Electronic Supplementary Material (ESI) for ChemComm.

This journal is © The Royal Society of Chemistry 2021

Supporting Information

A combined DFTB nanoreactor and reaction network generator approach for mechanism of hydrocarbon combustion

Jiawei Bai [abc], Xingchen Liu [ac]*, Tingyu Lei [ac], Botao Teng [d], Xiaodong Wen [abc]*

a. State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan 030001, China.

b. National Energy Center for Coal to Liquids, Synfuels China Co., Ltd., Beijing 101400, China.

c. University of Chinese Academy of Sciences, Beijing 100049, China.

d. College of Chemical Engineering and Materials Science, Tianjin University of Science and Technology, Tianjin 300457, P. R. China.

***Corresponding authors: liuxingchen@sxicc.ac.cn, wxd@sxicc.ac.cn**

Methodology

All the DFTB-NMD simulations were performed using an in-house developed version of deMon-Nano[1] software package. In DFTB-NMD, the movement of molecules was confined within a sphere by a boundary potential *V (r, t)* and the nanoreactor was periodically compressed and expanded due to the coupling with a virtual piston. The radius of this sphere *r* fluctuated between r_1 and r_2 to enhance the collision probability and sampling efficiency.

$$V(r,t) = f(t)U(r,r_1,k_1) + (1 - f(t))U(r,r_2,k_2) \quad (1)$$

$$U(r,r_0,k) = \frac{mk}{2}(r - r_0)^2\theta(r - r_0) \quad (2)$$

$$f(t) = \theta(\left\lfloor\frac{t}{T}\right\rfloor - \frac{t}{T} + \frac{\tau}{T}) \quad (3)$$

Here, k_1 = 1.0 kcal mol^{-1} Å^{-2}, r_1 = 14.0 Å, k_2 = 0.5 kcal mol^{-1} Å^{-2}, r_2 = 8.0 Å, τ = 1.5 ps, *T*= 2.0 ps. ⌊⌋ is the floor function and θ is the Heaviside step function. The function *f(t)* is a rectangular wave that oscillates between one (duration τ) and zero (duration *T*–τ), and *U (r, r_0, k)* is a radial potential that is zero inside the prescribed radius r_0 and harmonic outside. The rectangular waveform switches the restraint potential between *U (r, r_1, k_1)* and *U (r, r_2, k_2)*, which forces the atoms with a radial position 8.0 < *r* <14.0 Å towards the center and causes them to collide. The molecules will fill up the volume again when the sphere expands again because of the high temperature.

We simulated the ethylene combustion system in the presence of different number of O_2. The amount of ethylene was constant (40 C_2H_4) and the amount of oxygen was varied from 0 to 100. The simulation time of pure ethylene system was 500 ps and the rest were 10 ps. The time scales for simulation process are long enough to observe the formation of the main products. Self-consistent DFTB (SCC-DFTB) matsci[2] set was used to describe atomic forces of reactants, products and reaction intermediates. The MD simulations were run with a time-step of 0.5 fs. The temperature was controlled with the Berendsen thermostat, with the target temperature of 1000 K.[3] All the DFTB-NMD simulations were repeated 5 times with different initial configurations to ensure the consistency of the results.

The initial configurations were created using the Packmol[4] code. Ethylene and oxygen molecules were randomly generated in a sphere with a radius of 14 Å, and the distance between each molecule was not less than 2 Å. DFT calculations were performed by Gaussian16[5] package at the M062x/6-311g (d, p) level[6]. The convergence threshold for the maximum force and maximum displacement was 0.000450 kcal/ (mol Å), and 0.0018 Å, respectively.

Table S1 The geometry parameter and bond dissociation energy of ethylene and acetylene obtained from three functionals with 6-311G(d,p) basis set.

