SUPPLEMENTARY MATERIAL

Impact of Molecular Configuration on the Bond Breaking Rates of Hydrocarbons: A Computational Study

Jiang Wang,*,† Zhiling Li,† and Wenli Zhang*,‡

†*College of Science, Guizhou Institute of Technology, Boshi Road, Dangwu Town, Gui'an New District, Guizhou, 550025, China*

‡*School of Transportation Engineering, Guizhou Institute of Technology, Boshi Road, Dangwu Town, Gui'an New District, Guizhou, 550025, China*

E-mail: cwangjiang@git.edu.cn; wenlizhang@git.edu.cn

SUPPLEMENTARY TEXTS

Simulation Workflow

Algorithm S1 Workflow of the ReaxFF molecular simulation.

procedure REAXFF-MD(k) ▷ The k-th simulation
    t=0 ▷ Lifetime t initialization
    i=1 ▷ Fragment ID i initialization
    Break_Flag=False
    while Break_Flag=False **do**
        Simulate the i-th fragment for 10 ps.
        Check if molecule dissociate.
        if Molecule does not dissociate **then**
            $t = t$ + 10 ps
            $i = i + 1$
        else if Molecule dissociate at time t_0 **then**
            $t = t + t_0$
            Break_Flag=True
            r = configuration of the molecule at t_0
        end if
    end while
    return t, **r**
end procedure

Due to the exponential distribution governing the lifetime of molecules, the dissociation time in each simulation varies. Imposing a short time limit in the simulation may not guarantee bond breaking, while setting a long simulation time limit might be inefficient as many molecules may dissociate early. To balance these considerations, we implemented the simulation workflow as listed in Algorithm S1: In each simulation, we conducted a series of fragment simulations, with each fragment lasting for 10 ps. After completing a fragment, we analyzed the data to determine if any molecular dissociation occurred. If the molecule remained intact in the backbone, we extended the simulation by initiating the next round of fragment simulation. This process continued until the moment of molecular dissociation was identified, at which point the iteration was halted. This simulation strategy

efficiently identifies the lifetime of both early and late dissociating molecules while minimizing computational costs. Three representative simulation trajectories in the Supplementary Material illustrate the dissociation of three molecules.

Diffusion Maps Nonlinear Manifold Learning

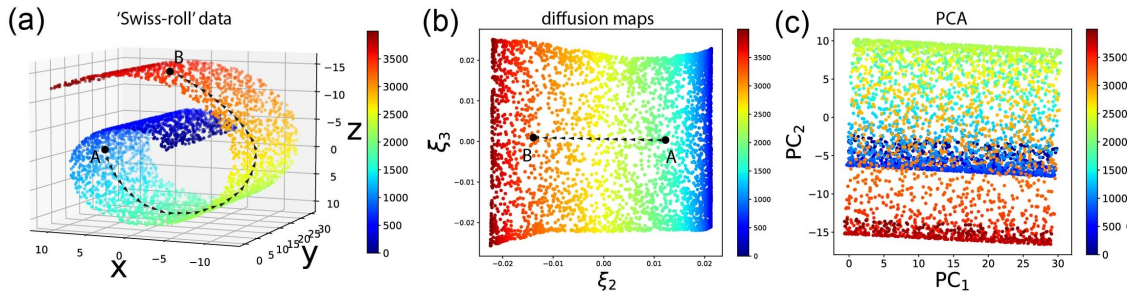

Figure S1: A toy model illustrates the distinction between diffusion maps and principal component analysis (PCA). In (a), the original data form an intrinsically 2-dimensional nonlinear 'Swiss-roll' structure within a 3-dimensional Euclidean space. The path connecting points A and B traverses the nonlinear manifold. (b) Diffusion maps (dMaps) effectively extracts the 2D intrinsic manifold and represents it using the top eigenvectors. (c) In contrast, traditional linear dimensionality reduction methods like principal component analysis (PCA) struggle to capture the nonlinear manifold, as evidenced by the presence of overlapping points in the space spanned by the top principal components.

