

## Supporting Information

### **An Effective Structural Descriptor to Quantify the Reactivity of Lattice Oxygen in CeO<sub>2</sub> Subnano-clusters**

Chuan Zhou,<sup>a</sup> Binghu Zhang,<sup>b</sup> P. Hu<sup>ac</sup>, Haifeng Wang<sup>a\*</sup>

<sup>a</sup>Key Laboratory for Advanced Materials, Research Institute of Industrial Catalysis and Centre for Computational Chemistry, East China University of Science and Technology, Shanghai 200237, China

<sup>b</sup>Research Institute of Applied Catalysis, School of Chemical and Environmental Engineering, Shanghai Institute of Technology, Shanghai 201418, China

<sup>c</sup>School of Chemistry and Chemical Engineering, The Queen's University of Belfast, Belfast BT9 5AG, U.K.

## S1 Computational details

In our work, we invoked the evolutionary algorithm within the Universal Structure Predictor Evolutionary Xtallography (USPEX) code to search for stable  $Ce_nO_{2n}$  ( $n=1-10$ ) clusters and optimize structures, by virtue of General Utility Lattice Program (GULP) and Vienna Ab Initio Simulation Package (VASP) based on density functional theory (DFT) together. The first generation of structures were produced using randomly selected point group symmetries and then subsequently relaxed and ranked by the total energy. All the new structures were produced from the old ones by heredity (50%), softmutation (20%), permutation (10%), and 20% of the new generation using the random symmetric algorithm with the target to find the global minimum energy configuration for each composition. In these calculations, we applied periodic boundary conditions with vacuum region of 12 Å around each  $Ce_nO_{2n}$  cluster. For the located optimized configuration for each composition, the molecular dynamics simulations (NVT,  $T=500K$ ,  $\sim 10$  ps) were also carried out to check its relative stability.

Subsequently, the structures were roughly optimized with Lewis force field<sup>1</sup> and conjugate gradient method in GULP. Herein, the Lewis force field includes Ce and O parameters, which was developed previously. Then each  $Ce_nO_{2n}$  cluster was placed in the center of a  $20 \times 20 \times 20$  Å cubic cell for the next step.

All structural relaxations and total energy calculations were again accurately optimized on VASP calculations based on DFT+U ( $U_{\text{eff}}=U-J=5$  eV)<sup>2-3</sup> method to obtain more precise results. The exchange-correlation functional was treated by the GGA-PW91. The project-augmented wave (PAW) method was applied to represent the core-valence electron interaction, and the valence electronic states were expanded in plane wave basis sets with energy cutoff at 400 eV. In both cases,  $1 \times 1 \times 1$  k-point mesh was used for Brillouin-zone integration. The force threshold for the optimization was 0.05 eV/Å.

The adsorption energies were calculated by using the following equation:

$$E_{\text{ads}} = E_{\text{H/Ce}_n\text{O}_{2n}} - \frac{1}{2}E_{\text{H}_2} - E_{\text{Ce}_n\text{O}_{2n}} \quad (\text{Eq-S1})$$

where  $E_{\text{H/Ce}_n\text{O}_{2n}}$  is the total energy of the interacting H/ $Ce_nO_{2n}$  system,  $E_{\text{H}_2}$  and  $E_{\text{Ce}_n\text{O}_{2n}}$  are energies of the gas-phase  $H_2$  molecule and the  $Ce_nO_{2n}$  cluster, respectively. With this definition, a negative value of adsorption energy suggests that the adsorption is stronger, while a positive value stands for a weaker bonding.

