Predicting Isotopologue Abundances in the Products of Organic Catagenesis with a Kinetic Monte-Carlo Model

Hao Xie¹*, Michael Formolo², John Eiler¹

¹Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
(*email address: hbx5051@psu.edu)
²ExxonMobil Upstream Integrated Solutions Spring, TX, USA

Abstract

Thermal cracking of complex organic matter can occur under many geological settings. However, the role of thermal cracking vs. other chemistries (e.g., metallic catalysis or Fischer-Tropsch-type reactions) in petroleum formation remains controversial. Realistic modeling of isotope effects in chemical reaction networks involving thermal cracking might shed light on this problem, especially given the recent progress on measurements of intramolecular stable isotope distributions in organic compounds. Previously published models of thermal cracking in petroleum formation are incapable of predicting the intramolecular isotope patterns of products because they do not incorporate realistic precursors, elementary reactions and patterns of inheritance. In this study, we develop a kinetic Monte-Carlo (kMC) model to address this problem. We simulate thermal breakdown of different types of organic matter that is represented by molecular models. At the onset of each simulation, we initialize the model parent organic molecules with isotopic substitutions, and then subject them to free radical chain reactions in a many-step process. Every simulation captures a possible route of thermal degradation and tallies the numbers of each unique isotopologue of all product molecules at the end. Although this model produces data that contain information for all molecules and isotopic forms, in this study we focus on the proportions of many of the isotopologues of all of the C₁⁻C₇ n-alkanes. We use two chemistry schemes that differ in complexity. The basic scheme (scheme A) includes only homolytic cleavage and capping of metastable radicals by hydrogen atoms. The more sophisticated scheme (scheme B) includes most reactions of importance in the free radical chain mechanism of thermal cracking. We find significant differences between the schemes A and B in predicted molecular and isotopic compositions of products. Scheme B is more consistent with natural data than scheme A, suggesting that full radical mechanisms should be considered in models of thermal cracking in natural hydrocarbon formation. Our model is further validated by reproducing results from hydrous pyrolysis experiments. For simulating natural petroleum formation, we found that the kMC model is acceptable at low maturity, but cannot match compositions and time estimates of mature gas formation, suggesting the influence of alternative mechanisms in late-stage, high-thermal-maturity
catagenesis. Using our model, we provide mechanistic explanations for some of the existing observations, such as the evolution of intramolecular and intermolecular carbon isotope compositions with thermal maturity. Our study also makes predictions of intramolecular isotope compositions for higher-order alkanes (C$_4$+) that have been little studied to date but present attractive targets for future measurements.

1. Introduction

Thermal cracking, the decomposition of organic molecules via bond cleavages and associated free radical reactions at elevated temperature, has been investigated for nearly a hundred years, in part because it was used to produce valuable petrochemical feedstocks in the oil refinement industry (Rice, 1931; Rice et al., 1932; Sung et al., 1945; Sadrameli, 2015). In natural systems, thermal cracking can potentially occur when organic matter is exposed to heat, such as during burial of sedimentary rocks (Tissot and Welte, 1978), hydrothermal activity at active oceanic spreading centers (Kvenvolden et al., 1986; Kawka and Simoneit, 1994; Simoneit and Fetzer, 1996), and accretion of planetary or meteoritic bodies (e.g. carbonaceous condrites; Huss et al., 2006). While it is widely accepted that thermal cracking is responsible for transformation of sedimentary organic matter to release petroleum during catagenesis, experimental studies of thermal cracking — the foundation of our quantitative understanding of the process — are generally performed at temperatures >300°C and timescale of minutes to hours, whereas natural catagenesis mostly takes place at temperatures <250°C and timescales of Myr to 100Myr. The hot and fast process of pyrolysis experiments promotes highly kinetically controlled cracking mechanisms, and it is questionable whether this chemistry takes place in the same way at natural catagenic conditions. Some previous studies have noted this difference between laboratory models of thermogenic petroleum formation and natural catagenesis, and have suggested that this difference in conditions and fundamental chemical processes might explain differences in chemical composition between experimental and natural products (e.g., in their methane and olefin contents). These studies challenge the thermal cracking paradigm of petroleum formation and propose that chemistry of organic catagenesis is catalytic and/or occurs via quasi-equilibrium processes (Mango, 1992; Helgeson et al., 1993; Mango et al., 1994; Corker et al., 1996; Helgeson et al., 2009). Reconciling these contradictions and providing a detailed mechanistic description of organic catagenesis are of great importance.

In this study, we aim to address these challenges by creating an explicit model of thermal cracking reactions that predicts chemical and isotopic properties of products as functions of the mechanisms, conditions and extents of cracking different source substrates, and by comparing these model predictions with measured properties of experimental and natural petroleum products. We use a statistical modeling
approach, the kinetic Monte Carlo method, to simulate free radical reactions of cracking with realistic depiction of elementary reactions and molecules. We embed the naturally occurring stable isotopes in the molecular representations of our model and describe the effects of these isotopes on reaction rates. This allows the outcome to include not only compositional and molecular average isotope ratios, but also isotopomer abundances, the importance of which will be detailed in the next section. We constrained our model with kinetic parameters for elementary reactions, including isotope effects on reaction rates, from established databases, which provide a basis for comparing independent, quantitative predictions of measured chemical and isotopic properties of experiment products. This set of models can be used to identify the atomistic mechanisms of cracking reactions and to explore the correlation between compositional and isotopic distributions of hydrocarbon products and the molecular structures of their precursors. We can also track the evolution of hydrocarbon chemical and isotopic composition with thermal maturation, linking model predictions to geological and planetary settings and processes.

2. Modeling Isotopomer Abundances of Catagenic Products

Stable isotope ratios of hydrocarbons are controlled by isotope ratios of precursor molecules and the chemical, physical and biological isotope effects associated with formation, transportation and destruction; for these reasons, they record valuable information regarding the origins and fates of these compounds. For several decades, bulk (material-average) and compound-specific (molecule-average) stable isotope compositions have been measured on geological hydrocarbons in efforts to understand their sources, thermal maturities, extents of biodegradation and other issues (Silverman and Epstein, 1958; Galimov, 1975; Schoell, 1980; Chung et al., 1988; Whiticar, 1996; Peters et al., 2004; Boreham et al., 2008; Dai et al., 2012). More recently, advances in several different analytical technologies have enabled study of abundances of individual isotopologues (i.e., compositionally and structurally unique isotopic forms) of hydrocarbon molecules. These measurements are often categorized into two types that measure different distinctive properties of isotopologues: position-specific and multiply-substituted. Such measurements can be approached using several different technologies, including Nuclear Magnetic Resonance (NMR), High-Resolution Mass Spectrometry (HR–MS; using both sector and Fourier-transform instruments), Infrared (IR) Absorption Spectroscopy, Gas Chromatography – Pyrolysis – Gas Chromatography – Isotope Ratio Mass Spectroscopy (GC-p-GC-irMS) and chemical or biochemical degradation followed by Isotope Ratio Mass Spectrometry (Gilbert et al., 2013; Stolper et al., 2014; Ono et al., 2014; Julien et al., 2016; Gilbert et al., 2016; Young et al., 2016; Piasecki et al., 2016; L. Gao et al., 2016; Clog et al., 2018; Li et al., 2018; Liu et al., 2018; Cesar et al., 2019; Gonzalez et al., 2019).
Interpretation of these new properties of isotopic distribution can help better trace the origins and fates of subsurface hydrocarbons.

Previous published models describing kinetic isotope effects associated with catagenesis generally do not predict abundances of specific isotopologues of hydrocarbons produced from thermal cracking of organic precursors (exceptions include previous models of the isotopologues of methane (Xia and Gao, 2019; Dong et al., 2021) and ethane/propane (Peterson et al., 2018)). However, there is a relatively extensive and longstanding body of quantitative models that simulate oil and gas generation from thermal cracking of kerogen (Sweeney et al., 1987; Ungerer, 1990; Burnham and Braun, 1990; Behar et al., 1997; Lorant et al., 1998; Cramer et al., 1998, 2001). These models typically represent a group of compounds by a general term (e.g., kerogen; oil; condensate; C1–C5 gas; coke). They then abstract cracking chemistry to a few ‘conversion’ reactions of those grouped compounds. Kinetics of these reactions can be parameterized by empirical distributions of activation energies that are fitted to high temperature cracking experiments. These models of reaction kinetics could be combined with quantum chemical calculations of associated isotope effects (Galimov, 1975; Tang et al., 2000; Xiao, 2001; Tang et al., 2005) to model compound-specific isotope ratios of petroleum components. However, significant additional development would be needed to modify or upgrade these models to describe production of specific isotopologues of generated hydrocarbons because this would require explicit descriptions of the rates of elementary reactions for specific molecular and isotopic structures of precursors.

Describing the full reaction paths from kerogen precursors to hydrocarbon products at the molecular level, where isotopic structure of precursor and product is specified and directly connected to reaction rate, demands an enormous amount of computational memory and processing power, beyond the capabilities of computers and programming during the period when the most well developed models of catagenesis were created (Ungerer, 1990; Behar et al., 1997). But, more recent kinetic modeling of explicit molecules and elementary reactions has been advanced to simulate more and more complicated reaction networks (Savage, 2000; You et al., 2009; Harper et al., 2011; C. W. Gao et al., 2016), especially for the systems of pyrolysis and combustion. These models represent the concentrations of species as variables and formulate relevant chemical reactions as components of ordinary differential equations of these variables. The differential equations are solved numerically (unless the model is simplistic enough for analytical solution) to provide time-evolving concentrations of all considered species. Ideally, we could follow this approach and introduce isotope labels to make these models describe the creation and destruction of specific isotopologues. However, there are two reasons that the methodologies used in these previous approaches of the kinetics of combustion can’t undertake the problem of isotope chemistry of thermal cracking of macromolecular substrates. First, molecular models of kerogen are much bigger
and more complex than the substrates that prior kinetic models typically consider. Attempts to model cracking of natural materials still often requires arbitrary rules to restrict molecular and reactional complexity (Dominé et al., 2002; Freund et al., 2007; Walters et al., 2007). Second, specification of isotopologues and kinetic isotope effects increases the demand of computational resources dramatically. Concentrations of isotopologues need to be tracked and stored as variables, and the number of these variables will be too large to be handled by any computer systems. For example, we can estimate the number of isotopologues for a kerogen molecule (typical substrate for catagenesis models) that has 5,000 carbon atoms (Bousige et al., 2016). Even if we only consider versions of this substrate molecule containing 0 to 3 $^{13}$C-substitutions (clearly a simplification, given that on average ~50 $^{13}$C’s will be present in each formula unit), there will be $2 \times 10^{10}$ symmetrically non-equivalent isotopic forms of that substrate molecule. For reference, simulation of the carbon isotope effects associated with pyrolysis of propane (6 carbon-isotope isotopologues in total) takes 44 core hours (Goldman et al., 2019).

Here we use a different approach to kinetic modeling, a statistical algorithm known as the kinetic Monte Carlo method. The strength of this method is that it circumvents the obstacle of storing and modifying the concentrations and reactions of astronomical numbers of species. Instead, it initializes the system with a relatively large unit of molecular structures composed of atoms (with isotopic information labeled) and bonds that connect atoms. It allows the system to evolve via a stepwise execution of reactions, where a new reaction is stochastically selected from an automatically generated reaction list at each step, based on the permitted chemistry of the system at its current state. Consequently, all isotopic forms of substrates and intermediates can be accounted for by randomized events, and we only extract information of interest (molecular and isotopic composition of select product compounds) from the system when the evolution ends.

### 3. The Kinetic Monte Carlo Model

The kinetic Monte Carlo (kMC) method was developed in the 1960s and 1970s to provide a numerical solution for systems that evolve with time (Voter, 2007). In this section, we review the concepts and development of the kMC method, and demonstrate how we adapt it for isotope effects in thermal cracking.

