Realistic molecular model of kerogen's nanostructure

Colin Bousige^{1,2}, Camélia Matei Ghimbeu³, Cathie Vix-Guterl³, Andrew E. Pomerantz⁴, Assiya Suleimenova⁴, Gavin Vaughan⁵, Gaston Garbarino⁵, Mikhail Feygenson⁶, Christoph Wildgruber⁷, Franz-Josef Ulm¹, Roland J.-M. Pellenq^{1,2,8} and Benoit Coasne^{1,2*}

Despite kerogen's importance as the organic backbone for hydrocarbon production from source rocks such as gas shale, the interplay between kerogen's chemistry, morphology and mechanics remains unexplored. As the environmental impact of shale gas rises, identifying functional relations between its geochemical, transport, elastic and fracture properties from realistic molecular models of kerogens becomes all the more important. Here, by using a hybrid experimental-simulation method, we propose a panel of realistic molecular models of mature and immature kerogens that provide a detailed picture of kerogen's nanostructure without considering the presence of clays and other minerals in shales. We probe the models' strengths and limitations, and show that they predict essential features amenable to experimental validation, including pore distribution, vibrational density of states and stiffness. We also show that kerogen's maturation, which manifests itself as an increase in the sp^2/sp^3 hybridization ratio, entails a crossover from plastic-to-brittle rupture mechanisms.

as recovery from organic-rich shales is quickly changing the United States economy, which raises both high hopes and environmental and political concerns. The gas (wet or dry) is partially adsorbed on the walls of the carbon backbone, kerogen, which intersects in small patches a tightly packed, finegrained sedimentary rock. Owing to the very low permeability of the rock^{1,2}, hydrocarbon recovery from kerogen requires stimulation techniques such as hydraulic fracturing, or 'fracking', that raise environmental and economic concerns³⁻⁵.

Yet, despite its importance, our knowledge about kerogen structure and properties is incomplete, owing to the difficulty of isolating kerogen from the source rocks under a bulk form amenable to geochemical and geophysical characterization⁶. Kerogen is an intrinsically complex material composed of an amorphous porous carbon skeleton with a pore size distribution spanning from a few Å to the μ m range, as probed by small-angle scattering⁷⁻¹⁰ and adsorption^{7,11,12}. Besides their known meso- and macroporosity, kerogens exhibit microporosity, which is thought to be responsible for their ultralow permeability, typically six orders of magnitude below that of conventional reservoirs^{13,14}, and unexplained fast hydrocarbon productivity declines¹⁵.

On a molecular scale, much of our current knowledge on kerogen relies on two-dimensional models which are based on the distribution of large complex carbonaceous molecules (as assessed from nuclear magnetic resonance (NMR) and X-ray photoelectron spectroscopy^{6,16-19}) or on molecular dynamics (MD) models describing kerogen as an assembly of molecules²⁰⁻²². Such models have allowed trends to be obtained for the heat capacity, enthalpy of formation, density, and adsorption in kerogen and in its soluble counterparts^{22–25}. However, these models of unconnected polyaromatic molecules are not designed to link chemical specificity to transport and mechanics, as they flow on pressure gradient or strain.

From a geochemical perspective, depending on its geographic origin, maturity and sediment history, kerogen exhibits a range of densities²⁶, chemical compositions (atomic contents and chemical functions), tortuosities, porosities, and so on. Given this variability, attempts have been made to classify kerogen types using maturity indicators such as vitrinite reflectance R_0 , aromatic/aliphatic ratio, and hydrogen-to-carbon (H/C) and oxygen-to-carbon (O/C) atomic ratios^{6,27}. When cross-plotted in the so-called van Krevelen diagram²⁸ (Fig. 1a), these indicators provide a means to distinguish different kerogen types in terms of depositional origin (lacustrine [type I], marine [type II], terrestrial [type III], and originating from residues [type IV]). In addition to geography, sediment history and specificity of biogenic structures remaining in kerogen play a role in kerogen variability; the type of deposited organisms and maturation kinetics through diagenesis and catagenesis^{6,27} affect the resulting structure. From a chemical viewpoint, maturation due to exposure to high temperature and pressure over geologic times results in a decrease of the H/C and O/C atomic ratios and the displacement towards the bottom-left corner of the van Krevelen diagram (Fig. 1a).

On the other hand, the geochemical specificity so-defined does not provide insight into the interplay between chemical, morphological, vibrational and mechanical properties. Herein, we adopt the perspective that, in addition to their chemical composition,

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ² (MSE)², UMI 3466 CNRS-MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ³Institut de Science des Matériaux de Mulhouse (IS2M), UMR 7360 CNRS - UHA, 15 rue Jean Starcky, BP 2488, 68057 Mulhouse cedex, France. ⁴Schlumberger-Doll Research, 1 Hampshire Street, Cambridge, Massachusetts 02139, USA. ⁵European Synchrotron Radiation Facility, BP-220, F-38043, Grenoble Cedex 9, France. ⁶Chemical and Engineering Materials Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. ⁷Instrument and Source Division (NScD), Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. ⁸CINaM, CNRS/Aix Marseille Université, Campus de Luminy, 13288 Marseille cedex 09, France. *e-mail: coasne@mit.edu

ARTICLES



Figure 1 | Kerogen in organic-rich shale formations. a, Van Krevelen diagram, that is, representation of our current understanding of the chemical evolution of immature kerogens of varying sources (Type I, II, III and IV) with increasing levels of maturity (based on ref. 53). The same colour code applies throughout the article: PYO2 (black), MarK (blue), EFK (red) and MEK (yellow). Levels of thermal maturity are usually measured using the vitrinite reflectance ($\% R_0$), a geochemical indicator integrating the effects of time and temperature during maturation of sediments. Many other maturity indicators are available, such as the aromatic/aliphatic ratio, pyrolysis temperature T_{max} , $sp^2/(sp^2 + sp^3)$ ratio, and aromatic C content (see Supplementary Table 1 for full sample characterization). In general, kerogen composition moves from the upper right of the figure to the lower left with increasing maturity. MEK is a Type IIS kerogen, whereas EFK and MarK are Type II marine kerogens. Their maturities increase with their vitrinite reflectance: $R_0 = 0.55$, 0.65 and 2.2%, respectively. Note that the uncertainty on the measured oxygen contents is large, but this does not affect the relevance of the discussion (see discussion in the Supplementary Information). **b**, X-ray microscopy image of raw shale showing inclusions of pyrite, clay, organic matter and other minerals. c, Molecular models of the four samples under study with a density of 1.2 g cm⁻³. Carbon, hydrogen and oxygen atoms are represented in grey, white and red, respectively. The box size is 50 Å in each direction.

structural features of kerogen drive many of its physical properties. We show that, once the structure is properly constrained by experimental diffraction data, composition and density, many morphological and physical properties follow in atomistic simulations. Based on a molecular dynamics—hybrid reverse Monte Carlo (MD–HRMC) reconstruction method²⁹, the present approach enables a bottom-up perspective on the broad science required for sustainable hydrocarbon recovery from shale. Whereas previous reconstruction techniques aimed at building supramolecular structures of kerogen which reproduce some chemical features^{6,16–22},

our strategy adopts a different standpoint by relying on molecular constraints only. This hybrid experimental/theoretical reconstruction technique generates molecular models using a classical yet realistic description of the chemistry at play while constraining its atomic structure to converge towards that of the experimental sample. Thanks to the use of a reactive force field, our strategy allows one to reconstruct molecular models without imposing a given distribution of chemical groups, carbon hybridization, and so on. Owing to the general nature of the proposed strategy (which relies on atomic structural and chemical composition data only), use of this method can be envisaged for any heterogeneous and disordered material. However, some a priori knowledge about the material chemistry and physics is always required to ensure that the method is robust. Moreover, despite the use of dual constraints over the structure and energy of the reconstructed models, some chemical defects (from a few % up to $\sim 10\%$) cannot be avoided owing to limitations in the choice of the reactive force field and quality of the phase space sampling. In particular, the quality of the phase space sampling attainable is limited and leads to a small amount of defects such as non-saturated atoms, isolated molecules, and possible polyaromatic arrangements explored only partially. Finally, the present strategy, which does not consider the presence of clays and minerals in shales, aims at developing a realistic model of mature and immature kerogen only.

To assess their relevance, our molecular models are compared against available experimental indicators such as the hybridization and distribution of chemical groups. Finally, our strategy is validated using a closure scheme in which the realistic models of mature and immature kerogens are shown to capture independent experimental data such as pore size distribution, vibrational density of states, and stiffness. These atom-scale models are then used to predict essential properties of kerogen such as an increase in the sp^2/sp^3 ratio on maturation. Such an ageing-induced chemical transition entails a crossover from plastic to brittle rupture mechanisms, which affects gas shale frackability-and thus recovery. Although kerogens are multiscale materials, this work aims at unravelling their unknown microscopic structure-that is, by decoding the 'genome' of kerogen, which is recognized as the cornerstone for understanding its complex adsorption, mechanical and transport properties. Despite some unavoidable defects (such as a few % of sp^1 carbons and unsaturated molecules) and limitations arising from the proposed strategy, the development of realistic molecular models of kerogen is the building block required to establish a multiscale approach. In particular, such atom-scale models should allow one to reconcile classical modelling³⁰⁻³² with molecular descriptions which properly account for strong adsorption in kerogen and the breakdown of hydrodynamics at the nm scale¹⁴.

Molecular reconstruction

The three elements to reconstruct realistic kerogen models are chemical composition, the experimental molecular texture parameter that constrains—on average—possible atomistic configurations, and density. Although the biologic origin of kerogen is not accounted for, as initial specific macromolecules are not imposed, this strategy avoids having to describe the maturation process over geologic times by focusing on the microscopic structure of kerogen at a given time.

The versatility of our approach is illustrated by considering three kerogens that differ in geologic origin, elemental composition (see Supplementary Table 1) and maturity, as shown in the van Krevelen diagram (Fig. 1a): an immature sulphur-rich marine kerogen from the Middle East (MEK); an immature marine kerogen from the carbonate-rich Eagle Ford Play (EFK), and a mature marine kerogen from the clay-rich Marcellus Play (MarK; ref. 33). A mineral-free shungite (PY02), a rather exceptional kerogen type from Russia, was also considered³⁴⁻³⁶. Because shungite and

kerogen have different origins and follow different maturations, PY02 is not a kerogen *per se*, but does share many features with overly mature kerogens, such as small heteroatom contents and large proportions of sp^2 carbons, mainly arranged in large aromatic clusters³⁶. The kerogen, which is embedded within various minerals in the organic-rich shale samples (Fig. 1b), was isolated by means of acid demineralization with the critical point drying method³⁷. This technique consists of removing the solvents used for demineralization at temperatures above their critical point (see details in the Supplementary Information). This preserves the

(see details in the Supplementary Information). This preserves the kerogen microstructure, as no solvent surface tension is involved on extraction. For the four samples, both diffraction data (S(Q)) and generalized phonon densities of states (GDOS) were obtained. Whereas the former is employed for reconstruction, the latter will be used for validation. All demineralized kerogens possess, to various extents, small amounts of pyrite, but almost no other minerals (see the diffraction patterns in Supplementary Fig. 1).

We then employ the MD-HRMC method to build the kerogen molecular models; an HRMC method enriched by MD relaxation steps in the framework of a simulated annealing procedure, that is at least one order of magnitude faster than the classical HRMC method for obtaining structures of similar quality²⁹. Like the classical HRMC method (refs 38,39), the MD-HRMC method consists of minimizing the configurational energy while constraining possible molecular configurations by means of the pair distribution function, G(r), obtained from the inverse Fourier transform of the experimental structure factor S(Q). Peaks in G(r) are related to the probability of finding an atom at a distance r from another atom. For the four samples, these G(r) functions are shown in Fig. 2. The only other constraints used in the MD-HRMC method are the experimental atomic ratios and densities. Figure 2 shows that the reconstruction is efficient and complete; the final structures exhibit G(r) close to their experimental counterparts. Figure 1c shows typical molecular configurations obtained for the four samples with a density 1.2 g cm^{-3} .

Finally, kerogen can exhibit a range of density values which is difficult to assess experimentally (see discussion in Supplementary Information). To overcome this difficulty, we generated for each sample (fixed chemical composition and G(r)) a database of molecular reconstructions with physical densities from 0.8 to 2.2 g cm⁻³. For each structure, we calculated a set of experimentally accessible properties: vibrational properties (GDOS), textural properties (pore size distributions and nitrogen adsorption isotherms), and elastic properties (bulk modulus *K*, shear modulus *G*, and so on). These properties are then compared with experimental values to determine the density that matches best the set of experiments.

Model validation against experiments

All kerogen models generated according to the strategy above are disordered porous carbons with a rich chemistry, including both sp² and sp³ carbons arranged in a main carbon backbone and some isolated molecules within the pores of the structure (Fig. 1c). Supplementary Table 2 shows the potential energy per atom for the four models at different densities. As expected, all energies are greater than that of graphite (-7.4 eV) and tend towards this asymptotic value as the heteroatom content decreases. This confirms that the OREBO potential accurately describes the energy of bonds involving C atoms⁴⁰. The structures are also found to be stable on energy minimization using the density-functional-based tight binding method (a semi-empirical quantum-chemistry method). Although it is difficult to estimate the contribution of unphysical defects in the reconstructed models (see below), they necessarily affect the potential energy, which is greater than that for graphite. On the other hand, the porous and amorphous nature of the kerogens, even without such defects, would also correspond to less stable configurations.





Figure 2 | Reconstruction of molecular models of kerogens. Experimental (lines) and simulated (points) C-C pair distribution functions *G*(*r*) for the four samples with $\rho = 1.2 \text{ g cm}^{-3}$. The colour code of Fig. 1a is used. The diffraction pattern for MarK is largely dominated by the crystalline peaks corresponding to pyrite, such that no structural information can be extracted from such diffraction measurements. However, as shown in Fig. 3a, PYO2 and MarK possess a very similar vibrational density of states, with comparable amounts of aromatic rings, alkyl chains, and *sp*² carbons. Consequently, we assumed that MarK and PYO2 have identical *G*(*r*) but different atomic compositions.

