoluble, and mostly aliphatic macromolecular structure, catagenesis during which kerogen starts releasing hydrocarbons (oil then wet gas) and converts to a predominantly aromatic structure, and metagenesis during which the aromatic carbon structure consolidates and light gas (methane) is produced. Kerogen in the diagenesis-early catagenesis stage, containing a signfificant proportion of aliphatic carbon is generally termed as immature, while aromatic kerogen, in the late catagenesis-metagenesis stage are termed as mature or post-mature (late metagenesis).

Atomistic simulations have proven extremely valuable in the elucidation of gas/kerogen equilibrium (adsorption[8–10] and poromechanics[11,12]) and dynamic (diffusion,[13] transport[14–16]) properties, as well as in the prediction of important kerogen properties like density,[17] porosity[18] and stiffness,[19,20] in the last decade. Yet, as discussed in a recent review, the assumptions made when creating the atomistic representation of kerogen may have a significant influence on the obtained results.[21] The most widely used kerogen models, the so-called "molecular" models in ref. 21, are obtained by packing a set of identical "kerogen" macromolecules as those proposed by Ungerer et al.,[17] constructed for various kerogen types and maturities on the basis of NMR and XPS data.[22] However, this is a rather strong preassumed constraint and the considered size for the molecules – 2400-3900 Da in Ungerer's models – is even below the accepted weight range, 5000-10000 Da, defining the insolubility limit.[7] Slightly later, type II kerogen models at various maturities and densities were built based on diffraction data using the hybrid reverse Monte Carlo method.[19] It was found that,

unlike for the molecular models, at the scale of the molecular simulation cell (5 nm), kerogen consists in a unique 3D connected network of covalent bonds, that percolates through the periodic boundaries of the cell.

Mimetic approaches in which kerogen and hydrocarbon formations and evolutions can be simulated directly from the parent organic matter, or from a low maturity state, are particularly attractive as they can in principle give access to the mechanisms underlying porosity and stiffness development, in addition to providing a dynamic path for the kerogen evolution, together with its associated fluid phase. However, because of the drastic difference in time scales between the geological timescales of kerogen evolution (million years) and the typical times accessible to reactive molecular dynamics (RMD) simulations (tens of nanoseconds at most), direct RMD simulation under geological temperatures and pressures remains inoperative. Some authors have considered performing simulations at much higher temperatures (> 2000 K) to compensate for the timescale difference.[23,24] Qian et al.[23] reported on the non-isothermal pyrolysis of the Siskin Green River kerogen model[25] using RMD simulations, progressively ramping up the temperature from 500 to ~ 3000 K. They observed the production of hydrocarbons, from about 2000 K, up to complete conversion of the kerogen "molecule" into gas. It has to be noted than such a gasification process is drastically different than the process of kerogen maturation. Similarly, Han et al.[24] used RMD to investigate the evolution of three molecular models of kerogen (from ref. 17) under isothermal pyrolysis at a series of temperatures up to 2600 K, reproducing reasonably well the van Krevelen diagram of type I, II and III kerogens. However, the authors noticed that the size of molecular fragments in their system decreased with increasing pyrolysis temperature and used the molecular fragments counting 17-40 C atoms only as constituents of kerogen. This clearly conflicts with the definition of kerogen as a solid material of increasing molecular weight (due to cross-linking) along maturation.[26] These works should thus not be considered as investigations on kerogen evolution.

To overcome the timescale limitation inherent to RMD simulations, Atmani et al. pro-

posed to couple RMD with the replica exchange molecular dynamics (REMD) framework. In REMD, the configurational sampling at a temperature of interest, which can be dramatically slow using conventional MD, is accelerated by incorporating states from simulations at higher temperatures, considerably enhancing the convergence of the simulation towards its equilibrium state. Applying this method to kerogen formation and evolution from simple models of type III OM – cellulose[27] and lignin[28] – atomistic models of kerogen were obtained from immature to very late metagenesis states,[29] with a good agreement with experimental kerogen yield, compared to the char yields obtained in pyrolysis experiments from the same precursors, fluid production and kerogen evolution pathway. To our knowledge, this constitutes so far, the only reliable effort to simulate kerogen evolution at the nanoscale.

In this work we present a REMD investigation for the geological evolution of a simple model of type I OM, which demonstrates more complexity in the evolution of both fluid content and kerogen structure, during maturation. We also discuss in details the build-up of the kerogen porosity and stiffness along the maturation pathway, which is subject to contradictory results in the literature. Indeed, the effects of kerogen maturity and density on stiffness have already been investigated by Bousige et al., using HRMC kerogen models,[19] and by Kashinath et al., using molecular models.[20] Both studies have shown that the dominent factor governing kerogen stiffness is density, with a modulus that increases with increasing density. This correlation was also found in the ultrasonic experiments of Yan et al. for Green River (type I) kerogen samples, who found bulk modulus increasing from 3.5 to 5 GPa when density increases from 1.05 to 1.25 g/cm^3.[30] The correlation between maturity and stiffness is less clear in recent literature. Investigations using molecular models show that at fixed density, stiffness decreases with increasing maturity,[20] which is somehow counterintuitive, while investigations using HRMC models show a significant increase in stiffness only at a very post-mature state (H/C $\sim$ 0).[19] Experimental data by Zargari et al.[31] indicates that the modulus is highest in the immature state (onset of oil generation), passing through a minimum in the mature state before increasing slightly in the over-mature (gas window)

state, yet data regarding density are not given. Conversely, other authors have observed a continous increase in (type II) kerogen stiffness with maturity, both exploring natural samples[32] and artificially matured kerogen.[33]

Methods

REMD simulation: principle

The REMD method consists in simulating in parallel many replicas of the same system at different temperatures, generally in the canonical (NVT) ensemble. Doing so, every replica tends towards its own equilibrium distribution of microstates at its given temperature. As schematized in Fig. 1(a), this leads to different distributions of, for instance, the potential energy, which shifts towards larger values as T increases. At the core of the REMD algorithm is a stochastic process that allows for exchanging the configurations (atomic positions) and velocities of replicas ran at neighboring temperatures. In this way, rare events, associated to the crossing of large (free-) energy barriers, can be efficiently sampled at high temperatures, and as pictured in Fig. 1(b), the obtained configurations can be progressively incorporated to the low temperature replica through successive exchanges. Note that for convenience we identify a replica to a constant T "trajectory" (horizontal lines in Fig. 1(b)), instead of the definition of continuous dynamics, albeit varying temperature (constant color in Fig. 1(b)), used by some authors.[34] In line with the formerly introduced parallel tempering Monte Carlo technique (PTMC),[35] it was shown that the canonical distribution is preserved in all replicas provided that configuration exchanges are accepted with the Metropolis-like criterion given in Eq. 1:[36]

$$A_{ij} = min\left\{1, exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(\mathcal{U}_i - \mathcal{U}_j)\right]\right\} \quad (1)$$

In this equation k_B is Boltzmann's constant, T_i and T_j are the temperatures, and $\mathcal{U}_i$ and $\mathcal{U}_j$ the potential energies, of two adjacent replicas i and j, respectively. Eq. 1 clearly highlights

an important feature of the REMD algorithm, namely that the acceptance rate of the swaps will strongly depend on the degree of overlap between the energy distributions of adjacent replicas (Fig. 1(a)).

It is also important to note that because of the lack of dynamic continuity in the replicas, due to exchanges, REMD has to be considered as a Monte Carlo simulation. In other words, the visited configurations are not ordered in time, and the REMD time is only a metric to measure the evolution of the system along the simulation, just like the MC steps or cycles in Metropolis Monte Carlo simulations. Therefore, no time-dependent properties can be produced from the analysis of REMD simulations.

Few adjustable parameters, listed in what follows, allow tuning the efficiency of a REMD simulation. The most trivial ones are the lower and upper replica temperatures, T_{low} and T_{up}, respectively. T_{low} is generally chosen as the temperature of interest, while T_{up} is set to be sufficiently high such that rare events are observed within a fraction of the total simulation time. More complex are the choices regarding the number of replicas n_{rep} and their spacing in temperature. Generally speaking, we aim at minimizing the simulation time needed to exchange configurations from the highest to lowest temperatures. This first implies that the adopted temperature spacing provides a similar degree of overlap for the potential energy distributions (Fig. 1(a)), defining the average acceptance ratio, over the entire replica path. The number of replicas needs to be adjusted such that this degree of overlap is sufficient to have a reasonable acceptance ratio, some authors mentioning 20 % as an ideal case scenario,[34] yet not too large to avoid increasing the complexity of the path (the more the number of replicas, the longest the exchange path), which would result in longer end-to-end exchange times, despite larger acceptance rates. In common systems, the average potential energy increases linearly with T ($3Nk_BT/2$, where N is the number of atoms, for an harmonic system), and the width of the energy distribution, assumed to be Gaussian with a standard deviation $\sigma_{\mathcal{U}}$, increases linearly with T and with the square root of N (assuming a constant

