Supporting Information for

Unveiling Fullerene Formation and Interconversion through Molecular Dynamics Simulations with Deep Neural Network Potentials

Yanbo Han,[a)] Mengyang Li,[b)] * Masahiro Ehara,[c)] and Xiang Zhao[a)], *

a) School of Chemistry, State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: xzhao@mail.xjtu.edu.cn

b) School of Physics, Xidian University, Xi'an 710071, China. E-mail: limengyang@xidian.edu.cn

c) Research Center for Computational Science, Institute for Molecular Science, Okazaki 444-8585, Japan

Note: large files (detailed. xyz coordinate files, interactive plotly graphs and etc.) are provided through figshare (10.6084/m9.figshare.28431026) as named **extra SI** in following.

1 MODEL PERFORMANCES AND FINETUNE

1.1 Zero-shot performances of DPA1 on C-C bond scan

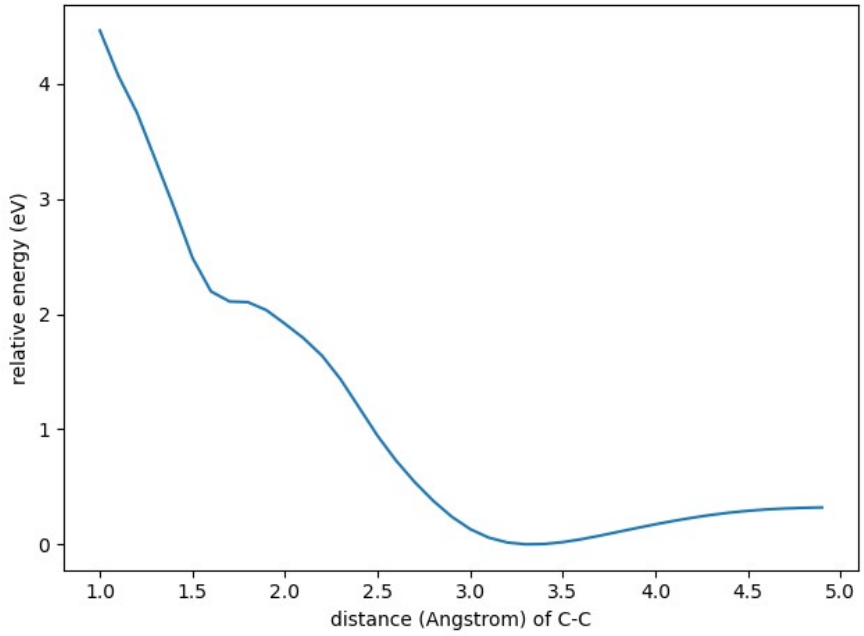


Figure 1: Energy scan of the molecule along the C-C bond length, using the pre-trained DPA1 model[1] from Aissquare.

1.2 Finetune details of spookynet

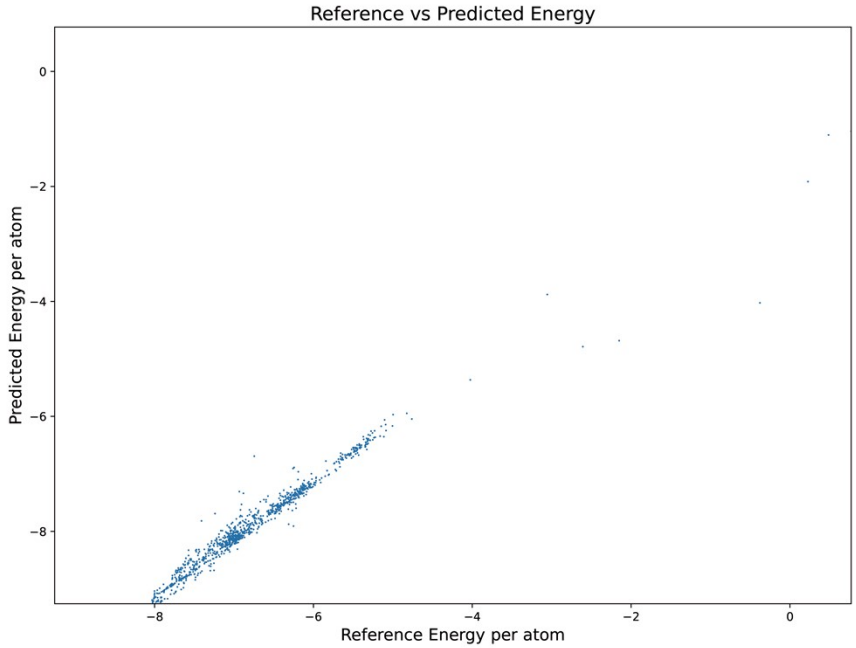
We finetune the pre-trained SpookyNet model[2] from SpookyNet repository and the pre-trained DPA1 model[1] from Aissquare. The finetune details are shown in Table S**1**. The finetune of DPA1 failed in our affordable resources, so only the finetune of SpookyNet is shown in Table S**1**. The finetune of DPA1 is similar to that of SpookyNet, except that the learning rate scheduler is Cyclical Learning Rate (CLR)[3] and with a warmup of 10 epochs.

