

Supplementary Materials

Exploring structures of nanoclusters combining high-dimensional neural network potentials with unsupervised machine learning algorithms

Huabing Cai¹, Aoni Xu², Dan Villamanca², Ang Cao^{1,3}, Jianhua Yan¹

¹State Key Laboratory of Clean Energy Utilization, College of Energy Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, China.

²School of Chemical and Biomolecular Engineering, The University of Sydney, Sydney 2006, Australia.

³Mongolia Daqingshan Laboratory, Hohhot 017000, Inner Mongolia, China.

Correspondence to: Dr. Ang Cao, State Key Laboratory of Clean Energy Utilization, College of Energy Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, China; Mongolia Daqingshan Laboratory, Hohhot 017000, Inner Mongolia, China. E-mail: angc@zju.edu.cn

1. The high-dimensional neural network^{1,2}

The main idea of this neural network is to represent the total energy as the sum of the energies of all the atoms,

$$E = \sum_i E_i \quad (1)$$

The E_i represents the energy of atom i . The atomic energies are trained by atomic neural networks using the symmetry functions to describe each atom's structural and chemical environment information. In this work, we choose the following three symmetry functions to represent the local chemical environment of all the atoms (oxygen, cerium atoms) from Cartesian coordinates, and the mathematical expressions of the three symmetry functions are as follows:

Radial symmetry function

$$G_i^{radial} = \sum_{j \neq i} e^{-\eta(r_{ij}-r_s)^2} f_c(r_{ij}) \quad (2)$$

Narrow angular symmetry function

$$G_i^{ang.n.} = 2^{1-\zeta} \sum_{\substack{j,k \neq i \\ j < k}} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta((r_{ij}-r_s)^2 + (r_{ik}-r_s)^2 + (r_{jk}-r_s)^2)} f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk}) \quad (3)$$

Wide angular symmetry function

$$G_i^{ang.w.} = 2^{1-\zeta} \sum_{\substack{j,k \neq i \\ j < k}} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta((r_{ij}-r_s)^2 + (r_{ik}-r_s)^2)} f_c(r_{ij}) f_c(r_{ik}) \quad (4)$$

And, $f_c(r_{ij})$ represents the cutoff function, which has the form

$$f_c(r_{ij}) = \begin{cases} 0.5 \times \left[\cos\left(\frac{\pi r_{ij}}{r_c}\right) + 1 \right] & \text{for } r_{ij} \leq r_c, \\ 0 & \text{for } r_{ij} > r_c. \end{cases} \quad (5)$$

r_{ij} represents the distance between atom i and atom j . r_c represents the cutoff radius.

To evaluate the accuracy of the neural network potential (NNP), the root-mean-square error (RMSE) of energies calculated by DFT and predicted by NNP is defined as:

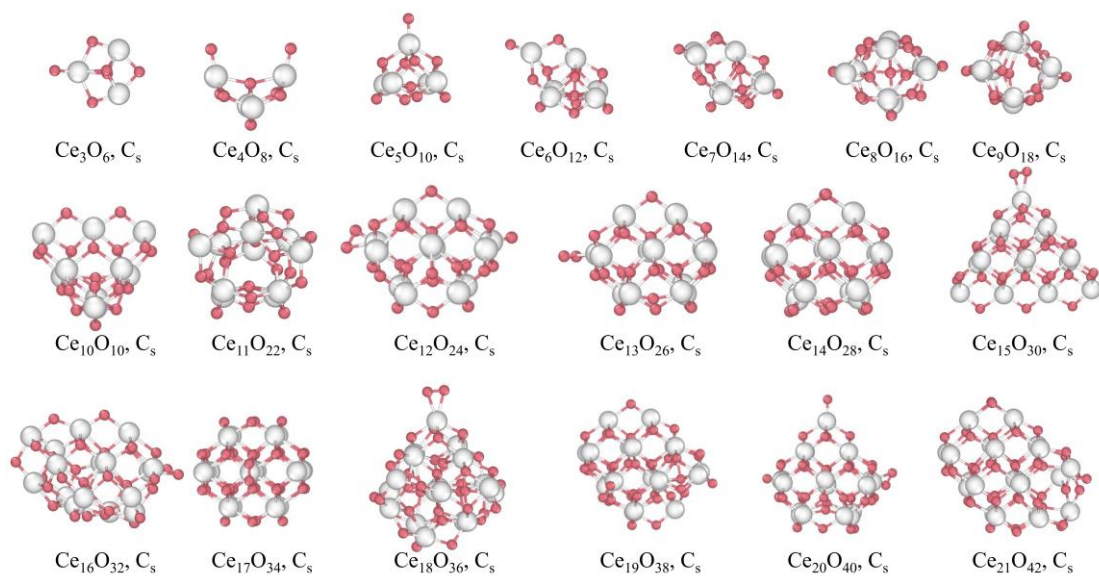
$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (E_{DFT}^{ref} - E_{NNP})^2} \quad (6)$$

M represents the total number of data sets.

Supplementary Table 1. Parameters of symmetry functions performed using the n2p2 code³

Radial			Angular				
η	r_s	r_c	η	λ	ζ	r_s	r_c
2	1.5	6	0.2222	-1	1	0	6
2	2	6	0.2222	1	1	0	6
2	2.5	6	0.2222	-1	6	0	6
2	3	6	0.2222	1	6	0	6
2	3.5	6	0.0408	-1	1	0	6
2	4	6	0.0408	1	1	0	6
2	4.5	6	0.0408	-1	6	0	6
2	5	6	0.0408	1	6	0	6
2	5.5	6	0.0165	-1	1	0	6
			0.0165	1	1	0	6
			0.0165	-1	6	0	6
			0.0165	1	6	0	6

2. The low-energy structures of cerium oxide nanoclusters



Supplementary Figure 1. The most stable structures and their corresponding symmetries of stoichiometric cerium oxide clusters ranging from Ce₃O₆ to Ce₂₁O₄₂. Color code: red, O atom; white, Ce atom.

Supplementary Table 2. Comparison of NNP-predicted energies and DFT-calculated energies for the lowest-energy structures

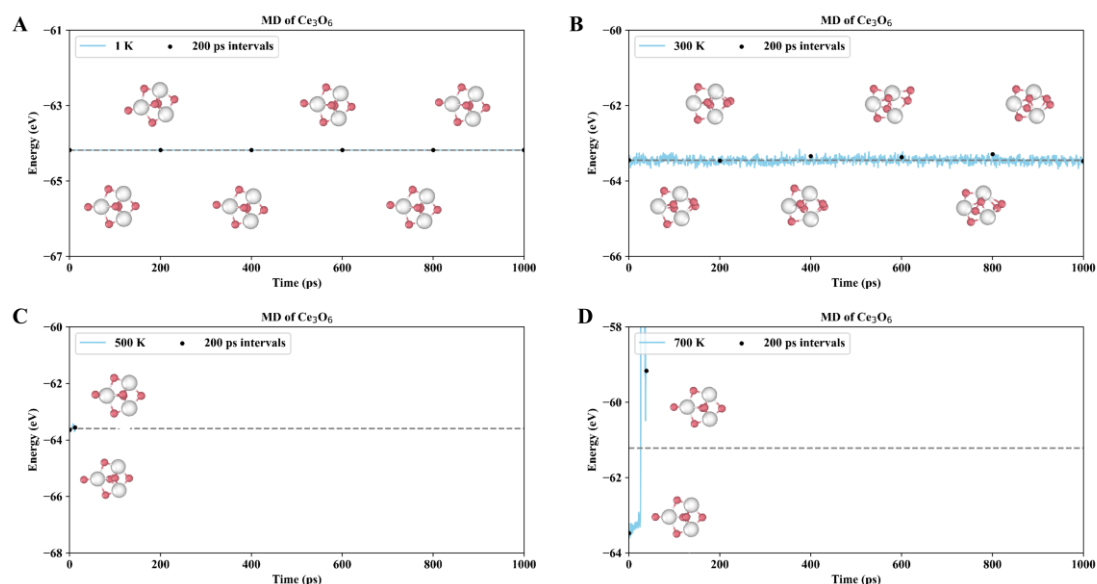
Nanoclusters	NNP-predicted energy (eV)	DFT-calculated energy (eV)	Absolute energy errors (eV)
Ce3O6	-64.094	-63.935	0.158
Ce4O8	-87.013	-86.995	0.018
Ce5O10	-110.283	-110.179	0.104
Ce6O12	-133.252	-133.179	0.073
Ce7O14	-156.483	-156.440	0.043
Ce8O16	-179.879	-180.062	0.183
Ce9O18	-203.045	-203.242	0.197
Ce10O20	-226.315	-226.679	0.364
Ce11O22	-249.062	-249.194	0.128
Ce12O24	-272.963	-273.062	0.099
Ce13O26	-296.164	-296.598	0.434
Ce14O28	-320.855	-320.756	0.099
Ce15O30	-343.747	-343.661	0.086
Ce16O32	-366.086	-366.354	0.268
Ce17O34	-388.722	-388.977	0.255
Ce18O36	-412.740	-412.807	0.067
Ce19O38	-436.624	-437.019	0.395
Ce20O40	-461.206	-461.373	0.167
Ce21O42	-484.147	-484.490	0.343

The stability function was defined as follows:

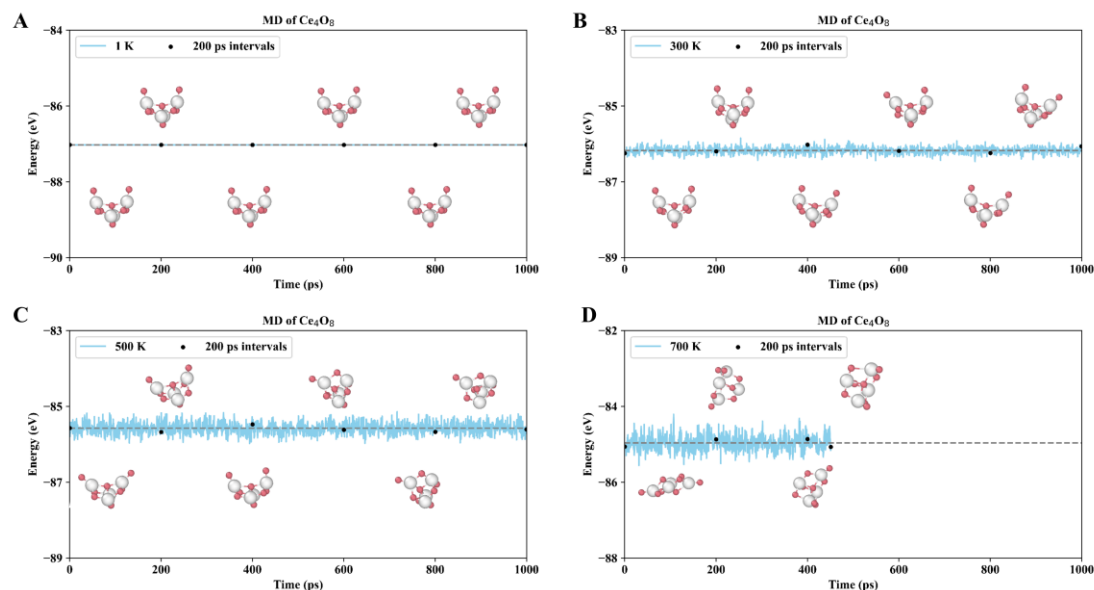
$$\Delta^2 E(Ce_n O_{2n}) = 2E_{tot}^{Ce_n O_{2n}} - E_{tot}^{Ce_{(n-1)} O_{2(n-1)}} - E_{tot}^{Ce_{(n+1)} O_{2(n+1)}} \quad (7)$$

Wherein the range of n is from 4 to 20, and $E_{tot}^{Ce_n O_{2n}}$ denotes the total energy of the cluster $Ce_n O_{2n}$.

