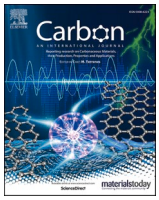
Carbon 225 (2024) 119086

Contents lists available at ScienceDirect

Carbon

journal homepage: www.elsevier.com/locate/carbon



The structure of appalachian coal: Experiments and *Ab initio* modeling

Chinonso Ugwumadu [a,*], David A. Drabold [a,**], Natasha L. Smith [b], Jason Trembly [c], Rudolph Olson III [b], Eric Shereda [d], Yahya T. Al-Majali [c,***]

[a] Department of Physics and Astronomy, Nanoscale and Quantum Phenomena Institute (NQPI), Ohio University, Athens, OH, 45701, USA
[b] CONSOL Innovations, Triadelphia, WV, 26059, USA
[c] Department of Mechanical Engineering, Institute for Sustainable Energy, and the Environment (ISEE), Ohio University, Athens, OH, 45701, USA
[d] CONSOL Energy, Canonsburg, PA, 15317, USA

ARTICLE INFO

Keywords:
Coal
NMR
FTIR
XPS
ReaxFF
DFT

ABSTRACT

This study focuses on the construction of small-scale atomistic representations of three Appalachian coals of different ranks– high volatile bituminous, low volatile bituminous, and semi-anthracite. Ultimate, proximate, ^{13}C NMR, FTIR, and XPS data were used to infer 2D and construct multiple 3D molecular models for each coal. A new simulation strategy for structure optimization, termed "Sectioned Optimization", is introduced to ensure energetically stable configurations of the 3D coal models. Density functional methods are applied to the models, improving their quality, and giving insight into the limitations of empirical simulations in this complex system. The structural models were validated *ex post facto* using a density functional code, VASP. We describe the distribution of local atomic environments in the models to emphasize the significant structural distinctions between the three coal ranks. We also report the first *ab initio* vibrational density of states calculations for coal and compare it to FTIR and Inelastic Neutron Scattering experiments and identify vibrational modes using *ab initio* calculations. We also provide animations of the vibrational normal modes in coal that are consistent with FTIR band assignments. We show that the specific heat capacity compares favorably with experimental data.

1. Introduction

Coal, a ubiquitous form of matter on Earth, exhibits diverse variations across locations and seams. The intricate molecular composition and structure of coal remains insufficiently understood. The evolving landscape of coal processing, extending beyond fuel production to materials engineering [1–3], necessitates a comprehensive understanding of its fundamental chemistry. An example of this evolving narrative is the "greening of coal", where coal finds utility as a filler material in composites with various plastics [4–7]. Another compelling example is the heightened research focus on coal graphitization [8–13]. This comes in response to the issue both natural and synthetic graphite availability, impacting the industrial production of devices such as lithium-ion batteries, lubricants, and refractories, among others [14]. If coal is going to be used to support the demand for graphite, these applications will require a chemically realistic understanding of coal's molecular composition and structure at the nanoscale, with a specific emphasis on the interatomic interactions within coal and interfacial dynamics between coal and other materials.

Following the pioneering work by van Krevelen [15], several coal models have been proposed [16–20], and there exists a body of work that employ these models in exploring useful coal properties and coal dynamics [21–23]. The prevailing view of coal's structure conceptualizes it as clusters of fused aromatics and hydroaromatic rings, potentially linked by weaker aliphatic bonds. Within these ring clusters there exist functional groups of non-carbon elements such as oxygen, sulfur, and nitrogen. Assuming a universally applicable structure for coal is misleading, as coal structures can vary drastically between coal beds; even within the same rank, there are significant variations. Consequently, it oversimplifies the complexity of the molecular composition and structure to adopt a proposed model for a particular coal solely based on its rank.

Proximate/ultimate analyses, cross-polarization/magic angle spinning (CP/MAS) ^{13}C nuclear magnetic resonance (NMR), Fourier

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: cu884120@ohio.edu (C. Ugwumadu), drabold@ohio.edu (D.A. Drabold), almajali@ohio.edu (Y.T. Al-Majali).

https://doi.org/10.1016/j.carbon.2024.119086
Received 28 January 2024; Received in revised form 19 March 2024; Accepted 27 March 2024
Available online 2 April 2024

transform infrared spectroscopy (FTIR), and X-ray photoelectron spectroscopy (XPS) are analytical techniques that provide valuable information on the molecular composition and structure of coal. Proximate/ultimate analyses offer key insights into the fundamental constituents of coal, aiding in its rank classification and understanding its combustion behavior. CP/MAS ^{13}C NMR elucidates the carbon framework of coal molecules by detecting the resonances of carbon nuclei in diverse environments [20,24–28]. FTIR identifies characteristic patterns corresponding to functional groups such as aliphatic and aromatic hydrocarbons, carbonyl, hydroxyl, nitrogen, and sulfur compounds, by measuring the absorption of infrared light [29,30]. XPS analyzes the photoemission of electrons from the coal surface when exposed to x-rays, discerning elements and their chemical environments, including functional groups and heteroatoms [31–33].

While ^{13}C NMR, FTIR and XPS offer valuable insights, each has specific limitations, and all suffer from the "scourge of averaging" (local information is lost by measuring spectra average over the entire sample) [34]. These shortcomings create difficulties in quantitative analysis due to overlapping peaks and limited spatial resolution, particularly when capturing localized structural differences [30,35,36]. Notably, both FTIR and XPS primarily provide surface-sensitive information, potentially leading to incomplete representations of the bulk composition [37,38]. Nonetheless, the synergistic analysis of results from these methods enhances the overall comprehension of coal molecular composition and structure, providing a more comprehensive characterization of both surface and bulk properties.

Inferring two-dimensional structures for coal through these analytical techniques is only a preliminary step. To truly comprehend the local chemistry of the coal, it is essential to leverage the inferred structure through the creation of realistic, small-scale three-dimensional structural models. These models are essential ingredients for molecular dynamics simulations, providing insights into the coal's behavior in all thermodynamic conditions and its interfacial interaction with other materials [4,12,20,23,39,40]. This study employs analyses of proximate, ultimate, ^{13}C CP/MAS NMR, FTIR, and XPS data, to characterize the chemical and molecular structure of coal. Realistic atomistic models are determined for three coals from Pennsylvania: high volatile bituminous (Pittsburgh seam No. 8; P8), low volatile bituminous (Keystone 325; KS-325), and semi-anthracite (Keystone 121; KS-121) coals. Fig. S1 in the supplementary material delineates the major bituminous and anthracite fields in Pennsylvania. The annotated regions, denoted as "A" and "B", specify the mining sites for P8 and KS-325 respectively, situated in Allegheny and Somerset County, Southwestern Pennsylvania. The location labeled as "C" identifies the site for KS-121 in Northumberland County, Northeastern Pennsylvania. The macromolecular structural models predicted from these methods are designed to comply with experimental NMR data, constrained by FTIR and XPS data. Additionally, recognizing the intricate nature of coal structure, a minimum of three models were generated for each coal, each exhibiting some unique structural characteristics while reproducing the experimental NMR spectra.

The models were converted into realistic 3D models with periodic boundaries using the reactive force field (ReaxFF) potential [41] as implemented in the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [42]. We introduce a structure optimization technique termed *sectioned optimization*, which ensures the generation of realistic coal molecular structures with minimum local atomic energies. To achieve chemically realistic structures, the bond configurations in the models were refined through density functional theory (DFT) energy minimization in VASP (Vienna *Ab Initio* Simulation Package) [43]. The variations in the local chemical environment of the DFT-treated models were assessed using the Smooth Overlap of Atomic Positions (SOAP) framework [44–46], providing a description and analysis for the atom-centered environment in the structures that remains invariant to permutations, translations, and rotations. The models were further explored by calculating the vibrational density of states (VDoS), serving as a theoretical benchmark for comparison with experimentally obtainable vibration data, including FTIR and Inelastic Neutron Scattering (INS), facilitating the identification of a variety of new predictions for vibrations in coal. INS almost directly provides the VDoS and typically offers more information than energy-dependent spectroscopic methods such as FTIR or Raman Spectroscopy [47,48]. It is also the preferred method for the analysis of hydrogenous species in a non-hydrogenous matrix [49]. The VDoS is also used to compute the specific heat capacity, which was compared with experimental specific heat data. We note that this is the first reported *ab initio* vibrational density of states calculation for coal.

The models derived from this study are suited for accurate computer simulations that can explain and predict structural, electronic, vibrational, and thermal properties, as well as combustion characteristics. Specifically, they can be applied to investigate interfacial dynamics in composite materials, where the matrix is comprised of coal and various plastics [4], as well as to study the electronic properties of coal-derived graphite-like materials [9,13]. In a previous work, we outlined a method to simulate carbonization and graphitization processes of coal-like materials [12], however, achieving accurate and reliable results necessitates chemically accurate and structurally robust models for the coal of interest. This is particularly relevant in research concerning the projection of the electronic conduction of a material in spatial grids, where local atomic configurations of atoms are the driving forces [50].

Another aspect of this work is our documentation of the procedures involved in analyzing, constructing, validating, and comparing coal models. This level of detail allows other researchers to replicate the process on different coals. Furthermore, we include the coal models derived in this study along with supplementary information to enhance accessibility and facilitate further research in this domain.

2. Methods

2.1. Experiments

Pittsburgh seam No. 8 (P8), Keystone 325 (KS-325), and Keystone 121 (KS-121) coal, sourced from coal mines in the Commonwealth of Pennsylvania, United States, were used for this study (refer to Fig. S1 for regions where the coals were mined). The D50 particle size for P8 is about 25 μm (500 mesh) and 34 μm (400 mesh ASTM E−11 sieve) for KS-325 and KS-121. Prior to analyses, the coal powder was dried in a convection oven at 110 °C for 24 h to decrease the moisture content to less than 1 wt%.

2.1.1. Proximate, ultimate and calorific analyses

A Blue-M drying oven was used to assess the coal's moisture content at ~105° C. The coal ash yield was examined employing a Thermcraft XSB-10 muffle furnace, heated to ~1100° C in air. The determination of fixed carbon and volatile matter yield was performed using a Hoskins vertical furnace. For the ultimate analysis, which included carbon, hydrogen, and nitrogen, a LECO 628 elemental determinator was employed, with sulfur content determined by incorporating a *Sulfur Add-on* Module. A PARR 6400 calorimeter was used for calorific analysis. Proximate analysis was conducted using the ASTM D3172-13 standard methods [51]. For the ultimate analysis, the testing method adhered to ASTM D3176-15 standards [52]. In detail, carbon, hydrogen, and nitrogen were analyzed following ASTM D5373-08 [53], while sulfur analysis followed the guidelines outlined in ASTM D4239-18 [54].

The United States classification system for medium- and high-rank coal, also used in this study, is based on volatile matter and fixed carbon content in the dry mineral-matter-free (dmmf) basis [55]. In the dmmf basis, the analyses and calculations are carried out assuming no moisture or mineral matter in the coal sample [56]. An estimate of the mineral-matter content (M_b) is obtained from the dry ash yield (A) using the Parr formula [57]:

$$M_b = (1.08 \times A) + (0.55 \times S_t) \quad (1)$$

where S_t is the total sulfur concentration (in wt.%) obtained from the ultimate analysis. The volatile matter yield on dmmf basis (V_{dmmf}) is then calculated as:

$$V_{dmmf} = \frac{100}{100 - M_b} \times \frac{100 \times V_d}{100 - A} \quad (2)$$

where V_d is the percentage of volatile matter in the dry basis. The fixed carbon in the dmmf basis is derived as 100 - V_{dmmf}.

2.1.2. Nuclear magnetic resonance spectroscopy

The ^{13}C Cross-Polarization Magic Angle Spinning Nuclear Magnetic Resonance Spectroscopy (*CP*-MAS NMR) experiments were conducted on a Bruker Avance III HD 600 MHz DNP spectrometer and was equipped with a triple-resonance HXY 3.2 mm DNP solid-state probe. The probe was tuned to ^{1}H and ^{13}C double mode. Each sample was packed into 3.2 mm rotors and spun at 10 kHz MAS at room temperature (~300 K) and collected with 24k scans (17 h) each. Pulses and chemical shifts were externally calibrated to *N*-acetyl-valine.

The NMR spectra are partitioned into three main regions: aliphatic carbon (0–90 ppm), aromatic carbon (91–170 ppm), and carboxyl/carbonyl group (171–225 ppm) [20,58]. In this study, the structural parameters (denoted by f_j^i) assigned to NMR regions are based on previous studies on coal NMR [20,25,26,58,59], and summarized in Table 1. Aliphatic carbons were classified into four sub-groups: terminal methyl carbon (f_{al}^* , 0–25 ppm), CH, CH_2 and quaternary carbon (f_{al}^H, 25–51 ppm), oxy-methylene or oxy-methine carbon (f_{al}^O, 51–75 ppm), and carbon adjacent to oxygen atoms in alcohols or aliphatic ethers (f_{al}^O, 75–90 ppm). The aromatic carbon region between 90 and 150 ppm is divided into aromatic carbon atoms bound to hydrogen (f_a^H, 90–129 ppm), bridgehead and substituted aromatic carbon atoms connected to hydroaromatic CH_2 carbon (f_a^B, 129–137 ppm), and alkyl side chains (f_a^S, 129–137 ppm). Oxygenated aromatic carbons linked to phenolic hydroxyl, methoxy, and aromatic ether (phenolic ether) oxygen atoms (f_a^P) are attributed to the chemical shift between 150 and 165 ppm. The bands ranging from 165 to 188 ppm correspond to carboxyl carbon, while those spanning 188–220 ppm were associated with carbonyl carbon (both denoted as f_a^C). Illustrations of aromatic clusters, bridges, loops, and side chains are presented in Fig. S2.

Seven additional structural parameters for the coal samples were derived from some of the structural parameters obtained directly from the ^{13}C NMR spectra. The calculated structural parameters, also summarized in Table 1, include the fraction of aromatic bridgehead carbon (X_b), average number of carbon atoms within aromatic clusters (n_{cl}), the number of attachments per cluster (σ), fraction of intact bridges per cluster (I_b), the number of bridges and loops per cluster (*BL*), number of side chains per cluster (*SL*), and the total molecular weight of the cluster (W_c). The methodology employed to obtain the additional parameters is based on the work of Solum, Pugmire, and Grant [24,25], and the equations are derived in Section S1.1 of the supplementary material.

Table 1
Structure parameters assignment for solid ^{13}C NMR spectra.

Parameters	Chemical Shift [ppm]	Assignment
f_{al}^*	0–25[a]	Terminal-CH_3 carbon
f_{al}^H	25–51[b]	CH, CH_2 and quaternary carbon
f_{al}^O	51–90[c]	Aliphatic carbon bonded to oxygen
f_{al}	0–90	Total aliphatic carbon
f_a^H	90–129	Protonated aromatic carbon
f_a^B	129–137	Aromatic bridgehead carbon
f_a^S	137–150	Alkylated aromatic carbon
f_a^P	150–165	Phenolic or phenolic ethers
f_a^N	129–165	Non-protonated aromatic carbon
f_a^C	165–220[d]	Carbonyl and carboxyl carbon
f_a	90–220	Total aromatic carbon
X_b	(129–137)/(90–165)	Fraction of aromatic bridgehead carbon
n_{cl}	See Equation S2 and S3	Number of carbon atoms per cluster
σ	See Equation S4	Cluster coordination number
I_b	See Equation S5	Fraction of intact bridges
BL	See Equation S6	Fraction of looped and aliphatic bridges
SC	See Equation S7	Side chains per cluster
W_c	See Equation S8	Molecular weight of clusters in g/mol

[a] Aliphatic terminal-CH_3 carbon (0–16 ppm), aromatic terminal-CH_3 carbon (16–25 ppm).
[b] CH or CH_2 carbon (25–36 ppm), quaternary carbon (36–51 ppm).
[c] Oxy-methylene or oxy-methine carbon (51–75 ppm), carbon adjacent to oxygen atoms in alcohols or aliphatic ether groups (75–90 ppm).
[d] Carboxyl carbon (165–188 ppm), carbonyl carbon (188–220 ppm).

2.1.3. Attenuated total reflectance fourier transform infrared spectroscopy

Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy (ATR-FTIR or simply FTIR) was conducted using a Thermo Scientific Nicolet 6700. The coal samples were tested using a diamond ZnSe crystal for a measurement range of 4000–400 cm^{-1} and a spectra resolution of 1.0 cm^{-1}. Three spectra were recorded using 256 scans for each coal sample.

2.1.4. X-ray photoelectron spectroscopy

X-ray Photoelectron Spectroscopy (XPS) was used to elucidate surface properties of the coal samples. The investigations were conducted on the NEXSA G2 instrument (Thermo Scientific) employing an Al Kα source (1486.6 eV). Initial surveys were conducted for all samples, employing a pass energy of 100 eV, Step size of 1.0 eV, and a dwell time of 10 m s. High-resolution spectra for C 1s, N 1s, O 1s, Al 2p, S 2p, and Cu 2p were then collected with a pass energy of 50 eV, maintaining the same step size and dwell time. In this study, only spectra for C, N, O, and S are considered. To compensate for charging effects, a flood gun with a current of 150 μA was employed.

The spectra for P8, KS-325, and KS-121 were calibrated to the aliphatic carbon feature with peaks at 284.3, 284.2, and 283.9 eV, respectively. OriginPro® was employed in fitting the peaks; utilizing the pseudo-Voigt function with the ratio of Gaussian to Lorentzian shape profile factors was allowed to vary between 20:80 % to 50:50 % [60]. Shirley or linear background subtraction methods were used for baseline correction of the spectra. Before peak fitting, the vertical offset was fixed at 0 to reduce the number of fitting variables. Due to the low noise-to-signal ratio for the nitrogen and sulfur XPS data, signal smoothing was applied before peak fitting.

2.2. Simulation

2.2.1. 2D model construction from experimental data

Methods for constructing coal models have been extensively reviewed by Matthews et al. [21] and by Liu et al. [23]. Our procedure for constructing 2D structural models for coal, outlined in the flowchart depicted in Fig. 1, adopts a bottom-up strategy. We commence with carbon skeletal structures comprising clusters and aliphatic loops/bridges, gradually incorporating non-carbon functional groups as the system size expands. The decision-making process regarding model expansion and the inclusion or exclusion of functional groups follows a systematic trial-and-error methodology. This involves comparing calculated NMR spectra from a proposed structure with experimental NMR spectra, until they align.

A similar flowchart to Fig. 1 is available in Figure S3 of the supplementary material, accompanied by a step-by-step elucidation of our method for constructing 2D coal model. Below, we outline the general procedure:

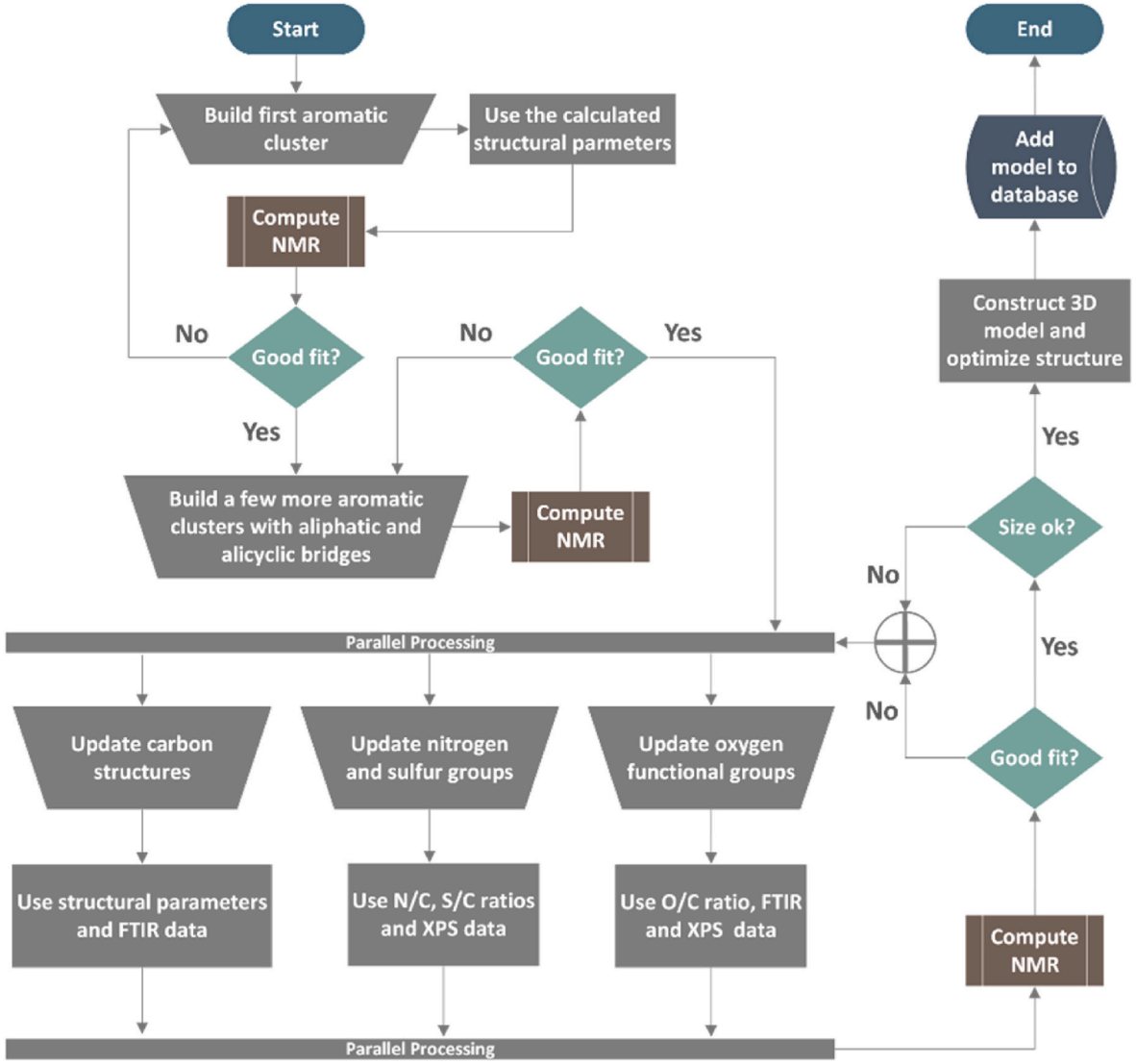


Fig. 1. Flowchart for the bottom-up, trial-by-error approach to constructing coal models using experimental data. Step-by-step elucidation of the method is provided in Fig. S3. (A colour version of this figure can be viewed online.)

1. Gather experimental information on aliphatic and aromatic carbon structures, along with various functional groups, obtained using the aforementioned methods.
2. Delineate 2D molecular structures of the coal using a molecular editor, such as ChemDraw®. We recommend models with at least 100 carbon atoms or system size above 200 atoms (depending on the coal rank)
3. Compute the NMR spectra for the initial structure, utilizing tools like ChemDraw® or MestRenova®. An example of the NMR chemical shifts computed using ChemDraw® is depicted in Figure S4.
4. The model's small volume, compared to real materials, causes the calculated NMR spectrum to appear as spikes at different chemical shifts, resembling δ functions. OriginPro® is employed to combine these spikes and construct a curve mirroring the experimental data, using peak convolution algorithms.
5. Iteratively refine the structure to coincide with the experimental ^{13}C NMR spectrum. Pay attention to the analysis from other experimental methods (FTIR, XPS, ultimate analysis etc.), considering them as limiting factors (see discussion in Figure S3).

In Fig. 1, all shapes within the flowchart adhere to their conventional meanings [61]. The *parallel processing* bar signifies that the three manual processes (represented by trapezoids) for updating the carbon, nitrogen, sulfur, and oxygen structures occur synchronously. The *Compute NMR* subroutine can be partially automated by using input scripts to execute tasks in MestReNova® or ChemDraw®. Similarly, the decision for a *good fit* can also be automated in OriginPro®, although human intervention still is necessary to update structures and assess whether the model aligns with the desired chemistry, based on experimental data.

