

# Supporting information for

## Machine Learning Quantitatively Characterizes the Deformation and Destruction of Explosive Molecules

Kaining Zhang, Lang Chen\*, Teng Zhang, Jianying Lu, Danyang Liu, Junying Wu

State Key Laboratory of Explosion Science and Technology, Beijing  
Institute of Technology, Beijing 100081, China

Corresponding Authors

\*E-mail: [chenlang@bit.edu.cn](mailto:chenlang@bit.edu.cn) (Lang Chen)

This PDF file includes the following:

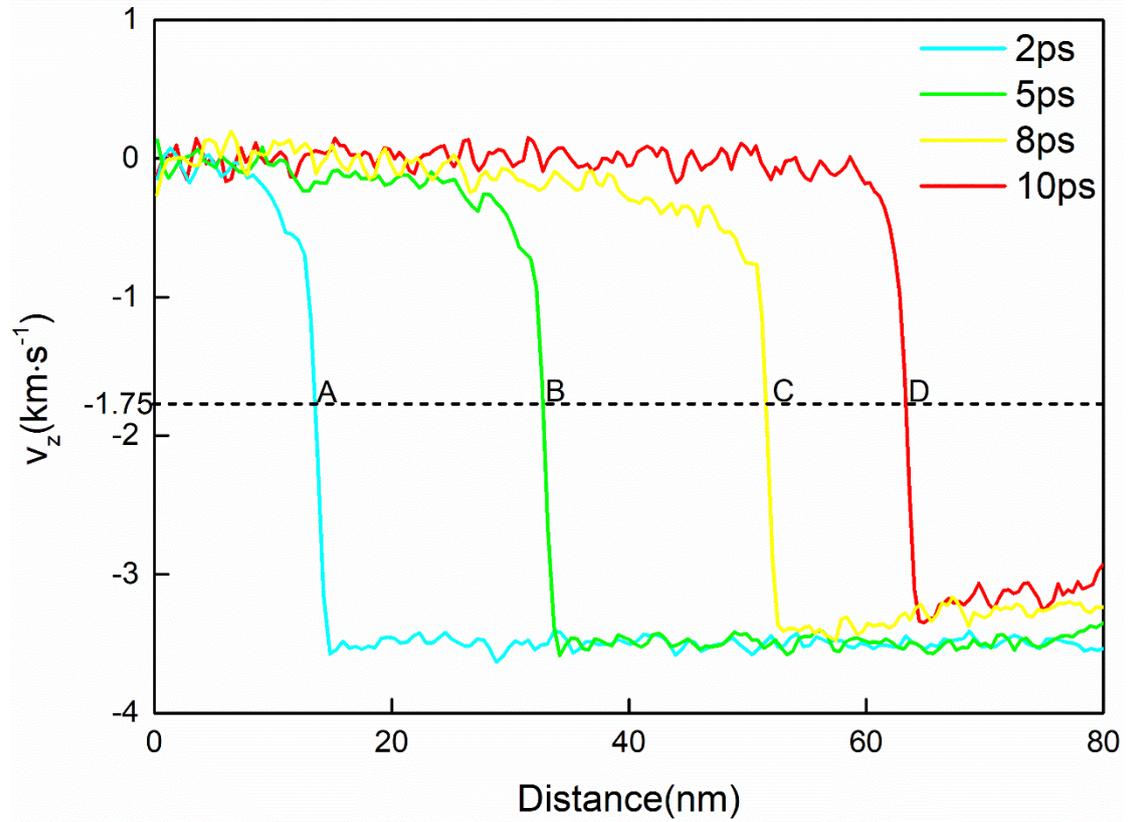
Figs. S1, S2, S3, S4, S5, S6, S7, S8, S9

Tables S1

Chemical Bond Telescopic Deflection,

Molecular Dynamics Simulation of Shock-loaded PETN and Training of the  
Machine Learning Model

**Figure S1:** Under a shock velocity of 3.5 km/s, the atom speed changes with distance along the z-direction at different moments.



## Chemical Bond Telescopic Deflection

Next, we quantify the chemical bond expansion and deflection.

Chemical bond length is:

$$dis = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (S1)$$

where  $(x_i, y_i, z_i)$  and  $(x_j, y_j, z_j)$  are the space coordinates of the atom pair.

The chemical bond expansion is given by

$$\Delta dis = dis_{t=t_{shock}} - dis_{t=0} \quad (S2)$$

where  $\Delta dis$  is the chemical bond expansion,  $dis_{t=t_{shock}}$  is the chemical bond expansion at the shocked moment, and  $dis_{t=0}$  is the chemical bond expansion at the initial moment.

The normal distribution of the chemical bond expansion change is fitted by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (S3)$$

where  $x$  is the chemical bond stretch value,  $\mu$  is the expectation and is the chemical bond stretch value corresponding to the highest peak of the normal curve, which represents the average level of the chemical bond stretch change, and  $\sigma$  is the standard deviation, which represents the degree of dispersion for the chemical bond stretch change. The group distance is set to 0.1 Å.

The angle between the chemical bond and the z-axis direction vector is given by

$$\alpha = \arccos \frac{\vec{A} \cdot \vec{k}}{|\vec{A}| \cdot |\vec{k}|} \quad (S4)$$

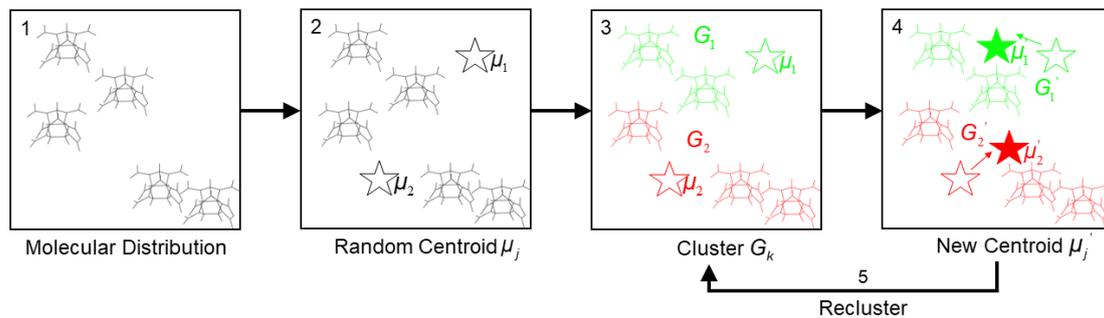
where  $\alpha$  is the angle between the chemical bond and the positive z-axis,  $\vec{A}$  is the chemical bond vector, and  $\vec{k}$  is the unit vector in the positive z-axis direction.