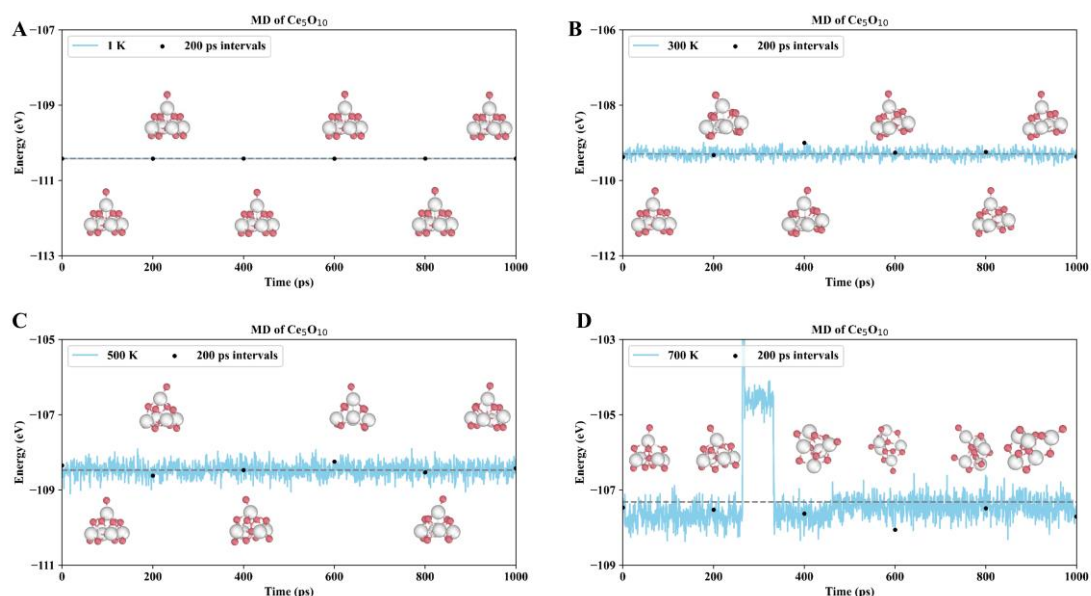
3. The NNPM simulations of the most stable nanoclusters



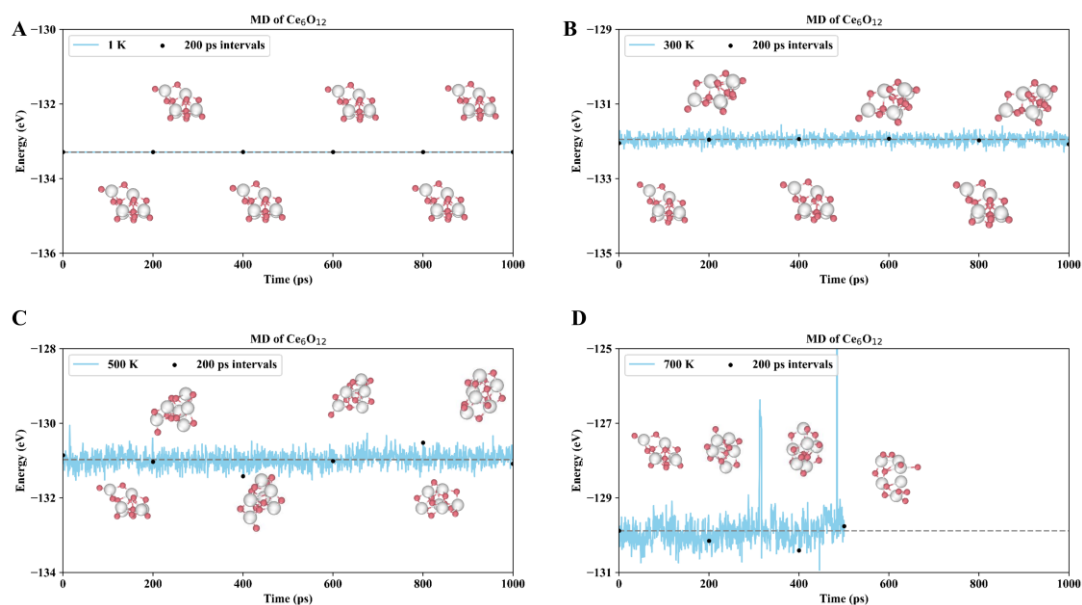
Supplementary Figure 2. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_3O_6 at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



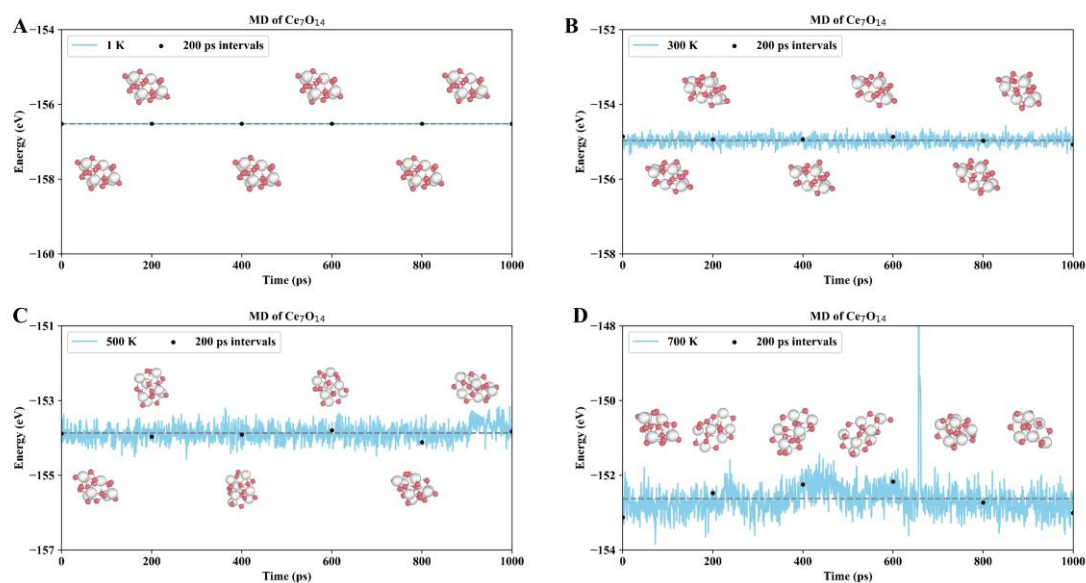
Supplementary Figure 3. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_4O_8 at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



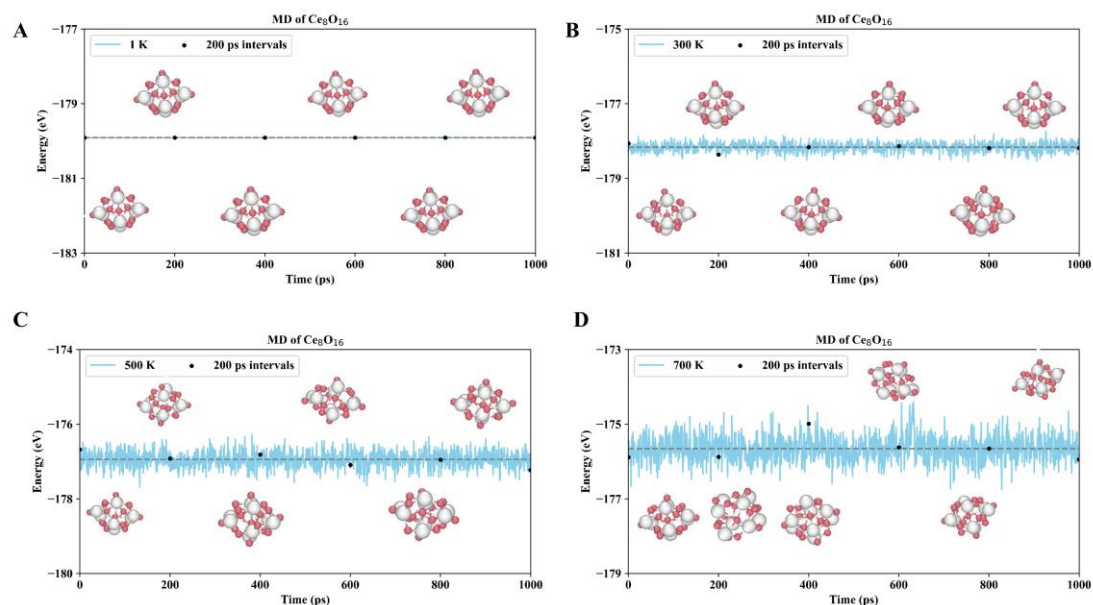
Supplementary Figure 4. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_5O_{10} at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



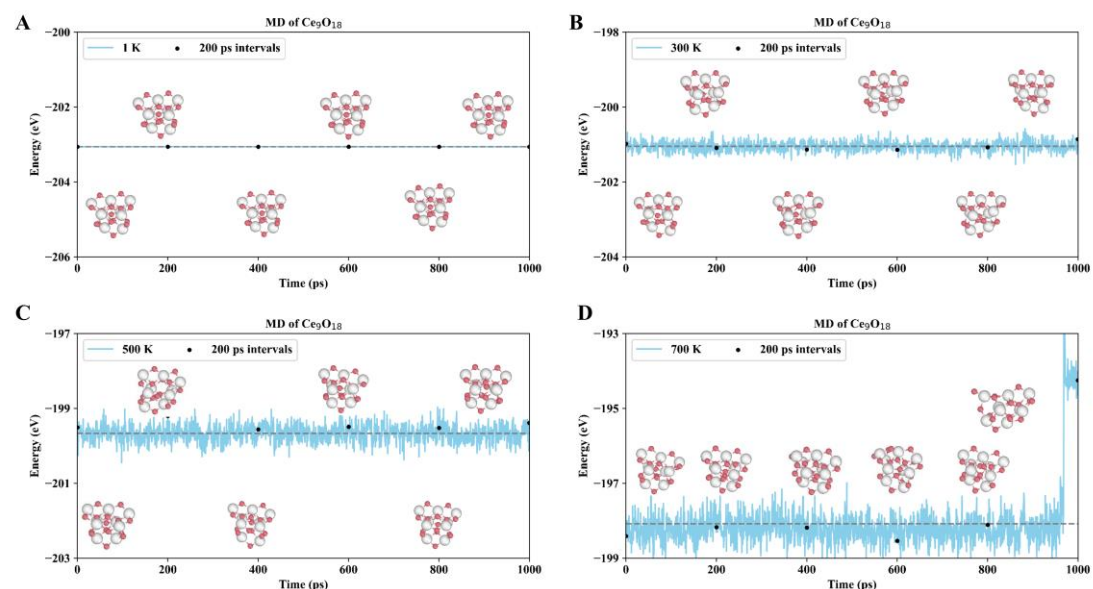
Supplementary Figure 5. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_6O_{12} at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



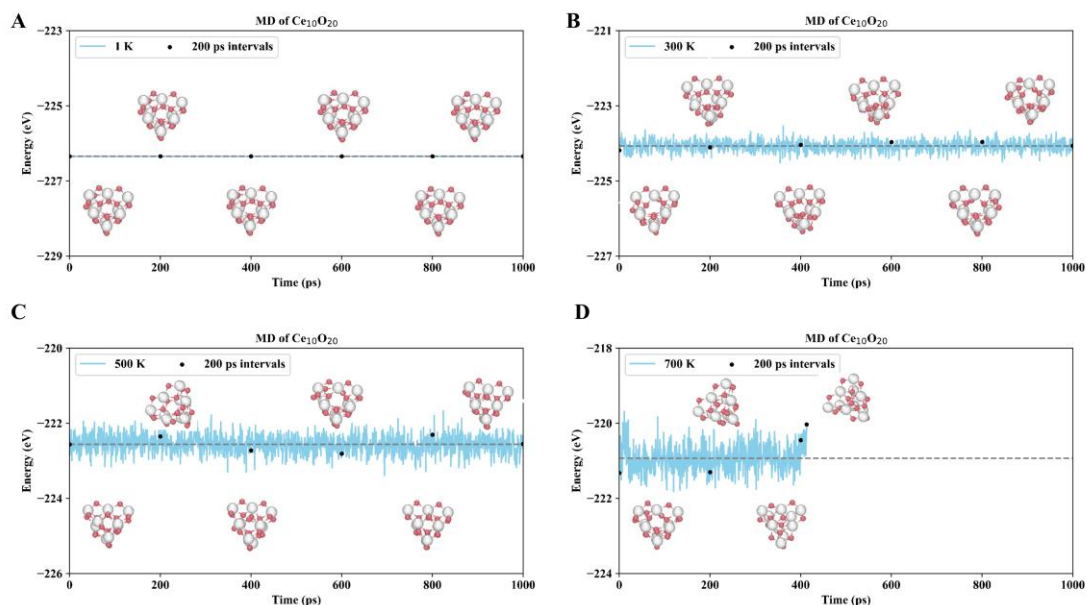
Supplementary Figure 6. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_7O_{14} at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



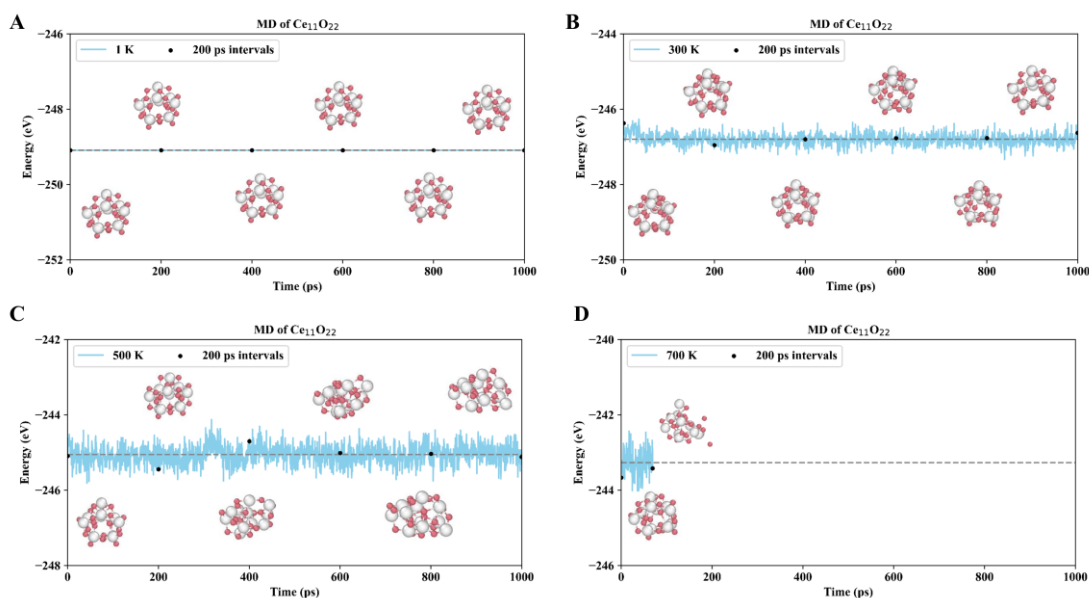
Supplementary Figure 7. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_8O_{16} at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