As expected, the carbon backbone in the molecular kerogen models contains aromatic clusters (which are larger in mature samples, as shown in Supplementary Fig. 2), alkyl chains, and various COH functions. Supplementary Fig. 4 shows the C-C-C bond angle distribution for the four kerogens. These data are consistent with the predominant sp^2 and sp^3 features in mature and immature kerogens, respectively. Immature samples (MEK and EFK) exhibit a peak around 110°, whereas mature samples (PY02 and MarK) exhibit a peak around 118°. Minor peaks around 60° correspond to a few three-membered rings, whereas the weak and broad feature between 140° and 180° in mature samples corresponds to a few unsaturated cumulene chains (sp1 chains): 1.0 at.% in EFK, 0.41 at.% in MEK, 6.1 at.% in MarK and 8.0 at.% in PY02—the fraction of *sp*¹ carbons for MarK is greater than that of sp³ owing to its maturity, which leads to a dominant fraction of sp² carbons. Nevertheless, Supplementary Fig. 4 shows that these unphysical features (three-membered rings and cumulene chains) are present in a small proportion. In addition to these defects, some isolated molecules (0 \sim 10 at.%) trapped in the kerogen pores were also found to coexist with the large carbon backbone. Full characterization of these defects is available in the Supplementary Information (Section III). Although most of these defects form small realistic molecules, we also observe free radicals. Owing to their limited occurrence, these isolated molecules do not affect the vibrational, mechanical, adsorption and fracture properties, but they necessarily contribute to the heat of formation, activation energy, and so on, which should therefore be considered with caution.

The first peak in the experimental G(r) (Fig. 2) reveals the difference in hybridization state and bonding between immature and mature samples. The peak at ~1.54 Å in immature kerogens (MEK and EFK) is characteristic of sp^3 carbons, whereas the peak at ~1.41 Å in mature samples (MarK and PY02) corresponds to sp^2 carbons, as recently predicted by Ungerer *et al.* using molecular dynamics relaxation of macromolecular kerogen models without constraining to recover the experimental G(r) (ref. 22). Furthermore, despite their disordered skeleton, the G(r) of immature samples exhibits broad peaks and shortrange order only (whereas mature samples show a more ordered structure with well-defined peaks up to ~15 Å). Maturity is thus

ARTICLES

recognized to affect both the carbon hybridization state and the long-range texture.

Whereas G(r) is used as a constraint in the MD–HRMC method, a true validation consists in comparing the GDOS of the molecular models with their experimental counterpart. In contrast to Raman and infrared spectroscopies, all vibrational modes are accessible to neutron spectroscopy, as there are no selection rules for the neutron-phonon interaction. Owing to the dominant scattering cross-section of hydrogen, the GDOS represents mostly vibrational modes involving hydrogen motion. Because the vibrational energy differs for hydrogens linked to sp² or sp³ carbons, a successful comparison of experimental and simulated GDOS modes is a means to independently validate the difference in hybridization state between mature and immature samples. To this end, experimental GDOS were obtained from inelastic neutron scattering, whereas simulated GDOS were calculated from the weighted Fourier transform of the atomic velocity autocorrelation function obtained by MD simulations (see Methods). The comparison between simulation and experimental GDOS (Fig. 3a) confirms the importance of the hybridization state in distinguishing mature from immature kerogens. In fact, the immature kerogens (EFK and MEK) possess almost identical GDOS: the C-H stretching modes around 2,960 cm⁻¹, the CH₂ scissor mode at 1,460 cm⁻¹, and the CH₃(s) mode at 1,375 cm⁻¹ are hallmarks of sp^3 carbons. Finally, the sharp peak at 720 cm⁻¹, which is typical of CH₂(s) rock modes in alkanes with at least seven carbons, indicates the expected presence of long alkane chains in immature samples. A similar consistency between experimental and simulated GDOS is found for the mature systems (MarK and PY02). Owing to their lower hydrogen content, mature kerogens exhibit lower intensities than immature ones. Specifically, above \sim 700 cm⁻¹, MarK and PY02 show very similar features (despite an intensity difference due to a higher H content in MarK); namely, an aromatic C-H stretching mode at 3,090 cm⁻¹, and aromatic in- and out-of-plane bending modes between 800 cm⁻¹ and 1,200 cm⁻¹. These spectra confirm that mature samples primarily contain sp^2 carbons, whereas immature samples primarily contain sp³ carbons (in agreement with ¹³C NMR data, as shown in Supplementary Table 1), a trend that is reproduced well by our models. PY02 exhibits an additional peak at around $\sim 600 \text{ cm}^{-1}$ (not observed for MarK), which indicates the presence of confined water within some closed porosity in PY02 (the water librational mode at \sim 560 cm⁻¹ is blueshifted to $\sim 600 \,\mathrm{cm}^{-1}$ when water is confined⁴¹). This result is in agreement with our molecular simulation study of adsorption, showing that 60% of the pores are closed in PY02 (see Supplementary Fig. 12, comparing adsorption simulations with and without accounting for pore connectivity). The isobaric heat capacity c_{p} of the kerogen models can be estimated from the DOS and simple thermodynamic data such as the thermal expansion coefficient α and compressibility β (Supplementary Information). Supplementary Fig. 9 shows that the heat capacity c_{p} of the kerogen models are in good agreement with experimental data for kerogens and coals $(c_{\rm p} \sim 0.75 - 2\,{\rm J\,K^{-1}\,g^{-1}})$ (refs 22,42,43). In particular, the molecular models of mature and immature kerogens capture the decrease in c_{p} on maturation.

The comparison between the simulated and experimental pore size distributions (Fig. 3b) allows further characterization and validation of the microtexture of our models of mature and immature kerogen. The pore size distribution of the molecular models is determined by calculating the maximum radii of test particles inside the pore cavities using a Monte Carlo scheme⁴⁴. These pore size distributions, which are consistent with those obtained from Molecular Dynamics²¹, are in agreement with the experimental pore size distributions determined from CO₂ adsorption experiments for the same samples (after demineralization to avoid possible microporosity in the mineral parts such as clays or at the interface with kerogen). The pore size distributions determined for our molecular



Figure 3 | **Experimental validation of the molecular models.** The colour code of Fig. 1a is used. **a**, Experimental (lines with error bars) and simulated GDOS. The vertical dashed line at 3,000 cm⁻¹ is a guide to the eye. The error bars are computed from the square root of the neutron count. **b**, Simulated (lines, $\rho = 1.2 \text{ g cm}^{-3}$) and experimental (bars) pore size distribution for EFK (red) and MarK (blue). The experimental pore size distributions are obtained for demineralized samples (after 24 h heating treatment at 80 °C under secondary vacuum) from CO₂ adsorption at room temperature to probe micropores ($\leq 10 \text{ Å}$).

models are also in agreement with the microporosity reported in unprocessed organic-rich shales^{7,11}. Owing to their nanometric size, our molecular models cannot capture mesoporosity (>2 nm), but the results above show that they capture the microporosity of real samples. Although advanced techniques such as small-angle scattering and/or combinations of various adsorption techniques have demonstrated the very wide pore size distributions in kerogen, recent experiments have established that microporosity amounts for up to 40–60% of the total pore volume in demineralized kerogen^{45,46}. In addition to this important microporosity, mesoporosity corresponding to distorted cylindrical-like or slit-like mesopores^{10,47-49} also represents a significant contribution to the total porous volume. Moreover, in the case of non-demineralized shales, clay also contains an intrinsic microporosity which contributes to the porous volume. Both our experimental data and those obtained for our models show that the structural order (and therefore the maturity) significantly affects the microporosity of kerogen. For a given density, mature samples possess larger micropores, and hence a larger amount of hydrocarbon, due to their more advanced maturation stage (kerogen decomposition). In contrast, immature samples exhibit smaller micropores (\sim 7 Å), and the presence of recoverable hydrocarbon is thus restricted to larger mesopores, which are beyond the focus of this investigation.

NATURE MATERIALS DOI: 10.1038/NMAT4541

Although G(r), GDOS and pore size distributions provide invaluable insights into the intimate interplay between hybridization state (namely the sp^2/sp^3 ratio), microtexture and maturity, the relation with kerogen density remains ambiguous; that is, they are necessary but not sufficient conditions to constrain possible molecular configurations to a single density. This observation calls for an additional closure condition at a scale representative of that of the kerogen density, which includes the solid carbon backbone and other heteroatoms embedded in the skeleton and the micropores. At this scale, acoustic and elasticity properties constitute a natural constraint, as they are associated with characteristic wavelengths that average the behaviour of both the solid and the pore space. The approach is illustrated in Fig. 4; for each sample (fixed chemical composition), using the MD-HRMC method we generate a number of molecular configurations that satisfy the G(r), but which differ in density in the range of $0.8-2.2 \,\mathrm{g \, cm^{-3}}$. For each sample, Fig. 4 shows both the data for the kerogen structures with and without the isolated molecules in kerogen porosity that were found to coexist with the carbon backbone. As expected, these molecules do not significantly affect the mechanical properties of the kerogen structures as they do not belong to the material skeleton. However, these molecules make a slight contribution to the density of the kerogen structures such that the corrected density is shifted towards lower density. The comparison between simulation and experimental elastic properties then ultimately provides an independent closure of the reconstruction approach, in the form of a kerogen sample density that relates to the elasticity properties. To prove this concept, we experimentally measured the bulk modulus of PY02 using X-ray diffraction under pressure (inset of Fig. 4), resulting in a value of $K_{\rm PY02} \simeq 13.7 \pm 2$ GPa. Then, using this value in the simulationgenerated stiffness-density plots in Fig. 4 provides an estimate of the sample density of 1.1 ± 0.1 g cm⁻³ (~1.4 g cm⁻³ if the effect of grain packing at mesoscale is taken into account⁵⁰).

In addition to serving as a closure scheme, the data in Fig. 4 show that, for a given chemical composition, the bulk modulus *K*, the shear modulus G and the Young's modulus E vary exponentially with density. On the other hand, the Poisson ratio $\nu \simeq 0.25$ is found to be insensitive to sample density, composition and maturity (see Supplementary Fig. 8). Although establishing accurate quantitative structure-property relations would require more samples (with different compositions, maturities, and so on), qualitative insights can be gained into the link between chemical composition, maturity and mechanical properties of kerogen. For a given density, no clear trend is observed between the H/C atomic ratio of a sample and its elastic moduli and Poisson ratio. In contrast, for a given density, a decrease in the elastic moduli is observed on increasing the O/C atomic ratios. The particular case of the mature samples (PY02 and MarK)—which share the same G(r) and thus a similar high sp^2/sp^3 ratio—suggests that the large O/C ratio for MarK is at the origin of the decrease in the bulk and shear moduli. This is thought to be due to the important role of heteroatoms on wave propagation in these disordered materials. This finding, together with the scaling relations between density and elastic properties, suggests that stiffness and density are two parameters that could be assessed to probe both the maturity and the chemical composition of kerogens.

Maturity-induced ductile-to-brittle crossover

The molecular models of immature and mature kerogens allow one to gain insights into the structural deformation and fracture mechanisms of kerogen, which play a critical role in the debate about sustainable hydrocarbon recovery using severe stimulation such as fracking. Investigation of molecular models contributes by answering whether the response to tensile loading is ductile or brittle. Ductile failure is associated with bulk plastic deformations, whereas brittle fracture localizes along surfaces—a difference which



Figure 4 | **Elastic properties of molecular models of kerogen.** Bulk modulus *K* (top), shear modulus *G* (middle) and Young modulus *E* (bottom) as a function of the sample density. The colour code of Fig. 1a is used. The filled and open symbols, respectively, are for kerogen structures with and without the trapped isolated molecules that coexist with the carbon backbone (see discussion on the trapped molecules in the Supplementary Information). The evolutions of *K*, *G* and *E* are fitted by an exponential function, whereas v is fitted by a linear function. The light blue area corresponds to the bulk modulus measured experimentally for PYO2 by means of X-ray diffraction. This value was obtained by fitting the sample volume V/V_0 as a function of pressure *P* against the Birch-Murnaghan equation of state (data shown in the inset). Each value is averaged over eight representative configurations, with an error bar corresponding to the standard deviation.

has immediate impact on hydrocarbon recovery from kerogen's micropores. We have simulated the stress-strain response of the samples in a strain-controlled confined tension (Fig. 5a, see Methods and Supplementary Information for details) using MD at 300 K with the reactive force field OREBO (ref. 40) to allow bond breaking and formation. The stress was determined from the virial expression and a strain rate of 109 s⁻¹ was chosen to minimize strain-rate effects⁵¹. The responses of all systems are a sequence of elastic loading interspersed with discrete stress drops reminiscent of strain localization events (Fig. 5a). Yet, a striking difference is to be noted between the immature (MEK and EFK) and mature (PY02 and MarK) kerogens in the pattern of the stress-strain seesaw features. In immature kerogens, the slope in the elastic loading branches varies little with increasing deformation. If extrapolated to zero stress, such a behaviour is associated with residual strains, reminiscent of a plastic behaviour. In contrast, mature kerogens exhibit elastic loading slopes that pivot around the origin of the stress-strain plot. These two behaviours are indicative, respectively,

ARTICLES



Figure 5 | **Fracture testing of the four kerogens with** $\rho = 1.2 \text{ g cm}^{-3}$. The colour code of Fig. 1a is used. **a**, Normalized stress-strain curves during a confined tensile test where the *x* direction is elongated while the *y* and *z* directions are kept fixed. The dashed lines are linear fits of several increasing portions. **b**, 3D representation of the atomic displacements longer than 1 Å computed at the stress drops marked by stars in **a**. Carbons are represented in grey and the oxygens in red, the orange planes are guides to the eyes to help see the planar localization of the atomic displacements. **c**, Evolution of the total number of *Sp*³ (filled symbols) and *sp*² (open symbols) carbon atoms N_{sp^x}/N_{sp^x} (initial) (right) during the confined tensile tests. EFK and MEK contain a majority of *sp*³ carbons, whereas PYO2 and MarK contain a majority of *sp*¹ carbons being rare in all samples.

of ductile behaviour for immature kerogens and brittle behaviour for mature kerogens. This is corroborated by the difference in atomic deformation mechanism in Fig. 5b; whereas the deformation in mature systems occurs localized within planes, indicative of fracture formation, significant atomic displacements (>1 Å) in the immature systems are apparently randomly dispersed throughout the sample.