The chemical bond deflection is given by

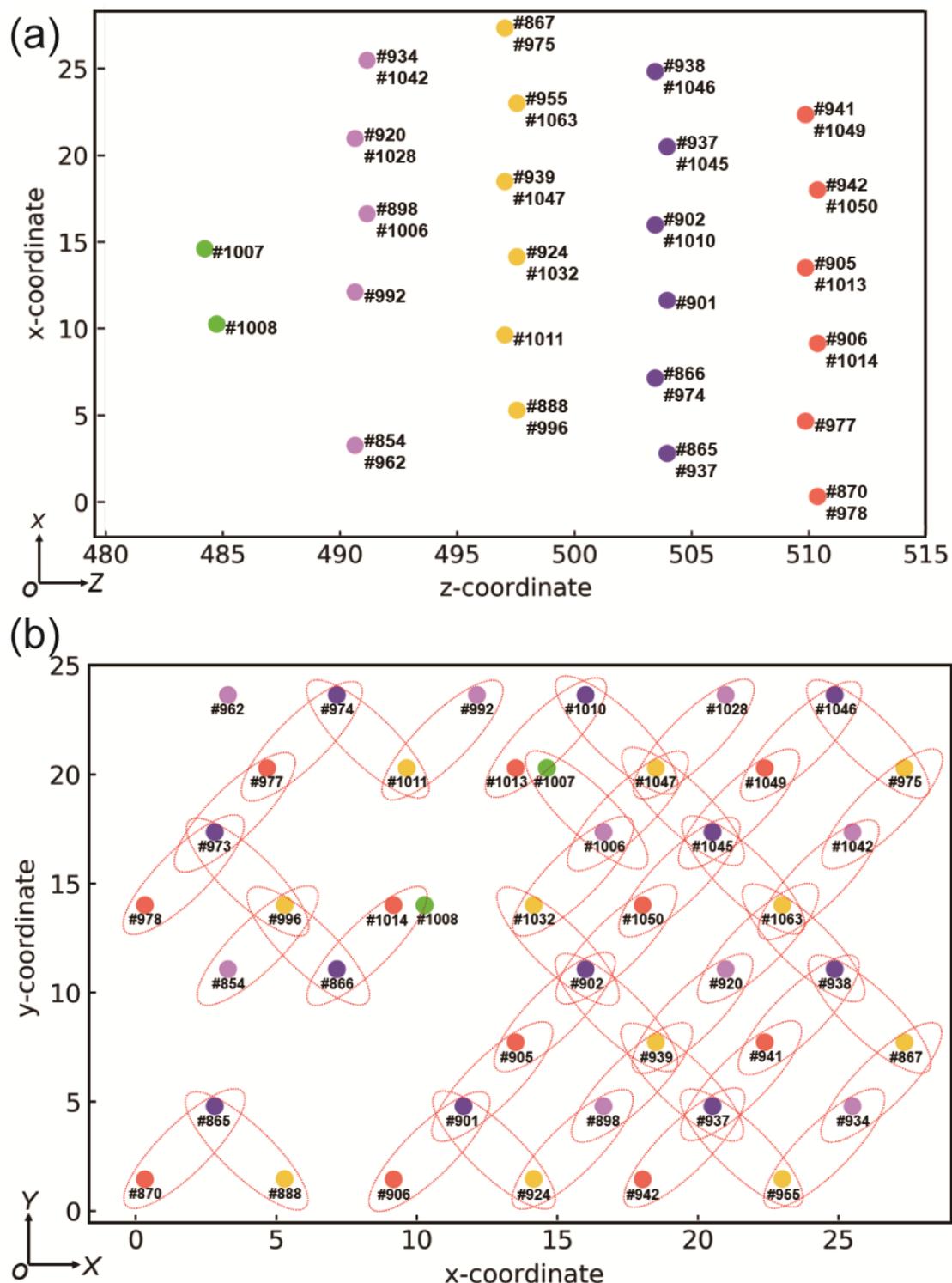
$$\Delta\alpha = \alpha_{t=t_{shock}} - \alpha_{t=0} \quad (S5)$$

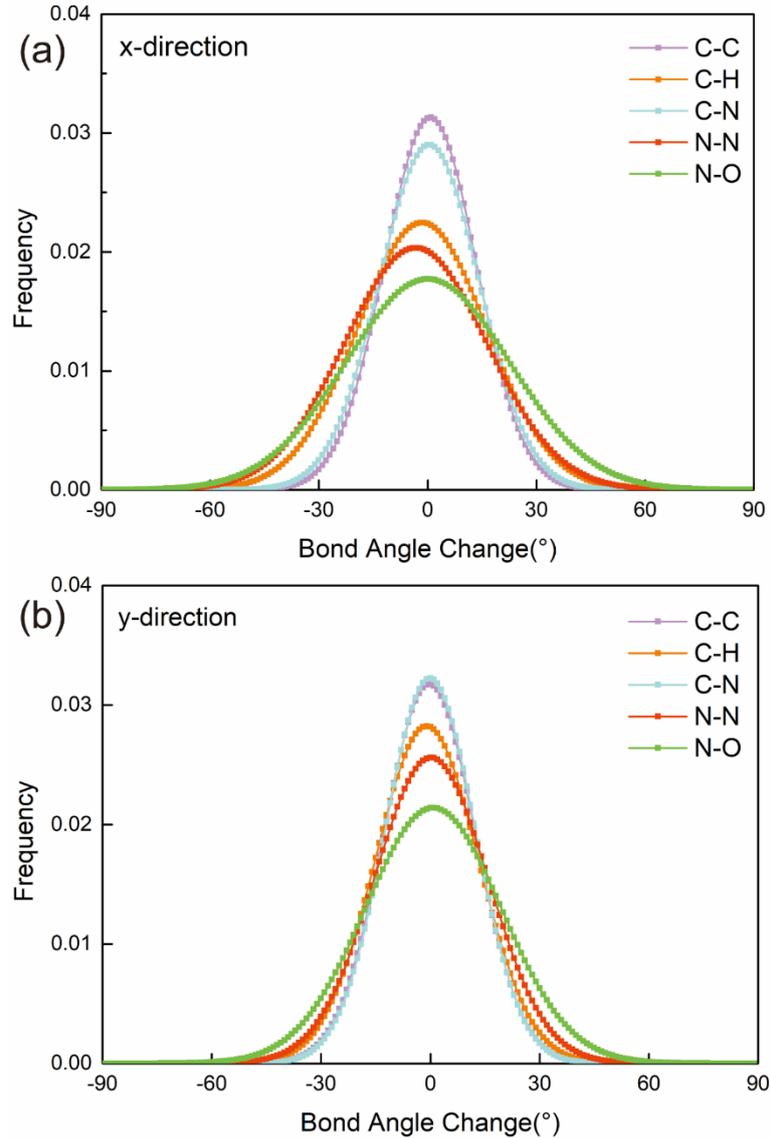
where  $\Delta\alpha$  is the chemical bond deflection,  $\alpha_{t=t_{shock}}$  is the chemical bond deflection at the shocked moment, and  $\alpha_{t=0}$  is the chemical bond deflection at the initial moment.