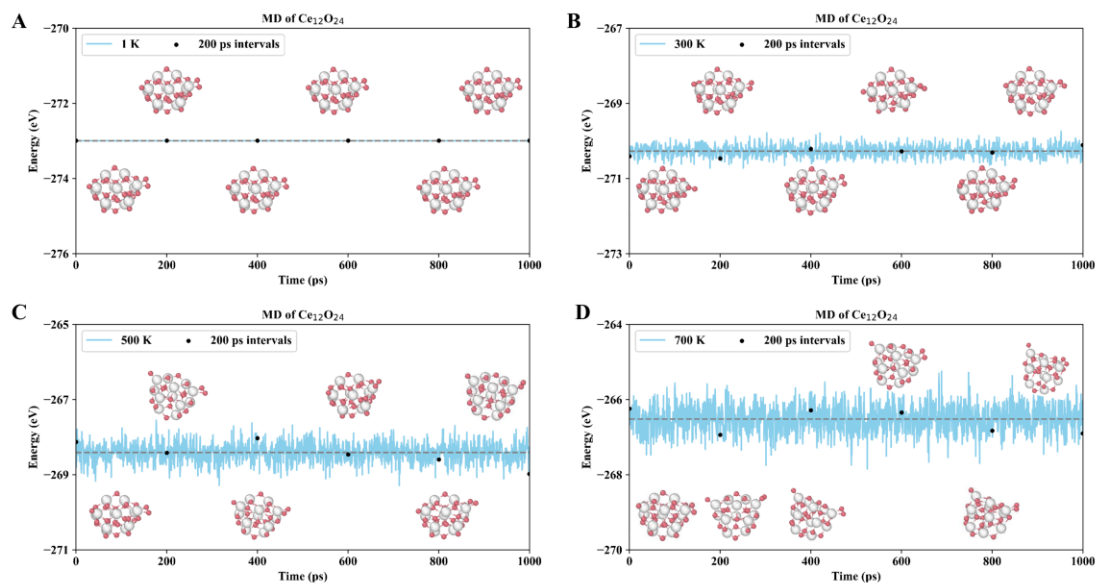
Supplementary Figure 8. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster Ce_9O_{18} at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



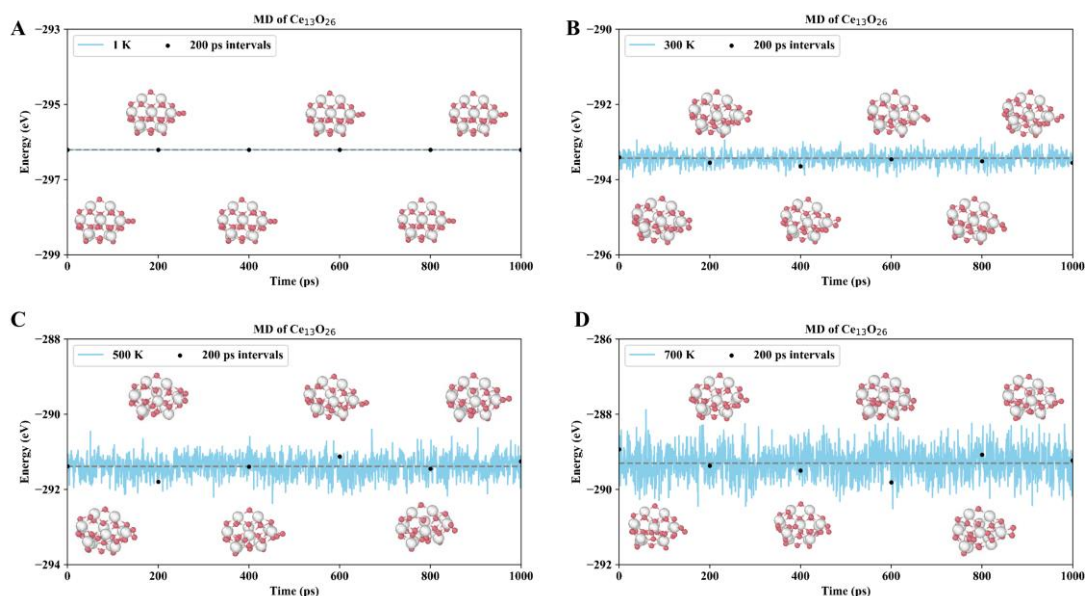
Supplementary Figure 9. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{10}\text{O}_{20}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



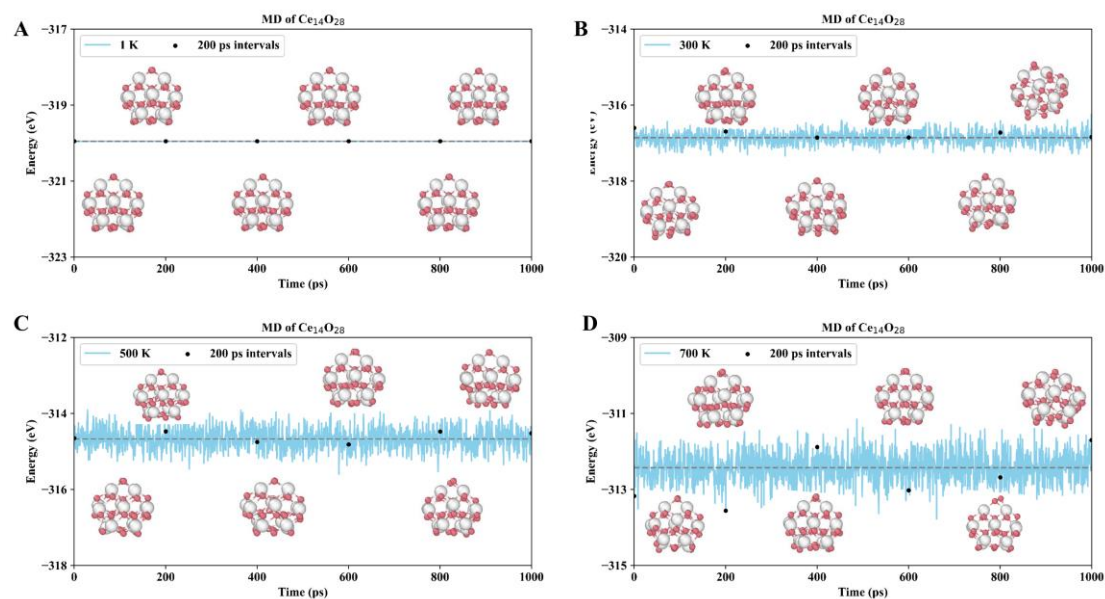
Supplementary Figure 10. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{11}\text{O}_{22}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



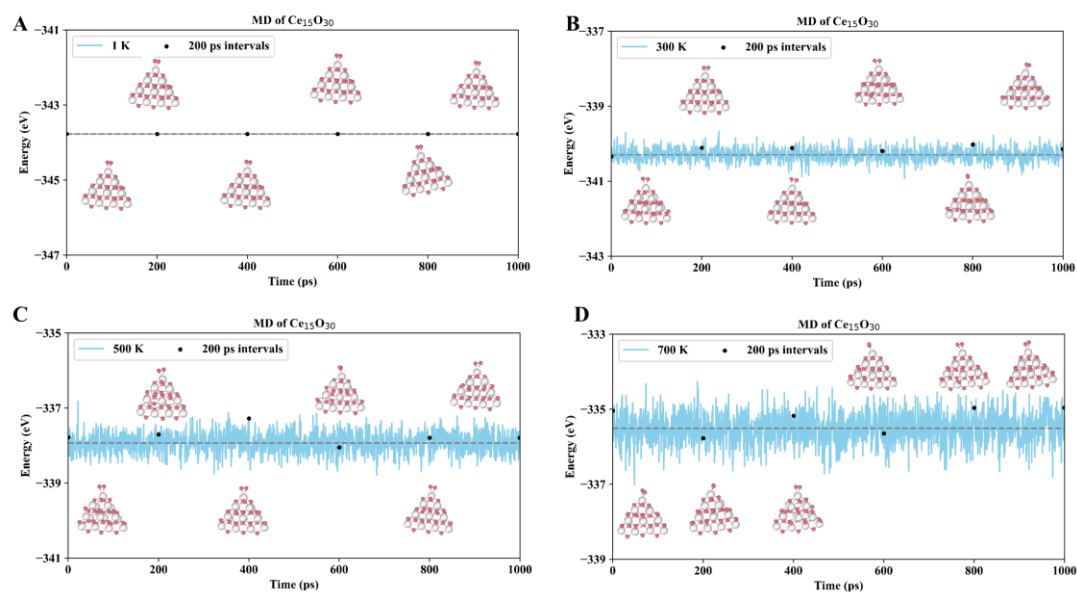
Supplementary Figure 11. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{12}\text{O}_{24}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



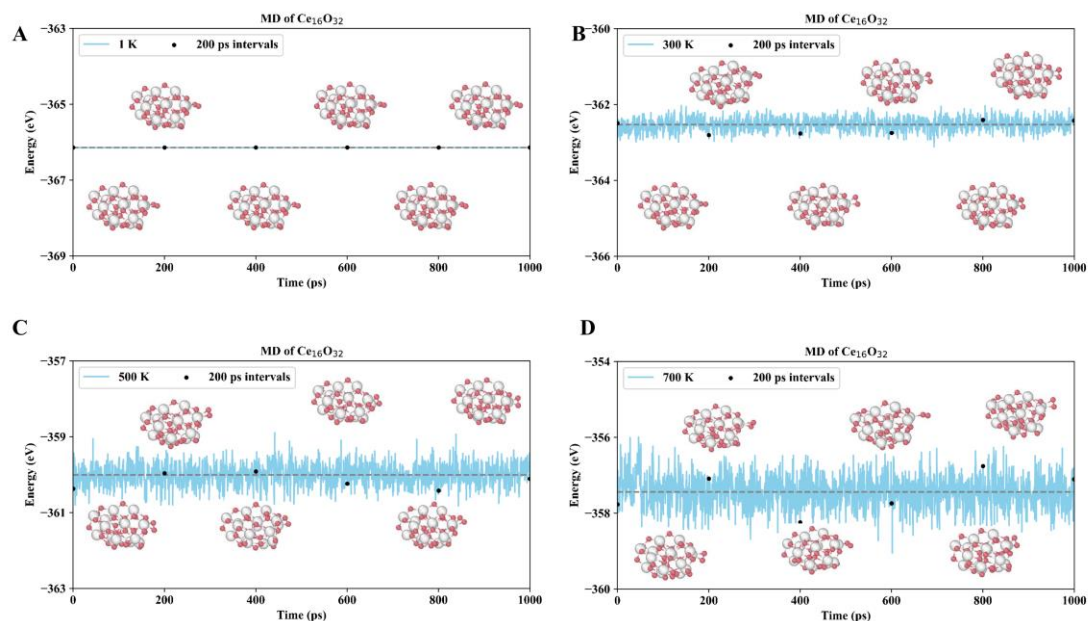
Supplementary Figure 12. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{13}\text{O}_{26}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



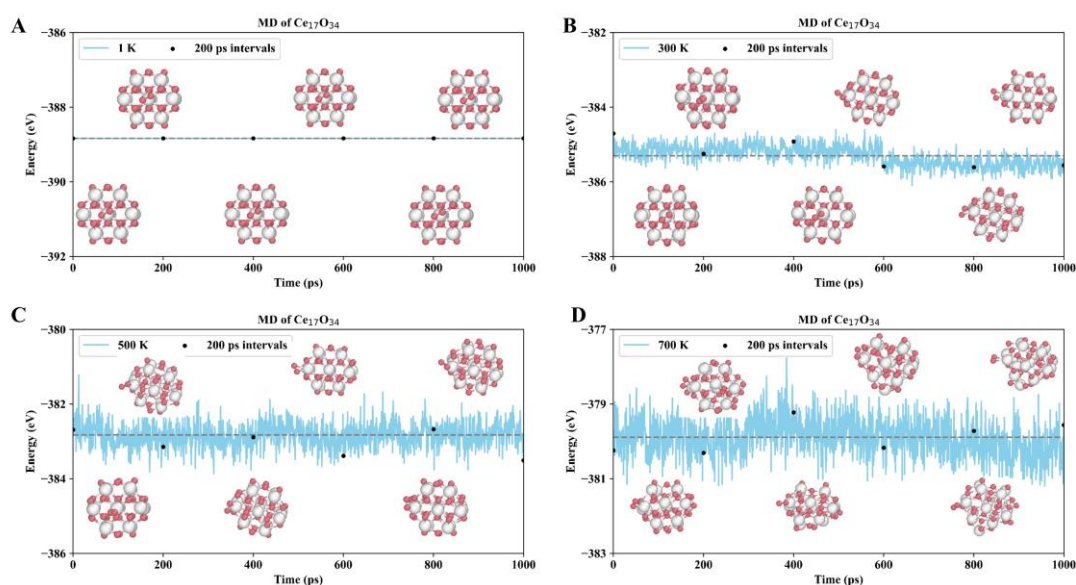
Supplementary Figure 13. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{14}\text{O}_{28}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