	Ethylene			Acetylene		Bond Dissociation Energy (Kcal/mol)		
	C-H (Å)	C=C (Å)	∠H-C=C (°)	C-H (Å)	C≡C(Å)	CH_2CH-H	HCC-H	HC≡CH
B3LYP	1.085	1.327	121.762	1.063	1.198	108.392	134.048	226.382
M062X	1.084	1.324	121.604	1.064	1.196	108.906	131.514	231.538
WB97XD	1.085	1.324	121.644	1.064	1.196	108.998	134.385	228.629
Exp[7]	1.087	1.339	121.300	1.060	1.203	111.207	132.908	230.636

In order to assess the quality of the functionals, the structure parameters and bond dissociation energy were calculated using B3LYP/M062X/WB97XD levels with 6-311G (d,p) basis set (Table S1). For comparison, we also

provide results obtained from previous experiments[7]. The result obtained from three kinds of functionals are consistent with experimental values. And the best results are obtained with the M062X level.

Table S2 The geometry parameter and bond dissociation energy of ethylene and acetylene obtained with three basis sets at M062X level of theory.

	Ethylene			Acetylene		Bond Dissociation Energy (Kcal/mol)		
	C-H (Å)	C=C (Å)	∠H-C=C (°)	C-H (Å)	C≡C(Å)	CH_2CH-H	HCC-H	HC≡CH
6-311G(d,p)	1.084	1.324	121.604	1.064	1.196	108.906	131.514	231.538
Def2TZVP	1.083	1.322	121.606	1.064	1.194	108.867	131.800	232.063
Def2QZVP	1.081	1.321	121.591	1.062	1.193	108.748	132.357	231.694
Exp[7]	1.087	1.339	121.300	1.060	1.203	111.207	132.908	230.636

Three kinds of basis sets, including 6-311G(d,p), Def2TZVP and Def2QZVP, were tested at M062X level (Table S2). And the calculational data was comparable with previous experimental values. The results illustrate that the accuracy of 6-311G(d,p) is adequate and the basis sets with better performance aren't necessary.

In conclusion, the M062X/6-311G(d,p) level of theory was selected due to its comparably low computational cost while maintaining in many cases a satisfactory accuracy. Previous literature studies have also confirmed that M062X gives accurate description of the structure and energy of hydrocarbons, and is suitable for the calculations in this study.[8, 9]

The ReactNetGenerator method[10] uses a two-state HMM model[11-13] to screen the noises from large-amplitude molecular vibrations and collisions in the simulations. In this way, much of the unstable structures produced by the periodical compression stages of the DFTB-NMD simulations to cross the barriers between stable species can be screened out. [12] The maximum number of species in the reaction matrix (*N*) was set large enough to ensure such that all the relevant species in the reaction processes were included for constructing the reaction networks. The detected species are then indexed by canonical simplified molecular input line entry specification (SMILES)[14], which from a reaction matrix was generated. The reaction network was constructed using a Fruchterman–Reingold force-directed algorithm.[15] The HMM algorithm was turned off when analyzing the local reaction networks for given species.

ReacNetGenerator is a tool that can analyze the trajectories of reactive MD simulations and automatically extract reaction network from it without predefined reaction coordinates and elementary reaction steps.[10] The essential input of ReacNetGenerator is atomic coordinates in the DFTB-NMD simulation trajectories, which contain numerous reaction events and molecular species. ReacNetGenerator can extract plentiful information from the DFTB-NMD trajectories, including molecular species and reaction events. The result of analysis is an interactive web page, which consists of the network, species and reactions. By default, it shows the reaction network formed by top 20 species which have the most reactions. The connectivity of the atoms is determined automatically from the atom coordinates in the input file. Then, molecules and radicals are detected according to the atomic connectivity.