It is intriguing to reconcile the observed ductile-to-brittle crossover with maturity indicators identified by G(r), GDOS and elasticity measurements. To this end, during the deformation of mature and immature kerogens we monitored the total number of C–C bonds (N_{CC}) as well as the number of sp^2 (N_{sp^2}) and sp^3 (N_{sp^3}) carbons (the hybridization state is determined from the number of neighbours, two atoms being considered neighbours if they are at a distance <1.9 Å). When plotted against the strain (Fig. 5c), $N_{\rm CC}$ increases in mature samples, whereas it decreases in immature samples. N_{sp^2} and N_{sp^3} reveal that the increase of N_{CC} in mature systems is due to the transformation of trivalent (sp^2) into tetravalent carbons (sp^3) . The opposite occurs in immature samples, as sp^3 bonds break into alkane chains and rearrange to form sp^2 bonds at the end of two resulting chains. This explains the decrease in $N_{\rm CC}$ and is suggested to be at the origin of dispersed atomic displacements; much akin to the dislocation mechanisms that drive the plastic hardening behaviour of metals⁵².

In summary, the overall picture which thus emerges is that kerogen's intrinsic fracture behaviour is predetermined by the hybridization state, namely the sp^2/sp^3 ratio, and thus kerogen's maturity. This result is likely to be general, as the increase in sp^2/sp^3 ratio with maturity is widely observed. Our molecular structures cannot be considered a microscopic model of shales (because minerals and large-scale porosities are missing), but it represents a fundamental building block from which a multiscale strategy can be developed to determine adsorption, transport and mechanical properties of such heterogeneous and disordered media. Although further work is needed to describe coexistence at different scales between clay, other minerals and kerogen, the present study illustrates in a bottom-up fashion the prospect of using advanced molecular reconstruction methods and statistical mechanics molecular modelling to reveal fundamental knowledge.

Methods

Methods and any associated references are available in the online version of the paper.

Received 27 January 2015; accepted 14 December 2015; published online 1 February 2016

References

- 1. Kerr, R. A. Natural gas from shale bursts onto the scene. *Science* **328**, 1624–1626 (2010).
- Cueto-Felgueroso, L. & Juanes, R. Forecasting long-term gas production from shale. *Proc. Natl Acad. Sci. USA* 110, 19660–19661 (2013).
- Osborn, S. G., Vengosh, A., Warner, N. R. & Jackson, R. B. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc. Natl Acad. Sci. USA* 108, 8172–8176 (2011).
- Howarth, R. W., Ingraffea, A. & Engelder, T. Natural gas: should fracking stop? Nature 477, 271–275 (2011).
- Vidic, R., Brantley, S., Vandenbossche, J., Yoxtheimer, D. & Abad, J. Impact of shale gas development on regional water quality. *Science* 340, 1235009 (2013).
- Vandenbroucke, M. & Largeau, C. Kerogen origin, evolution and structure. Org. Geochem. 38, 719–833 (2007).
- Clarkson, C. *et al.* Pore structure characterization of north American shale gas reservoirs using USANS/SANS, gas adsorption, and mercury intrusion. *Fuel* 103, 606–616 (2013).
- Mastalerz, M., He, L., Melnichenko, Y. B. & Rupp, J. A. Porosity of coal and shale: insights from gas adsorption and SANS/USANS techniques. *Energy Fuels* 26, 5109–5120 (2012).
- Thomas, J. J., Valenza, J. J., Craddock, P. R., Bake, K. D. & Pomerantz, A. E. The neutron scattering length density of kerogen and coal as determined by CH₃OH/CD₃OH exchange. *Fuel* 117, 801–808 (2013).
- Gu, X., Cole, D. R., Rother, G., Mildner, D. F. R. & Brantley, S. L. Pores in Marcellus shale: a neutron scattering and FIB-SEM study. *Energy Fuels* 29, 1295–1308 (2015).
- Wang, Y., Zhu, Y., Chen, S. & Li, W. Characteristics of nanoscale pore structure in northwestern Hunan shale gas reservoirs using field emission scanning electron microcopy, high-pressure mercury intrusion, and gas adsorption. *Energy Fuels* 28, 945–955 (2014).

NATURE MATERIALS DOI: 10.1038/NMAT4541

- Firouzi, M., Rupp, E. C., Liu, C. W. & Wilcox, J. Molecular simulation and experimental characterization of the nanoporous structures of coal and gas shale. *Int. J. Coal Geol.* **121**, 123–128 (2014).
- Aguilera, R. et al. Flow units: from conventional to tight-gas to shale-gas to tight-oil to shale-oil reservoirs. SPE Reservoir Eval. Eng. 17, 190–208 (2014).
- Falk, K., Coasne, B., Pellenq, R. J.-M., Ulm, F.-J. & Bocquet, L. Subcontinuum mass transport of condensed hydrocarbons in nanoporous media. *Nature Commun.* 6, 6949 (2015).
- 15. Baihly, J. D. *et al. SPE Annual Technical Conference and Exhibition* (Society of Petroleum Engineers, 2010).
- Behar, F. & Vandenbroucke, M. Chemical modelling of kerogens. Org. Geochem. 11, 15–24 (1987).
- Siskin, M. et al. Composition, Geochemistry and Conversion of Oil Shales 143–158 (Springer, 1995).
- Lille, Ü., Heinmaa, I. & Pehk, T. Molecular model of Estonian kukersite kerogen evaluated by ¹³C MAS NMR spectra. *Fuel* 82, 799–804 (2003).
- Kelemen, S. *et al.* Direct characterization of kerogen by X-ray and solid-state ¹³C nuclear magnetic resonance methods. *Energy Fuels* 21, 1548–1561 (2007).
- Orendt, A. M. *et al.* Three-dimensional structure of the Siskin Green River oil shale kerogen model: a comparison between calculated and observed properties. *Energy Fuels* 27, 702–710 (2013).
- Collell, J. et al. Molecular simulation of bulk organic matter in type II shales in the middle of the oil formation window. Energy Fuels 28, 7457–7466 (2014).
- Ungerer, P., Collell, J. & Yiannourakou, M. Molecular modelling of the volumetric and thermodynamic properties of kerogen: influence of organic type and maturity. *Energy Fuels* 29, 91–105 (2015).
- Yiannourakou, M. *et al.* Molecular simulation of adsorption in microporous materials. *Oil Gas Sci. Technol.* 68, 977–994 (2013).
- Mullins, O. C. *et al.* Advances in asphaltene science and the Yen–Mullins model. *Energy Fuels* 26, 3986–4003 (2012).
- Sedghi, M., Goual, L., Welch, W. & Kubelka, J. Effect of asphaltene structure on association and aggregation using molecular dynamics. *J. Phys. Chem. B* 117, 5765–5776 (2013).
- 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* 19, 2495–2499 (2005).
- Tissot, B. & Welte, D. (eds) Petroleum Formation and Occurrence (Springer, 1984).
- van Krevelen, D. W. Coal: Typology, Chemistry, Physics, Constitution (Elsevier, 1961).
- Bousige, C., Boţan, A., Ulm, F.-J., Pellenq, R. J.-M. & Coasne, B. Optimized molecular reconstruction procedure combining hybrid reverse Monte Carlo and molecular dynamics. *J. Chem. Phys.* 142, 114112 (2015).
- Firouzi, M., Alnoaimi, K., Kovscek, A. & Wilcox, J. Klinkenberg effect on predicting and measuring helium permeability in gas shales. *Int. J. Coal Geol.* 123, 62–68 (2014).
- Bažant, Z. P., Salviato, M., Chau, V. T., Viswanathan, H. & Zubelewicz, A. Why fracking works. J. Appl. Mech. 81, 101010 (2014).
- Monteiro, P. J., Rycroft, C. H. & Barenblatt, G. I. A mathematical model of fluid and gas flow in nanoporous media. *Proc. Natl Acad. Sci. USA* 109, 20309–20313 (2012).
- Pomerantz, A. E. *et al.* Sulfur speciation in kerogen and bitumen from gas and oil shales. *Org. Geochem.* 68, 5–12 (2014).
- Melezhik, V., Filippov, M. & Romashkin, A. A giant palaeoproterozoic deposit of shungite in NW Russia: genesis and practical applications. Ore Geol. Rev. 24, 135–154 (2004).
- Melezhik, V. A. *et al.* Petroleum surface oil seeps from a palaeoproterozoic petrified giant oilfield. *Terra Nova* 21, 119–126 (2009).
- Kovalevski, V., Buseck, P. R. & Cowley, J. Comparison of carbon in shungite rocks to other natural carbons: an X-ray and TEM study. *Carbon* 39, 243–256 (2001).
- Suleimenova, A. *et al.* Acid demineralization with critical point drying: a method for kerogen isolation that preserves microstructure. *Fuel* 135, 492–497 (2014).
- Opletal, G. *et al.* Hybrid approach for generating realistic amorphous carbon structure using Metropolis and reverse Monte Carlo. *Mol. Simul.* 28, 927–938 (2002).

- Jain, S., Pellenq, R., Pikunic, J. & Gubbins, K. Molecular modeling of porous carbons using the hybrid reverse Monte Carlo method. *Langmuir* 22, 9942–9948 (2006).
- Ni, B., Lee, K.-H. & Sinnott, S. B. A reactive empirical bond order (REBO) potential for hydrocarbon-oxygen interactions. *J. Phys. Condens. Matter* 16, 7261–7275 (2004).
- 41. Bellissent-Funel, M.-C. Status of experiments probing the dynamics of water in confinement. *Eur. Phys. J. E* 12, 83–92 (2003).
- 42. Tomeczek, J. & Palugniok, H. Specific heat capacity and enthalpy of coal pyrolysis at elevated temperatures. *Fuel* **75**, 1089–1093 (1996).
- Savest, N. & Oja, V. Heat capacity of kukersite oil shale: literature overview. Oil Shale 30, 184–192 (2013).
- Bhattacharya, S. & Gubbins, K. E. Fast method for computing pore size distributions of model materials. *Langmuir* 22, 7726–7731 (2006).
- Rexer, T. F., Benham, M. J., Aplin, A. C. & Thomas, K. M. Methane adsorption on shale under simulated geological temperature and pressure conditions. *Energy Fuels* 27, 3099–3109 (2013).
- Rexer, T. F. T., Mathia, E. J., Aplin, A. C. & Thomas, K. M. High-pressure methane adsorption and characterization of pores in Posidonia shales and isolated kerogens. *Energy Fuels* 28, 2886–2901 (2014).
- Chen, C., Hu, D., Westacott, D. & Loveless, D. Nanometer-scale characterization of microscopic pores in shale kerogen by image analysis and pore-scale modeling. *Geochem. Geophys. Geosyst.* 14, 4066–4075 (2013).
- Guo, X., Li, Y., Liu, R. & Wang, Q. Characteristics and controlling factors of micropore structures of the Longmaxi shale in the Jiaoshiba area, Sichuan Basin. *Nat. Gas Ind. B* 1, 165–171 (2014).
- Klaver, J., Desbois, G., Littke, R. & Urai, J. L. BIB-SEM characterization of pore space morphology and distribution in postmature to overmature samples from the Haynesville and Bossier shales. *Mar. Petrol. Geol.* 59, 451–466 (2015).
- Vandamme, M., Ulm, F.-J. & Fonollosa, P. Nanogranular packing of C–S–H at substochiometric conditions. *Cement Concrete Res.* 40, 14–26 (2010).
- Mi, X. & Shi, Y. *MRS Proceedings* Vol. 1224, 1224-FF10-10 (Cambridge Univ. Press, 2009).
- He, G., Eckert, J., Löser, W. & Schultz, L. Novel Ti-base nanostructure–dendrite composite with enhanced plasticity. *Nature Mater.* 2, 33–37 (2002).
- Seewald, J. S. Organic-inorganic interactions in petroleum-producing sedimentary basins. *Nature* 426, 327–333 (2003).

Acknowledgements

This work was supported by the X-Shale project enabled through MIT's Energy Initiative, with sponsorship provided by Shell and Schlumberger. Additional support was provided by the ICoME2 Labex (ANR-11-LABX-0053) and the A*MIDEX projects (ANR-11-IDEX-0001-02) co-funded by the French programme 'Investissements d'Avenir' managed by ANR, the French National Research Agency. The authors thank M. Hubler (MIT) and J. Gelb (Carl Zeiss X-ray Microscopy) for providing the X-Ray Microscopy image of raw shale, and A. Saul, J. M. Leyssale, H. Van Damme and A. Archereau for fruitful discussions. The neutron scattering experiments were carried out at the Spallation Neutron Source, which is sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, US Department of Energy, under Contract No. DE-AC05-00CR22725 with Oak Ridge National Laboratory.

Author contributions

C.B., B.C., R.J.-M.P. and F.-J.U. designed the work. C.B. performed the simulations and data treatment. A.S. and A.E.P. performed the acid demineralization with critical point drying. A.S., A.E.P., C.M.G. and C.V.-G. measured the nitrogen and CO₂ adsorption isotherms. C.B. and B.C. performed the scattering measurements with G.V., G.G., M.F. and C.W. as instrument scientists. C.B., B.C., R.J.-M.P. and F.-J.U. wrote the paper.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.C.

Competing financial interests

The authors declare no competing financial interests.