**Figure S2:** k-means clustering algorithm calculating molecular adjacent relationship schematic diagram.



**Figure S3:** (a) Molecular centroid distribution in the  $XOZ$  plane at the initial moment (note that the z-coordinates for the molecules of the same color are roughly similar) and (b) The molecular centroid distribution in the  $XOY$  plane at the initial moment (note that the molecules in the same circle are three-direction adjacent molecules)





**Figure S4:** Chemical bond deflection normal distribution fitting curve in the x direction (a) and y direction (b).

We loaded shocks on CL-20 in the x-direction and y-direction respectively and compared the deflection degrees of chemical bonds related to nitrogen atoms and chemical bonds related to carbon atoms.

Under shock conditions in three directions, the deflection degree of chemical bonds related to nitrogen atoms is greater than that of chemical bonds related to carbon atoms, which indicates that the radial movement space of chemical bonds related to nitrogen atoms is larger than that of chemical bonds related to carbon atoms. Combined with the snapshot diagram, we believe that

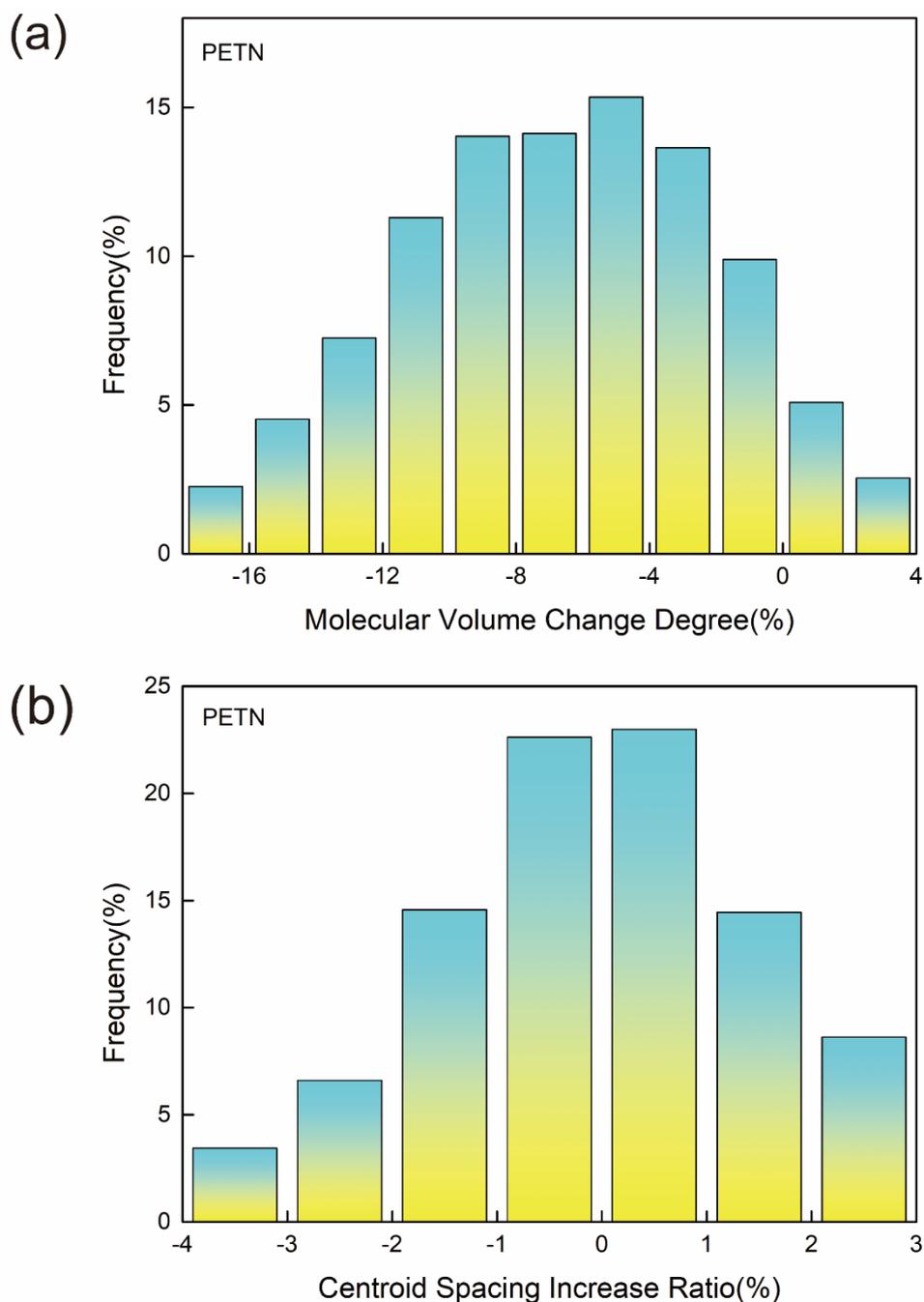
this is because the C-C and C-N chemical bonds are distributed on the cage structure, and the N-N and N-O chemical bonds are located on the extension branch of the cage structure. The chemical bonds on the cage structure are connected in a closed loop, the movement of the chemical bonds is mutually restricted, and the deflection movement space is small; the epitaxial chemical bond is mainly intermolecular movement, which has a larger movement space. Therefore, the degree of deflection of chemical bonds related to N-O and N-N nitrogen atoms is greater than that of chemical bonds related to C-H, C-N, and C-C carbon atoms.