2.2.2. 3D model construction and ReaxFF optimization

The final 2D coal models are first converted into 3D structures using the Chem3D feature in ChemDraw®. Optionally, the MM2 force field [62], implemented in Chem3D, can be used to for crude optimization of the structure through steric energy minimization. To build realistic *three-dimensional coal models* that are useful for local chemical analysis, especially at classical molecular dynamics and density functional theory (DFT) levels, the 3D models obtained from ChemDraw® were optimized using the ReaxFF interatomic potential [41]. ReaxFF, designed for systems with chemical reactions like coal [12], was implemented in LAMMPS [42], encompassing interactions for hydrogen, carbon, nitrogen, oxygen, and sulfur.

In LAMMPS, the conventional procedure for creating 3D coal models involves placing the model in a cubic box, and then heating it under isobaric-isothermal (NPT) conditions to optimize the structure and density. This is followed by energy minimization (using the conjugate gradient algorithm) to attain a minimum energy stereochemical configuration. However, this conventional method often fails, particularly if the starting configuration is far from the equilibrium conformation. In practice, this often leads to highly unrealistic models. To address this, we introduced a new structure optimization strategy that we have termed "sectioned optimization". The procedure involves cycles of: (1) heating the system in a canonical (NVT) ensemble for atom rearrangement, (2) continuing this heating process in an NPT ensemble, (3) minimizing the energy of the intermediate structure using conjugate gradient algorithm, and then (4) repeating steps 1–3 until convergence in the energy and density of the system is achieved. The key point here is that the first iteration of the NVT (step 1) and NPT (step 2) simulation should be executed over shorter runtimes (about 25–50 ps). Subsequently, longer runs (≥100 ps) can be employed until density and energy convergence are achieved. This approach guides the system to an optimized energy conformation.

An implementation of sectioned optimization on a representative KS-121 coal model is depicted in Fig. 2. The region "a" in the density plot identifies steps 1–3 and region "b" is the conventional method. The unmarked area between a and b represents the transition region where the simulation runtime is increased. The system pressure variation is illustrated in the lower panel of Fig. 2, using moving averages (MA) with periods of 0.25 (brown) and 1 ps (gray). To obtain the optimal models in this work, the temperature and pressure were fixed at 500 K and 0.4 GPa. This is a general burial temperature and pressure for coal [12]. Temperature and pressure were controlled using the Nosé-Hoover thermostat and barostat, respectively [63–65], employing a time step of 0.25 fs. The barostat was set to act independently along the three dimensions (x, y, and z), so that the diagonal components of the stress tensors (P_{xx}, P_{yy}, and P_{zz}) are not coupled. The damping constants (τ) [66] were set to 100 and 1000 timesteps for the thermostat and barostat, respectively. This means that the temperature and pressure are rescaled every 25 and 250 fs, respectively.

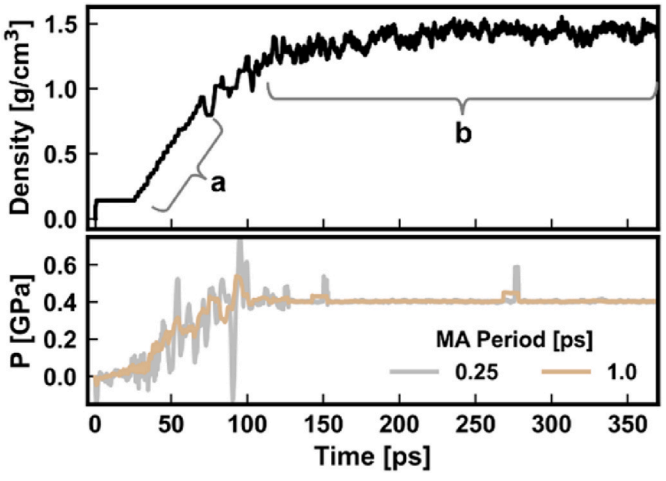


Fig. 2. Time development for the density and pressure (P) convergence at 0.4 GPa for a KS-121 coal model. The sectioned optimization region is labeled "a" in the density plot. The region "b" corresponds to the conventional optimization method. The unlabeled area between a and b is the transition region. The lower panel illustrates pressure variations, employing moving averages (MA) with periods of 0.25 (brown) and 1 ps (gray). (A colour version of this figure can be viewed online.)

2.2.3. DFT post-treatment, vibrations and specific heat capacity

To obtain the highest level of energetic stability in the coal models, the energy and structure of the models were validated *ex post facto* via DFT, implemented within VASP. Zero-pressure (conjugate gradient) relaxation of the models was carried out until energy convergence with a tolerance of 10^{-5} eV. The vibrational density of states (VDoS) of the DFT-relaxed models were computed from the force constants (Hessian matrix), using the finite difference approach – the dynamical matrix was constructed, diagonalized, and used to determine the phonon modes and frequencies of the models. Within the harmonic approximation, the VDoS ($g(\omega)$) is given as:

$$g(\omega) = \frac{1}{3N}\sum_{k=1}^{3N}\delta(\omega - \omega_k) \quad (3)$$

where, N and ω represent the number of atoms and the eigen-frequencies of normal modes, respectively. The δ function is approximated by a Gaussian with a standard deviation of 1.5 % of the maximum frequency. To quantify the localization of the vibrational normal modes, vibrational inverse participation ratio (VIPR) was also calculated as:

$$V(\omega_n) = \frac{\sum_i^N |u_n^i|^4}{\left(\sum_{i=1}^N |u_n^i|^2\right)^2} \quad (4)$$

Here, u_n^i is the displacement vector of the i^{th} atom for the n^{th} normal mode frequency (ω_n). VIPR ranges from $1/N$ to 1, where VIPR → $1/N$ indicates that several atoms contribute to the normal mode and VIPR → 1 indicates fewer atoms contribute to that eigen frequency, signifying a highly localized mode.

Exploiting the vibrational density of states, the temperature-dependent specific heat capacity, denoted as $C(T)$ was computed, using the following equation derived from elementary statistical mechanics [67,68]:

$$C(T) = \int_0^{E_{max}} \left(\frac{h\omega}{K_B T}\right)^2 \frac{\exp\left(h\omega/K_B T\right)}{\left[\exp\left(h\omega/K_B T\right) - 1\right]^2} g(\omega)\, d\omega \quad (5)$$

A single k-point (Γ) and periodic boundary conditions were used for all DFT calculations, and the interactions between the electrons and ions were described using the Perdew–Burke–Ernzerhof projected-augmented-wave pseudopotential [69,70]. We set a cutoff energy of 400 eV for the plane-wave-basis used to expand electronic wave functions during the MD simulation. For the vibrational density of states calculations, the tolerance for the total energy convergence was reduced to 10^{-6} eV.

2.2.4. The Smooth Overlap of Atomic Positions (SOAP) framework

Identifying clear structural differences among coal structures poses challenges due to shared features like polyaromatic hydrocarbons across various seams and ranks. Without a localized metric to confirm the uniqueness of a specific coal structure in comparison to others, meaningful progress in developing distinct coal structures for the same type or different ranks remain elusive. To address this challenge, we have adopted the Smooth Overlap of Atomic Positions (SOAP) framework, originally developed by Bartók, Kondor, and Csányi [44,45], to differentiate between structures and confirm their unicity. SOAP derives a permutation, translation, and rotation invariant description of the atomic environment in given coal structure (henceforth referred to as descriptors). It captures atomic geometry regions by employing a local expansion involving a Gaussian-smoothed atomic density, utilizing orthonormal functions derived from real (tesseral) spherical harmonics, $Y_l^m(\theta,\varphi)$, and radial basis functions, $\rho_n(r)$ [71]. The DScribe software package [46] was employed to obtain SOAP descriptors, which is a power spectrum vector P, defined as:

$$P_{nnl}^{Z_1Z_2} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m d_{nlm}^{Z_1} d_{nlm}^{Z_2}$$

$$d_{nlm}^{Z} = \iiint_{\mathscr{R}^3} dV\, \rho_n(r) Y_l^m(\theta,\varphi) g^Z(r) \quad (6)$$

$$g^Z(r) = \sum_i^{|Z_i|} e^{-1/2\sigma^2 |r-R_i|^2}$$

where n and n' are radial basis functions (up to some n_{max}), l is the angular degree of the spherical harmonics. Z_1 and Z_2 are atomic species. The function $g^z(r)$ is a Gaussian smoothed density of atoms with atomic number Z. To characterize the local environments of atoms in this study, a cutoff radius of 3 Å was applied. The maximum radial basis function and the angular degree of spherical harmonics, Y_l^m, were set to 9. Employing these parameters, SOAP descriptors for each coal structure were generated as a $N \times L$ dimensional matrix, where N corresponds to the number of atoms in the structure and L is the descriptor vector for each atom, comprising 10,350 elements.

Utilizing the SOAP descriptors derived for the atom-centered environments (α and β) for two coal structures (A and B), a measure of (dis) similarity between the structures is determined through the average kernel approach ($\overline{K}(A,B)$) given as [45]:

$$\overline{K}(A,B) = \frac{1}{NM} \sum_{\alpha\beta} C_{\alpha\beta}(A,B) \quad (7)$$

Where N and M denote the number of atoms in structure A and B, respectively. The term $C_{\alpha\beta}(A,B)$ represents a pairwise metric, chosen as the radial basis function (rbf) kernel [72] in this study. The value of $\overline{K}$ ranges from 0 to 1, with $\overline{K} \rightarrow 1$ signifying that the two structures being compared have similar or the same local environment, and $\overline{K} \rightarrow 0$, meaning the opposite. $\overline{K}$ is a symmetric matrix with the number of rows corresponding to the number of structures being compared.

3. Results

Experimental data from four techniques: (1) proximate and ultimate analyses, (2) nuclear magnetic resonance (NMR) spectroscopy, (3) attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy, and (4) x-ray photoelectron spectroscopy (XPS) were collectively utilized in estimating the elemental composition and molecular structural characteristics of P8, KS-325, and KS-121 coals. The subsequent sub-sections provide discussions of the experimental results.

3.1. Proximate and ultimate analyses

The proximate analysis is presented in Table 2. The ash yield in KS-121 is twice that of the bituminous coals (P8 and KS-325) and contains the lowest percentage of volatile matter. The weight percentage of fixed carbon in KS-325 coal is slightly higher (~4.5 %) than KS-121, with P8 coal having the lowest fixed carbon content at ~53 %. From Equation (2), the volatile matter yield on the dmmf basis for P8, KS-325, and KS-121 coals is 46.63, 19.10, and 14.78 wt%, respectively. The elemental concentration of carbon, hydrogen, nitrogen, and sulfur in the coal samples were determined from ultimate analysis. The element to carbon ratios for the total hydrogen (H/C), nitrogen (N/C), oxygen (O/C), and sulfur (S/C) in the coal samples is presented in Table 2.

3.2. Nuclear magnetic resonance spectroscopy

The ^{13}C NMR spectra for P8, KS-325, and KS-121 is illustrated in the left panel of Fig. 3. The NMR peaks were fitted into 8 major peaks using the OriginPro® software (version 2023b). The peaks are labeled and color-coded "a – h" with the corresponding functional groups responsible for those peaks displayed in the right panel of Fig. 3, with the contributing carbon group colored red. The relative ratios of aliphatic, aromatic, and oxygenated carbon in the coal samples are presented in Table 3. The (gray) shaded columns in Table 3 are the structural parameters, which have been described previously in Table 1. The calculated structural parameters [24] are also presented. Additional details, such as the area of the peaks and the modal position of the NMR chemical shift position for each coal, are provided in Table S1.

3.3. Attenuated total reflectance fourier transform infrared spectroscopy

The fitted peaks obtained from ATR-FTIR spectroscopy for P8, KS-325, and KS-121 are shown in the first, second, and third columns in Fig. 4, respectively. The rows labeled I–IV represent frequency regions corresponding to different chemical environments in the coal. The frequency regions considered are (I) 700–900 cm^{-1}, (II) 1000–1800 cm^{-1}, (III) 2800–3100 cm^{-1}, and (IV) 3100–3600 cm^{-1}. Table 4 presents the relative percentage ratios calculated for the peaks in each frequency region, with peak assignments drawn from the work of Painter et al. [28]. For ease of reference, the average peak positions calculated for all the coals are presented in Table 4 and utilized in subsequent discussions. However, precise peak positions, as well as the peak areas for each coal, are presented in Tables S2 to S5 for the frequency regions I to IV, respectively. Furthermore, the complete frequency range of the FTIR spectra for all the coals is depicted in Figure S5.

The peaks a – g in Row I of Fig. 4 (700–900 cm^{-1} region in Table 4) corresponds to aromatic substitutions present in the coals. The number of adjacent hydrogens in aromatic rings offers an estimate of the degree

Table 2
Proximate and ultimate analyses (on dry basis) of the coal samples.

Coal Samples	Proximate [%]			Ultimate Analysis [%]					Atomic Ratio			
	Ash	Volatile Matter	Fixed Carbon	Carbon	Hydrogen	Nitrogen	Oxygen	Sulfur	H/C	N/C	O/C	S/C
P8	8.16	38.40	53.44	75.90	5.18	1.35	6.66	2.75	0.8	0.02	0.1	0.01
KS-325	7.71	16.06	76.23	82.19	3.92	1.20	3.98	1.00	0.6	0.01	0.04	0.005
KS-121	18.83	9.49	71.68	72.47	2.78	1.05	3.88	0.99	0.5	0.01	0.03	0.01

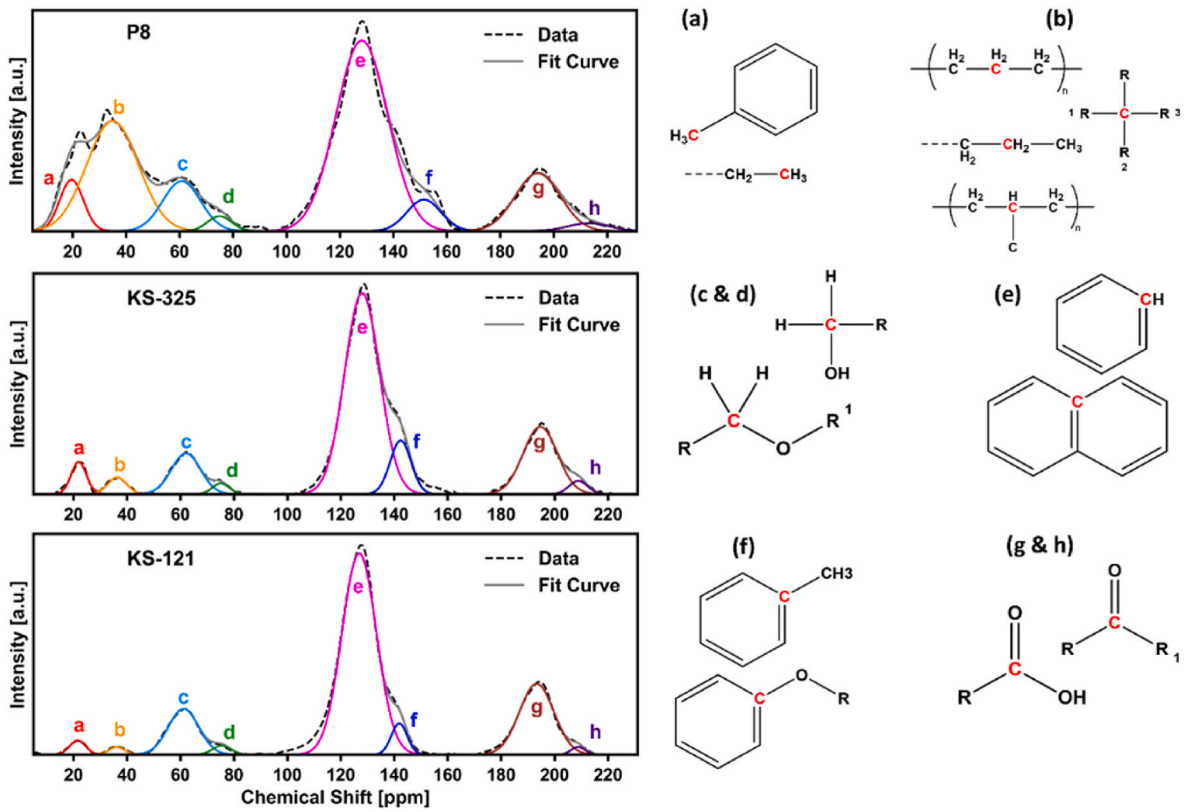


Fig. 3. (Left Panel) Solid state ^{13}C NMR spectra of the coals. The right panel shows the carbon atoms (colored in red) in the chemical environment responsible for the peaks (labeled a - h). The structural parameters assigned to the peaks are: (a) f_{al}^{*}, (b) f_{al}^{H}, (c & d) f_{al}^{O}, (e) $f_a^H + f_a^B + f_a^S$, (f) $f_a^S + f_a^P$, (g & h) f_a^C.

Table 3
Specific carbon structural distribution of the coal samples from solid^{13}C NMR. The chemical shift range associated to each structural parameter is provided in ppm (e.g., 0–16 ppm, 25–36 ppm, etc.).

Coal	Relative Ratios of Aliphatic Carbon Groups									
			f_{al}^{*}			f_{al}^{H}			f_{al}^{O}	f_{al}
	0–16	16–25	0–25	25–36	36–51	25–51	51–75	75–90	51–90	0–90
P8	0.03	0.08	0.11	0.11	0.12	0.23	0.09	0.01	0.10	0.44
KS-325	0.01	0.03	0.04	0.01	0.01	0.02	0.09	0.01	0.10	0.16
KS-121	0.01	0.02	0.03	0.01	0.01	0.02	0.10	0.01	0.11	0.17
	Relative Ratios of Aromatic Carbon Groups									
	f_a^H	f_a^B	f_a^S	f_a^P	f_a^N	f_a^C	f_a			
	90–129	129–137	137–150	150–165	129–165	165–188	188–220	165–220	90–220	
P8	0.23	0.12	0.09	0.03	0.24	0.06	0.03	0.09	0.56	
KS-325	0.31	0.20	0.12	0.01	0.33	0.02	0.18	0.20	0.84	
KS-121	0.38	0.18	0.07	–	0.25	0.02	0.18	0.20	0.83	
	Relative Ratios of the Additional Structural Parameters									
	X_b	n_{cl}	σ	I_b	BL	SC	W_c [g/mol]			
P8	0.26	13	2.9	0.3	0.7	2.2	388.1			
KS-325	0.31	16	2.5	0.8	1.9	0.6	278.3			
KS-121	0.28	14	1.2	0.6	0.7	0.5	279.5			

of aromatic substitution. Among these peaks, peak "a," associated with ortho-aromatic substitutions (four adjacent aromatic C–H), is the dominant environment across all the coals. Peaks "b" and "c" correspond to trisubstituted aromatic groups, while peaks "d" and "g" are attributed to isolated C–H groups, although peak d (~812 cm^{-1}) has also been suggested by Kuehn et al. [73] to be associated with vibrations from two adjacent aromatic C–H. Concerning the weak para-aromatic substitution (peak e), P8 and KS-121 exhibit relative ratios of 2 and 3 %, respectively, while no peak was found for KS-325. Only P8 shows a peak around 855 cm^{-1} (peak f), assigned to tetrasubstituted aromatic groups, at ~10 %.

Except for peak "d" in Row II of Fig. 4, attributed to the deformation vibrations of CH_2 and CH_3, the remaining peaks are associated with oxygen-containing functional groups. Notably, the relative ratios observed for the bituminous coals are almost evenly distributed (refer to Table 4). KS-121 exhibits a dominant peak "a" (~54%), assigned to aliphatic ethers and alcohols, with only 8 % relative ratio for C–O

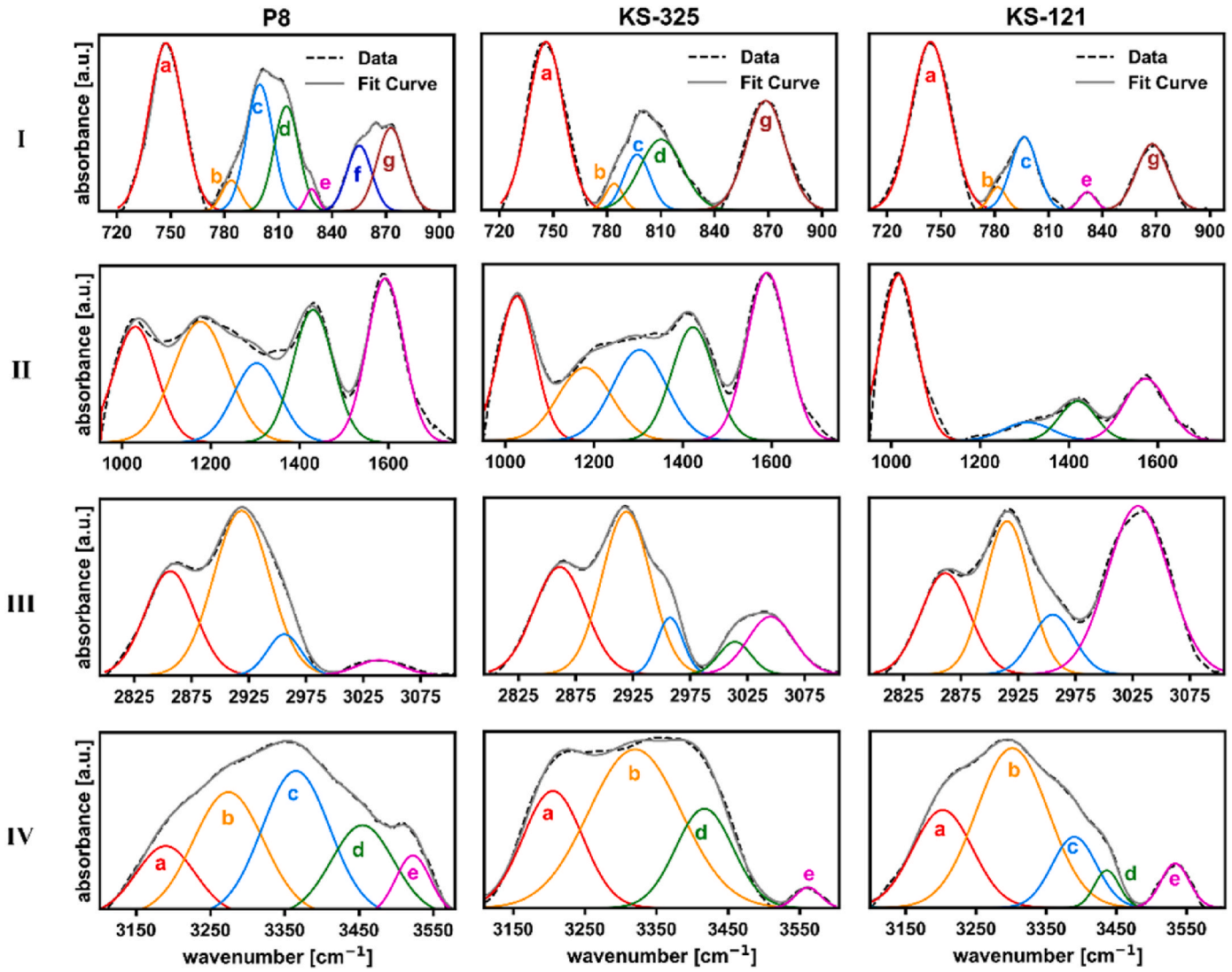


Fig. 4. ATR-FTIR spectra of the coal samples in the left, center, and right panels, revealing distinct chemical environments across key frequency regions. The first, second, third, and fourth rows, denoted as I, II, III, and IV, represent peaks within the ranges 700 to 900 cm^{-1}, 1000–1800 cm^{-1}, 2800–3100 cm^{-1}, and 3100–3600 cm^{-1}, respectively. The fitted curves are color-coded and the peaks are labeled "a – g" using the assigned colors in rows I and IV. (A colour version of this figure can be viewed online.)

stretching and O–H bending in phenolic structures (peak c). In some studies [26,29], bands around 1600 cm^{-1} have been associated with aromatic ring stretching mode. However, an alternative interpretation links this peak to highly conjugated hydrogen-bonded carbonyl groups [65,66], and from Table 4, the average position of peak "e" for all three coals is ~1590 cm^{-1} (close to 1600 cm^{-1}). However, analysis of the peak areas supports the assignment to highly conjugated hydrogen-bonded carbonyl groups. This is evidenced by the expectation that peaks attributed to aromatic ring in high-rank coals, like low volatile bituminous and semi-anthracites, should display larger areas. However, this trend is not observed in KS-325 and KS-121, which demonstrates a relative ratio of only 27 and 24 % for peak "e".