Proposed by Coifman et al.,[1–4] diffusion maps model data as a random walk or diffusion process on an intrinsic low-dimensional manifold. This manifold can be nonlinear and curved, embedded within a high-dimensional Euclidean space, as illustrated in Figure S1(a): data points diffusing on the intrinsic 2D nonlinear Swiss-roll manifold in 3D Euclidean space. The dMaps algorithm extracts intrinsic dimensions corresponding to the slowest diffusion modes, as shown in Figure S1(b). For instance, from point A to B, the slowest diffusion mode follows the arrow-dashed path in the Swiss-roll nonlinear manifold. In this regard, dMaps outperforms regular linear approaches like principal component analysis (PCA), as depicted in Figure S1(c). Previously, both our work and that of others have successfully applied dMaps to various molecular dynamics (MD) simulation data, including protein and alkane folding simulations, self-assembly of asphaltene, and patchy colloid particles.[5–13]

In this research, we would apply dMaps to the equilibrated n-tridecane simulation data, to find the effective dimensionality of its configuration and extract the low dimensional collective variable. The first step is to compute the pairwise distance d_{ij} between each two configurations $\mathbf{r}_i, \mathbf{r}_j$ from equilibrium simulation trajectory, and $i, j = 1, 2, 3, \cdots, N_c$, where N_c is the number of configurations being sampled. We use the rotational minimized root mean square distance (RMSD) between $\mathbf{r}_i$ and $\mathbf{r}_j$ as the pairwise distance, Kabsch algorithm is applied to carry out the calculation of d_{ij}.[14]

The second step involves creating the hopping probability matrix, $\mathbf{A}$, by incorporating a Gaussian kernel into the distance matrix: $A_{ij} = \exp\left(-d_{ij}^2/2\epsilon\right)$, where ϵ is the bandwidth.[1–4] This operation ensures that if two configurations $\mathbf{r}_i, \mathbf{r}_j$ have a large distance between them ($d_{ij} \to \infty$), their hopping probability becomes 0 ($A_{ij} \to 0$), while adjacent pairs have substantial hopping probabilities ($d_{ij} \to 0 \Rightarrow A_{ij} \to 1$).

The third step is to obtain the Markov matrix, $\mathbf{M}$, defined as $\mathbf{M} = \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{D}$ is the normalization matrix—a diagonal matrix with elements $D_{ii} = \sum_{k=1}^{N_C} A_{ik}$.

The last step involves performing an eigen decomposition of the Markov matrix $\mathbf{M}$. This process yields a series of eigenvalues in descending order: $[\lambda_1, \lambda_2, \ldots, \lambda_{N_C}] \in (0, 1]$ where $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N_C} > 0$. Associated with these eigenvalues are eigenvectors: $[\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_{N_C}]$. Note that the Markov matrix $\mathbf{M}$ is a $N_C \times N_C$ matrix, and eigenvector $\vec{\phi}_i$ has $N_C c$ components. Eigenvectors corresponding to large eigenvalues represent the slowest diffusion modes, capturing the most important intrinsic dimensions. It's worth noting that among all eigenvalues, $\lambda_1 = 1$, and the corresponding eigenvector $\boldsymbol{\xi}_1$ is an all-one vector, signifying the steady state, which can be disregarded.

For dimensionality reduction, each high-dimensional state point is projected into the low-dimensional space spanned by the top k eigenvectors. For instance, the i-th configuration from the simulation will have a $k-1$ dimensional coordinate: $\text{configuration}_\text{i} \to [\boldsymbol{\xi}_2(i), \boldsymbol{\xi}_3(i), \ldots, \boldsymbol{\xi}_k(i)]$, where $\boldsymbol{\xi}_k(i)$ denotes the i-th component of the k-th eigenvector $\boldsymbol{\xi}_k$. The value of k can be determined based on the location of the gap in the eigenvalue spectrum.