The master equation for the evolving system can be formulated as (Fichthorn et al., 1991; Chatterjee and Vlachos, 2007):

$$\frac{\partial Pr(\sigma, t)}{\partial t} = \sum_{\sigma'} W(\sigma' \rightarrow \sigma, t) Pr(\sigma', t) - \sum_{\sigma} W(\sigma \rightarrow \sigma', t) Pr(\sigma, t) \tag{1}$$
Where \( Pr(\sigma, t) \) denotes the probability that the system (or an element of the system) is in state \( \sigma \) at time \( t \), and \( W(\sigma \rightarrow \sigma', t) \) is the probability per unit time (probability density function) of the system transitioning from \( \sigma \) to \( \sigma' \), and vice versa. In this system, both the next state and transition probabilities depend on the previous state. When dealing with a complex system having many interacting mechanisms, equation (1) cannot be solved analytically or by deterministic differential equation methods. The kinetic Monte Carlo method provides a feasible numerical solution to this equation via sampling possible paths and states stochastically. This method simulates the reactive system with individual atoms and bonds rather than the ‘concentration’ or ‘activity’ descriptions in deterministic formulations. The kMC method tracks a step by step evolution of the system, where it discretizes time into steps of variable duration that jump the system from state to state. If the object of study progresses by chemical reactions, a jump is selected amongst all probable chemical reactions at that given time. If all probable reactions could be enumerated, the probability of a certain reaction happening in the next step is (Gillespie, 1976):

\[
P(\tau, \mu) = W_\mu \ast \exp \left( -\sum \nu W_\nu \tau \right) \quad (2)
\]

\( P(\tau, \mu) \) indicates the probability density (over time) that the next reaction would happen between times \( \tau \) and \( \tau + d\tau \) and would be reaction \( \mu \), where \( \mu \) is an integer indexing the reaction number. \( W \) is the stochastic formulation of reaction rates in deterministic chemical kinetics. Using basic principles of probability, it could also be written as:

\[
P(\tau, \mu) = P(\tau) \ast P(\mu | \tau). \quad (3)
\]

Here \( P(\tau) \) is the probability density that the next reaction would happen between times \( \tau \) and \( \tau + d\tau \) and \( P(\mu | \tau) \) is the probability of that reaction being reaction \( \mu \), given that it is happening between \( \tau \) and \( \tau + d\tau \). Since the probability of more than one reaction happening over the time span of \( d\tau \) is \( o(d\tau) \), we could use the addition theorem of probabilities to get:

\[
P(\tau) = \sum \nu P(\tau, \nu). \quad (4)
\]

Combining equation (3) and (4) gives

\[
P(\mu | \tau) = \frac{P(\tau, \mu)}{\sum \nu P(\tau, \nu)}. \quad (5)
\]

Substituting \( P(\tau, \mu) \) and \( P(\tau, \nu) \) from equation (2) yields:

\[
P(\tau) = W_{tot} \ast \exp (-W_{tot} \tau) \quad (6)
\]
\[ P(\mu | \tau) = \frac{W_\mu}{W_{\text{tot}}}, \quad (7) \]

where \( W_{\text{tot}} = \sum W_v \). Equation (6) and (7) provides the method for propagating chemical reactions. Equation (6) gives the probability density function of time for the next reaction, which allows a passing time length (time interval between the current reaction and the next reaction) to be generated randomly for a given reaction step. This means that the discretized series of events in the evolution could be projected to the continuous time vector. Equation (7) indicates that the probability of a reaction to be selected is proportional to its rate (stoichiometric rate with unit of time\(^{-1}\)).

Based on equation 6, the time passed to a kMC time step can be estimated with the exponential function using a random number \( p_2 \) : \( \tau = \ln \left( \frac{1}{p_2} \right) / \sum W_v \), where \( p_2 \in (0, 1) \) and is uniformly distributed, and \( \sum W_v \) is the summation of rates of all possible reactions (Fichthorn et al., 1991; Peterson et al., 2018).

Integrating this function over the possibility range yields that the average time passed for such a step is \( \tau = 1/\sum W_v \). Therefore, we calculate the time increments per step with the average time equation and constrain the time flow during a kMC simulation.

Peterson et al. (2018) first applied the kMC method to isotope fractionation in thermal cracking, where they studied carbon isotope fractionation of cracking \( n\text{-C}_{18} \) alkane (octadecane). In their model, the chemistry scheme is simplified to C-C bond cleavages only. Here we extend this concept to explore isotope effects involving multiple elements in thermal cracking of different substrates and with multiple reaction types. Specifically, we advance the notion of a thermal cracking kMC model in four ways: First, we consider a variety of molecular models, including structural models of kerogens of different types and oil compounds as substrates. Second, we use reaction schemes that are more complete and realistic than homolytic cleavage of C–C bonds alone. Third, we implement accurate kinetic parameters, including rate constants of elementary reactions and KIE, derived from existing \textit{ab initio} and empirical calculations. Finally, our model considers both \(^{13}\text{C}\) and D substitution and provides abundances of site-specific and multiply-substituted isotopologues for \( \text{C}_1\text{-C}_7 \) product molecules. Note that Xie et al. (2020) present a preliminary and simplified version of this model to calculate position-specific hydrogen isotope ratios of propane produced by catagenesis of kerogen and oil compounds.

### 3.1. Initialization

Structural modeling of kerogen and coal has a long history; the earliest study to hypothesize a representative molecular structure of coal was from the 1940s and of kerogen from the 1960s (reviewed by Vandenbroucke and Largeau, 2007). However, these works rely on limited information provided by bulk organic matter analysis and pyrolysis and oxidative degradation experiments. The emergence and
improvement of other analytical techniques (e.g., infrared spectroscopy, nuclear magnetic resonance, X-ray) improved chemical characterization of kerogen by providing richer information on bonding environments and moieties (Ganz and Kalkreuth, 1987; Ibarra et al., 1996; Kelemen et al., 2007; Salmon et al., 2011). Moreover, advances in computational chemistry have enabled calculation of mechanical and thermodynamic properties of very large molecular models (Ungerer et al., 2015). Combinations of various analytical observations and bottom-up molecular simulations result in more realistic molecular models. In this study, we use a model of an immature type I kerogen, kIA, which was developed from Ungerer et al. to simulate Green River Shale kerogen.

In addition to these models of kerogens and coals, we also created a model of mixed alkanes to mimic the saturated-hydrocarbon fraction of crude oil. The molecular model of saturated oil contains C_{10}–C_{30} n-alkanes and pristane and phytane. Further details regarding the composition of this model can be found in the Appendix A. Supplementary Material.

In most of the calculations we present, we combine (repeat) many units of the molecular model in each system such that it contains around 10^{5} to 10^{7} carbon atoms. \textsuperscript{13}C and \textsuperscript{2}H substitutions are then populated across the structure. A random number is generated for each position to decide whether an isotope label will be assigned. The probability of having a \textsuperscript{13}C substitution for a carbon position equals to the \textsuperscript{13}C abundance (as a fraction of all carbon atoms) in the initial substrate. For hydrogen isotopes, the symmetry number ratio of that position and \textsuperscript{1}H abundances for non-substituted atoms are also multiplied (e.g., the probability of a methyl group containing \textsuperscript{2}H will be 3F_{2H}(1-F_{2H})^{2}). We assume that the isotopes in the precursor are distributed randomly, without position-specific preferences (equal probability of isotopic labeling across all molecules and positions). This likely is not a realistic assumption, as we know that some biomolecule precursors of kerogen have heterogeneous stable isotope distributions (Abelson and Hoering, 1961; Monson and Hayes, 1980; Hayes, 2001). However, we are not aware of any measurements of natural materials, models or experiments that attempt to associate stable isotope compositions with structural moieties in kerogen molecules, so it is not possible to devise a credible mapping of the intramolecular isotope compositions of kerogens without substantial new work. It is also possible that some of the heterogeneity in isotopic compositions of biomolecules does not lead to systematic position-specific isotopic variations in kerogens or oil compounds simply because the latter are statistically averaged across many precursors sampled through many intermediate reactions. For example, the previous model of Xie et al. (2020) shows that differences in \delta^{13}C between odd and even carbon sites in \textit{n}-C_{17} alkane precursors have little impact on position-specific isotope values of propane formed by thermal cracking (Xie et al., 2020). Yet we acknowledge that this problem is understudied and encourage future research to provide better constraints on internal isotope distributions in kerogen (e.g., see Keppler
et al., 2004, 2007; Lloyd et al., 2021 for studies of $^{13}$C and D distribution in lignin). We assign water with a typical formation water isotope value of $\delta^2$H=-20 ‰ (Engle et al., 2020). In the next few subsections, we detail the organization of our kMC model, which is also visualized in Figure 1.

### 3.2. Chemistry schemes

Despite the overwhelming complexity of the chemistry of thermal cracking, it is possible to reduce it into schemes consisting of a limited number of elementary reaction types that act on one of the many structurally unique sites in the precursors. The most common description of thermal cracking processes in the field of stable isotope petroleum geochemistry assumes that homolytic cleavage of single bonds dominates the formation of low molecular weight alkanes (Chung et al., 1988; Tang et al., 2000; Galimov, 2006). Considering only homolytic cleavage is parsimonious, yet it has been widely applied to explain molecular and isotopic compositions of natural gases (Waples and Tornheim, 1978b; Tang et al., 2000). More sophisticated analysis of cracking chemistry identifies several types of reactions to be important: (1) homolytic cleavage, (2) β-scission, (3) H-abstraction, (4) radical recombination, and (5) radical isomerization (Ranzi et al., 1995; Xiao, 2001; Yuan et al., 2011). In this study, we explore thermal cracking models with two chemistry schemes: In scheme A, molecules break down exclusively via homolytic cleavage, and all radicals are immediately capped with H after cleavage (yielding long-lived stable products). In scheme B, we attempt to depict free radical chemistry more fully by implementing several of the most important reaction types. Figures 2 and 3 and the following paragraphs detail these two schemes:

#### Scheme A

Under scheme A, each reaction ‘event’ is a single bond dissociation (homolytic cleavage). All single bonds between non-hydrogen atoms are considered, including those with non-carbon atoms. Each bond dissociation event generates two radicals that are each subsequently capped with a H atom. This H atom is assumed to be derived from water (via a H-abstraction mechanism).

#### Scheme B

We devise a free radical reaction network in scheme B, which includes six reaction types that can be broadly categorized into initiation, propagation, and termination reactions.

1. **Initiation**: homolytic cleavage
2. **Propagation**: β-scission, radical isomerization, H-abstraction
3. **Termination**: radical recombination.
Examples of theses reaction classes are shown in Figure 2. Reactions of homolytic cleavage in scheme B are treated in the same way as scheme A, except that product radicals are not quenched immediately by hydrogen capping; instead, radicals produced by homolytic cleavage are labeled in the model and subject to radical reactions in the next model time step. β-scission is the radical decomposition reaction where a second nearest (‘β’) ‘C—C’ single bond away from the radical position dissociates, and the nearest ‘C—C’ single bond becomes a double bond (e.g. 1-butyl→ethylene+ethyl). Radical isomerization reactions are intra-molecular hydrogen migration reactions. A H atom on an aliphatic chain containing an initial radical carbon site migrates to cap that initial radical, forming a new radical carbon site where that hydrogen originated (e.g. 1-octyl→5-octyl). A H abstraction reaction occurs when a radical molecule abstracts an H atom from another carbon position or water, resulting in a transfer of radical (e.g., 1-ethyl+octane→ethane+2-octyl). Radical recombination is the combination reaction between two radicals (e.g., ethyl+ethyl→butane).

The reaction networks of schemes A and B are displayed in Figure 3. Note that selected reaction types under scheme B represent those that are most relevant to alkane formation, while the rest of the free radical mechanisms are implicit. For example, we do not consider reactions specific to the unsaturated bonds. It has been shown that double/triple bonds can polymerize and condense into aromatic rings (Cypres, 1987; Frenklach, 2002). Another potential path is to become oxidized by an OH radical formed from H-abstracted water (e.g., Dagaut et al., 1990). Since it is unlikely for these hydrocarbon radicals to become hydrogenated to yield saturated hydrocarbons (alkanes) in either path, their chemistry has no direct impacts on our products of interest. However, future development of the kMC model can explore these reactions, both to make the model chemical scheme more complete and realistic and to enable prediction of isotopic structures of non-alkane molecules.

### 3.3. Iteration

Once the model system is initialized by defining the chemical and isotopic structure of its initial substrates, we then allow the system to evolve through chemical reactions described in a series of distinct time steps. Every iteration step executes a single selected elementary reaction, or ‘event’, chosen randomly from the list of all possible elementary reactions determined by the system state at that time. The probability of an event being chosen for a given time step is equal to the reaction rate coefficients (in unit of time⁻¹) for that event divided by the summation of all rate constants for all possible events, determined by the system state at that time.
Rate coefficients of listed reactions are estimated with a modified Arrhenius form: 
\[ k = A T^n e^{-\frac{E_a}{RT}} \]
where \( A \) is pre-exponential factor, \( n \) is the temperature exponent (for expressing the temperature dependence of the pre-exponential factor in the original Arrhenius equation) and \( E_a \) is activation energy. Reactions in the event list are assigned with parameters retrieved from an embedded mini-kinetic library. The mini library is organized by reaction types (homolytic cleavage, \( \beta \)-scission, radical isomerization, H-abstraction and radical recombination) and reaction attributes, which includes atoms (CHNOS) associated with the reaction, degrees of carbon atoms (defined by number of non-hydrogen atoms connected: primary, secondary, tertiary and quaternary). The program processes the reaction information to classify it and acquire its designated parameters (\( A, n \) and \( E_a \)) from the mini library. Parameters in the mini-kinetic library is established by directly obtaining parameters for different reaction types from the Reaction Mechanism Generator database, under RMG-Py rate rules (C. W. Gao et al., 2016). These parameters were not modified or optimized. We enclose the full set of parameter data for mini-kinetic library in the tables of the Appendix A. Supplementary Material.