ARTICLES

NATURE MATERIALS DOI: 10.1038/NMAT4541

Methods

X-ray and neutron diffraction. X-ray scattering experiments were performed on ID11 and ID27 at ESRF in Grenoble. Ambient temperature and pressure measurements were performed on ID11 with an incoming X-ray energy of 42 keV (0.29523 Å), and data were collected in the [0.9, 17.5] Å⁻¹ range. An incoming X-ray energy of 36 keV (0.344 Å) was used on ID27, which allows the measurement of reliable data in the [0.8, 16] Å⁻¹ range. The sample PY02 was placed in a diamond anvil cell together with a ruby (to measure the pressure by fluorescence) and the pressure was increased up to 5.1 GPa. The neutron diffraction measurements were performed on NOMAD (ref. 54) at SNS in Oak Ridge, Tennessee, and the G(r) functions were computed using data in the [0.35, 31.4] Å⁻¹ range. Neutron diffraction is highly sensitive to hydrogen impurities (hydrogen is a purely incoherent scatterer, whereas diffraction is a coherent process), which results in a strong incoherent background. The X-ray data were therefore preferred for the three kerogens, whereas the neutron data were used for PY02-a wider range in wavevector is preferable to get the best resolution for the pair distribution function. The G(r) obtained from the Fourier transform of the scattering experimental data is a total pair distribution function which encompasses all atom-atom contributions (C, O, and H). However, owing to the fact that O atoms are present in small amounts in the samples (at most $\sim 10\%$ for MarK), the simulated G(r)considered in the reconstruction strategy are assumed to correspond to the C-C contribution only (whereas contributions involving H atoms are necessarily very small, as H is an incoherent neutron scatterer and has a negligible X-ray scattering cross-section). Neglecting the C-O contribution in the total pair correlation function is a reasonable approximation as it remains within the error bar induced by the various corrections made to the diffraction data (Fourier transform, corrections for the apparatus geometry, pyrite removal from the data, and so on). To validate this approximation, we have also verified that including these O-O and O–C contributions to the G(r) has little impact on the resulting G(r)functions (Supplementary Fig. 6). Full data treatment is reported in the Supplementary Information.

Inelastic neutron scattering. We have performed inelastic neutron scattering measurements on the VISION spectrometer at SNS, to determine the phonon density of states (GDOS) in the [100, 4,000] cm⁻¹ energy range with a 1.5% $\Delta E/E$ resolution. Data were collected at 5 K, and the collection times and sample masses were 39 h for PY02 (525 mg), 3 h for MEK (584 mg), 3 h for MarK (497 mg) and 5 h 30 for EFK (223 mg). The data are corrected for the sample holder scattering (vanadium cylindrical can) and normalized to monitor neutron count and sample mass. In the case of powders, the only 'selection rule' is an intensity factor, the scattered intensity being proportional to the ratio of the scattering length and the mass of the scatterer. This ratio is dominant in the case of the incoherent scattering from hydrogen⁵⁵, so that the intensity is dominated by modes involving hydrogen motion. Data treatment is reported in the Supplementary Information.

MD-HRMC. We reconstructed molecular models of mature and immature kerogens in organic-rich shales using a technique that combines molecular dynamics and hybrid reverse Monte Carlo simulations. This method, referred to as MD–HRMC (ref. 29), allows one to produce structures satisfying the experimental pair distribution function as well as the atomic chemical composition and density. The starting configuration is a cubic box (with periodic boundary conditions) containing a random distribution of atoms matching the experimental density and atomic ratios. Following the MD–HRMC procedure in ref. 29, a simulated annealing procedure of HRMC blocks is performed from 10⁵ K down to 300 K, with temperature steps such as $T_{new} = 0.9T_{old}$. Between each temperature change, a MD relaxation is performed at $T_{\rm MD} = 2,000$ K. During the HRMC block at T_x , the atoms are displaced randomly and the move is accepted with the probability:

$$P_{C}^{acc}(T_{\chi}) = \min\left[1, \exp\left(-\frac{1}{T_{\chi}}\left(\Delta\chi^{2} + \frac{\Delta U}{\omega}\right)\right)\right] \text{ for C atoms}$$

$$P_{het}^{acc}(T_{\chi}) = \min\left[1, \exp\left(-\frac{\Delta U}{\omega T_{\chi}}\right)\right] \text{ for heteroatoms}$$

The use of a combined minimization technique relying on the G(r) function and the energy is needed as many structures can be built that reproduce the G(r) (or

equivalently the structure factor). Finally, many stable configurations reproducing the experimental pair distribution function are still possible, but these structures are assumed to be representative of the same experimental system.

The procedure above was implemented into LAMMPS (ref. 56), and the total energy was computed using the bond order reactive force field OREBO (ref. 40). Although reactive force fields are often developed for compounds that form at high temperatures or pressures, OREBO was found to capture the structure and chemical stability of molecules and polymers, as well as the chemical reactions involved in these systems. Consequently, although it was not specifically designed to model the formation and maturation of kerogen, we assume that it can also be applied to this class of complex carbonaceous compounds. At the end of the MD-HRMC procedure, the simulated structures can be further relaxed using MD simulations at 300 K. Supplementary Fig. 6 shows that, on relaxation, some of the correlations in the G(r) are modified—as noted in ref. 29. Such local relaxation after switching off the HRMC procedure cannot be avoided, as the force field makes the structure converge towards a different free energy minimum. Nevertheless, although such post-MD-HRMC relaxation is needed to obtain fully relaxed structures with realistic chemical, physical and mechanical properties, these local relaxations do not drastically modify the structure and morphology of the resulting models, which remain very close to those initially obtained.

Molecular dynamics. To simulate the GDOS of our models, we performed MD using the semi-*ab initio* force field COMPASS (ref. 57), as provided with the Forcite module in Materials Studio 6.1 (ref. 58; designed to provide reliable vibrational properties in hydrocarbons). The sp^2/sp^3 character was attributed on the basis of a distance cutoff and number of neighbours criterion, but the latter does not allow one to determine the aromaticity. The GDOS is obtained using nMoldyn 4.0 (ref. 59) as the Fourier transform of the velocity autocorrelation function. Details are available in the Supplementary Information.

We computed the elastic properties of the molecular models of kerogens using non-equilibrium MD simulations with LAMMPS (ref. 56). The temperature is fixed using a Nosé-Hoover thermostat and the atomic interactions are computed using the force field OREBO (ref. 40). Because the experimental diffraction measurements were performed at ambient pressure and temperature, the residual stress is relaxed using a MD run (50 ps) in the NPT ensemble at 300 K and 1 bar. The bulk modulus K is computed from the slope of the pressure P as a function of the volume V for volume variations smaller than 1%. The shear modulus G and the Young modulus E are computed from the low-strain ($\epsilon < 0.05$) region of the stress-strain evolution during, respectively, a NVT simulation with shear xy deformation, and a NPT simulation with elongation in the *x* direction and a null pressure fixed in the y and z directions to allow contraction. The strain rate was 109 s⁻¹, as it was shown that, for amorphous carbons with REBO-based potentials, such a strain rate is low enough⁵¹. Fracture is studied by performing a NVT simulation with an elongation in the x direction while the y and z directions are kept fixed ('confined tensile test'). These simulations are performed using several configurations and box sizes to obtain averaged elastic constants. Details are available in the Supplementary Information.

Code availability. The different structural models of mature and immature kerogens are available on request by email.

References

- Neuefeind, J., Feygenson, M., Carruth, J., Hoffmann, R. & Chipley, K. K. The nanoscale ordered MAterials diffractometer NOMAD at the Spallation Neutron Source SNS. *Nucl. Instrum. Methods* 287, 68–75 (2012).
- 55. NIST, Neutron scattering lengths and cross sections; http://www.ncnr.nist.gov/resources/n-lengths
- Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. J. Comp. Phys. 117, 1–19 (1995).
- Sun, H. Compass: an *ab initio* force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *J. Phys. Chem. B* 102, 7338–7364 (1998).
- 58. Materials Studio Modelling Environment Version 6.1 (Accelrys, 2013).
- Hinsen, K., Pellegrini, E., Stachura, S. & Kneller, G. R. nMoldyn 3: using task farming for a parallel spectroscopy-oriented analysis of molecular dynamics simulations. *J. Comp. Chem.* 33, 2043–2048 (2012).

(1)

mature materials

Realistic molecular model of kerogen's nanostructure

Colin Bousige,^{1,2} Camélia Matei Ghimbeu,³ Cathie Vix-Guterl,³ Andrew E. Pomerantz,⁴

Assiya Suleimenova,⁴ Gavin Vaughan,⁵ Gaston Garbarino,⁵ Mikhail Feygenson,⁶ Christoph

Wildgruber,⁷ Franz-Josef Ulm,¹ Roland J.-M. Pellenq,^{1,2,8} and Benoît Coasne^{1,2,*}

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

²<MSE>², UMI 3466 CNRS-MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

³Institut de Science des Matériaux de Mulhouse (IS2M), UMR 7360 CNRS - UHA,

15 rue Jean Starcky, BP 2488, 68057 Mulhouse cedex, France

⁴Schlumberger-Doll Research, 1 Hampshire St., Cambridge, MA, 02139, USA

⁵European Synchrotron Radiation Facility,

BP-220, F-38043, Grenoble Cedex 9, France

⁶Chemical and Engineering Materials Division,

Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA ⁷Instrument and Source Division (NScD),

Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA ⁸CINaM, CNRS/Aix Marseille Université,

Campus de Luminy, 13288 Marseille cedex 09, France

I. SAMPLES

All kerogen samples considered are type II kerogens, which contain large hydrogen amounts (when immature) and thus generate oil and gas in large volumes [1]. The maturity of the various samples was assessed through the usual experimental indicators gathered in Tab. SI1. The type II nature of MarK was determined from the study of immature areas in the same Marcellus formation, and is consistent with other formations such as Barnett shale [2]. The fact that MarK appears in the van Krevelen diagram in the type IV region (Fig. 1a) is likely to be due to a too high oxygen content measured in the elemental composition. In turn, this is believed to arise from the fact that kerogen samples were contaminated with ralstonite, an oxygen-containing mineral that can precipitate during the demineralization process and is detected during the pyrolysis measurement

^{*}Electronic address: coasne@mit.edu

of oxygen used here. Formation of ralstonite during demineralization is common and is described for example by Pernia *et al.* [3]. Assigning quantitative error bars is challenging because the abundance of ralstonite is unknown and ralstonite is difficult to remove.

Tab. SI1 also shows the sulfur content for each kerogen sample. Considering the molar mass of S (~ 32 g/mol), the S wt% content in Tab. SI1 corresponds in fact to a molar fraction equivalent to that of O, *i.e.* x = 0.03 for MEK and EFK and 0.14 for MarK (note that PY02 does not contain sulfur). In the case of MarK, XANES data show that about 90% of these S atoms correspond to sulfur from pyrite, which cannot be removed despite the demineralization undergone by the experimental samples (see section II). This is supported by the diffraction measurements reported in Fig. SI1 showing that MarK and EFK contain pyrite. Considering that the uncertainty over the sulfur content is high and that most sulfur is assumed to belong to pyrite, we did not account for sulfur in our kerogen models. Finally, Tab. SI1 also shows the nitrogen content in Tab. SI1 corresponds to a molar fraction below 1% in all samples and we have therefore omitted the nitrogen atoms in our reconstruction strategy.



Figure SI1: Diffraction patterns of the four samples after PDFgetX2 corrections. The pyrite signal is simulated for different crystallite sizes and subtracted from MEK and EFK experimental data. The resulting data are then fitted using 8 Gaussian functions (dashed lines).

Table SI1: Experimental characterization of the samples. All samples but PY02 contain pyrite (FeS₂) that cannot be removed. All samples but PY02 were subjected to sample demineralization prior to characterization. Data for PY02 are incomplete due to its very large carbon atom content and very low proton content, which prevents performing dipolar dephasing NMR experiments. The very low amount of oxygenated functional groups (CO and CO₂) for PY02, which is comparable with that of graphitic carbons, was confirmed by means of temperature programmed desorption; While these experiments demonstrated the presence of chemisorbed and physisorbed water, only small amounts of CO and CO₂ groups and no amounts of sulfur or nitrogen-based groups were observed to desorb upon sample heating.

		MEK	EFK	MarK	PY02
Elemental analysis	Carbon (wt%)	64.5	54.88	36.27	91.20
	Hydrogen (wt%)	7.84	5.5	1.39	0.29
	Oxygen (wt%)	7.19	7.36	5.25	0.30
	Nitrogen (wt%)	1.3	0.95	0.73	~ 0
	Sulfur (wt%)	13.3	11.9	26.3	~ 0
	$\rm H/C$	1.45	1.19	0.46	0.09
	O/C	0.08	0.10	0.11	0.01
Bulk chemistry	R_0 (%)	0.55	0.65	2.20	
	TOC (wt%) Kerogen/Bitumen	15.4	7.3	5.1	
	yield (wt%)	7.3	1.3	0	
Rock Eval	S1 (mg/g)	5.21	2.02	0.27	
	S2 (mg/g)	111.34	45.52	0.14	
	S3 (mg/g)	1.95	0.99	0.86	
	T_{max} (°C)	410	419		
¹³ C-NMR	$sp^2/(sp^2+sp^3)$ ratio	0.21	0.28	0.95	0.84
	aromatic (%)	20	28	90	84
	aliphatic (%)	79	72	5	8
	carbons/cluster	19	20	54	
	aromatic cluster size (nm)	0.7	0.7	1.2	—
	carboxyl + carbonyl + amide (FaC) (%) phenoxy/phenolic +	1	0	5	
	alkyl-substituted aromatic (%)	10	13	23	11
	bridgehead aromatic $(\%)$	8	12	60	
	nonprotonated aromatic $(\%)$	17	25	83	—
	protonated aromatic $(\%)$	2	3	7	
	alcohol/ether $(\%)$	1	1	1	1.5
	methoxy (%)	2	1	0	0
	mole fraction of bridgehead C	0.39	0.42	0.67	_
	% of aromatic carbons with attachments average aliphatic carbon	48	49	25	
	chain length (nbr of C)	9.1	6	0.25	

II. SAMPLE DEMINERALIZATION

The process of kerogen isolation from shale involved removal of soluble organic matter (bitumen), inorganic minerals, and water. First, 20 g aliquots of shale were crushed in a Shatterbox to achieve a particle size of less than 100 microns. The crushed samples were washed with a mixture of dichloromethane:methanol (90:10 v/v) in a Soxhlet extraction apparatus until no additional colored material was extracted from the shale. Bitumen-free shale samples were dried in air at room temperature. Inorganic mineral matter was then removed from the shale by an acid demineralization process [4]. Hydrochloric acid was used to dissolve carbonates and sulfates, and hydrofluoric acid to remove silicates. Boric acid was used to dissolve any fluorides that formed during the reaction of hydrofluoric acid with the minerals or residual ions. The acid treatment was repeated several times to dissolve most of the minerals. Pyrite is not removed from the samples during this process.