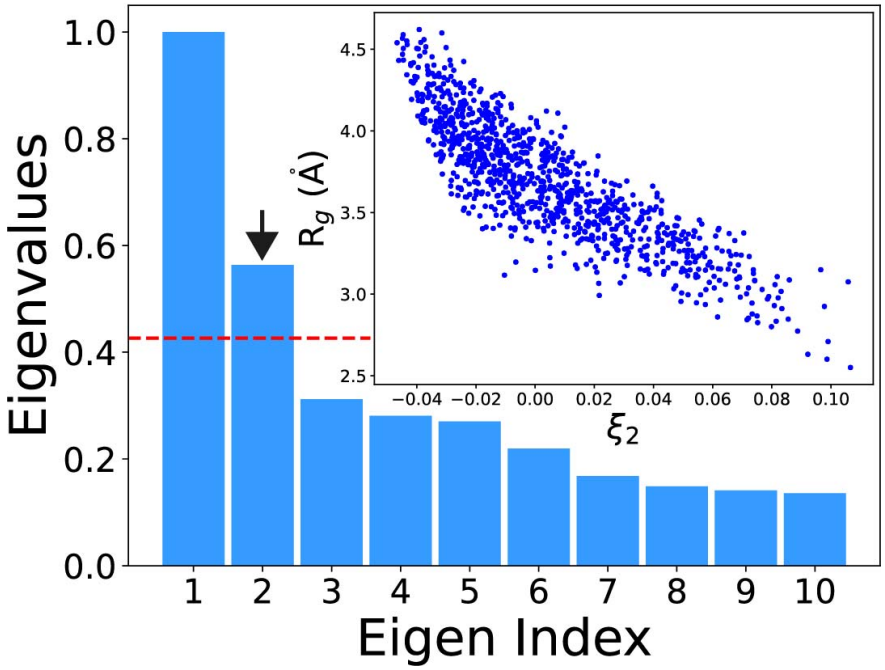

Figure S2: Eigenvalue spectrum of the dMaps, and the inset shows the relation between ξ_2 and the radius of gyration R_g.

In Fig. S2, the eigenvalue spectrum resulting from dMaps analysis is presented. A discernible gap is observed between the second and third eigenvalues, highlighted by the horizontal dashed line. This gap signifies that the effective dimensionality of the molecule's configuration is one, and the low-dimensional collective variable can be adequately represented using $\boldsymbol{\xi}_2$, as indicated by the black arrow. The inset illustrates the relationship between $\boldsymbol{\xi}_2$ and the radius of gyration R_g, showcasing a negative correlation with a Pearson correlation coefficient of -0.84. This fairly strong correlation suggests that R_g can effectively capture the overall molecular configuration, and span the configurational dependent bond breaking rate, so $\mathbf{r}$ could be replaced by R_g in all equations: $\mathbf{r} \rightarrow R_g$.

Comparison with DFT Results

Besides the simulation with the oxygen moleucles at present, we also conduct pyrolysis simulation of three molecules in the gas-phase, where there is no oxygen molecules, the

obtained results are shown in Figure S3, Figure S4, and Figure S5.

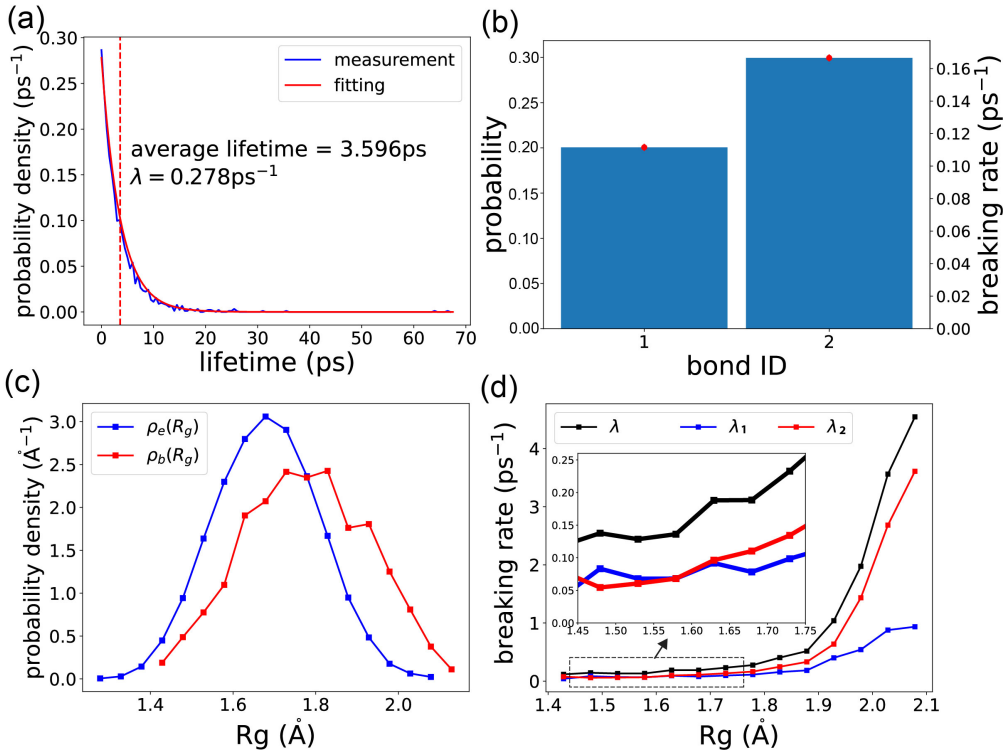

Figure S3: Results of the pyrolysis simulation of n-pentane. (a) The lifetime distribution of the modified data. The vertical dashed line marks the average lifetime $\bar{t} = 3.596$ ps, the red curve represents the exponential distribution with the corresponding rate constant $\lambda = 1/\bar{t} = 0.278$ ps^{-1}. (b) Overall probabilities and corresponding rate constants for single bond 1 and single bond 2. (c) The equilibrium distribution of the R_g, $\rho_e(R_g)$; and the distribution of the R_g under the condition that the molecule just breaks, $\rho_b(R_g)$. (d) Bond dissociation rate for the entire molecule ($\lambda(R_g)$, black curve); for bond 1 ($\lambda(R_g, 1)$, blue curve); for bond 2 ($\lambda(R_g, 2)$, red curve).