## **Molecular Dynamics Simulation of Shock-loaded PETN and Training of the Machine Learning Model**

### **Molecular Dynamics Simulation of Shock-loaded PETN**

The molecular dynamics simulation process of shock-loaded Pentaerythritol tetranitrate (PETN) is the same as that of shock-loaded CL-20. We obtained the stable PETN unit cell structure from the Cambridge Crystal Database (CCDC 669555). In the unit cell structure, there are nine PETN single molecules. The lengths of the unit cells along the a, b, and c directions are 9.38, 9.38, and 6.7 Å, respectively. Then, the unit cell is expanded by three times, three times, and thirty-two times in the A, B, and C directions, respectively, and finally forms an orthogonal supercell structure containing 576 PETN single-molecules (16704 atoms). After geometrically relaxing the orthogonal supercell, the supercell structure was expanded by four times along the C direction to obtain the large orthogonal PETN supercell. It contains 2304 PETN single molecules (66816 atoms). With the NVE ensemble, the terminal boundary of the supercell along the z-direction is set as a fixed reflecting wall, and the starting boundary is set as a free surface<sup>1</sup>. All atoms with a velocity of 3.5 km/s hit the reflecting wall in the z-direction, generating a shock wave. The total calculation time is 10 ps. We performed 20 parallel simulations and the calculation process for obtaining the preprocessing data is the same as that for CL-20.

### **Molecular Volume Changes and Molecular Spacing Changes**



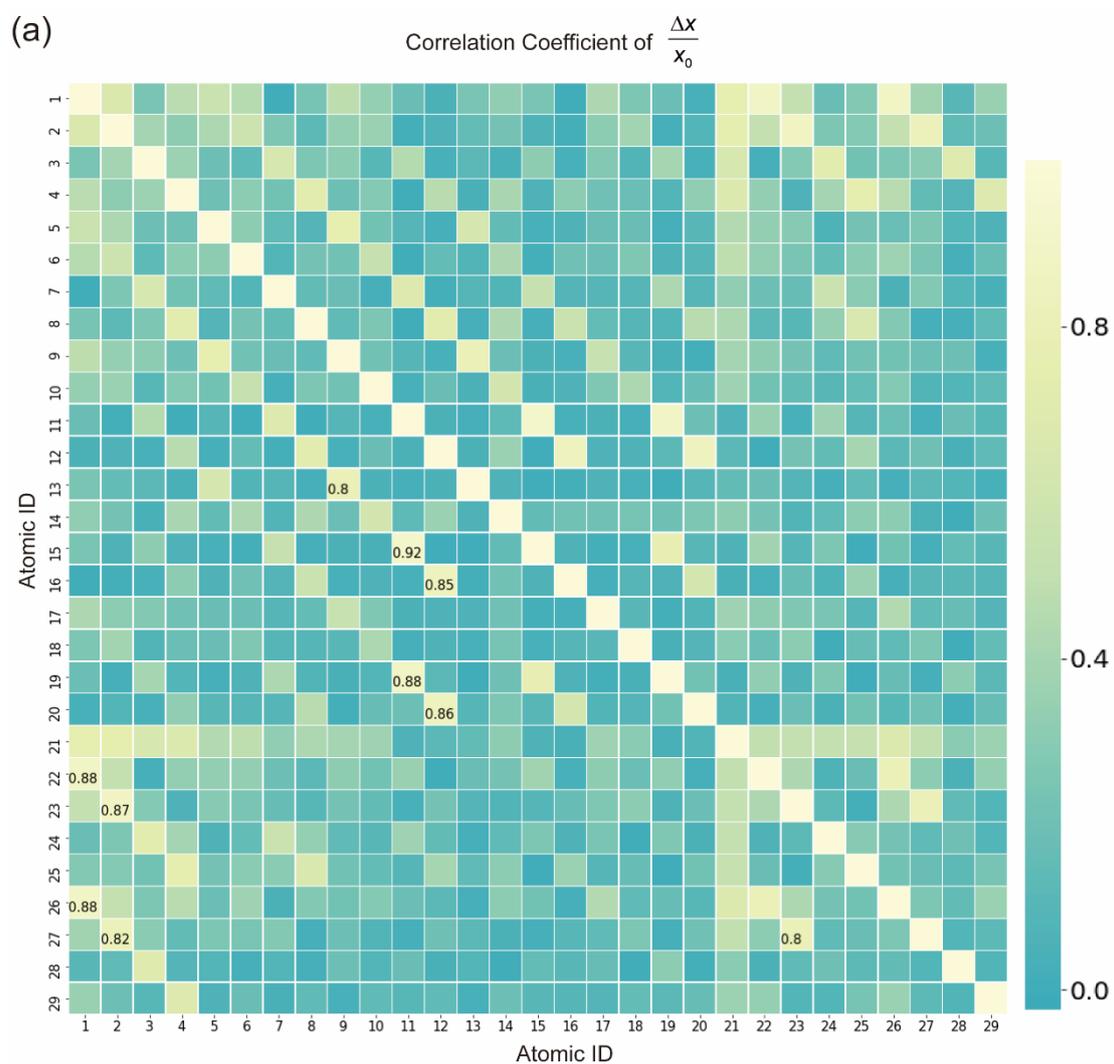
**Figure S5:** (a) Molecular volume change degree frequency distribution histogram and (b) Adjacent molecular centroid spacing increase rate frequency distribution histogram.