For kinetic parameterization of homolytic cleavage, the algorithm labels all reactions by one of five categories: ‘C(aliphatic) —C(aliphatic)’, ‘C(aliphatic) —C(aromatic)’, ‘C—O’, ‘C—N’, ‘C—S’ and others (single bonds between two non-carbon atoms). We group all dissociation reactions of single bonds between two non-carbon atoms into the ‘other’ category and set their rates to be zero because these bonds are rare and do not affect the model’s generation of hydrocarbons directly. The ‘C—H’ bond is not permitted to undergo homolytic cleavage because its rate coefficient is at least 5 orders of magnitude slower than that for ‘C—C’ bonds (at 500K). Since ‘C(aliphatic) —C(aliphatic)’ bonds are the most diverse and common single bonds in kerogen and their bond dissociation energies (BDE) are sensitive to the degree of carbon positions (i.e. primary/secondary/tertiary/quaternary), we subdivide this category by degrees of both carbon positions (i.e. bonds are sorted into carbon degree pairs: p-p, p-s, p-t, p-q, s-s, etc. p:primary; s:secondary; t:tertiary.). We use a different set of kinetic parameters for an unusual type of ‘C(aliphatic) —C(aliphatic)’ bond that is adjacent to a ‘C(aliphatic) —C(aromatic)’ bond (i.e., the bond is on an alkyl chain attached to an aromatic structure). Theoretical calculations have shown that these type bonds have considerably lower dissociation energies than common ‘C(aliphatic) —C(aliphatic)’ bonds (Xiao 2001). Therefore, we assign these bonds with a separate set of kinetic parameters regardless of their degrees of carbon positions. The algorithm then assigns calculated kinetic rate constants to all single bonds based on their structural category.

For \( \beta \)-scission, we assign a universal set of parameters for all reactions because studies have shown that kinetics of \( \beta \)-scission reactions do not vary strongly among different structures (Xiao, 2001). Ratkiewicz and Truong (2012) show that rate coefficients of \( \beta \)-scission reactions occurring for different molecular
sizes and degrees of carbon positions do not differ by more than an order of magnitude at relevant temperature (500K). This universal set of kinetic parameters in the library is set equal to that for \( \beta \) scission of 1-octyl radical.

For radical isomerization (H-migration), we assign kinetic parameters to each reaction based on its migration distance. We allow migration from the radical position (1) to the 4, 5, 6, or 7 positions (number indexed by distance to the radical position), because the migration of other distances are at least 3 orders of magnitude slower (Ratkiewicz et al., 2010; Sirjean et al., 2012; Ratkiewicz, 2013). The library data contain sets of kinetic parameters specific for migrations to the 4, 5 and 6 positions.

The reaction types explained in the preceding three paragraphs are unimolecular reactions, which have rate coefficients expressed in \( \text{s}^{-1} \). H–abstraction and radical recombination are bimolecular reactions that have rate coefficients in concentration\(^{-1}\cdot\text{s}^{-1}\) (e.g., molecules\(^{-1}\cdot\text{cm}^{3}\cdot\text{s}^{-1}\)). The kMC theory requires using rate constants with time\(^{-1}\) units (equation (1-7)). Therefore, we multiply the bimolecular rate coefficients by concentration of one of the reactants, which we specify in the following paragraph for each of the relative reaction types.

In H-abstraction (abstracting organic hydrogen) reactions, we permit two reaction paths: abstracting H from organic hydrogen and abstracting H from water. For both reactions, the kinetic parameters are assigned based on the degree of carbon at the radical position (0°, 1°, 2°, 3°). In the organic H abstraction reaction, we then multiply rate coefficients by the concentration of total aliphatic carbons in the system to obtain a concentration-independent rate constant. We used kinetic parameters for typical carbon radicals (0°, 1°, 2°, 3°) abstracting hydrogen from octane, so the concentration of the aliphatic structure is normalized to the equivalent octane concentration (converted by mols carbon). Estimation of the concentration of the aliphatic structure is detailed in the Appendix A. Supplementary Material. If a H-abstraction reaction occurs, a hydrogen donor is randomly selected with uniform possibility across all hydrogen containing positions on the substrate structure. This unbiased selection is a simplification in our model and is used to address two practical limitations: First, hydrogen donor positions that are physically close to the radical would be abstracted more easily, but we do not know their geometrical locations because our molecular representation is topological (see Section 3.6.2.). Second, assigning specific kinetics for each hydrogen donor requires assessment of their chemical environments, which would demand high computational costs due to the magnitude of assigned kinetics required to accurately represent the system. For the abstracting hydrogen from water, we multiply rate coefficients by a typical water concentration in shale (see Appendix. A. Supplementary Material for details).

In radical termination reactions, we enumerate a reaction list where each member is a group of two different radicals in the system. Since reaction rates for different types of radicals do not vary
significantly, we apply one set of kinetic parameters for all scenarios. The calculated kinetic rate constant is multiplied by a radical concentration factor, $C_r$. Since we consider one pair of radicals per reaction, the volume that each radical occupies is the volume of the system. Radical concentration factor in this context equals the molar concentration (in mol/m$^3$) of organic matter in the system. Therefore, we can estimate $C_r$ with $C_r = \rho \cdot \text{TOC} / M_s$, where $\rho$ is shale density, TOC is total organic matter content, $M_s$ is molecular weight of the system (including all atoms).

Homolytic cleavage, $\beta$-scission and radical isomerization are the unimolecular reactions in our model. We note that rate coefficients of these nominally unimolecular reactions can be pressure dependent. This stems from the requirement of an inert third body to collide with the excited reactant and remove excess energy (e.g., Baulch et al., 2005). Multiple theoretical methods are available for estimating pressure-dependence on rate-coefficients (Allen et al., 2012), but we use the high-pressure limit kinetic data from the RMG database (RMG-py rate rules) because it should be approached in the relatively high pressures of the catagenetic environment we are attempting to describe (Wong et al., 2003). We note that there is a significant gap between typical conditions of oil and gas generation (120–250 °C and 1000–3000 bar) and what the RMG database is commonly intended for, which is industrial pyrolysis and combustion (500–2100 °C and 0.01–50 bar).

Following assignment of rate coefficients based on reaction types and bond environments, reactions affected by isotope substitution on the substrate are modified with kinetic isotope effects (KIE). In this study, we concentrate on $^{13}$C and $^{2}$H substitutions. We generally calculate KIE with the Arrhenius formula of 

$$KIE = \frac{A^*}{A} \cdot e^{-\frac{\Delta E_a}{RT}},$$

where $A^*/A$ and $\Delta E_a$ are the pre-exponential factor ratio and activation energy difference between substituted and unsubstituted reactions; exceptions to this treatment are specified. Detailed explanation of KIE calculation and full set of parameters are included in the Appendix A. Supplementary Material.

### 3.4 Temperature and Time

We enable specification of the temperature in our model and allow it to change during cracking in order to simulate burial conditions and histories in sedimentary basins. The rates of all reactions considered by the model are updated continuously as the temperature changes from one time step to the next. Exact temperature settings are specified for each model experiment.
### 3.5 Model output

When the number of time steps reaches the target number chosen in the initialization of the program, iteration is terminated and the system is parsed to tally the final product molecular constituents and their isotopologues. Note that we do not analyze or store information for molecules and isotopologues during the iteration steps. As a result, our data represent cumulative gas generation instead of instantaneous, the latter of which might produce more extreme compositional and isotopic variations. Furthermore, we do not move molecules in or out of the system in simulations of this paper, essentially maintaining a closed-system scenario. We note that these condition settings (cumulative vs. instantaneous generation and open vs. closed systems) can be easily modified in the model, so future work could explore these options.

An example of a count of product molecules and isotopologues from the computed cracking of aliphatic oil is given in Appendix C. The relative abundances of isotopologues of each compound are then converted into position-specific isotope ratios, clumped isotope values and compound-specific isotope ratios. For example, the position-specific isotope ratio for the m position is calculated with $R_m = \frac{\sigma_m C_m}{\sigma_n C_n}$, where $\sigma_m$, $\sigma_n$ are symmetry numbers of m-substituted isotopologue and non-substituted isotopologue, and $C_m$, $C_n$ are concentrations of the m-substituted isotopologue and non-substituted isotopologue. In our current application, we focus on carbon and hydrogen isotopologues of C$_1$–C$_7$ n-alkanes, as they are the main components of natural gas and condensates, for which there is abundant stable isotope data from cracking experiments and natural gases. We note that our model could be easily modified to report information beyond these molecules, such as the isomers of C$_{4+}$ alkanes, oil compounds and sulfur isotope substitutions.

### 3.6 Programming Details

#### 3.6.1 Program organization

A simulation made using the kMC model is composed of initialization of system, iteration of time steps and processing output. The total number of time steps depends on the size of the system as well as the extent of reactions. Under scheme A, this number can be converted to a percentage ‘reaction progress’ by dividing by the total number of single bonds in the system (maximum number of steps that can be taken). It is more complicated to relate the total number of steps with reaction progress with scheme B due to the complexity of reaction pathways, but we will discuss how we approach it in the results section. A single simulation does not produce enough n-alkane molecules to be statistically useful for estimating proportions of isotopologues. A simulation starts with a system that has $10^5$ to $10^7$ carbon atoms, so given natural abundances of rare isotopes only $10^2$–$10^4$ $^{13}$CH$_4$ molecules will be formed if 10% carbon is turned
into methane, which leads to unacceptable uncertainty in the isotope ratios of products. The uncertainty of the proportions of isotopologues in product molecules of interest computed by the model follows counting statistics (i.e., error scales with the reciprocal of the square root of the quantity of each isotopologue of interest). Reaching precisions that are useful for answering geochemical questions using this method demands counting large numbers of each isotopologue of interest; e.g., one million of a given isotopologue must be observed to achieve a 1‰ (1 standard error) relative uncertainty in its abundance. Therefore, we repeat the simulation under the same settings (initialization and number of time steps) for many repetitions and combine the outputs. Typically, \(10^3\text{–}10^4\) repetitions can result in satisfying precision. These repetitions allow us to parallelize the program for faster overall execution on a computation cluster facility. In this study, we use Caltech’s High-Performance Cluster. Each experiment uses 2,000–10,000 core hours. Implementing a Monte Carlo method in parallel computing can cause problems as the pseudorandom number generator used to initialize each model and select reactions at each time step of each model can have seeding biases (Rosenthal and Carlo, 2000). We avoid this problem by seeding the generator using current time at each computing thread.

### 3.6.2 Data structure

The topological molecular structures of chemical compounds in the model are stored as graphs. Each node represents an atom and each edge represents a covalent bond. Edge weights are assigned to reflect bond orders of their respective bonds. Properties associated with the atoms are stored in separate arrays, which include the element type (C, H, O, N or S), whether that site is a radical, and any isotope (\(^{13}\)C and \(^2\)H) substitution.

When a simulation ends, we begin to process the computational output. The first step is to divide the system into separate molecules. This is operated by separating the system graph into subgraphs that are not in connection (covalent bonds) with other subgraphs. We extract and store these subgraphs from the original graph. Next, we examine the properties of these subgraphs for identification of compounds. These subgraphs are sorted by node size (number of nodes) and then we test whether each subgraph of a given size represents an \(n\)-alkane, as we report data for \(n\)-alkanes in this study. We constructed a set of reference graphs representing \(C_1\text{–}C_7\) \(n\)-alkanes. The target graph is then compared with one of the reference graphs (of same size) to examine whether the two conform to a graph isomorphism relationship. The graph isomorphism relationship here also restricts node atom elements and edge bond weights; i.e., the graph of a selected model product molecule has to have identical atoms and bond weights in the identical topological relationships as in the pre-constructed reference graphs. If graph isomorphism exists between the target subgraph and a reference graph, the subgraph is recognized as the \(n\)-alkane compound.
represented by the reference graph. We then examine the isotope substitutions in each selected target subgraph to categorize its isotopologue identity.