In the compared DFT paper,[15] parameters A, n, E_a in the modifyed Arrhenius equation ($k = AT^n \exp(-E_a/RT)$) are fitted and reported, as indicated in the Figure S6, which is copied from the referenced paper.[15] Note that parameters for only bond 1 and 2, which producing CH_3 and C_2H_5 radicles are reported, as the dissociation rate constant for other C-C bonds are similar to bond 2, which produce C_2H_5. Using these parameters, the bond dissociation rate constant at our simulation temperature (3500K) could be predicted. Table 1 shows all rate constants predicted from the paper and from our approach.

Compare the results and conclusions from the paper with our results, we can conclude

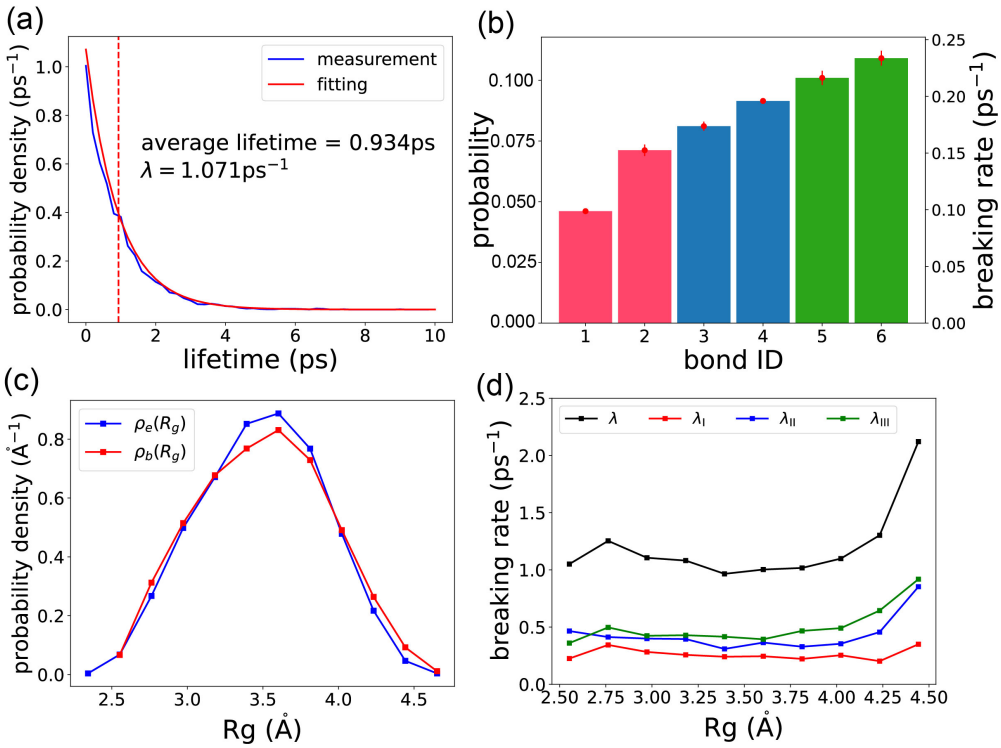

Figure S4: Results of the pyrolysis simulation of n-pentane. (a) The lifetime distribution of the modified data. The vertical dashed line marks the average lifetime $\bar{t} = 0.934$ ps, the red curve represents the exponential distribution with the corresponding rate constant $\lambda = 1/\bar{t} = 1.071$ ps^{-1}. (b) Overall probabilities and corresponding rate constants for each single bond. (c) The equilibrium distribution of the R_g, $\rho_e(R_g)$; and the distribution of the R_g under the condition that the molecule just breaks, $\rho_b(R_g)$. (d) Bond dissociation rate for the entire molecule ($\lambda(R_g)$, black curve); for unified bond I ($\lambda(R_g, 1)$, red curve); for unified bond II ($\lambda(R_g, 1)$, blue curve); for unified bond III ($\lambda(R_g, 1)$, green curve).