To be honest, the SpookyNet model is not a common finetune-able model with encoder-decoder structure. It is a model with a complex encoder structure and a simple decoder structure. The decoder, or fitting network, is a simple fully connected neural network with only 128 nodes, which is too much simple and hard to finetune, so we finally continue the train of the whole model, including the encoder and the decoder.

Table S1: Finetune details of SpookyNet.

Hyperparameter	**SpookyNet**
Optimizer	RAdam
Learning rate max	0.0001
Learning rate min	1e-8
CycleLR **exp_range** γ	0.9999
CycleLR **step_size_up**	50
CycleLR **step_size_down**	50
Epochs for finetune	5000
Epochs for warmup	10
Batch size	16*

*Using **accumulate_grad_batches=4** with help of pytorch-lightning, so the real batch size for backpropagation is 64.



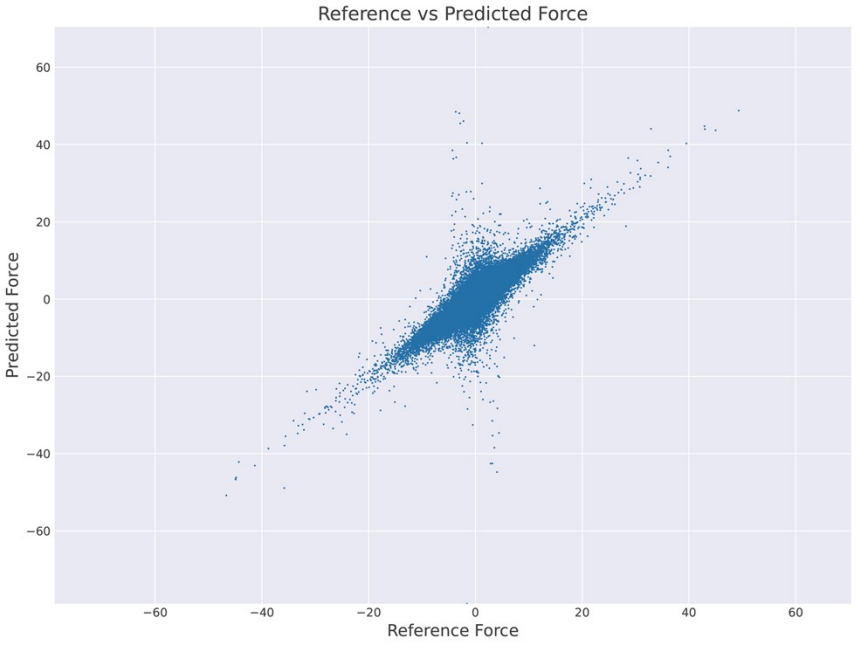


Figure S2: Finetune results on SpookyNet.
(All value units are in atomic unit, Energy: eV, Force: eV/Å)