3.6.3 Random event selection

An array of rates of all possible reactions is calculated at each iteration step: \([r_1, r_2, \ldots, r_N]\). A uniformly distributed random number \(p \in (0, 1)\) is generated at each iteration step to select one of these reactions to occur, using the Mersenne Twister pseudorandom number generator. To do so, we calculate a cumulative summation array for the array of all possible reaction rates, by the order of the reaction list. The first reaction from the cumulative summation array to reach \(p*S\), where \(S\) is the summation of all rates, is the selected reaction.

3.6.4 Algorithmic optimization

The kMC method has high demand for computational resources (both CPU and memory). We implemented several approaches to improve the computational efficiency of our kMC model. These approaches are reported in Appendix. B.

4 Results

4.1 Accuracy test for a simple system

The accuracy of results from kMC simulations can be tested by applying that simulation to an analytically solvable problem and comparing the simulation output to the corresponding analytical solution. Here we take this approach by simulating and analytically solving the outcome of cracking butane under scheme A, where the possible reaction paths are small in number and so relatively tractable to analytical description. We monitor the position-specific isotope distribution in residual butane, which is expected to evolve over time as homolytic cleavage reactions progress. In the following paragraph we derive the analytical solution to these predicted effects, which we then compare with results of our kMC simulation of this scenario.

Under chemistry scheme A, isothermal \(n\)-butane cracking initiates with dissociation of one of its three \(C—C\) bonds. Therefore, the concentration of non-isotope-substituted \(n\)-butane will evolve over time by the following loss equation:

\[
\frac{dB(t)}{dt} = - R_{1,2} - R_{2,3} - R_{3,4}. \quad #(8)
\]

\[
B(t) = B(0)e^{-(R_{1,2} + R_{2,3} + R_{3,4})t}. \quad #(9)
\]
\( R_{a,b} \) denotes rate constants for homolytic cleavage of the bond between carbon atom \( a \) and carbon atom \( b \).

Isotope effects modify this loss equation based on the primary, secondary and tertiary isotope effects. For example, in the case of 1-deuterium \( n \)-butane (\( n \)-butane containing a single deuterium substitution on one of its first carbon positions, noted as \( B_1 - D \)), cleavage of its carbon-carbon bond 1,2 will be affected by a secondary hydrogen kinetic isotope effect and cleavage of its carbon-carbon bond 2,3 will be affected by a tertiary hydrogen kinetic isotope effect. Therefore,

\[
B_{1-D}(t) = B_{1-D}(0)e^{-\left(\frac{\partial KIE}{2}R_{1,2} + \frac{\partial KIE}{2}R_{2,3} + R_{3,4}\right)t}. \tag{10}
\]

where \( \frac{\partial KIE}{2} \) denotes a secondary deuterium KIE from a substituted \( 1^\circ \) carbon for homolytic cleavage of C—C bonds; defined as ratio between rate constants of substituted and unsubstituted molecules (kD/kH).

Similarly, we can define the rate of loss of 2-deuterium \( n \)-butane (one deuterium substituted at the second carbon position) as:

\[
B_{2-D}(t) = B_{2-D}(0)e^{-\left(\frac{\partial KIE}{2}R_{1,2} + \frac{\partial KIE}{2}R_{2,3} + \frac{\partial KIE}{2}R_{3,4}\right)t}. \tag{11}
\]

Because the \( n \)-butane molecule is symmetrical, equations 10 and 11 provide enough information to solve for the evolution in the site-preference of the D/H composition of \( n \)-butane undergoing destruction by homolytic cleavage:

\[
\epsilon D_{2-1} = \frac{2B_{2-D}(t)}{3B_{1-D}(t)} - 1 = \frac{2B_{2-D}(0)}{3B_{1-D}(0)}e^{-\left(\frac{\partial KIE}{2}R_{1,2} + \frac{\partial KIE}{2}R_{2,3} + \frac{\partial KIE}{2}R_{3,4}\right)t} - 1. \tag{12}
\]

The ratio 2/3 derives from the symmetry number ratio between 1-D \( n \)-butane and 2-D \( n \)-butane. If we assume the starting substrate \( n \)-butane has a homogeneous deuterium distribution,

\[
\epsilon D_{2-1} = e^{-\left(\frac{\partial KIE}{2}R_{1,2} + \frac{\partial KIE}{2}R_{2,3} + \frac{\partial KIE}{2}R_{3,4}\right)t} - 1. \tag{13}
\]

Combining eqn. (9) and (13),

\[
\epsilon D_{2-1} = f^{-\left(\frac{\partial KIE}{2}R_{1,2} + \frac{\partial KIE}{2}R_{2,3} + \frac{\partial KIE}{2}R_{3,4}\right)t} - 1, \tag{14}
\]

where \( f = \frac{B(t)}{B(0)} \). \( \tag{15} \)

Similarly, for \( ^{13}C \),

\[
\epsilon^{13} C_{2-1} = f^{-\left(13\frac{KIE}{2}R_{1,2} + 13\frac{KIE}{2}R_{2,3} + 13\frac{KIE}{2}R_{3,4}\right)t} - 1. \tag{16}
\]

Note that \( R_{1,2} = R_{3,4} \) because of \( n \)-butane’s molecular symmetry.
We ran a kMC experiment of n-butane cracking at 500K until 20% of all single bonds dissociate, which corresponds to 60.03% loss of n-butane. At 500K, the parameters in equation (14) and (16) are: $D_2KIE_2 = 0.8635; D_3KIE_2 = 1.0113; D_2KIE_1 = 0.8153; D_3KIE_1 = 0.9756; 1^{13}C KIE_2 = 0.9810; 1^{13}C KIE_2 = 0.9946; 1^{13}C KIE_1 = 0.9806; 1^{13}C KIE_1 = 0.9983; R_{1,2} = R_{3,4} = 4.6243 \times 10^{-22} s^{-1};$ and $R_{2,2} = 8.6033 \times 10^{-21} s^{-1}$. The analytical model predicts that the residual butane will have site-specific H and C isotope structures characterized by values of $\epsilon D_{2-1}=93.0 \text{‰}$ and $\epsilon^{13}C_{2-1}=14.4 \text{‰}$. The kMC results for both carbon and hydrogen isotopes match the analytical values to within statistical uncertainty in the kMC experiments (Figure 4).

**4.2 Isotope ratios of C$_1$-C$_7$ n-alkanes**

In this section, we report intramolecular and intermolecular differences in isotope ratios of products from experiments in which we executed the kMC model to a fixed reaction progress at 500K, using either aliphatic oil or kerogen model kIA as precursor. This reaction progress is defined differently for each scheme. For scheme A, we stop when 20% of the initial bonds have undergone cleavage (number of steps divided by total initial number of single bonds), which is selected to produce a relatively even distribution of C$_1$–C$_7$ n-alkanes. It is unclear how we should relate bond-cleavage progress in scheme A (20% completion) to that in scheme B, because scheme B has more reaction classes and the conversion from macromolecular substrate to gaseous hydrocarbons and residue is not a linear function of the number of time steps or reaction events. We decided to use an empirical parameter, methane transformation ratio, for standardizing reaction progress. It is defined as:

$$F_{C1} = \frac{[C1]}{maximum \ [C1]} \times 100\%. \ (#17)$$

[C1] is the yield of methane molecules per kerogen (or other substrate) unit. Maximum yield of C1 is computed by running the kMC model to full completion. The methane transformation ratio grows monotonically with reaction progress in both scheme-A and scheme-B models, making it a good measure of reaction progress. For 20% bond cleavage completion under scheme A, $F_{C1}$ is calculated to be 1% — corresponding to an early maturation stage. We thereby present data from running scheme B that yield a methane transformation ratio around 1%. Molecular and isotopic data for these runs are deposited in Appendix C.

We computed elapsed real time for our experiments by methods illustrated in section 3.5. For experiments with scheme A, time passed is $1.22\pm0.03 \times 10^{19}$s (387 Byr) for cracking the model kIA kerogen and $3.66\pm0.04 \times 10^{19}$s (1160 Byr) for cracking model oil. These are many orders of magnitude higher than typical
burial times of geological strata ($10^{13}$–$10^{15}$ s or $10^0$–$10^2$ Myr). Under scheme B, the elapsed time of designated reaction progress is computed to be $1.05\pm0.02\times10^{15}$ s (33 Myr) for cracking model kIA kerogen and $2.40\pm0.06\times10^{18}$ s (76.2 Byr) for cracking model oil; the first of these is closer to a realistic burial time scale whereas the second remains far longer than any relevant geological timescales.

The substantial differences between model times for significant reaction progress of both scheme A models and the Scheme B model of oil cracking vs. those of relevant geological processes expose the over-simplification of our treatment of rates and reactions, particularly for scheme A. Thermal cracking can be accelerated by relatively fast radical reactions but is always slow when rate limited by homolytic cleavage (section 3.6.4). The kMC model we present that most closely matches geological timescales of catagenesis (scheme B cracking kerogen) succeeds in this respect because the substrate contains abundant radical initiators — moieties, such as heteroatoms (especially sulfur) and single bonds adjacent to aromatic rings, that are more susceptible to cleavage and generate radicals. In contrast, scheme B cracking of oil demands long times because such radical initiators are absent in the simplified model substrate. And, all scheme A models lack radical chemistry and are correspondingly slow. These observations echo experiments showing that sulfur kickstarts radical initiation in cracking and thereby accelerates overall rates of catagenesis (Lewan, 1998). We also recognize other factors that might have caused our model time to exceed realistic natural times of catagenesis: (1) The kinetic database that we are using (RMG rate rules) might be inaccurate in the temperature range of catagenesis, as they have not been well calibrated to experiments at these temperatures. However, these inaccuracies can be canceled to some extent for predicting molecular and isotopic abundances, as they affect all type of reactions and could lead to systematic errors in overall model timescales. (2) Scheme B is also a simplification of true radical mechanisms. For example, reactions of unsaturated species and oxidation are overlooked. And (3) we have not considered effects on overall catagenetic rates of potential catalysts such as clay minerals, transition metals and water (Seewald, 2003). Therefore, we focus on reaction progress rather than time when discussing the relationship between our model results and natural catagenetic processes (see also section 5.3).

We present compound-specific carbon and hydrogen isotope compositions calculated for C$_1$ through C$_5$-$n$-alkanes produced by our kMC model of cracking the aliphatic oil and kerogen kIA, respectively. Note that methane transformation ratio $F_{C1}$ is 1% and $\delta^{13}$C is -25‰ and $\delta^2$H is -100‰ for the initial substrate in these experiments. Carbon isotope data are illustrated using the ‘Chung plot’, or a plot of $\delta^{13}$C vs 1/n (where n=carbon chain length; Figure 5). Chung et al. (1988) devised this diagram based on the hypothesis that one position per alkane is isotopically fractionated by KIE’s during bond dissociation reactions by which these compounds were generated. The chemistry of either of the schemes in our model
is more varied than this simplifying assumption, such that there are many possible sites in these molecules that can be influenced by bond-dissociation KIE’s. However, the Chung diagram is widely used to interpret measured isotopic compositions of natural gas components, so it is useful to examine our model outputs in this plot. We find that overall range of compound-specific carbon isotope compositions of C$_1$–C$_5$ n-alkanes from our model agree reasonably well with empirical field of data for natural gases generated from sapropelic substrates. We fit our data linearly and show that 3 out of 4 kMC model scenarios conform closely to a linear trend (R$^2$≥0.95), whereas one scenario, cracking model kerogen kIA with scheme A chemistry, is poorly fit by a line in this composition space (R$^2$=0.49; Figure 5). This exception is caused by a high $\delta^{13}$C of methane at -33.9‰. In comparison, $\delta^{13}$C of methane is -48.8‰ for cracking aliphatic oil under scheme A, and -55.5‰ and -53.6‰ for cracking kIA kerogen and aliphatic oil under scheme B, respectively. Excluding the result for model kIA kerogen with scheme A chemistry, the Chung-plot slopes for the remaining three scenarios range from -37.7 to -30.8‰. These slopes resemble those typical of suites of natural gases, and the overall positions of the trends overlap the empirical field of natural gas, suggesting our model captures first-order features of measured carbon isotopic properties of natural gases — an encouraging result given that our model was not tuned in any way to match isotopic data for natural gases or experiments simulating catagenesis.