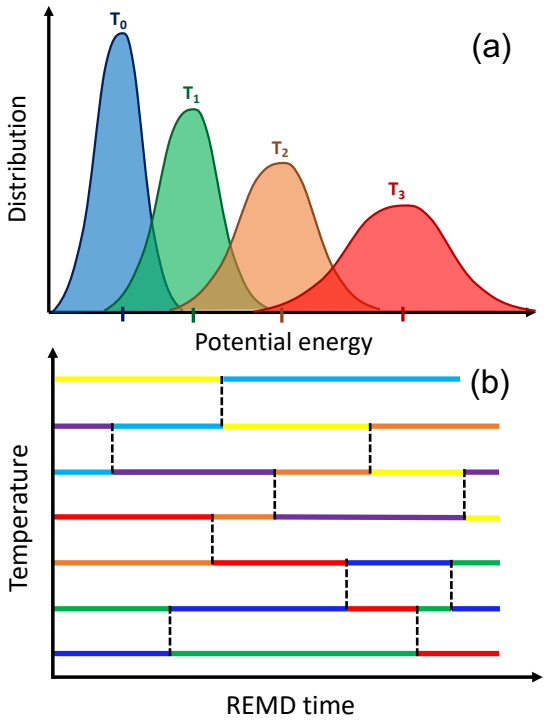

Figure 1: Schematics of the REMD method. (a) Sketch of the potential energy distribution at different temperatures, showing the broadening of the distributions with increasing T. (b) Sketch of the exchange dynamics showing that during REMD a continuous trajectory (displayed with a unique color) travels through different temperatures via a series of T exchange. The initial configuration (for exemple, the dark blue line) can thus be brought to higher temperatures, where it undergoes several chemical reactions, and then travel back to low temperatures. Here for convenience of the definition, we define a replica as a state of constant temperature (i.e. horizontal lines in (b)) which appears as a trajectory with successive configuration exchanges with replicas at close T.

heat capacity C_V):[37]

$$\sigma_{\mathcal{U}} = T\sqrt{k_B\left(C_V - \frac{3}{2}Nk_B\right)}. \quad (2)$$

In this situation, a spacing such that $T_{i+1}/T_i = \alpha$, where α is a constant (>1) is recommended,[34] with n_{rep} (determining the value of α) chosen such that the acceptance rate of configuration swaps is about 20 %. The choice of the exchange attempt frequency is arbitrary, yet it has to be frequent enough to ensure a reasonable characteristic time for the end-to-end exchange timescale, which has to remain a fraction of the total simulation time, yet should, in theory, be lower than the characteristic frequency of the thermostats used to fix the replicas temperatures. In practice, as the velocities are rescaled to fit with the current replicas temperatures after each successful exchange,[36] canonical sampling is expected to be achieved almost whatever the exchange frequency. Another well-known method in which a canonical distribution is produced using MC updates based on MD trajectories of arbitrary duration (and accuracy) is the so called hybrid Monte-Carlo method, also called as Hamiltonian Monte Carlo.[38]

Another important feature to mention is that as for conventional MD and MC methods, which have been initially designed to sample equilibrium distributions, REMD is also often used to actually charecterize the evolution of a system, initially out-of-equilibrium, possibly in a metastable state, towards its equilibrium. In this case, which is the situation covered in the manuscript, the REMD "trajectory", as would a MC "trajectory", provides a thermodynamic path of decreasing free energy towards the equilibrium state (state of minimum free energy).

Kerogen precursor

A mixture of five unsaturated fatty acids, listed in Fig. 2 was used to model algaenan organic matter.[39] We first placed manually one exemplar of each molecule in a 3D periodic cubic box with a sufficiently large lattice parameter (initial density of ~ 0.1 g/cm^3) to avoid

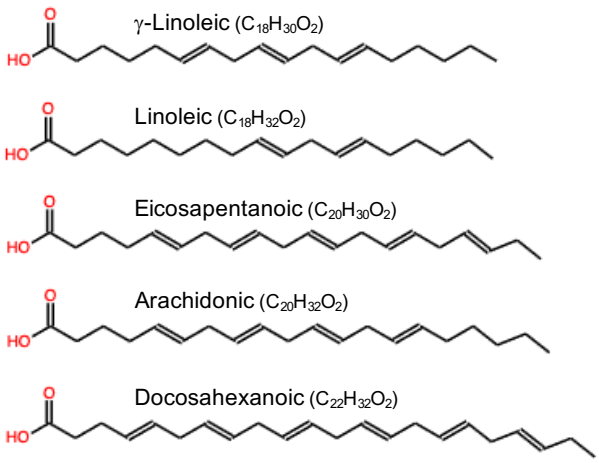

Figure 2: Molecular structure of the five fatty acids used as a simplified model of algaenan organic matter.

overlaps between the molecules. This box was then replicated three times in each Cartesian direction (twenty seven copies of each molecule). This corresponds to a system of 2646 C atoms, 4212 H atoms and 270 O atoms with H/C = 1.59 and O/C = 0.1, i.e. typical of early diagenesis type I OM.[21] This system was chosen to be as large as possible, to account as much as possible of the complexity of the evolution process, while still keeping computationally tractable, knowing that computational cost increases considerably with system size.

MD and REMD simulation details

MD and REMD simulations are performed with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package[40] using the Atmani ReaxFF force field[27] used in former works.[27–29,41] A detailed description of the forcefield, including the full list of parameters in LAMMPS format is available in Ref. 41. The equations of motion are integrated using the velocity-Verlet integrator[42] with a time step of 0.1 fs and the temperature and pressure are controlled using a Nosé-Hoover thermostat (both MD and REMD) and a Nosé-Hoover-Andersen barostat (MD only), operating with time constants of 50 and 250 fs, respectively.[43]

The system was first equilibrated using MD in the canonical ensemble – i.e under fixed

number of atoms (N), temperature (T) and volume (V) – for 20 ps at a temperature of 423 K. Then, it was compressed to a pressure of 50 MPa at the same temperature in the isothermal-isobaric – NPT, where P stands for pressure – ensemble. These conditions correspond to the temperature and pressure at $\sim$ 5 km depth. The 100 ps long MD simulation was sufficient to stabilize the energy and density of the cell.

The obtained liquid phase, having a density of 1.19 g/cm^3, was then subjected to REMD for 10 ns. In this simulation, 120 replicas were used, covering temperatures ranging from 423 to 3500 K, with a spacing scheme such that $T_{i+1}/T_i = 1.0179$. Attempts to exchange configurations from replicas operating at closeby temperatures were performed every 10 fs. These adopted number of replicas and spacing scheme allowed to reach an average acceptance rate of about 20-25 % for the exchange attempts, i.e. close to the recommended value of 20 %. We acknowledge that the lag time between swap attempts is shorter than the relaxation time of the thermostat (50 fs), yet this was absolutely necessary as reducing the exchange frequency would have increased the REMD time needed to achieve the same OM evolution path. This was nos possible, considering that the 10 ns long simulation presented here used about 8×10^6 CPU hours. Nevertheless, the rescaling of velocities after every successful exchange allows maintaining a proper dynamical and thermodynamic behavior in the replicas. We indeed checked that the proper canonical distributions of velocity components and kinetic energy were not affected by this high exchange frequency. The distribution of the instantaneous temperature in the 423 K replica (Fig. S1), shows the expected Gaussian distribution around the target value.

Conversely to what was performed in Atmani et al.,[27,28] where the REMD was interrupted at intervals to relax pressure using MD in the NPT ensemble, here, as in Leyssale et al.,[29] the REMD simulation, which showed a slower evolution compared to the case of type III OM precursors, was entirely performed in the NVT ensemble (i.e. without interruption).

Results analysis

Atomistic configurations of the 423 K replica were stored every 28 ps during the REMD simulation and analysed as described in former work.[29] First, the different molecular entities constituting each configuration are identified using a standard clustering algorithm based on bond distances. First, a neighbor list is formed to determine all the chemical bonds in the system based on the following cutoff distances: 1.9 Å (C-C, C-O and O-O), 1.5 Å (C-H), 1.35 Å (O-H) and 1.1 Å (H-H). An individual molecule ID (integer value) is attributed to all the atoms. Then, looping over the neighbor list, every time two bonded atoms have different molecule IDs, the lowest ID number is attributed to both atoms. The loop process is repeated until no more change in ID is observed. A this point, all the molecules have been determined with a unique ID. Independent molecules are then classified according to their constituting number of carbon atoms as gas (0-4), light tar (5-13), heavy tar (14-40) and kerogen (> 40). Note that the used definition for kerogen was mostly adopted to comply with simulation literature, yet remains somehow unrealistic as the actual constituents of kerogen should have masses at least two orders of magnitude larger to reach insolubility. For each configuration, the obtained list of molecules can be straightforwardly converted into a mass distribution (i.e. number of molecules of mass m), or spectrum. Then, all light species, including gas, light tar and heavy tar, were removed from the kerogen and the latter was relaxed at 423 K and 50 MPa using MD in the NPT ensemble for 5 ns. Equilibrium properties were reached in about 2 ns and the remaining 3 ns were used to compute average properties.

Investigated properties include the bulk modulus K characterizing kerogen stiffness, computed from the volume fluctuactions:

$$\frac{1}{K} = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle}, \tag{3}$$

where k_B is Boltzmann's constant and $\langle ... \rangle$ indicate time averages. The total geometric

porosity and pore size distributions (PSDs) were computed, as in ref. 29, using a grid of 0.1 Å lattice spacing, and averaged over five independent configurations. Assuming van der Waals diameters of 3.36, 3.17 and 2.42 Å for the C, O and H atoms of the kerogen, the total (probeless) porosity is defined as the fraction of the grid points that are not overlapping with the kerogen atoms. For every grid point that was found to belong to the porosity, the pore size is defined as the largest spherical pore that can contain this grid point without overlapping with a matrix atom. For this, tentative pore centers are thrown randomly in the neighborhood of the considered grid point until 500 successive attempts are reached, the pore size at the grid point is then given by the diameter of the largest sphere. PSD is then obtained as the distribution of the pore size over all the grid points in the porosity. A detailed description of the algorithm is described in Ref.,[28] the only difference is that, here, we adopt a probe radius equal to zero.