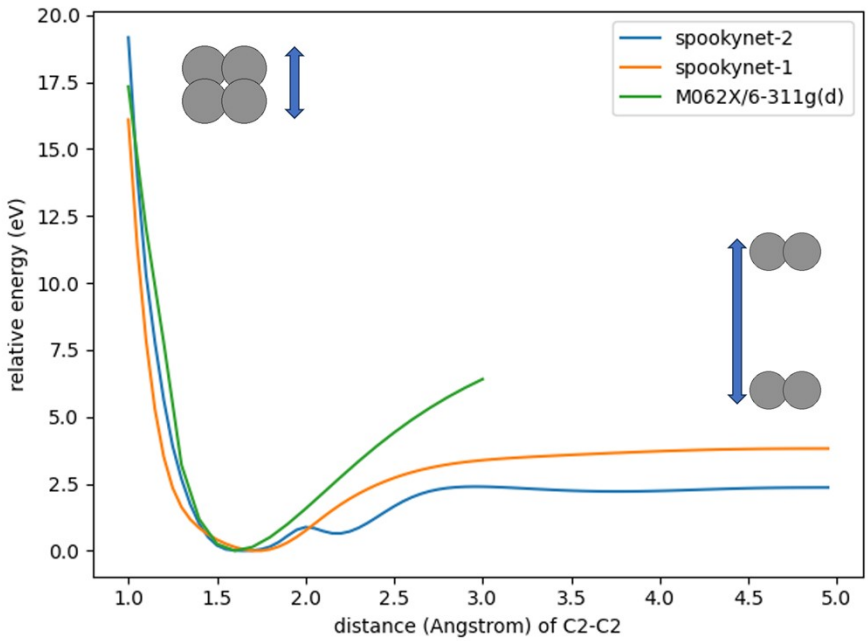


Figure S3: Energy scan of the - cluster with different early-stop models, and compared with the DFT reference at M06-2x/6-311G(d) level. The models trained on the GAP20 datasets[4], which is calculated at optB88-vdW level with VASP, the energy of the DFT reference is just to show it is expected to be smooth.
Calculated with Gaussian 16 package[5].

2 THE MOLECULAR DYNAMICS SIMULATION RESULTS OF FULLERENE GENERATION

All of the molecular dynamics are produced with LAMMPS packgage[6], but with different potential files, and deal with LibOVITO python interface[7] and Plotly package[8]. Key rings (pentagons, hexagons, and heptagons) in the structures are found with the help of Parallel Minimum Cycle Basis Library(https://github.com/d-michail/parmcb), which has been integrated within the Fullerenetool package(https://github.com/saltball/fullerenetool) as Cython interface.

Table S2: Simulation details of studied cases.

Directory id*	Cases	Simulation details	Notes
1	**Carbon cluster temperature**	**10976 carbon atoms in box of 150*150*150 Å^3** - Relax 50ps(6000K) - 100ps (call 'short') and 1ns (call 'long') (6000K to 2500K and 2500K to 300K) (totally 200ps and 2ns, respectively) **1372 carbon atoms simulations in box of 70*70*70 Å^3 and 150*150*150 Å^3** - Relax 50ps (6000K) - 2.5ns (6000K to 2500K) or 1ns(150*150*150 Å^3) - 1ns (2500K to 300K)	(Tdamp=0.1 ps, timestep=0.001 ps) 3 simulations with different random seeds for 'short' and 1 for 'long' respectively (Tdamp=0.1 ps, timestep=0.001 ps) **1372 carbon atoms simulations** are established with spookynet-1 potential(file `spookynet-finetune/spookynetgap20_190.pt`)
2	**Carbon cluster temperature (lower density)**	**10976 carbon atoms in box of 300*300*300 Å^3** - Relax 50ps (6000K) - 2ns (6000K to 2500K) - 2ns (2500K to 300K)	(Tdamp=0.1 ps, timestep=0.001 ps) 3 simulations with different random seeds
3	**Cluster nanoreactor-like dynamics**	**Initial 37 structures are from lower density simulation snapshots at 2500K, use clusters (C-C bond cutoff=2.1Å) with atoms not less than 60. Using `fix wall/piston zlo pos 1.0 vel 10.0 units box` in lammps** - Relax 10ps (from 1000K) - 50ps/100ps/200ps (2500K) - 50ps(to 300K)	(timestep=0.001 ps) Using `fix wall/piston zlo pos 1.0 vel 10.0 units box` in lammps 3*37 simulations with different random seeds for three simulation durations
4	**Fullerene growth in carbon vapor**	**2572 carbon atoms in box of 150*150*150 Å^3, including 20 fullerene cages** - Relax 50ps (from 1000K) - 100ps (4000K/3000K/2000K) - 50ps(to 300K) And - Relax 50ps (from 1000K) - 100ps/200ps (2500K) - 100ps(to 300K)	(Tdamp=0.1 ps, timestep=0.001 ps) 2*3 simulations at 2000/3000/4000K 3 simulations at 2500K
5	**Fullerene growth with iron cluster**	*See caption and following description of Fig. S10*	

* Directory id: The name of directory in Extra SI to store trajectories of LAMMPS result .xyz files related to the cases.