However, many "noise" molecules that are unstable or even unreasonable in energy or the structure will appear because reactive MD simulation normally contains large amplitude molecular vibrations and collisions. Therefore, to generate human readable skeletal reaction networks (such as Fig. 4, and Figure S3 and S4 of the SI), by default, ReacNetGenerator uses a two-state hidden Markov model (HMM) [11-13] to filter these "noise". The filtering ability of the HMM model is determined by the HMM parameters (the off-diagonal elements of Matrix **A** and **B**). In this work, we used the default parameters where $\mathbf{A}=\begin{pmatrix}0.999 & 0.001\\ 0.001 & 0.999\end{pmatrix}$ and $\mathbf{B}=\begin{pmatrix}0.6 & 0.4\\ 0.4 & 0.6\end{pmatrix}$.[10] For this

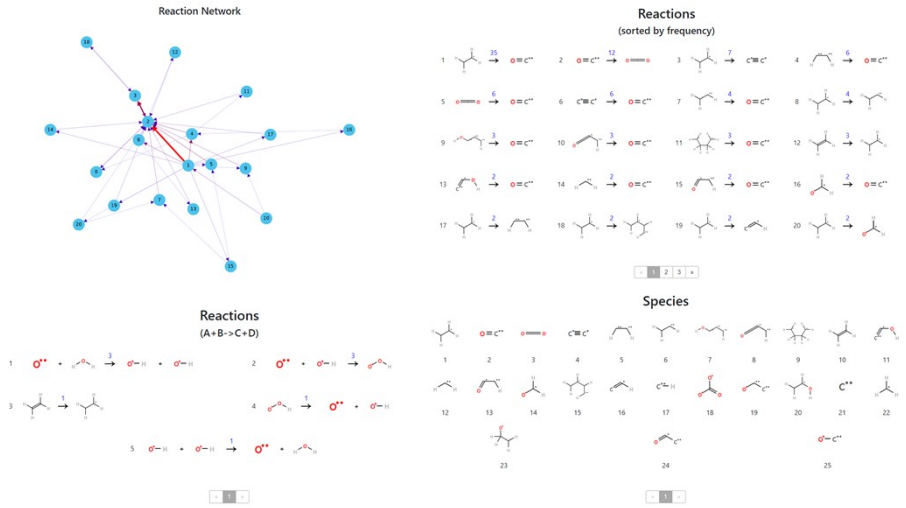

Figure S1 The analysis results of MD trajectories (100 O_2 and 40 C_2H_4) by ReacNetGenerator with HMM-based filter.

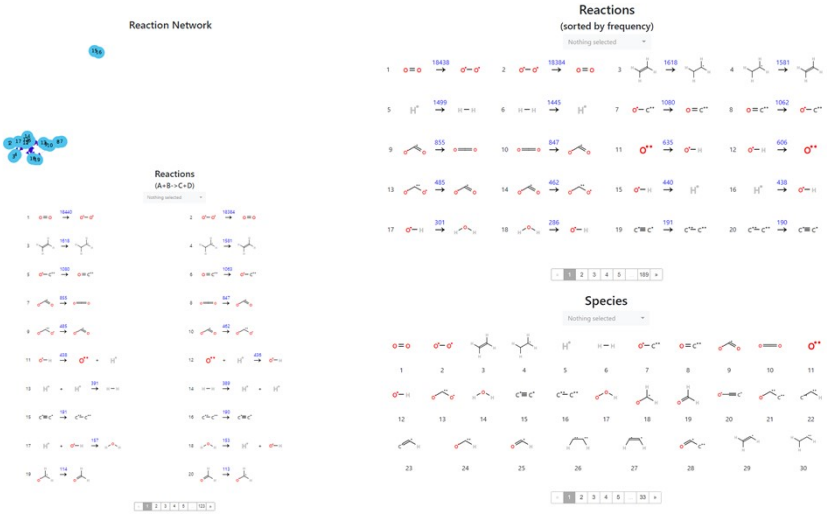

Figure S2 The analysis results of MD trajectories (100 O_2 and 40 C_2H_4) by ReacNetGenerator without HMM-based filter.

reason, the number of species and reactions in the reaction network with HMM-based filter is usually far less than the number without filtration. Hence the HMM-filtered skeletal reaction network only represents the conversion relationship between each species. To demonstrate the unstable species that has been eliminated, the analysis results of MD trajectories (100 O_2 and 40 C_2H_4) by ReacNetGenerator with and without HMM-based filter are obtained and compared in Figure S1 and Figure S2 (The number of species included in the reaction network was set to 20). The total number of species in Figure S1 is only 25. However, when the HMM is turned off, over 900 species appear as shown in Figure S2. The ones that are filtered by the HMM includes species such as $H_2C\dot{C}H{=}O$, :CHCH=O, and HC≡CH.