Table S1: Bond dissociation rate constants predicted from Ref[15] and from our calculation for n-pentane and n-tridecane. Dissociation of two bonds are included: bond 1: $R-CH_3 \rightarrow R + CH_3$, and bond 2: $R - C_2H_5 \rightarrow R + C_2H_5$. Where $R - CH_3$ and $R - C_2H_5 = C_5H_{12}$ or $C_{13}H_{28}$. The unit of the rate constant are s^{-1}.

	n-pentane bond 1 $R - CH_3$	n-pentane bond 2 $R - C_2H_5$	n-tridecane bond 1 $R - CH_3$	n-tridecane bond 2 $R - C_2H_5$
Ref	8.87e+10	4.04e+11	2.96e+10	1.63e+11
Our	1.09e+11	1.62e+11	9.33e+10	1.51e+11

that, based on the following four aspects, ReaxFF simulation (combined with our calculating method) could qualitatively capture bond breaking properties:

1. DFT shows that all C-C bonds at different positions in one alkane chain have similar

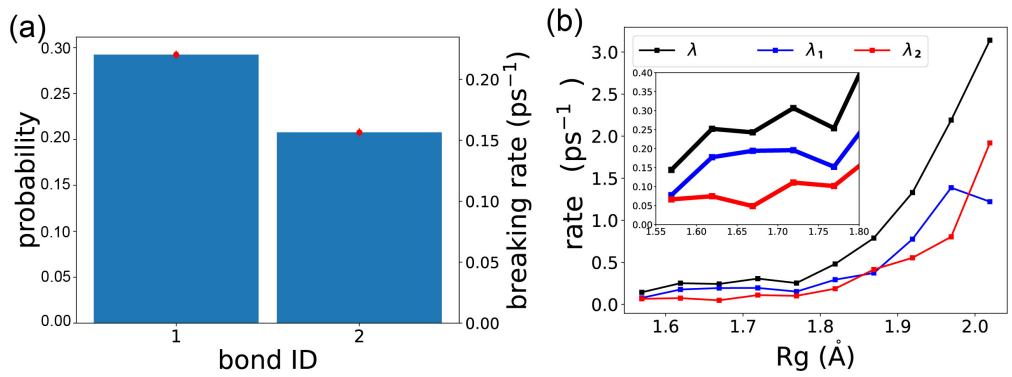

Figure S5: Results of the pyrolysis simulation of 1,3-Propanediol. (b) Overall probabilities and corresponding rate constants for single bond 1 and single bond 2. (b) Bond dissociation rate for the entire molecule ($\lambda(R_g)$, black curve); for bond 1 ($\lambda(R_g, 1)$, blue curve); for bond 2 ($\lambda(R_g, 2)$, red curve).

dissociation rate constants, the difference is less than 1 order of magnitude, and this is consistent with our result.

2. Comparing our rate constant values with the values predicted from the referenced paper in Table 1, although they are not closely similar, they exhibit a comparable order of magnitude. This suggests that ReaxFF qualitatively captures the dissociation kinetics.

Two following trends are also consistent:

3.Even though bond dissociation rate constant are similar to each other within the same alkane chain, the dissociation of the bond 1 ($R-CH_3 \rightarrow R+CH_3$) is harder than bonds at other locations, this is consistent with our results, with bond 1 has significantly smaller rate constant than bond 2.

4. For the bond at the same position, as the alkane size increase, its dissociation rate constant decreases.

All the above consistencies indicate that ReaxFF incorporated with our method could qualitatively estimate bond dissociation rate constant, and capture alkane cracking properties.

In fact, our approach can be seen as a 'data-driven' method for estimating rate constants, which differs from existing approaches such as VTST. In VTST theory, one must identify all potential reactions, locate the transition states, and compute rate constants for each pathway

Table 3. Rate parameters of reaction $RCH_3 \rightarrow R + CH_3$ (RCH_3 = n-C_nH_{2n+2}, n = 2-17) in modified Arrhenius form: A (s^{-1}), n and Ea (kcal/mol).

n-Alkanes	A	n	E_a	n-Alkanes	A	n	E_a
C_2H_6	3.54E+22	−1.57	89.0	$C_{10}H_{22}$	1.93E+26	−2.86	86.0
C_3H_8	4.63E+22	−1.43	87.4	$C_{11}H_{24}$	4.85E+26	−2.83	85.2
C_4H_{10}	8.15E+23	−1.84	88.2	$C_{12}H_{26}$	1.87E+26	−2.74	84.4
C_5H_{12}	7.93E+23	−1.87	87.7	$C_{13}H_{28}$	1.44E+27	−2.98	84.0
C_6H_{14}	3.41E+24	−2.02	87.3	$C_{14}H_{30}$	1.71E+27	−3.07	83.7
C_7H_{16}	2.04E+24	−2.06	86.7	$C_{15}H_{32}$	8.78E+26	−3.11	83.1
C_8H_{18}	1.44E+26	−2.63	87.0	$C_{16}H_{34}$	3.91E+26	−2.93	82.2
C_9H_{20}	3.59E+26	−2.72	86.3	$C_{17}H_{36}$	3.51E+26	−2.98	81.7