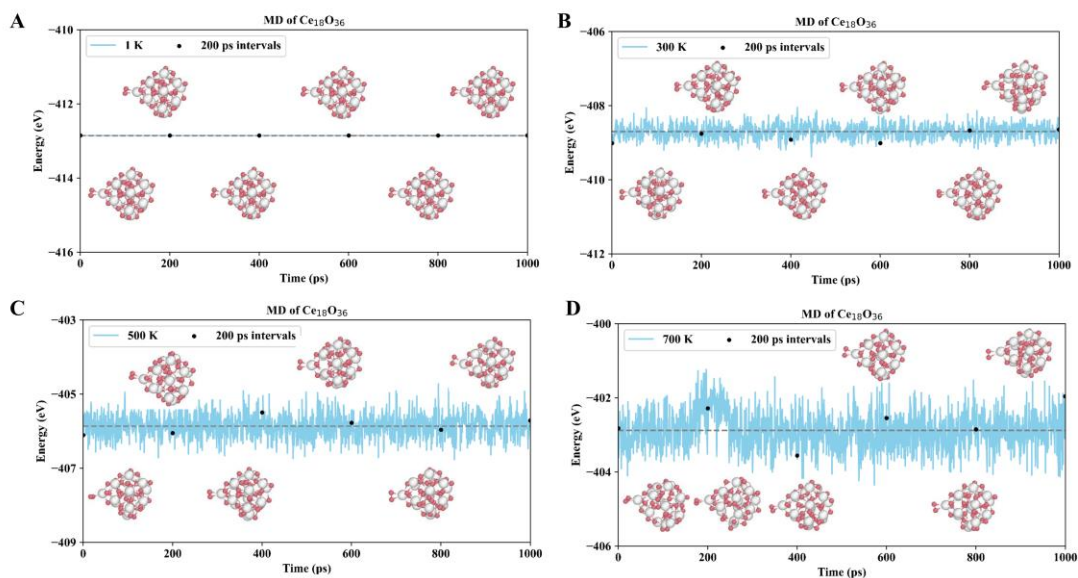
Supplementary Figure 14. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{15}\text{O}_{30}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



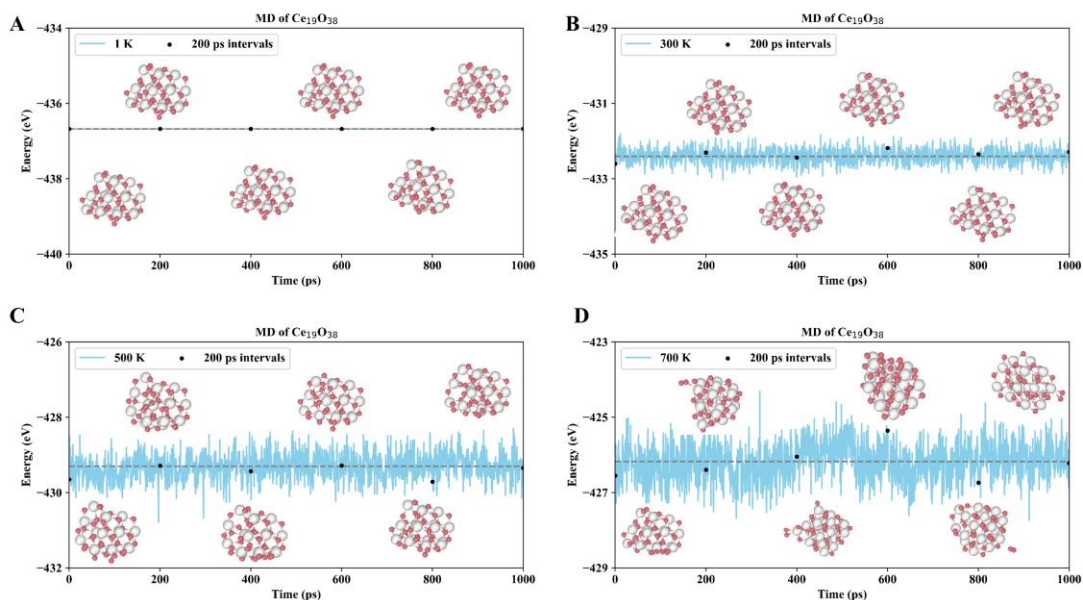
Supplementary Figure 15. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{16}\text{O}_{32}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



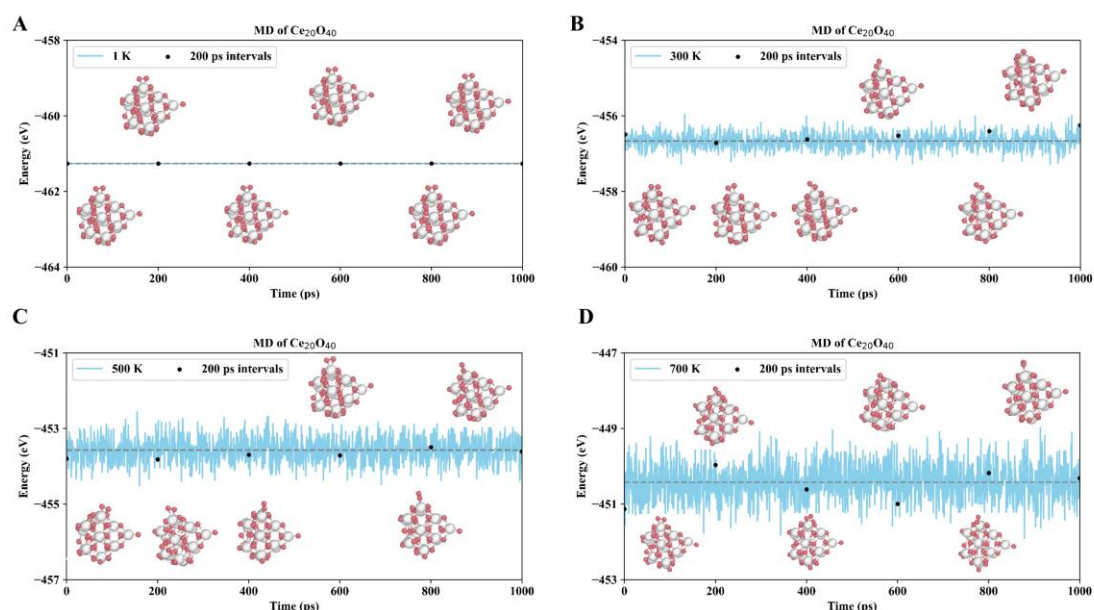
Supplementary Figure 16. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{17}\text{O}_{34}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



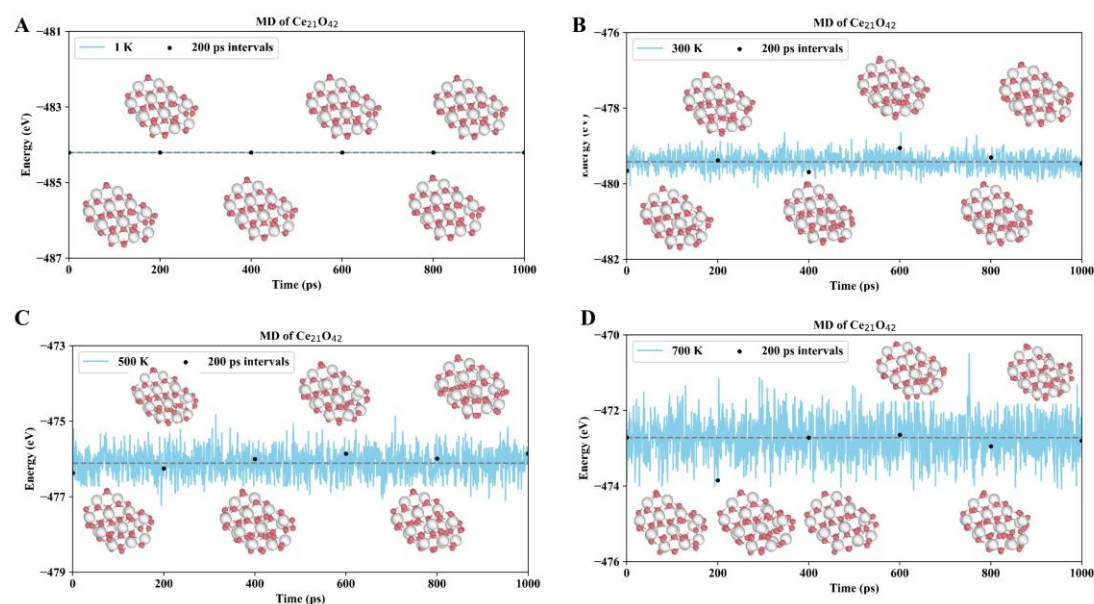
Supplementary Figure 17. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{18}\text{O}_{36}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.



Supplementary Figure 18. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{19}\text{O}_{38}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.

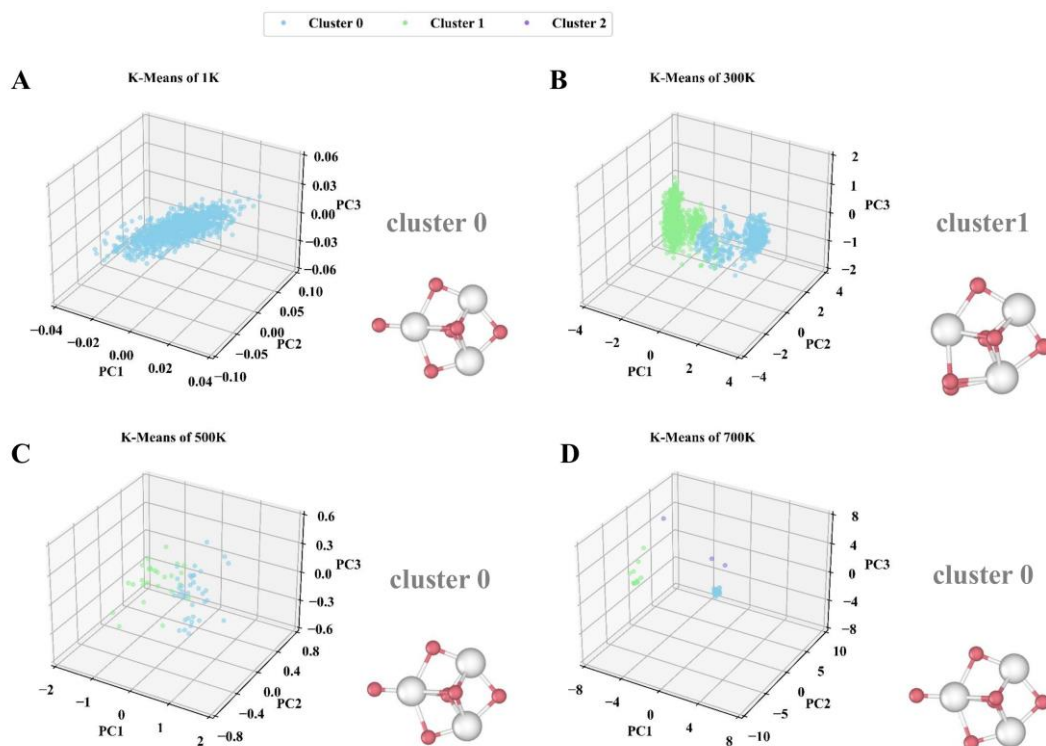


Supplementary Figure 19. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{20}\text{O}_{40}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.