## S2 The bond order conservation-Morse potential(BOC-MP) method

The BOC-MP method<sup>4-6</sup> was based on three postulates. (i) In a many-body system, all forces are spherical in that they depend on distance only. (ii) Each two-body A-B interaction is described by a Morse potential(MP):

$$E(x(r)) = a(x^2(r) - 2x(r)) \quad (\text{Eq-S2})$$

where the variable  $x(r)$ , called an A-B bond order(BO), depends on the bond distance,  $r$ , as

$$x(r) = \exp\{-(r - r_0) / b\} \quad (\text{Eq-S3})$$


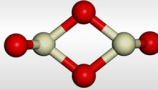
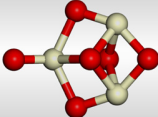
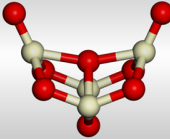
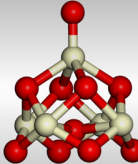
The ‘ $a$ ’ parameter is the A-B bond energy at the equilibrium distance,  $r_0$ , when the equilibrium bond order,  $x_0$ , is unity by definition. The ‘ $b$ ’ parameter is a distance scaling constant. (iii) In a many-body system, the total bond order,  $X$ , of all interacting two-center bonds is conserved at unity:

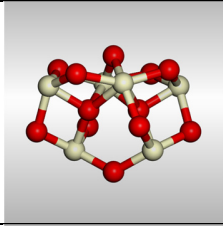
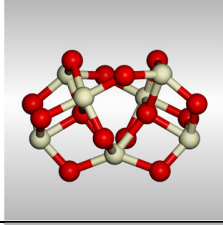
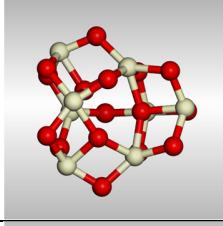
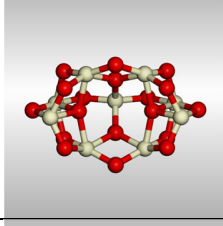
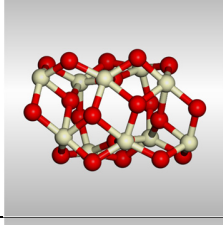
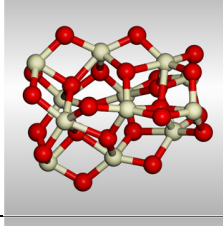
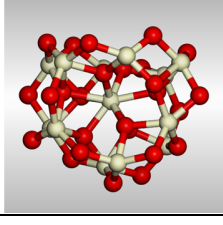
$$X = \sum_i x_i = 1 \quad (\text{Eq-S4})$$

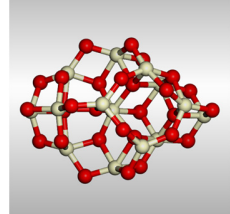
### S3 H-adsorption energies on $Ce_nO_{2n}$ subnano-clusters

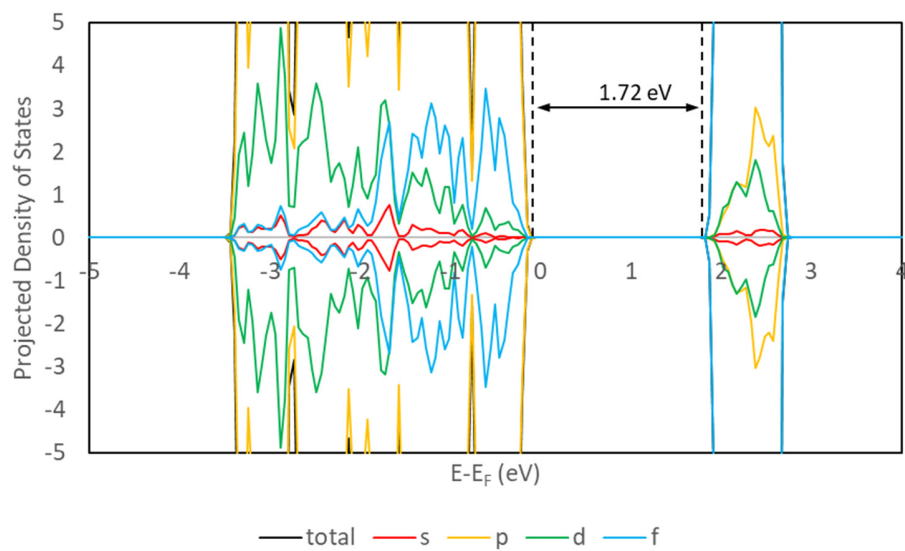
The most stable configurations of a series of  $Ce_nO_{2n}$  ( $n=1-10, 12-14$ ) subnano-clusters using genetic-algorithm-based global optimization method at the DFT level can be seen in Table S1, which includes also the optimized energy with VASP, point group symmetry and band gap of each cluster. Figure S1 shows the projected density of states of  $Ce_9O_{18}$  subnano-cluster for an example, giving a HOMO-LUMO gap of 1.72 eV.