2.1 Carbon cluster temperature

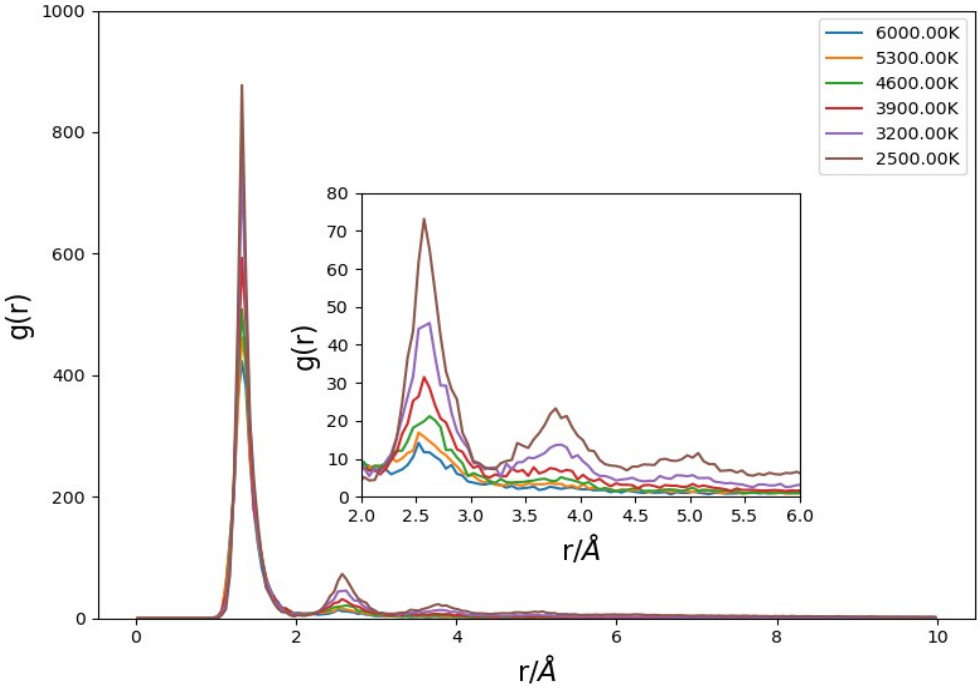


Figure S4: Radial distribution function (RDF) of carbon atoms in a fullerene simulation with 10,976 atoms in a 150 Å cubic box at various temperatures, calculated using LibOVITO. Left axis shows RDF values.
Inset: Magnified view of the RDF curves' second peaks and the third ones.

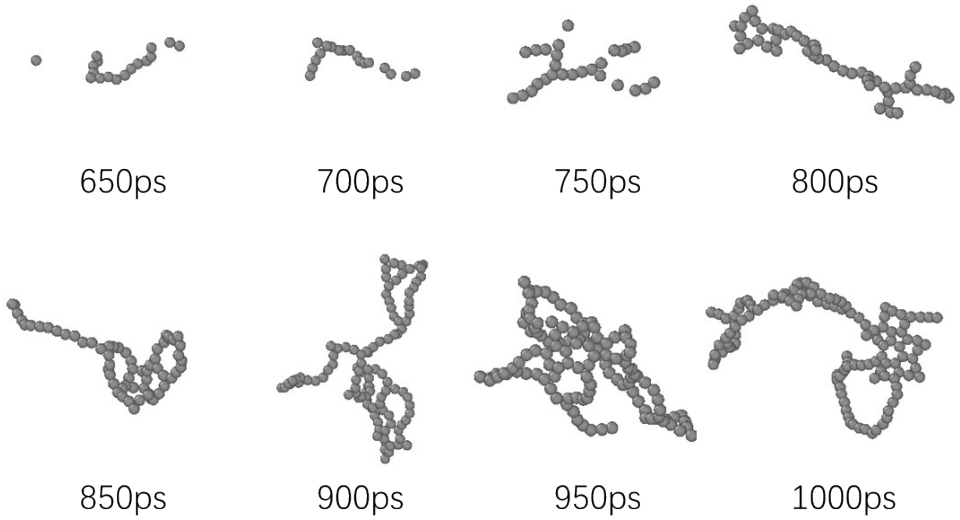


Figure S5. Visualization of the key steps in the formation of clusters from the file `0_5.xyz`. This includes only the atoms that are ultimately part of the final cluster, so there appears to be a loss of atoms in the earlier images. This appearance is due to the fact that those atoms do not remain in the final cluster.

Figure S5 illustrates a typical case of cluster formation obtained from low-density simulations. It includes the form of carbon chains present at higher temperatures, which gradually transform into

rings and interconnected rings as the temperature decreases. It is also noticeable that, in the early stages, atoms within the chains are still exchanging or, in other words, are still being significantly lost. This loss almost ceases once a significant number of rings have formed. More cases can be found in the xyz files available in the attached `cluster_atoms` directory attached with SI.