The samples were washed with deionized water after each treatment. To remove residual water, isolated kerogen samples were dried using a critical point dryer – a novel method, which was proposed to preserve kerogen microstructure. The detailed description of the method is given elsewhere [5]. As opposed to conventional drying of kerogen in an oven, water is vaporized without exerting capillary stress on pore walls in critical point drying. During critical point drying, liquid is converted to a supercritical fluid which is then converted to gas without creating surface tension. As a result, an isolated kerogen sample has a representative microstructure [5]. The critical point dryer uses liquid carbon dioxide as a displacement fluid (critical pressure of 73.9 MPa and critical temperature of 304.25 K). Since CO₂ is not miscible with water, an intermediate fluid was required, which could mix both with water and CO_2 . Anhydrous ethanol (99.5%) was used as an intermediate fluid to remove water from the kerogen samples. Dehydration of the samples was done in steps by gradually increasing the ethanol concentration from 30% to 99.5%. The critical point drying of the isolated kerogen samples was performed at the Imaging and Analysis facility of the Center for Nanoscale Systems at Harvard University using an AutoSamdri 815, Series A from Tousimis. While CO_2 can swell the kerogen upon critical drying, we assume it will relax to its un-swollen state at the end of the drying process when CO_2 is released. Considering the sample treatment and the conditions used to analyze the samples afterwards, there should be little CO_2 left in the kerogen when the structural measurements are performed.

III. MODELS

5

The atomic ratios used in the reconstructions, which were determined by elementary composition analysis, are gathered in Tab. SI1. The reconstructed structures have the following dimensions: $50 \times 50 \times 50$ Å³. The experimental density of the kerogen samples cannot be measured accurately for two reasons. First, the isolated samples contain significant quantities of pyrite (in some cases > 20%), a dense mineral that is not removed by the demineralization procedure. Due to the high density of pyrite (around 5 g/cm^3), this contamination results in a large error in density measurement. Pyrite can be removed by chromous chloride treatment, but that treatment cannot be combined with the critical point drying procedure needed to preserve the kerogen microstructure. Second, the density relevant to this work does not correspond to any of the densities that are measured with traditional instrumentation. The most common measured densities for shale samples are the envelope density (in which case the volume includes the volume of the solids and the volume of all intragranular pores) and the matrix density (in which case the volume includes the solid only). Our manuscript describes a model of what is often referred to as the "solid" kerogen, where the maximum pore size is 1.5 nm, below the limit observable by scanning electron microscopy. In addition to the small pores considered here, kerogen also contains larger pores (up to hundreds of nm) that are routinely observed by scanning electron microscopy but are beyond the scope of our molecular model. Traditional matrix density measurement would not include the small pores we are modeling while traditional envelope density measurements would additionally include the large pores we are not modeling, such that neither traditional density measurement is relevant for our model. Therefore, we believe that using a closure scheme (mechanical properties) is the most accurate method to estimate the relevant density. To overcome this difficulty, we generated for each sample (fixed chemical composition and G(r)) a database of molecular reconstructions with densities ranging from 0.8 to to 2.2 g/cm³, with the aim of finding a posteriori the samples that best fit a set of measurements amenable to both experiments and simulations, namely the generalized density of states, pore size distribution, and elastic properties.

As discussed in the article, the principal features of the measured (and simulated) pair distribution functions are the following: the mature samples (PY02 and MarK), although amorphous, show an order extending over ~ 20 Å (distance at which peaks are still discernible in the G(r)). Combined with a first neighbor peak located in 1.41 Å – typical of sp^2 carbons – this is characteristic of the presence of relatively large aromatic clusters (or "graphene-like" sheets) in these mature samples, as shown in Fig. SI2. In immature samples (EFK and MEK), the first neighbor peak is



Figure SI2: Visualization of the aromatic clusters in the kerogen samples ($d = 1.2 \text{ g/cm}^3$): only carbon and oxygen atoms belonging to 5- or 6-membered rings are shown, and each aromatic cluster is represented with a different color. Note that all aromatic clusters are connected by atoms which form a connecting backbone (not drawn for the sake of clarity). For the sake of clarity, we show here the structures obtained for kerogen models of a size 25 Å (but similar cluster structures are present for larger molecular models).

located in 1.54 Å, which is typical of sp^3 carbons, and the G(r) only contains broad peaks that quickly smear out, all order being lost after 6 Å.

Chemical characterization of the models was performed based on an algorithm looking for the number and type of neighbor atoms. The results are shown in Fig. SI3 for each molecular model of a density 1.2 g/cm³. Both chemical functions centered on oxygen atoms and on carbon atoms are considered as they are sufficient to finely characterize the experimental and simulated samples. The C-centered functions can be compared with ¹³C-NMR experiments (green data in Fig. SI3). The results show that the $sp^2/(sp^2 + sp^3)$ ratios of the models are in very good agreement with ¹³C-NMR experimental data. Moreover, in fair agreement with experimental data, the molecular models show that the samples exhibit very little or no carboxylic functions. Finally, the proportion of R–O–R functions found in our molecular models is consistent with the experimental data.

A few isolated molecules (0 ~ 10 at.%) trapped in the pores of kerogen were also found to coexist with the large C-backbone. In most cases, these molecules form small realistic molecules (including H₂, H₂O, acetylene, methane, ethane, methanol, dimethyl ether, etc.), which can be



Figure SI3: Chemical characterization of the four kerogen samples considered in this work. The color code of the article is used: PY02 in black, MarK in blue, EFK in red and MEK in yellow. The green data correspond to ¹³C-NMR experiments. "X-centered functions" designates the proportion of atoms of type X involved in a given function, with X = O or C. Error bars are estimated from the statistics over several molecular configurations.

seen as indicative of the gas/oil that would remain trapped in kerogen (see Tab. SI3 for a list of these molecules for each kerogen structure). Although it should be emphasized that our procedure does not physically capture the complex chemical reactions at the heart of molecule generation upon kerogen evolution, gaseous products including CO_2 , CO, H_2O , H_2 , CH_4 and other short aliphatic hydrocarbons are known to be produced upon thermal maturation [6]. Moreover, as will be discussed later, some experimental samples exhibit molecules that remain trapped in the closed porosity of kerogen. On the other hand, a few of these molecules generated in our molecular models (3.6 at.% at most) are defective with unsaturated bonds (therefore somewhat similar to the cumulene chains discussed above). With longer simulation times and larger samples, such molecules should diffuse to recombine by reacting with the kerogen structure and/or with other molecules to form more stable molecules. Nevertheless, as shown in Table SI3, such defective molecules represent a very small number with only one or two occurrences for most isolated molecules. In any case, we checked that these isolated molecules do not affect the results reported in this work since,



Figure SI4: C–C–C bond angle distribution in the four molecular models of kerogen. The color code of the article is used: PY02 in black, MarK in blue, EFK in red and MEK in yellow. The vertical dotted lines indicate angles of 110° and 118° which correspond to sp^3 and sp^2 carbons, respectively.

Table SI2: Average energy per atom U_{at} (eV/at) for the different samples considered in this work (the number in parentheses are the same energies but in kJ/mol). These data differ from the standard heats of formation reported in Fig. SI10; the former correspond to the potential energy per atom of the structures while the latter are the heat of formation defined with respect to the standard state as defined from the IUPAC.

Sample	U_{at} (eV/at)
EKF	-4.53(-437)
MEK	-4.33(-418)
MarK	-5.46(-527)
PY02	-6.46(-623)

owing to their limited occurrence, they do not contribute significantly to the vibrational, mechanical, adsorption, and fracture properties (see also the G(r) functions calculated with and without these molecules in Fig. SI6). As pointed out earlier, the defects above occur because of sampling limitations in the reconstruction technique for large systems – the reconstruction procedure runs for several months for systems with a size of 5 nm, a compromise between computation time and quality of the reconstruction has to be made.



Figure SI5: Visualization of some defects in a kerogen sample (in this case, MarK $d = 1.2 \text{ g/cm}^3$): (1) 3membered ring, (2) unsaturated free molecule, and (3) unsaturated cumulene chain in the carbon backbone.

IV. MD-HRMC

The MD-HRMC method has been described in details elsewhere [7]. During the HRMC block at temperature T_{χ} , the atoms are displaced randomly and the MC move is accepted with the probability $P^{acc}(T_{\chi})$:

For C atoms:

$$P_C^{acc}(T_{\chi}) = \min\left[1, exp\left(-\frac{1}{T_{\chi}}\left(\Delta\chi^2 + \frac{\Delta U}{\omega}\right)\right)\right]$$
(SI1)

For heteroatoms (O and H):

$$P_{het}^{acc}(T_{\chi}) = \min\left[1, exp\left(-\frac{\Delta U}{\omega T_{\chi}}\right)\right]$$

where ω is a weighing parameter, $T_{\chi} = k_B/\omega T$ the effective temperature (*T* is the thermodynamical temperature), *U* the total energy, and χ^2 is an error function defined as [7]:

$$\chi^{2} = \frac{\sum_{i} (G_{simul}(r_{i}) - G_{exp}(r_{i}))^{2}}{\sum_{i} (G_{exp}(r_{i}))^{2}}$$
(SI2)

Following the discussion in Ref. [7] for systems containing a non-negligible number of non-carbon atoms, ω is defined as $\omega = \frac{N_{total}^2}{N_{carbon}} \times 7 \text{ eV}$, where N_{carbon} and N_{total} are the number of carbons in the structure and the total number of atoms, respectively. Such a definition ensures that $\Delta \chi^2$ and ΔU are given an equivalent weight despite the fact that hydrogen atoms contribute to the total energy and not to χ^2 . In order to prevent the formation of H₂ molecules, the maximum displacements for H are allowed to be larger than for C and O (typically 200 times larger). Table SI3: Isolated molecules coexisting with the C-backbone for each kerogen structure considered in this work. While molecules in blue are physical molecules, molecules in red are under-saturated or oversaturated – the latter representing less than $\sim 4\%$ of the atoms at most. For each molecule type, the number in parenthesis represents the atomic percentage of atoms involved in these molecules (molecules with atomic proportion less than 0.1% were omitted as they correspond to 1 or 2 molecules at most). The total percentage of the atoms involved in under/over -saturated molecules is indicated for each molecular model in the first row next to the sample density.

EFK						
$0.8 \text{ g/cm}^3 (3.07\%)$	$1.0 \text{ g/cm}^3 (3.26\%)$	$1.2 \text{ g/cm}^3 (1.43\%)$	$1.4 \text{ g/cm}^3 (1.08\%)$			
H_2 (0.33), CH_4O (0.05), CH_4O_2	H_2 (16.24), H_2O (0.04), CO (0.01),	${\rm H}_2 \ (3.28), {\rm H}_2 {\rm O} \ (0.07), {\rm CH}_4 \ (0.16),$	H_2 (16.10), H_2O (0.03), H_2O_2			
$(0.06), C_2H_4O (0.63), C_2H_6O$	$\rm CH_2O~(0.11),~\rm CH_4O_2~(0.05),~\rm C_2H_2$	$\rm CH_4O~(0.10),~C_2H_4~(0.07),~C_2H_4O$	$(0.06), CH_2O (0.04), CH_4O (0.08),$			
$(0.66), C_3H_4O_2 (0.07), C_4H_6O_2$	$(0.03), C_2H_4 (0.04), C_2H_4O (0.74),$	$(0.31), C_2H_6O (0.79), CH_3 (0.42),$	C_2H_4 (0.03), C_2H_4O (0.20),			
$(0.10), C_4H_8O_3 (0.12), C_5H_{10}O$	$C_2H_4O_2$ (0.11), C_2H_6O (0.18),	C_2H_5O (0.44), C_3H_7O (0.12)	$C_2H_4O_2$ (0.04), C_2H_6O (0.08),			
(0.13), CH ₃ (0.10), CH ₃ O (0.16),	C ₂ H ₈ O (0.07), H ₃ (0.43), CHO		C_2H_8O (0.05), $C_3H_5O_2$ (0.05),			
C_2H_3O (0.20), C_2H_5O (1.58),	$(0.28), CH_3 (0.32), CH_3O (0.30),$		$C_3H_8O_2$ (0.06), H_3 (0.31)			
C_4H_7O (0.10), $C_4H_7O_2$ (0.32),	C_2H_2O (0.23), C_2H_3O (0.35),					
C_4H_8O (0.11), $C_7H_{11}O_3$ (0.17)	C_2H_5O (0.42), $C_4H_4O_2$ (0.13)					
	M	EK				
$0.8 \text{ g/cm}^3 (1.35\%)$	$1.0 \text{ g/cm}^3 (1.18\%)$	$1.2 \text{ g/cm}^3 (0.92\%)$	$1.4 \text{ g/cm}^3 (1.95\%)$			
H_2 (0.80), H_2O (0.02), CH_4 (0.04),	H_2 (22.76), H_2O (0.06), H_2O_2	H_2 (6.46), H_2O (0.26), H_2O_2	H_2 (6.22), H_2O_2 (0.15), CH_4O			
CH_4O (0.39), CH_4O_2 (0.06),	$(0.08), CH_2O (0.03), CH_4 (0.07),$	$(0.07), CH_2O (0.07), CH_4 (0.11),$	$(0.45), C_2H_4O (0.79), C_2H_6$			
C_2H_4O (0.17), C_2H_6O (1.11), CH_3	$\rm CH_4O~(0.08),~C_2H_4~(0.04),~C_2H_4O$	CH_4O (0.16), CH_4O_2 (0.04),	$(0.30), C_2H_6O (1.01), C_3H_8O$			
$(0.26), CH_3O (0.16), C_2H_5O$	$(0.05), C_2H_4O_2 (0.05), C_2H_4O_3$	$C_2H_4O_2$ (0.04), C_2H_6 (0.04),	$(0.45), C_3H_8O_2 (0.97), H_3 (0.22),$			
$(0.66), C_5H_7 (0.10)$	$(0.06), C_2H_6O (0.06), C_3H_6O$	C_2H_6O (0.10), CH_3 (0.57), CH_3O	CH_3 (0.45), C_2H_5O (0.90),			
	$(0.07), C_4H_6O_2 (0.08), H_3 (0.57),$	$(0.11), C_2H_5O (0.13)$	$C_2H_5O_2$ (0.34)			
	$C_2H_3O(0.12)$					
MarK						
$0.8 \text{ g/cm}^3 (2.11\%)$	$1.0 \text{ g/cm}^3 (0.60\%)$	$1.2 \text{ g/cm}^3 (1.22\%)$	$1.4 \text{ g/cm}^3 (0.60\%)$			
H_2O (0.33), H_2O_2 (0.16), CH_4O_2	H_2O (0.43), H_2O_2 (0.13), CH_4	H_2O (0.58), H_2O_2 (0.17), CH_4O	H_2O (0.11), H_2O_2 (0.02), CH_4			
$(0.11), C_2H_2$ (2.83), C_3H_4 (0.11),	$(0.13), C_2H_2 (0.76), C_2H_4 (0.04),$	$(0.03), CH_4O_2 (0.04), C_2H_2$	$(0.18), C_2H_2 (0.99), C_2H_4 (0.03),$			
CHO (0.61), CHO ₂ (0.13), C ₂ O	$C_2H_4O_2$ (0.05), C_3H_4 (0.13)	$(0.67), C_2H_4 (0.03), C_2H_4O_2$	CHO (0.22)			
$(0.47), C_2O_2$ (0.13), C_3HO (0.28)		$(0.04), C_3H_4 (0.04), C_4H_6O_2$				
		(0.06), CHO (0.24)				
PY02						
$0.8 \text{ g/cm}^3 (0.03\%)$	$1.0 \text{ g/cm}^3 (0.00\%)$	$1.2 \text{ g/cm}^3 (0.00\%)$	$1.4 \text{ g/cm}^3 (0.00\%)$			
$C_{2}H_{2}$ (0.03)						

 $_{2}H_{2}$ (0.03)

Because of the limited Q range of the experimental measurements of S(Q), the target function has to be considered with a binning (spacing between two experimental points) that is larger or equal to the experimental resolution $\partial r \sim 2\pi/(Q_{max} - Q_{min})$. For computational reasons, the number of bins has to be an integer defined as $N_{bin} = 10.2/\partial r$. We used a resolution $\partial r = 0.3$ Å for MEK and EFK, and $\partial r = 0.2$ Å for PY02 and MarK.