Figure S5(a) shows the molecular volume change rate frequency distribution histogram. The volume change rate of PETN is mostly 2-10% shrinkage, followed by a compression rate of 0-2% and 10%-12%. Very few molecules compress more than 12% or expand in volume. Figure S5(b) shows

the adjacent molecular spacing increase rate frequency distribution histogram. The distance change rate between adjacent molecules is mostly compressed by -1%-1%. The maximum compression rate for the adjacent molecular spacing ranges between 3% and 4%, and the maximum expansion rate is within 3%, but the number of these two intervals is very small.

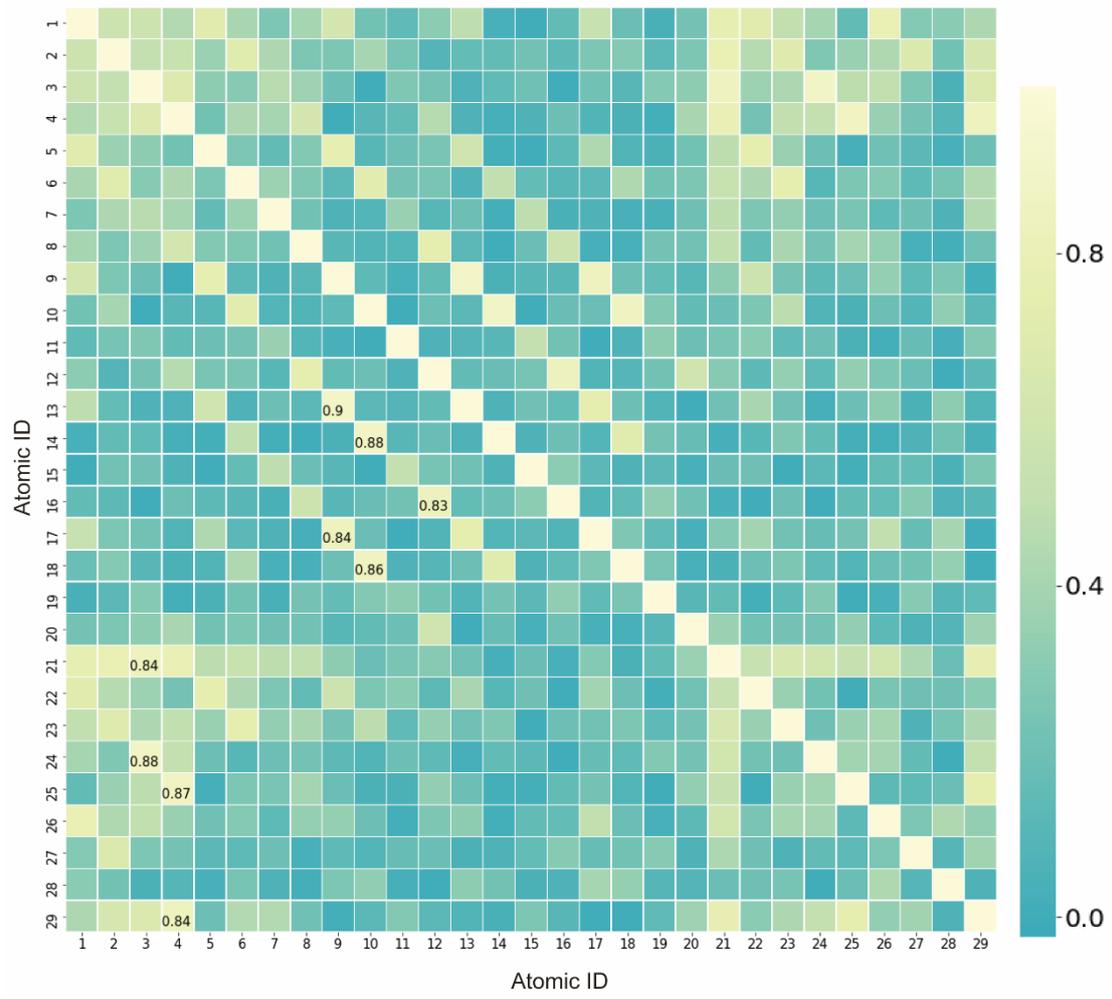
## Training of the Machine Learning Model