2.2 Cluster nanoreactor-like dynamics

Note: See `nano_reactors/nanoreactor_{1,2,3}/plot_{0-37}.html` in **extra SI** for interactive graph with plotly. (One may need to enable network and javascript of html file browser to get plotly cdn files.) Coordinates for each simulation are provided as `coord_{0-37}.xyz` in those directories.

In interactive graph and following, pentagons are in red, hexagons are in light blue, and heptagons are in light yellow. And carbon atoms are shown with tiny points, which with two neighbors are in red, with three neighbors are in green, and other atoms are in black.

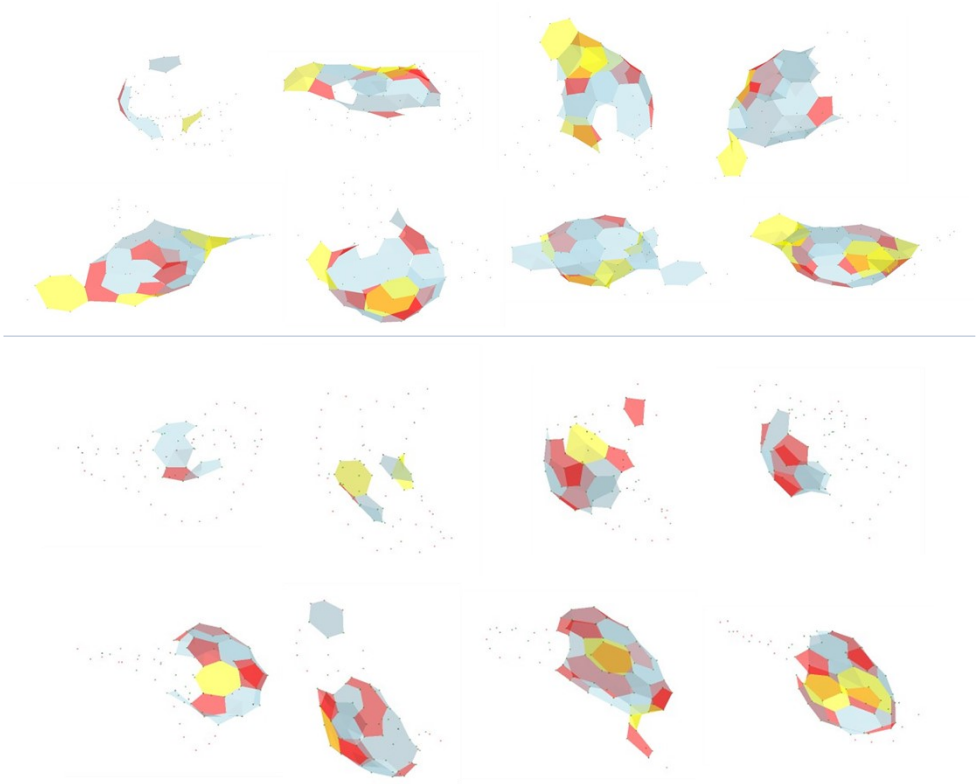


Figure S6: Two example for carbon cluster simulation with piston. In similar simulation time, some cluster (the upper, as example) fast generate into a fullerene, while some cluster (the bottom, as example) still keep the cluster structure, but the cluster structure is not stable, and change to a fullerene cage in further longer simulations. Pentagons are in red, hexagons are in light blue, and heptagons are in light yellow. Due to the angle and rotation, some of the rings appear to be deformed. (See extra SI for interactive graph.)