The limitation of the above HMM-filtered skeletal reaction network is that some crucial intermediates may be filtered out. To identify the elementary reaction steps that make up the entire reaction pathway starting from the reactant all the way to the product, the local reaction network of specific species on the reaction pathway can be identified in ReacNetGenerator by mouse-clicking.

In our current study, we found the skeletal reaction network of stable carbonous species by trajectory analysis with HMM-based filter, and three complete reaction pathways of the conversion from C_2H_4 to CO_2 in

ethylene combustion by continuous local reaction network without HMM-based filter.

Table S3 Chemical composition of system species at the corresponding time from a typical DFTB-NMD simulation of ethylene combustion.

Time(ps)	Chemical Composition
0	100 O_2 + 40 C_2H_4
1	100 O_2 + 38 C_2H_4 + 1 C_4H_8
2.5	76 O_2 + 31 H_2 +17 CO +12 H +7 OH +7 O + 6 C_2H_4 + 4 C_2H_2 + 3 H_2O + 3 CH_2O + 3 C_2OH + 3 C_2H_3 + 2 O_2H + 1 CO_2H + 1 CH_3 + 1 CH_2 + 1 C_6H_3 + 1 C_5H_3 + 1 C_3H_3 + 1 C_3H + 1 C_2H_3OH + 1 C_2O + 1 C_2H + 1 C_2
5	58 H_2 + 50 O_2 + 46 CO + 22 OH + 12 H + 10 O + 6 CO_2 + 6 C_2O + 2 H_2O + 2 C_3O + 1 CH_2 + 1 C_5OH + 1 C_3OH + 1 C_3H_2 + 1 C_2
7.5	62 CO + 56 H_2 + 30 O_2 + 29 OH + 15 CO_2 + 10 O + 7 H +5 H_2O + 2 O_2H + 1 C_3
10	65 CO + 47 H_2 + 30 H + 27 O_2 + 27 O + 21 OH + 12 CO_2 + 7 H_2O + 1 COH + 1 C_2O

Table S4 The reactions that occurred during the oxidation of C_2H_4 to CO_2 from a typical DFTB-NMD simulation.