Sampling of intermediate and long range orders can be improved by using $r^2 (G(r) - 1)$ as a target function instead of G(r) in HRMC techniques [8]. However, in the present work, we decided to keep G(r) as the target function since the short-range order plays a predominant role in the chemistry, physics and mechanics of kerogen.

V. X-RAY AND NEUTRON DIFFRACTION

A.
$$G(r)$$

For x-ray data, the integrated data were normalized and corrected for the various techniquespecific factors using the software PDFgetX2 [9]. Then, the obtained structure factor S(Q) is linked to the atomic pair distribution function G(r) via a Fourier transform:

$$G(r) = 1 + \frac{4\pi}{\rho_0 (2\pi)^3} \int_0^\infty Q^2 \left(S(Q) - 1 \right) \operatorname{sinc}(Qr) dQ,$$
(SI3)

where ρ_0 stands for the atomic density of the sample. In practice, the experimental Q-range is not infinite, which results in non physical oscillations in the G(r) function after Fourier transform. The actual transformation from S(Q) to G(r) is thus performed as:

$$G(r) = 1 + \frac{4\pi}{\rho_0(2\pi)^3} \int_{Q_{min}}^{Q_{max}} Q^2 \left(S(Q) - 1\right) \operatorname{sinc}(Qr) L(\Delta, Q) dQ$$
(SI4)

where $\Delta = Q_{max} - Q_{min}$, and $L(\Delta, Q) = \operatorname{sinc}(\pi Q/\Delta)$ is a Lorch function that allows eliminating the oscillations in the Fourier transform due to the limited experimental Q range that is used [10] (the limitations, indicated in the Methods section of the article, arise mainly from the fact that the high Q scattering intensity is very low due to the sample form factor, and obtaining a good signal-to-noise ratio at high Q is often unfeasible).

All demineralized kerogens show, to various extents, the presence of pyrite impurities – but no other minerals. For the MarK sample, the pattern is largely dominated by the pyrite signal, and no structural information about the MarK kerogen can be extracted from these measurements. In the case of EFK and MEK, however, the intensity of the pyrite signal is comparable to that of the kerogen, the first being characterized by sharp and intense peaks and the latter by broad and weak peaks. The pyrite powder diffraction pattern can be simulated from its tabulated structural information. The simulated pattern has also to take into account the broadening of the peaks due to instrumental resolution as well as finite size of the crystallites. In practice, this is done by convoluting the simulated pyrite signal with a Lorentzian function of width given by the Scherrer equation [11]:

$$FWHM(\theta) = \frac{K\lambda}{\tau\cos(\theta)}, \text{ with } Q = \frac{4\pi}{\lambda}\sin(\theta),$$
 (SI5)

where λ is the x-ray wavelength, 2θ the scattering angle, τ the mean particle size and K is Scherrer shape factor whose typical value is 0.89. The best subtractions are performed with pyrite crystallites sizes of $\tau = 19$ nm for EFK and $\tau = 13$ nm for MEK. After subtraction of the simulated pyrite signal, the kerogen signal remains noisy and some reminiscence of pyrite peaks are still present. Since no sharp peak is expected in the diffraction pattern of kerogen because it is an amorphous system, we fit the subtracted signal with a panel of 8 Gaussians functions in order to obtain a smooth pattern (see dashed lines in Fig. SI1) (a sharp peak from an unknown impurity, present around 1.2 Å⁻¹, is fitted and removed from the diffraction pattern).



Figure SI6: Experimental (lines) and simulated (symbols) G(r) for each sample considered in this work. The color code of Fig. SI1 is used: PY02 in black, MarK in blue, EFK in red and MEK in yellow. The circles are G(r) computed with C–C contributions only with the closed and open symbols for the models with and without the isolated molecules in the porosity. The open squares correspond to G(r) computed with all contributions (C–C, C–O, and O–O). Open triangles correspond to the G(r) obtained after switching off the RMC procedure and a 10 ps MD relaxation at 300 K in the NVT ensemble.

B. Bulk modulus

The diffraction under pressure measurements were performed on the beam line ID27 at the European Synchrotron Research Facility (ESRF) in Grenoble, France. The sample PY02 was placed in a diamond anvil cell together with a ruby in order to measure the pressure by fluorescence (pressure was increased up to 5.1 GPa). The inset in Fig. 4 presents the pressure evolution of the

position Q of the first diffraction peak for PY02, or rather $(Q_0/Q)^3 \propto V/V_0$ where the subscript 0 stands for the room pressure value. The pressure evolution of the relative change in volume allows us to fit the bulk modulus K of PY02 using a first-order Birch-Murnaghan equation of state:

$$\frac{V}{V_0} \propto \left(1 + P \frac{K'}{K}\right)^{1/K'},\tag{SI6}$$

which yields a bulk modulus of $K_{PY02} = 13.7 \pm 2$ GPa when fixing K' = 4.

VI. INELASTIC NEUTRON SCATTERING

In a neutron experiment, the quantity measured is proportional to the so-called differential scattering cross section $\left(\frac{\partial^2 \sigma}{\partial E \partial \Omega}\right)$, giving the number of incident neutrons with wave vector \mathbf{k}_i scattered by the sample in a solid angle $\partial \Omega$ around the solid angle Ω and with an energy range ∂E around E (*i.e.*, having a scattered wave vector \mathbf{k}_f or a wave vector transfer $\mathbf{Q} = \mathbf{k}_f - \mathbf{k}_i$). Due to the peculiar neutron-nucleus interaction, this quantity can be split into a coherent and an incoherent cross section:

$$\left(\frac{\partial^2 \sigma}{\partial E \partial \boldsymbol{\Omega}}\right) \propto \frac{k_f}{4\pi k_i} \left[S_{coh}(\boldsymbol{Q}, \omega) + S_{inc}(\boldsymbol{Q}, \omega)\right]$$
(SI7)

Because the measured samples are all under the form of powder, the incoherent approximation is valid and the same formalism can be used to describe the coherent and incoherent "one-phonon" dynamical structure factors $S(\mathbf{Q}, \omega)$ [12]:

$$S(\boldsymbol{Q},\omega) \propto e^{-2W(Q)} \frac{n^{\pm}(\omega,T)}{\omega} G(\boldsymbol{Q},\omega)$$
 (SI8)

where the exponent \pm refers to neutron energy gain (+, anti-Stokes) or neutron energy loss (-, Stokes) scattering processes. On the spectrometer VISION at the Spallation Neutron Source (Oak Ridge, TN), experiments are performed at the temperature of 5 K and only Stokes processes are considered. Q is the norm of the scattering vector \mathbf{Q} , $n^{\pm}(\omega, T)$ the Bose population factor, $2W(Q) = Q^2 \langle u^2 \rangle$ the Debye-Waller factor, $\langle u^2 \rangle$ the total mean squared displacement. The generalized density of states $G(\mathbf{Q}, \omega)$ is expressed as:

$$G(\boldsymbol{Q},\omega) = \sum_{(j,\boldsymbol{q})} \delta\left[\omega - \omega_j(\boldsymbol{q})\right] \sum_{\alpha} \frac{\sigma_{coh/inc}^{\alpha}}{m_{\alpha}} |\boldsymbol{Q} \cdot \boldsymbol{e}_{\alpha}(j,\boldsymbol{q})|^2$$
(SI9)

where (j, q) denotes the phonon modes of frequency $\omega_j(q)$ and polarization vector $\boldsymbol{e}(j, q) = \sum_{\alpha} \boldsymbol{e}_{\alpha}(j, q)$, while $\sigma_{coh/inc}^{\alpha}$ and m_{α} stand, respectively, for the coherent/incoherent cross section and the mass of the atom α in the unit cell. In the case of a powder, where the

intensities are naturally integrated over all orientations of the Q vectors, one can write $\sum_{\alpha} |\mathbf{Q} \cdot \mathbf{e}_{\alpha}(j, \mathbf{q})|^2 = Q^2 \sum_{\alpha} |\mathbf{e}_{\alpha}(j, \mathbf{q})|^2$ and $G(\mathbf{Q}, \omega) = Q^2 \times G(\omega)$.

The four considered samples contain a majority of carbon, hydrogen and oxygen with natural isotopic ratio, as well as traces of nitrogen and sulfur. The ratio σ/m is largely dominant in the case of the incoherent scattering from hydrogen [13] so that it prevails in the observed intensity.

The data collected on VISION are corrected for the sample holder scattering (vanadium cylindrical can), and normalized to monitor neutron count and sample mass. Due to the peculiar geometry of VISION, the measurements are performed with a constant angle, and therefore with a Q vector depending on the energy transfer $\hbar\omega$ as:

$$Q^{2} = \frac{2m_{n}}{\hbar^{2}} \left(E_{f} + \hbar\omega \right) + \frac{2m_{n}}{\hbar^{2}} E_{f} - \frac{4m_{n}}{\hbar^{2}} \cos\left(\theta\right) \sqrt{E_{f}(E_{f} + \hbar\omega)}$$
(SI10)

where m_n is the neutron mass, E_f the energy of the scattered neutron and θ the scattering angle. On VISION, there are two banks of detectors, with $\theta = 135^{\circ}$ and $\theta = 45^{\circ}$ (back scattering and transmission, respectively).

Fig. 3.a reports the scattered intensity $S(Q, \omega)$ on the transmission bank for the four samples – the two banks give the exact same signal for energy transfers above 600 cm⁻¹ but the transmission bank provides better statistics. In addition to these normalizations, the data are corrected by a factor $Q^2 n(\omega)/\omega$ to obtain the GDOS $G(\omega)$ [Eqs. (SI8) to (SI10)].

VII. MOLECULAR DYNAMICS – SIMULATION OF THE GDOS

In order to simulate the GDOS, we performed MD simulations using the semi-*ab initio* forcefield COMPASS [14] as provided with the Forcite module provided in Materials Studio 6.1 [15]. This forcefield was selected at is was designed to provide reliable vibrational properties in hydrocarbons and that no reactivity is expected from the structure at this point of the study. While OREBO captures most of the vibrational features in carbon-based compounds, we decided to use COMPASS as it provides an improved description of the vibrational modes [14]. Fig. SI7 compares the experimental GDOS for the MarK sample with the simulated spectra obtained using COMPASS [14] and OREBO [16] (for further comparison, we also show the data obtained with ReaxFF forcefield [17] although it was not used in the present work). Fig. SI7 shows that both reactive force fields (OREBO and ReaxFF) provide a reasonable description of the vibrational spectrum in kerogen but fail to reproduce quantitatively the vibrational energies (especially for the C–H stretching mode at 3000 cm^{-1} but also in the lower energy range). In contrast, COMPASS is found to be in good agree-

15

ment with the experimental data, both in the high and low energy ranges. Note that discrepancies between the experimental and simulated peak amplitudes mainly arise from the limited statistics attainable with nm-scale molecular models.



Figure SI7: Effect of the forcefield for the simulation of the GDOS: the top panel shows the experimental GDOS measured for the MarK sample, while the bottom panel compares the simulated GDOS – for the same MarK molecular model – with COMPASS (blue), ReaxFF (black) and OREBO (red).