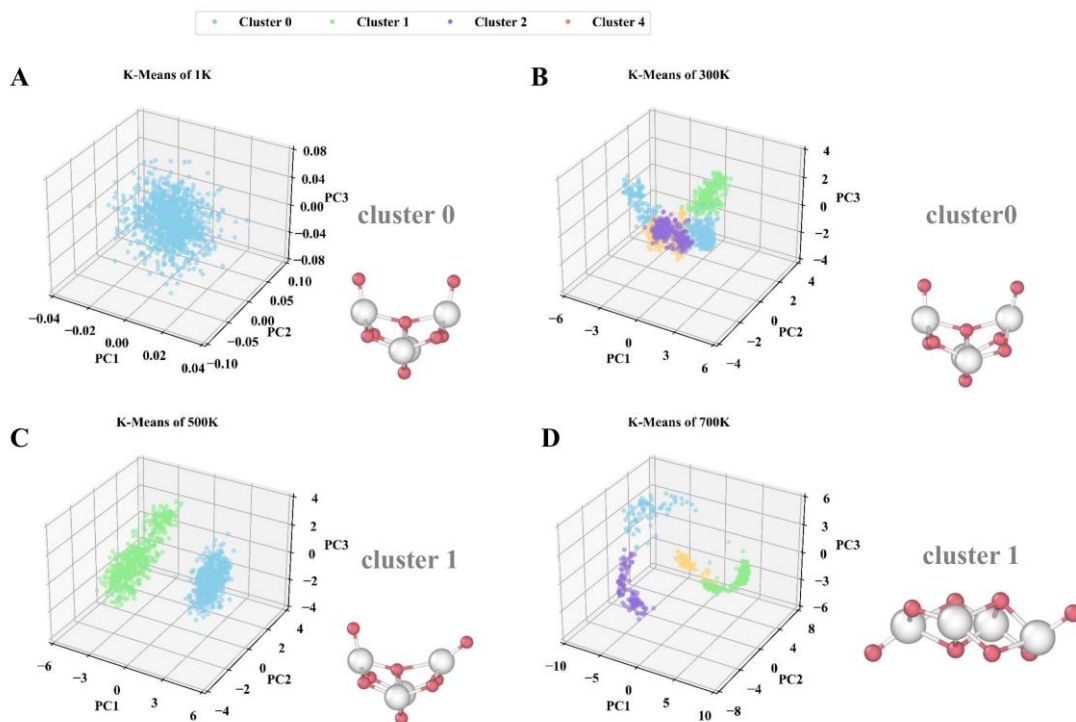


Supplementary Figure 20. Molecular dynamics (MD) simulations of the most stable cerium oxide nanocluster $\text{Ce}_{21}\text{O}_{42}$ at different temperatures, including energies during the simulations and snapshots taken every 200 ps. A: 1 K; B: 300 K; C: 500 K; D: 700 K. Color map: white, Ce atoms; pink, O atoms.

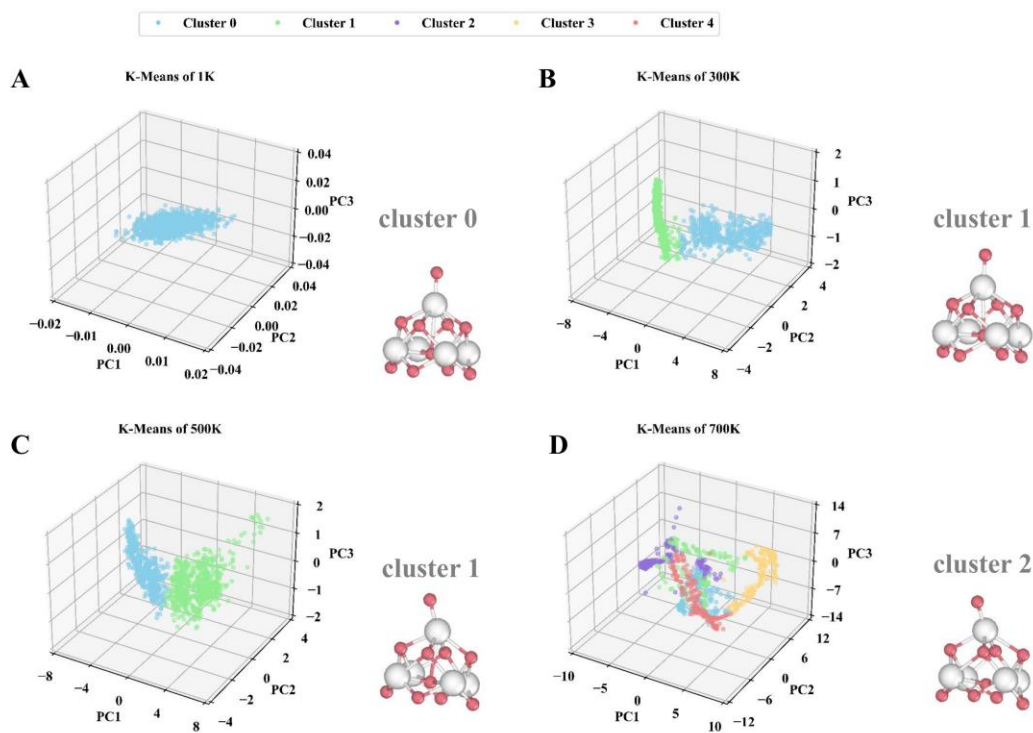
4. The NNPM analysis using unsupervised machine learning algorithms



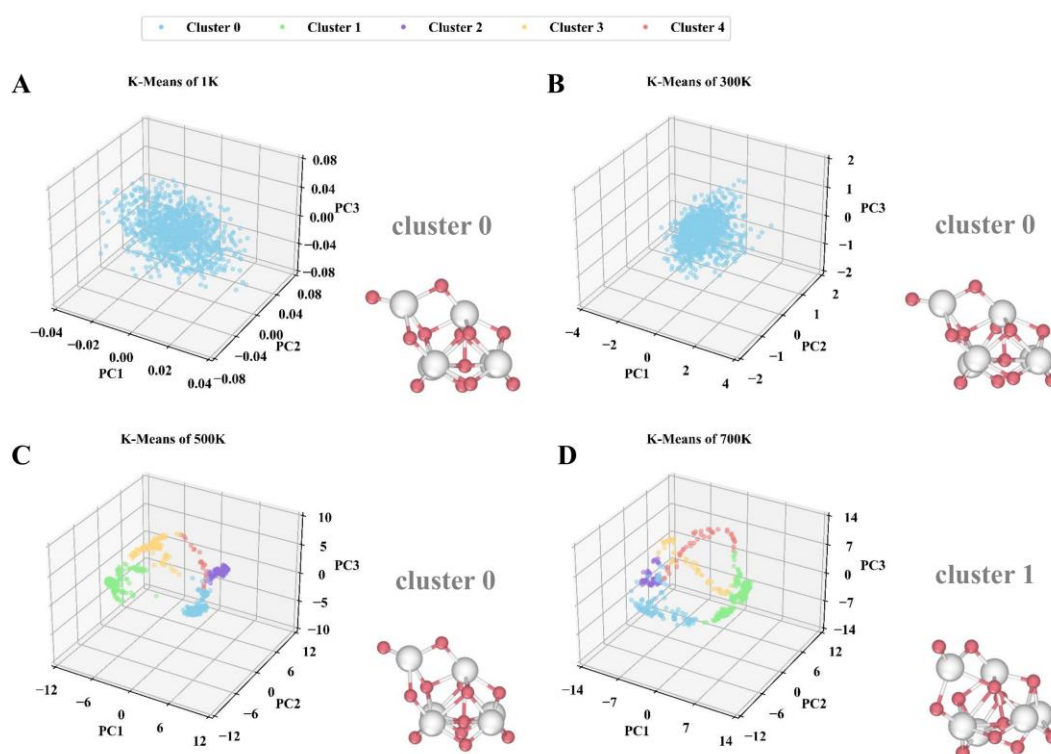
Supplementary Figure 21. Visualization of clustered nanocluster Ce_3O_6 at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms.



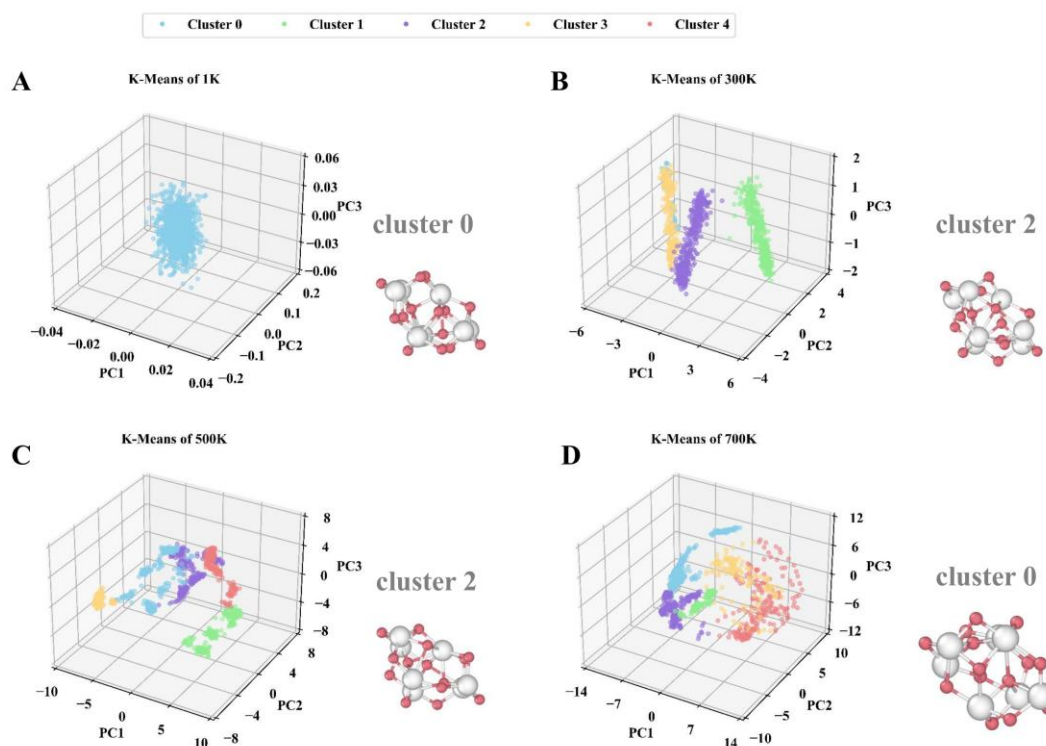
Supplementary Figure 22. Visualization of clustered nanocluster Ce_4O_8 at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



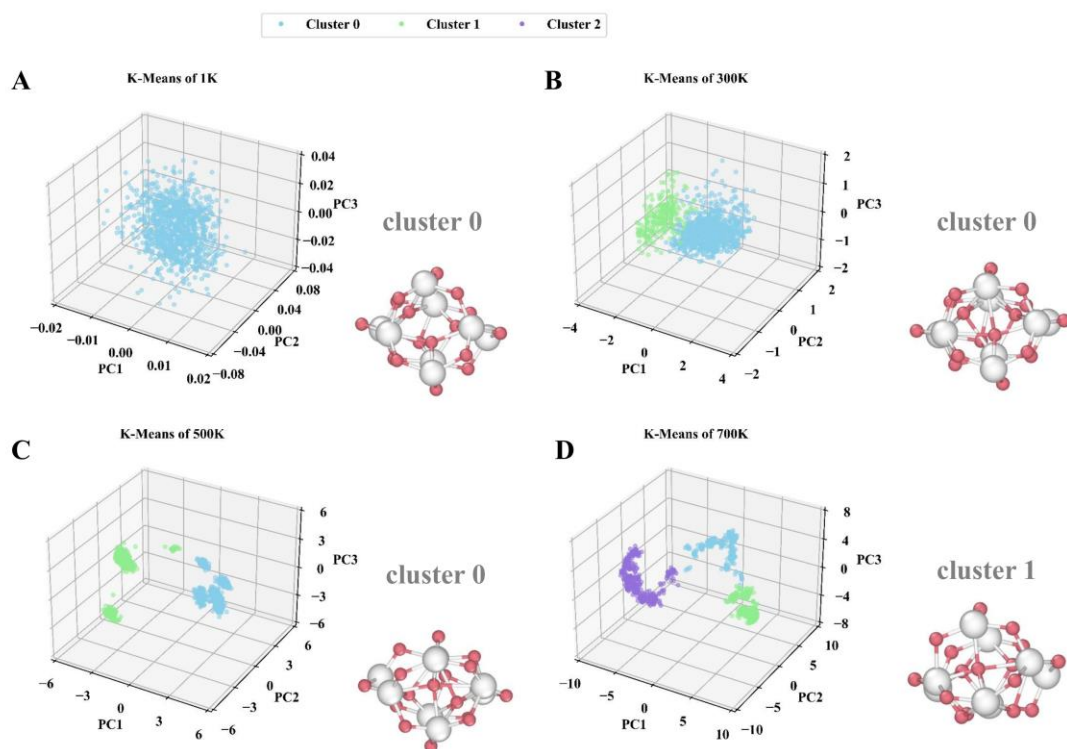
Supplementary Figure 23. Visualization of clustered nanocluster Ce_5O_{10} at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