Row III in Fig. 4 is linked to C–H vibrations in both aliphatic and aromatic groups. The aliphatic C–H stretching bands fall between 2800 and 3000 cm^{-1} (peaks a, b, and c), while aromatic C–H groups are found between 3000 and 3100 cm^{-1} (peaks d and e). The position of peak "a" (P8: 2857 cm^{-1}, KS-325: 2861 cm^{-1}, KS-121: 2862 cm^{-1}) makes it challenging to assign it to either symmetric vibrations of CH_2 or CH_3 [29, 73]. However, since peak "a" contributes only 20 % in P8 and 30 % in KS-325 and KS-121, it is disregarded when constructing the coal models. P8 and KS-325 display a higher ratio for CH_2 bridges (peak b) at ~ 50 and 42%, respectively, while this peak is only about 28% in KS-121. Peak c, with a mean position at ~2956 cm^{-1}, is attributed to aromatic terminal $-CH_3$, although CH_2 in some hydrocarbons also absorbs around 2950 cm^{-1} [29]. Peak d and e are associated with aromatic C–H vibrations. Peak d is fitted separately at 3014 cm^{-1} for KS-325 because highly strained aliphatic rings also have bands around that frequency (~3000 cm^{-1}) [29]. However, experimental evidence for such vibrations in coal is lacking. Later in the discussion section, evidence from the vibrational density of states calculations will confirm that Peak d is indeed due to aromatic *C*–H stretching mode. Row IV (3100–3600 cm^{-1}) is associated with hydrogen bonding groups in the coals. The primary feature for all the coals is hydrogen bonded to oxygen functional groups (OH–O–R) in peaks b and c, while the bituminous coals display a higher peak area for hydrogen bonding between OH groups (OH–OH) compared to KS-121 in peak d.

3.4. *X-ray photoelectron spectroscopy*

Fig. 5 displays the chemical species obtained from fitting the C 1s, N 1s, O 1s, and S 2p XPS signals for the coal samples, with the peak binding energy and relative area percentage ratios summarized in Table 5.

Table 4
Relative percentage ratio of the peak area for regions in the ATR-FTIR spectra. The average peak positions calculated using all three coals are also presented.

ATR-FTIR Region [cm^{-1}]	Coal Samples	Relative Percentage Ratio of Fitted ATR-FTIR Peaks [%]						
		a	b	c	d	e	f	g
700–900 [1]	P8	35.2	3.5	20.3	15.8	1.6	9.7	13.9
	KS-325	39.7	2.9	9.6	19.9	–	–	27.9
	KS-121	58.3	3.4	17.3	–	2.6	–	18.4
Average Peak Position [cm^{-1}]		745.9	783.2	797.6	812.5	830.2	855.2	869.8
1000–1800 [2]	P8	18.6	25.0	13.7	20.1	22.6		
	KS-325	20.6	15.3	19.0	18.6	26.5		
	KS-121	54.3	–	8.0	13.2	24.4		
Average Peak Position [cm^{-1}]		1024.4	1178.2	1304.5	1424.5	1590.8		
2800–3100 [3]	P8	31.7	55.8	8.4	–	4.2		
	KS-325	30.0	42.4	7.8	5.6	14.2		
	KS-121	20.0	27.9	10.2	–	41.8		
Average Peak Position [cm^{-1}]		2859.8	2917.9	2956.3	3014.0	3038.0		
3100–3600 [4]	P8	13.5	28.6	32.4	18.6	6.9		
	KS-325	25.1	52.7	–	20.4	1.8		
	KS-121	25.8	50.7	14.0	4.1	5.5		
Average Peak Position [cm^{-1}]		3199.2	3298.7	3377.6	3436.0	3539.3		

[1] Aromatic CH out-of-plane bending modes. (a) 1,2-substituted aromatic. (b & c) trisubstituted aromatic. (d) Isolated CH and/or two neighboring CH. (e) Weak 1,4-substituted aromatic. (f) Tetrasubstituted aromatic. (g) Isolated aromatic CH.

[2] (a) aliphatic ethers & alcohols. (b & c) C–O stretch and OH bend in phenoxy & ethers. (d) Deformation vibration of CH_2 & CH_3. (e) Highly conjugated hydrogen bonded C═O.

[3] (a) symmetric stretching modes of CH_2 and/or CH_3. (b) Telescopic vibration of asymmetric CH_2 bridges only (i.e., not CH_2 hydroaromatic). (c) CH_3 attached directly to aromatic rings (although CH_2 in certain hydroaromatic structures also absorb at ~ 2950 cm^{-1}). (d) Highly strained aliphatic rings or (more likely) aromatic CH vibrations. (e) Aromatic CH vibrations.

[4] (a) cyclic H-bond. (b & c) OH–O H-bond. (d) H-bonded OH groups (i.e., OH–OH hydrogen bond). (e) OH–π H-bond

Deconvoluting the narrow scan of C 1s spectra facilitates calculating the contributions from C–C and C–H components in coal [31]. Additional C 1s peaks, indicative of oxidized carbon states, such as C–O and C═O, are also discerned [33]. Notably, the π–π* transition, characterized by a distinctive shake-up line (satellite peak), is observed in KS-121 and KS-325, originating from carbon in aromatic compounds excited by the departing photoelectrons [74].

For easy reference, Fig. S2 illustrates representations of some nitrogen and sulfur functional groups used in this study. Deconvolution of the N 1s signal enables the determination of pyridinic, pyrrolic, quaternary, and oxidized nitrogen species in coal [31,75]. Oxygen species (C═O, –OH, and –COOH) are derived from the O 1s signal, in addition to the analysis based on the C 1s signal. The S 2p signal is resolved into peaks corresponding to aliphatic sulfur, thiophenes, sulfones, sulfonates, and sulfates [76,77]. From the S 2p XPS signal, the organic sulfur groups commonly observed in all three coals were aliphatic sulfur (likely in form of sulfides) and thiophene, while the inorganic sulfur exists as sulfate. The bituminous coals, P8 and KS-325, show peaks for sulfonates with relative content of 14 and 19 %, while KS-121 contains a small proportion of sulfone (4 %).

4. Discussion

4.1. Characterization of the coal samples

Volatile matter yield in dry mineral-matter-free basis (V_{dmmf}), a key indicator of coal rank [55], is typically low in high-rank coals. The V_{dmmf} obtained for P8 (46.63 %), KS-325 (19.10 %), and KS-121 (14.78 %) align with classifications as high volatile bituminous, low volatile bituminous, and semi-anthracite coals, respectively [56]. Frazer's fuel ratio (fixed carbon/volatile matter) also yields similar classification for the coals [78]. In dry ash-free (daf) basis [79], the fuel ratio for P8, KS-325, and KS-121 is 1.4, 4.7, and 7.6 respectively. P8 is additionally classified as high volatile B (hvBb), based on its gross calorific (heating) value of 13,555 Btu/lb, in the moist mineral-matter-free (mmmf) basis [56]. From Table 2, despite higher carbon content in bituminous KS-325 compared to semi-anthracite KS-121, the hydrogen-to-carbon (H/C) atomic ratio suggests a higher carbon-to-hydrogen ratio in KS-121 than in KS-325. A similar trend is observed for the oxygen-to-carbon (O/C) ratio.

4.2. Functional groups in the coal samples

The chemical structure of coal can be estimated by analyzing the relative amounts of carbon, oxygen, nitrogen, and sulfur types derived from the experimental analyses in this study. All experimental data has been rounded to integer values and applied to a standard coal structure based on 100 carbon atoms for clarity. The analysis primarily focuses on NMR data for carbon and oxygen, complemented by XPS data for nitrogen and sulfur. FTIR data provides additional insights into potential hydrogen bonding environments, oxygen functional groups, and aromatic ring substitutions. Ultimate analysis results ensure that atomic ratios for the coals remain consistent. Later in this section, the vibrational band assignments from the FTIR analysis emerge as useful for post-validation of the model. They aid in understanding and comparing vibrational signatures observed in the model with the vibrational density of states and its associated inverse participation ratio.

4.2.1. Carbon skeletal structure of the coals

Assignments of the carbon groups, delineated by the ^{13}C NMR spectra in Fig. 3, draw primarily from the work of Solum et al. [24] and Yoshida et al. [58]. The analysis of carbon structures is categorized into aliphatic and aromatic groups.

4.2.1.1. Aliphatic long-chain and cyclic structures. The 0–16 ppm region in Table 3 indicates 3 terminal methyl carbon (–CH_3) per 100 carbon atoms in P8, and 1 each for KS-325 and KS-121. Consequently, P8 likely contains 3 terminal ethyl carbons (–H_2C–CH_3) per 100 carbon atoms, while KS-325 and KS-121 contain 1 each. The 25–36 ppm spectral region is considered to reflect only CH_2 carbons, and not CH carbons due to the lower occurrence of aliphatic CH groups in coal structure [29]. Regarding aliphatic chains, the broad peak obtained from ^{13}C NMR data does not precisely indicate the length of the aliphatic –CH_2– chains within the 25–36 ppm region. For example, the prediction of eleven CH_2 carbons per 100 carbon atoms for P8 does not directly imply the presence of 1 undecane bridge every 100-carbon atom (more details in Figure S4b). However, the presence of aliphatic CH_2 bridges can be inferred from the FTIR data since the "b" peak around 2918 cm^{-1} in Row

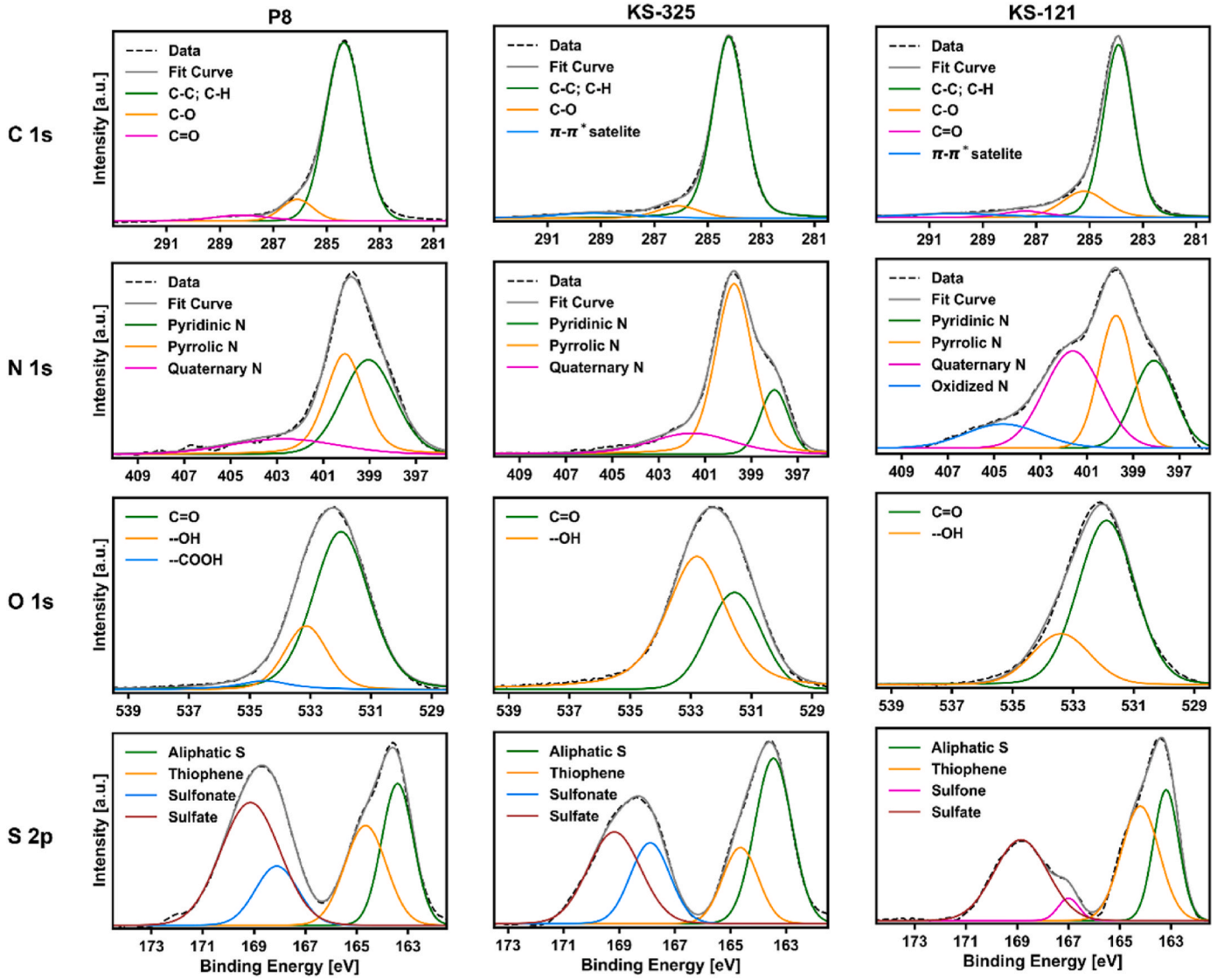


Fig. 5. Cabon 1s, Nitrogen 1s, and Sulfur 2s species obtained from x-ray photoelectron spectroscopy analysis. (A colour version of this figure can be viewed online.)

Table 5
XPS binding energy (in eV) peak position and the relative area percentage ratio.

XPS Signal	Chemical Species	Coal Samples					
		P8		KS-325		KS-121	
		Energy	% Ratio	Energy	% Ratio	Energy	% Ratio
C 1s	C–C; C–H	284.3	87.1	284.2	85.4	283.9	72.1
	C–O	286.1	9.0	286.1	7.8	285.2	16.5
	C═O	288.2	3.9	–	–	287.4	4.5
	π – π*	–	–	289.2	6.9	290.1	6.9
N 1s	Pyridinic N	398.7	45.6	398.0	16.8	398.1	23.4
	Pyrrolic N	400.1	39.3	399.7	65.6	399.7	29.3
	Quaternary N	402.7	15.2	401.6	17.6	401.6	35.8
	Oxidized N	–	–	–	–	404.6	11.5
O 1s	C═O	531.9	72.3	531.6	33.7	531.9	76.3
	–OH	533.1	22.4	532.8	66.3	533.4	23.7
	–COOH	534.5	5.2	–	–	–	–
S 2p	Aliphatic S	163.4	24.4	163.4	36.4	163.2	26.6
	Thiophene	164.7	22.5	164.6	16.1	164.2	34.9
	Sulfone	–	–	–	–	167.0	3.8
	Sulfonate	168.1	13.7	167.9	18.9	–	–
	Sulfate	169.1	39.4	169.2	28.6	168.9	34.8

III (Fig. 4) indicates contribution from methylene bridges only. Bellamy suggested that methyl groups directly attached to aromatic groups also exhibit strong bands near 2920 cm^{-1} [80], but the small number of terminal methyl groups observed from ^{13}C NMR (16–25 ppm in Table 3) implies that the predominant contributor should be methylene bridges in the coals.

While the FTIR data confirms the presence of methylene bridges, the challenge of determining the length (n) of the methylene chains, $(CH_2)_n$, persists. To address this, we utilized the lower limit for the $(CH_2)_n$ content in coals of different ranks, reported by Calkins [81,82]. We assigned value for n between 1 and 3 for P8, and 1 or 2 for KS-325 and KS-121, subject to fitting of the experimental NMR spectra. Notably, one CH_2 is also predicted for KS-325 and KS-121 from the NMR data. The NMR Peak "c" (51–75 ppm) is attributed to oxygenated $–CH_2–$ groups (51–75 ppm), and is proposed to marginally extend the average aliphatic chain length by acting as crosslinks between aromatic clusters [16–18]. The limited presence of aliphatic ether groups (C–O–C) in the 75–90 ppm region (refer to Table 3) across all three coals indicates minimal to no aliphatic ether linkages. This is consistent with NMR reports for aliphatic ether bridges in similar coal ranks from other studies [18,27,83], reinforcing the likelihood that oxygenated $–CH_2–$ primarily functions as crosslinks between aromatic (Ar) clusters, forming $Ar–O–CH_2–Ar$ aromatic bridges. Consequently, the aliphatic bridges considered for the coal models in this study include $–CH_2–$, $–(CH_2)_2–$, $–(CH_2)_3–$, $–O–CH_2–$.

Alicyclic or saturated hydroaromatic carbon in coal function as connecting bridges (loops) between unsaturated polyaromatic structures. Fig. 6 illustrates commonly identified polycyclic aromatic hydrocarbons (PAH) found in coal and used in this study. Dihydroanthracene and dihydrophenanthrene are the most frequently proposed PAHs for most coal structures [9,24]. From Table 3, the number of bridges and loops (BL) calculated for KS-121 and P8 is one per aromatic cluster (indicative of structures like dihydroanthracene or dihydrophenanthrene). KS-325, on the other hand, averages two per

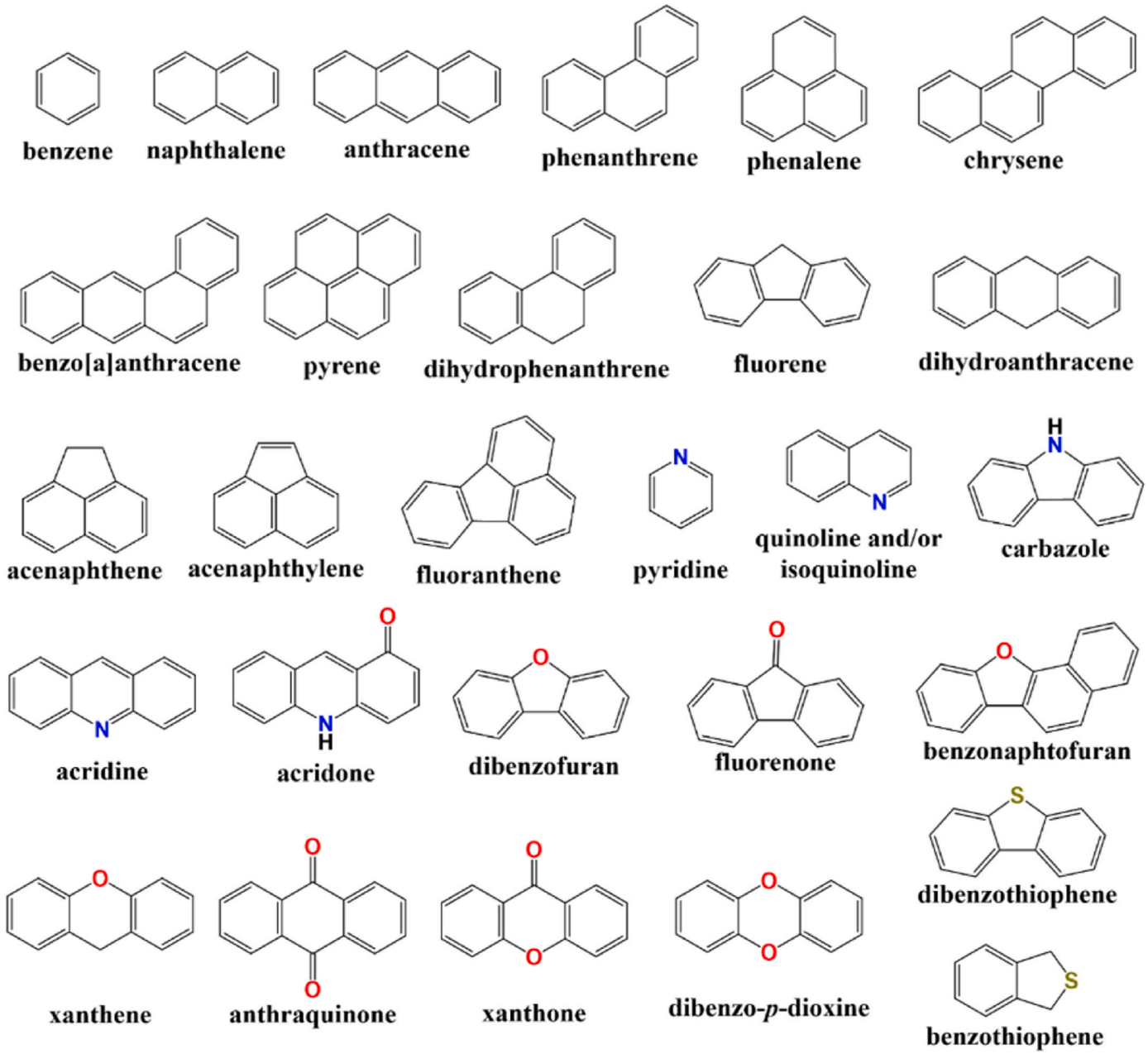


Fig. 6. Commonly identified polycyclic aromatic hydrocarbons (PAH) in coal and considered in this study. (A colour version of this figure can be viewed online.)

cluster. It is worth noting that, I_b, used to separate cluster coordination (σ) into bridges/loops and side chains (*SC*), can introduce errors as high as 25 % in the coal model. This is due to I_b considering all side chains as single methyl groups, neglecting other terminating groups such as OH and terminal methyl chains.

4.2.1.2. Aromatic carbon groups. In Table 3, the n_{cl} structural parameter values for P8 (13), KS-325 (16), and KS-121 (14) suggest a prevalence of linearly catenated aromatic carbon, aligning with the $C_{4m+2}H_{2m+4}$ family of polyaromatic hydrocarbons structures. Therefore, the average number of fused aromatic rings (rounded to an integer) in P8, KS-325, and KS-121 is 3, 4, and 3, respectively. This indicates that, along with benzene-like isolated rings, phenanthrene, anthracene, and possibly chrysene (for KS-325) are the primary PAH components to be considered (refer to PAH structures in Fig. 6). Circular catenated PAH is also plausible; PAH like phenalene and pyrene are considered in the coals, as these structures align with the calculated number of fused aromatic rings. Pyrene is likely present in KS-325, considering its n_{cl} value is precisely 16 carbon atoms, equivalent to a cluster of 4 benzene-like rings.

The distribution of aromatic substitutions, extracted from the 700–900 cm^{-1} FTIR spectra region, is also useful in determining the type of PAH suitable for representation, and *vice versa*. For example, anthracene can be chosen to represent isolated CH groups, although isolated CH groups can also result from substitution of other functional groups into adjacent carbon atoms in an aromatic ring. However, the latter can be checked using XPS, FTIR, or the elemental ratio from ultimate analysis (for non-carbon functional groups). Apart from using PAHs or terminal ethyl, aromatic terminal $–CH_3$ in the 16–25 ppm NMR spectral region (8 for P8, 3 for KS-325, and 2 for KS-121) can also facilitate required aromatic substitutions. However, all these considerations are contingent upon fitting of the experimental NMR spectra.

From the C 1s XPS data, the subtle π-π* satellite peak in KS-325 and KS-121 indicates the occurrence of sp^2 carbon in these coals, reminiscent of graphite. While quantitatively determining sp^2 carbon solely based on the satellite state might be ambitious, a qualitative comparison reveals the presence of the π-π* in KS-325 and KS-121, but not in P8. This indicates that KS-325 and KS-121 have a higher number of fused aromatic rings compared to P8.

4.2.2. Oxygen functional groups

Balancing oxygen counts from ^{13}C NMR parameters (f_{al}^O, f_a^P, and f_{al}^C) with the O/C ratio from ultimate analysis proves challenging. In Table 2, the O/C ratio reveals low oxygen counts for all the coals: 5, 4, and 3 oxygen atoms per 100 carbon atoms for P8, KS-325, and KS-121, respectively. This contrasts with the higher oxygen content derived from NMR by summing values for f_{al}^O, f_a^P, and f_{al}^C. While acknowledging potential double counting of oxygen in NMR (e.g., methoxy groups attached to aromatic rings are counted in both f_{al}^O and f_a^P), the low ratio (3 and 1 in P8, KS-325, and KS-121, respectively) or absence (in KS-121) of phenol or phenolic ether indicates minimal impact on the overall oxygen concentration. Given that the elemental oxygen concentration is derived through an elimination procedure in ultimate analysis [52], the oxygen concentration based on NMR, supported by FTIR and XPS, is considered for coal structure predictions.