Table 4. Rate parameters of reaction $RC_2H_5 \rightarrow R + C_2H_5$ (RC_2H_5 = n-C_nH_{2n+2}, n = 4-17) in modified Arrhenius form: A (s^{-1}), n and Ea (kcal/mol).

n-Alkanes	A	n	E_a	n-Alkanes	A	n	E_a
C_4H_{10}	9.82E+21	−1.01	85.3	$C_{11}H_{24}$	1.48E+26	−2.49	83.7
C_5H_{12}	1.78E+23	−1.51	86.3	$C_{12}H_{26}$	1.65E+26	−2.51	83.1
C_6H_{14}	8.0E+21	−1.10	85.1	$C_{13}H_{28}$	5.25E+25	−2.40	82.3
C_7H_{16}	5.94E+24	−1.88	85.4	$C_{14}H_{30}$	8.43E+25	−2.43	81.8
C_8H_{18}	1.26E+24	−1.94	85.1	$C_{15}H_{32}$	3.26E+28	−3.30	82.8
C_9H_{20}	3.06E+24	−1.94	84.2	$C_{16}H_{34}$	1.42E+24	−2.03	80.2
$C_{10}H_{22}$	9.15E+23	−1.99	83.7	$C_{17}H_{36}$	2.05E+25	−2.4	80.2

Figure S6: Fitted values for the modified Arrhenius equation for varies alkane chain sizes. This table is copied from Ref.[15]

before combining them to determine the final reaction or dissociation rate. This requires a clear depiction of each reaction.

In our 'data-driven' approach, we bypass the need for detailed reaction knowledge. Instead, we simulate all potential reactions using reactive MD or QM/MD. With a sufficiently large number of trajectories, all possible reaction paths are sampled. From these statistical results, we derive reaction rate constants in a 'top-down' fashion, contrasting with the traditional 'bottom-up' approach. The accuracy of capturing all potential reactions in simulations becomes crucial. To achieve this, methods such as Car–Parrinello molecular dynamics and QM/MM techniques integrate equation of motions into QM, ensuring accurate representation of reactions derived from QM.

In this research paper, we utilize ReaxFF simulation rather than QM for its speed and effectiveness as a prototype to validate our method. Our results demonstrate that ReaxFF yields qualitatively accurate outcomes. In future research, we intend to integrate our method with more precise QM/MD simulation data.