R1: $CH_2CH_2 \rightarrow H_2C{=}C^{\cdot}H + H\cdot$	R29: $:C{=}C{=}O \rightarrow :C: + CO$
R1: $CH_2CH_2 \rightarrow H_2C{=}C^{\cdot}H + H\cdot$	R30: $:C{=}C{=}O + \cdot OH \rightarrow CO + :^{\cdot}COH$
R3: $CH_2CH_2 + H\cdot \rightarrow H_2C{=}C^{\cdot}H + H_2$	R31: $:C{=}C{=}O + O_2 \rightarrow CO + :^{\cdot}COO\cdot$
R4: $CH_2CH_2 + \cdot OH \rightarrow H_2C{=}C^{\cdot}H + H_2O$	R32: $:C{=}C{=}O + H\cdot \rightarrow CO + :^{\cdot}CH$
R5: $CH_2CH_2 + O_2 \rightarrow H_2C{=}C^{\cdot}H + HOO\cdot$	R33: $H_2C{=}C^{\cdot}H + \cdot OH \rightarrow H_2C{=}CHOH$
R6: $H_2C{=}C^{\cdot}H \rightarrow H_2C{=}C: + H\cdot$	R34: $H_2C{=}CHOH + O: \rightarrow H_2C^{\cdot}CH{=}O\cdot + \cdot OH$
R7: $H_2C{=}C^{\cdot}H + O: \rightarrow H_2C{=}C: + \cdot OH$	R35: $H_2C^{\cdot}CH{=}O\cdot + O: \rightarrow H_2C{=}C{=}O + \cdot OH$
R8: $H_2C{=}C^{\cdot}H + H\cdot \rightarrow H_2C{=}C: + H_2$	R36: $H_2C^{\cdot}CH{=}O\cdot \rightarrow H_2C{=}C{=}O + H\cdot$
R9: $H_2C{=}C^{\cdot}H + \cdot OH \rightarrow H_2C{=}C: + H_2O$	R37: $H_2C^{\cdot}CH{=}O\cdot \rightarrow :CHCH{=}O + H$
R10: $H_2C{=}C^{\cdot}H + O_2 \rightarrow H_2C{=}C: + HOO\cdot$	R38: $H_2C{=}C{=}O + O: \rightarrow HC^{\cdot}{=}C{=}O + \cdot OH$
R11: $H_2C{=}C^{\cdot}H \rightarrow HC{\equiv}CH + H\cdot$	R39: $H_2C{=}C{=}O \rightarrow HC^{\cdot}{=}C{=}O + H\cdot$
R12: $H_2C{=}C^{\cdot}H + O: \rightarrow HC{\equiv}CH + \cdot OH$	R40: $:CHCH{=}O \rightarrow HC^{\cdot}{=}C{=}O + H\cdot$
R13: $H_2C{=}C^{\cdot}H + H\cdot \rightarrow HC{\equiv}CH + H_2$	R41: $HC^{\cdot}{=}C{=}O + \cdot OH \rightarrow :C{=}C{=}O + H_2O$
R14: $H_2C{=}C^{\cdot}H + \cdot OH \rightarrow HC{\equiv}CH + H_2O$	R42: $HC^{\cdot}{=}C{=}O + O_2 \rightarrow :C{=}C{=}O + HOO\cdot$
R15: $H_2C{=}C^{\cdot}H + O_2 \rightarrow HC{\equiv}CH + HOO\cdot$	R43: $HC^{\cdot}{=}C{=}O + H\cdot \rightarrow :C{=}C{=}O + H_2$
R16: $H_2C{=}C: \rightarrow HC{\equiv}C\cdot + H\cdot$	R44: $CH_2CH_2 + O_2 \rightarrow \cdot CH_2CH_2OO\cdot$
R17: $H_2C{=}C: + H\cdot \rightarrow HC{\equiv}C\cdot + H_2$	R45: $\cdot CH_2CH_2OO\cdot \rightarrow \cdot CH_2CH_2O\cdot + O:$
R18: $HC{\equiv}CH \rightarrow HC{\equiv}C\cdot + H\cdot$	R46: $\cdot CH_2CH_2O\cdot \rightarrow CH_2O + :CH_2$
R19: $HC{\equiv}CH + H\cdot \rightarrow HC{\equiv}C\cdot + H_2$	R47: $CH_2O + O_2 \rightarrow \cdot CHO + HOO\cdot$
R20: $HC{\equiv}CH + O: \rightarrow HC{\equiv}C\cdot + \cdot OH$	R48: $CH_2O + H\cdot \rightarrow \cdot CHO + H_2$
R21: $HC{\equiv}C\cdot \rightarrow \cdot C{\equiv}C\cdot + H\cdot$	R49: $CH_2O \rightarrow \cdot CHO + H\cdot$
R22: $HC{\equiv}C\cdot + O: \rightarrow \cdot C{\equiv}C\cdot + \cdot OH$	R50: $\cdot CHO + H\cdot \rightarrow CO + H_2$
R23: $HC{\equiv}C\cdot + H\cdot \rightarrow \cdot C{\equiv}C\cdot + H_2$	R51: $\cdot CHO + O: \rightarrow CO + \cdot OH$
R24: $HC{\equiv}C\cdot + \cdot OH \rightarrow \cdot C{\equiv}C\cdot + H_2O$	R52: $\cdot CHO + O_2 \rightarrow CO + HOO\cdot$
R25: $\cdot C{\equiv}C\cdot + \cdot OH \rightarrow \cdot C{\equiv}COH$	R53: $\cdot CHO \rightarrow CO + H\cdot$
R26: $\cdot C{\equiv}COH \rightarrow :C{=}C{=}O + H\cdot$	R54: $CO + O: \rightarrow CO_2$
R27: $\cdot C{\equiv}COH + H\cdot \rightarrow :C{=}C{=}O + H_2$	R55: $CO + \cdot OH \rightarrow CO_2 + H\cdot$
R28: $\cdot C{\equiv}C\cdot + O: \rightarrow :C{=}C{=}O$	