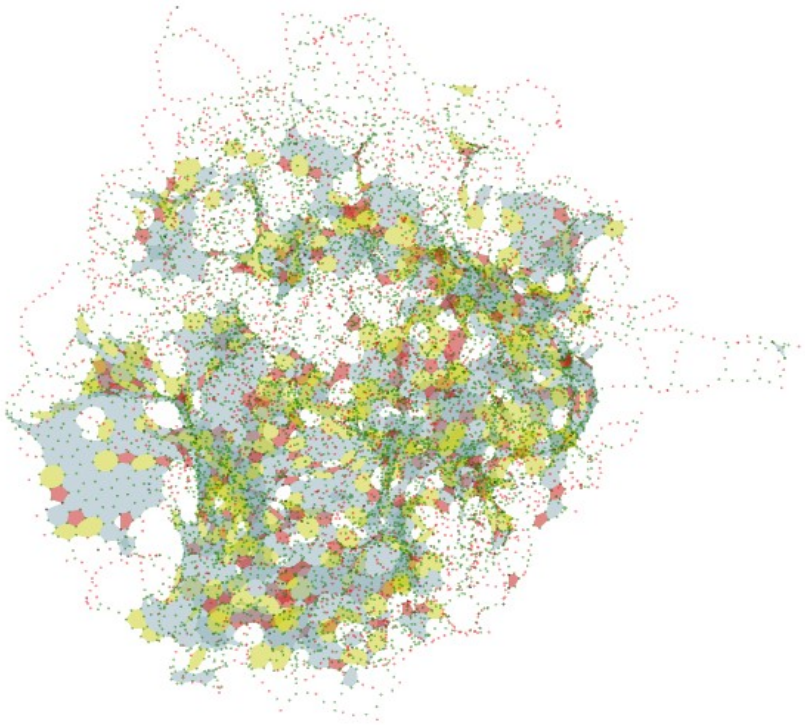


Figure S7: The snapshot of the carbon amorphous structure with 10976 carbon atoms in 150 angstrom cubic box after cooling from 6000 K to 2500 K. There is no obvious free carbon atoms(six atoms as three C_2 molecules in this snapshot) out of the amorphous structure. Some graphene-like structure(light blue) can be found in the amorphous structure.

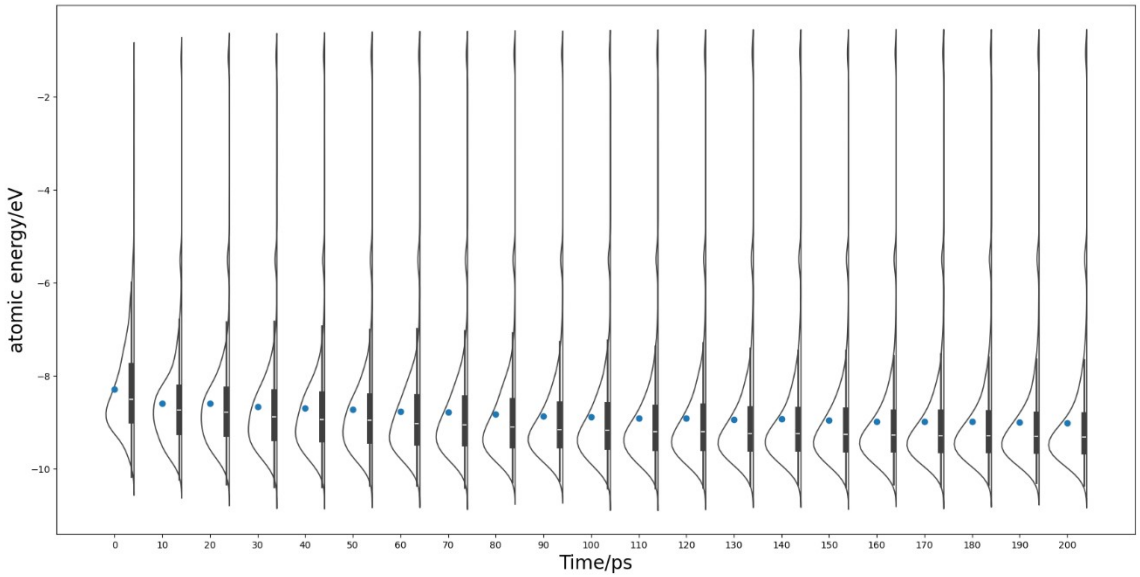


Figure S8: In the 200 ps piston nanoreactor-like simulation where the single-atom energies of all carbon atoms were calculated using Spookynet, a violin plot was generated to show the distribution of these energies at different time points. The plot consists of kernel density estimate (KDE) curves and box plots. The white short horizontal line in the middle of each box indicates the position of the median, while additional blue dots represent the positions of the data means.

Using a deep neural network potential that decomposes the total system energy based on atomic energies, it is possible to calculate the energy of each atom in its local environment within the system. This can serve as a simple descriptor. Figure S8 shows the statistics of carbon atom energies across all clusters during the 200 ps simulation. Overall, the trend indicates a decrease in energy, and the distribution becomes less skewed. This can be interpreted as the system evolving towards a more stable state. Additionally, the products in these simulations (referred to as nanoreactor_3 in the extra SI) are predominantly nearly closed cage-like structures, characteristic of fullerenes.