The kerogen models are also analyzed in terms of composition (O/C and H/C atomic ratios), and of the nature of their carbon backbone. This includes the standard determination of the coordination number, based on the previously introduced bonding distances, and the characterization of medium-range order, characterizing the carbon (and oxygen) backbone.[27–29] The first step here is the determination of ring statistics according to the shortest path ring definition of Franzblau.[44] Then, ring clusters are determined on the basis that two bond-sharing rings belong to the same cluster. Finally, C and O atoms that are not involved in rings are defined as "chain" atoms, and clustered into chain fragments: two bonded chain atoms being defined as belonging to the same chain fragment. An iterative algorithm, similar to the one used previously to identify the molecules present in the system, is used to determine ring clusters and chain fragments.

Finally, X-ray diffraction (XRD) patterns were computed with Debyer,[45] using the last kerogen configuration of the relaxation simulations. The Cu K_α radiation wavelength (0.1542 nm) was adadopted to convert scattering vectors (Q) to diffraction angles (2θ). All the analyses but the XRD calculations were performed with in-house codes.

Results

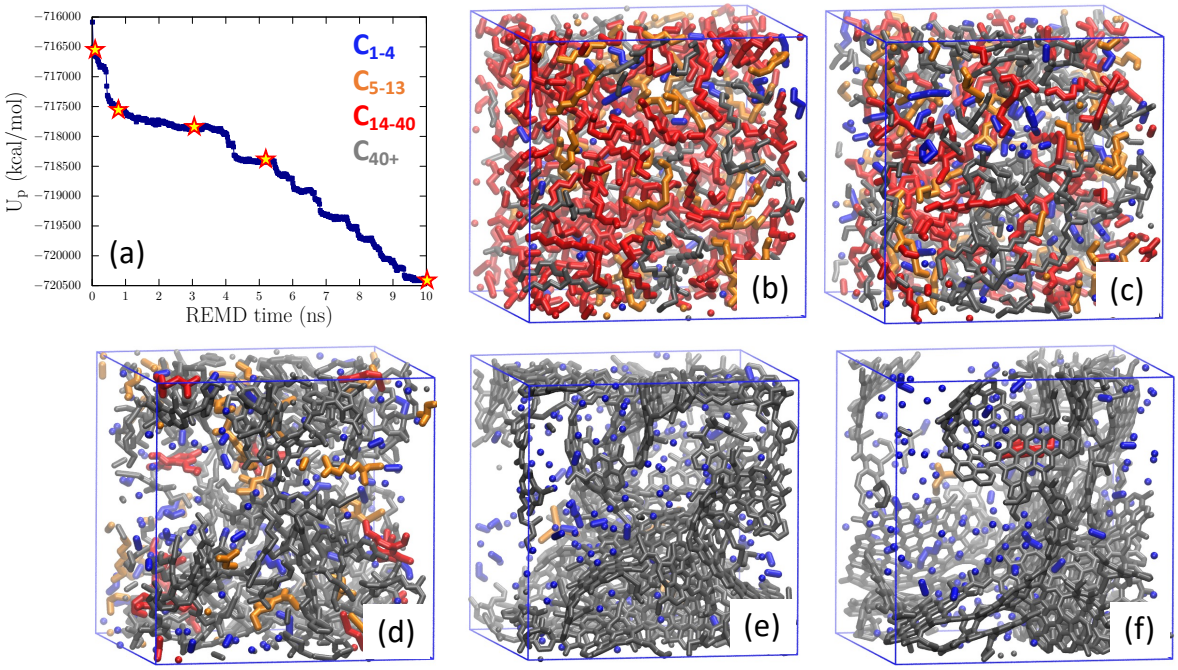

Figure 3: (a) Evolution of the potential energy U_p as a function of simulation time. (b-f) Snapshots of the system at (b) 0.17, (c) 0.7, (d) 2.9, (e) 5.2 and (f) 10 ns, showing the contributions of gas (blue), light tar (orange), heavy tar (red) and kerogen (grey). For clarity, only the organic C atoms are shown.

Fig. 3 shows the evolution of the total potential energy U_p of the system, and some selected snapshots, during the REMD simulation. As in former reports,[27–29] a significant transformation occurs within the OM. The mixture of fatty acids progressively converts into both smaller molecular entities, containing hydrocarbons and inorganic molecules (not shown), and larger species attributed here to kerogen. The process is clearly exoenergetic, all along the simulation, and the respective amouts of kerogen and fluid seem relatively stable above 1 ns. However, a clear transition takes place between early states, characteristic of the oil window (before 1 ns), and states in the gas window (above 3 ns). This transition is concomitant to the evolution of the kerogen structure from immature (aliphatic) to mature (aromatic). Some porosity which appears here as empty, clearly visible in the snapshots at high maturity, actually corresponds to the presence of inorganic gases, mostly water and

hydrogen, which are not shown for clarity.

Overall, the total amount of heat released during the 10 ns long simulation is of about 4500 kcal/mol, corresponding to 0.11 kcal/g. This can be compared to the 0.50 kcal/g obtained with lignin as a proxi for type III OM, during a simulation of same duration and yielding to a similar advancement in the kerogen evolution process.[29] The significantly larger heat production during type III OM evolution can be associated to a much larger oxygen content in the OM, and thus to a larger amount of produced water, which is exothermic.[27]

A consequence of performing the simulation at fixed volume, mimicking closed system pyrolysis, is that pressure increases significantly during the process, up to about 900 MPa at the end of the simulation (Fig. S2). This results from the production of fluid and its progressive evolution to an increasing number of small molecules. We assume here, as in former work,[29] that it has a limited effect on chemistry. Yet it may significantly influence the density and porosity of the obtained kerogen. This topic is addressed in the discussion section.

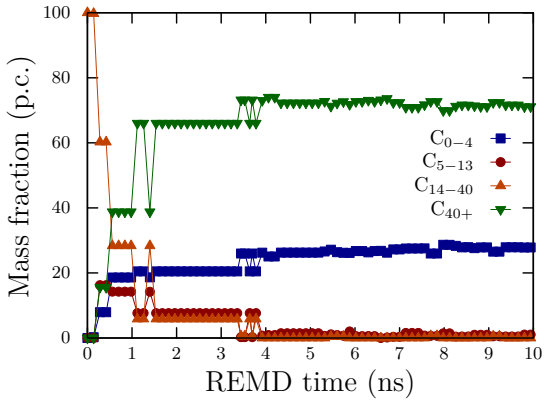

Figure 4: Mass distribution as a function of simulation time. gas (squares), light tar (circles), heavy tar (up triangles), kerogen (down triangles).

The evolution of the system in terms of mass fraction of gas (C_{0-4}) light tar (C_{5-13}), heavy tar (C_{14-40}) and kerogen (C_{40+}), is shown in Fig. 4, as a quantitative complement to the snaphots in Fig. 3. In agreement with the visual impression from the snaphots in Fig. 3, we see that the initial mixture of fatty acids (in the C_{14-40} category) breaks down in the

early steps of the simulation. A first stage in the evolution process is observed from 0.3 to 1 ns, in which the kerogen part accounts for $\sim$ 39 % of the total mass, followed by heavy tar (28 %), gas (19 %) and light tar (14 %). Then, from 1 to 4 ns, the kerogen part has grown to account for $\sim$ 66 % of the mass, the gas part represents $\sim$ 20 %, and the light and heavy tar contents are of around 6 and 8 %, respectively. Then, above 4 ns, tar contents decrease to very low ($\leq$ 1 %) values, indicative of the oil to gas window transition, and the gas and kerogen contents fluctuate around plateau values of 27 and 72 %, respectively. Interestingly, the obtained kerogen yield is significantly larger to those obtained in previous investigations using lignin (58 %) and cellulose (27 %) as proxies for type III OM.[28] One can also compare the predicted kerogen yield to the recent work by Zhang et al. who performed pyrolysis experiments of type II kerogen confined in a gold tube and reported kerogen yields of about 60% in the metagenesis stage.[46]

Fig. 5 gives a more detailed picture of the gas content evolution during the simulation. As for type III OM precursors,[28] the gas content in mass is dominated by water, with more than 100 mg produced per gram of precursor, and water is mostly released in the early steps of the evolution process. Yet, the amount of released water remains lower than those obtained with type III OM precursors: $\sim$ 320 and 230 mg/g with cellulose and lignin, respectively.[28,29] Besides water, large quantities of methane (70 mg/g) and hydrogen (45 mg/g) are also produced, yet most of them appear at a later stage, i.e. after $\sim$ 4 ns. Other hydrocarbons, such as C_2H_6 and C_2H_4 counting for $\sim$ 10 mg/g, C_3H_6, counting for $\sim$ 4 mg/g, and C_3H_8 and C_4H_8 ($\sim$ 1 mg/g) are also present at the end of the simulation. Among the gas species, one observes that some appears as intermediates during the evolution process, being first produced, then consumed. This includes butane, reaching about 3 mg/g at around 6 ns, then reducing to zero, and CO_2 which is only present in the first ns of simulation. This is also the case of CO, counting for about 25 mg/g at around 1 ns of simulation and then decreasing to $\sim$ 3 mg/g at 10 ns.