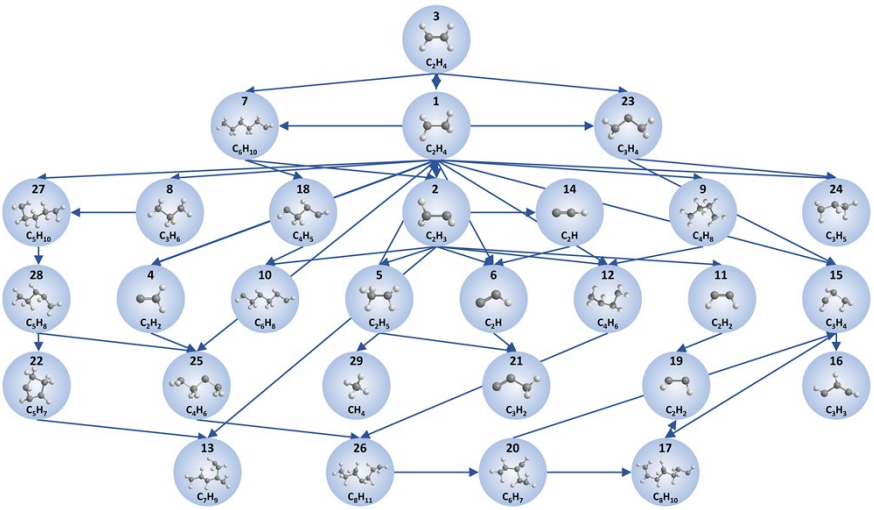

Figure S3 The reaction network formed by the species which contain carbon in the first 10 ps trajectory of ethylene polymerization simulation. Species are shown in blue spheres, and the chemical reactions are indicated using blue arrows. The thicker the line indicates more reactions between species. Hydrogen, carbon atoms are shown in white, gray, respectively.

In the simulation of ethylene polymerization, the important reactions leading to the formation of PAH occurred within 10 ps of the simulations, and the file size that can be read by ReacNetGenerator is limited. Therefore, we analyzed the trajectory from the first 10ps of the simulation as shown in Figure S3. This reaction network is composed of carbon-containing species. The species that are unstable or even unreasonable were filtered out by the HMM algorithm in ReacNetGenerator, hence the reaction network only represents the conversion relationship between each species (the blue arrows). The thickness of the lines represents the frequency of the conversion between species. The most frequently occurring conversions are the activation of ethylene (3→1) and the formation of vinyl through the dehydrogenation of activated ethylene (1→2). One can find a number of reaction pathways from ethylene to short carbon chains, such as C_2H_4→C_4H_8→C_4H_6→C_8H_{11}. With the occurrence of the above C-C coupling reactions, the length of the carbon chains was continuously increased. Eventually, all C atoms formed a large PAH and the H atoms existed in the form of H_2 and dissociative H.