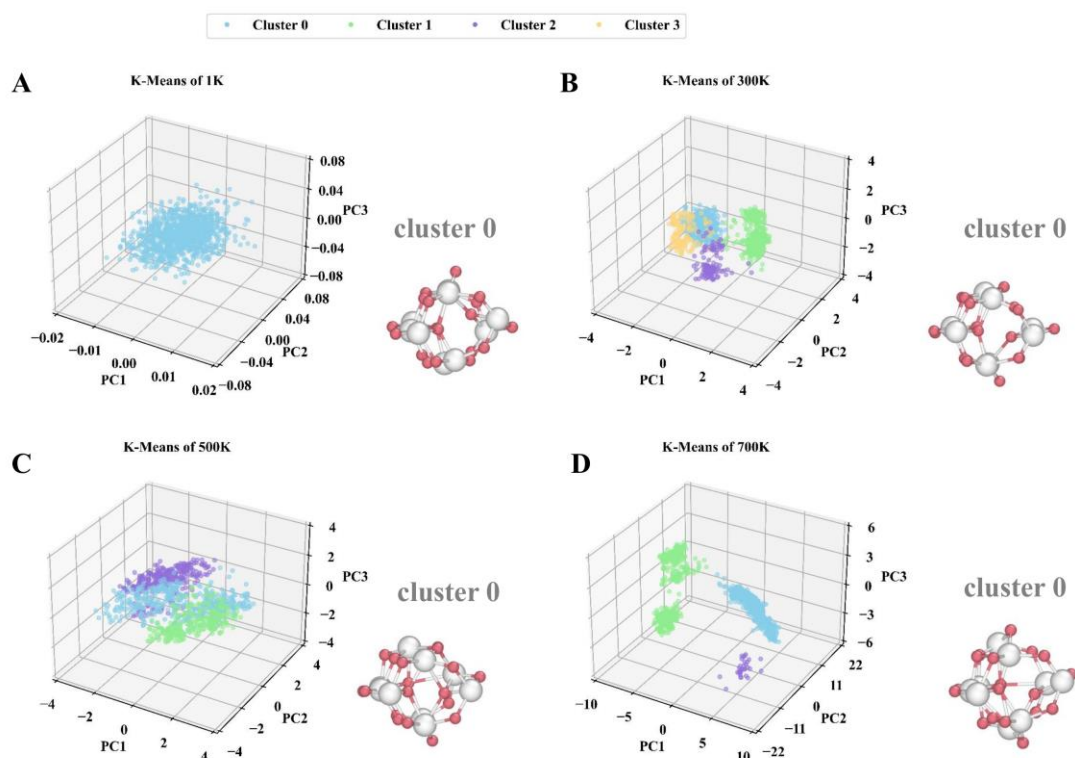
Supplementary Figure 24. Visualization of clustered nanocluster Ce_6O_{12} at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



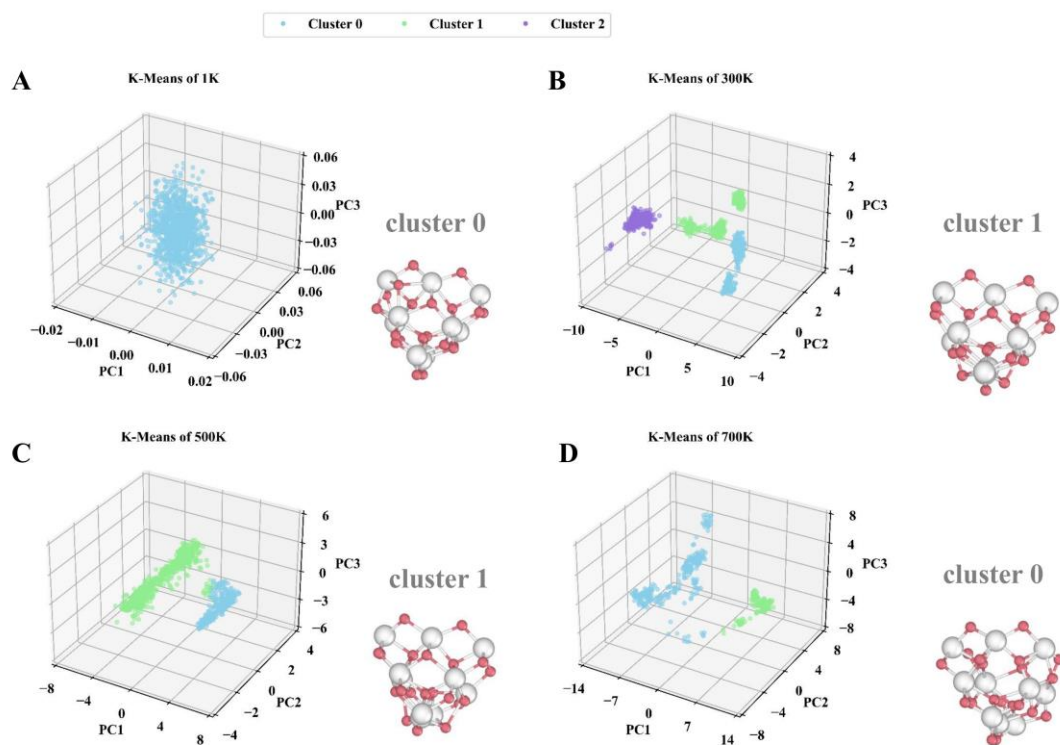
Supplementary Figure 25. Visualization of clustered nanocluster Ce_7O_{14} at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



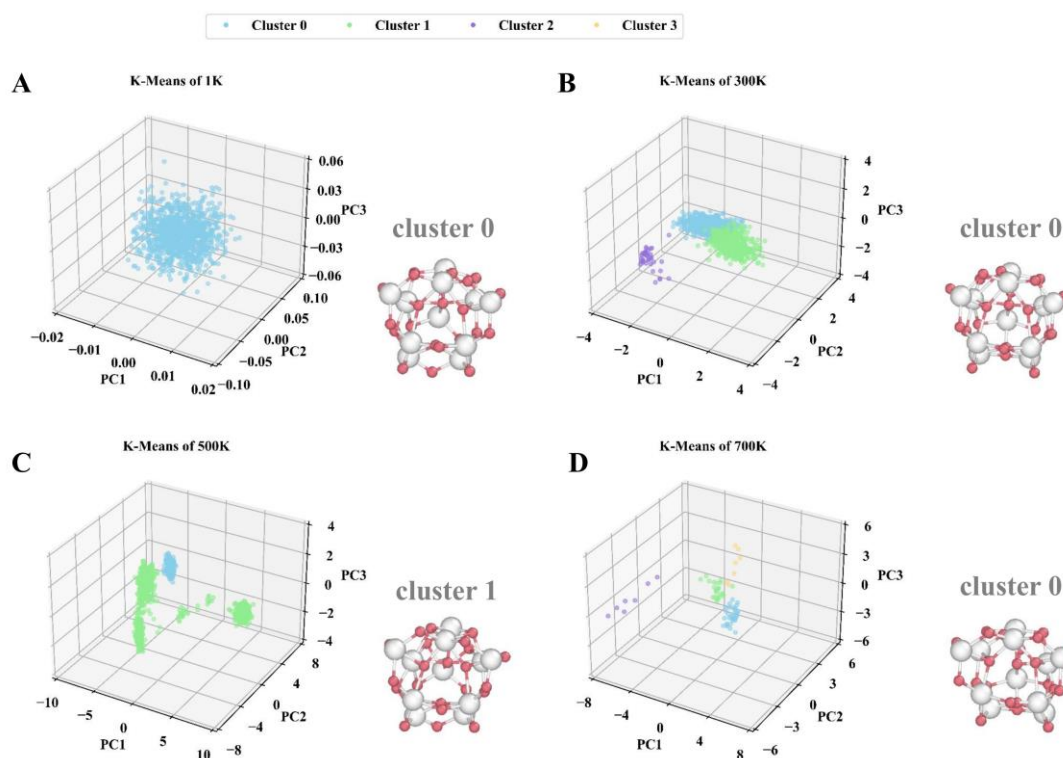
Supplementary Figure 26. Visualization of clustered nanocluster Ce_8O_{16} at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



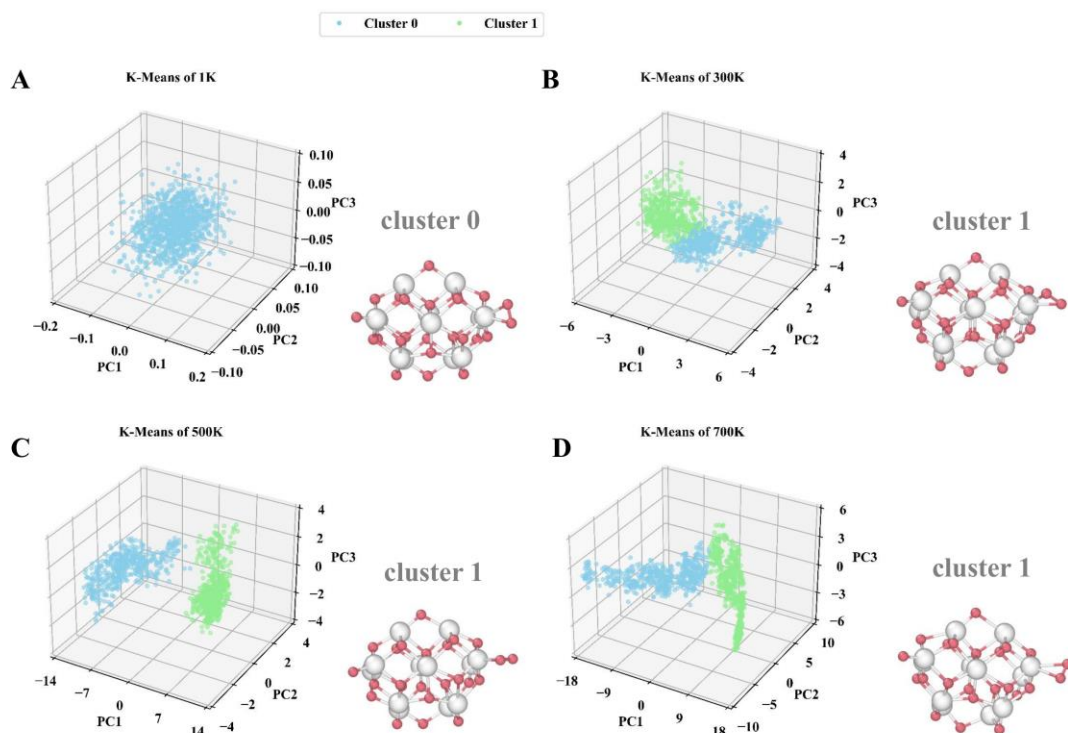
Supplementary Figure 27. Visualization of clustered nanocluster Ce_9O_{18} at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



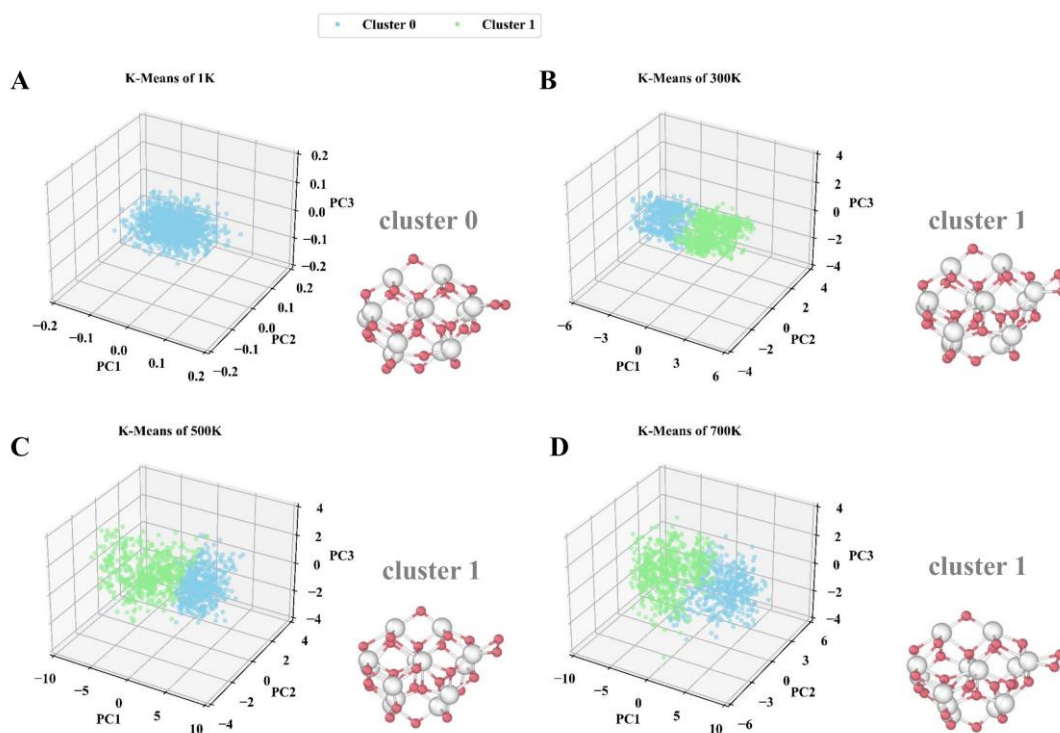
Supplementary Figure 28. Visualization of clustered nanocluster $\text{Ce}_{10}\text{O}_{20}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