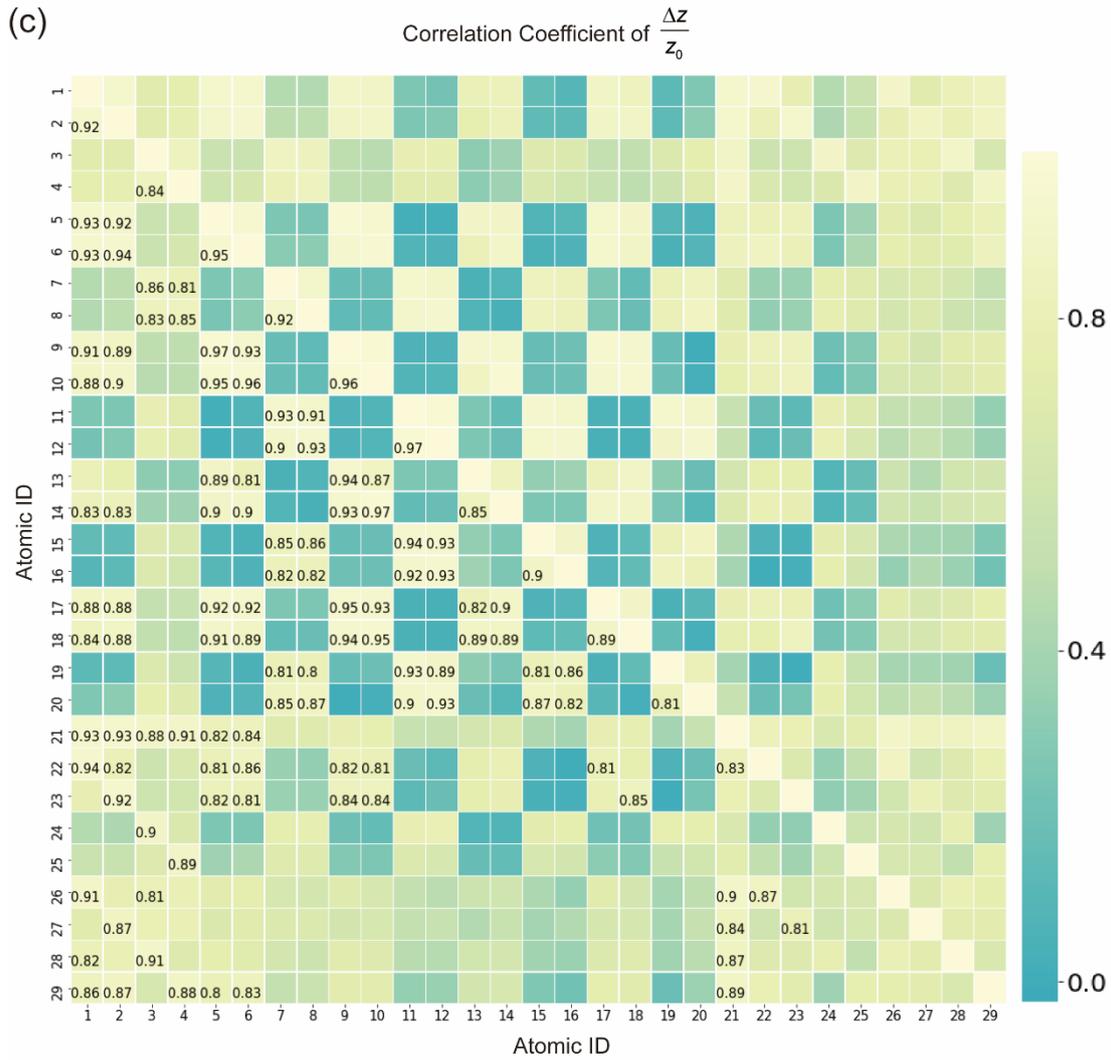
### Correlation Calculation



(b)

Correlation Coefficient of  $\frac{\Delta y}{y_n}$





**Figure S6:** (a) Correlation coefficient of the atomic displacement rate in the x-direction, (b) Correlation coefficient of the atomic displacement rate in the y-direction, and (c) correlation coefficient of atomic displacement rate in the z-direction.

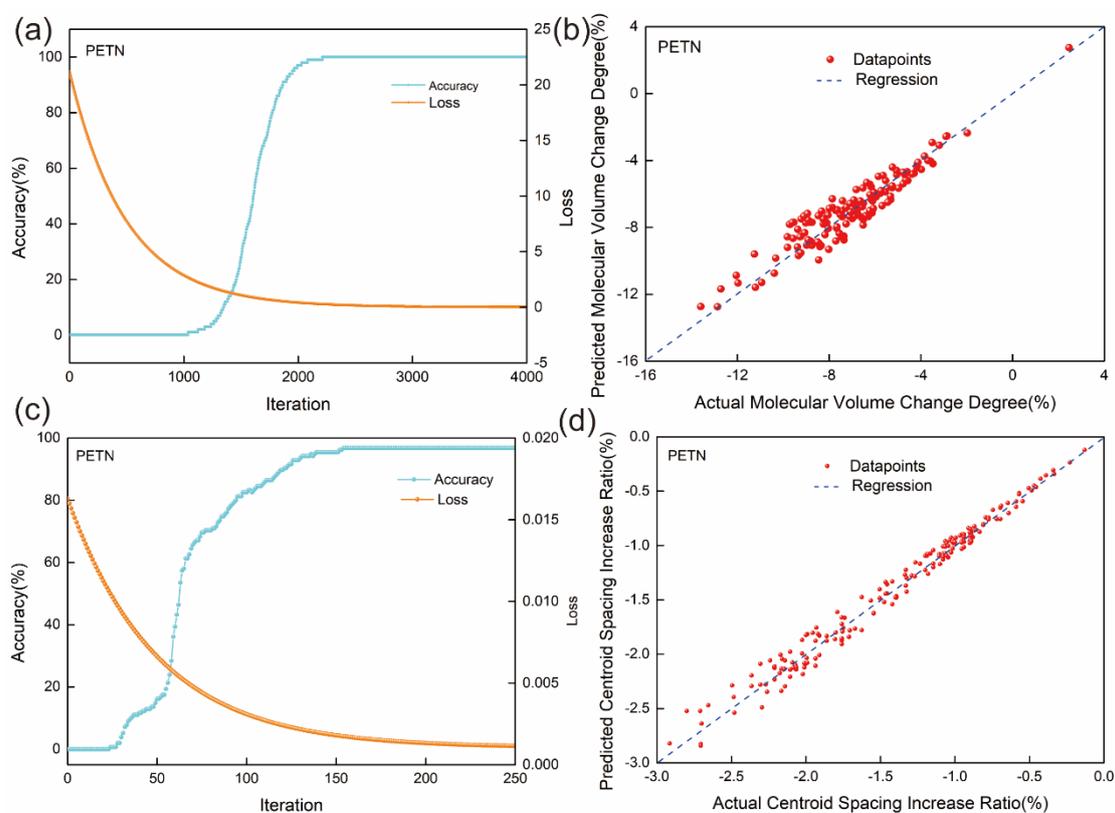
Figure S6(a) shows the correlation coefficient of the atomic displacement rate in the x-direction. The correlation coefficients of the atomic displacement rate in the x-direction between No. 1 and No. 22, No. 1 and No. 26 are 0.88. There is a very strong correlation between the atomic displacement rate in the x-direction between No. 1 and No. 22 and No. 26. Therefore, we can use the No. 1 atomic displacement rate in the x-direction to represent that for No. 22 and 26. The correlation coefficients for the atomic displacement rate in the x-

direction between No. 2 and No. 23, No. 2 and No. 27, are 0.87 and 0.82, respectively. We can use the No. 2 atomic displacement rate in the x-direction to represent that for No. 23 and 27. The correlation coefficient for the atomic displacement rate in the x-direction between No. 9 and No. 13 is 0.8. We can use the No. 9 atomic displacement rate in the x-direction to represent that for No. 13. The correlation coefficients for the atomic displacement rate in the x-direction between No. 11 and No. 15, No. 11 and No. 19, are 0.92 and 0.88, respectively. We can use the No. 11 atomic displacement rate in the x-direction to represent that for No. 15 and 19. The correlation coefficients for the atomic displacement rate in the x-direction between No. 12 and No. 16, No. 12 and No. 20, are 0.85 and 0.86, respectively. We can use the No. 12 atomic displacement rate in the x-direction to represent that for No. 23 and 27. In addition, we retain the x-direction offset rate for other atomic IDs because they do not have a strong correlation with the x-direction offset rates for the other atoms. In summary, we use the atomic displacement rate in the x-direction except No. 13, 15, 16, 19, 20, 22, 23, 26, and 27 to represent that for all atoms.