**Table S1.** The optimized energy, point group symmetry and HOMO-LUMO gap of  $Ce_nO_{2n}$  ( $n=1-10, 12-14$ ) subnano-clusters.

Clusters	Structures	The optimized energy (eV)	Point group symmetry	HOMO-LUMO gap (eV)
$CeO_2_1$		-19.22	$C_{2h}$	2.57
$CeO_2_2$		-41.39	$C_{2h}$	2.12
$CeO_2_3$		-64.12	$C_s$	1.94
$CeO_2_4$		-87.21	$C_{2v}$	1.68
$CeO_2_5$		-110.55	$C_4$	1.20

CeO <sub>2</sub> _6		-133.56	C <sub>2</sub>	1.46
CeO <sub>2</sub> _7		-156.98	C <sub>2</sub>	1.82
CeO <sub>2</sub> _8		-180.17	C <sub>1</sub>	1.72
CeO <sub>2</sub> _9		-203.23	C <sub>2v</sub>	1.72
CeO <sub>2</sub> _10		-226.41	C <sub>2</sub>	1.76
CeO <sub>2</sub> _12		-273.72	C <sub>1</sub>	1.91
CeO <sub>2</sub> _13		-296.32	C <sub>1</sub>	1.67

CeO <sub>2</sub> _14		-320.06	C <sub>1</sub>	1.42
----------------------	---	---------	----------------	------

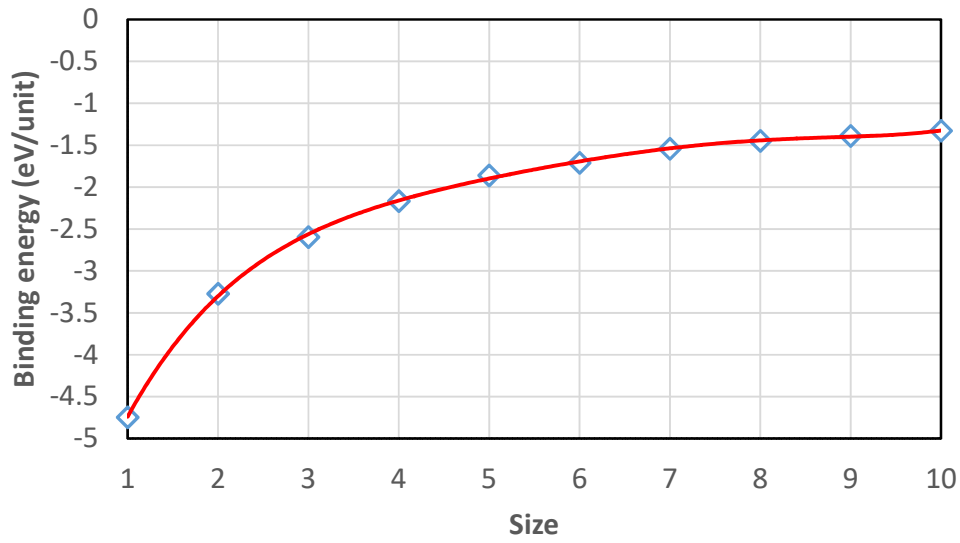


**Figure S1.** Projected DOS of Ce<sub>9</sub>O<sub>18</sub> subnano-cluster.

The binding energy of  $Ce_nO_{2n}$  ( $n=1-10$ ) is defined with the following equation, which refers to the bulk  $CeO_2$  unit cell:

$$E_{b,cluster} = (n \cdot E_{bulk}(CeO_2) - E_{cluster})/n \quad (\text{Eq-S5})$$

where  $E_{cluster}$  and  $E_{bulk}(CeO_2)$  are the total energy of cluster and bulk  $CeO_2$ , respectively. It is found that as the cluster size increases, the binding energies ascend rapidly and then increase slowly, indicating that the  $CeO_2$  clusters become more and more stable and may tend to the bulk  $CeO_2$ , as shown in Figure S2.