Position-specific isotope ratios of C$_3$–C$_7$ alkanes from scheme A are shown in Figure 6. Note that we set isotope ratios of initial substrates to be $\delta^{13}$C=-25‰ and $\delta^2$H=-100‰ for all experiments included in this study. Alkanes produced from cracking oil (aliphatic fraction) have distinctive intramolecular stable isotope patterns characterized by heavy-isotope depletion on the terminal methyl group. The terminal groups have $\delta^{13}$C values lower than the rest of the molecule by 10–20‰ and $\delta^2$H values lower by 230–250‰. The non-terminal positions have relatively uniform hydrogen isotope ratios, while the methylene groups neighboring terminal positions have carbon isotope ratios lower than inner positions and higher than terminal positions. This pattern is consistent from propane to heptane. Alkanes produced from cracking model kIA kerogen substrate have stable isotope distributions generally similar to those from model aliphatic oil but differ from them in two noteworthy respects: 1) Carbon isotope fractionations between positions are generally smaller in gas from kerogen cracking. The range of C$_3$–C$_7$ alkanes’ position-specific $\delta^{13}$C is 6–32‰ for aliphatic oil, and 2–11‰ for kerogen. 2) Position-specific hydrogen isotope ratios of n-heptane (C$_7$H$_{16}$) are not consistent with the terminal-depletion pattern described earlier. The 3rd,4th and 5th positions on n-heptane have lower D/H ratio than the 2nd and 6th position on heptane, by around 150‰.

Experiments using scheme B at extents of reaction progress similar to the preceding results for scheme A are presented in Figure 7. Position-specific carbon isotope results for alkanes exhibit a closely similar
pattern to those produced from scheme A (i.e., dip in $\delta^{13}$C values at the terminal position and its neighboring position). A subtle change is observed in $n$-hexane produced by cracking kerogen, where the 2nd and 5th positions become more negative in $\delta^{13}$C compared to that from scheme A. The ranges of intramolecular $\delta^{13}$C of $n$-alkanes larger than propane produced by cracking kerogen from scheme B is slightly higher than those produced from scheme A, varying between 23 to 40‰. Position-specific hydrogen isotope distributions are generally similar to those from scheme A, but show more diverse features, especially for the experiment using model kIA kerogen as substrate. For example, the 3rd position of $n$-pentane produced from scheme B is more than 50‰ lower in $\delta^2$H than the 2nd, 4th positions, but that from scheme A is higher in $\delta^2$H than the 2nd and 4th positions. Similarly, 3rd and 5th position of $n$-heptane become more depleted in hydrogen isotopes in scheme B.

4.3 Effects of reaction progress

We examined the computed isotopic compositions and intramolecular isotopic structures of product compounds of interest at several different stages of reaction progress for a model using the kIA kerogen molecule as substrate and either chemistry scheme A or B (Table 1). This exercise shows what the kMC model predicts for the trend in gas isotope chemistry over the course of natural burial and thermal maturation. We find that the $\delta^{13}$C values of alkanes increase with reaction progress (denoted by the methane transformation ratio $F_{\text{C1}}$) (Figure 8). The $\delta^{13}$C of ethane and propane under scheme A rises very fast, and $\delta^{13}$C of propane even becomes exceptionally high after $F_{\text{C1}} > 30\%$. The $\delta^{13}$C of ethane and propane under scheme B are less variable. The $\delta^{13}$C of methane under scheme A decreases slightly in the early stages of reaction progress and then increase in the later stages. Similarly, proportions of wet gas components ($C_{2+}$) decrease monotonically with reaction progress under scheme B but decrease at early stage and increase later under scheme A (Figure 9). Under chemistry scheme A, the intramolecular carbon isotope differences of propane (i.e. center-terminal fractionations, $\varepsilon^{13}$C$_{C-T}$) increase with reaction progress; specifically, $\varepsilon^{13}$C$_{C-T}$ rises from 5‰ at 1% methane transformation ratio (20% cleavage completion) to 25‰ at 28% methane transformation ratio (60% cleavage completion) (Figure 10). Under chemistry scheme B, the center-terminal carbon isotope fractionation of propane remains relatively constant (total range of 4 ‰, with no consistent temporal trend) over the course of reaction, and remains similar to the value observed in scheme A at low total reaction progress (i.e., early in the model, or low in thermal maturity). Again, note the reaction progress here is expressed as a percentage of bonds broken for scheme A, and at equivalent methane yield to scheme A for scheme B. We find the estimated realistic time increases drastically with reaction progress. In the maturation series of cracking model kIA at 500K using scheme B, the estimated elapsed time is barely acceptable for early-stage gas generation ($\sim 10^7$ yrs at
F_{C1}=1\%; \text{ section 4.2) but becomes } 3.6\times10^{13} \text{ yrs at } F_{C1}=78\%. \text{ This reflects a poorer performance in simulating late stage catagenesis than early stage, which we will discuss in section 5.3.}

5 Discussion

5.1 Influence of precursor chemical structure on isotopic and molecular properties of products

It has been previously demonstrated that the intramolecular isotopic fractionation of the products of hydrocarbon cracking is largely controlled by precursor chemical structure (Xie et al., 2020). The rationale is that isotopic depletion on certain sites of products is usually associated with expression of KIE’s of bond dissociation reactions. The single bond dissociation reactions (either homolytic cleavage or β-scission) in hydrocarbon chain moieties of substrates selects for \(^{12}\text{C}\) over \(^{13}\text{C}\) and for \(^{1}\text{H}\) over \(^{2}\text{H}\) at the bond dissociation site with a strong primary isotope KIE, thereby making that position in the product depleted in the heavy isotope. The adjacent positions will also be isotopically selected, but with a milder secondary isotope KIE. In this study, the terminal groups of alkanes tend to be isotopically depleted (in both C and H) relative to the internal positions, regardless of reaction schemes. We suggest that this results from the prevalence of unbranched straight-chains in the chemical structures of precursor substrates, either aliphatic oil or model kIA kerogen. The terminal positions are mostly sourced from cleaved positions when cracking linear unbranched chains, but branched chains favor cleavage on branching positions (due to lower energies of bonds connecting higher degree carbons), in which scenario the cleaved positions become internal positions in produced alkanes. Therefore, more branched chemical structures lead to less relative depletion on terminal positions of produced alkanes, and vice versa. When the aliphatic portion (part of the kerogen structure composed of acyclic alkyl chains, a main source of volatile hydrocarbons) of the substrate is dominantly composed of linear unbranched alkyl chains instead of branched ones, such as the aliphatic oil model and kerogen model in this study, the produced \(n\)-alkanes have relatively negative isotopic dips at terminal positions and relatively uniform isotopic ratios in the internal positions for product hydrocarbons. For carbon isotopes, the impact extends inward to the neighbors of terminal groups, which can be explained by a relatively significant secondary carbon isotope KIE (at 500K, \(1000\ln\left(\frac{^{13}\text{C}}{^{12}\text{C}}KIE_{2}\right) = -19\) and \(1000\ln\left(\frac{^{13}\text{C}}{^{2}\text{H}}KIE_{2}\right) = -5.4\)). A few compounds in the results using model kIA kerogen as substrate have some of the internal positions depleted too, reflecting the higher proportion of branched chains in model kIA.

The abundance of radical initiators (moieties that are more susceptible to cleavage and generate radicals, such as heteroatoms (especially sulfur) and single bonds adjacent to aromatic rings) in the substrate molecular structure also impacts kMC results. In our study, the oil model does not have radical initiators,
whereas the kerogen kIA model does. As presented in section 4.2, the estimated time for cracking oil is much longer than for cracking kerogen. Under scheme A, we also noticed that the kerogen model produces methane with a carbon isotope ratio that is higher than that produced by oil cracking (for early-produced gas at \( FC_1 = 0.01 \); Figure 5). As a matter of fact, \( \delta^{13}C \) of methane is higher than \( \delta^{13}C \) of ethane, exhibiting an ‘isotopic reversal’ that is usually observed in over-mature natural gases from deep formations (Burruss and Laughrey, 2010; Tilley and Muehlenbachs, 2013; J. Dai et al., 2014; Milkov et al., 2020). We note that it is uncommon for natural gases to exhibit reversal in early maturation of source rock. We believe that this model anomaly could be attributed to the relatively weak carbon isotope KIE in homolytic cleavage of carbon-heteroatom single bonds. For example, \( 1000 \ln (KIE) \) for a primary carbon is -9.6 in a C—S bond and -19.6 in a C—C bonds. Cleavage of these heteroatom bonds could form methane with relatively high \( \delta^{13}C \) at early stage of catagenesis. As reaction progress increases, this enrichment becomes diluted with isotopically lighter methane generated from cracking C—C bonds that have stronger carbon isotope KIE. Since cracking under scheme B can proceed via other radical chain reactions than homolytic cleavage, such positive anomalies in methane’s \( \delta^{13}C \) are attenuated as observed from results.

The lack of polar and aromatic fractions in our model oil caused high time estimates and potentially changes in molecular and isotopic composition of gas. Similarly, variations in chemical structures and compositions of kerogen (e.g. sapropelic vs. humic origins) will also affect catagenetic products. Future work should explore these effects.

### 5.2 Scheme A vs. scheme B chemistry

In this section, we compare model results using scheme A vs. scheme B chemistry, and ask how similar each is to natural gases.

The intramolecular isotopic distributions are generally similar between scheme A and B (Figure 6 and 7), except that scheme B tends to produce greater H isotope depletions in the internal positions. For models using chemistry scheme A, the internal positions become more heavy-isotope enriched relative to terminal positions with increasing reaction progress (Figure 10). We recognize two main causes for this trend. (1) Secondary cracking of the product \( n \)-alkane molecules increases residual molecules’ heavy isotope ratio and does so unevenly across the non-equivalent molecular sites. This is because internal positions have two C—C bonds and thus have higher probability of participating in bond dissociation events and being affected by KIE’s. (2) The hydrocarbons produced early in catagenesis (at low reaction progress) are more commonly formed via one bond dissociation events. Because \( n \)-alkanes are symmetrical, the isotope
ratio of a terminal position averages one fractionated position and one unaffected position, such that the average heavy-isotope depletion of terminal sites is diluted by about a factor of 2 relative to the actual KIE. The late stage C\textsubscript{1}-C\textsubscript{7} hydrocarbons (those still present after high reaction progress) are often products of two bond dissociation events, enhancing the overall magnitude of heavy-isotope depletion in the terminal positions. In contrast, in models that use chemistry scheme B, the center-terminal fractionations vary little over a range of reaction progress. This discrepancy likely stems from the addition of radical chemistry in scheme B, although the specific mechanism is not clear. One of the possible mechanisms for moderating the development of stronger site-specific fractionations as reaction progress increases is reformation of propane via radical recombination (e.g., methyl+ethyl \rightarrow \text{propane}). These reactions provide a source of propane that does not have high central-terminal isotopic contrasts. Central-terminal fractionation for propane increases slightly or does not change with thermal maturity in natural gas samples (Piasecki et al., 2018; Liu et al., 2018; He et al., 2019; Zhao et al., 2020; Julien et al., 2020) and pyrolysis experiments (Piasecki et al., 2018; Gilbert et al., 2019). The maximum range of $\varepsilon^{13}$C\textsubscript{C-T} variation in propane is less than 12‰ in these pyrolysis experiments and natural samples, but is 22‰ in our simulation of the kIA kerogen model under scheme A. Therefore, the trend in site preference with increasing thermal maturity calculated using scheme B chemistry is much closer to analogous trends documented in natural gases.