C-H Bond Breaking Rate Constant

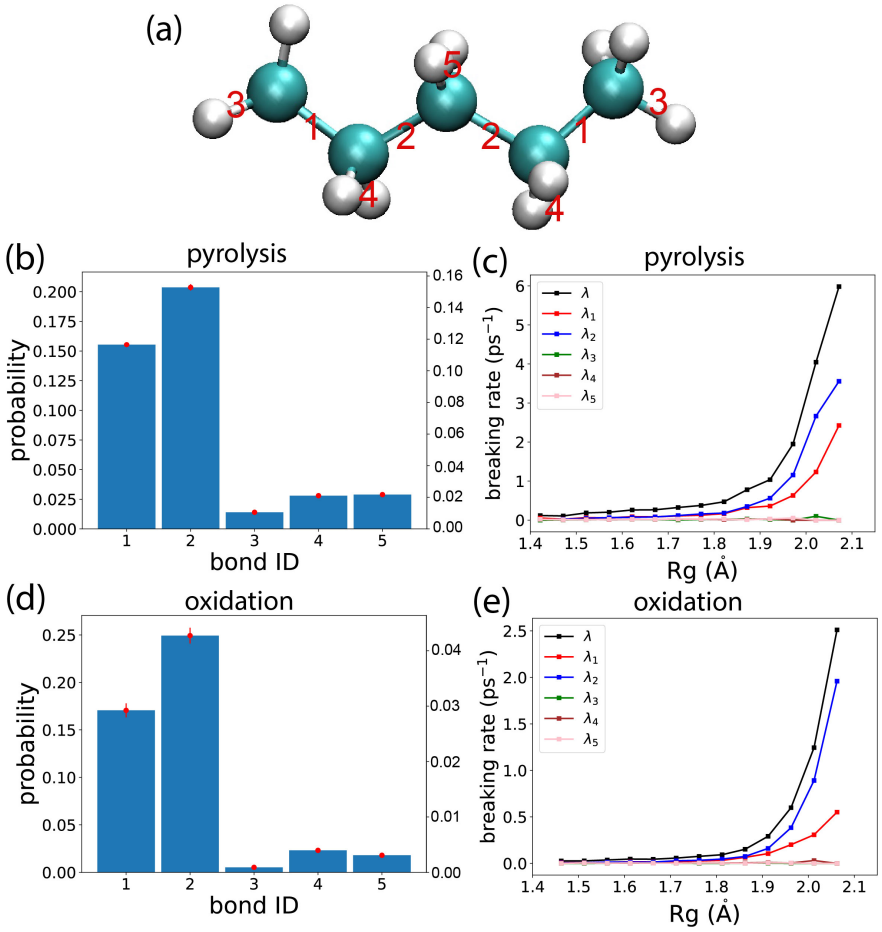

Figure S7: (a) Indices for each bond, symmetrical bonds are assigned the same index. There are five types of bonds in total: Bonds 1 and 2 are C-C bonds; Bond 3 is the C-H bond on the edge carbon atom; Bonds 4 and 5 are C-H bonds on the central carbon atoms. (b) Single bond breaking probabilities and breaking rate in the pyrolysis simulation (with no oxygen molecule). (c) Rate constants for each bond as a function of the radius of gyration. (d) Single bond breaking probabilities and breaking rate in the oxidation simulation (with oxygen at present). (e) Rate constants for each bond as a function of the radius of gyration in an oxygenated environment.

To compare how C-H bonds break relative to C-C bonds in *n*-pentane, we conducted additional simulations in pyrolysis and oxidation, considering C-H bond breakage. There are five bond types, accounting for symmetry: two C-C bonds and three C-H bonds, as shown in Figure S7a. Figure S7b demonstrates that C-H bonds have significantly lower overall dissociation rate constants than C-C bonds, with terminal H atoms having the smallest

breaking rates. Figure S7c illustrates the relationship between breaking rate constants and molecular configurations, showing that C-C bonds break more readily when the molecule is stretched, while C-H bonds remain consistently flat and low. Figures S7d and S7e show simulations with oxygen present; the trends are similar, but the breaking rates are lower in the oxygen environment.

SUPPLEMENTARY FIGURES

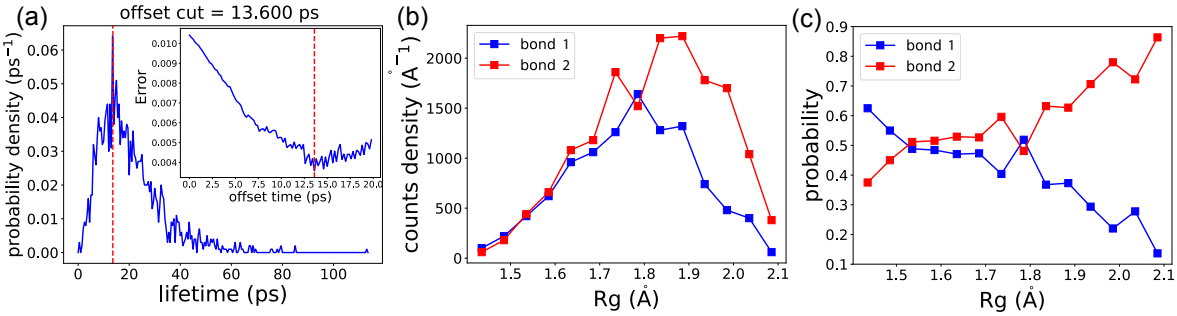

Figure S8: (a) Lifetime distribution for all simulation data for n-pentane, the inset shows the fitting error of the data with exponential distribution as a function of the offset cut, the best offset cut is determined to be 13.600 ps. (b) Counting number density for bond 1 and 2 as a function of R_g, $\rho_i(R_g)$, where i = 1,2. (c) Probabilities among bond 1 and 2 at different R_g, $P_i(R_g)$, and i = 1,2.