**Figure S2.** The variation of binding energy of  $Ce_nO_{2n}$  ( $n=1-10$ ) subnano-clusters with cluster size.

We selected 44 oxygen sites for H adsorption to elucidate the structure-activity relationship, among them there are 4 O<sub>1c</sub> sites, 36 O<sub>2c</sub> sites and 4 O<sub>3c</sub> sites, and the H adsorption energies span from -2.18 to -0.64 eV.

**Table S2.** 44 oxygen sites for H adsorption.

Clusters	sites	CN <sub>O</sub>	$\sum$ CN <sub>M</sub>	$E_{ads}(eV)$
CeO <sub>2</sub> _1	1	1	2	-1.30
CeO <sub>2</sub> _2	2	1	3	-1.72
	3	2	6	-0.68
CeO <sub>2</sub> _3	4	3	13	-0.77
	5	2	9	-1.66
	6	2	8	-1.36
	7	2	9	-1.66
CeO <sub>2</sub> _4	8	1	4	-1.93
	9	2	8	-1.32
	10	2	8	-0.99
CeO <sub>2</sub> _5	11	1	5	-2.18
	12	3	15	-1.35
	13	2	10	-1.51
CeO <sub>2</sub> _6	14	2	10	-1.59
	15	2	10	-1.40
	16	2	8	-1.06
	17	2	8	-1.11
	18	2	10	-1.54
CeO <sub>2</sub> _7	19	2	9	-1.25
	20	2	11	-1.55
	21	2	10	-1.47
	22	2	8	-1.14
	23	2	9	-1.33
CeO <sub>2</sub> _8	24	2	9	-1.31
	25	2	11	-1.52
	26	2	8	-1.07
	27	3	14	-0.68
	28	2	11	-1.53
	29	2	9	-1.34
	30	2	9	-1.38
	31	2	10	-1.45



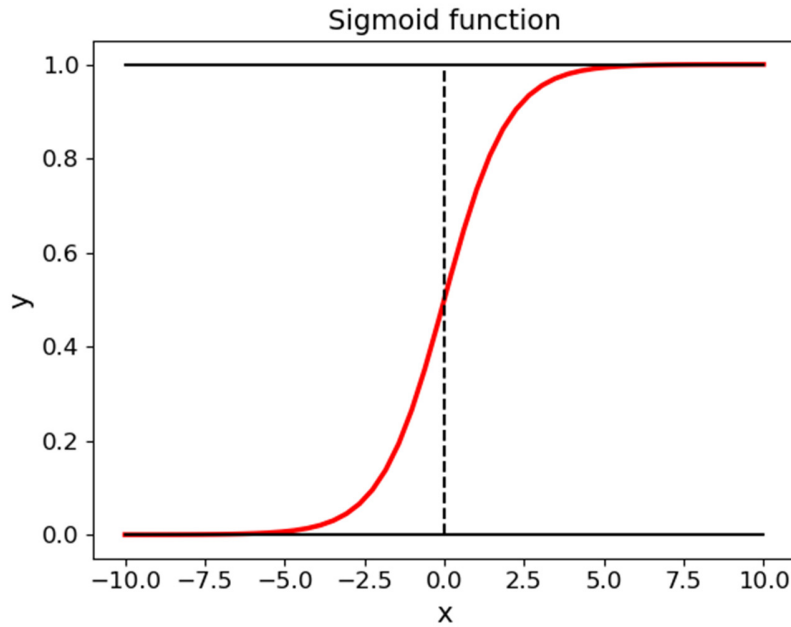
<b>Clusters</b>	<b>sites</b>	<b>CN<sub>O</sub></b>	<b><math>\sum</math>CN<sub>M</sub></b>	<b><math>E_{ads}(eV)</math></b>
	32	2	10	-1.47
	33	2	8	-1.08
	34	2	10	-1.42
CeO <sub>2</sub> _9	35	2	9	-1.22
	36	2	10	-1.42
	37	2	9	-1.36
	38	3	14	-0.64
	39	2	9	-1.21
CeO <sub>2</sub> _10	40	2	9	-1.34
	41	2	11	-1.66
	42	2	9	-1.33
	43	2	9	-1.18
	44	2	9	-1.25