Fig. 6 shows the mass spectra, in the 0-350 g/mol range, computed at different simulation

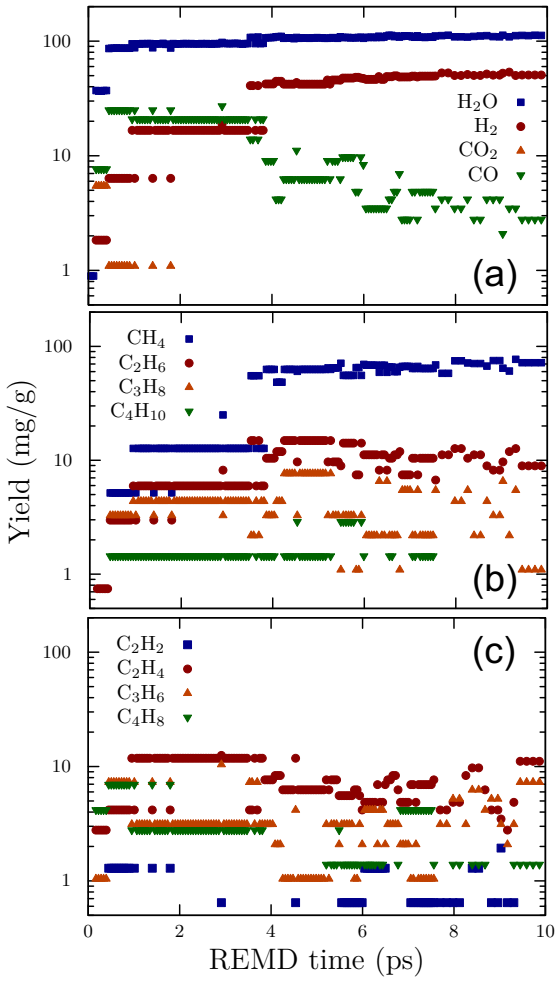

Figure 5: Evolution of the gas production with simulation time. Shown are the main (a) inorganic gases, (b) saturated hydrocarbons, (c) unsaturated hydrocarbons.

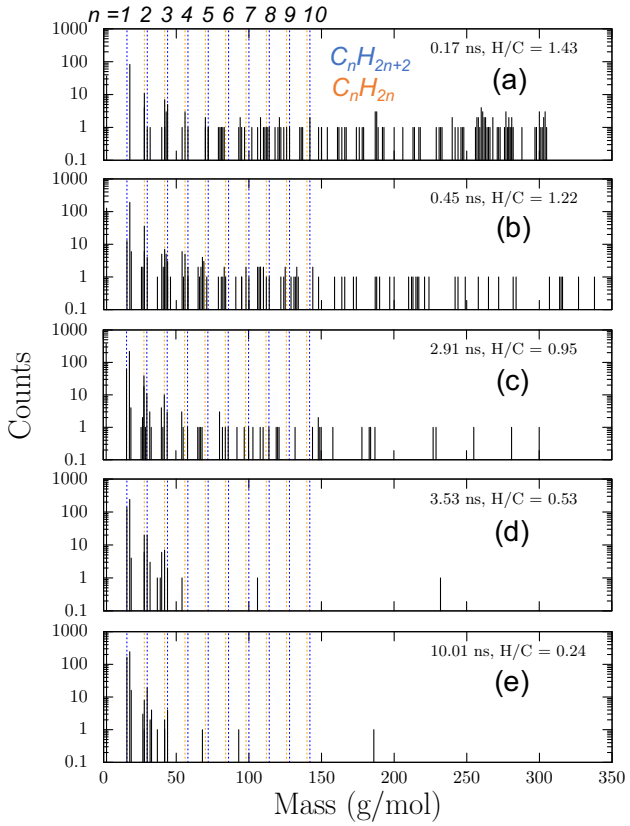

Figure 6: Mass spectra of the fluid phase at various simulation times. The corresponding kerogen maturity level is indicated with the H/C ratio and the vertical dotted lines indicate the first alkanes (blue) and alkenes (orange).

times. Note that, as common indicators of thermal maturity in experimental reports such as the vitrinite reflectance (R_0) can obviously not be determined from our atomistic models, we use the atomic H/C ratio, as a maturity indicator throughout the manuscript. H/C is expected to decrease when R_0 increases. At 0.17 ns of simulation, the forming kerogen is at a very immature stage (H/C = 1.43) and the fluid phase is characterized by a broad mass distribution corresponding to gas species, with masses typically lower than 50 g/mol, liquid hydrocarbons, from 50 to 200 g/mol, and heavier tar species with masses larger than 200 g/mol (Fig. 6(a)). Most alkanes (C_nH_{2n+2}) and alkenes (C_nH_{2n}) with $2 \leq n \leq 8$ are present at this stage, yet no methane molecule is observed. One also note that the molecules observed at 128 and 142 g/mol, correspond to functionalized hydrocarbons, $C_8H_{16}O$ and $C_8H_{14}O_2$, respectively, and not to nonane and decane as one could think. When increasing simulation time, kerogen maturity significantly increases as indicated by the H/C ratio decreasing progressively down to 0.24 at 10.01 ns, and the mass distribution in the fluid phase changes drastically. First, one clearly observes that the mass distribution narrows down towards light molecules with increasing time. Heavy tar species, which were numerous at low kerogen maturities (Fig. 6(a-b)), becomes rare when the kerogen H/C reaches values close to unity (Fig. 6(c), see also the snapshot in Fig. 3 at 2.91 ns), indicating the transition from oil window to gas window. At this stage, the hydrocarbon content is clearly dominated by methane yet species up to octane are still present in the fluid phase. At larger kerogen maturity values when H/C < 0.6 (Fig. 6(d-e)), heavy species have almost entirely disappeared from the fluid phase (see also Fig. 3 above 3 ns) and the alkane and alkene content is limited to species with $n \leq 3$, with abundancy clearly decreasing with increasing n. Although the parent OM and conditions slightly differ, the gas phase obtained in this work in the metagenesis stage – $\sim$ 80 mg/g of CH_4 and 40 mg/g of C_{2-5} species – can be compared to the recent results of Wang et al. – 80-250 mg/g of CH_4 and 0-160 mg/g of C_{2-5} – in the dry gas stage of confined pyrolysis experiments on type II kerogen.[33]

Fig. 7 shows the evolution of the isolated kerogen on a van Krevelen diagram (atomic H/C

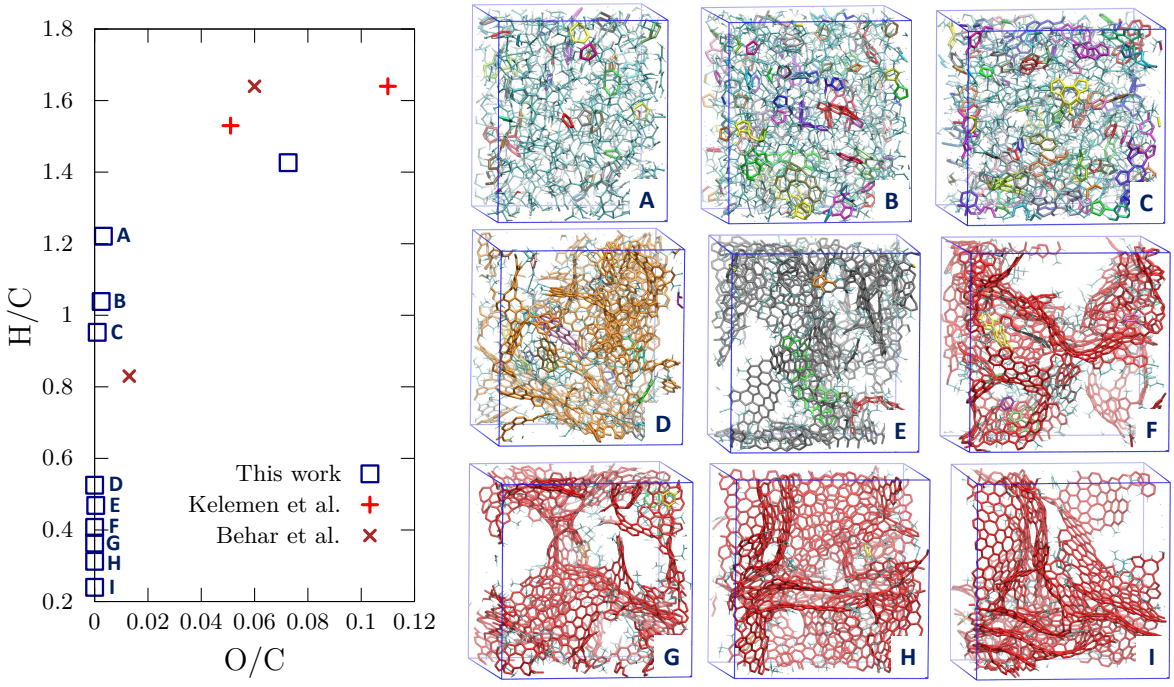

Figure 7: Van Krevelen diagram comparing the kerogen evolution obtained in this work to analyzes made by Kelemen et al.[22] and Béhar and Vandenbroucke[26] on natural type I kerogen samples. Snapshots of the kerogen models, identified with capital letters A-I, are shown on the right hand side with a thin ball and stick representation of carbon (cyan), hydrogen (white) and oxygen (red), and a superimposed thicker stick representation of ring clusters colored with one unique color per cluster.

versus O/C). The first identified kerogen, with H/C = 1.43 and O/C=0.073, has a similir composition to the immature type I kerogen characterized in Kelemen et al.[22] and Béhar et al.[26] Strictly speaking, this material should rather be considered as asphaltene, rather than kerogen, as it is constituted by 8 independant molecules of molecular masses ranging from 588 to 1284 Da, well below the insolubility threshold, located in the 5000-10000 Da range.[7] Also, because of its small size (only 551 C atoms), this model is not considered for further analyses. The nine other materials, named "A" to "I" by decreasing values of H/C and for which snaphots are provided on the right hand side of Fig. 7, are described in details in what follows.