The use of COMPASS [14] requires the determination of the hybridization state of each atom. The hybridization of each carbon atom was assigned based on a criterion relying both on a distance cutoff and the number of neighbors. The distance criterion first attributes the bond type (single, double, triple), while the number of neighbors determines the hybridization state: sp^1 for linear carbons (the small number of sp^1 carbons were found to correspond to cumulenes, *i.e.* consecutive double bonds), sp^2 for trigonal, and sp^3 for tetrahedral. All these different carbon types were taken into account in our vibrational analysis (GDOS) through the use of the semi-*ab initio* force field COMPASS [14]. As the GDOS is largely dominated by modes involving hydrogen motion, these sp^1 atoms (< 5%) do not affect the GDOS. However, the criterion described above does not allow determining the aromaticity, which explains the redshift of the simulated C–H stretching mode around 3000 cm⁻¹ for PY02 and MarK with respect to the experimental GDOS (Fig. 3.a of the paper). A first relaxation of 100 ps in the NVT ensemble (thermalization) is performed at 300 K, followed by a 20 ps relaxation in the NVE ensemble (production), with a time step of 1 fs and storing the positions and velocities every 4 fs. From the trajectory, the velocity autocorrelation function $C_{\alpha\alpha}(t)$ (VACF) is computed for all atoms α :

$$C_{\alpha\alpha}(t) = \frac{1}{N_{\alpha}} \sum_{i=0}^{N_{\alpha}} \boldsymbol{v}_{\alpha,i}(0) \cdot \boldsymbol{v}_{\alpha,i}(t), \qquad (SI11)$$

where N_{α} is the number of atoms α in the box and $\boldsymbol{v}_{\alpha,i}(t)$ the velocity of the i^{th} atom α at time t. The generalized density of states $G(\omega)$ is then computed from the Fourier transform of the VACF, $\tilde{C}_{\alpha\alpha}(\omega)$, the atomic concentration c_{α} and the atomic neutron scattering cross section σ_{α} :

$$G(\omega) = \sum_{\alpha} \sigma_{\alpha} c_{\alpha} \tilde{C}_{\alpha\alpha}(\omega).$$
 (SI12)

The VACF is a self correlation function and therefore – by definition – corresponds to incoherent neutron scattering. However, in the frame of the incoherent approximation [12], the coherent and incoherent functions can be computed using the sole incoherent formalism (replacing σ_{inc} by σ_{coh}). Given the different atomic cross sections involved [13], the GDOS is however dominated by the hydrogen incoherent signal. The GDOS were obtained from the trajectories using nModlyn 4.0 [18]. The simulated $G(\omega)$ are then directly comparable to the experimental data.

VIII. MOLECULAR DYNAMICS – ELASTIC CONSTANTS

We computed the elastic properties of the kerogen molecular models using non-equilibrium MD simulations with the LAMMPS molecular simulation package [19]. The temperature was fixed using a Nosé-Hoover thermostat and the atomic interactions were computed using the forcefield OREBO [16]. First, the residual stress within the structures is relaxed with a 50 ps dynamics in the NPT ensemble at 300 K and 1 bar. Then, various deformations with a strain rate of 10^9 s^{-1} are applied to the relaxed structures; isostatic compression/dilatation, shear deformation, and uniaxial elongation. The various elastic constants are then fitted from the slope in the low-strain region ($\varepsilon < 0.05$) of the obtained stress-strain curves.

The bulk modulus $K = -V \left(\frac{\partial P}{\partial V}\right)_T$ is computed from the slope of the pressure P as a function of the volume V for volume isotropic variations smaller than 1% during two different simulations; a first simulation in the NVT ensemble in which the volume is increased or decreased with a rate of 10^9 /s and a second simulation in the NPT ensemble where the pressure is increased from 0 to 1 GPa at the same rate. The shear modulus G is computed from the fit in the low-strain region of $\sigma_{xy}/\varepsilon_{xy}$ during a NVT simulation with xy shear deformation. The Young modulus E is computed from the fit in the low-strain region of $\sigma_{xx}/\varepsilon_{xx}$ during a partial NPT simulation with elongation in the x direction and a null pressure fixed in the y and z directions to allow contraction.

Confined tensile tests, in which the x direction is stretched but the y and z directions are kept fixed, allow an indirect determination of the bulk and shear moduli, and a direct determination of the Poisson ratio. The slopes at the origin of the various stresses are indeed given by:

$$\frac{\sigma_{xx}}{\varepsilon_{xx}} = K + \frac{4}{3}G$$

$$\frac{\sigma_{yy}}{\varepsilon_{xx}} = \frac{\sigma_{zz}}{\varepsilon_{xx}} = K - \frac{2}{3}G$$

$$\frac{\sigma_{yy}}{\sigma_{xx}} = \frac{\sigma_{zz}}{\sigma_{xx}} = \frac{\nu}{1 - \nu}$$
(SI13)

The Poisson ratio is also computed from the following relations:

$$\nu = \frac{3K - E}{6K}$$

$$\nu = \frac{3K - 2G}{2(3K + G)}$$
(SI14)

These different ways to compute the elastic constants applied to the different sample configurations provide a statistical distribution (with mean values and error bars plotted in Fig. SI8).



Figure SI8: Poisson ratio ν as a function of the sample density as computed from Eqs. (SI13) and (SI14). The color code of Fig. SI1 is used: PY02 in black, MarK in blue, EFK in red and MEK in yellow.

IX. HEAT CAPACITY AND FORMATION ENTHALPY

The isochoric and isobaric heat capacities, C_v and C_p , are defined as:

$$C_v = \left(\frac{\partial U}{\partial T}\right)_V \qquad \qquad C_p = \left(\frac{\partial H}{\partial T}\right)_P \qquad (SI15)$$

and are linked via:

$$\frac{1}{m}\left(C_p - C_v\right) = c_p - c_v = T\frac{\alpha^2}{\rho\beta} \tag{SI16}$$

where $\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_P$ is the thermal expansion coefficient, ρ the density, and $\beta = -\frac{1}{V} \left(\frac{\partial V}{\partial P}\right)_T$ the compressibility (*i.e.* the reciprocal of the bulk modulus). Experimentally, the isobaric heat capacity c_p is generally more easily measured than the isochoric heat capacity c_v since the measurements are performed at constant temperature and pressure.

In the case of solids, the difference between c_v and c_p is often neglected as the thermal expansion α is usually small. However, considering that the compressibility of solids is also small, the difference between c_v and c_p (Eq. (SI16)) can be of the order of 30 % [20]. In our case, the difference between the isobaric and isochoric heat capacities can be neglected: for coals (close to kerogens), α is generally of the order of $4 \cdot 10^{-5}$ K⁻¹ [21], and we compute a bulk modulus of the order of $1/\beta \simeq 10$ GPa for a density of 1.2 g/cm³ (Fig. 4). At 300 K, this yields an approximate difference between c_p and c_v of $4 \cdot 10^{-3}$ J/K/g, which is negligible compared to c_v . The following discussion will therefore focus on the isochoric heat capacities c_v which are assumed to be equal to c_p .

The heat capacity $C_v = mc_v = \left(\frac{\partial U}{\partial T}\right)_V$ (where *m* is the sample mass) can be computed from the internal energy *U*, which is related to the phonon density of states $G(\omega)$ and the Bose population $n(\omega) = 1/\left[\exp\left(\hbar\omega/k_BT\right) - 1\right]$ as:

$$U(T) = \int_{-\infty}^{\infty} d\omega \ G(\omega) n(\omega) \hbar \omega$$
 (SI17)

For inhomogeneous systems containing N_{α} atoms of type α (of mass m_{α} and velocity v^{α}), the phonon density of states is defined as the sum of the mass-weighted Fourier transform of the partial velocity auto-correlation functions:

$$G(\omega) = \frac{1}{k_B T} \sum_{\{\alpha\}} m_\alpha \sum_{i=1}^{N_\alpha} \int_{-\infty}^{\infty} dt \ \langle \boldsymbol{v}_i^\alpha(t) \cdot \boldsymbol{v}_i^\alpha(0) \rangle e^{i\omega t}$$
(SI18)

With $N = \sum_{\{\alpha\}} N_{\alpha}$ the total number of atoms, $G(\omega)$ is normalized so that:

$$\int_0^\infty d\omega \ G(\omega) = 3N \tag{SI19}$$

Going back to the definition of C_v , we have:

$$C_v(T) = \left(\frac{\partial U}{\partial T}\right)_V = \frac{1}{k_B T^2} \int_0^\infty d\omega \ G(\omega) \left(\hbar\omega\right)^2 \times \frac{e^{\hbar\omega/k_B T}}{\left(e^{\hbar\omega/k_B T} - 1\right)^2}$$
(SI20)

which tends towards $3Nk_B$ when $T \to \infty$.

The DOS used in this work are those obtained using MD simulations for the study of the inelastic neutron scattering (but weighted differently). Fig. SI9.a shows $c_p(T) = c_v(T)$ for the four samples, and Fig. SI9.b shows the evolution of c_p at 300 K as a function of the H/C ratio. The isobaric heat capacities vary linearly with the maturity parameter H/C.



Figure SI9: (a) Isobaric heat capacities c_p of the four samples as a function of the temperature as computed from Eq. (SI20). The horizontal dashed lines indicate the limit $\frac{1}{m}3Nk_B$, with m the mass of the model and N the total number of atoms. – (b) Isobaric heat capacities at 300 K as a function of the H/C atomic ratio. For the four molecular models, the color code of Fig. SI1 is used: PY02 in black diamond, MarK in blue triangle, EFK in red square and MEK in yellow circle. Empty symbols correspond to experimental measurements of alkane chains [22], various coals and kerogens [23, 24], and diamond and graphite [25]. The dashed line is a linear fit $c_p = 0.689 \times H/C + 0.718 \text{ J/k/g}.$



Figure SI10: Standard heat of formation ΔH_f for kerogen as a function of its O/C and H/C atomic ratios. The filled circles correspond to the kerogen samples considered in this work (black = PY02, blue = MarK, yellow = MEK, and red = EFK) while the empty squares are the samples considered in [26]. The empty blue circle corresponds to the data expected for MarK based on its vitrinite reflectance shown in Fig. 1 (this method is thought to provide an accurate estimate for the O/C ratio). For each sample, the number in black is the theoretical value obtained using the Group-Contribution method for our samples and using MOPAC calculations for the other samples. The number in blue corresponds to the value determined from Boie correlation (see text). The background color code with the scale shown on the right corresponds to the values obtained from Boie correlation [27]. This correlation states that the heat of combustion ΔH_c and the standard heat of formation ΔH_f can be determined as ΔH_c (kJ/kg) = $351.7w_C + 1162w_H - 111.0w_O + 104.7w_N + 27w_S - 439.6$ and ΔH_f (kJ/kg) = $\Delta H_c - (1418w_H + 327.8w_C + 92.6w_S - 24.2w_N)$ where w_X are the weight percentage of atom X.

In addition to heat capacity, the heat of formation directly reflects the chemistry and maturity of the kerogen [26]. Heats of formation include a contribution arising from the quantization of the vibrational degrees of freedom, which cannot be estimated using classical simulations. Therefore, in order to estimate the standard heat of formation for the different kerogens, we use the groupcontribution method which accurately describes the thermochemistry of organic compounds [28, 29, 29, 30]. This technique allows determining the thermochemistry of organic compounds through its decomposition into simple elementary groups [29, 29, 30]. This group-contribution method is usually used for small compounds where group assignment can be done easily. Considering the size of our systems (more than 10,000 atoms), this cannot be done manually and a specific code was written in order to decompose the samples into the following groups (CH_x, aromatic C, aromatic C–CH_x, –OH, R–COO–, CH_x=CH_x, CH_x=C=CH_x, H_x \equiv CH_x, etc. [29]. Several versions of the group-contribution method exist [29, 30]; while they lead to different values (within a few percents), the different versions lead to the same trends (effects of maturity, chemistry, etc.).

Fig. SI10 shows the standard heat of formation for our samples as a function of their chemical composition (O/C and H/C ratios). Fig. SI10 also shows the standard heats of formation obtained for 5 different kerogens using semi-empirical ab initio calculations in [26]. Both data series show that immature kerogens (EFK and MEK for this work) possess a negative heat of formation which is a common feature to linear and cyclic alkanes. In contrast, owing to their aromatic clusters, mature kerogens (MarK and PY02 for this work) exhibit a positive standard heat of formation. Upon increasing the O/C ratio, the standard heat of formation decreases as fewer aromatic compounds are present in favor of organic molecules in immature kerogens. Similarly, upon increasing the H/C ratio, samples exhibit fewer sp^2 elements and larger sp^3 groups such as alkyl chains. For each sample, Fig. SI10 also reports the standard heat of formation as obtained from its stoichiometric coefficients (C, H, O, N, and S) using Boie correlation [27]. Such a relationship, which is found to reasonably capture the change in the standard heat of formation upon varying the H/C and O/C ratios, is a very useful tool that allows probing the maturity and chemistry of kerogens.

X. ADSORPTION ISOTHERMS AND GCMC

Following previous works [31, 32], nitrogen adsorption at 77.4 K in the molecular models of mature and immature kerogens were simulated using the Grand Canonical Monte Carlo (GCMC) simulation technique as well as methane adsorption at 450 K. In GCMC simulations, we consider a system at constant volume V (the pore with the adsorbed phase) in equilibrium with an infinite reservoir of molecules imposing its chemical potential μ and temperature T. For given T and μ , the adsorbed amount is given by the ensemble average of the number of adsorbed molecules versus the pressure P of the gas reservoir (the latter is obtained from the chemical potential according to the equation of state for the bulk gas). The adsorption isotherm is simulated by increasing or decreasing the chemical potential of the reservoir; the final configuration of a simulation is the initial state for the next point. Periodic boundary conditions were used in the 3 directions of space. In our simulations, the nitrogen molecules are simply described as a single Lennard-Jones sphere, and we used the same Lennard-Jones parameters as ref. [32, 33]. These parameters are reported in Tab. SI4 for the like-atom interactions. The cross interactions are computed from the Lorentz-Berthelot rules:

$$\sigma_{xy} = \frac{1}{2} \left(\sigma_{xx} + \sigma_{yy} \right) \qquad \varepsilon_{xy} = \sqrt{\varepsilon_{xx} \varepsilon_{yy}} \tag{SI21}$$

Table SI4: Carbon, oxygen, nitrogen, hydrogen and methane Lennard-Jones parameters used in the GCMC and MD simulations of the adsorption isotherms.

	$\mathbf{C}\mathbf{C}$	00	NN	HH	CH_4	
σ_{xx} (Å)	3.4	3.0	3.75	2.42	3.81	
ε_{xx} (K)	28.0	85.6	80.26	15.1	148.1	



Figure SI11: Experimental (lines) and simulated (full symbols) N_2 adsorption isotherms at 77 K for the densities (in g/cm³) that best reproduce the measurements: 2.2 for PY02, 1.0 for MarK, 1.4 for EFK and MEK (the empty symbols correspond to densities of 1.0). The color code of Fig. SI1 is used.