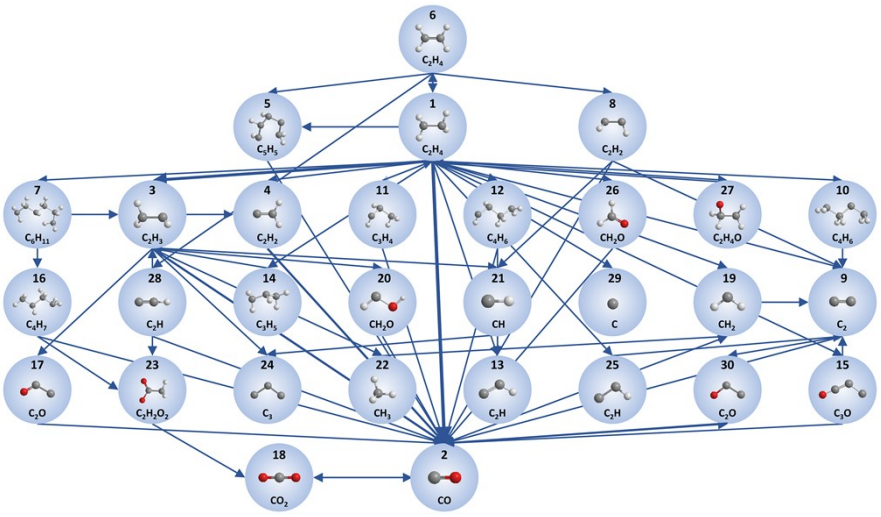

Figure S4 The reaction network formed by the species which contain carbon in trajectory of oxygen-lean combustion of ethylene simulation. Species are shown in blue spheres. Hydrogen, oxygen and carbon atoms are shown in white, red and gray, respectively.

Figure S4 depicts the reaction network extracted from oxygen-lean ethylene combustion simulation. The formation of CO (2) from various carbonaceous species such as activated ethylene (1), vinyl (3), and vinylidene (4), was the most frequent conversions. Another frequent conversion is the activation of ethylene (6→1). In the beginning, ethylene had been constantly vibrating, twisting and deforming due to high temperature. Subsequently, activated ethylene molecules (1) generated with the broken of double bonds, which is an important intermediate. The main reactions in the simulation process were dehydrogenation and C-C coupling reactions. The occurrence of reactions was random so that short carbon chains with different numbers of C and H were formed (7,10,11…). It can be seen from the reaction network that the reaction process is mainly featured by the polymerization and rearrangement reaction of several carbonaceous radicals to form short carbon chains. Free oxygen radicals in the system were attached to the end of the unsaturated carbon chain to prevent the further growth of the carbon chains and generate C_nH_mCO. Then CO was generated by the collision-induced cleavage of C-C bond (15→2). Subsequently, CO was converted to CO_2 by reacting with O: or ·OH radicals.

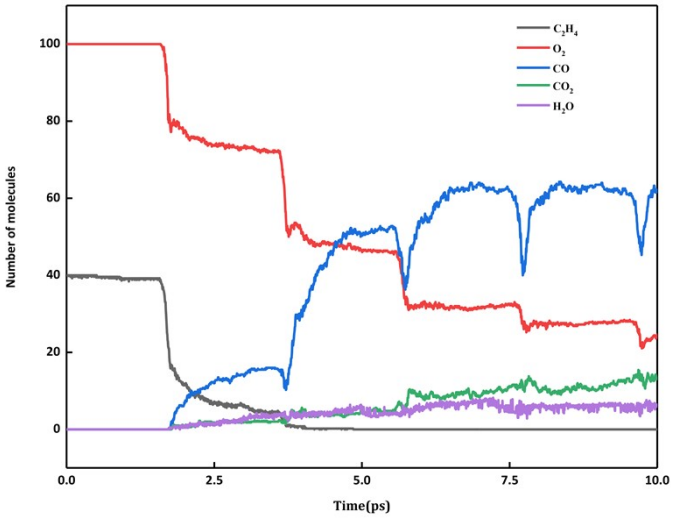

Figure S5 The number of important reactants, products, and intermediates evolved with simulation time. These curves are smoothed to make them look better and clearer. All the data are averaged from 5 parallel trajectories.