Unlike P8 and KS-325, assigning phenolic groups to the semi-anthracite KS-121 coal is improbable. This is supported by the absence of an NMR peak for phenolic group in the 150–165 ppm region, further supported by the absence of a clear "b" and "c" peak in the FTIR absorption band in the 1000–1800 cm^{-1} range (Row II in Fig. 4). The low concentration of the phenolic group in KS-121 validates the use of I_b in estimating the number of sidechains available in the semi-anthracite coal, which is particularly relevant as the ratio of aromatic –OH terminations and crosslinking reactions is predominantly linked to f_a^P [58].

In Fig. 3 and Table 3, the NMR spectra of KS-325 and KS-121 show minimal curve area between 165 and 188 ppm, indicating a low presence of carboxyl carbon (2 in 100 carbon atoms) in both coals. This implies that the carbonyl functional group is predominant. XPS analysis in Table 5 corroborates this, showing no carboxylic (–COOH) peak in the O 1s signal for KS-325 and KS-121. However, P8 exhibits 6 carboxyl carbons per 100 carbon atoms, with a discernible carboxyl peak in the O 1S signal. The consistent identification of carboxyl carbon through both XPS and NMR analyses for P8 confirms that carboxyl and carbonyl should be considered in the structural composition of P8 coal. These findings for P8 contradict Solomon's conclusions regarding the absence of carbonyl and carboxyl functional groups in their P8 model [18], which were based on FTIR data and CO_2 output from thermal decomposition.

4.2.3. Nitrogen and sulfur functional groups

NMR and FTIR provide insights into the carbon and oxygen groups, while XPS is preferred for inferring nitrogen and sulfur functional groups in coal. The N 1s scan in Table 5 indicates the presence of pyridinic, pyrrolic, and quaternary (or graphitic) nitrogen in all three coals. P8 has a lower proportion of quaternary nitrogen (15.2 %) compared to the other two nitrogen functional groups. Pyrrolic nitrogen dominates in KS-325 (65 %), with pyridinic and quaternary nitrogen close at 16.8 and 17.6 %, respectively. KS-121 shows an additional peak attributed to oxidized nitrogen at 11.5 %, while the other functional groups show relatively equal ratios.

From the S 2p XPS scan, the low sulfone content observed in KS-121, along with its absence in the bituminous coals, aligns with findings from Zhang et al. [76]. In coal modeling, key emphasis is placed on functional groups such as thiophene and aliphatic sulfur (predominantly sulfides and thiols) [17–20]. In KS-325, the aliphatic sulfur to thiophene ratio is approximately 2:1, whereas in P8 and KS-121, these functional groups are nearly equivalent. This suggests that when modeling P8 and KS-121, equal consideration should be given to both thiophene and aliphatic sulfur, while KS-325 leans towards a preference for aliphatic sulfur.

4.3. Coal models, DFT post-treatment, and characterization

While techniques like ultimate and proximate analyses, ^{13}C NMR, FTIR, and XPS provide valuable insights into coal structure, their inherent limitations, including approximation and bias from curve-fitting methods, must be acknowledged. Therefore, relying solely on a single coal model for analysis can lead to inaccurate estimations. A more robust approach, used in this study, involves constructing an ensemble of models aligned with experimental data, offering a statistically averaged portrayal for accurate coal characterization. To this end, four distinct models were constructed for each coal to conform with the experimental NMR data. Details of the models are presented in Table 6. The 3D structure files, in *.xyz* format, for all 12 models are provided as supplementary material. Additionally, the 2D structure files (in *.chxml*) can be accessed from our website [84]. For conciseness, subsequent discussions will focus exclusively on the first 3 models (for each coal), and where applicable, individual models will be identified using the code name provided in the second column of Table 6.

Fig. 7 illustrates the experimental NMR spectra alongside the calculated spectra and their corresponding structure for Model 1 (M1) of the coals. Similar representations for Models 2 (M2) and 3 (M3) are presented in Figure S6, S7, and S8 for P8, KS-325, and KS-121, respectively. Discrepancies between the carbonyl and carboxyl region of the experimental NMR spectra and the calculated spectra are evident across all models. This disparity arises because the experimental peak height reflects the actual molecular size of the coal, which is not entirely captured by the model due to finite-size effects. Despite this limitation, the qualitative shape of the peak remains consistent.

The H/C ratio calculated in the models closely mirrors the range obtained from ultimate analysis, typically within ±0.1 deviation. However, the O/C ratio derived from ultimate analysis proves too low to

Table 6
Details of the models constructed for P8, KS-325, and KS-121 coals.

Coal	Model Code Name[a)]	Model Properties		
		Chemical Composition	Molecular Weight [g/mol]	Density [g/cm^3][b)]
P8	P8_M1	$C_{207}H_{158}N_4O_{15}S_2$	3005.67	1.42
	P8_M2	$C_{206}H_{151}N_3O_{15}S_3$	3004.66	1.42
	P8_M3	$C_{239}H_{175}N_3O_{18}S_3$	3473.21	1.41
	P8_M4	$C_{184}H_{137}N_3O_{15}S_2$	2694.25	1.41
KS-325	KS325_M1	$C_{159}H_{104}N_2O_{17}S_2$	2378.70	1.40
	KS325_M2	$C_{175}H_{109}N_3O_{16}S_3$	2605.98	1.42
	KS325_M3	$C_{145}H_{88}N_4O_{11}S_3$	2158.50	1.40
	KS325_M4	$C_{170}H_{111}N_3O_{16}S_2$	2515.88	1.39
KS-121	KS325_M1	$C_{154}H_{94}N_2O_{13}S_2$	2244.57	1.43
	KS325_M2	$C_{180}H_{107}N_3O_{10}S_4$	2600.09	1.43
	KS325_M3	$C_{124}H_{78}N_2O_{11}S_3$	1868.17	1.44
	KS325_M4	$C_{149}H_{86}N_2O_{10}S_3$	2160.51	1.43

[a] M1, M2, M3, and M4 indicated the first, second, third, and fourth models constructed for the coals. All models are included in the supplementary material. Only the first 3 models are discussed in this study.
[b] Density of final structure post-DFT treatment.

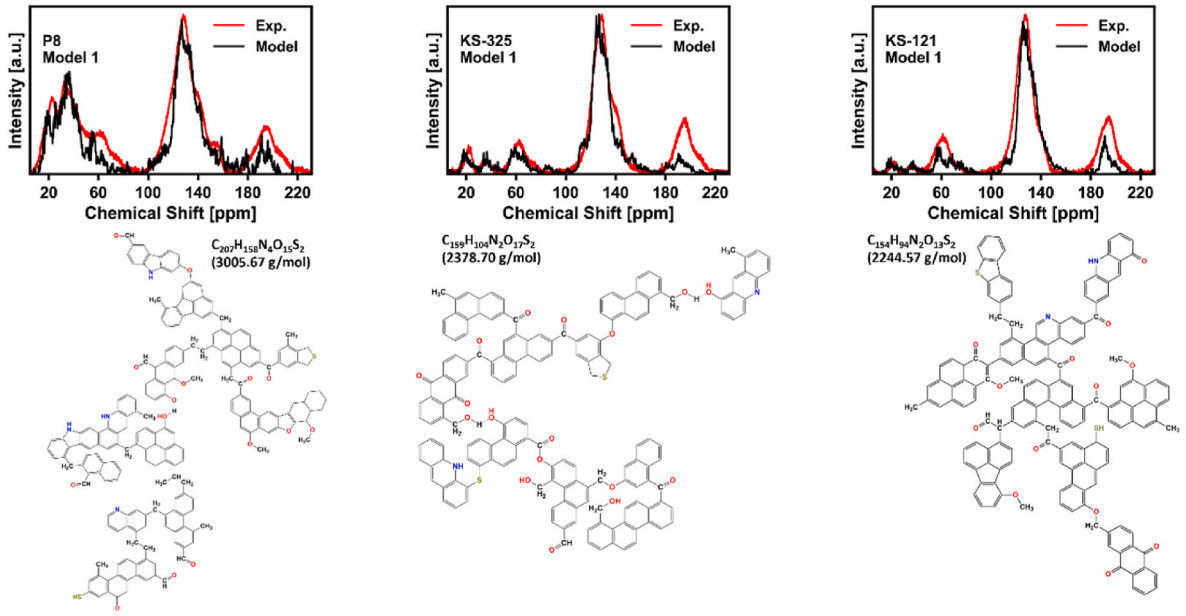


Fig. 7. Representative models constructed for P8, KS-325, and KS-121 coals. The experimental NMR spectra is plotted along with calculated NMR spectra for Model 1 of all the coals. Similar figures for Models 2 and 3 are available in the supplementary material. (A colour version of this figure can be viewed online.)

accurately match the NMR spectra. Therefore, the O/C ratio in the models was determined based on NMR and FTIR analysis. Nitrogen and sulfur functional groups were adjusted to align with XPS analysis conclusions. P8 models feature larger system sizes compared to KS-325 and KS-121, as shown in Table 6, due to different constraints during model construction. The NMR spectra for P8 exhibits a larger relative area for aliphatic groups, requiring more atoms to align with the experimental data. Conversely, KS-325 and KS-121 require more aromatic structures than aliphatic ones, resulting in relatively smaller system sizes for these models. The final densities of all the models are within range of typical coal density (1.1–1.8 g/cm^3) [85].

4.3.1. *Structural properties of the models*

To ensure the chemical realism of the coal models, validation was performed through DFT-based energy/structural minimization. Conjugate gradient energy minimization was employed to optimize atom position, bond length, bond angles, and cell shape of the models obtained from ReaxFF (henceforth referred to as "ReaxFF" models). This approach aims to minimize constraints during calculation, assuring that the DFT-relaxed models (henceforth referred to as "DFT" models) are truly optimized. To demonstrate the benefits of DFT post-treatment, we compared the structural attributes of the ReaxFF and DFT models by using total radial distribution function (RDF), bond length distribution (BLD), and bond angle distribution (BAD) computations [86]. Only the results obtained for KS121_M1 is presented in Fig. 8, for brevity, while similar comparisons for all 9 models are provided in Figure S9.

The RDF plot in Fig. 8a demonstrates consistent peak matching between the ReaxFF and DFT model, indicating no emergence of new peaks, post-DFT treatment. This pattern is consistent across the other 8 models, as illustrated in Figure S9. The absence of new peaks in the RDF, combined with the absence of bond-breaking or forming throughout the DFT relaxation, indicates that the atoms are arranged within energetically stable and realistic local environments in configuration space. However, the narrower RDF peaks in the DFT model indicate that the configuration of the ReaxFF model is not the optimal local minimum. The DFT model exhibits a more defined local structure compared to the ReaxFF model, as observed in the BLD and BAD plots in Fig. 8b and c, respectively, as well as in Figure S9, for the other models. In the BLD plot, the prominent peak "i" corresponds to C–H bond at ~1.09 Å [87, 88], while the small peak "ii" (~1.24 Å), which is less resolved in the

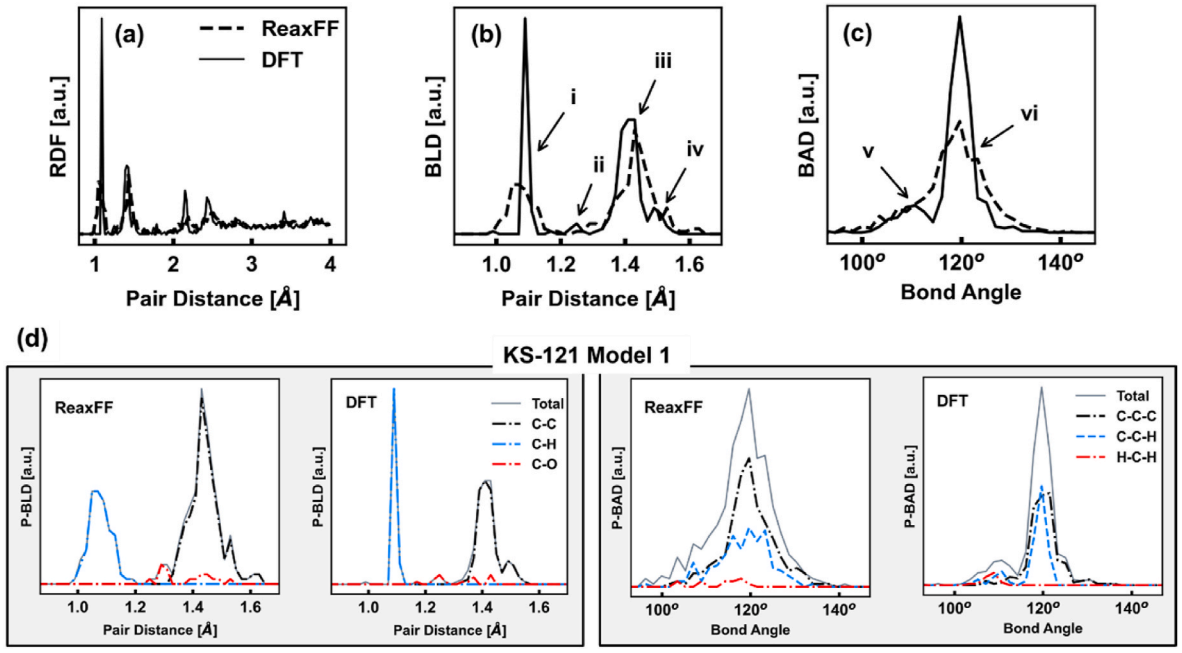


Fig. 8. Structure analysis of ReaxFF and DTF treatment of KS121_M1. (a), (b) and (c) depict the radial distribution function (RDF), bond length distribution (BLD) and bond angle distribution (BAD) for the KS121_M1, before and after DFT treatment. The atom-projected BLD (*P*-BLD) and BAD (*P*-BAD) of the model is illustrated in (d). (A colour version of this figure can be viewed online.)

ReaxFF model, represents C–O bonding [12]. Additionally, peaks "ii" (~1.41 Å) and "iv" (~1.5 Å) for C═C and C–C bonds [88,89], respectively, are more clearly distinguished in the DFT model. Similarly, the BAD peaks in Fig. 8c exhibit better resolution for the DFT model. The broad peak observed in the ReaxFF model is separated into 2 peaks in the DFT models, "iv" (~110°) and "v" (~120°), representing C–C–H and H–C–H bond angles, with C–C–C bond angle also contributing to peak "v" [90–92]. Further exploration of structural differences involved examining the BLD and BAD projections onto atomic species, as presented in the atom-projected bond length distribution (*P*-BLD) and the atom-projected bond angle distribution (*P*-BAD) plots in Fig. 8d. The ReaxFF model shows a broader distribution of bond lengths and angles,

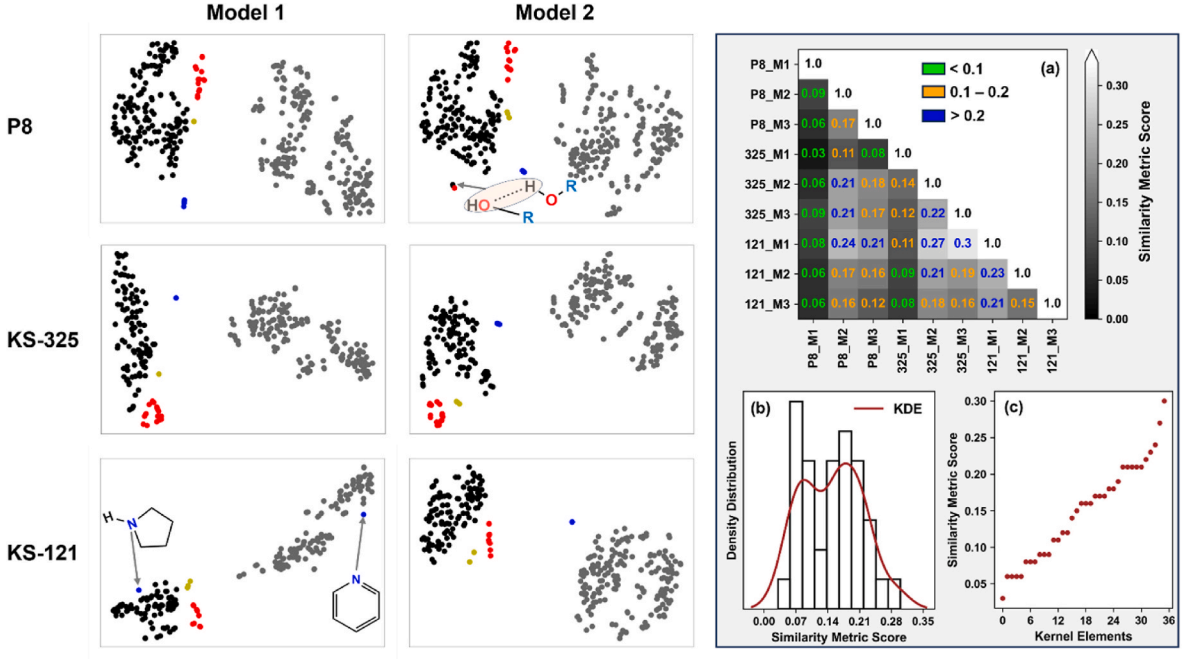


Fig. 9. (Left Panel) Configurational subtleties within coal structures are visualized through *t*-SNE, displaying results for Model 1 and 2 only. The $N\times$ 10, 350 SOAP descriptors are projected onto an $N \times 2$ *t*-SNE co-ordinate. Hydrogen, carbon, nitrogen, oxygen, and sulfur are colored black, gray, blue, red, and gold colors, respectively. Chemical structures annotated in P8_M2 and KS121_M1 illustrate the atomic environments corresponding to the indicated data points. [Right Panel] (a) Similarity metric scores of the coal models. Scores are color-coded: values < 0.1 in green, between 0.1 and 0.2 in yellow, and > 0.2 in blue. The "KS" in KS-121 and KS-325 is omitted for convenience. (b) A density plot with the associated Gaussian kernel density estimate (brown curve) of the similarity metric scores. (c) Frequency distribution of the elements of in the lower triangular $\overline{K}$ matrix (without the diagonal elements). Each brown dot represents one of 36 elements.

resulting in less defined peaks. This trend remained consistent across all nine models, as evidenced by the RDF, BLD, and BAD plots in Figure S9, along with the *P*-BLD and *P*-BAD plots for P8_M1, illustrated in Figure S10.

Therefore, it is evident that DFT post-treatment improves the ReaxFF models, ensuring chemically accurate representations crucial for scaling coal models [12,20,23,27]. Inaccuracies in local conformations can compromise credibility, emphasizing the need for precise structural features simulations. While ReaxFF offers energetically stable models, DFT treatment enhances structural fidelity, vital for intricate multi-species simulations like coal-plastic interactions for advanced materials [4,5,7], exploring electronic or electrochemical properties in coal-derived graphite [9,13], or utilizing coal in carbon-metal composites for ultra-conducting wires, akin to graphene applications [93–97]. In all cases, accurate coal structure and chemistry is indispensable.

4.3.2. Unicity analysis for the models using SOAP

The models constructed for a particular coal must exhibit distinct and distinguishable local structures to ensure completeness of any conclusions drawn from their utilization. Structural orthogonality among the models enables exploration of a wider range of interactions and properties that may not be discerned using models of similar representation (from supercells constructed from single representative model). The unicity of the DFT models was investigated using the SOAP technique. The SOAP descriptor for each model is an $N\times 10,350$ matrix, where N is the number of atoms in the model. Given the difficulty in visualizing or interpreting such high-dimensional data, we transformed the $N \times 10,350$ matrix into an $N \times 2$ matrix using the "t-distributed stochastic neighbor embedding (*t*-SNE)" algorithm, implemented using the Scikit-learn Python module [98]. *t*-SNE is a nonlinear dimensionality reduction technique used for visualizing high-dimensional data in lower dimensions [99]. Recently it has been applied to discern structural features in a million-atom model of silicon [100]. More on *t*-SNE is provided in Section S1.4 of the supplementary material.

The scatter plot in the left panel of Fig. 9 projects the 10,350 SOAP descriptor elements of each atom onto a 2D *t*-SNE coordinate, preserving maximum structural information. Hydrogen, carbon, nitrogen, oxygen, and sulfur atoms are represented by black, gray, blue, red, and gold colors, respectively. Due to the nature of the *t*-SNE algorithm, direct reconstruction of the original high-dimensional (10,350) co-ordinates is challenging [99]. Nevertheless, the distance between points in *t*-SNE space reflects their structural dissimilarity in SOAP space. The plot distinctly clusters for each atomic species, illustrating clear separation even among atoms of the same chemical species. This underscores the accuracy of SOAP implementation in discerning similarity within the local atomic environment of each model. A similar plot incorporating data for Model 3 of the coals is presented in Figure S11. Moreover, certain outliers were identified in the *t*-SNE representations (refer to insets in left panel of Fig. 9). In P8_M2, hydrogen and oxygen atoms involved in hydrogen bonding between two phenolic groups are in proximity. Also, the pyrrolic and pyridinic nitrogen in KS121_M1 are separated and positioned near the hydrogen and carbon clusters, respectively. This could stem from the limited number of nitrogen atoms in the model or related to the presence of hydrogen in pyrrolic structures, as well as the nitrogen substitution of carbon in a hexagonal ring (a major feature of KS-121) in pyridinic structures.

In the right panel of Fig. 9, similarity metric scores between pairs of models are depicted in a gray color scale (Fig. 9a). Only the unique elements, derived from the 9×9 $\overline{K}$ matrix (using Equation (7)), are displayed. As each model is self-similar, the diagonal elements each have a score of 1 (the upper limit of $\overline{K}$) and are not considered in subsequent analysis. We annotate and color-code the similarity metric scores: green for scores less than 0.1, yellow for scores between 0.1 and 0.2, blue for scores above 0.2. The maximum metric score is 0.3, hence the gray scale extends only to 0.34. In Fig. 9b, the density plot, along with the Gaussian kernel estimate (KDE), reveals a bimodal distribution for the 36 unique elements of $\overline{K}$, with primary and secondary modes centered around at ~0.18 and ~0.09, respectively. The first, second, and third quartiles from the metric score distribution are 0.09, 0.16, and 0.21, respectively. Furthermore, the scatter plot in Fig. 9c illustrates the frequency distribution of the similarity metric scores for the unique $\overline{K}$ elements in ascending order. The first dot, representing the kernel element with a score of 0.03, highlights the most distinct local environment between P8_M1 and KS325_M1. Conversely, the last dot at 0.30 represents the metric score between KS325_M3 and KS121_M1 – indicating the most similar local environments. The number of dots at any position in the plane indicates the number of $\overline{K}$ (kernel) elements having that metric score.

Considering models from the same rank, the maximum similarity score obtained for P8, KS-325 and KS-121 is 0.17, 0.18, and 0.23, respectively. The low statistical values obtained, the substantial deviation of the similarity metric scores from the maximum value of 1, along with the evident structural dissimilarity in the 2D *t*-SNE plot, strongly suggests that the models are unique and simulation-ready.

4.3.3. Vibrational signatures and specific heat capacity

We extend the DFT analysis to computing the vibrational density of states (VDoS) of the models, as it provides accurate and fundamental information about the low temperature vibrational excitations. It also serves as a metric for comparing simulated vibrational normal mode frequencies with experimentally obtainable data from FTIR or inelastic neutron scattering (INS) analysis. The upper-right panel of Fig. 10 shows the VDoS plot for KS121_M1, specifically the 1680–2800 cm^{-1} region where a phonon energy gap is observed. The full VDoS for KS121_M1, with frequency ranging between 0 and 3800 cm^{-1}, is presented in Figure S12. While direct comparison between the VDoS and FTIR or INS data are intricate, the spectra from all three methods exhibit remarkable similarities in their shape profiles. Notably, the absence of FTIR peaks or band assignments in the 1680–2800 cm^{-1} frequency region, consistent with the findings in this study and others [20,27,29,30,49,101], highlights a pattern that is also reflected in the dynamical structure factor, $S(Q,\omega)$, obtained from INS analysis of different coal ranks [49].