2.3 Fullerene growth in carbon vapor

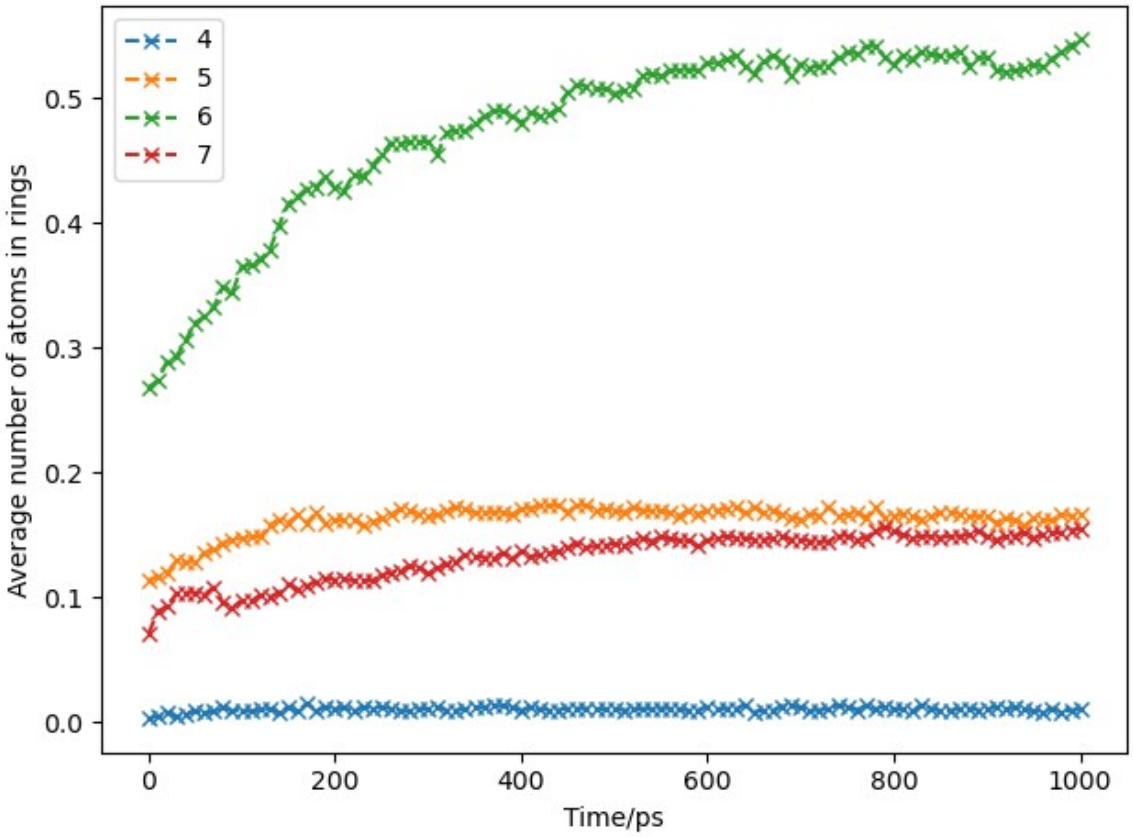


Figure S9. The average number of atoms in Rings in simulations of C_{60} fullerene exposed to 2500 K C_2 vapor. 4, 5, 6, and 7 denote the average number of atoms in four-membered rings, five-membered rings, six-membered rings, and seven-membered rings, respectively. The definition of the average number of atoms can be found in Equation S1.

As shown in Figure S9, the average number of atoms per ring is calculated using the following expression:

$$A(r_N) = \frac{r_N \times N}{N_{tol}}, r_N = 4, 5, 6, 7 \ \#(S1)$$

Given assumption that carbon atoms participate in three rings under extreme conditions, A(6) = 1. Therefore, the number N of each ring r_N(4, 5, 6, and 7-membered rings) type

multiplied by the number of atoms in those rings(r_N), divided by the total number of atoms (N_{tol}) in the system yields this average number of atoms value.

This metric reflects the variation in the participation of individual atoms in forming rings within the system. For C_{60}, where each carbon atom contributes to one five-membered ring and two six-membered rings, A(6)=0.67, while A(5)=0.33. A higher A(6) indicates a greater proportion of six-membered rings within the system, whereas a lower A(5) suggests that even when closed-cage structures form, they tend to be larger than C_{60}. In the later stages of this simulation, there is an increase in the formation of seven-membered rings, similar to five-membered rings, both tending towards convergence. This observation implies that during the growth of C_{60} in C_2 vapor, there is a tendency for more six-membered rings to form, indicating a preference for expanding the cage size beyond that of C_{60}.