All these models have considerably lower O/C (i.e. below 0.01), indicating a post-diagenesis stage.[7] Model A, B and C, with H/C ranging from 1.22 to 0.95, are typical of the oil window, or catagenesis (see the fluid mass spectra corresponding to model A and C in Fig. 6(b) and (c), respectively), while models D-I, for which H/C < 0.6, correspond to the late gas window (metagenesis). Model A is also constituted by different (12) independent molecules, with masses in the 500-3000 Da range, and thus, most certainly, soluble in some organic solvents. However, starting from model B, all models are formed by a connected network of covalent bonds, percolating though the cell periodic boundary conditions, mimicking a material with macroscopic mass. In a recent investigation, Lee et al.[47] have presented the mass distribution of some pre-oil to oil window type II kerogen samples from the Bakken formation. Using matrix assisted laser desorption ionization - time of flight (MALDI-TOF) spectroscopy, they observed that the distribution center shifts to larger mass values with increasing maturity. Quantitatively speaking, their most mature sample has its maximum distribution at ~ 1.5 kDa. Although not reported, their kerogen H/C ratio can be estimated to slightly above 1.2 by comparing the ratio of sp^2 over sp^3 carbon atoms (based on ^{13}C NMR data in Ref. 48) to those of our models (see Fig. 8). This is about the value of model A which shows a very similar mass distribution. For more mature kerogen, when H/C $<$ 1.2, in agreement with the kerogen models obtained by Bousige et al.[19] using the hybrid

reverse Monte Carlo (HRMC) method, our results show that the size of independent kerogen constituents is above the one of the simulated systems. This invalidates the molecular representation of mature kerogen proposed in Ungerer et al.[17] We also note that in model C, three smaller entities, with masses 631, 678 and 1167 Da are also present, in addition to the large percolating structure. Also, noticeable, model C is actually reasonably similar in composition to the end of catagenesis green river kerogen (H/C = 0.83, 0/C = 0.013) characterized by Béhar and Vandenbroucke.[26]

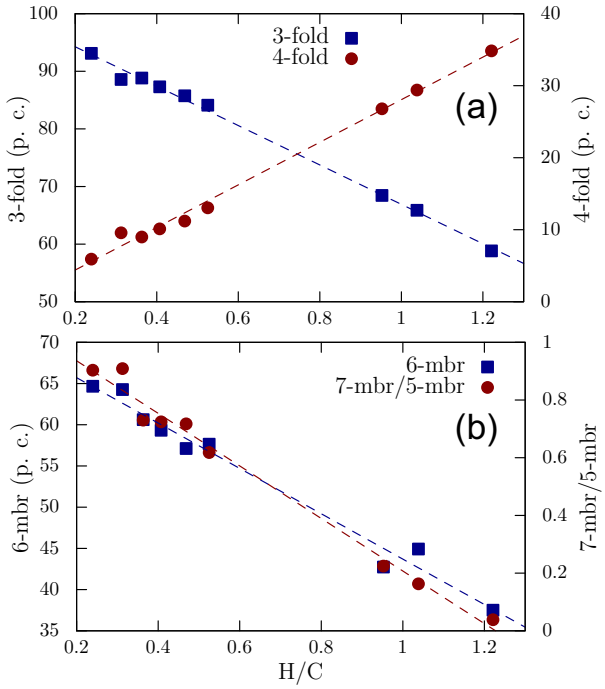

Figure 8: Evolution with the kerogen H/C ratio of (a) the fractions of (left) sp^2 and (right) sp^3 C atoms, and (b) the fraction of (left) hexagonal rings and (right) the ratio of the numbers of heptagons over pentagons.

Carbon atoms coordination numbers and ring statistics are given as a function of the H/C ratio in Fig. 8(a) and 8(b), respectively. The fraction of 3-fold (sp^2) C atoms increases linearlry from 59% for model A (H/C = 1.22) to 93% for model I (H/C = 0.24), while, conversely, the fraction of 4-fold atoms (sp^3) decreases from 35 to 6%, the remaining atoms (1-6%) being 2-fold atoms (either sp or sp^2 radicals). Noticing that 3-fold and 4-fold fractions are equivalent to the aromatic and aliphatic fractions derived from ^{13}C NMR experiments,

respectively, one observes that model C (68% sp^2 and 27% sp^3 C) has close coordination fractions to the end of catagenis kerogen analyzed by Béhar and Vandenbroucke[26] (64% sp^2 and 36% sp^3).

Fig. 8(b) shows that the fraction of six-membered rings (hexagons) increases significantly with decreasing H/C, from 37 % for model A to 65 % for model I. As obtained for type III kerogen,[29] the remaining rings are mostly pentagons at low maturities, implying curved ring clusters. As maturity increases, the numbers of pentagons and heptagons tend to even, which is consistent with the formation of flat aromatic flakes.

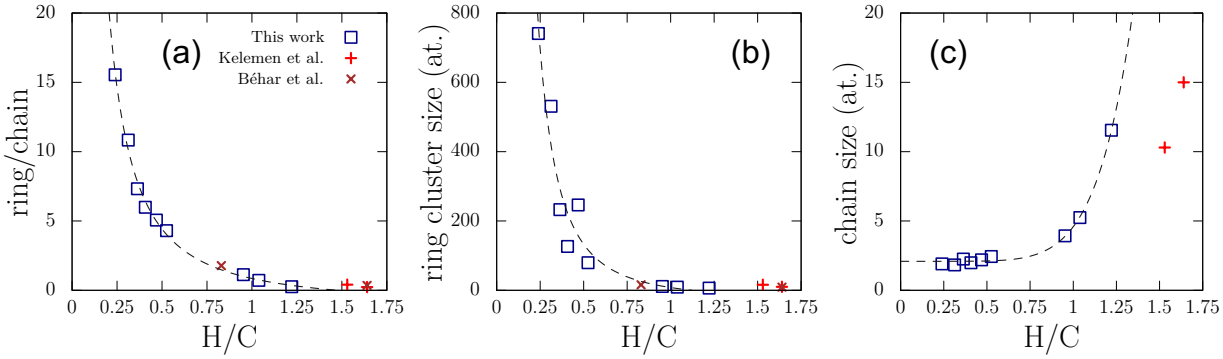

Figure 9: Evolution with the kerogen H/C ratio of (a) the ratio of C and O atoms in rings and chains, (b) the average ring cluster size, and (c) the average chain fragment size. Experimental data extracted from ^{13}C NMR data in Kelemen et al.[22] and Béhar and Vandenbroucke[26] are given for comparison. Dashed grey lines are power law fits to the data, merely serving as guides to the eye.

Fig. 9 shows the evolution with the H/C ratio of some parameters that allows for characterizing the carbon skeleton of the kerogen. The ratio of atoms (excluding hydrogen) in ring and chain structures (Fig. 9(a)) increases significantly from 0.26 in the most immature kerogen (A) to 15.5 in the most mature (I), indicating a transition from highly aliphatic to highly aromatic. As H/C decreases, we also observe that the average size of ring clusters considerably increases from about 6 atoms (i.e. one ring) in model A to 741 atoms (301 rings) in model I (Fig. 9(b)). Meanwhile, the average size of chain-like fragments decreases from 12 atoms (A) to 2 atoms (I). All these data are consistant with the experimental data of Kelemen et al.[22] and Béhar and Vandenbroucke,[26] asides maybe from the chain size of the

most immature models, which appear sligthly larger than expected at given H/C values for immature type I kerogen. However, the data reported by Béhar and Vandenbroucke for the end of categenesis kerogen sits almost perfectly on the curves fitted from the REMD data.

Finally, looking at all these data and as can be visually inferred in Fig. 7, the structure of the kerogen in the catagenesis stage (models A to C) can be described as small ring fragments crosslinked by relatively long chains, with growing ring content and increasing size of ring fragments as maturity increases. Conversely, in the metagenesis stage (models D to I), the ring structure has coalesced into a large ring cluster percolating through the simulation cell periodic boundaries. Chain fragments are found at ring cluster edges, some of them cross-linking the main ring fragment to smaller ones. The increase in average ring cluster size during maturation is mostly due to the inclusion of these smaller clusters into the main one, via suppression of the chain-like cross-links, thus diminishing the number of independent clusters.