The VIPR (refer to Equation (4)) was employed to further characterize the vibrational normal modes. As a reminder, VIPR $\rightarrow 1/N$ indicates that several atoms contribute to the normal mode and VIPR $\rightarrow 1$ indicates fewer atoms contribute to that eigen frequency, signifying a highly localized mode. Lower frequency modes (<1680 cm^{-1}) are distributed over carbon atoms, while higher frequencies above ~2950 cm^{-1}, characterized by high VIPR (>0.6), are predominantly linked to hydrogen atoms in polyaromatic structures. Near the energy gap (1680–2800 cm^{-1}), six localized modes labeled "1–6" are discerned. The chemical environment contributing to these vibrational normal modes are illustrated in the upper-left panel in Fig. 10. Apart from mode "3", which is located around 2612 cm^{-1} and associated with a stretching vibration of the hydrogen atom in a phenyl-mercaptan, the other modes exhibit vibrational characteristics consistent with the FTIR band assignment in this study and by Painter et al. [29].

Mode "1" (~1667 cm^{-1}) and mode "2" (~1717 cm^{-1}) localize on a different C═O bonds in a carboxyl and aldehyde functional group, respectively. Mode "4" (~2896 cm^{-1}) is associated with a hydrogen atom in a terminal aldehyde. Mode "5" (~2919 cm^{-1}) corresponds to a hydrogen atom in a bridging CH_2, and mode "6" (~2934 cm^{-1}) is localized on a hydrogen atom connected to aliphatic CH bridge. Notably, these frequencies fall within the FTIR band assignment ranges for the corresponding coal functional groups [29,49,60,101]. An animation (gapModes.mp4) illustrating all six localized gap modes is provided in the supplementary material. The animations are also available on our website [84].

Additional vibrational normal modes, corresponding to FTIR band assignments, were identified. For example, pure bending and stretching

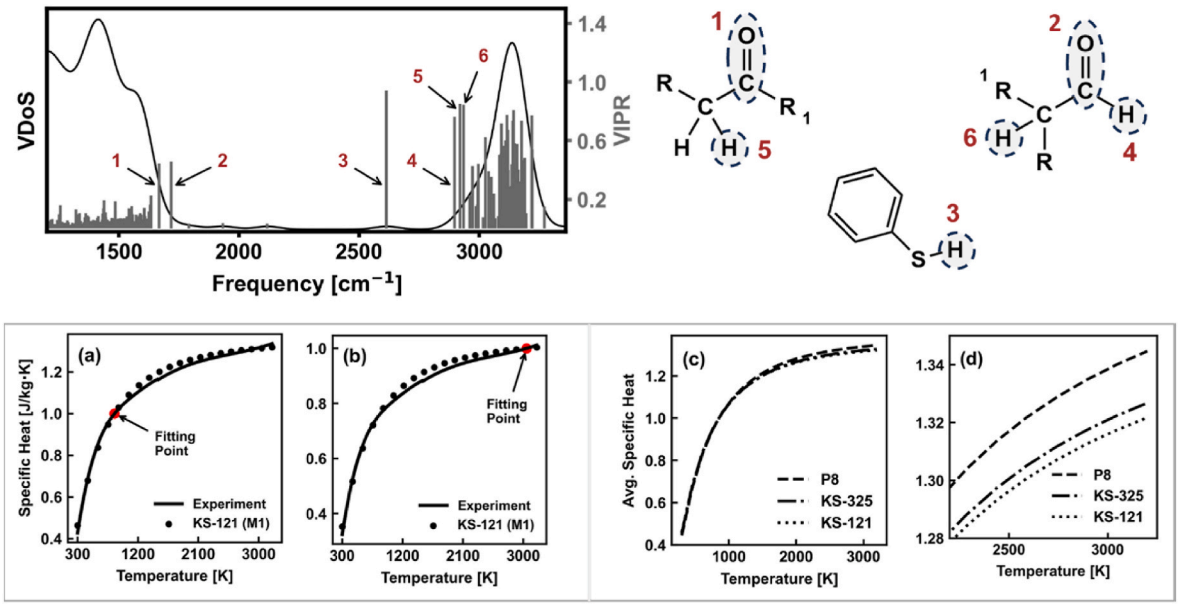


Fig. 10. [Upper Panel] The VDoS and VIPR for KS121_M1 are presented. Chemical environments contributing to the localized vibrational states in the gap region (annotated 1 to 6) are highlighted with dashed circles. [Lower Panel] Specific heat capacity as a function of temperature from 300 to 3200 K. The calculated specific heat for KS121_M1 is compared with experimental data [102]. The comparison includes (a) points in the curve region with high gradients and (b) their maximum values (approximating the Dulong-Petit limit at "high temperatures"). These points are annotated in the figure. Additionally, (c) compares the specific heat capacity obtained by averaging the values obtained for models constructed for each coal, and (d) zooms in on their differences at elevated temperatures. (A colour version of this figure can be viewed online.)

vibration normal modes in aromatic carbon rings were identified at ~426 and ~1600 cm^{-1} [26], respectively. This vibrational mode has also been observed in amorphous graphite [12,91]. Out-of-plane vibrations for two neighboring hydrogen atoms and CH in polyaromatic hydrocarbons were observed at ~819, and ~884 cm^{-1}, respectively [29]. Breathing modes for CH_3 attached to aromatic rings were observed at 2970 and at 2994 cm^{-1}. This agrees with our assignment for peak c in the 2800–3000 cm^{-1} FTIR band region in Fig. 4 and Table 4. Correspondingly, the ambiguity surrounding peak "d" in the same band region, regarding whether it originates from strained aliphatic rings or aromatic C–H stretching modes (discussed in Section 3.3), was resolved by the vibrational modes observed around 3115 cm^{-1}. This confirmed that the high frequency vibrations (>3000 cm^{-1}) indeed stem from aromatic C–H stretching. These additional normal modes are displayed in an animated video (otherModes.mp4), also provided in the supplementary material. The shape of the VDoS is consistent across all the models, regardless of rank. For comparison, the VDoS and VIPR for Model 3 for all the coals are presented in Figure S13. Notably, one new localized *C*–H mode was observed in KS325_M3 at ~2798.4 cm^{-1}, for a carbon atom bonded to the nitrogen in an acridine PAH (see PAH structures in Fig. 6).

On a different note of general relevance to disordered materials, the vibration typically observed in FTIR experiments at ~860 cm^{-1} is commonly assigned as pure vibrations involving isolated aromatic CH [29,60]. Our observation revealed mode mixing of vibrations form isolated hydrogen atoms and deformation vibrations in aromatic C═C, thus indicating that care is needed in mode identification, as mixing between modes is important in disordered (amorphous) systems [103].

The specific heat capacity provides an additional criterion for evaluating computational models against real-world materials. In the lower panel of Fig. 10, the computed specific heat is compared with experimental specific heat data obtained by one of the authors (R. Olson III) [102]. The computed and experimental specific heat capacity profiles reveal a significant coincidence across the temperature spectrum ranging from 300 to 3200 K. To validate this agreement, the plots were normalized based on two distinct reference points: A region characterized by the steepest gradient (Fig. 10a) and at the highest temperature end (Fig. 10b) – which crudely approximates the Dulong-Petit limit [104]. The specific heat capacities for models of P8 and KS-325 were calculated, yet they could not be compared to experimental data due to its unavailability. However, in Fig. 10c, the specific heat capacity values for all three models within the same rank were averaged and normalized to the fitting point region used in Fig. 10a for comparison. At high temperatures, discernible differences in the average specific heat capacity values across coal ranks are evident, with P8 exhibiting the highest values, followed by KS-235 and KS-121 (refer to Fig. 10d). Conversely, at lower temperatures, the specific heat capacities among all models closely converge. It is worth noting that models within the same rank exhibit closely aligned values, as demonstrated in Figure S14 of the supplementary material.

5. Conclusion

This study constructs multiple 3D models of three coals from Pennsylvania. The coals are of different ranks including high volatile bituminous (P8), low volatile bituminous coal (KS-325), and semi-anthracite coal (KS-121) coal. Proximate, ultimate, ^{13}C NMR, FTIR, and XPS experiments were conducted to gather comprehensive data. Structural parameters governing the carbon skeletal structure of the coals were calculated from fitted NMR spectra, supplemented by FTIR data when necessary. Oxygen functional groups were predicted using NMR, FTIR, and XPS, with the latter serving as the primary indicator for sulfur and nitrogen functionalities proposed for coal. The structural models for the coals underwent iterative updates to align with the experimental NMR spectra, while considering insights from other experimental analyses.

To accommodate the complexity of coal structure and potential biases in experimental curve-fitting methods, at least three models were constructed for each coal. Each model displayed distinct features while still matching the experimental NMR spectra. The 2D coal structures were further transformed into realistic 3D structures for molecular simulations and chemical analysis. The REAXFF models were validated *ex post facto* via DFT codes implemented within VASP. To ensure the

uniqueness of the models, the SOAP method encoded the local environment of atoms in each coal structure. A pairwise similarity kernel metric was then applied to compare atom-centered environments for each coal, providing a robust validation of the distinctiveness of the constructed models. As the *pièce de résistance*, the vibrational density of state of the DFT-validated models were computed and were found to have striking similarities to the shape profiles from FTIR and INS. Furthermore, the specific heat capacity, derived from elementary statistical mechanics, employing the vibrational density of states, exhibited remarkable congruence with experimentally determined specific heat values. This marks the first instance of *ab initio* calculations of the vibrational density of states for coal reported in literature.

To the best of our knowledge, there is no publication on coal structural models for coals from the semi-anthracite and low volatile bituminous coal regions from which KS-325 and KS-121 are mined. Conversely, a model exists for P8 coal, as posited by Solomon [18]. However, our experimental findings and analysis for P8 coal indicate an underestimation of carbonyl and carboxyl functional groups in Solomon's coal model. It is essential to emphasize that this does not discredit Solomon's proposed structure, nor does it assert superior accuracy for ours, but highlights the inherent non-uniqueness in describing coal structures, a phenomenon influenced by both the nature of the coal itself and potential biases in experimental techniques. Therefore, our prediction of the P8 coal structure only provides an alternative perspective.

Funding

This material is based upon work supported by the U.S. Department of Energy under award number DE-FE0032143. This work also used computational resources at the Pittsburgh Supercomputing Center (Bridges-2 Regular Memory) through allocation dmr190008p and phy230007p from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, funded by the U.S. National Science Foundation (NSF) grant number: 2138259, 2138286, 2138307, 2137603, and 2138296.

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

CRediT authorship contribution statement

Chinonso Ugwumadu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **David A. Drabold:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Natasha L. Smith:** Writing – review & editing. **Jason Trembly:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Rudolph Olson:** Writing – review & editing, Validation, Resources, Project administration, Investigation, Funding acquisition. **Eric Shereda:** Funding acquisition, Project administration, Resources, Validation, Writing – review & editing. **Yahya T. Al-Majali:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

C.U. expresses appreciation to the Nanoscale & Quantum Phenomena Institute (NQPI) for the financial support conferred through the NQPI research fellowship (2023). C.U. thank Dr. Dan Conroy, Dr. Kody Wolfe, and Dr. Khalifa Yehia for helpful discussions regarding NMR, FTIR and XPS data analyses, respectively. The authors thank Grace Baranack for her valuable assistance in providing the coals for the experimental part of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carbon.2024.119086.

References

[1] V.C. Hoang, M. Hassan, V.G. Gomes, Coal derived carbon nanomaterials – recent advances in synthesis and applications, Appl. Mater. Today 12 (2018) 342–358, https://doi.org/10.1016/j.apmt.2018.06.007.
[2] R. Atkins, Coal in a new carbon age: powering a wave of innovation in advanced products & manufacturing, National Coal Council Reports (2019) 1–80. https://researchrepository.wvu.edu/ncc_reports/50.
[3] Y. Zhang, H. Li, T. Ramirez Reina, J. Liu, Coal chemistry industry: from production of liquid fuels to fine chemicals to carbon materials, Energy Fuel. 37 (2023) 17754–17764, https://doi.org/10.1021/acs.energyfuels.3c02661.
[4] L.E. Veley, C. Ugwumadu, J.P. Trembly, D.A. Drabold, Y. Al-Majali, 3D printing of sustainable coal polymer composites: study of processing, mechanical performance, and atomistic matrix–filler interaction, ACS Appl. Polym. Mater. 5 (2023) 9286–9296, https://doi.org/10.1021/acsapm.3c01784.
[5] Y.A. Al-Majali, C.T. Chirume, E.P. Marcum, D.A. Daramola, K.S. Kappagantula, J. P. Trembly, Coal-filler-based thermoplastic composites as construction materials: a new sustainable end-use application, ACS Sustain. Chem. Eng. 7 (2019) 16870–16878, https://doi.org/10.1021/acssuschemeng.9b04453.
[6] Y.T. Al-Majali, E.S. Alamiri, B. Wisner, J.P. Trembly, Mechanical performance assessment of sustainable coal plastic composite building materials, J. Build. Eng. 80 (2023) 108089, https://doi.org/10.1016/j.jobe.2023.108089.
[7] L. Veley, J. Trembly, Y. Al-Majali, 3D printing of sustainable coal polymer composites: thermophysical characteristics, Mater. Today Commun. 37 (2023) 106989, https://doi.org/10.1016/j.mtcomm.2023.106989.
[8] A. Paul, R. Magee, W. Wilczewski, N. Wichert, C. Gula, R. Olson, E. Shereda, Y. Al-Majali, J.P. Trembly, K. Wolfe, J. Staser, T.R. Garrick, Application of the multi-species, multi-reaction model to coal-derived graphite for lithium-Ion batteries, J. Electrochem. Soc. (2024), https://doi.org/10.1149/1945-7111/ad2061.
[9] M. Shi, C. Song, Z. Tai, K. Zou, Y. Duan, X. Dai, J. Sun, Y. Chen, Y. Liu, Coal-derived synthetic graphite with high specific capacity and excellent cyclic stability as anode material for lithium-ion batteries, Fuel 292 (2021) 120250, https://doi.org/10.1016/j.fuel.2021.120250.
[10] H. Zhang, K. Li, J. Sun, Z. Sun, L. Yuan, Q. Liu, The structural evolution and mutation of graphite derived from coal under the influence of natural igneous plutonic intrusion, Fuel 322 (2022) 124066, https://doi.org/10.1016/j.fuel.2022.124066.
[11] A. Gharpure, R.L. Vander Wal, Enhancing graphitic carbon precursors from coal pyrolysis: a comparative analysis of microwave plasma and conventional thermal upgradation methods, ACS Omega 8 (2023) 40587–40599, https://doi.org/10.1021/acsomega.3c05382.
[12] C. Ugwumadu, R.O. III, N.L. Smith, K. Nepal, Y. Al-Majali, J. Trembly, D. A. Drabold, Computer simulation of carbonization and graphitization of coal, Nanotechnology 35 (2023) 095703, https://doi.org/10.1088/1361-6528/ad1058.
[13] F. Islam, A. Tahmasebi, R. Wang, J. Yu, Structure of coal-derived metal-supported few-layer graphene composite materials synthesized using a microwave-assisted catalytic graphitization process, Nanomaterials 11 (2021) 1672, https://doi.org/10.3390/nano11071672.
[14] J. Zhang, C. Liang, J.B. Dunn, Graphite flows in the U.S.: insights into a key ingredient of energy transition, Environ. Sci. Technol. 57 (2023) 3402–3414, https://doi.org/10.1021/acs.est.2c08655.

[15] D.W. van Krevelen, J. Schuyer, Coal Science; Aspects of Coal Constitution, Elsevier Pub. Co., Amsterdam, 1957.

[16] J.H. Shinn, From coal to single-stage and two-stage products: a reactive model of coal structure, Fuel 63 (1984) 1187–1196, https://doi.org/10.1016/0016-2361(84)90422-8.

[17] W.H. Wiser, Conversion of bituminous coal to liquids and gases: chemistry and representative processes, in: L. Petrakis, J.P. Fraissard (Eds.), Magn. Reson. Introd. Adv. Top. Appl. Foss. Energy, Springer, Netherlands, Dordrecht, 1984, pp. 325–350, https://doi.org/10.1007/978-94-009-6378-8_12.

[18] P.R. Solomon, New Approaches Coal Chem. 169 (1981) 61–71, https://doi.org/10.1021/bk-1981-0169.ch004.

[19] P.H. Given, Structure of bituminous coals: evidence from distribution of hydrogen, Nature 184 (1959) 980–981, https://doi.org/10.1038/184980a0.

[20] K. Shi, J. Chen, X. Pang, F. Jiang, S. Hui, S. Zhang, H. Pang, Y. Wang, D. Chen, X. Yang, B. Li, T. Pu, Average molecular structure model of shale kerogen: experimental characterization, structural reconstruction, and pyrolysis analysis, Fuel 355 (2024) 129474, https://doi.org/10.1016/j.fuel.2023.129474.

[21] J.P. Mathews, A.C.T. van Duin, A.L. Chaffee, The utility of coal molecular models, Fuel Process. Technol. 92 (2011) 718–728, https://doi.org/10.1016/j.fuproc.2010.05.037.

[22] Y. Xia, R. Zhang, Y. Cao, Y. Xing, X. Gui, Role of molecular simulation in understanding the mechanism of low-rank coal flotation: a review, Fuel 262 (2020) 116535, https://doi.org/10.1016/j.fuel.2019.116535.

[23] S. Liu, L. Wei, Q. Zhou, T. Yang, S. Li, Q. Zhou, Simulation strategies for ReaxFF molecular dynamics in coal pyrolysis applications: a review, J. Anal. Appl. Pyrolysis 170 (2023) 105882, https://doi.org/10.1016/j.jaap.2023.105882.

[24] M.S. Solum, R.J. Pugmire, D.M. Grant, Carbon-13 solid-state NMR of Argonne-premium coals, Energy Fuel. 3 (1989) 187–193, https://doi.org/10.1021/ef00014a012.

[25] M.S. Solum, R.J. Pugmire, Coal Liquefaction Process Streams Characterization and Evaluation: 13C-NMR Analysis of CONSOL THF-Soluble Residual Materials from the Wilsonville Coal Liquefaction Process, CONSOL, Inc., Library, PA (United States); Utah Univ., Salt Lake City, UT (United States), 1992, https://doi.org/10.2172/6782684.

[26] H. Kawashima, Y. Yamashita, I. Saito, Studies on structural changes of coal density-separated components during pyrolysis by means of solid-state 13C NMR spectra, J. Anal. Appl. Pyrolysis 53 (2000) 35–50, https://doi.org/10.1016/S0165-2370(99)00058-3.

[27] J. Jia, Q. Yang, B. Liu, D. Wang, Structural characterization and macromolecular structure construction of non-caking coal in Chicheng mine, Sci. Rep. 13 (2023) 16931, https://doi.org/10.1038/s41598-023-44045-2.

[28] Q. Wei, Y. Tang, 13C-NMR study on structure evolution characteristics of high-organic-sulfur coals from typical Chinese areas, Minerals 8 (2018) 49, https://doi.org/10.3390/min8020049.

[29] P. Painter, M. Starsinic, M. Coleman, Determination of functional groups in coal by Fourier transform interferometry, Fourier Transform Infrared Spectrosc. 4 (2012) 169–240.

[30] P.C. Painter, M.M. Coleman, R.W. Snyder, O. Mahajan, M. Komatsu, P.L. Walker, Low temperature air oxidation of caking coals: Fourier transform infrared studies, Appl. Spectrosc. 35 (1981) 106–110.

[31] M. Kozłowski, XPS study of reductively and non-reductively modified coals, Fuel 83 (2004) 259–265, https://doi.org/10.1016/j.fuel.2003.08.004.

[32] M.A. Olivella, J.M. Palacios, A. Vairavamurthy, J.C. del Rio, F.X.C. de las Heras, A study of sulfur functionalities in fossil fuels using destructive- (ASTM and Py–GC–MS) and non-destructive- (SEM–EDX, XANES and XPS) techniques, Fuel 81 (2002) 405–411, https://doi.org/10.1016/S0016-2361(01)00198-3.

[33] T. Grzybek, K. Kreiner, Surface changes in coals after oxidation. 1. X-ray photoelectron spectroscopy studies, Langmuir 13 (1997) 909–912, https://doi.org/10.1021/la9510893.

[34] R. Thapa, D.A. Drabold, Ab initio simulation of amorphous materials, in: At. Simul. Glas., John Wiley & Sons, Ltd, 2022, pp. 30–59, https://doi.org/10.1002/9781118939079.ch2.

[35] G.H. Major, T.G. Avval, B. Moeini, G. Pinto, D. Shah, V. Jain, V. Carver, W. Skinner, T.R. Gengenbach, C.D. Easton, A. Herrera-Gomez, T.S. Nunney, D. R. Baer, M.R. Linford, Assessment of the frequency and nature of erroneous x-ray photoelectron spectroscopy analyses in the scientific literature, J. Vac. Sci. Technol. A 38 (2020) 061204, https://doi.org/10.1116/6.0000685.

[36] O. Exner, How to get wrong results from good experimental data: a survey of incorrect applications of regression, J. Phys. Org. Chem. 10 (1997) 797–813, https://doi.org/10.1002/(SICI)1099-1395(199711)10:11<797::AID-PCA951>3.0.CO;2-K.

[37] Y. Chen, C. Zou, M. Mastalerz, S. Hu, C. Gasaway, X. Tao, Applications of micro-Fourier transform infrared spectroscopy (FTIR) in the geological sciences—a review, Int. J. Mol. Sci. 16 (2015) 30223–30250, https://doi.org/10.3390/ijms161226227.

[38] D.N.G. Krishna, J. Philip, Review on surface-characterization applications of X-ray photoelectron spectroscopy (XPS): recent developments and challenges, Appl. Surf. Sci. Adv. 12 (2022) 100332, https://doi.org/10.1016/j.apsadv.2022.100332.

[39] G.A. Carlson, Computer simulation of the molecular structure of bituminous coal, Energy Fuel. 6 (1992) 771–778, https://doi.org/10.1021/ef00036a012.

[40] G. Li, F. Zheng, Q. Huang, J. Wang, B. Niu, Y. Zhang, D. Long, Molecular insight into pyrolysis processes via reactive force field molecular dynamics: a state-of-the-art review, J. Anal. Appl. Pyrolysis 166 (2022) 105620, https://doi.org/10.1016/j.jaap.2022.105620.

[41] A.C.T. van Duin, S. Dasgupta, F. Lorant, W.A. Goddard, ReaxFF: a reactive force field for hydrocarbons, J. Phys. Chem. A 105 (2001) 9396–9409, https://doi.org/10.1021/jp004368u.

[42] A.P. Thompson, H.M. Aktulga, R. Berger, D.S. Bolintineanu, W.M. Brown, P. S. Crozier, P.J. in 't Veld, A. Kohlmeyer, S.G. Moore, T.D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, S.J. Plimpton, Lammps - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comput. Phys. Commun. 271 (2022) 108171, https://doi.org/10.1016/j.cpc.2021.108171.

[43] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169–11186, https://doi.org/10.1103/PhysRevB.54.11169.

[44] A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments, Phys. Rev. B 87 (2013) 184115, https://doi.org/10.1103/PhysRevB.87.184115.

[45] S. De, A.P. Bartók, G. Csányi, M. Ceriotti, Comparing molecules and solids across structural and alchemical space, Phys. Chem. Chem. Phys. 18 (2016) 13754–13769, https://doi.org/10.1039/C6CP00415F.

[46] L. Himanen, M.O.J. Jäger, E.V. Morooka, F. Federici Canova, Y.S. Ranawat, D. Z. Gao, P. Rinke, A.S. Foster, DScribe: Library of descriptors for machine learning in materials science, Comput. Phys. Commun. 247 (2020) 106949, https://doi.org/10.1016/j.cpc.2019.106949.