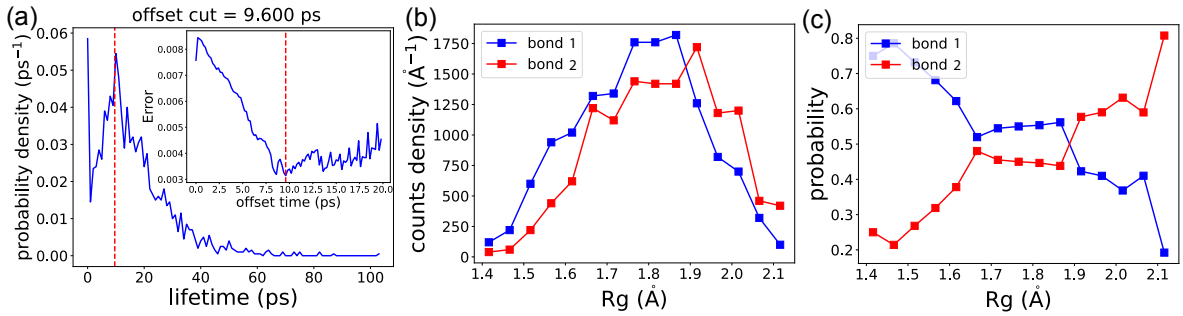

Figure S9: (a) Lifetime distribution for all simulation data for 1,3-Propanediol, the inset shows the fitting error of the data with exponential distribution as a function of the offset cut, the best offset cut is determined to be 9.600 ps. (b) Counting number density for bond 1 and 2 as a function of R_g, $\rho_i(R_g)$, where $i = 1,2$. (c) Probabilities among bond 1 and 2 at different R_g, $P_i(R_g)$, and $i = 1,2$.

SUPPLEMENTARY VIDEOS

movieS1.mp4 shows one typical ReaxFF simulation trajectory of the *n*-tridecane, and the dissociation of the *n*-tridecane occurs approximately at 5.00 ps.

movieS2.mp4 shows one typical ReaxFF simulation trajectory of the *n*-pentane, and the dissociation of the *n*-pentane occurs approximately at 7.00 ps.

movieS3.mp4 shows one typical ReaxFF simulation trajectory of the 1,3-Propanediol, and the dissociation of the 1,3-Propanediol occurs approximately at 9.00 ps.

References

(1) Coifman, R. R.; Lafon, S. Diffusion maps. *Applied and computational harmonic analysis* **2006**, *21*, 5–30.

(2) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the national academy of sciences* **2005**, *102*, 7426–7431.

(3) Coifman, R. R.; Kevrekidis, I. G.; Lafon, S.; Maggioni, M.; Nadler, B. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Modeling & Simulation* **2008**, *7*, 842–864.

(4) Rohrdanz, M. A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *The Journal of chemical physics* **2011**, *134*.

(5) Ferguson, A. L.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Kevrekidis, I. G. Systematic determination of order parameters for chain dynamics using diffusion maps. *Proceedings of the National Academy of Sciences* **2010**, *107*, 13597–13602.

(6) Long, A. W.; Ferguson, A. L. Nonlinear machine learning of patchy colloid self-assembly pathways and mechanisms. *The Journal of Physical Chemistry B* **2014**, *118*, 4228–4244.

(7) Wang, J.; Ferguson, A. L. Nonlinear reconstruction of single-molecule free-energy surfaces from univariate time series. *Physical Review E* **2016**, *93*, 032412.

(8) Wang, J.; Gayatri, M. A.; Ferguson, A. L. Mesoscale simulation and machine learning of asphaltene aggregation phase behavior and molecular assembly landscapes. *The Journal of Physical Chemistry B* **2017**, *121*, 4923–4944.

(9) Wang, J.; Ferguson, A. L. A study of the morphology, dynamics, and folding pathways of ring polymers with supramolecular topological constraints using molecular simulation and nonlinear manifold learning. *Macromolecules* **2018**, *51*, 598–616.

(10) Wang, J.; Gayatri, M.; Ferguson, A. L. Coarse-grained molecular simulation and nonlinear manifold learning of archipelago asphaltene aggregation and folding. *The Journal of Physical Chemistry B* **2018**, *122*, 6627–6647.

(11) Wang, J.; Ferguson, A. L. Recovery of protein folding funnels from single-molecule time series by delay embeddings and manifold learning. *The Journal of Physical Chemistry B* **2018**, *122*, 11931–11952.

(12) Wang, J.; Ferguson, A. Nonlinear machine learning in simulations of soft and biological materials. *Molecular Simulation* **2018**, *44*, 1090–1107.

(13) Boninsegna, L.; Gobbo, G.; Noé, F.; Clementi, C. Investigating molecular kinetics by variationally optimized diffusion maps. *Journal of chemical theory and computation* **2015**, *11*, 5947–5960.

(14) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **1976**, *32*, 922–923.

(15) Xu, W.; Fan, H. A DFT study on recombination of alkyl radicals to C2-C17 normal alkanes & branched C8 alkanes and corresponding CC bond pyrolysis reaction. *Molecular Physics* **2020**, *118*, e1773002.