Alkane gas generated using chemistry scheme A is wetter (lower in C\textsubscript{1}/C\textsubscript{2+}, mol proportions) than that produced using scheme B (Figure 9). The C\textsubscript{1}/(C\textsubscript{2}+C\textsubscript{3}) ratio for gases generated at low reaction progress (F\textsubscript{C1}<15%) using scheme A ranges from 0.13 to 0.51, which is exceptionally low compared to data for natural gas samples. Conventionally, it has been shown that laboratory experiments tend to produce gases that are wetter than natural gases (Mango et al., 1994; Price and Schoell, 1995). However, the gas produced at early-maturation under scheme A is so wet that it has more wet-gas components than almost every analyzed natural or laboratory gas. Milkov and Etiope (2018) show that C\textsubscript{1}/(C\textsubscript{2}+C\textsubscript{3}) ratios of 99% of recorded natural gas samples are greater than 1. Almost all gases from laboratory experiments also have C\textsubscript{1}/(C\textsubscript{2}+C\textsubscript{3})>0.5 (Evans and Felbeck, 1983; Saxby and Riley, 1984; Seewald et al., 1998; Pan et al., 2008; Lewan et al., 2008; Liguo et al., 2009; Shao et al., 2018). The extremely low methane contents for our model-calculated gases generated using scheme A could be attributed to the mechanism and kinetics of homolytic-cleavage. For example, in thermal cracking of nC\textsubscript{6} under scheme A, we could estimate hydrocarbon composition after the first reaction, which corresponds to 20% completion of bond dissociations of a single n-hexane molecule (1 out of 5 bonds; note that real kMC modeling starts with many units of molecules instead of one). The rate of cleaving internal bonds ($8.6\times10^{-21}$s\textsuperscript{-1} at 500K) will be faster than the terminal bonds ($4.6\times10^{-22}$s\textsuperscript{-1} at 500K), because internal bonds connect two secondary carbon positions whereas the terminal bonds connect one primary carbon and one secondary carbon.
Consequently, the produced gas can be calculated to be 1.7% C$_1$, 32% C$_2$, 32% C$_3$, 32% nC$_4$, and 1.7% nC$_5$, which has a C$_1$/C$_2$+C$_3$ ratio of 0.03. This example shows that pure homolytic cleavage of linear alkyl structures tends to produce very high proportions of wet components.

Conversely, molecular breakdown under scheme B has other paths that suppress generation of ultra-wet gas. C$_1$/C$_2$+C$_3$ ratios of gases produced using scheme B increase from 1.7 to 7.7 as reaction progresses. This range overlaps with the typical results from thermal cracking experiments and early to main stage maturity natural gases (Figure 9). Some of the possible reaction events will favor production of methyl radicals, which can be capped to form methane. For example, a propyl radical could undergo a β-scission reaction to form a methyl radical. Similarly, a 1-hexyl radical has a very high probability to isomerize into a 3-hexyl radical, which could undergo β-scission to form a methyl radical. Another important reason why scheme B produces more methane-rich gases is that methyl radicals have higher reactivity with an alkyl-H or water (compared to other alkyl radicals), so it is more favorably de-radicalized than ethyl and other longer radicals. However, we note that methane contents from scheme B are also generally closer to the low-end in the range of thermogenic gas deposits. We will explore the potential cause and meaning of this in section 5.3.

As discussed in section 5.1, compound-specific δ$^{13}$C in alkanes produced from kIA kerogen at an early stage of cracking (F$_{C1}=1\%$) under scheme A is characterized by anomalously high δ$^{13}$C of methane. However, gases produced early in the reaction progress from both kerogen and oil models under chemistry scheme B have compound-specific carbon isotope ratios generally similar to those observed for low-maturity thermogenic natural gas, such as gas from Paris Basin (Pinti and Marty, 1995; Prinzhofer et al., 2009) or Julia field (Thiagarajan et al., 2020a). δ$^{13}$C of alkanes increases with reaction progress under both schemes (Figure 8). In particular, ethane and propane become progressively enriched in $^{13}$C rapidly at the middle- to late stage under scheme A, and their δ$^{13}$C values exceed 0‰, which is above the typical range of natural gas deposits. The same phenomenon was discovered in kMC model results by Peterson et al. (2018). We attribute this to cracking of non-methane alkanes that leave the residual pool more heavy-isotope enriched, which has been hypothesized for longer alkanes (Tang et al., 2005). Under chemistry scheme A, cracking is a unidirectional process of breaking larger molecules down to smaller fragments. At near completion of each model experiment, longer alkanes are decomposed to high extents, so the heavy isotope contents of the remaining fractions become significantly elevated by distillation. The carbon isotopes of these alkanes in the maturation sequence of scheme B do not increase as much as they do in scheme A—they do not turn positive even at the end of the examined maturity range. This is likely a result of alkane inter-conversion mediated by radical reactions. For example, a methyl radical can form from a methane molecule via H-abstraction by another radical, and the methyl radical can recombine with
another radical to form a higher-order alkane. This reformed alkane inherits carbon from the isotopically lighter smaller alkanes. Therefore, this mechanism can potentially ‘buffer’ the carbon isotope composition of higher order alkanes.

Finally, we tested the level of hydrogen incorporation into C$_{1}$-C$_{7}$ product hydrocarbons from water. In a series of low-temperature catagenesis experiments, Wei et al. (2019) found that around 6–14% of methane H derives from water at 200°C. In the kMC model, water hydrogen becomes incorporated into alkane molecules via H-abstraction reaction between alkyl radicals and water. It is difficult to resolve where hydrogen comes from directly, as it carries similar isotopic signatures as organic hydrogen. Therefore, we did a short kMC experiment with water’s $\delta^{2}$H increased to 10,000‰. The $\delta^{2}$H increment in produced methane indicates that 27% of H in methane originate from water under scheme A and 10% under scheme B (500K, 1% methane transformation, model kerogen k1A). Our results demonstrate that scheme B also better mimics lab experiments in origins of alkane hydrogen. However, we note that water-alkane hydrogen isotope exchange is suggested to play an important role in higher maturity catagenesis (Sessions et al., 2004; Xie et al., 2020; Xie et al., 2021).

The results summarized above strongly suggest that scheme B is superior to scheme A in simulating natural gas formation at catagenic conditions, in three main ways: First, its estimated timespan of kerogen cracking is much closer to a realistic time range. Second, its prediction of gas wetness is closer to natural gas molecular compositions. Third, its compound-specific and position-specific isotopes of alkanes evolve with thermal maturation more like to trends in natural gases. Conversely, our results suggest that pure homolytic cleavage is a poor choice of mechanism for interpreting the molecular and isotopic compositions of gases generated by cracking kerogen or oil; this assumption has been used often by previous studies (e.g., Waples and Tornheim, 1978b; Waples and Tornheim, 1978a; Tang et al., 2000; Tang et al., 2005; Ni et al., 2011; Peterson et al., 2018, among many others), suggesting much of the mechanistic interpretation of gas chemistry may be incorrect (at least in some key details) and should be re-examined.

5.3. Low temperature catagenesis vs. high temperature pyrolysis experiments

Previously, we showed that the kMC model with chemistry scheme B could generally reproduce compositional and isotopic trends in natural gas evolution using kerogen model as substrate. However, there are still some inconsistencies. Notably, simulation of middle- to late stage catagenesis (mature to overmature stage) also yield relatively wetter gas (Figure 11). Even at very high reaction progress (methane transformation ratio ~ 80%), simulated gas molecular composition still corresponds to an early maturation (‘oil window’) natural gas. It does not seem to reproduce the wet-gas to dry-gas transition that
we see in relatively late, ‘gas window’ natural catagenesis. It is possible that this discrepancy is related to physical processes. It has been proposed that expulsion loss, desorption, migration and entrapment could alter molecular composition of gases in nature (Galimov, 1988; Price and Schoell, 1995; Rooney et al., 1995). Besides, we note that natural gas is often formed under open or semi-open conditions, while our model results effectively describe catagenesis in a close system, which might also contribute to this discrepancy. Specifically, the loss of early-formed wetter gas in an open system might decrease wetness of retained gas. However, we note that these physical processes are likely not enough to create widespread overmature dry gas deposits in both conventional and unconventional reservoirs (Milkov and Etiope, 2018; Milkov et al., 2020), because they don’t provide a universal means to remove wet gas components almost entirely in all of these geological environments that host overmature dry gas.

We think that sluggish decomposition of \( n \)-alkanes is at least partially responsible for discrepancies in wetness between our models and natural gases at the mature stage of catagenesis, and may also contribute to the longer time frames of our model calculations. Ethane is a particularly illuminating case as it can only crack via homolytic cleavage of a relatively strong C(primary)—C(primary) bond (while higher alkanes could decompose via relatively fast \( \beta \)-scission). At 500K, the rate constant for ethane’s homolytic cleavage is \( 3.4 \times 10^{-23} \text{s}^{-1} \) (\( 1.0 \times 10^{15} \text{yr}^{-1} \)), which is about 7–9 orders of magnitude too slow to efficiently degrade ethane thermally or convert wet gas into dry gas via thermal cracking under catagenetic conditions. As previously noted, rate coefficients could potentially be inaccurate at 500K as RMG database has not been calibrated to this temperature range, yet it is unlikely that this kinetic data error could fully account for the big discrepancy in time. This effect does not impact early-stage catagenesis as much, because radical initiation at this early stage could be promoted by homolytic cleavage of weaker single bonds (bonds involving heteroatoms S/O, higher degree carbons, and single bonds adjacent to aromatic rings) and a chain of radical reactions following that. But, in overmature gases ethane decreases in abundance in natural systems for reasons that seem not to be represented in even our relatively sophisticated scheme B chemistry.

This evidence suggests a role of alternative pathways (of thermal cracking) during natural gas formation, such as catalysis or quasi-equilibrium chemistry (Mango, 1992; Helgeson et al., 2009; Mango et al., 2009; Thiagarajan et al., 2020b). We can test this idea by examining the apparent ‘Chung plot’ carbon isotope slopes from the kMC model, thermodynamic equilibrium, and natural gases (Figure 12). Histograms of such slopes for natural gases exhibit a peaked, unimodal distribution. We compiled kMC model results for cracking kerogen model kIA at \( F_{cl}=1–29\% \) (500K; see Table 1 for results) and determined a range of -42.4--27.5‰. We note that the range of kMC results can be potentially underestimated as other molecular models and temperatures were not explored. The equilibrium results were determined using \( \delta^{13} \text{C} \) of \( C_1-C_5 \).
n-alkanes under thermodynamic equilibrium, calculated using theoretical estimates (Thiagarajan et al., 2020b) at 100–250°C and $\delta^{13}$C_{methane}=-45‰. We found that the kMC results are on the lower side (steeper slopes) of the peak in the natural data, and equilibrium predictions are on the higher side (flatter slope). The equilibrium model overlaps with a higher density area of the natural sample histogram than does the kMC model. A significant part of natural gas data is between kMC and equilibrium values, which could be interpreted as being in a transition from kinetic control to thermodynamic control. This observation is consistent with the hypothesis that both kinetic cracking and quasi-equilibrium processes are relevant in natural gas generation, and that quasi-equilibrium processes become increasingly important at the highest thermal maturities. By this interpretation, the decrease in ‘Chung plot’ slopes characteristic of the most mature natural gases could be thought of as a signature of the onset of quasi-equilibrium hydrocarbon chemistry.

Additionally, to test the performance of the kMC model on indisputable thermal cracking systems, we performed a set of kMC experiments to simulate laboratory pyrolysis experiments. Since we previously found that chemistry scheme A is too simplified to simulate cracking chemistry with satisfactory accuracy, we only apply chemistry scheme B in this modeling activity. We reference our experiments to a recent publication, Wu et al. (2019), where they conducted a series of hydrous pyrolysis experiments to simulate thermal maturation of an immature marine kerogen and measured molecular and isotopic composition of natural gas. They used a type I kerogen from calcareous shale in the Liaohe Basin, analogous to the kIA molecular model that we use. In their experiments, thermal maturation of kerogen is controlled by experimental temperature. We choose to apply temperature and time of Wu et al.’s experiments for setting reaction progress in kMC. This differs from the previous experiments in this study, where we used methane production to control reaction progress.

Gas dryness of both experimental and model results trend upwards with increasing temperature (and maturity), but the range of variation is greater in experimental results (Figure 13). Experimental gas is wetter than modeled gas for lower temperature experiments, and crosses over to compositions dryer than the model at higher temperature. We note that the discrepancy here is much smaller than in the comparison of model to natural data, where natural gas is an order of magnitude dryer than kMC data in the mature stage. The evolution of $\delta^{13}$C and $\delta^2$H values of methane and ethane show close agreement between pyrolysis experiments and kMC model, i.e., all increase with temperature at similar rates. The ability to reproduce stable isotope trends from pyrolysis experiments while using the laboratory temperature-time program as input is encouraging, as our model is not calibrated to any kerogen pyrolysis experiments. On this basis, our thermal cracking model seems to capture the kinetic rates and reproduce gas chemistry of high temperature pyrolysis better than low temperature catagenesis, although the model
with scheme B chemistry describes maturity trends in stable isotope compositions equally well for both natural and experimental gases.

Overall, the kMC model (scheme B) mimics laboratory cracking experiments better than geological formation of petroleum. For the latter scenario, kMC results are generally comparable to natural data in the pre-mature stage but fail to match gas wetness and time evolution. (A previously noted corollary of this observation is that experiments generally over-predict wetness of natural gases; e.g., Mango et al. (1994)) One way to reconcile this discrepancy requires a paradigm shift in our thinking about the mechanisms of natural gas generation, and their evolution with increasing maturation. That is to say, cracking dominates low-temperature catagenesis, but thermally activated alternative mechanisms dominate at high temperatures, where they reform natural gas and convert wet gas into dry gas (Thiagarajan et al., 2020b; Xie et al., 2021). Potential mechanisms could include mineral or metal catalyzed reactions and radical reaction mechanisms omitted in this study (e.g., explicit reactions of unsaturated molecules and in oxidation/reduction chemistry).