[47] S.N. Taraskin, S.R. Elliott, Connection between the true vibrational density of states and that derived from inelastic neutron scattering, Phys. Rev. B 55 (1997) 117–123, https://doi.org/10.1103/PhysRevB.55.117.

[48] R. Haworth, G. Mountjoy, M. Corno, P. Ugliengo, R.J. Newport, Probing vibrational modes in silica glass using inelastic neutron scattering with mass contrast, Phys. Rev. B 81 (2010) 060301, https://doi.org/10.1103/PhysRevB.81.060301.

[49] F. Fillaux, R. Papoular, A. Lautié, J. Tomkinson, Inelastic neutron-scattering study of the proton dynamics in coals, Fuel 74 (1995) 865–873, https://doi.org/10.1016/0016-2361(95)00017-Y.

[50] K.N. Subedi, K. Prasai, D.A. Drabold, Space-projected conductivity and spectral properties of the conduction matrix, Phys. Status Solidi B 258 (2021) 2000438, https://doi.org/10.1002/pssb.202000438.

[51] ASTM International D3172-13 (2021)e1, Standard practice for proximate analysis of coal and coke. https://doi.org/10.1520/D3172-13R21E01, 2013.

[52] ASTM International D3176-15, Standard practice for ultimate analysis of coal and coke. https://doi.org/10.1520/D3176-15R23, 2015.

[53] ASTM International D5373-08, Standard test methods for determination of carbon, hydrogen, and nitrogen in laboratory samples of coal. https://doi.org/10.1520/D5373-08, 2008.

[54] ASTM International D4239-18e1, Standard test method for sulfur in the analysis sample of coal and coke using high-temperature tube furnace combustion. https://doi.org/10.1520/D4239-18E01, 2018.

[55] ASTM International D121-15, Standard terminology of coal and coke. https://doi.org/10.1520/D0121-15, 2015.

[56] ASTM International D388-23, Standard Classification of Coals by Rank, 2023, https://doi.org/10.1520/D0388-23.

[57] S.W. Parr, The Classification of Coal, University of Illinois, 1928. https://books.google.com/books?id=LjAkAQAAMAAJ.

[58] T. Yoshida, Y. Maekawa, Characterization of coal structure by CP/MAS carbon-13 NMR spectrometry, Fuel Process. Technol. 15 (1987) 385–395, https://doi.org/10.1016/0378-3820(87)90060-9.

[59] M.J. Trewhella, I.J.F. Poplett, A. Grint, Structure of Green River oil shale kerogen: determination using solid state 13C NMR spectroscopy, Fuel 65 (1986) 541–546, https://doi.org/10.1016/0016-2361(86)90046-3.

[60] JoséV. Ibarra, E. Muñoz, R. Moliner, FTIR study of the evolution of coal structure during the coalification process, Org. Geochem. 24 (1996) 725–735, https://doi.org/10.1016/0146-6380(96)00063-0.

[61] L.A. Schultheiss, E.M. Heiliger, Techniques of flow-charting, Clin. Libr. Appl. Data Process. 1st (1963) 1963. https://hdl.handle.net/2142/743. (Accessed 9 March 2024).

[62] N.L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms, J. Am. Chem. Soc. 99 (1977) 8127–8134, https://doi.org/10.1021/ja00467a001.

[63] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms, J. Chem. Phys. 101 (1994) 4177–4189, https://doi.org/10.1063/1.467468.

[64] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190, https://doi.org/10.1063/1.328693.

[65] W. Shinoda, M. Shiga, M. Mikami, Rapid estimation of elastic constants by molecular dynamics simulation under constant stress, Phys. Rev. B 69 (2004) 134103, https://doi.org/10.1103/PhysRevB.69.134103.

[66] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690, https://doi.org/10.1063/1.448118.

[67] S.M. Nakhmanson, D.A. Drabold, Low-temperature anomalous specific heat without tunneling modes: a simulation for a-Si with voids, Phys. Rev. B 61 (2000) 5376–5380, https://doi.org/10.1103/PhysRevB.61.5376.

[68] B. Bhattarai, D.A. Drabold, Amorphous carbon at low densities: an ab initio study, Carbon 115 (2017) 532–538, https://doi.org/10.1016/j.carbon.2017.01.031.

[69] P.E. Blöchl, Projector augmented-wave method, Phys. Rev. B 50 (1994) 17953–17979, https://doi.org/10.1103/PhysRevB.50.17953.

[70] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868, https://doi.org/10.1103/PhysRevLett.77.3865.
[71] M.O.J. Jäger, E.V. Morooka, F. Federici Canova, L. Himanen, A.S. Foster, Machine learning hydrogen adsorption on nanoclusters through structural descriptors, npj Comput. Mater. 4 (2018) 1–8, https://doi.org/10.1038/s41524-018-0096-5.
[72] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, London, England, 2006, in: https://gaussianprocess.org/gpml/chapters/.
[73] D.W. Kuehn, R.W. Snyder, A. Davis, P.C. Painter, Characterization of vitrinite concentrates. 1. Fourier Transform infrared studies, Fuel 61 (1982) 682–694, https://doi.org/10.1016/0016-2361(82)90240-X.
[74] D. Pantea, H. Darmstadt, S. Kaliaguine, C. Roy, Electrical conductivity of conductive carbon blacks: influence of surface chemistry and topology, Appl. Surf. Sci. 217 (2003) 181–193, https://doi.org/10.1016/S0169-4332(03)00550-6.
[75] S.R. Kelemen, M.L. Gorbaty, P.J. Kwiatek, T.H. Fletcher, M. Watt, M.S. Solum, R. J. Pugmire, Nitrogen transformations in coal during pyrolysis, Energy Fuel. 12 (1998) 159–173, https://doi.org/10.1021/ef9701246.
[76] L. Zhang, Z. Li, Y. Yang, Y. Zhou, J. Li, L. Si, B. Kong, Research on the composition and distribution of organic sulfur in coal, Molecules 21 (2016) 630, https://doi.org/10.3390/molecules21050630.
[77] F. Gao, Z. Jia, M. Qin, X. Mu, Y. Teng, Y. Li, Q. Bai, Effects of organic sulfur on oxidation spontaneous combustion characteristics of coking coal, Energy Explor. Exploit. 40 (2022) 193–205, https://doi.org/10.1177/01445987211049045.
[78] E.E. Smith, Coals of the State of Washington, Department of the Interior, U.S. Geological Survey, 1911. https://books.google.com/books?id=he8JAAAAIAAJ.
[79] ASTM International D3180-15, Standard practice for calculating coal and coke analyses from as-determined to different bases. https://doi.org/10.1520/D3180-15, 2015.
[80] L.J. Bellamy, The Infra-red Spectra of Complex Molecules, Springer Netherlands, Dordrecht, 1975, https://doi.org/10.1007/978-94-011-6017-9.
[81] W.H. Calkins, Coal flash pyrolysis: 3. An analytical method for polymethylene moieties in coal, Fuel 63 (1984) 1125–1129, https://doi.org/10.1016/0016-2361(84)90199-6.
[82] W.H. Calkins, B.K. Hovsepian, G.R. Dyrkacz, C.A.A. Bloomquist, L. Ruscic, Coal flash pyrolysis: 4. Polymethylene moieties in coal macerals, Fuel 63 (1984) 1226–1229, https://doi.org/10.1016/0016-2361(84)90429-0.
[83] T. Yoshida, T. Kazuaki, M. Yosuke, Liquefaction reaction of coal: 1. Depolymerization of coal by cleavages of ether and methylene bridges, Fuel 64 (1985) 890–896, https://doi.org/10.1016/0016-2361(85)90138-3.
[84] D.A. Drabold, C. Ugwumadu, Webpage containing the structure files for the coal models constructed in this study, Mater. Theory Group - Ohio Univ (2024). https://people.ohio.edu/drabold/coal_models/. (Accessed 15 January 2024).
[85] W.E. White, C.H. Bartholomew, W.C. Hecker, D.M. Smith, Changes in surface area, pore structure and density during formation of high-temperature chars from representative U.S. coals, Adsorpt. Sci. Technol. 7 (1990) 180–209, https://doi.org/10.1177/026361749000700401.
[86] R. Thapa, C. Ugwumadu, K. Nepal, D.A. Drabold, M.T.M. Shatnawi, Ab initio simulation of amorphous GeSe3 and GeSe4, J. Non-Cryst. Solids 601 (2023) 121998, https://doi.org/10.1016/j.jnoncrysol.2022.121998.
[87] G.E. Bacon, N.A. Curry, S.A. Wilson, R. Spence, A crystallographic study of solid benzene by neutron diffraction, Proc. R. Soc. Lond. Ser. Math. Phys. Sci. 279 (1997) 98–110, https://doi.org/10.1098/rspa.1964.0092.
[88] D.R. Lide (Ed.), CRC Handbook of Chemistry and Physics, 88th edition, CRC Press, Boca Raton, Fla, 2007.
[89] L. Pauling, L.O. Brockway, Carbon—carbon bond distances. The electron diffraction investigation of ethane, propane, isobutane, meopentane, cyclopropane, cyclopentane, cyclohexane, allene, ethylene, isobutene, tetramethylethylene, mesitylene, and hexamethylbenzene. Revised values of covalent radii, J. Am. Chem. Soc. 59 (1937) 1223–1236, https://doi.org/10.1021/ja01286a021.
[90] K.K. Onchoke, C.M. Hadad, P.K. Dutta, Density functional theoretical study of nitrated polycyclic aromatic hydrocarbons, Polycycl. Aromat. Comp. 24 (2004) 37–64, https://doi.org/10.1080/10406630490277443.
[91] R. Thapa, C. Ugwumadu, K. Nepal, J. Trembly, D.A. Drabold, Ab initio simulation of amorphous graphite, Phys. Rev. Lett. 128 (2022) 236402, https://doi.org/10.1103/PhysRevLett.128.236402.
[92] C. Ugwumadu, K. Nepal, R. Thapa, Y.G. Lee, Y. Al Majali, J. Trembly, D. A. Drabold, Simulation of multi-shell fullerenes using machine-learning Gaussian approximation potential, Carbon Trends 10 (2023) 100239, https://doi.org/10.1016/j.cartre.2022.100239.
[93] K.S. Kappagantula, J.A. Smith, A.K. Nittala, F.F. Kraft, Macro copper-graphene composites with enhanced electrical conductivity, J. Alloys Compd. 894 (2022) 162477, https://doi.org/10.1016/j.jallcom.2021.162477.
[94] K.N. Subedi, K. Nepal, C. Ugwumadu, K. Kappagantula, D.A. Drabold, Electronic transport in copper–graphene composites, Appl. Phys. Lett. 122 (2023) 031903, https://doi.org/10.1063/5.0137086.
[95] A. Nittala, J. Smith, B. Gwalani, J. Silverstein, F.F. Kraft, K. Kappagantula, Simultaneously improved electrical and mechanical performance of hot-extruded bulk scale aluminum-graphene wires, Mater. Sci. Eng. B 293 (2023) 116452, https://doi.org/10.1016/j.mseb.2023.116452.
[96] K. Nepal, C. Ugwumadu, K.N. Subedi, K. Kappagantula, D.A. Drabold, Physical origin of enhanced electrical conduction in aluminum-graphene composites, Appl. Phys. Lett. 124 (2024) 091902, https://doi.org/10.1063/5.0195967.
[97] K. Nepal, C. Ugwumadu, A. Gautam, K. Kappagantula, D.A. Drabold, Electronic conductivity in metal-graphene composites: the role of disordered carbon structures, defects, and impurities, J. Phys. Mater. 7 (2024) 025003, https://doi.org/10.1088/2515-7639/ad261a.
[98] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[99] L.J.P. van der Maaten, G.E. Hinton, Visualizing high-dimensional data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[100] J.D. Morrow, C. Ugwumadu, D.A. Drabold, S.R. Elliott, A.L. Goodwin, V. L. Deringer, Understanding defects in amorphous silicon with million-atom simulations and machine learning, Angew. Chem. int. Ed. (2023) e202403842, https://doi.org/10.1002/anie.202403842.
[101] S. Fujii, Y. Ōsawa, H. Sugimura, Infra-red spectra of Japanese coal: the absorption bands at 3030, 2920 and 1600 cm−1, Fuel 49 (1970) 68–75, https://doi.org/10.1016/0016-2361(70)90009-8.
[102] R. Olson III, Experimental specific heat capacity of carbon foam made from coal., CONSOL Innovations LLCTriadelphia, WV, U.S.A, Unpublished Results..
[103] D.A. Drabold, J. Dong, Band tail states and the Anderson transition in amorphous silicon, J. Non-Cryst. Solids 227–230 (1998) 153–157, https://doi.org/10.1016/S0022-3093(98)00030-1.
[104] M. Laing, M. Laing, Dulong and Petit's law: we should not ignore its importance, J. Chem. Educ. 83 (2006) 1499, https://doi.org/10.1021/ed083p1499.

Supporting Information

The Structure of Appalachian Coal: Experiments and *Ab initio* Modeling

Chinonso Ugwumadu [1,], David A. Drabold [1,*], Natasha L. Smith [2], Jason Trembly [3], Rudolph Olson III [2], Eric Shereda [4], Yahya T. Al-Majali [3,*]*

1) Department of Physics and Astronomy, Nanoscale and Quantum Phenomena Institute (NQPI), Ohio University, Athens, OH 45701, USA.
2) CONSOL Innovations, Triadelphia, WV 26059, USA
3) Department of Mechanical Engineering, Institute for Sustainable Energy, and the Environment (ISEE), Ohio University, Athens, Ohio 45701, USA.
4) CONSOL Energy, Canonsburg, PA 15317, USA

*Corresponding Authors e-mail: cu884120@ohio.edu; drabold@ohio.edu; almajali@ohio.edu

Keywords: Coal, NMR, FTIR, XPS, ReaxFF, DFT

Abstract

This study focuses on the construction of small-scale atomistic representations of three Appalachian coals of different ranks– high volatile bituminous, low volatile bituminous, and semi-anthracite. Ultimate, proximate, ^{13}C NMR, FTIR, and XPS data were used to infer 2D and construct multiple 3D molecular models of each coal. A new simulation strategy for structure optimization, termed "Sectioned Optimization", is introduced to ensure energetically stable configurations of the 3D coal models. Density functional methods are applied to the models, improving their quality, and giving insight into the limitations of empirical simulations in this complex system. The structural models were validated *ex post facto* using a density functional code, VASP. We describe the distribution of local atomic environments in the models to emphasize the significant structural distinctions between the three coal ranks. We also report the first *ab initio* vibrational density of states calculations for coal and compare it to FTIR and Inelastic Neutron Scattering experiments and identify vibrational modes using *ab initio* calculations, offering animations for the vibrational modes in coal that are consistent with FTIR band assignments. We show that the specific heat capacity compares favorably with experimental data.

Table of Contents

S1. Supplementary Notes and Comments S3

S1.1 Formulas to calculate some ^{13}C NMR Structural Parameters S3

S1.2 2D and 3D Structural Models of the KS-121, KS-325 and P8 Coals. S4

S1.3 Animations of Vibrational Modes in the Coal Model S4

S1.4 t-distributed stochastic neighbor embedding (t-SNE) S4

S2. Tables S5

S2.1 ^{13}C Nuclear Magnetic Resonance Measurement S5

Table S1. Peak area (A), chemical shifts (δ in ppm), relative area percentage ratio (r%), and structural parameters determined from ^{13}C NMR analysis of the coals. S5

S2.2 Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy S6

Table S2. Peak area (A), Peak position (μ), relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 700 – 900 cm^{-1}. S6

Table S3. Peak area (A), Peak position (μ), relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 1000 – 1800 cm^{-1}. S7

Table S4. Peak area (A), Peak position (μ), relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 2800 – 3100 cm^{-1}. S8

Table S5. ATR-FTIR Peak area (A), Peak position (μ), and relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 3100 – 3600 cm^{-1}. S9

S3. Figures S10

References S21

S1.Supplementary Notes and Comments

S1.1 Formulas to calculate some ^{13}C NMR Structural Parameters

The methodology employed to derive all seven parameters is based on the work of Solum, Pugmire, and Grant [1,2]. The average number of carbon atoms per aromatic cluster or polyaromatic hydrocarbon (n_{cl}) is calculated based on the fraction of bridgehead carbon atoms (X_b), which can be expressed as:

$$X_b = \frac{f_a^B}{f_a^H + f_a^N} \quad \textbf{S1}$$

To determine n_{cl}, the cluster size model introduced by Solum and coworkers[1] provides a numerical solution through the following equations:

$$X_b = 0.5 \times \{(1 - \tanh\beta)X_b' + (1 + \tanh\beta)X_b''\} \quad \textbf{S2}$$

$$\beta = \frac{n_{cl} - C_0}{m}; \quad X_b' = \frac{n_{cl} - 6}{2 \times n_{cl}}; \quad X_b'' = \frac{1 - \sqrt{6}}{\sqrt{n_{cl}}} \quad \textbf{S3}$$

C_0 and m are the shifting and scaling parameters, set to their optimal values of 19.57 and 4.15 [1]. X_b' and X_b'' are the limits for linear and circular of carbon. Taking $\boldsymbol{f_a^s + f_a^p}$ to be the total number of aromatic ring attachment, the cluster coordination number (σ) of the number of attachments per cluster can be calculated as [1,2]:

$$\sigma = \frac{(f_a^s + f_a^p)}{(f_a/n_{cl})} \quad \textbf{S4}$$

If broken bridges are taken to end with a single methyl group, then one can approximate the fraction of intact bridges in a cluster (I_b) as [2]:

$$I_b = \frac{(f_a^s + f_a^p - f_{al})}{(f_a^s + f_a^p)} \quad \textbf{S5}$$

The fraction of aliphatic bridges and loops (hydroaromatic structure like the dihydroanthracene structure), denoted as *BL*, is:

$$BL = \sigma \times I_b \quad \textbf{S6}$$

While the side chains per cluster (*SC*) is defined as:

$$SC = \sigma - L_b \quad \textbf{S7}$$

An illustration of bridges, loops and side chains is presented in **Figure S2**. Given the percentage of carbon in the coal from the ultimate analysis (%carbon), the total molecular weight (W_c) of the cluster, including all the side chains and half of the bridge and looped aliphatic carbon is calculated as:

$$W_c = \frac{n_{cl} \times 12.01}{f_a \times \%carbon \times 0.01} \quad \text{S8}$$

S1.2 2D and 3D Structural Models of the KS-121, KS-325 and P8 Coals.

The structure files for the coal models developed in this study can be accessed from our website (see link). The 2D structure is available in ChemDraw® XML format (*.chxml*), while the 3D models are stored in XYZ formats. The coals are organized in folders using the code names from the manuscript (P8, KS325, and KS121), and accompanying coal structure images are included. For convenience, we also offer the 3D XYZ files as supplementary material.

S1.3 Animations of Vibrational Modes in the Coal Model

We present two video files displaying vibrational normal modes attributed to specific functional groups within the coal models. Figure S12 illustrates the vibrational density of states (VDoS) along with its corresponding inverse participation ratio (VIPR). The videos, which are also available in the *KS121_Animations* folder on our website (see link), are as follows:

1. **gapModes.mp4:** This video displays vibrational normal modes associated with functional groups situated within the phonon energy gap region spanning from 1600 cm^{-1} to 29500 cm^{-1}.
2. **otherModes.mp4:** This video highlights vibrational modes occurring in both the lower and higher frequency regions, outside the phonon energy gap.

In the animations, hydrogen, carbon, nitrogen, oxygen, and sulfur are color coded in white, gray, blue, red, and yellow, respectively.

S1.4 t-distributed stochastic neighbor embedding (t-SNE).

t-distributed Stochastic Neighbor Embedding (t-SNE) is an algorithm for visualizing high-dimensional data in lower dimensions [3]. In t-SNE, the coordinates denote the positions of data points in a reduced-dimensional space, typically 2D or 3D. Unlike principal component analysis (PCA), which is effective for linear data [4], t-SNE is specifically designed to handle non-linear data. The algorithm begins by computing pairwise similarities or distances between data points in the original high-dimensional space. It then aims to embed these points in a lower-dimensional space (often 2D or 3D) while preserving their pairwise similarities as much as possible. The resulting coordinates in t-SNE represent the positions of the data points in the lower-dimensional space post-embedding. Each data point is assigned a specific set of coordinates corresponding to its location in the reduced-dimensional space. These coordinates are determined through the optimization process of the t-SNE algorithm, which endeavors to minimize the divergence between the pairwise similarities in the original high-dimensional space and those in the lower-dimensional space.

S2.Tables

S2.1 ^{13}C Nuclear Magnetic Resonance Measurement

Table S1. Peak area (A), chemical shifts (δ in ppm), relative area percentage ratio (r%), and structural parameters determined from ^{13}C NMR analysis of the coals.

Coal		Fitted NMR Peaks							
		a	b	c	d	e	f	g	h
P8	*A*	1.21E+04	5.69E+04	1.89E+04	3.60E+03	9.89E+04	1.07E+04	2.60E+04	3.26E+03
	δ	19.6	34.9	60.8	75.0	128.2	151.4	194.1	213.1
	r%	5.24	24.69	8.21	1.56	42.96	4.63	11.30	1.41
KS-325	*A*	3.96E+03	2.53E+03	9.65E+03	1.45E+03	5.75E+04	9.10E+03	1.87E+04	1.97E+03
	δ	21.96	36.34	62.09	75.13	128.00	142.38	194.66	208.87
	r%	3.78	2.41	9.21	1.39	54.80	8.68	17.85	1.88
KS-121	*A*	2.24E+03	1.17E+03	1.41E+04	1.64E+03	7.05E+04	4.98E+03	2.46E+04	1.07E+03
	δ	21.5	36.2	61.2	75.5	126.9	141.9	193.3	208.8
	r%	1.73	0.91	10.91	1.27	54.59	3.86	19.07	0.83
Structural Parameters		f_{al}^{*} (0 – 25)	f_{al}^{H} (25 – 51)	f_{al}^{O} (51 – 90)		$f_a^H + f_a^B$ (90 – 150)	$f_a^S + f_a^P$ (137 – 165)	f_a^C (165 – 220)	

S2.2 Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy

Table S2. Peak area (A), Peak position (μ), relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 700 – 900 cm^{-1}.

Coal		Fitted ATR-FTIR Peaks (700 – 900 cm^{-1}) [a]						
		a	b	c	d	e	f	g
P8	*A*	0.27	0.03	0.16	0.12	0.01	0.08	0.11
	μ	747.5	783.9	799.7	814.6	828.7	855.2	872.8
	r%	35.2	3.5	20.3	15.8	1.6	9.7	13.9
KS-325	*A*	0.71	0.05	0.17	0.36	–	–	0.50
	μ	746.0	784.0	796.8	810.3	–	–	868.6
	r%	39.7	2.9	9.6	19.9	–	–	27.9
KS-121	*A*	0.96	0.06	0.28	–	0.04	–	0.30
	μ	744.2	781.8	796.5	–	831.6	–	867.9
	r%	58.3	3.4	17.3	–	2.6	–	18.4

[a] ATR-FTIR peak Assignment: Aromatic CH out-of-plane bending modes. (a) 1,2-substituted aromatic group. (b and c) trisubstituted aromatic group. (d) isolated CH and/or two neighboring CH. (e) weak 1,4-substituted aromatic group. (f) tetrasubstituted aromatic group. (g) isolated aromatic CH group.

Table S3. Peak area (A), Peak position (μ), relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 1000 – 1800 cm^{-1}.