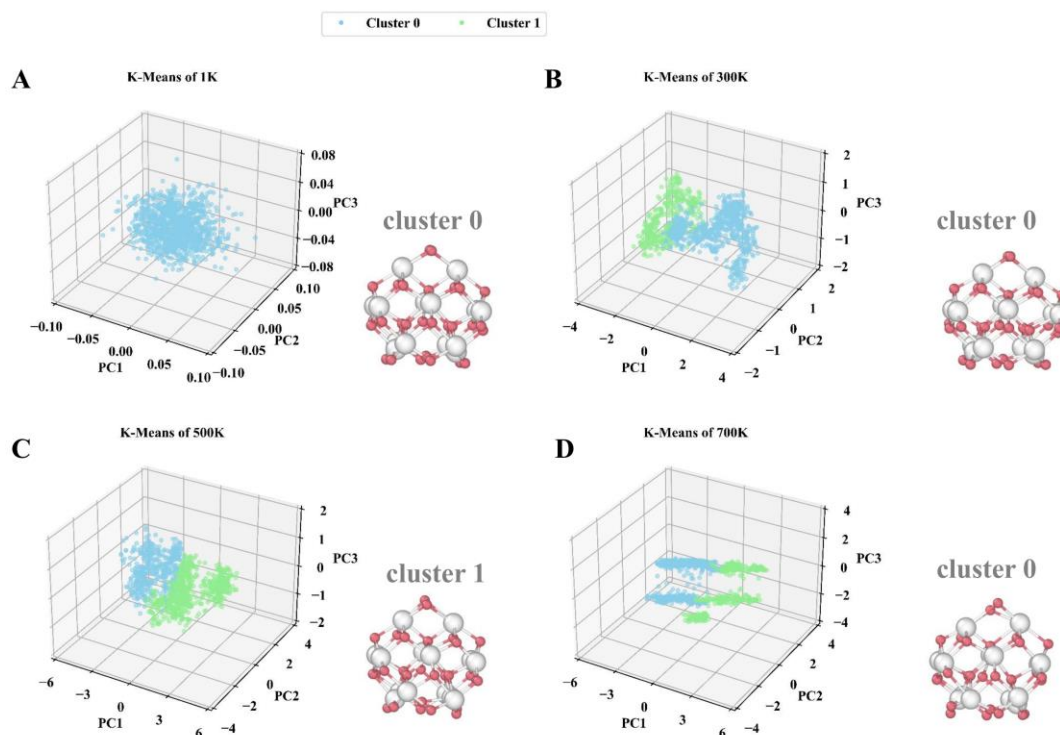
Supplementary Figure 29. Visualization of clustered nanocluster $\text{Ce}_{11}\text{O}_{22}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



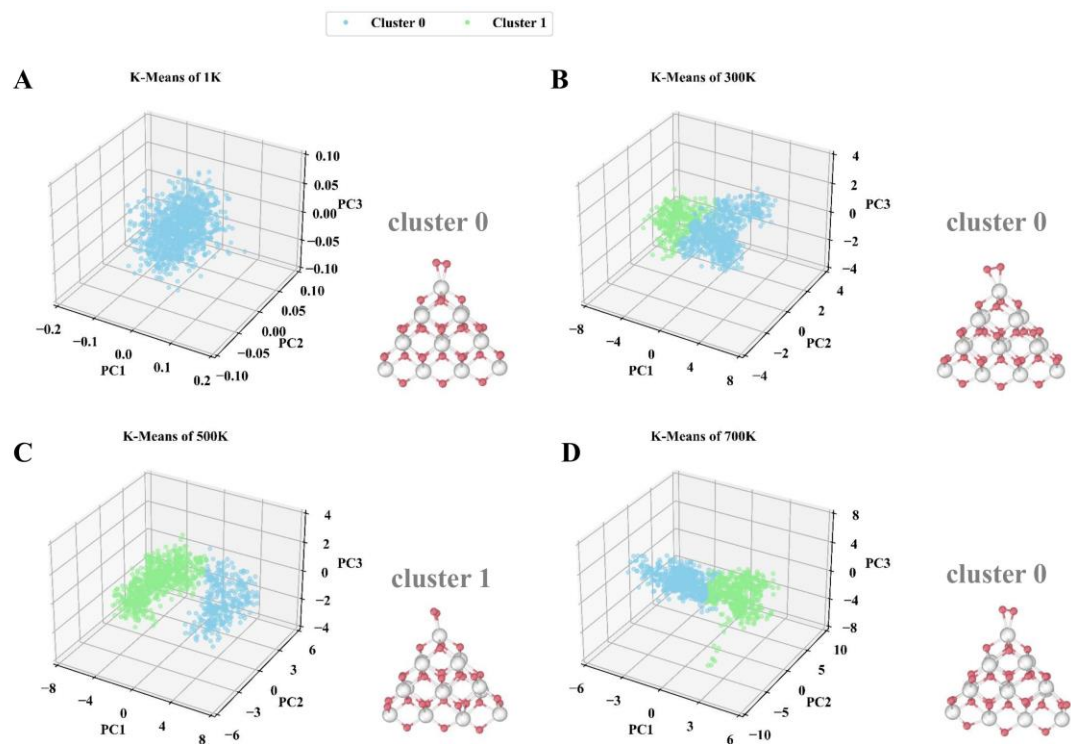
Supplementary Figure 30. Visualization of clustered nanocluster $\text{Ce}_{12}\text{O}_{24}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



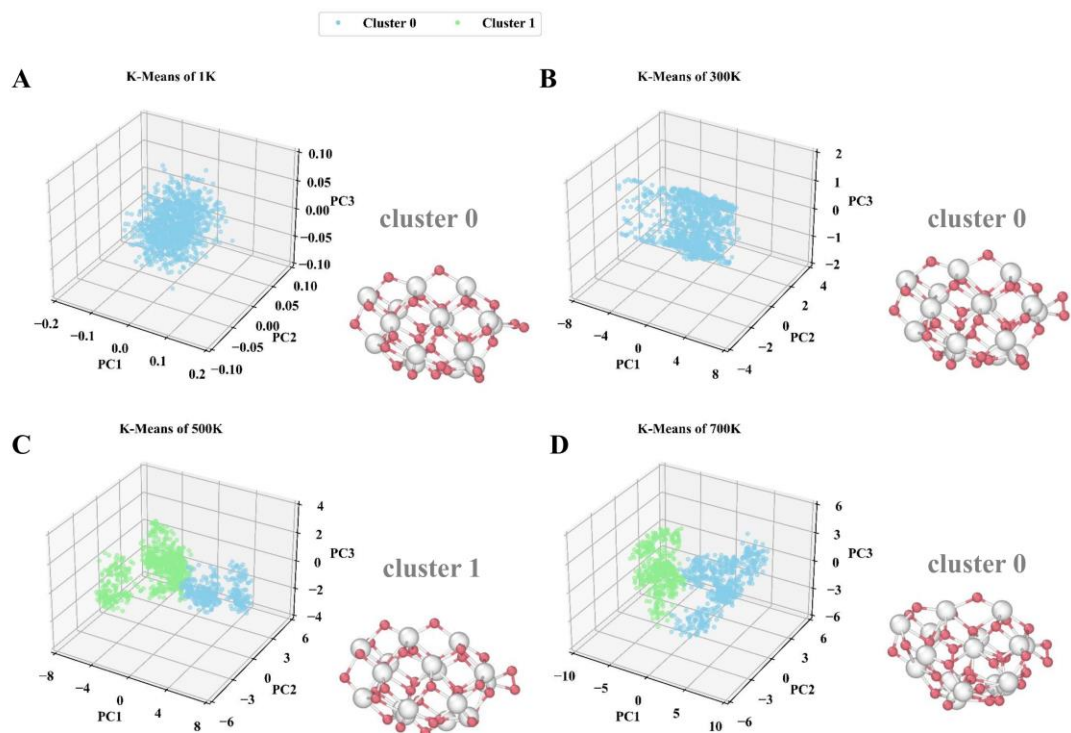
Supplementary Figure 31. Visualization of clustered nanocluster $\text{Ce}_{13}\text{O}_{26}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



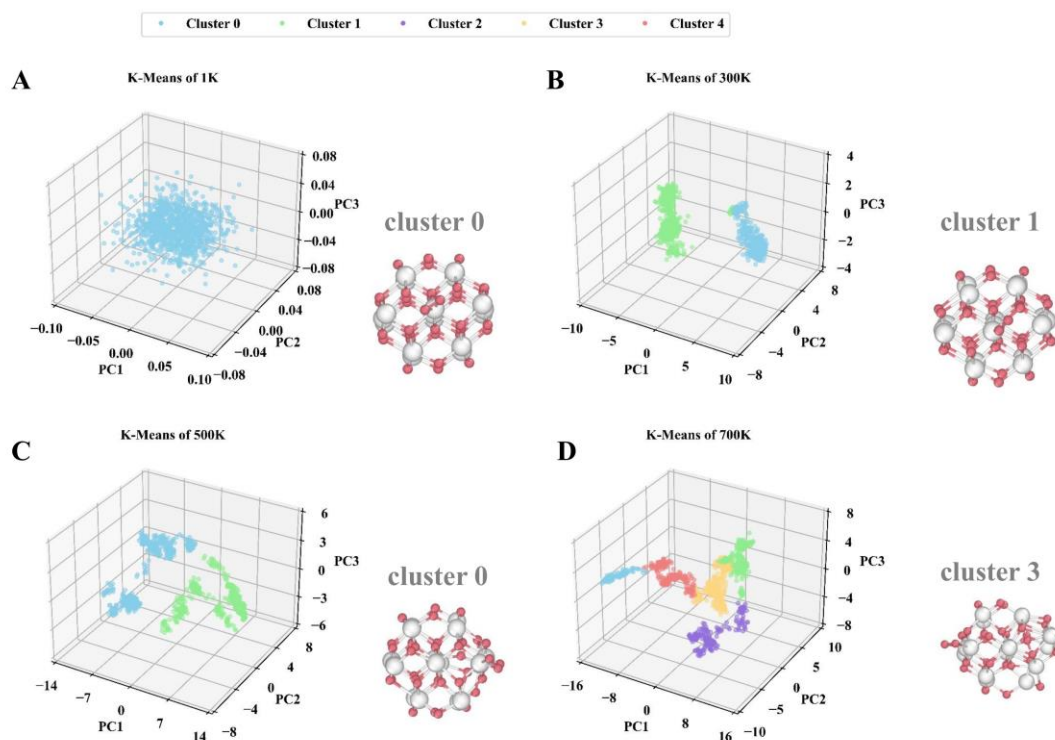
Supplementary Figure 32. Visualization of clustered nanocluster $\text{Ce}_{14}\text{O}_{28}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



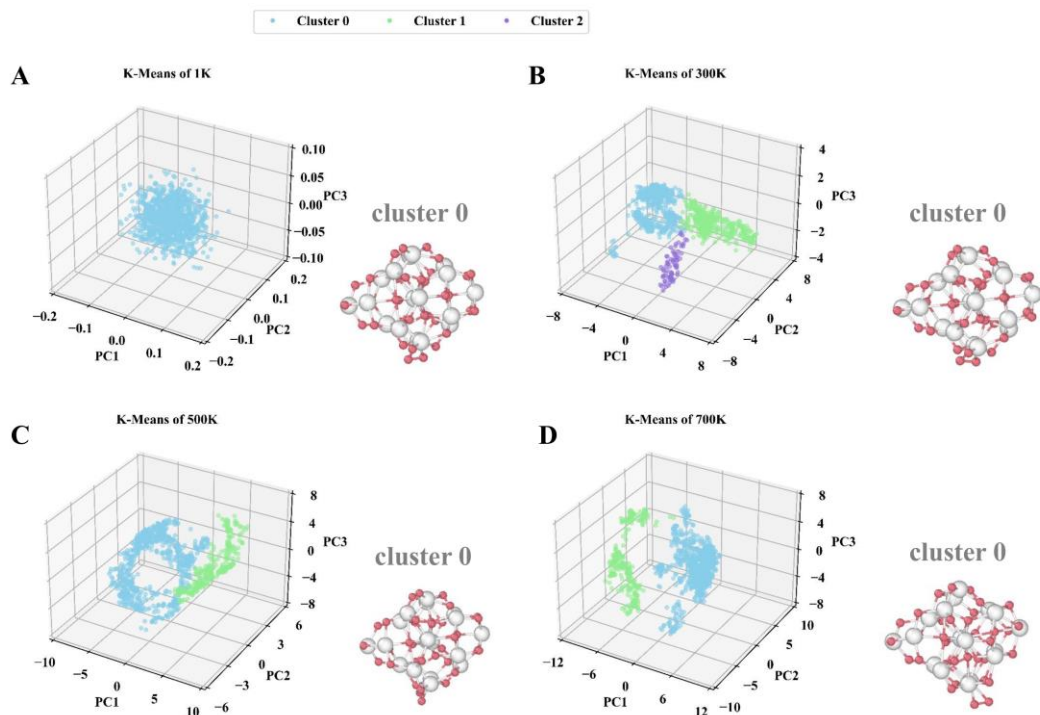
Supplementary Figure 33. Visualization of clustered nanocluster $\text{Ce}_{15}\text{O}_{30}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