#### S4 The characters of sigmoid function

The sigmoid function, also called logistic function, has been usually used for neural networks as activation function, which is originally defined as:

$$S(x) = \frac{1}{1 + e^{-x}} \quad (\text{Eq-S6})$$

Its function behavior can be illustrated in Figure S3. The two key characters of this function are: (i) it decays to 0 and 1 as  $x$  approaches to negative and positive infinity, respectively; (ii) near  $x=0$ ,  $S(x)$  changes greatly and monotonously, which reflect the variation trend of a typical bond strength around the equilibrium bond length. Inspired by these characters, we proposed to use the adjusted sigmoid function (see eqn (3) in main text) to calculate the effective coordination number for a specific bond.



**Figure S3.** The variation behavior of sigmoid function  $S(x)$ .

## S5 The effective coordination number approach of Hoppe

We firstly adopt the effective coordination number (ECN) approach of Hoppe et al.<sup>7-8</sup>, which takes into account the number of Ce around a given oxygen, where the individual bond distance  $R_i$  are weighted according to the average distance  $R_{av}$ :

$$\text{ECN} = \sum_i \exp\left(1 - \left(\frac{R_i}{R_{av}}\right)^6\right) \quad (\text{Eq-S7})$$

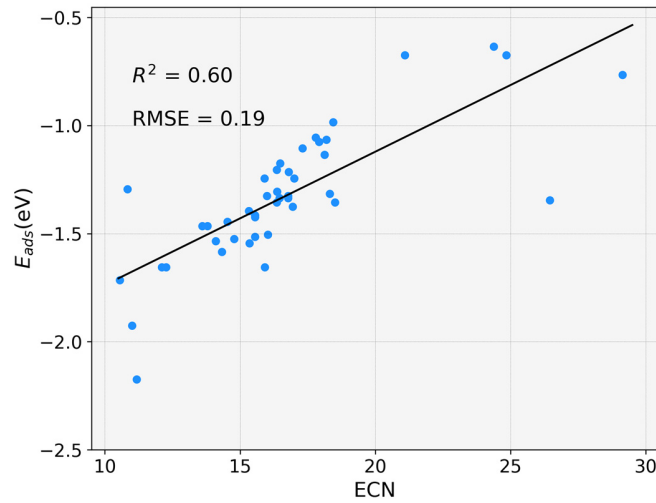
The average distance  $R_{av}$  is calculated with respect to the minimum Ce–O bond distances ( $R_{min}$ ):

$$R_{av} = \frac{\sum_i R_i \exp\left(1 - \left(\frac{R_i}{R_{min}}\right)^6\right)}{\sum_i \exp\left(1 - \left(\frac{R_i}{R_{min}}\right)^6\right)} \quad (\text{Eq-S8})$$

The relationship between ECN and adsorption energy ( $E_{ads}$ ) of adsorbate can be constructed as follows:

$$E_{ads} = a \cdot \text{ECN} + b \quad (\text{Eq-S9})$$

Then, we obtained the linear correlation with  $R^2=0.60$ , as shown in Figure S4.



**Figure S4.** Linear correlation between ECN of Hoppe and H-adsorption energy for  $\text{Ce}_n\text{O}_{2n}$  ( $n=1-10$ ) subnano-clusters, the correlation coefficient  $R^2$  is 0.60 and the RMSE is 0.19 eV.