Coal		Fitted ATR-FTIR Peaks (1000 – 1800 cm^{-1}) [a)]				
		a	b	c	d	e
P8	*A*	7.51	10.08	5.55	8.11	9.15
	μ	1030.3	1177.2	1303.5	1430.5	1593.1
	r%	18.6	25.0	13.7	20.1	22.6
KS-325	*A*	7.06	5.23	6.50	6.38	9.06
	μ	1026.0	1179.1	1303.1	1422.9	1590.1
	r%	20.6	15.3	19.0	18.6	26.5
KS-121	*A*	12.71	–	1.88	3.10	5.71
	μ	1016.8	–	1307.0	1420.0	1589.3
	r%	54.3	–	8.0	13.2	24.4

[a)] ATR-FTIR peak Assignment: (a) aliphatic ethers and alcohols. (b and c) C–O stretch and OH bend in phenoxy structures and ethers. (d) deformation vibration of CH_2 and CH_3. (e) highly conjugated hydrogen bonded C=O.

Table S4. Peak area (A), Peak position (μ), relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 2800 – 3100 cm^{-1}.

Coal		Fitted ATR-FTIR Peaks (2800 – 3100 cm^{-1}) [a]				
		a	b	c	d	e
P8	*A*	2.27	3.99	0.60	–	0.30
	μ	2856.5	2918.8	2956.0	–	3039.0
	r%	31.7	55.8	8.4	–	4.2
KS-325	*A*	2.03	2.87	0.53	0.38	0.96
	μ	2861.3	2919.3	2957.5	3014.0	3045.3
	r%	30.0	42.4	7.8	5.6	14.2
KS-121	*A*	0.59	0.82	0.30	–	1.23
	μ	2861.5	2915.5	2955.4	–	3029.8
	r%	20.0	27.9	10.2	–	41.8

[a] ATR-FTIR peak Assignment: (a) symmetric stretching modes of CH_2 and/or CH_3 groups. (b) telescopic vibration of asymmetric CH_2 bridges only (i.e., no contribution from CH_2 in hydroaromatic structures). (c) CH_3 attached directly to aromatic rings (although CH_2 of certain hydroaromatic structures also absorb around 2950 cm^{-1}). (d) strained aliphatic rings or (more likely) aromatic CH vibrations. (e) aromatic CH vibrations

Table S5. ATR-FTIR Peak area (A), Peak position (μ), and relative area percentage ratio (r%), and functional groups assigned to ATR-FTIR spectra from 3100 – 3600 cm^{-1}.

Coal		Fitted ATR-FTIR Peaks (3100 – 3600 cm^{-1}) [a]				
		a	b	c	d	e
P8	*A*	0.83	1.77	2.00	1.15	0.42
	μ	3189.7	3274.2	3365.1	3454.6	3522.6
	r%	13.5	28.6	32.4	18.6	6.9
KS-325	*A*	0.62	1.31	–	0.51	0.05
	μ	3205.4	3320.0	–	3416.9	3561.6
	r%	25.1	52.7	–	20.4	1.8
KS-121	*A*	0.65	1.28	0.35	0.10	0.14
	μ	3202.6	3301.8	3390.1	3436.5	3533.8
	r%	25.8	50.7	14.0	4.1	5.5

[a] ATR-FTIR peak Assignment: (a) cyclic hydrogen bond. (b and c) OH--O hydrogen bond. (d) Hydrogen-bonded OH groups (i.e., OH--OH hydrogen bond). (e) OH--π hydrogen bond

S3. Figures

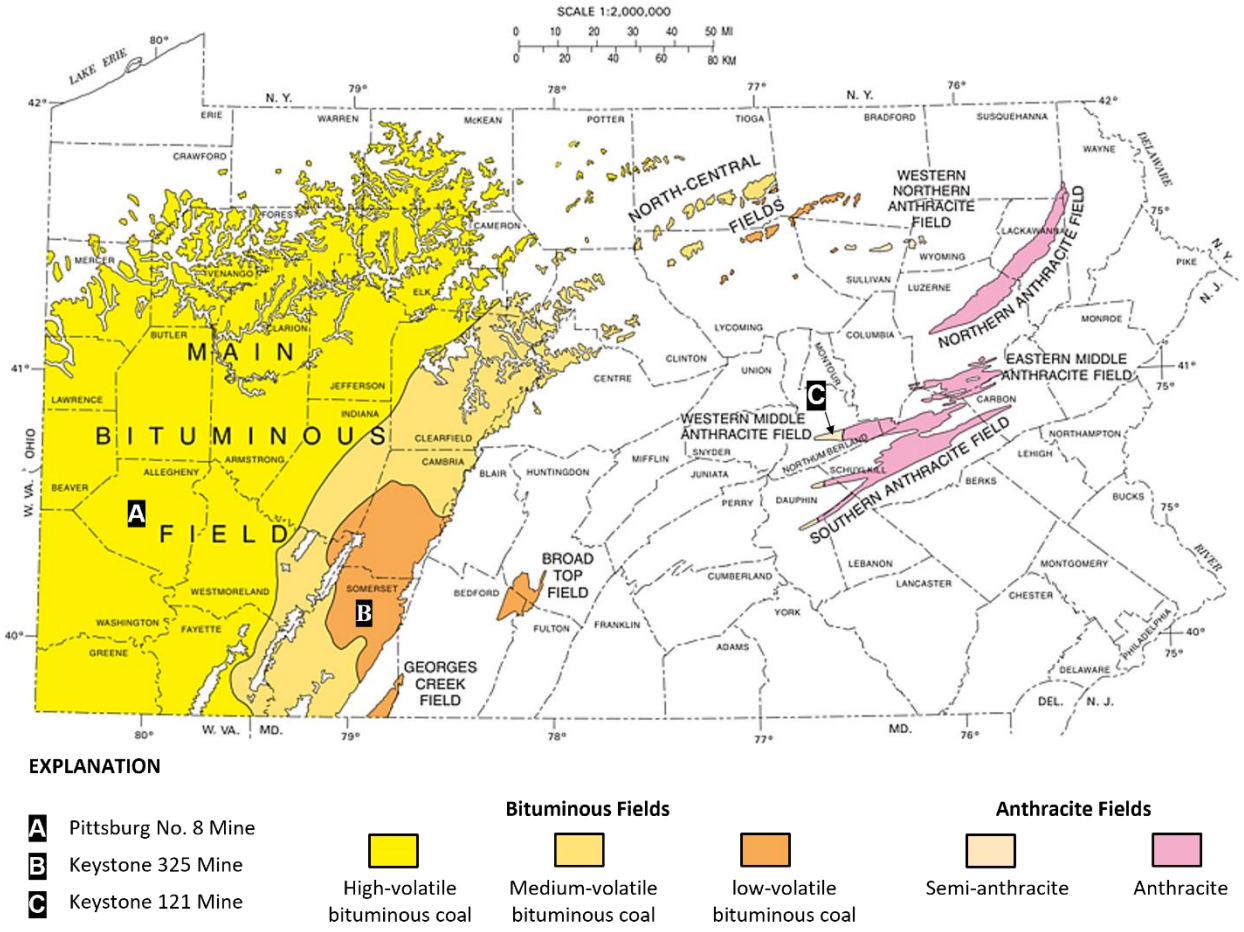


Figure S1. **(**Adapted from Reference [5]) Distribution of Pennsylvania coals showing the locations of the five major anthracite coal fields and the four major bituminous coal fields. Areas within the fields are color coded to one of five coal ranks: high-, medium-, or low-volatile bituminous, semi-anthracite, and anthracite. A, B and C show the regions where P8 (Allegheny County), KS-325 (Somerset County), and KS-121 (Northumberland County) coals are mined.

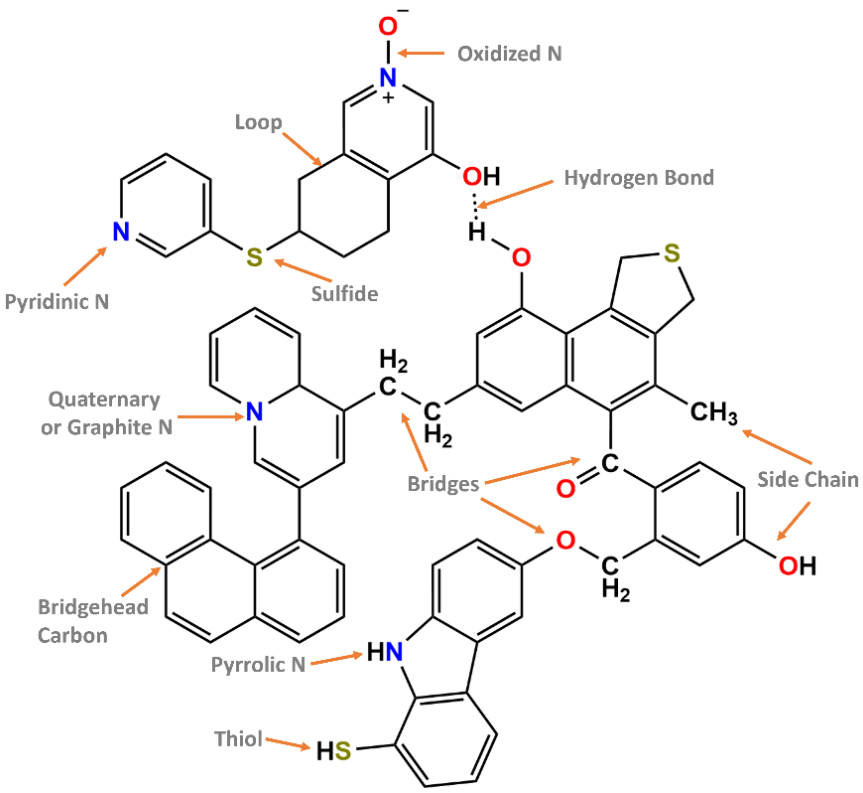


Figure S2. Some representative chemical structures that can be identified using NMR, FTIR and XPS analysis of coal. This structure is then used to create coal models.

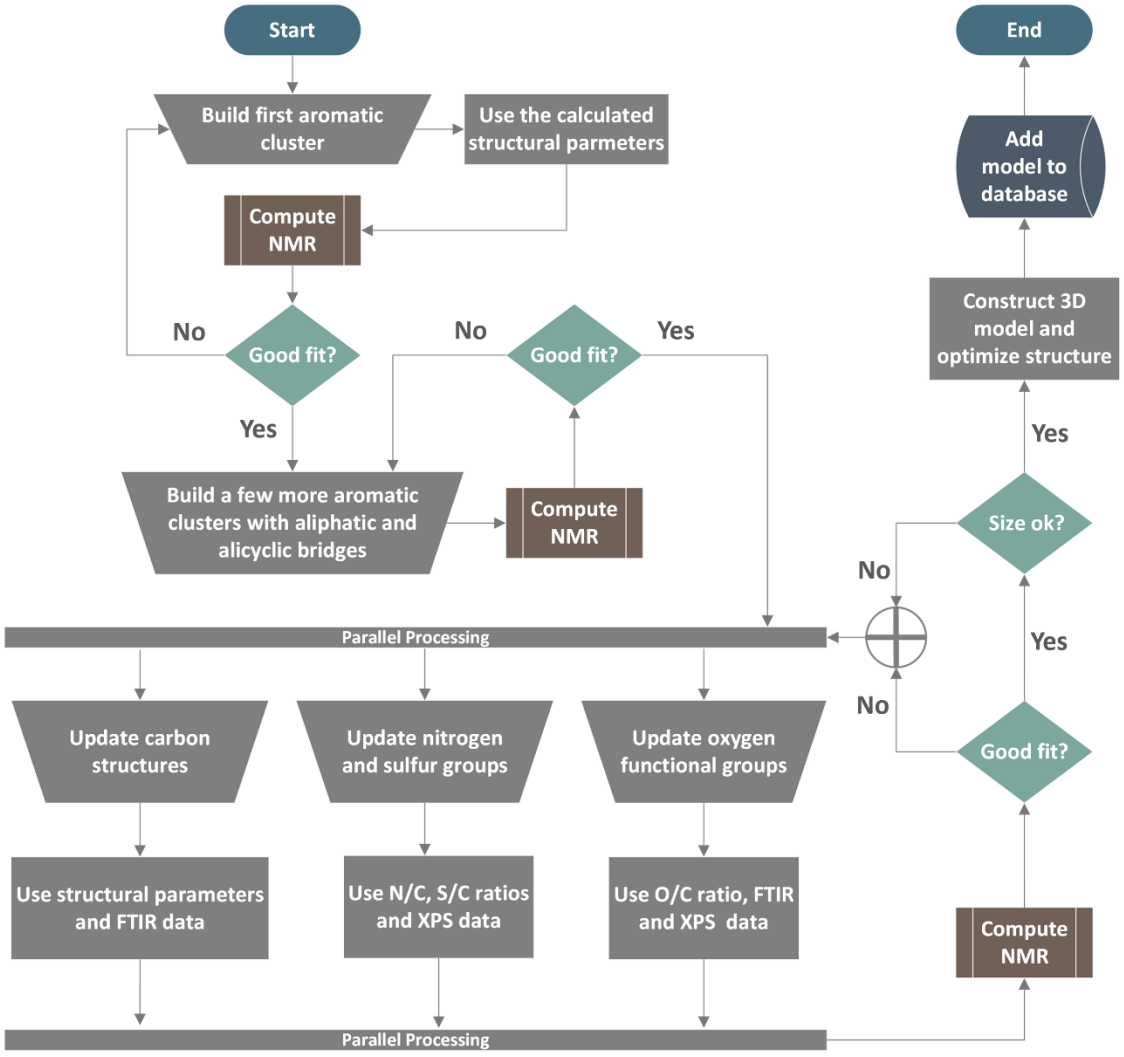


Figure S3. Flow Chart for the bottom-up approach to constructing coal models using experimental data.

Discussion on Figure S3

Our method for creating 2D structural models for coal, as illustrated in the flowchart shown in Figure S3, employs a bottom-up approach. We initiate the process with small carbon skeletal structures containing clusters and aliphatic loops/bridges, progressively integrating non-carbon functional groups as the system size increases. The decision-making process for expanding the model and determining the inclusion or exclusion of functional groups follows a systematic trial-and-error method. This includes comparing calculated NMR spectra from a proposed structure with experimental NMR spectra.

All shapes in the flowchart adhere to their conventional meanings [6]. The `parallel processing` bar signifies that all three manual processes (represented by trapezoids) for adding carbon, nitrogen, sulfur, and oxygen structures

occur synchronously. The `Compute NMR` subroutine (or predefined process) is facilitated by MestReNova®, although ChemDraw® can also be utilized. It is important to note that while the trapezoid shape suggests `manual processes`, scripting within MestReNova® and ChemDraw® enables automation up to NMR calculation and data extraction. Additionally, OriginPro® scripts can be employed to qualify and compare peaks. However, human intervention remains necessary for structural updates and to ensure the model aligns with desired chemistry.

Conditional blocks connected to the `logical OR` ensure the appropriate system size and NMR fit is achieved after iterations before proceeding to 3D model construction using ChemDraw® followed by ReaxFF structure optimization and DFT post-treatment.

For constructing 2D representative coal models from scratch, we offer additional insights to consider. We will augment the manuscript outline, incorporating pertinent points, as necessary. The procedure is as follows:

1. Gather experimental information on aliphatic and aromatic carbon structures, along with various functional groups, obtained using aforementioned methods.
    - The key experiments are NMR, FTIR, XPS and ultimate analysis. If other elemental analysis methods are available those could be useful too. We only used the four methods mentioned in this work.
2. Delineate 2D molecular structures of the coal using a molecular editor, such as ChemDraw® or MestRenova®.
    - From the flowchart, the first step is to build aromatic clusters. For this step, the calculated structural parameters, especially the average number of carbon atoms within aromatic clusters (n_{ci}) and the total molecular weight of the cluster (W_c), are most helpful. They provide information regarding the number of carbon atoms and aromatic rings expected in the cluster.
    - Concurrently, it is beneficial to begin computing the NMR spectra at this stage. This allows for a preliminary understanding of how the calculated NMR peaks correspond to the aromatic structures. The fitting process can be done using OriginPro® or similar tools. Scripting these processes optimizes efficiency, but they can also be performed manually.

3. Compute the NMR spectra for the initial structure, utilizing tools like ChemDraw® or MestRenova®.
    - This step follows the subsequent flowchart process of incorporating additional carbon skeletal structures, including aliphatic and alicyclic bridges. At this stage, the other five calculated parameters prove beneficial, as well as the main structural parameters. It is worth considering developing a straightforward script in a preferred programming language to monitor atom positions and analyze how alterations in atom-atom interactions impact NMR assignments. This monitoring is advantageous because NMR assignments are a localized property. Understanding which structures or functionalities result in incorrect fits can assist in avoiding such issues in future iterations.
4. The model's small volume, compared to real materials, causes the calculated NMR spectrum to appear as spikes at different chemical shifts, resembling δ functions. OriginPro® is employed to combine these spikes and construct a curve mirroring the experimental data, using peak convolution algorithms.
    - Consistency is key when performing deconvolution or spectrum manipulation. For example, when using OriginPro®, it is advisable to process δ functions calculated from NMR using the `convolution` package within the signal processing library. Set the convolution type to circular and ensure to select the entire range of the NMR spectrum (usually 0 ~ 235 ppm). Maintaining uniform settings across all fits ensures consistency throughout the process.
5. Iteratively refine the structure to coincide with the experimental ^{13}C NMR spectrum. Pay attention to the analysis from other experimental methods (FTIR, XPS, etc.), at least considering them as limiting factors.
    - This addresses non-carbon functional groups. As outlined in the flowchart, it is essential to factor in conclusions drawn from other methods, particularly FTIR for oxygen and XPS for nitrogen and sulfur. Additionally, verifying the atomic ratios obtained from ultimate analysis proves beneficial. However, if the ratio from ultimate analysis is insufficient and no other elemental analysis results are available, it is prudent to prioritize FTIR, XPS, and NMR analyses.
6. Constructing and utilizing an ensemble of structurally distinct models is advised.

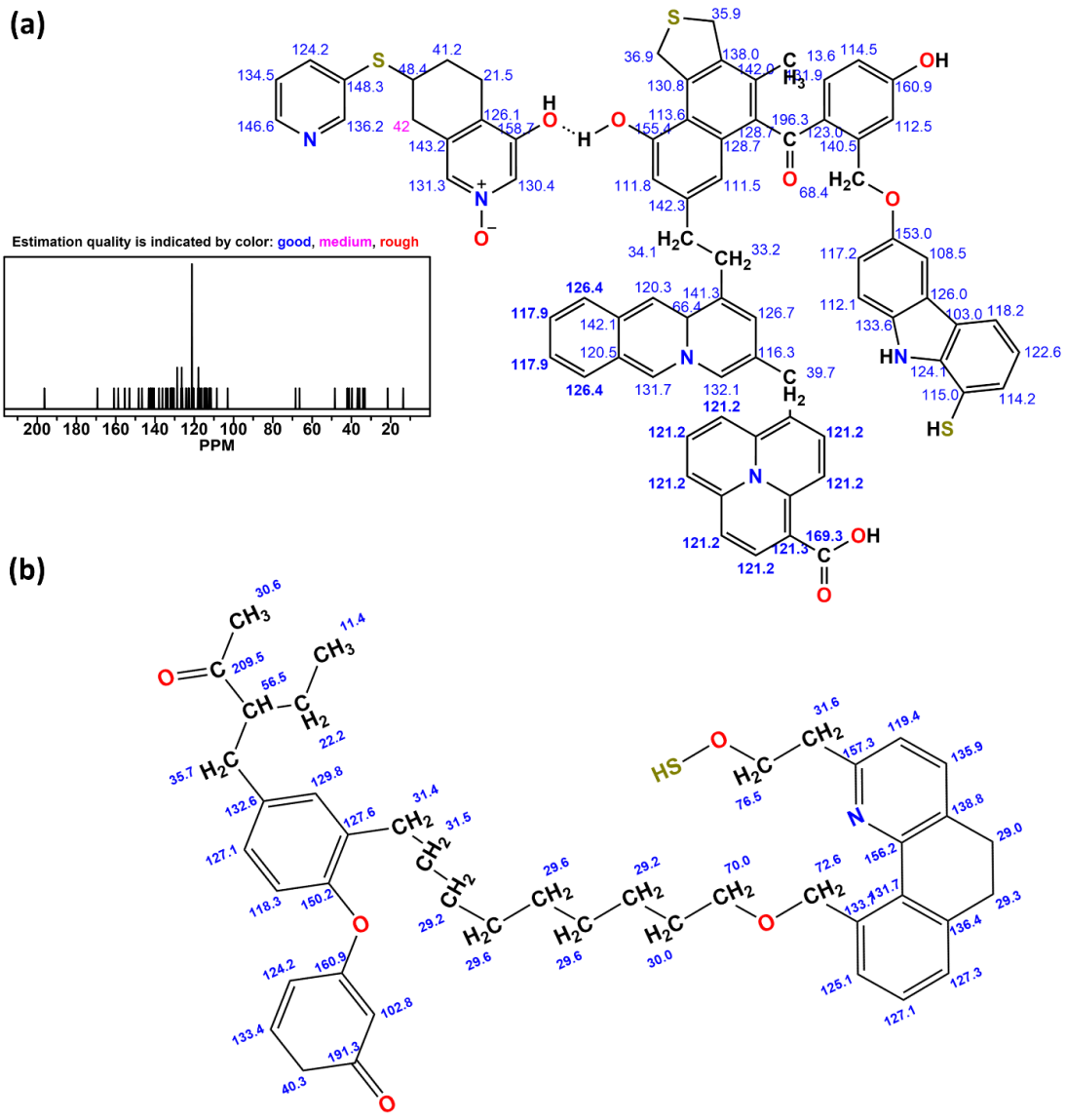


Figure S4. NMR computations were conducted for a representative structure using the ^{13}C ChemNMR package within ChemDraw® molecular editor software. Comparable analyses can be achieved using software such as MestReNova®. The chemical shifts in (b) demonstrate how NMR resolution impacts the prediction of –CH_2– chains in coal. While computed NMR spectra are well-resolved (resembling δ functions), real coal's experimental data appears broader. Hence, resolving the 25 – 36 ppm NMR peak region to determine the length of –CH_2– chains is not straightforward. Experimental NMR provides the average number of CH_2 but doesn't necessarily indicate the length of –CH_2– bridges or chains.

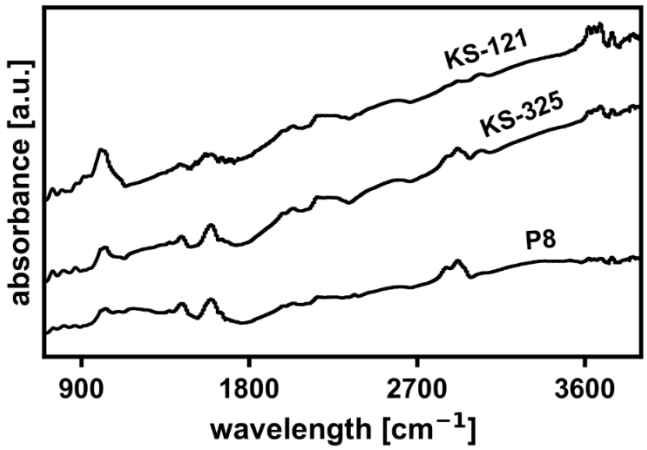


Figure S5. Full ATR-FTIR Spectra for coal samples. This allows for comparison of the intensities of the different regions considered in the manuscript.