Figure S6(b) shows the correlation coefficient of the atomic displacement rate in the y-direction. The correlation coefficients for the atomic displacement rate in the y-direction between No. 3 and No. 21, No. 3 and No. 24, No. 4 and No. 25, No. 4 and No. 29, No. 9 and No. 13, No. 9 and No. 17, No. 10 and No. 14, No. 10 and No. 18, No. 12 and No. 16 are 0.84, 0.88, 0.87, 0.84, 0.9, 0.84, 0.88, 0.86 and 0.83, respectively, which is a very strong correlation. We can use the No. 3 atomic displacement rate in the y-direction to represent that for Nos. 21 and 24, the No. 4 atomic displacement rate in the y-direction to represent that for No. 25 and 29, the No. 9 atomic displacement rate in the y-direction to represent that for No. 13 and No. 17, the No. 10 atomic displacement rate in the y-direction to represent that for No. 14 and No. 18, and the No. 12 atomic displacement rate in the y-direction to represent that for No. 16. In summary, we use the atomic displacement rate in the x-direction except No. 13, 14, 16, 17, 18, 21, 24, 25 and 29 to represent that for all atoms.

Figure S6(c) shows the correlation coefficient of the atomic displacement rate in the z-direction. The correlation coefficients for the atomic displacement rate in the z-direction between most of the two atoms are greater than 0.8, which shows a strong correlation. We use the atomic displacement rate in the z-direction for atoms No. 1 and No. 3 as the feature set to represent that for all cage structure atoms.

### Gradient Descent Algorithm Training Process



**Figure S7:** (a) Change in the loss and accuracy of predicting the molecular volume change degree of PETN with the number of iterations. (b) Comparison between the predicted value and the true value of the molecular volume change degree of PETN. (c) Change in the loss and accuracy of predicting the molecular centroid spacing increase ratio of PETN with the number of iterations. (d) Comparison between the predicted value and the true value of the molecular centroid spacing increase ratio of PETN.

Table S1 Feature set and coefficients for the molecular volume change rate determined by the gradient descent algorithm

Feature	$(\frac{\Delta x}{x_0})_1$	$(\frac{\Delta x}{x_0})_2$	$(\frac{\Delta x}{x_0})_3$	$(\frac{\Delta x}{x_0})_4$	$(\frac{\Delta x}{x_0})_5$	$(\frac{\Delta x}{x_0})_6$	$(\frac{\Delta x}{x_0})_7$	$(\frac{\Delta x}{x_0})_8$	$(\frac{\Delta x}{x_0})_9$
$\theta$	-4.43 E-03	-2.58 E-03	2.76 E-03	5.72 E-03	-5.19 E-03	3.49 E-03	6.65 E-03	-9.51 E-03	7.86 E-03
Feature	$(\frac{\Delta x}{x_0})_{10}$	$(\frac{\Delta x}{x_0})_{11}$	$(\frac{\Delta x}{x_0})_{12}$	$(\frac{\Delta x}{x_0})_{14}$	$(\frac{\Delta x}{x_0})_{17}$	$(\frac{\Delta x}{x_0})_{18}$	$(\frac{\Delta x}{x_0})_{21}$	$(\frac{\Delta x}{x_0})_{24}$	$(\frac{\Delta x}{x_0})_{25}$
$\theta$	7.09 E-03	5.45 E-03	2.37 E-03	8.47 E-03	-6.84 E-03	7.45 E-03	3.02 E-03	4.23 E-03	3.69 E-03
Feature	$(\frac{\Delta x}{x_0})_{28}$	$(\frac{\Delta x}{x_0})_{29}$	$(\frac{\Delta y}{y_0})_1$	$(\frac{\Delta y}{y_0})_2$	$(\frac{\Delta y}{y_0})_3$	$(\frac{\Delta y}{y_0})_4$	$(\frac{\Delta y}{y_0})_5$	$(\frac{\Delta y}{y_0})_6$	$(\frac{\Delta y}{y_0})_7$
$\theta$	-4.92 E-03	-5.29 E-03	-3.09 E-03	9.27 E-04	-1.10 E-02	8.53 E-04	3.89 E-03	-8.36 E-03	2.89 E-03
Feature	$(\frac{\Delta y}{y_0})_8$	$(\frac{\Delta y}{y_0})_9$	$(\frac{\Delta y}{y_0})_{10}$	$(\frac{\Delta y}{y_0})_{11}$	$(\frac{\Delta y}{y_0})_{12}$	$(\frac{\Delta y}{y_0})_{15}$	$(\frac{\Delta y}{y_0})_{19}$	$(\frac{\Delta y}{y_0})_{20}$	$(\frac{\Delta y}{y_0})_{22}$
$\theta$	-3.25 E-03	-9.02 E-03	1.33 E-03	-4.94 E-03	-9.01 E-03	-3.07 E-03	-5.34 E-03	2.31 E-03	-6.32 E-03
Feature	$(\frac{\Delta y}{y_0})_{23}$	$(\frac{\Delta y}{y_0})_{26}$	$(\frac{\Delta y}{y_0})_{27}$	$(\frac{\Delta y}{y_0})_{28}$	$(\frac{\Delta z}{z_0})_1$	$(\frac{\Delta z}{z_0})_3$			
$\theta$	-1.88 E-03	-5.93 E-03	1.10 E-03	-4.16 E-03	1.08 E-02	-6.51			