## S6 Three schemes

The optimized parameters and linear correlations of the three schemes discussed in the main text are shown as follows:

**Table S3.** The parameters of three schemes. Note that the parameters of schemes(i) are same with ACN method.

Schemes	k	$\alpha_1$	$a$	$b$	$R^2$
i	0	7.869787	0.176064	-2.50056	0.80
ii	2.7	9.304059	0.217273	-4.07707	0.88
iii	0.8	7.806322	0.288074	-3.10924	0.94

## S7 the parameters of the quartic equation

We applied a quartic equation to describe the relation between  $E_{\text{ads}}$  and  $f(\text{SCN})$ , as follows:

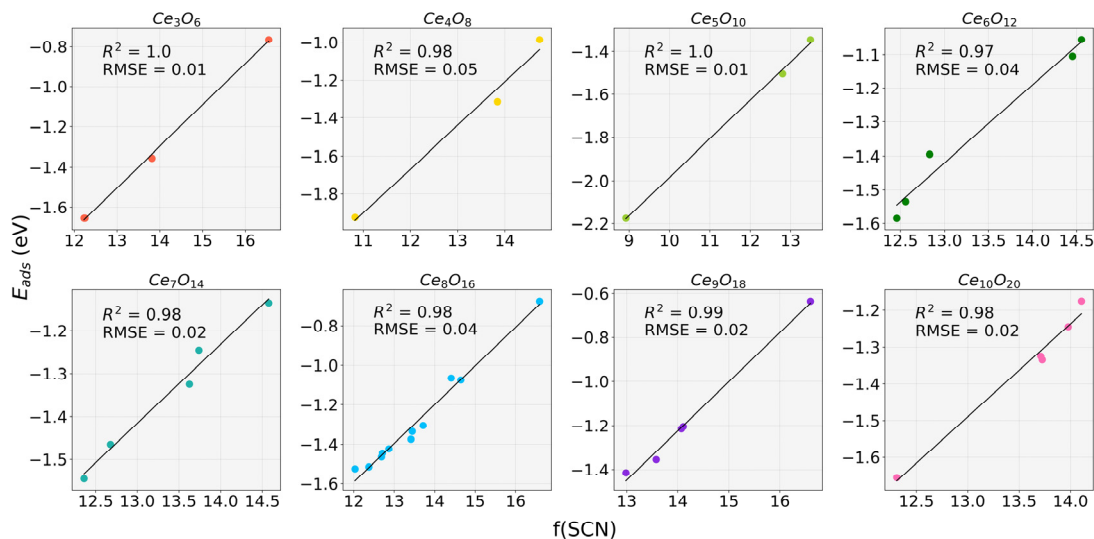
$$f(\text{SCN}) = \alpha_1 \cdot \text{SCN}_0 + \alpha_2 \cdot \text{SCN}_0^2 + \alpha_3 \cdot \text{SCN}_0^3 + \alpha_4 \cdot \text{SCN}_0^4 - (\beta_1 \cdot \sum \text{SCN}_M + \beta_2 \cdot (\sum \text{SCN}_M)^2 + \beta_3 \cdot (\sum \text{SCN}_M)^3 + \beta_4 \cdot (\sum \text{SCN}_M)^4) \quad (\text{Eq-S10})$$

$$E_{\text{ads}} = a \cdot f(\text{SCN}) + b \quad (\text{Eq-S11})$$

**Table S4.** The optimal parameters of the quartic equation.

$k$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$a$	$b$
0.800	11.275	2.477	-1.232	0.145	-1.760	0.859	-0.081	0.002	0.208	-4.137

## S8 Linear correlation on each $Ce_nO_{2n}$ ( $n=3-10$ ) subnano-cluster

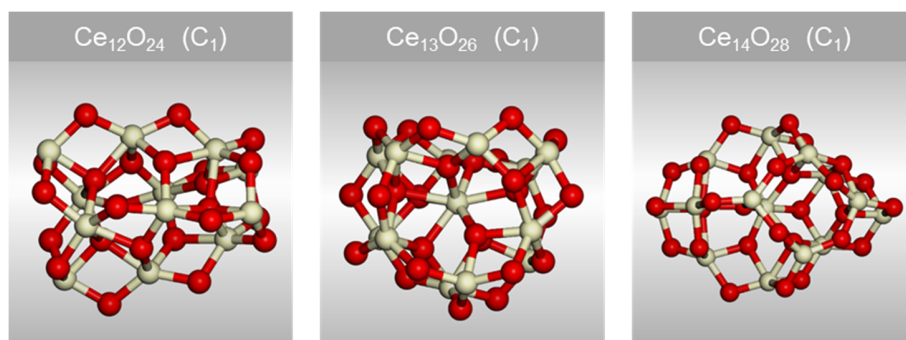


**Figure S5.** Linear correlation between  $f(\text{SCN})$  and H-adsorption energy on each  $Ce_nO_{2n}$  ( $n=3-10$ ) subnano-cluster. On average, the RMSE is 0.03 eV with a  $R^2$  value of 0.99.