2.4 Fullerene growth with iron cluster

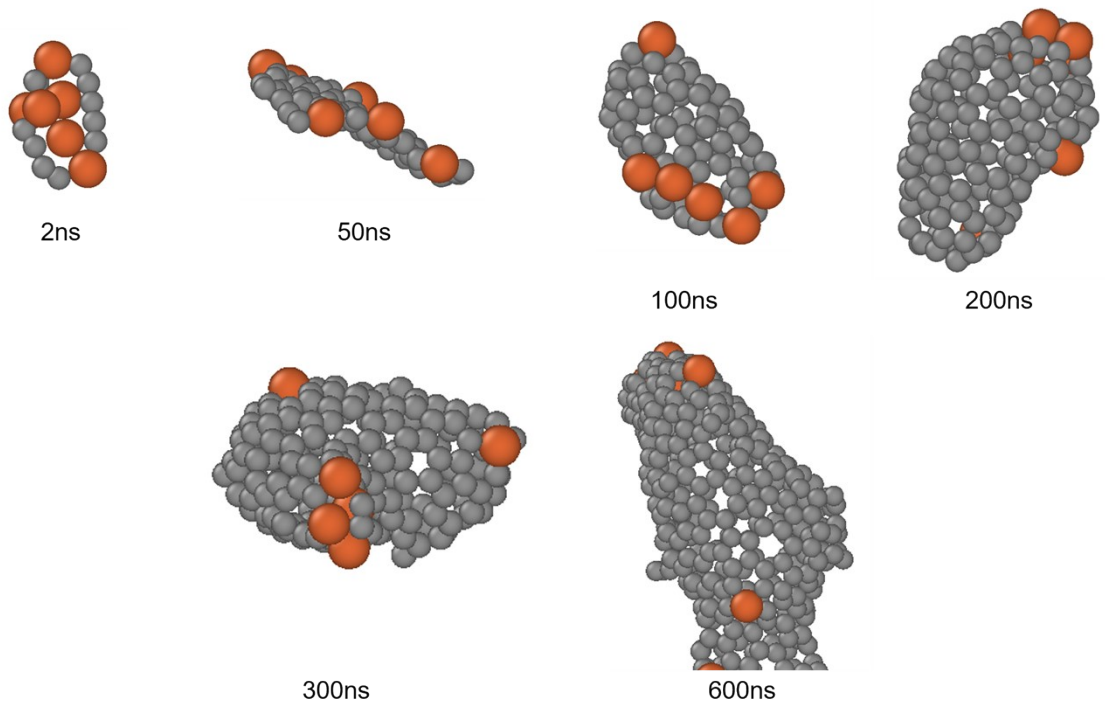


Figure S10: Snapshot from the molecular dynamics simulation of the Fe-C system. Larger red spheres represent iron atoms, while smaller gray spheres represent carbon atoms.

In the molecular dynamics simulations of the Fe-C system shown in **Fig. S10**, a 100 Å large box with 6 iron atoms and 1 carbon atom is used as the initial system. The simulations were performed using the NVT ensemble in LAMMPS with a time step of 0.002 ps. Carbon atoms were added to the system at a rate of 5 atoms per ns until the system contained 40 carbon atoms, and then at a rate of 1 atom every 2 ns until the system contained 460 carbon atoms.

References

[1] D. Zhang, "DPA_1_OC2M," (2022). use the file OC_10M.pb download from https://www.aissquare.com/models/detail?pageType=models&name=DPA_1_OC2M

[2] O.T. Unke, S. Chmiela, M. Gastegger, K.T. Schütt, H.E. Sauceda, and K.-R. Müller, "**SpookyNet**," (2022). https://github.com/OUnke/SpookyNet

[3] L.N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (IEEE, 2017), pp. 464–472.

[4] P. Rowe, V.L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelides, "An accurate and transferable machine learning potential for carbon," The Journal of Chemical Physics **153**(3), (2020).

[5] Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

[6] A.P. Thompson, H.M. Aktulga, R. Berger, D.S. Bolintineanu, W.M. Brown, P.S. Crozier, P.J. in 't Veld, A. Kohlmeyer, S.G. Moore, T.D. Nguyen, R. Shan, M.J. Stevens, J. Tranchida, C. Trott, and S.J. Plimpton, "**LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales**," Comp. Phys. Comm. **271**, 108171 (2022).

[7] A. Stukowski, "**Visualization and analysis of atomistic simulation data with OVITO–the open visualization tool**," Modelling and Simulation in Materials Science and Engineering **18**(1), 015012 (2009).

[8] Plotly Technologies Inc. Collaborative data science. Montréal, QC, 2015. https://plot.ly.