Fig. SI11 compares the simulated and experimental N_2 adsorption isotherms for the different samples considered in this work, while Tab. SI5 shows the corresponding measured and simulated BET surfaces. In the case of the simulated adsorption isotherms, the reported data correspond to the sample density leading to the best agreement with the experimental data (the experimental data have been corrected to account for the presence of remaining minerals). However, experimental adsorption suffers from several drawbacks such as its inability to probe the closed porosity in Table SI5: Comparison of the measured and computed BET surfaces of the four samples. S_{BET} are the specific surfaces obtained by simulating the adsorption isotherms using the GCMC technique. S_{BET}^* are the specific surfaces corrected for the pore connectivity (as probed using MD simulations). The experimental specific surface could not be estimated reliably for PY02 due to too large error bars over the N₂ adsorption isotherm.

	EFK	MEK	MarK	PY02
$S_{BET}(\text{Exp.}) \ (\text{m}^2/\text{g})$	21 ± 10	22 ± 10	178 ± 40	
$S_{BET}(\text{Simul}) \ (\text{m}^2/\text{g})$	195	124	222	280
S_{BET}^* (Simul) (m ² /g)	47	17	177	116

samples. This is clearly seen for the sample PY02 which exhibits closed porosity as evidenced by the presence of confined water (see the generalized density of states in Fig. 3.a). In order to reproduce the very low adsorbed amounts in PY02, an unrealistic sample density of 2.2 g/cm³ has to be used (*i.e.* the density of graphite) to match the simulated and experimental adsorbed amounts. As discussed in the section relative to the mechanical testing of our models, such a high density is incompatible with the mechanical properties of shungite. The experimental bulk modulus of $K \sim 13.7 \pm 2$ GPa for PY02 is indeed incompatible with the simulated value $K \simeq 138$ GPa for a density of 2.2 g/cm³ (even when accounting for the effect of grain – or pore – packing [34]). Moreover, correcting the simulated adsorbed amounts for the large porosity (> 2 nm) not accounted for in our molecular model suggests that the density of PY02 should be even larger. Similarly, densities larger than 1.4 g/cm³ are required to match the low experimental adsorbed amounts for EFK and MEK (whose molecular models also exhibit significant closed porosity). In contrast, the simulated adsorption isotherm for MarK is found to be in reasonable agreement with the experimental data when a physical density of 1.0 g/cm^3 is used. Moreover, the PSD for this sample (Fig. 3.b) shows that the microstructure contains large pores (up to 13.5 Å) connected to each other. The results above suggest that, unless the whole porosity of a sample is connected to the external environment, gas adsorption provides incomplete and non-conclusive characterization of the porosity in kerogen. While this technique still provides valuable information such as the PSD (whose shape and position are not affected by the total pore volume or surface), such a drawback constitutes a severe limitation for evaluating available resources in a given shale formation. This result shows that adsorption experiments must be combined with mechanical measurements to assess the structural and textural properties of kerogen.

In order to take into account the effect of pore connectivity, we have also determined the methane adsorption isotherms using Molecular Dynamics simulations reflecting the setup of real



Figure SI12: Effect of pore connectivity on adsorption in kerogen. (a) Methane adsorption isotherms at 450 K using GCMC (lines) and MD (symbols) in the four kerogen samples of density 1.2 g/cm³. The color code of the article is used: PY02 in black, MarK in blue, EFK in red and MEK in yellow. (b) Comparison of the amount of adsorbed molecules at a given pressure between GCMC and MD adsorption simulations. The dashed line indicates the results expected for fully connected pore networks, $N_{ads}(MD) = N_{ads}(GCMC)$. The inset shows a high-pressure configuration of the MD methane adsorption simulation. The blue, white and red sticks show the bonds between carbon, hydrogen and oxygen atoms in the kerogen matrix (here MarK), while the methane molecules are represented as dark blue spheres.

adsorption experiments. The initial structure is a kerogen membrane of a size $100 \times 100 \times 50$ Å³. This membrane is in contact with two bulk reservoirs of a size $100 \times 100 \times 100$ Å³ containing methane molecules (see inset of Fig. SI12). At a time t = 0, the system is allowed to relax for 5 ns in the NVT ensemble at T = 450 K with a Nosé-Hoover thermostat. The methane molecules get adsorbed in the material as they diffuse through the connected pores only. At the end of the 5 ns, equilibrium is reached with adsorbed methane in equilibrium with the methane molecules left

get adsorbed in the material as they diffuse through the connected pores only. At the end of the 5 ns, equilibrium is reached with adsorbed methane in equilibrium with the methane molecules left in the bulk reservoirs. While the adsorbed amount is readily obtained by integrating the number N of molecules within the kerogen model, the final equilibrium pressure P is obtained from the density ρ of methane molecules in the bulk reservoirs (the pressure is determined through the virial expansion, $P/k_BT = \rho + B_2(T)\rho^2$, where $B_2(T)$ is the second virial coefficient and k_B is Boltzmann's constant). Fig. SI12 compares the adsorption isotherm for the four kerogen models with a density of 1.2 g/cm^3 , as obtained using GCMC simulations and using the MD strategy above. In agreement with our previous conclusion, these additional simulations show that pore connectivity is poor for all samples (about 80% of the pores are closed in immature models and about 20% and 60% in MarK and PY02). Such a low connectivity for dense kerogens explains the low adsorption levels observed experimentally and confirms the discussion above. Thanks to these data, we can correct the BET surfaces shown in Tab. SI5 for this connectivity factor. Such corrected values S_{BET}^* (Tab. SI5) show that pore accessibility drastically affects the specific surface area assessed by nitrogen adsorption, leading in particular to very small BET surfaces. While this correction is obviously negligible for well connected (mature) kerogens, this correction is crucial for immatures samples (EFK and MEK). In the latter case, when BET surfaces are corrected for pore accessibility, our molecular models recover the order of magnitude observed experimentally for the specific surface areas (a few tens of m^2/g). Considering the large uncertainty over the experimental BET surfaces, this agreement between experimental samples and molecular models can be considered satisfactory. Note that in the case of PY02 the experimental specific surface could not be estimated reliably due to too large error bars.

In order to determine the effect of pore accessibility on the characterization of kerogen, we computed the PSD of the different molecular models when restricted to accessible pores (as probed from the MD simulations described above). Fig. SI13 reports the comparison of the PSD with and without taking into account pore accessibility. While accessible pores tend to be slightly larger (since they are statistically more likely to be connected), they are of the same order of magnitude as those observed in the PSD not corrected for pore accessibility. In particular, the corrected PSD confirms that, for a given density, mature kerogens tend to exhibit larger pores than immature



Figure SI13: Pore size distributions for MarK (top) and EFK (bottom). As in Fig. 3.b., the bars report experimental measurements of the PSD from CO_2 adsorption performed on demineralized samples. The full lines represent the PSD computed for the kerogen molecular models with a density of 1.2 g/cm³. The dashed lines are the PSD for the same molecular models but corrected for pore accessibility, *i.e.* the PSD restricted to pores accessible from the external surface.

kerogens.

- Vandenbroucke, M. & Largeau, C. Kerogen origin, evolution and structure. Org. Geochem. 38, 719–833 (2007). http://dx.doi.org/10.1016/j.orggeochem.2007.01.001.
- Bernard, S., Wirth, R., Schreiber, A., Schulz, H.-M. & Horsfield, B. Formation of nanoporous pyrobitumen residues during maturation of the Barnett Shale (Fort Worth Basin). Int. J. Coal Geol. 103, 3–11 (2012). http://dx.doi.org/10.1016/j.coal.2012.04.010.
- [3] Pernia, D., Bissada, K. A. & Curiale, J. Kerogen based characterization of major gas shales: Effects of kerogen fractionation. Org. Geochem. 78, 52 – 61 (2015). http://dx.doi.org/10.1016/j.orggeochem.2014.10.014.
- [4] Durand, B. & Nicaise, G. Kerogen: Insoluble organic matter from sedimentary rocks (Editions Technip, 1980).
- [5] Suleimenova, A. et al. Acid demineralization with critical point drying: A method for kerogen isolation that preserves microstructure. Fuel 135, 492–497 (2014). http://dx.doi.org/10.1016/j.fuel.2014.07.005.
- [6] Redding, C., Schoell, M., Monin, J. & Durand, B. Hydrogen and carbon isotopic composition of coals

and kerogens. *Phys. Chem. Earth* **12**, 711–723 (1980). http://dx.doi.org/10.1016/0079-1946(79)90152-6.

- Bousige, C., Boţan, A., Ulm, F.-J., Pellenq, R. J.-M. & Coasne, B. Optimized molecular reconstruction procedure combining hybrid reverse Monte Carlo and molecular dynamics. J. Chem. Phys. 142, 114112 (2015). http://dx.doi.org/10.1063/1.4914921.
- [8] Leyssale, J.-M., Costa, J.-P. D., Germain, C., Weisbecker, P. & Vignoles, G. Structural features of pyrocarbon atomistic models constructed from transmission electron microscopy images. *Carbon* 50, 4388 – 4400 (2012). http://dx.doi.org/10.1016/j.carbon.2012.05.015.
- [9] Qiu, X., Thompson, J. W. & Billinge, S. J. PDFgetX2: a GUI-driven program to obtain the pair distribution function from x-ray powder diffraction data. J. Appl. Cryst. 37, 678–678 (2004). http://dx.doi.org/10.1107/S0021889804011744.
- [10] Soper, A. K. & Barney, E. R. On the use of modification functions when fourier transforming total scattering data. J. Appl. Cryst. 45, 1314–1317 (2012). http://dx.doi.org/10.1107/S002188981203960X.
- [11] Cullity, B. D. Elements of x-ray diffraction. American Journal of Physics 25, 394–395 (1957). http://dx.doi.org/10.1119/1.1934486.
- [12] Squires, G. Thermal neutron scattering (Cambridge University Press, Cambridge, 1978).
- [13] NIST. URL http://www.ncnr.nist.gov/resources/n-lengths/.
- [14] Sun, H. Compass: An ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. J. Phys. Chem. B 102, 7338–7364 (1998). http://dx.doi.org/10.1021/jp980939v.
- [15] Materials Studio Modelling Environment v. 6.1. Accelrys Inc.. San Diego, CA (2013).
- [16] Ni, B., Lee, K.-H. & Sinnott, S. B. A reactive empirical bond order (REBO) potential for hydrocarbonoxygen interactions. J. Phys.: Cond. Mat. 16, 7261 (2004). http://dx.doi.org/10.1088/0953-8984/16/41/008.
- [17] van Duin, A. C., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: a reactive force field for hydrocarbons. J. Phys. Chem. A 105, 9396–9409 (2001). http://dx.doi.org/10.1021/jp004368u.
- [18] Hinsen, K., Pellegrini, E., Stachura, S. & Kneller, G. R. nMoldyn 3: Using task farming for a parallel spectroscopy-oriented analysis of molecular dynamics simulations. J. Comp. Chem. 33, 2043–2048 (2012). http://dx.doi.org/10.1002/jcc.23035.
- [19] Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. J. Comp. Phys. 117, 1–19 (1995). URL http://lamps.sandia.gov. http://dx.doi.org/10.1006/jcph.1995.1039.
- [20] Smith, N. O. The difference between Cp and Cv for liquids and solids. J. Chem. Edu. 42, 654 (1965). http://dx.doi.org/10.1021/ed042p654.
- [21] Macrae, J. & Ryder, C. Thermal expansion of coal. Nature 176, 265 (1955). http://dx.doi.org/10.1038/176265a0.
- [22] Huang, D., Simon, S. L. & McKenna, G. B. Chain length dependence of the thermodynamic properties of linear and cyclic alkanes and polymers. J. Chem. Phys. 122, 084907 (2005).

http://dx.doi.org/10.1063/1.1852453.

- [23] Tomeczek, J. & Palugniok, H. Specific heat capacity and enthalpy of coal pyrolysis at elevated temperatures. Fuel 75, 1089 – 1093 (1996). http://dx.doi.org/10.1016/0016-2361(96)00067-1.
- [24] Camp, D. W. Oil shale heat-capacity relations and heats of pyrolysis and dehydration (1987).
- [25] NIST. URL http://webbook.nist.gov/cgi/cbook.cgi?ID=C7782425&Mask=2.
- [26] Ungerer, P., Collell, J. & Yiannourakou, M. Molecular modelling of the volumetric and thermodynamic properties of kerogen: influence of organic type and maturity. *Energy & Fuels* 29, 91–105 (2015). http://dx.doi.org/10.1021/ef502154k.
- [27] Sherritt, R., Jia, J., Purnomo, M. & Schmidt, J. Advances in steady-state process modeling of oil shale retorting. Vacuum 425, 600 (2009).
- [28] Van Speybroeck, V., Gani, R. & Meier, R. J. The calculation of thermodynamic properties of molecules. *Chem. Soc. Rev.* **39**, 1764–1779 (2010). http://dx.doi.org/10.1039/B809850F.
- [29] Marrero, J. & Gani, R. Group-contribution based estimation of pure component properties. Fluid Phase Eq. 183, 183–208 (2001). http://dx.doi.org/10.1016/S0378-3812(01)00431-9.
- [30] Joback, K. G. & Reid, R. C. Estimation of pure-component properties from group-contributions. Che. Eng. Comm. 57, 233–243 (1987). http://dx.doi.org/10.1080/00986448708960487.
- [31] Coasne, B., Galarneau, A., Pellenq, R. J. & Di Renzo, F. Adsorption, intrusion and freezing in porous silica: the view from the nanoscale. *Chem. Soc. Rev.* 42, 4141 (2013). http://dx.doi.org/10.1039/C2CS35384A.
- [32] Billemont, P., Coasne, B. & De Weireld, G. Adsorption of carbon dioxide, methane, and their mixtures in porous carbons: effect of surface chemistry, water adsorption, and pore disorder. *Langmuir* 29, 3328–3338 (2013). http://dx.doi.org/10.1021/la3048938.
- [33] Pikunic, J. et al. Structural modeling of porous carbons: constrained reverse monte carlo method. Langmuir 19, 8565–8582 (2003). http://dx.doi.org/10.1021/la034595y.
- [34] Vandamme, М., Ulm, F.-J. & Fonollosa, Ρ. Nanogranular packing of C-S-H at substochiometric conditions. Cement andConcrete Research **40**, 1426(2010).http://dx.doi.org/10.1016/j.cemconres.2009.09.017.