The computed X-ray diffraction patterns for the nine kerogen models are shown in Fig. 10. In the early maturation stage, kerogen is mostly amorphous, the only noticeable feature being a a broad band located at ~ 20 degrees, often termed as γ peak in immature kerogen and attributed to aliphatic moities,[49,50] and a barely visible features at the carbon 10 band location (42 degrees). Conversely, in the metagenesis stage, all the characteristic features of turbostratic carbon can be observed. A detailed analysis of the patterns of models B to I, after baseline removal, is provided in Fig. S3, with identification of the 002, 10 and 004 peaks. The evolutions with H/C of the in-plane (L_a) and out-of-plane (L_c) coherence lengths (or crystallite sizes), and of the interlayer spacing (d_{002}), are given in Fig. S4. Model D, with H/C = 0.53, is the one with smaller coherence lengths, namely $L_a = 1.1$ nm and $L_c = 0.69$ nm, corresponding to 2–3 layers stacks of small (20–30 rings) graphene flakes. With further maturation, L_a and L_c increase to plateau values of about 2.3 and 1.0 nm, respectively. At this stage, common to models E to I, the structure of crystallites can be viewed as stacks of 3 to 4 graphene flakes of about 70-80 rings. We note that similar coherence lengths

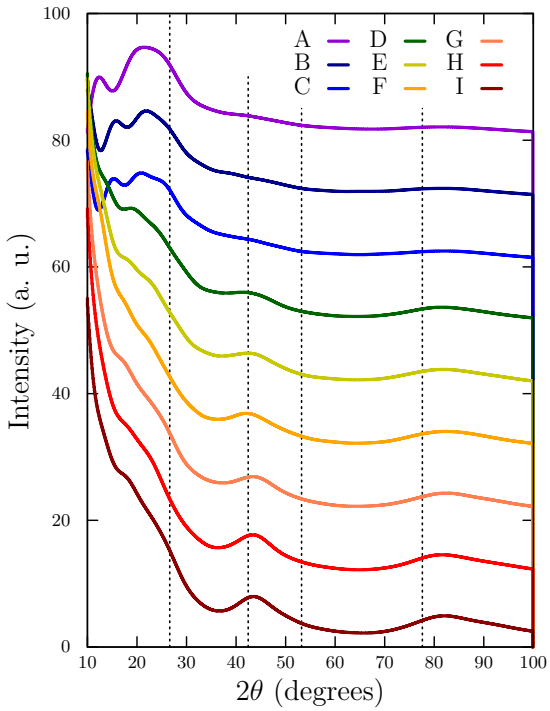

Figure 10: Xray-diffraction patterns computed for the nine kerogen models (A to I from top to bottom). Corresponding graphite peak locations in graphite are indicated with black dotted lines as a guide to the eye (002, 10, 004 and 11 from left to right).

were obtained for type III kerogen at the same level of maturity,[29] confirming that in the metagenesis stage, no significant difference in composition and structure remains between keogens of different origins. Also, as in ref. 29, d_{002} is found at $\sim$ 3.85 Å, a value that is common for low temperature, non-graphitizable, carbons as chars and blacks.[51]

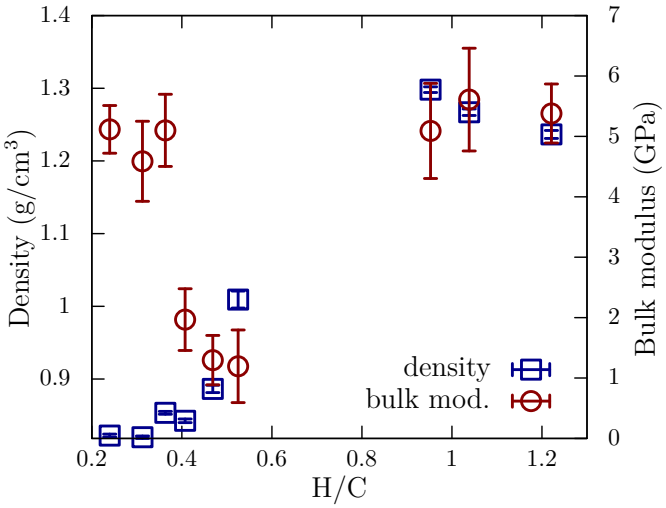

Figure 11: Evolution with H/C of the kerogen density (squares, left scale) and bulk modulus (circles, right scale).

Fig. 11 shows the evolutions of the kerogen density and bulk modulus as a function of the H/C ratio. Here, we assume that the kerogen models are isotropic and the bulk modulus is computed from the volume fluctuations at equilibrium, using Eq. 3. A full analysis of kerogen elasticity, including the Young, shear and bulk modulus as well as Poisson coefficients for type I (this work) and type III (Ref. 29) kerogen models will be published in a forthcoming article. In the most immature states (models A-C), corresponding to H/C $>$ 0.9, density ranges from 1.2 to 1.3 g/cm^3 and the bulk modulus is of about 5 GPa. When reaching the metagenesis stage, for H/C in the 0.4-0.6 range (models D-F), a significant drop in density is observed, down to $\sim$ 0.84 g/cm^3 for model F. The bulk modulus also significantly decreases, showing values of $\sim$ 1-2 GPa. Finally, in the late metagenesis stage (models G-I), density gets stabilized at $\sim$ 0.82 g/cm^3 and the bulk modulus is again of about 5 GPa. The evolution of the total (probeless) kerogen porosity with H/C follows the exact opposite to

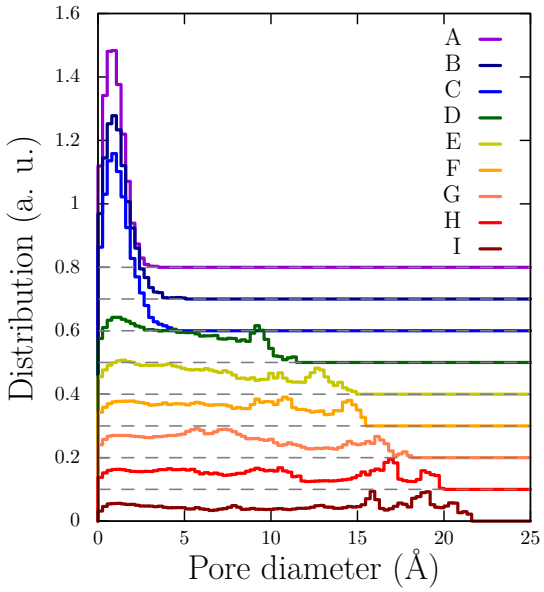

Figure 12: Normalized pore size distributions (PSDs) of the kerogen models.

the density (Fig. S5). Starting at 20-25 % for models A-C, porosity increases significantly with decreasing H/C up to a value of 62 % for model I. More interesting is the evolution of the keorgen PSD with maturity shown in Fig. 12. In agreement with the snapshots in Fig. 7, with no pores larger than 3–4 Å, models A-C present no to little accessible porosity to common adsorbates. However, when reaching the metagenesis stage (model D) a rather flat PSD, extending up to about 1 nm is observed. The PSD further broadens with additional maturation, reaching pore sizes of about 2 nm in model I. Both increases in porosity and pore sizes are clearly visible in Fig. 7.

Discussion

The evolution with maturation of the kerogen density, porosity and bulk modulus observed here is extremely similar to former results on type III kerogen models.[29] Despite slightly different values, both studies exhibit a significant decrease in density and increase in porosity during the catagenesis stage, before stabilization in the metagenesis stage. Also, in both works, the bulk modulus is shown to pass through a minimum value at the late catageneis

or early metagenis stage. In what follows we attempt to rationalize these evolutions with respect to the simulated maturation conditions and give some comparison with existing experimental data.

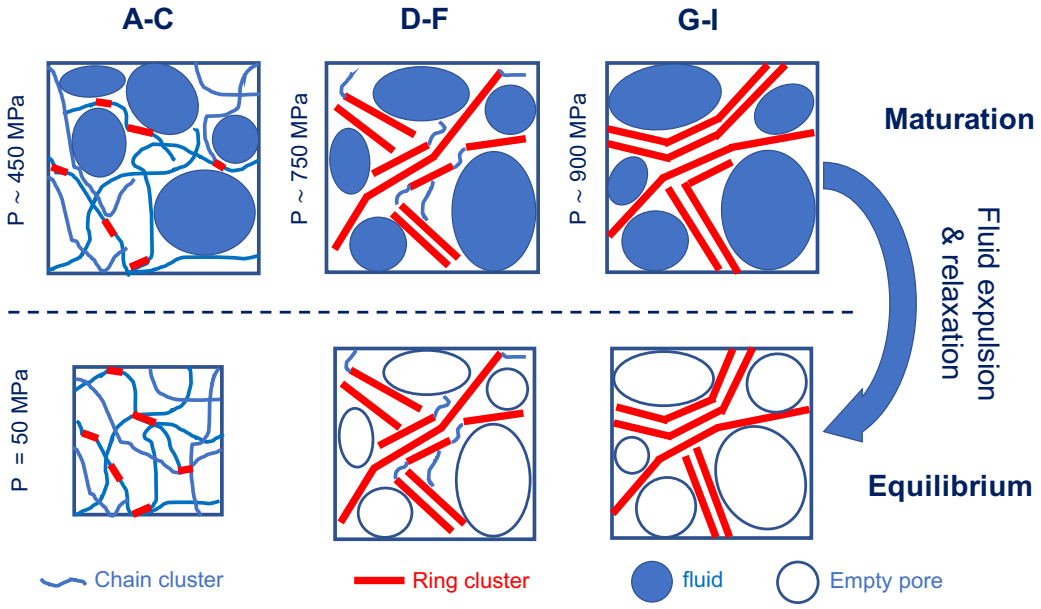

Figure 13: Sketch of the kerogen evolution during maturation in the REMD simulation (top) and of its corresponding microstructure after fluid expulsion (bottom).