We separately calculated  $f(\text{SCN})$  and  $E_{\text{ads}}$  in each  $Ce_nO_{2n}$  subnano-cluster except  $Ce_nO_{2n}$  ( $n=1-2$ ). It can be seen that for each structure, the linear relationship between the  $f(\text{SCN})$  and the H-adsorption energy is grossly strong. On average, the RMSE is 0.03 eV with an  $R^2$  value of 0.99.

## S9 Model prediction

The most stable  $Ce_nO_{2n}$  ( $n=12-14$ ) subnano-clusters were shown in Figure S6. We selected total 10 points in these structures, see Table S5.



**Figure S6.** The most stable  $Ce_nO_{2n}$  ( $n=12-14$ ) subnano-clusters. The value in parentheses represents point group. Red and white balls indicate the O and Ce, respectively.

**Table S5.** The H adsorption energies from DFT calculations compared with the predicted adsorption energies as given by eqn (Eq-S10)-(Eq-S11).

Clusters	sites	CN <sub>O</sub>	$\sum CN_M$	$E_{ads-DFT}(eV)$	$E_{ads-Predicted}(eV)$
CeO <sub>2</sub> _12	1	2	9	-1.26	-1.30
	2	2	10	-1.46	-1.45
	3	2	9	-1.25	-1.30
CeO <sub>2</sub> _13	4	2	9	-1.30	-1.30
	5	2	11	-1.59	-1.58
	6	2	10	-1.44	-1.45
	7	2	9	-1.29	-1.29
CeO <sub>2</sub> _14	8	2	10	-1.39	-1.40
	9	2	10	-1.54	-1.55
	10	2	9	-1.34	-1.35

## Reference

1. Lewis G V, C. C. R. A., Potential models for ionic oxides. *J. Phys. C Solid State Phys.* **1985**, *18* (6), 1149.
2. Wang, H. F.; Li, H. Y.; Gong, X. Q.; Guo, Y. L.; Lu, G. Z.; Hu, P., Oxygen vacancy formation in CeO<sub>2</sub> and Ce<sub>1-x</sub>Zr<sub>x</sub>O<sub>2</sub> solid solutions: electron localization, electrostatic potential and structural relaxation. *Phys. Chem. Chem. Phys.* **2012**, *14* (48), 16521-16535.
3. Wu, X. P.; Gong, X. Q., Unique Electronic and Structural Effects in Vanadia/Ceria-Catalyzed Reactions. *J. Am. Chem. Soc.* **2015**, *137* (41), 13228-31.
4. Sellers, E. S. H., The UBI-QEP method: a practical theoretical approach to understanding chemistry on transition metal surfaces. *Surf. Sci. Rep.* **1998**, *30* (1), 1-119.
5. Maestri, M.; Reuter, K., Semiempirical rate constants for complex chemical kinetics: first-principles assessment and rational refinement. *Angew. Chem. Int. Ed.* **2011**, *50* (5), 1194-7.
6. Shustorovich, E., Chemisorption Energetics and Surface Reactivity: UBI-QEP versus DFT Projections. *Russ. J. Phys. Chem. B* **2007**, *1* (4), 307-329.
7. Hoppe, R., et al., A new route to charge distributions in ionic solids. *J. Less Common Met.* **1989**, *156* (1), 105-122.
8. Woodley, S. M.; Catlow, R., Evolutionary structure prediction and electronic properties of indium oxide nanoclusters. *Phys. Chem. Chem. Phys.* **2010**, *12* (30), 8436-7.