6 Summary and Conclusion

Realistic modeling of explicit radical reaction networks in thermal cracking has been hindered by the computational and material complexity of the system. In this study, we employ the kinetic Monte Carlo method to circumvent the complexity problem by modeling a stochastic process of radical chain reactions. We designed a model that directly takes the molecular structure of precursor organic matter as input, temperature and time as environmental condition variables, and outputs the composition and isotopic values of hydrocarbon products. The model outputs contain position-specific stable isotope data as it tracks the isotopic substitution at specific molecular positions. Such information could serve as additional constraints on the molecular structure and composition of precursor substrates as well as physical conditions (such as temperature) of catagenesis. We verified the reliability and accuracy of this methodology in a confined system of butane cracking.

In order to test the necessity of including more complete radical reaction types, we designed two chemistry schemes, a relatively simple one that only considers homolytic cleavage and capping, and a more complicated one that further includes β-scission, radical isomerization, H-abstraction, and radical recombination. We conducted experiments to model catagenetic processes with both schemes and found that their results are distinct from each other. The more complicated scheme B outperforms the simpler scheme A in reproducing several aspects of natural gas data, including gas wetness, trend of intramolecular isotope fractionations, compound-specific isotope compositions, and timescales of
reaction. We conclude any mechanistic interpretation of catagenetic natural gas formation must include
the extended radical reaction types, whether one is trying to describe yield, gas composition or gas stable
isotope properties.

The favorable fit between kMC results and data from hydrous pyrolysis experiments demonstrates our
model’s accuracy in simulating thermal cracking chemistry, especially since it used the actual
experimental temperature-time program and was not fitted to experiments in any way. For application
towards geological burial conditions, our model can reproduce the isotopic evolution in natural gases but
fails to yield corresponding time estimates and wetness for mature stage of catagenesis. This supports the
hypothesis that mature gases are formed or modified via non-cracking mechanisms.

Our model employs chemical kinetics to form a link between three types of information: chemical
structure of precursors, temperature/time conditions, and quantity/composition/isotopes of hydrocarbon
products. We envision that one could infer any one of these three types of information by feeding the
model with data from the other two. For example, it could be used to predict petroleum generation
potential for given organic matter. Conversely, it will also allow inferring the origin of oil and gas with
measurable geochemical properties. We also acknowledge that exploration of catagenetic chemistry with
the kMC method is still in its infancy and the results presented here face several limitations. Most
importantly, our results did not fully consider temperature variations in natural thermal maturation of
source rocks, expulsion of hydrocarbons in semi-open shale rocks, catalytic processes, and various types
of precursor materials. We recognize these issues as plausible directions for future work.

Acknowledgement

The computations were conducted on the Caltech High Performance Cluster, partially supported by a
grant from the Gordon and Betty Moore Foundation. We thank David Wang, Michael Lawson, Tom
Miller, Alex Sessions, Kenneth Farley, Guannan Dong and Nivedita Thiagarajan for helpful discussions.
We thank Alexie Milkov, Kenneth Peters, and an anonymous reviewer for reviewing our manuscript.

Research Data

All scripts and data used in this paper are available at https://github.com/1995123xh/kMC.
Appendix A. Supplementary Material

Parameters used for in chemical kinetics and KIE, and composition of molecular model of oil (saturated fraction)

Appendix B. Methods for optimizing computational efficiency

In this appendix section, we explain the treatments employed to reduce the computational cost without sacrificing accuracy.

(1) We made hydrogen atoms implicit in the model. Numbers of hydrogen atoms are implied by the bonding of carbon (and presence or absence of radicals). This step reduces the number of nodes and edges of the molecular graph and therefore saves memory. We record the presence/absence of $^2\text{H}$ for each carbon atom instead of hydrogen atom (multiple $^2\text{H}$ substitution on one carbon atom is neglected in the current version of the model). When a reaction involving hydrogen occurs (H-abstraction, radical isomerization), we treat the hydrogen atoms on the reacted position explicitly. We analyze the degrees of carbon (by bond connections) and unpaired electrons to determine number of hydrogen atoms, and check the $^2\text{H}$ substitution table to determine isotopes.

(2) We arbitrarily enriched the rare-isotope content in the starting model substrates, since most of the error is contributed by the uncertainty in counting product molecules that contain one or more minor isotope substitutions. We apply an enrichment factor relative to natural abundances of 5 for $^{13}\text{C}$ content and 50 for D content. We do not attempt higher level of enrichment to avoid doubly and multiply substituted isotopologues overwhelming the $\text{C}_2\text{H}_6\text{alanes}$, as those isotopologues are not identified in output processing. For example, if we use 10 times natural abundance for $^{13}\text{C}$ content, the doubly substituted $^{13}\text{C}$ isotopologue of heptane would be as abundant as its singly substituted isotopologues. We divide the isotope ratios of the output with these factors so that results can be compared with natural materials.

(3) For models that employ chemistry of scheme B, we divide the reaction types into three tiers: fast reactions, medium reactions and slow reactions. Fast reactions are radical isomerization reactions to the fourth and fifth position that have rate constants of $10^4$–$10^5\text{s}^{-1}$ at 500K. Medium reactions include $\beta$-scission, radical recombination, H-abstraction and capping. These reactions have radicals in the reactant and rate constants of $10^0$–$10^3\text{s}^{-1}$ at 500K. Slow reactions are homolytic cleavage reactions that have rate constants of $10^{-27}$–$10^{-16}\text{s}^{-1}$ at 500K. Because fast reactions have orders of magnitude higher probability of happening, almost all the computational resources would be allocated to calculating fast reactions unless we intervened by artificially bypassing, suppressing or replacing them. However, repetitive propagation
of the radical isomerization reactions does not promote overall progress (measured by time passed) of thermal cracking, nor does it improve the quality of Monte Carlo results. We thereby place a limit of repetition number by introducing an ‘inhibitor’ on consecutive execution of radical isomerization steps. The length of a radical isomerization sequence is confined by a random integer generated at the first isomerization reaction of that sequence. When that number is reached, the sequence is interrupted to execute a reaction from the list of medium and slow reactions. We conducted a test experiment to constrain the best choice for the sample interval of the uniformly distributed random number. The goal is to have isomerization sequences long enough (i.e., high maximum and minimum boundaries) so that the probabilistic distribution of final radical position is indistinguishable from that of uninhibited execution of isomerization events, which represents a metastable steady-state scenario because of the large number of reactions. On the other hand, we also aim to force these sequences to be as short as possible to be computationally efficient. We ran this test on an \( n \)-nonane (\( C_{9}H_{20} \)) molecule, as its size is close to a typical alkyl chain on the substrates that we are studying (the average alkyl chain length in the kIA kerogen model is 8.0). For each combination of the maximum and minimum numbers, we repeat a chain of the isomerization step (until interrupted) for many times to accumulate a distribution of final radical position. The outcome of uninhibited execution of isomerization is estimated with very large interval numbers. We found that a combination of a maximum number of 30 and minimum number of 10 is high enough to reproduce the outcome of an uninhibited isomerization chain. (Figure A1)

Although medium reactions are also orders of magnitude faster than slow reactions, the intrinsic design of the radical chain reactions makes it unnecessary to limit medium reactions. When a system keeps executing medium and fast reactions, radicals automatically quench due to radical-loss through the medium reactions of H-abstraction with water and radical recombination. Therefore, homolytic cleavage (slow reaction) is the only reaction type that does not require a radical as a reactant and becomes the only available option when the system has no radicals. Effectively, the model proceeds by executing a homolytic cleavage reaction, then executing a series of one or more fast and medium reactions until all radicals are consumed, then making the next homolytic cleavage reaction, and so forth.

(4) We also modified models employing chemistry scheme B by representing molecular graphs with two matrices, a connection matrix and a bond matrix, during the chemical evolution of simulations in which the graph size is big (>100,000 nodes). This is especially important for chemistry scheme B because the entire system has to be stored in one graph in order to allow the two-body reaction of radical recombination. The computational costs of graph operations, such as addition/removal of edges and search for node neighbors, scale with the number of nodes and edges in a graph. In these cases, we replace the graph data structure with two \( n \times 4 \) matrices (where \( n \)=number of nodes). In the connection
matrix, each row lists the indices of target nodes in connections with the row node (the maximum number of connections is 4 and any vacant element is left equal to 0). In the bond matrix, each row lists the bond orders (where aromatic bonds are assigned an order of 1.5) at the same position of the connection matrix. This representation is more efficient compared with the more common way of using an $n \times n$ adjacency matrix. For example, the complexity of a search algorithm (i.e., to find neighbors) using this method is $O(1)$, vs. $O(n)$ when using an adjacency matrix.

**Appendix C. Data for Cracking Model Oil and Kerogen at $F_{C1}=1\%$.**

**Figures**

Figure 1: A flow-chart of the kMC model. The pie chart shows an example of iteration that selects from a pool of many permitted elementary reactions.

Figure 2: Illustration of reaction types included in our models.

Figure 3: Chemical paradigm of scheme A and scheme B.

Figure 4: Cumulative average central-terminal isotope fractionation of residual butane in a butane cracking test under scheme A at 500K. X-axis is the number of simulations. The red line shows analytical solution. The dashed black lines envelope +/- 1 standard error of the mean for the final result (last point in the series), estimated using the counting statistics of all simulations.

Figure 5: Compound-specific carbon isotope plot for natural gas cracked from different substrates under scheme A and scheme B, at methane transformation ratio $F_{C1}=1\%$. Gray area shows common isotope composition of natural gas derived from type II kerogen (Zou et al., 2007). Precursor material was assigned have $\delta^{13}C= -25 \text{‰}$ and $\delta^2H= -100 \text{‰}$.

Figure 6: Position-specific carbon and hydrogen isotope values of $n$-alkanes produced at early stage (bond cleavage=20% and $F_{C1}=1\%$) from substrates under scheme A, scaled in VPDB and VSMOW respectively. Oil: modeled saturated fraction of oil; kIA: kIA kerogen model. Temperature is at 500K. Precursor material was assigned $\delta^{13}C= -25 \text{‰}$ and $\delta^2H= -100 \text{‰}$.

Figure 7: Position-specific carbon and hydrogen isotope values of $n$-alkanes produced at early stage (bond cleavage=20% and $F_{C1}=1\%$) from substrates under scheme B, scaled in VPDB and VSMOW respectively. Oil: modeled saturated fraction of oil; kIA: kIA kerogen model. Temperature is at 500K. Precursor material was assigned $\delta^{13}C= -25 \text{‰}$ and $\delta^2H= -100 \text{‰}$. 
Figure 8: Evolution of compound-specific $\delta^{13}$C in a series of experiments with varying reaction progress. Kerogen molecular model kIA is used as substrate. Precursor material was assigned $\delta^{13}$C = -25‰ and $\delta^2$H = -100‰.

Figure 9: Evolution of gas dryness ratio, C1/(C2+C3), in a series of kMC experiments vs. reaction progress. Kerogen molecular model kIA is used as substrate. Bars show the range of variation for gases from pyrolysis experiments (Evans and Felbeck, 1983; Saxby and Riley, 1984; Seewald et al., 1998; Pan et al., 2008; Lewan et al., 2008; Liguo et al., 2009; Shao et al., 2018) and thermogenic natural gas deposits (Sherwood et al., 2017; Milkov and Etiophe, 2018; Milkov et al., 2020), respectively.

Figure 10: Carbon isotope difference between the central and terminal positions of propane (see Xie et al. (2020) for definition) vs. reaction progress. Kerogen molecular model kIA was used as substrate. Bars show range of variation for gases from pyrolysis experiments (Piasecki et al., 2018; Gilbert et al., 2019) and non-biodegraded thermogenic gas deposits (Piasecki et al., 2018; Liu et al., 2018; Liu et al., 2019; Zhao et al., 2020; Julien et al., 2020).

Figure 11: Evolution of C1/(C2+C3) ratio and methane $\delta^{13}$C in experiments varying reaction progress. Kerogen molecular model kIA was used as substrate. Open circles represent natural gas data from a few major shale gas plays compiled from literature (Rodriguez and Philp, 2010; Sherwood et al., 2017; Byrne et al., 2018; Liu et al., 2019; Abrams and Thomas, 2020).