Fig. 13 shows a schematics of the kerogen evolution during the REMD simulation. Three important evolution stages have been covered in the REMD simulation: (i) early catagenesis during which an aliphatic kerogen (models A-C) is formed as well as an oil-rich fluid phase ; (ii) early metagenesis during which the kerogen has evolved to a predominantly aromatic network (models D-F) and the fluid has matured towards gaseous hydrocarbons; and (iii) late metageneis during which the kerogen backbone (models G-I) has become almost entirely aromatic. Unfortunately, late catagenesis configurations were not observed in the 423 K replica of the REMD process, preventing their investigation here. This is a clear drawback of this method which, albeit ensuring a progressive and rapid (with respect to simulation time) evolution of the replica of interest towards equilibrium, does not permit an easy, and a priori, control of the visited intermediate states. Indeed, configurations typical of the

catagenesis stage were observed in some low temperature replicas, including some close to the temperature of interest like $\sim$ 460 K, yet these configurations were exchanged with more mature (and stable) configurations coming from higher T replicas, before reaching the 423 K replica. This results from a competition between the exchange dynamics, and the dynamics of evolution in high T replicas, the latter being certainly the fastest in our case. Solving this issue would either require to increase the frequency of exchange, which may bias the accuracy of the results, as our exchange frequency is probably already slightly too fast, compared to the thermostat frequency (see the discussion in the introduction), or decrease the maximum replica temperature T_{up}, to slow down the evolution. We are currently performing a REMD investigation with T_{up} set to 2456 K, yet it inevitably slows down the evolution at 423 K and we will probably have to limit the simulation to the catagenesis stage only.

An other limitation of the method is that, for practical reasons, the REMD simulation had to be performed at fixed volume and mass, in order not to slow down the replica exchange process, hence mimicking closed system pyrolysis, while natural kerogen evolution rather operates under relatively constant pressure, and is subjected to geological events that lead to partial expulsion of the fluid content. In the REMD simulation, we have observed that the fluid production and maturation induce a signficant increase in the system's pressure, from $\sim$ 450 MPa in the early catagenesis stage to $\sim$ 900 MPa in the late metagenesis stage. Kerogen models were then analyzed after total fluid expulsion and relaxation at 423 K and 50 MPa ($\sim$ 5 km depth conditions). During the relaxation, the immature kerogen models (A-C) densify due to compaction, a well-known phenomenon following the expulsion of free water in larger pores.[52] The obtained density (1.25 g/cm^3) is typical of values expected for immature kerogen,[17,52] and the models show almost no porosity after relaxation, which is consistent with the very low amounts of micropores ($<$ 0.006 cm^3/g) reported for immature type I kerogen.[53] The $\sim$ 5 GPa modulus of these models thus corresponds to the intrinsic stiffness of the immature kerogen backbone. Models in the early (D-F) and late (G-I) metagenesis states retain most of the porosity that existed prior to fluid expulsion, indicating that the

mature kerogen backbone is actually stiffer than the one of immature kerogen. However, the bulk modulus of models D-I is lower (D-F) or equal (G-I) to the one of models (A-C), because of the presence of porosity, giving rise to a much lower density in these models.

This predominant effect of density (or porosity) on modulus is consistent with former computational[19,20] and experimental investigations.[30] Regarding the correlation between maturity and stiffness, as discussed above, our results clearly suggest that the kerogen "wall" stiffness is significantly larger for mature kerogen than for immature kerogen, contradicting the results obtained with molecular models,[20] which cannot capture the rigidification induced by cross-linking.

Finally, these findings provide an interesting picture of porosity development in kerogen. Under geological conditions, and in the absence of adsorbed fluid, immature kerogen (diagenesis - early catagenesis) is essentially non-porous and the porosity of immature kerogen is expected to be proportional to the amount of adsorbed fluid. Conversely, the porosity of mature kerogen seems to be extremely dependent on the amount of produced fluid, and on whether this fluid was released before the rigidification of the kerogen skeleton, which takes place during the late catagenesis. Reported experimental data on kerogen density, in the 1.0–1.4 g/cm^3 range,[30,54] being larger than the ones reported here for mature kerogen, suggests that fluid expulsion actually occurs, at least in part, prior to kerogen rigidification. Especially, Okiongbo et al. have analyzed natural samples and have reported that density increases with maturity,[54] yet the burial depth, also increasing with maturity, could also have an important effect on density.

Further work is definitely needed to address kerogen formation and evolution in more realistic conditions, and account more accurately for porosity and stiffness buildups. The most straighforward way to achieve this will be, as in former works on type III kerogen,[27,28] to interrupt the REMD simulation whenever the pressure in the T_{low} replica significantly depart from the expected pressure or fluid content. This replica could then be relaxed at the suited conditions (P and fluid content) and the REMD relaunched with all the replicas

reinitialized with a copy of the relaxed system. This brings out two important difficulties: i) one we'll have to make a somehow arbitrary choice about when to expell fluid, which fluid fraction and which fluid molecules to expell, as this is mostly unknown and may vary from a shale play to another; ii) this would add a significant computational cost by considerably slowing down the system evolution, as, every time the REMD is interrupted, all the replicas at $T > T_{low}$, some being considerably more evolved, are lost.

Conclusion

Despite some important limitations, above summarized, REMD, combined with a reactive potential, remains a very effective – and so far the only – atomistic simulation technique to simulate kerogen evolution and we have presented here a complete REMD investigation of type I kerogen evolution at 423 K, covering most of the diagenesis, early catageneis and full metageneis stages. Despite the use of a very rude model of the parent OM, described as a mixture of five fatty acids, and the constraint of working under closed conditions, a consistent set of results describing fluid production, kerogen structure evolution as well as density and stiffness build-ups were obtained, and compared to existing experimental and simulation data. An important result, compared to former REMD simulations on type III kerogen evolution, using cellulose and lignin as type III OM precursors, is the presence of the oil window in the simulations, with a fluid phase rich in long hydrocarbons, including alkanes up to octane, and an aliphatic immature kerogen. Although further work would be required to reach more realistic maturation conditions, especially working under constant pressure conditions and exploring the effects of (partial) fluid release, the produced kerogen models can now be used to investigate the effects of maturity on gas adsorption and transport in type I kerogen.

CRediT authorship contribution statement

Jean-Marc Leyssale: Conceptualization, Methodology, Computational investigation, Formal analysis, Data curation, Writing - original draft. **Pierre-Louis Valdenaire**: Computational investigation, Data curation. **Kévin Potier**: Formal analysis, Data curation, Writing - review & editing. **Roland J.-M. Pellenq**: Conceptualization, Formal analysis, Writing - review & editing.

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.

Supporting Information

Additional results regarding the validation of the methodology, pressure development during the REMD simulation, and characterization of the struture and porosity of the kerogen models.

Acknowledgements

References

(1) Kerr, R. A. Natural Gas From Shale Bursts Onto the Scene. *Science* **2010**, *328*, 1624–1626.

(2) Hackley, P. C.; Cardott, B. J. Application of Organic Petrography in North American Shale Petroleum Systems: A Review. *Int. J. Coal Geol.* **2016**, *163*, 8–51.

(3) Jia, B.; Tsau, J.-S.; Barati, R. A Review of the Current Progress of CO_2 Injection EOR and Carbon Storage in Shale Oil Reservoirs. *Fuel* **2019**, *236*, 404–427.

(4) Yuan, J.; Luo, D.; Feng, L. A Review of the Technical and Economic Evaluation Techniques for Shale Gas Development. *Appl. Energy* **2015**, *148*, 49–65.

(5) Sovacool, B. K. Cornucopia or Curse? Reviewing the Costs and Benefits of Shale Gas Hydraulic Fracturing (fracking). *Renew. Sustain. Energy Rev.* **2014**, *37*, 249–264.

(6) Cane, R. F. Geological semantics. *Nature* **1970**, *228*, 1009.

(7) Vandenbroucke, M.; Largeau, C. Kerogen Origin, Evolution and Structure. *Org. Geochem.* **2007**, *38*, 719 – 833.

(8) Collell, J.; Galliero, G.; Gouth, F.; Montel, F.; Pujol, M.; Ungerer, P.; Yiannourakou, M. Molecular simulation and modelisation of methane/ethane mixtures adsorption onto a microporous molecular model of kerogen under typical reservoir conditions. *Microporous Mesoporous Mater.* **2014**, *197*, 194–203.

(9) Falk, K.; Pellenq, R.; Ulm, F. J.; Coasne, B. Effect of Chain Length and Pore Accessibility on Alkane Adsorption in Kerogen. *Energy & Fuels* **2015**, *29*, 7889–7896.

(10) Zhao, T.; Li, X.; Zhao, H.; Li, M. Molecular simulation of adsorption and thermodynamic properties on type II kerogen: Influence of maturity and moisture content. *Fuel* **2017**, *190*, 198–207.

(11) Obliger, A.; Valdenaire, P.-L.; Capit, N.; Ulm, F. J.; Pellenq, R. J.-M.; Leyssale, J.-M. Poroelasticity of Methane-Loaded Mature and Immature Kerogen from Molecular Simulations. *Langmuir* **2018**, *34*, 13766–13780.

(12) Li, Z.; Yao, J.; Firoozabadi, A. Kerogen Swelling in Light Hydrocarbon Gases and Liquids and Validity of Schroeder's Paradox. *J. Phys. Chem. C* **2021**, *125*, 8137–8147.