Note:  $\theta_0 = -8.63E-04$

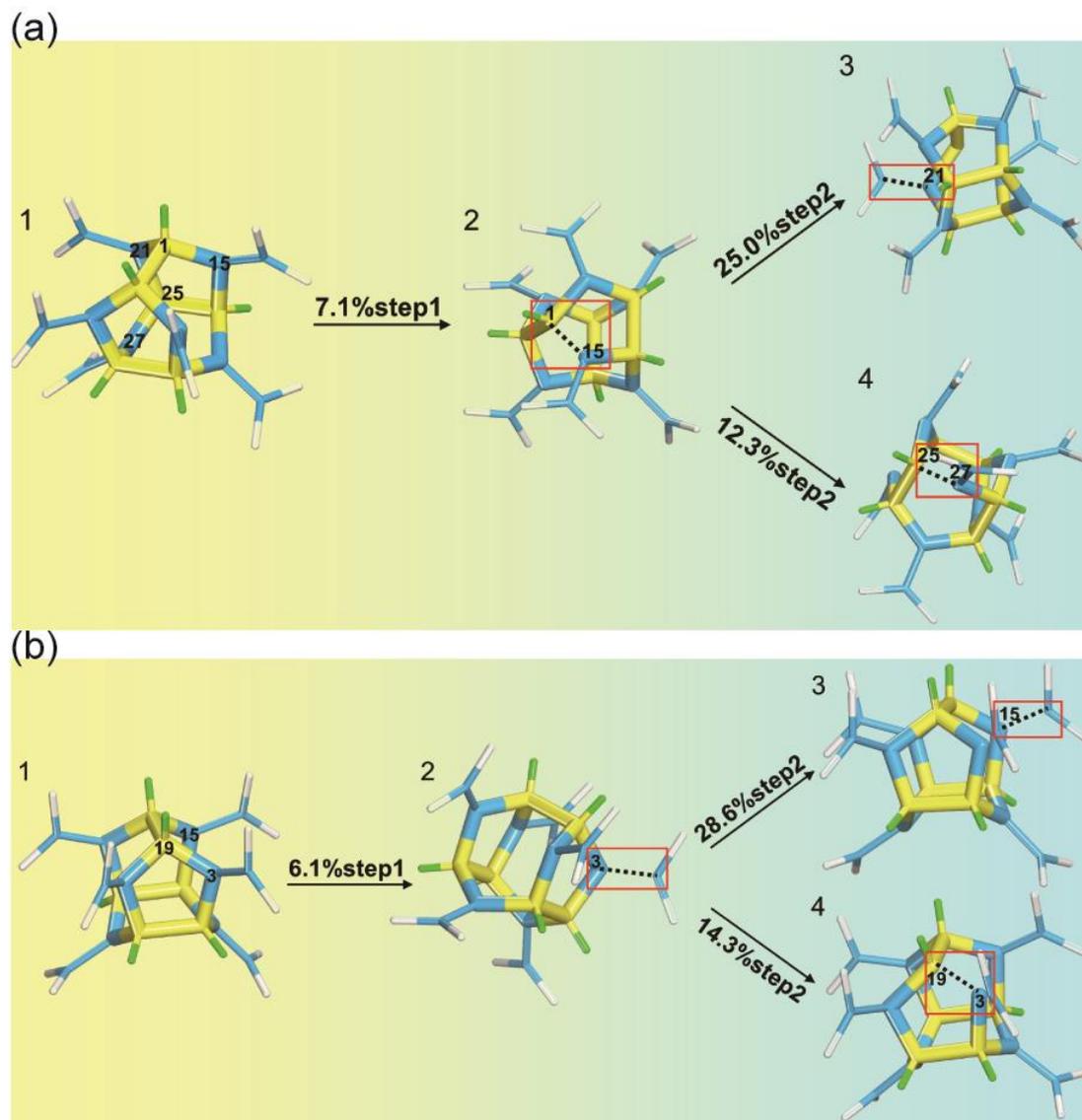
We applied the molecular volume change degree machine learning model of CL-20 to the PETN data, and the learning rate of the gradient descent algorithm for predicting the degree of volume change of the PETN molecule is the same as those of the CL-20 molecule. Figure S7(a) shows the change in the loss and accuracy of predicting and calculating the molecular volume change degree of PETN with the number of iterations. After 4000 iterative calculations, the loss value tended to 0 and the accuracy rate tended to 1. This shows that the machine learning model for predicting the degree of molecular volume change of PETN has good prediction ability. Finally, the feature set and coefficients for the cage structure volume change rate determined by the gradient descent algorithm are shown in Table 1. Figure S7(b) shows a

comparison between the predicted value and the true value of the molecular volume change degree of PETN. The points composed of the true value and the predicted value are close to the optimal model line, which indicates that the predicted value of the molecular volume change degree is close to the true value. This demonstrates that the calculation model can accurately predict the degree of molecular volume change of PETN according to the molecular conformation change.

Figure S7(c) shows the change in the loss and accuracy of predicting the molecular centroid spacing increase ratio of PETN with the number of iterations. After 250 iterative calculations, the loss value tended to 0 and the accuracy rate reached 96.13%. This shows that the calculation model for predicting the degree of centroid spacing increase ratio of PETN has good prediction ability. The feature set and coefficients for the centroid spacing increase ratio determined by the gradient descent algorithm are shown in equation (S1). Figure S7(d) shows a comparison between the predicted value and the true value of the centroid spacing increase ratio of PETN. The points composed of the true value and the predicted value are close to the optimal model line, which indicates that the predicted value of the centroid spacing increase ratio is close to the true value. This demonstrates that the calculation model can accurately predict the degree of centroid spacing increase ratio of PETN according to the molecular conformation change.

$$f\left(\frac{\Delta V}{V_0}\right) = 2.12 \times 10^{-3} \times \left(\frac{\Delta V}{V_0}\right)_1 - 0.18 \cdot \left(\frac{\Delta V}{V_0}\right)_2 - 4.12 \times 10^{-3} \quad (\text{S1})$$

**Figure S8:** Initial destruction path with the second probability (a) and with the third probability (b), the dotted line is the position of the broken bond.



**Figure S9:** (a) N-O and N-N damage point change with time and (b) Cage structure damage point change with time, the dotted line is the position of the broken bond.