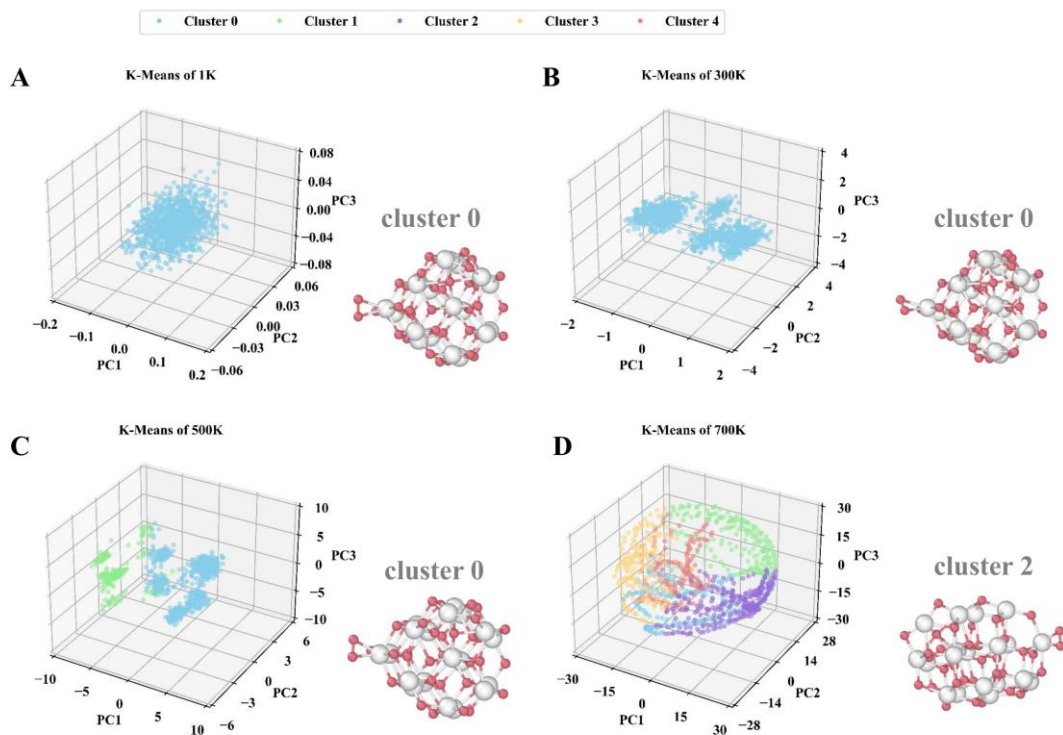
Supplementary Figure 34. Visualization of clustered nanocluster $\text{Ce}_{16}\text{O}_{32}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



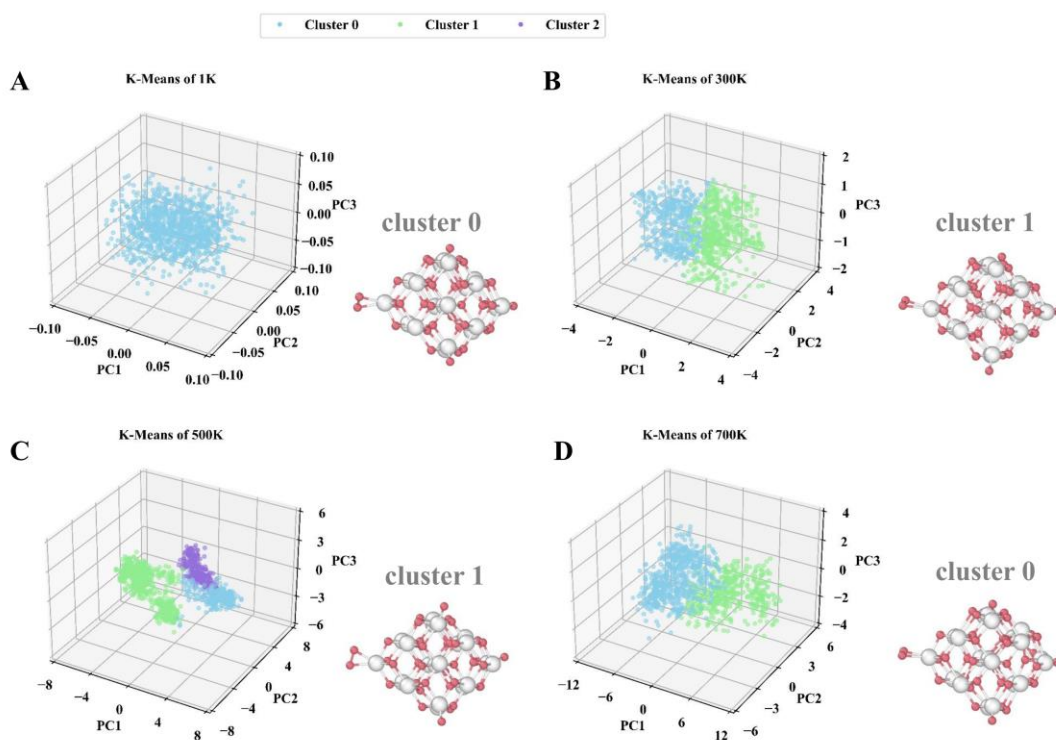
Supplementary Figure 35. Visualization of clustered nanocluster $\text{Ce}_{17}\text{O}_{34}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



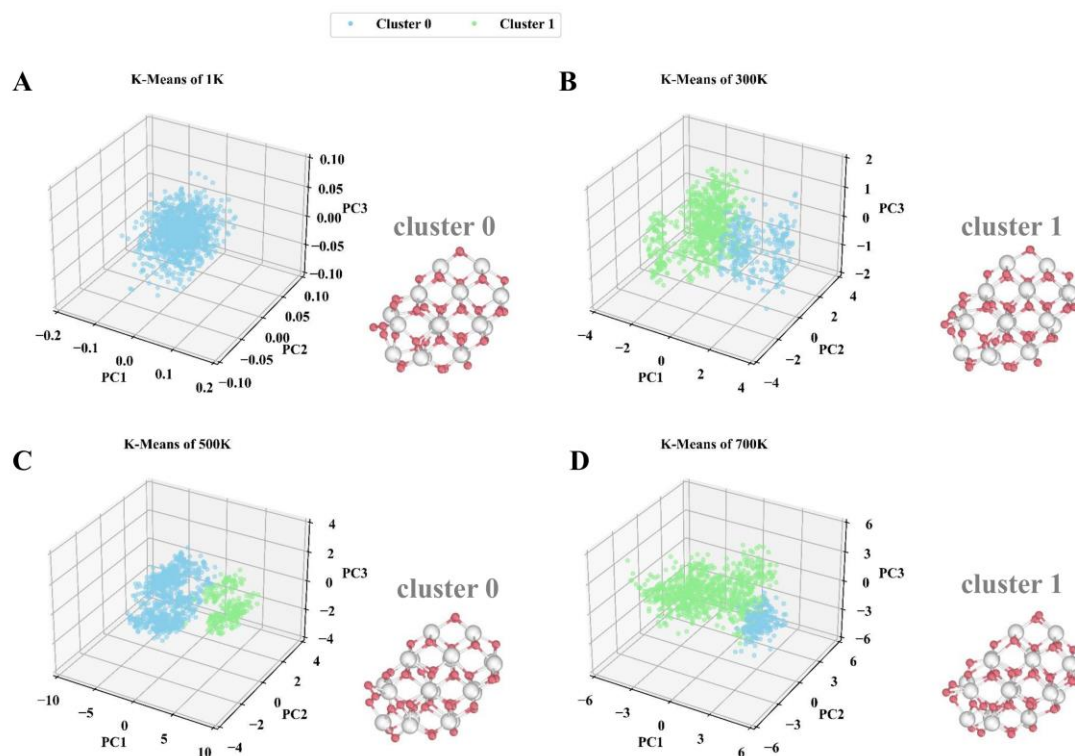
Supplementary Figure 36. Visualization of clustered nanocluster $\text{Ce}_{18}\text{O}_{36}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



Supplementary Figure 37. Visualization of clustered nanocluster $\text{Ce}_{19}\text{O}_{38}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms.



Supplementary Figure 38. Visualization of clustered nanocluster $\text{Ce}_{20}\text{O}_{40}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms



Supplementary Figure 39. Visualization of clustered nanocluster $\text{Ce}_{21}\text{O}_{42}$ at different temperatures via PCA dimensionality reduction, along with the representative structure of the largest cluster. A, B, C, and D correspond to 0 K, 300 K, 500 K, and 700 K, respectively. Color map: white, Ce atoms; pink, O atoms

Supplementary Table 3. Average energy, RMSD, and number of each largest cluster after clustering

	1 K				300 K			
	Avg Etot	Avg Ep	Avg RMSD	Num	Avg Etot	Avg Ep	Avg RMSD	Num
Ce3O6	-64.18	-64.18	0.087	1001	-64.079	-64.129	1.71	553
Ce4O8	-87.026	-87.026	0.033	1001	-86.899	-86.963	0.352	328
Ce5O10	-110.42	-110.421	0.054	1001	-110.231	-110.326	0.252	509
Ce6O12	-133.289	-133.289	0.036	1001	-133.1	-133.195	0.328	718
Ce7O14	-156.521	-156.522	0.043	1001	-156.328	-156.436	0.47	397
Ce8O16	-179.914	-179.915	0.088	1001	-179.711	-179.836	0.354	786
Ce9O18	-203.065	-203.065	0.043	1001	-202.769	-202.925	0.557	391

Ce10O20	-226.348	-226.349	0.028	1001	-226.095	-226.222	0.785	290
Ce11O22	-249.096	-249.097	0.05	1001	-248.768	-248.933	0.445	553
Ce12O24	-272.996	-272.997	0.08	1001	-272.528	-272.76	0.57	509
Ce13O26	-296.215	-296.216	0.039	1001	-295.736	-295.98	0.453	579
Ce14O28	-319.956	-319.957	0.05	1001	-319.522	-319.742	0.221	635
Ce15O30	-343.78	-343.781	0.032	1001	-343.267	-343.524	0.708	652
Ce16O32	-366.148	-366.149	0.047	1001	-365.519	-365.833	0.333	545
Ce17O34	-388.843	-388.844	0.142	1001	-388.336	-388.589	0.359	607
Ce18O36	-412.861	-412.862	0.078	1001	-412.129	-412.496	0.344	534
Ce19O38	-436.686	-436.688	0.03	1001	-436.051	-436.369	0.231	449
Ce20O40	-461.273	-461.275	0.088	1001	-460.63	-460.952	0.451	508
Ce21O41	-484.217	-484.219	0.041	1001	-483.538	-483.878	0.319	721

	500 K				700 K			
	Avg Etot	Avg Ep	Avg RMSD	Num	Avg Etot	Avg Ep	Avg RMSD	Num
Ce3O6	-64.178	-64.179	0.138	37	-64.176	-64.178	0.229	76
Ce4O8	-86.794	-86.911	0.694	532	-86.77	-86.899	2.128	188
Ce5O10	-110.1	-110.259	0.664	552	-110.071	-110.245	0.832	381
Ce6O12	-132.987	-133.138	0.615	481	-132.946	-133.116	1.19	244
Ce7O14	-156.198	-156.37	0.821	251	-155.988	-156.264	1.791	306
Ce8O16	-179.578	-179.769	0.48	602	-179.233	-179.597	1.577	385
Ce9O18	-202.573	-202.825	0.721	451	-202.293	-202.681	0.669	576
Ce10O20	-225.725	-226.037	1.195	707	-225.768	-226.058	2.357	270
Ce11O22	-248.453	-248.775	0.634	474	-248.523	-248.815	1.156	32
Ce12O24	-272.232	-272.613	0.459	566	-271.917	-272.467	2.193	507
Ce13O26	-295.383	-295.801	0.475	532	-295.281	-295.855	0.568	544
Ce14O28	-319.066	-319.514	0.459	569	-319.057	-319.548	0.333	664
Ce15O30	-342.827	-343.305	0.569	671	-342.629	-343.206	0.569	551

Ce16O32	-365.209	-365.681	0.437	660	-365.034	-365.592	0.922	540
Ce17O34	-388.037	-388.439	1.147	565	-387.576	-388.257	2.229	282
Ce18O36	-411.862	-412.359	0.853	427	-411.299	-412.081	0.87	763
Ce19O38	-435.761	-436.274	0.402	770	-435.303	-436.042	3.288	291
Ce20O40	-460.339	-460.822	0.452	539	-459.786	-460.643	0.581	587
Ce21O41	-483.103	-483.66	0.45	779	-482.474	-483.36	0.534	819

References

1. Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters*. **2007**, *98*, 146401. DOI: 10.1103/PhysRevLett.98.146401
2. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics*. **2011**, *134*, 074106. DOI: 10.1063/1.3553717
3. Singraber, A.; Morawietz, T.; Behler, J.; et al. Parallel Multistream Training of High-Dimensional Neural Network Potentials. *Journal of Chemical Theory and Computation*. **2019**, *15*, 3075-3092. DOI: 10.1021/acs.jctc.8b01092