(13) Obliger, A.; Valdenaire, P.-L.; Ulm, F.-J.; Pellenq, R. J.-M.; Leyssale, J.-M. Methane Diffusion in a Flexible Kerogen Matrix. *J. Phys. Chem. B* **2019**, *123*, 5635–5640.

(14) Collell, J.; Galliero, G.; Vermorel, R.; Ungerer, P.; Yiannourakou, M.; Montel, F.; Pujol, M. Transport of Multicomponent Hydrocarbon Mixtures in Shale Organic Matter by Molecular Simulations. *J. Phys. Chem. C* **2015**, *119*, 22587–22595.

(15) Obliger, A.; Ulm, F.-J.; Pellenq, R. Impact of Nanoporosity on Hydrocarbon Transport in Shales' Organic Matter. *Nano Lett.* **2018**, *18*, 832–837.

(16) Vasileiadis, M.; Peristeras, L. D.; Papavasileiou, K. D.; Economou, I. G. Transport properties of shale gas in relation to kerogen porosity. *J. Phys. Chem. C* **2018**, *122*, 6166–6177.

(17) Ungerer, P.; Collell, J.; Yiannourakou, M. Molecular Modeling of the Volumetric and Thermodynamic Properties of Kerogen: Influence of Organic Type and Maturity. *Energy Fuels* **2015**, *29*, 91–105.

(18) Vasileiadis, M.; Peristeras, L. D.; Papavasileiou, K. D.; Economou, I. G. Modeling of bulk kerogen porosity: Methods for control and characterization. *Energy Fuels* **2017**, *31*, 6004–6018.

(19) 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. Realistic Molecular Model of Kerogen's Nanostructure. *Nat. Mater.* **2016**, *15*, 576–582.

(20) Kashinath, A.; Szulczewski, M.; Dogru, A. H. Modeling the Effect of Maturity on the Elastic Moduli of Kerogen Using Atomistic Simulations. *Energy Fuels* **2020**, *34*, 1378–1385.

(21) Obliger, A.; Bousige, C.; Coasne, B.; Leyssale, J.-M. Development of Atomistic Kerogen Models and Their Applications for Gas Adsorption and Diffusion: A Mini-Review. *Energy Fuels* **2023**, *37*, 1678–1698.

(22) 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. Direct Characterization of Kerogen by X-ray and Solid-State 13C Nuclear Magnetic Resonance Methods. *Energy & Fuels* **2007**, *21*, 1548–1561.

(23) Qian, Y.; Zhan, J.-H.; Lai, D.; Li, M.; Liu, X.; Xu, G. Primary understanding of non-isothermal pyrolysis behavior for oil shale kerogen using reactive molecular dynamics simulation. *Int. J. Hydrog. Energy* **2016**, *41*, 12093–12100.

(24) Han, Q.; Li, M.; Liu, X.; Xiao, H.; Ren, J.; Guo, C. A maturation scale for molecular simulation of kerogen thermal degradation. *Org. Geochem.* **2023**, *175*, 104507.

(25) Siskin, M.; Scouten, C.; Rose, K.; Aczel, T.; Colgrove, S.; Pabst Jr., R. In *Composition, Geochemistry and Conversion of Oil Shales*; Snape, C., Ed.; Kluwer Academic Publishers: Dordrecht, 1995; pp 143–158.

(26) Béhar, F.; Vandenbroucke, M. Chemical Modelling of Kerogens. *Org. Geochem.* **1987**, *11*, 15–24.

(27) Atmani, L.; Bichara, C.; Pellenq, R. J.-M.; Van Damme, H.; van Duin, A. C. T.; Raza, Z.; Truflandier, L. A.; Obliger, A.; Kralert, P.; Ulm, F. J.; Leyssale, J.-M. From Cellulose to Kerogen: Molecular Simulation of a Geological Process. *Chem. Sci.* **2017**, *8*, 8325–8335.

(28) Atmani, L.; Valdenaire, P.-L.; Pellenq, R. J.-M.; Bichara, C.; Van Damme, H.; van Duin, A. C. T.; Ulm, F. J.; Leyssale, J.-M. Simulating the Geological Fate of Terrestrial Organic Matter: Lignin vs Cellulose. *Energy Fuels* **2020**, *34*, 1537–1547.

(29) Leyssale, J.-M.; Valdenaire, P.-L.; Potier, K.; Pellenq, R. J.-M. Replica Exchange Molecular Dynamics Simulation of Organic Matter Evolution: From Lignin to Overmature Type III Kerogen. *Energy Fuels* **2022**, *36*, 14723–14733.

(30) Yan, F.; hua Han, D. *SEG Technical Program Expanded Abstracts 2013*; 2013; pp 2778–2782.

(31) Zargari, S.; Wilkinson, T. M.; Packard, C. E.; Prasad, M. Effect of thermal maturity on elastic properties of kerogen. *Geophysics* **2016**, *81*, M1–M6.

(32) Emmanuel, S.; Eliyahu, M.; Day-Stirrat, R. J.; Hofmann, R.; Macaulay, C. I. Impact of thermal maturation on nano-scale elastic properties of organic matter in shales. *Mar. Pet. Geol.* **2016**, *70*, 175–184.

(33) Wang, J.; Liu, Y.; Yang, C.; Jiang, W.; Li, Y.; Xiong, Y.; Peng, P. Evolution of mechanical properties of kerogen with thermal maturity. *Mar. Pet. Geol.* **2022**, *145*, 105906.

(34) Earl, D. J.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys* **2005**, *7*, 3910–3916.

(35) Geyer, C. J. *Computing Science and Statistics Proceedings of the 23rd Symposium on the Interface*; American Statistical Association: New York, 1991; p 156.

(36) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(37) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press, 1987.

(38) Duane, S.; Kennedy, A.; Pendleton, B. J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222.

(39) Becker, E. W. *Microalgae: Biotechnology and Microbiology*; Cambridge University Press, Cambridge, 1994.

(40) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(41) Leyssale, J.-M.; Galvez, M. E.; Valdenaire, P.-L.; Pellenq, R.; van Duin, A. C. Atomic-Scale Mechanism of Carbon Nucleation from a Deep Crustal Fluid by Replica Exchange Reactive Molecular Dynamics Simulation. *Geochim. Cosmochim. Acta* **2022**, *329*, 106–118.

(42) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.

(43) Shinoda, W.; Shiga, M.; Mikami, M. Rapid Estimation of Elastic Constants by Molecular Dynamics Simulation Under Constant Stress. *Phys. Rev. B* **2004**, *69*, 134103.

(44) Franzblau, D. S. Computation of Ring Statistics for Network Models of Solids. *Phys. Rev. B* **1991**, *44*, 4925–30.

(45) Debyer, Freely Available Diffraction Software. https://github.com/wojdyr/debyer.

(46) Zhang, D.; Cao, H.; Lei, Y.; Wu, H.; Wang, X.; Guo, X.; Yan, J.; Wang, S.; Shi, T.; Li, H.; Zou, Y.; Peng, P. A study on molecular structural evolution of type II kerogen in a gold tube thermal system: Insights from solid-state 13C NMR. *Fuel* **2023**, *331*, 125898.

(47) Lee, H.; Abarghani, A.; Liu, B.; Shokouhimehr, M.; Ostadhassan, M. Molecular weight variations of kerogen during maturation with MALDI-TOF-MS. *Fuel* **2020**, *269*, 117452.

(48) Lee, H.; Oncel, N.; Liu, B.; Kukay, A.; Altincicek, F.; Varma, R. S.; Shokouhimehr, M.; Ostadhassan, M. Structural Evolution of Organic Matter in Deep Shales by Spectroscopy (1H and 13C Nuclear Magnetic Resonance, X-ray Photoelectron Spectroscopy, and Fourier Transform Infrared) Analysis. *Energy Fuels* **2020**, *34*, 2807–2815.

(49) Wang, Q.; Yao, C.; Zhou, Y.; Che, Y.; Liu, H.; Zhao, X.; Zhang, Y. The relationship between the oil shale density and the structural features of corresponding kerogens. *Carbon Resour. Convers.* **2022**, *5*, 84–91.

(50) Wang, D.; Liu, Y.; Zhang, T.; Christopher, D. H.; Bature, N. S.; Fan, T.; Guo, H. Structural and quantitative evolution of organic matters in oil shale during two different retorting processes. *AIChE J.* **2021**, *67*, e17278.

(51) Gharpure, A.; Vander Wal, R. Improving graphenic quality by oxidative liberation of crosslinks in non-graphitizable carbons. *Carbon* **2023**, 118010.

(52) Liu, B.; Wang, Y.; Tian, S.; Guo, Y.; Wang, L.; Yasin, Q.; Yang, J. Impact of thermal maturity on the diagenesis and porosity of lacustrine oil-prone shales: Insights from natural shale samples with thermal maturation in the oil generation window. *Int. J. Coal Geol.* **2022**, *261*, 104079.

(53) Liu, B.; Mohammadi, M.-R.; Ma, Z.; Bai, L.; Wang, L.; Xu, Y.; Ostadhassan, M.; Hemmati-Sarapardeh, A. Evolution of porosity in kerogen type I during hydrous and anhydrous pyrolysis: Experimental study, mechanistic understanding, and model development. *Fuel* **2023**, *338*, 127149.

(54) Okiongbo, K. S.; Aplin, A. C.; Larter, S. R. Changes in Type II Kerogen Density as a Function of Maturity: Evidence from the Kimmeridge Clay Formation. *Energy Fuels* **2005**, *19*, 2495–2499.

Graphical TOC Entry