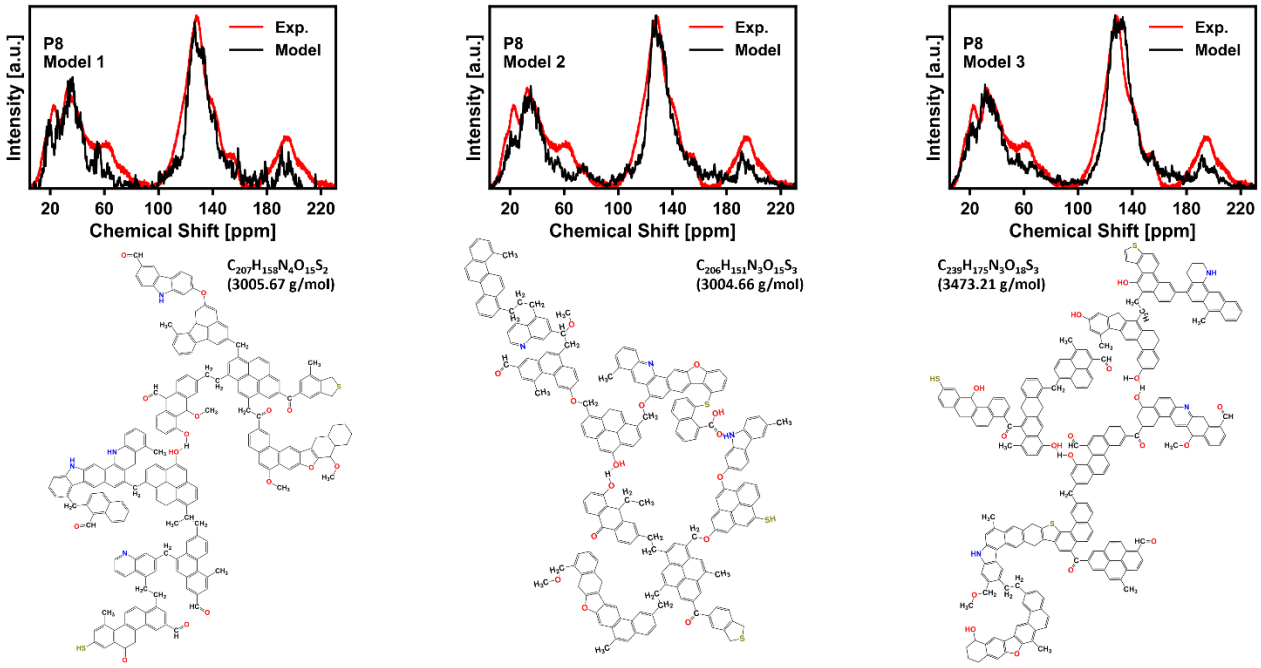


Figure S6. Representative models for P8 coal.

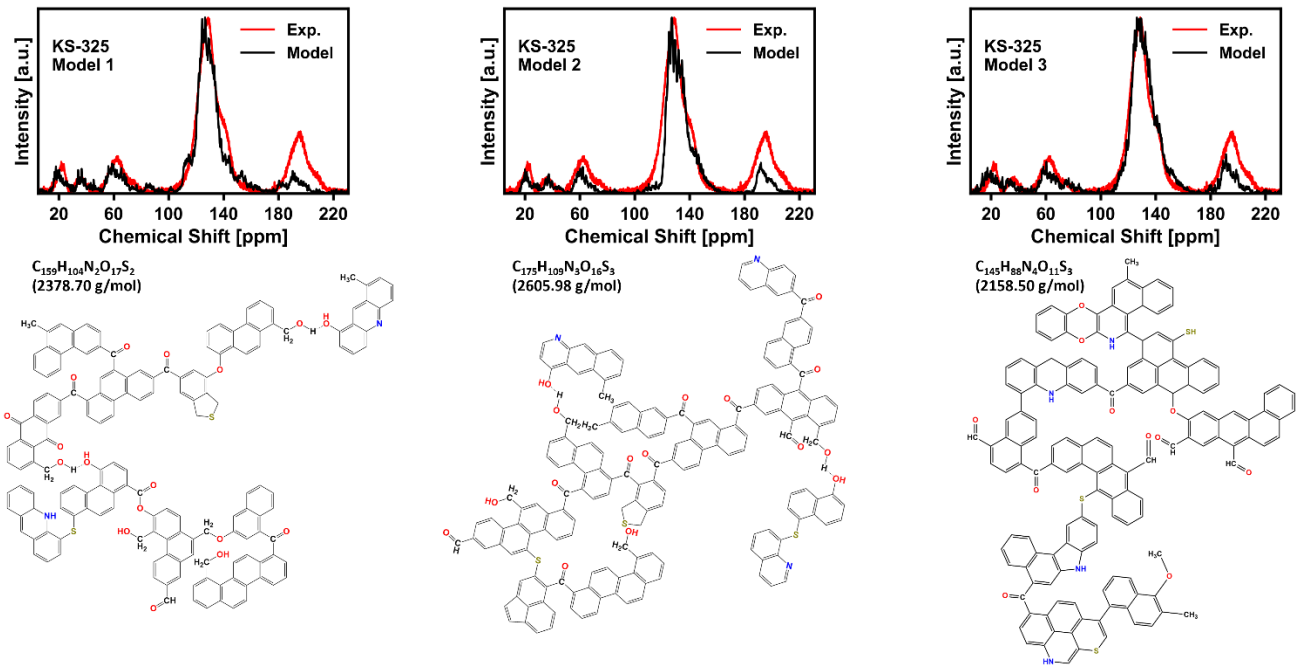


Figure S7. Representative models for KS-325 coal.

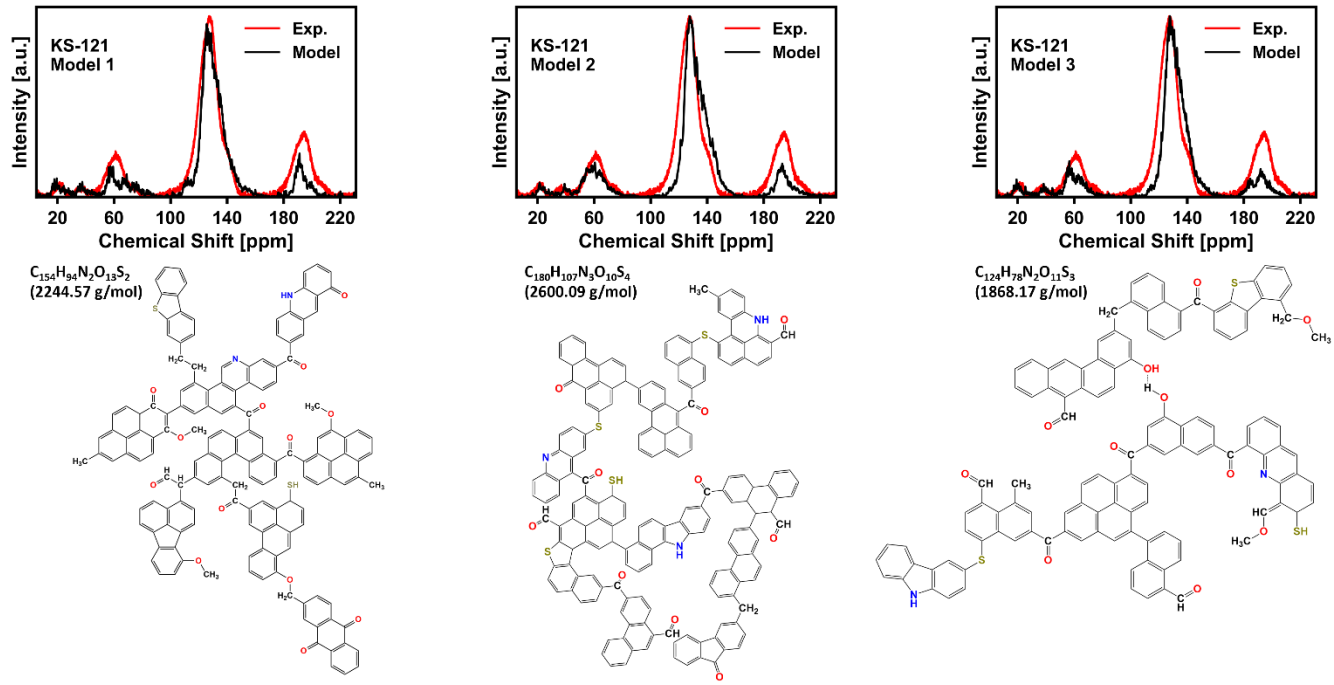


Figure S8. Representative models for KS-121 coal.

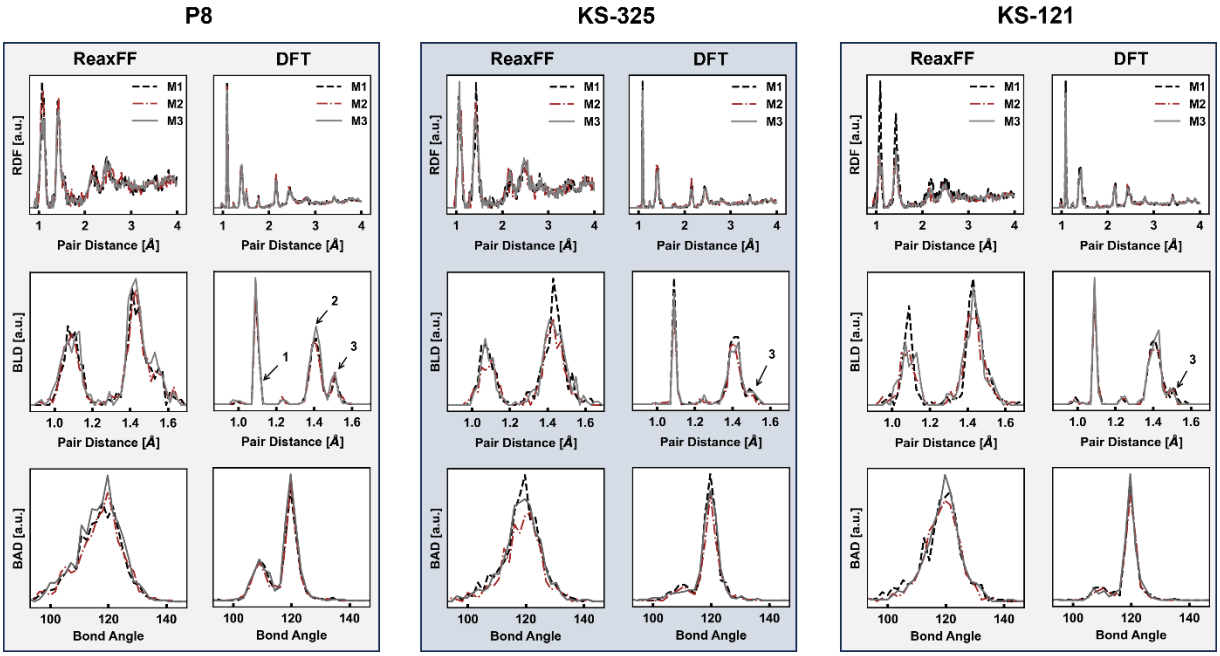


Figure S9. Comparison of structure features among Model 1, 2 and 3, denoted as M1, M2 and M3, for P8, KS-325, and KS-121. Plots are presented for the radial distribution function (RDF), bond length distribution (BLD) and bond angle distribution (BAD). Results are provided for structures obtained after ReaxFF simulation and subsequent DFT post-treatment. While the structural features of the ReaxFF models for all three coals are similar, notable differences emerge in the DFT models. Specifically, the third peak in the bond length distribution of the DFT models exhibits a more pronounced peak "3" (corresponding to C=C bonds [7]) in P8 compared to KS-121 and KS-325. This feature is less prominent in the ReaxFF models. Moreover, the P8 DFT models display more intense bond angles at ~109° compared to KS-325 and KS-121. This peak is attributed to contributions from C–C–H and H–C–H bond angles [8]. The atomic species contributing to the differences in BLD and BAD are consistent across all coals and are depicted in the atom-projected BLD and BAD, labeled as P-BLD and P-BAD, as shown in shown in Figure S10.

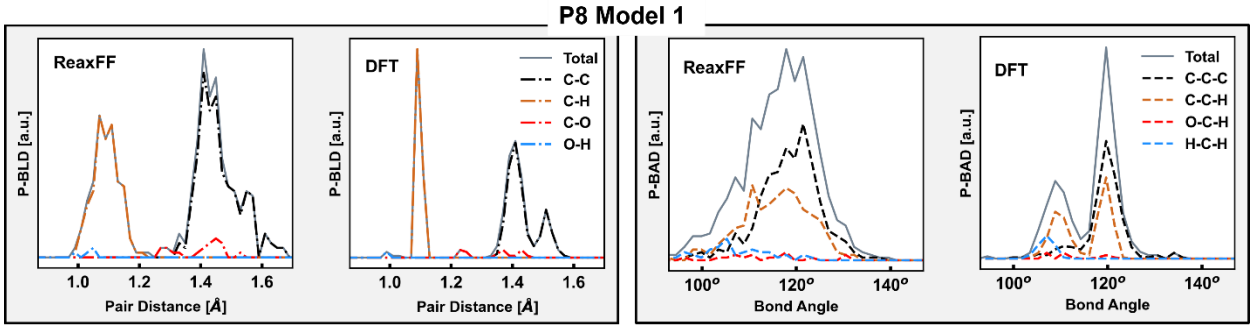


Figure S10. Atom-projected BLD (P-BLD) and BAD (P-BAD) of the ReaxFF and DFT models for Model 1 of P8. The atomic species contributing to the structural features are identified.

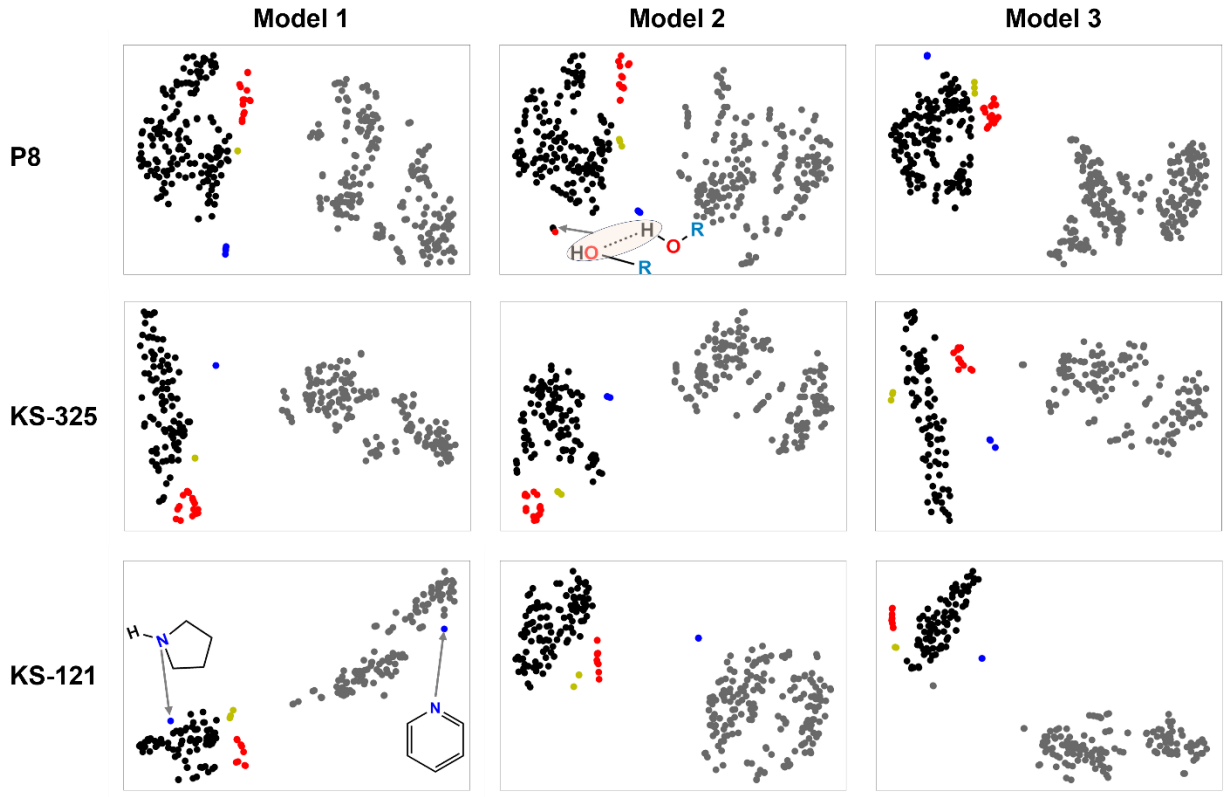


Figure S11. Configurational subtleties within coal structures are rendered through t-SNE. The SOAP 10,350 descriptors are embedded into a map with 2 unitless t-SNE co-ordinates. Hydrogen, carbon, nitrogen, oxygen, and sulfur are colored black, gray, blue, red, and gold, respectively. The chemical structures in selected panels depict the atomic environments corresponding to the indicated data points.

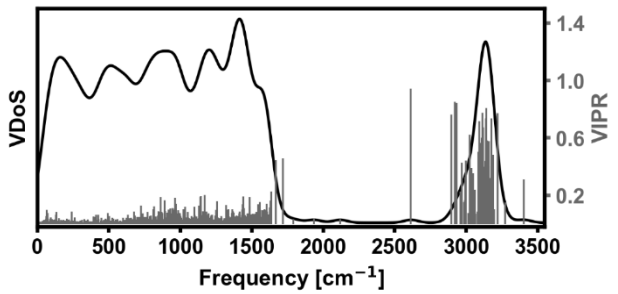


Figure S12. The vibrational density of states (VDoS) and its corresponding inverse participation ratio (VIPR) for Model 1 of the KS-121 models (KS121_M1), discussed in Figure 9 of the manuscript. This shape profile is consistent across all the models (see Figure S13 for comparison with Model 3).

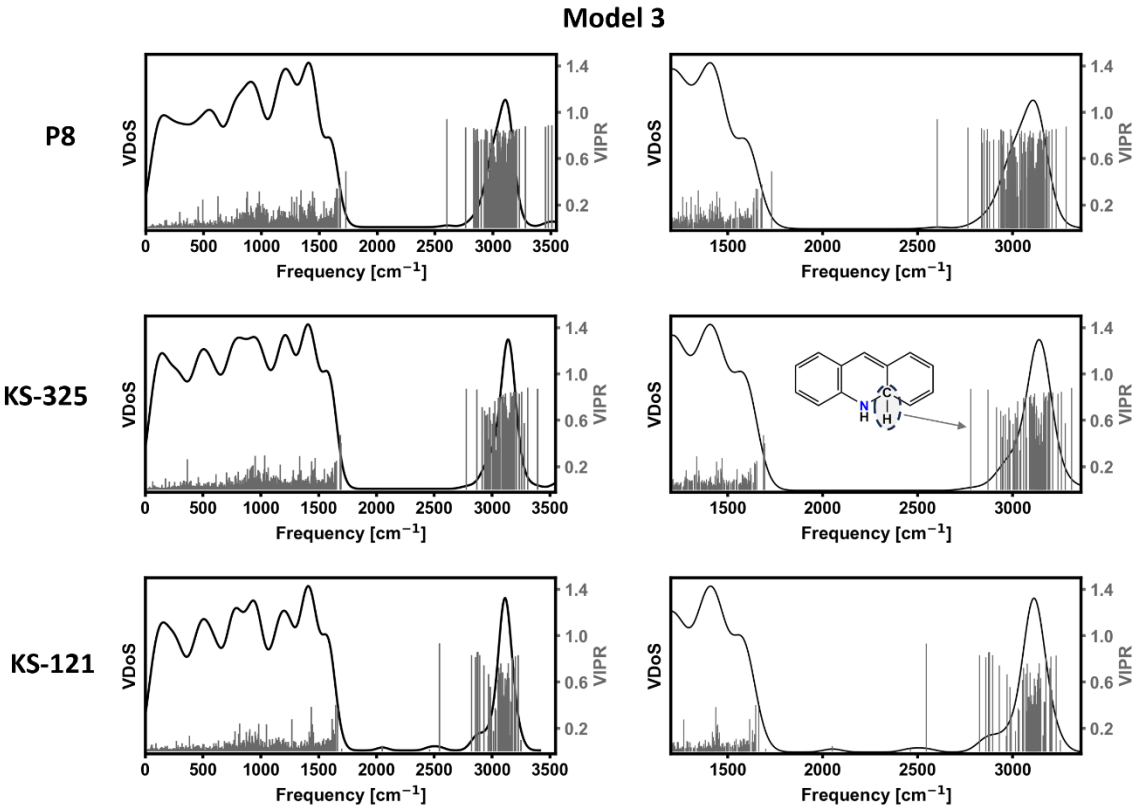


Figure S13. The plot for the VDoS and VIPR for Model 3 of all the coals is presented. A consistent overall shape profile is observed. This pattern persists even in Model 3 (not depicted). However, a new vibration mode, C-H bond for carbon bonded to the nitrogen in an acridine structure, is observed at ~2798.4 cm^{-1} for the KS-325 model. The environment responsible for the rest of the high VIPR values are same as discussed for Model 1 of KS-121 in the manuscript.

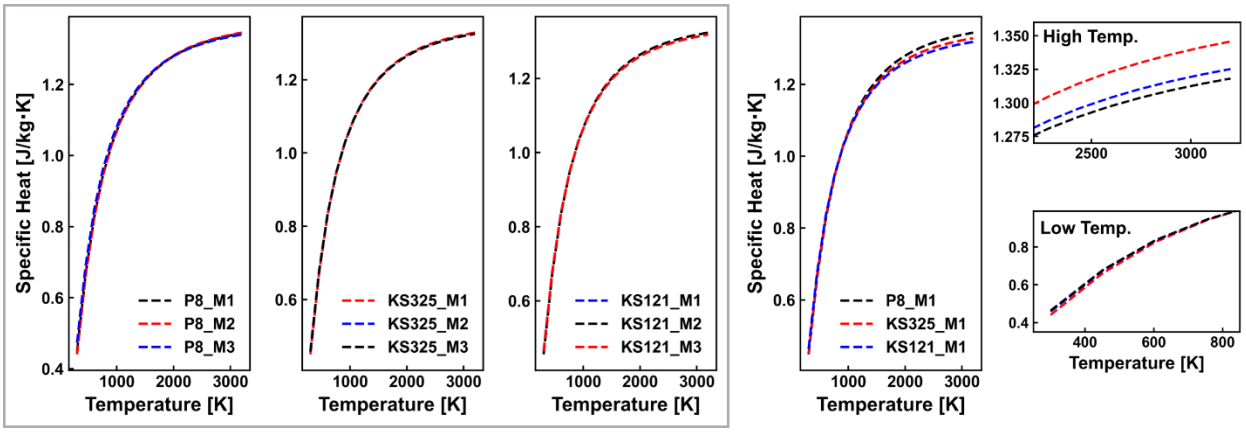


Figure S14. Specific heat capacity for all models, plotted as a function of temperature (300 to 3200 K). Within the same rank, models exhibit closely aligned values, while slight differences emerge when comparing models from different coal ranks. At high temperatures, distinctions in specific heat capacity values are observed, with P8 having the highest values, followed by KS-235 and KS-121, as demonstrated using Model 1. Conversely, at lower temperatures, the specific heat values among all models closely converge.

References

[1] M.S. Solum, R.J. Pugmire, D.M. Grant, Carbon-13 solid-state NMR of Argonne-premium coals, Energy Fuels 3 (1989) 187–193. https://doi.org/10.1021/ef00014a012.

[2] M.S. Solum, R.J. Pugmire, Coal liquefaction process streams characterization and evaluation: 13C-NMR analysis of CONSOL THF-soluble residual materials from the Wilsonville coal liquefaction process, CONSOL, Inc., Library, PA (United States); Utah Univ., Salt Lake City, UT (United States), 1992. https://doi.org/10.2172/6782684.

[3] L.J.P. van der Maaten, G.E. Hinton, Visualizing high-dimensional data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

[4] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 374 (2016) 20150202. https://doi.org/10.1098/rsta.2015.0202.

[5] Pennsylvania Geological Survey, Distribution of Pennsylvania coals (3rd ed., revised), (2000). https://maps.dcnr.pa.gov/publications/Default.aspx?id=718 (accessed November 30, 2023).

[6] L.A. Schultheiss, E.M. Heiliger, Techniques of flow-charting, Clin. Libr. Appl. Data Process. 1st 1963 (1963). https://hdl.handle.net/2142/743 (accessed March 9, 2024).

[7] C. Ugwumadu, R.O. III, N.L. Smith, K. Nepal, Y. Al-Majali, J. Trembly, D.A. Drabold, Computer simulation of carbonization and graphitization of coal, Nanotechnology 35 (2023) 095703. https://doi.org/10.1088/1361-6528/ad1058.

[8] H.A. Bent, Molecules and the chemical bond: An Introduction to conceptual valence bond theory, Trafford Publishing, 2013.