Figure S5 shows the time-dependent changes in the number of reactants and main products. At the beginning of the simulation, the number of ethylene remained basically invariant. At 1.7 ps, the reaction with radicals and O_2 was the major cause of a large amount of consumption of ethylene during the first compression of the system, and the number of ethylene dropped steeply to 17 and then slowly decreased. At about 3.86 ps, there was no more ethylene in the system. During periodic compression, as the density of molecules increased, and the possibility of a reaction became higher. Oxygen was consumed rapidly. The decrease curve of oxygen number was stepped. At 17.65 ps, the first CO was generated. The number of CO ends up being 65. The number of CO was fluctuant in every compression period. CO_2 and H_2O were slowly generated with time. Eventually, the numbers of CO_2 and H_2O were 12 and 7, respectively.

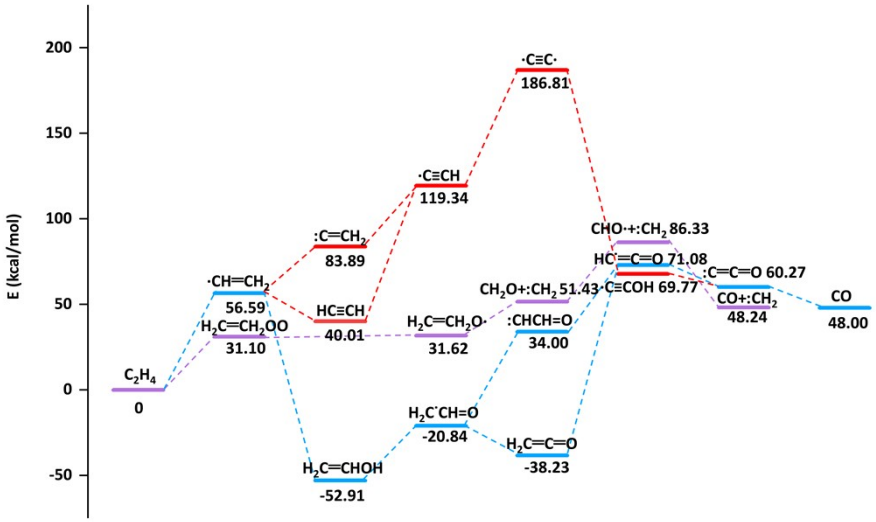

Figure S6 Potential energy diagram of three reaction pathways of ethylene oxidation obtained from DFT.

References

1. R. M. Heine T, Patchkovskii S, Frenzel J, Koester AM, Calaminici P, Escalante S, Duarte HA, Flores R, Geudtner G, Goursot A, 2009.
2. O. A. F. Frenzel J, Jardillier N, Heine T, Seifert G, *TU-Dresden*, 2004-2009.
3. P. J. Berendsen HJC, Vangunsteren WF, Dinola A, Haak JR, *J. Chem. Phys.*, 1984, **81**, 3684 − 3690.
4. L. Martinez, R. Andrade, E. G. Birgin and J. M. Martinez, *J. Comput. Chem.*, 2009, **30**, 2157-2164.
5. F. M. e. al., *(Gaussian Inc, Wallingford CT,2016)*, 2019.
6. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215-241.
7. D. R. Lide, *CRC Handbook of Chemistry and Physics*, 2005.
8. M. H. Keshavarz, M. Zamani, F. Atabaki and K. Hosseini Monjezi, *Comput. Theor. Chem.*, 2013, **1011**, 30-36.
9. Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157-167.
10. J. Zeng, L. Cao, C.-H. Chin, H. Ren, J. Z. H. Zhang and T. Zhu, *Phys. Chem. Chem. Phys.*, 2020, **22**, 683-691.
11. L. R. Rabiner, *Proc. IEEE*, 1989, **77**, 257-286.
12. L. P. Wang, R. T. McGibbon, V. S. Pande and T. J. Martinez, *J. Chem. Theory Comput.*, 2016, **12**, 638-649.
13. S. Russell and P. Norvig, 2002.
14. D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31-36.
15. T. M. Fruchterman, E. M. J. S. P. Reingold and experience, *Software: Practice and Experience*, 1991, **21**, 1129-1164.