Figure 12: Distribution of apparent ‘Chung plot’ slopes for global thermogenic natural gases globally. We compiled data from several publications (Dai et al., 2012; J. X. Dai et al., 2014; Ni et al., 2015; Sherwood et al., 2017; Thiagarajan et al., 2020b) and each individual sample by linear regression. For criteria of thermogenic origin, refer to Xie et al. (2021). Samples with at least four compound-specific $\delta^{13}$C measurements are considered (i.e., minimum requirement is C1–C4 isotope data). kMC bar shows range of slopes from kMC results for cracking kIA kerogen model at reaction progress $F_{C1}$=1–29% under chemistry scheme B (500K) (Table 1). Equilibrium bar shows range of slopes fitted from $\delta^{13}$C values of C1–C5 n-alkanes under thermodynamic equilibrium, constructed using theoretical estimates (Thiagarajan et al., 2020b) at 100–250°C and $\delta^{13}$Cmethane=-45‰.

Figure 13: Molecular and isotopic trends of hydrous pyrolysis experiment (Wu et al., 2019), and a series of kMC experiments using the same temperature-time conditions and kIA as substrate. Hydrous pyrolysis experiments are green diamonds while kMC results are red circles. Precursor material was assigned $\delta^{13}$C = -28‰ and $\delta^2$H = -140‰.

Figure A1: Results of testing the appropriate range of the radical isomerization chain limit number. These panels show the distribution of the final radical position after consecutive execution of isomerization steps.
with different configurations of maximum and minimum step numbers. An inhibitor is selected from that interval with uniform probability. In each test, we repeat the isomerization chain 50,000 times to generate a distribution. The radical position is initialized at the second carbon on an n-nonane molecule (i.e., 2-nonyl radical) for all tests. T=500K.

Tables
Table 1. Molecular and isotopic compositions of \( n \)-alkanes produced by cracking kerogen at different extent of reaction.

Reference cited


Schoell M. (1980) The hydrogen and carbon isotopic composition of methane from natural gases of


Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
<table>
<thead>
<tr>
<th>Scheme</th>
<th>( F_{c1} )</th>
<th>( C1/(C2+C3) )</th>
<th>( C1 )</th>
<th>( C2 )</th>
<th>( C3 )</th>
<th>( nC4 )</th>
<th>( nC5 )</th>
<th>( C1 )</th>
<th>( C2 )</th>
<th>( C3 )</th>
<th>( nC4 )</th>
<th>( nC5 )</th>
<th>( \varepsilon_{13C_{CT}} )</th>
<th>( C3 )</th>
<th>( 1\sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.4%</td>
<td>0.5</td>
<td>34.4</td>
<td>29.2</td>
<td>30.0</td>
<td>28.4</td>
<td>44.6</td>
<td>314.0</td>
<td>351.1</td>
<td>259.5</td>
<td>317.8</td>
<td>295.4</td>
<td>3.0</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1%</td>
<td>0.2</td>
<td>33.5</td>
<td>36.3</td>
<td>30.2</td>
<td>32.7</td>
<td>28.8</td>
<td>318.0</td>
<td>332.1</td>
<td>285.3</td>
<td>272.6</td>
<td>242.1</td>
<td>5.1</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.2%</td>
<td>0.1</td>
<td>43.0</td>
<td>37.0</td>
<td>29.7</td>
<td>24.5</td>
<td>21.4</td>
<td>387.6</td>
<td>346.8</td>
<td>290.0</td>
<td>245.6</td>
<td>211.8</td>
<td>5.1</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.7%</td>
<td>0.2</td>
<td>39.4</td>
<td>31.2</td>
<td>25.5</td>
<td>12.3</td>
<td>-5.7</td>
<td>420.4</td>
<td>337.1</td>
<td>241.3</td>
<td>167.3</td>
<td>111.9</td>
<td>6.9</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.4%</td>
<td>0.3</td>
<td>38.3</td>
<td>26.5</td>
<td>15.9</td>
<td>-</td>
<td>-</td>
<td>427.7</td>
<td>311.1</td>
<td>183.5</td>
<td>-</td>
<td>-</td>
<td>7.0</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28.4%</td>
<td>0.8</td>
<td>37.4</td>
<td>21.3</td>
<td>29.9</td>
<td>-</td>
<td>-</td>
<td>434.4</td>
<td>277.4</td>
<td>103.5</td>
<td>-</td>
<td>-</td>
<td>25.0</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52.6%</td>
<td>2.2</td>
<td>34.6</td>
<td>14.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>441.3</td>
<td>217.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>77.5%</td>
<td>6.9</td>
<td>33.8</td>
<td>-1.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>429.4</td>
<td>103.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1.0%</td>
<td>1.7</td>
<td>53.6</td>
<td>43.2</td>
<td>32.1</td>
<td>29.8</td>
<td>26.0</td>
<td>341.5</td>
<td>272.4</td>
<td>227.4</td>
<td>214.9</td>
<td>182.9</td>
<td>4.5</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.1%</td>
<td>1.7</td>
<td>53.7</td>
<td>43.5</td>
<td>29.1</td>
<td>27.4</td>
<td>27.6</td>
<td>341.2</td>
<td>272.5</td>
<td>220.5</td>
<td>212.0</td>
<td>180.9</td>
<td>3.7</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.4%</td>
<td>1.9</td>
<td>52.8</td>
<td>43.1</td>
<td>29.5</td>
<td>27.3</td>
<td>27.6</td>
<td>339.8</td>
<td>268.5</td>
<td>215.6</td>
<td>204.0</td>
<td>180.6</td>
<td>3.4</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.7%</td>
<td>2.1</td>
<td>50.0</td>
<td>41.7</td>
<td>27.4</td>
<td>21.9</td>
<td>25.2</td>
<td>336.6</td>
<td>267.4</td>
<td>206.7</td>
<td>203.8</td>
<td>175.6</td>
<td>4.5</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.6%</td>
<td>2.2</td>
<td>49.7</td>
<td>39.7</td>
<td>25.6</td>
<td>25.7</td>
<td>27.7</td>
<td>332.8</td>
<td>255.4</td>
<td>191.8</td>
<td>184.7</td>
<td>158.7</td>
<td>4.8</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.9%</td>
<td>2.2</td>
<td>49.4</td>
<td>39.6</td>
<td>24.2</td>
<td>24.3</td>
<td>22.0</td>
<td>330.6</td>
<td>250.0</td>
<td>185.0</td>
<td>173.6</td>
<td>152.5</td>
<td>1.5</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.8%</td>
<td>2.7</td>
<td>46.3</td>
<td>37.9</td>
<td>21.4</td>
<td>23.5</td>
<td>28.1</td>
<td>309.1</td>
<td>216.5</td>
<td>133.5</td>
<td>-99.3</td>
<td>-72.9</td>
<td>-0.3</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73.6%</td>
<td>6.2</td>
<td>37.6</td>
<td>29.6</td>
<td></td>
<td></td>
<td></td>
<td>267.2</td>
<td>159.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77.7%</td>
<td>7.0</td>
<td>36.9</td>
<td>30.6</td>
<td></td>
<td></td>
<td></td>
<td>263.9</td>
<td>150.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78.5%</td>
<td>7.7</td>
<td>38.0</td>
<td>28.8</td>
<td></td>
<td></td>
<td></td>
<td>263.9</td>
<td>142.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: A flow-chart of the kMC model. The pie chart shows an example of iteration that selects from a pool of many permitted elementary reactions.
Figure 2: Illustration of reaction types included in our models.
Figure 3: Chemical paradigm of scheme A and scheme B.
Figure 4: Cumulative average central-terminal isotope fractionation of residual butane in a butane cracking test under scheme A at 500K. X-axis is the number of simulations. The red line shows analytical solution. The dashed black lines envelope +/- 1 standard error of the mean for the final result (last point in the series), estimated using the counting statistics of all simulations.
Figure 5: Compound-specific carbon isotope plot for natural gas cracked from different substrates under scheme A and scheme B, at methane transformation ratio $F_{CI}=1\%$. Gray area shows common isotope composition of natural gas derived from type II kerogen (Zou et al., 2007). Precursor material was assigned have $\delta^{13}C = -25 \%$ and $\delta^{2}H = -100 \%$. 
Figure 6: Position-specific carbon and hydrogen isotope values of \( n \)-alkanes produced at early stage (bond cleavage=20% and \( F_{Cl}=1\% \)) from substrates under scheme A, scaled in VPDB and VSMOW respectively. Oil: modeled saturated fraction of oil; kIA: kIA kerogen model. Temperature is at 500K. Precursor material was assigned \( \delta^{13}C = -25 \% \) and \( \delta^2H = -100 \% \).
Figure 7: Position-specific carbon and hydrogen isotope values of \( n \)-alkanes produced at early stage (bond cleavage=20\% and \( F_{\text{Cl}} = 1\% \)) from substrates under scheme B, scaled in VPDB and VSMOW respectively. Oil: modeled saturated fraction of oil; kIA: kIA kerogen model. Temperature is at 500K. Precursor material was assigned \( \delta^{13}\text{C} = -25 \% \) and \( \delta^2\text{H} = -100 \% \).
Figure 8: Evolution of compound-specific $\delta^{13}C$ in a series of experiments with varying reaction progress. Kerogen molecular model kIA is used as substrate. Precursor material was assigned $\delta^{13}C = -25$‰ and $\delta^2H = -100$‰.
Figure 9: Evolution of gas dryness ratio, C1/(C2+C3), in a series of kMC experiments vs. reaction progress. Kerogen molecular model kIA is used as substrate. Bars show the range of variation for gases from pyrolysis experiments (Evans and Felbeck, 1983; Saxby and Riley, 1984; Seewald et al., 1998; Pan et al., 2008; Lewan et al., 2008; Liguoz et al., 2009; Shao et al., 2018) and thermogenic natural gas deposits (Sherwood et al., 2017; Milkov and Etiope, 2018; Milkov et al., 2020), respectively.
Figure 10: Carbon isotope difference between the central and terminal positions of propane (see Xie et al. (2020) for definition) vs. reaction progress. Kerogen molecular model kIA was used as substrate. Bars show range of variation for gases from pyrolysis experiments (Piasecki et al., 2018; Gilbert et al., 2019) and non-biodegraded thermogenic gas deposits (Piasecki et al., 2018; Liu et al., 2018; Liu et al., 2019; Zhao et al., 2020; Julien et al., 2020).
Figure 11: Evolution of C1/(C2+C3) ratio and methane $\delta^{13}$C in experiments varying reaction progress. Kerogen molecular model kIA was used as substrate. Open circles represent natural gas data from a few major shale gas plays compiled from literature (Rodriguez and Philp, 2010; Sherwood et al., 2017; Byrne et al., 2018; Liu et al., 2019; Abrams and Thomas, 2020).
Figure 12: Distribution of apparent ‘Chung plot’ slopes for global thermogenic natural gases globally. We compiled data from several publications (Dai et al., 2012; JinXing Dai et al., 2014; Ni et al., 2015; Sherwood et al., 2017; Thiagarajan et al., 2020b) and each individual sample by linear regression. For criteria of thermogenic origin, refer to Xie et al. (2021). Samples with at least four compound-specific δ^{13}C measurements are considered (i.e., minimum requirement is C_1–C_4 isotope data). kMC bar shows range of slopes from kMC results for cracking kIA kerogen model at reaction progress F_{C_1}=1–29% under chemistry scheme B (500K) (Table 1). Equilibrium bar shows range of slopes fitted from δ^{13}C values of C_1–C_4 n-alkanes under thermodynamic equilibrium, constructed using theoretical estimates (Thiagarajan et al., 2020b) at 100–250°C and δ^{13}C_{methane} = 45%.
Figure 13: Molecular and isotopic trends of hydrous pyrolysis experiment (Wu et al., 2019), and a series of kMC experiments using the same temperature-time conditions and kIA as substrate. Hydrous pyrolysis experiments are green diamonds while kMC results are red circles. Precursor material was assigned $\delta^{13}C = -28\%$ and $\delta^2H = -140\%$. 
Figure A1: Results of testing the appropriate range of the radical isomerization chain limit number. These panels show the distribution of the final radical position after consecutive execution of isomerization steps with different configurations of maximum and minimum step numbers. An inhibitor is selected from that interval with uniform probability. In each test, we repeat the isomerization chain 50,000 times to generate a distribution. The radical position is initialized at the second carbon on an n-nonane molecule (i.e., 2-nonyl radical) for all tests. T=500K.