Supporting Information for

Understanding the Zinc-Doped Hydroxyapatite Structures Using the First-Principles Calculations and Convolutional Neural Network Algorithm

Jing Wang[1], Ruihan Wang[1], Mingli Yang[2] and Dingguo Xu[1,2]*

[1]MOE Key Laboratory of Green Chemistry and Technology, College of Chemistry, Sichuan University, Chengdu, Sichuan, 610064, PR China

[2]Research Center for Material Genome Engineering, Sichuan University, Chengdu, Sichuan, 610065, PR China

* To whom correspondence should be addressed: dgxu@scu.edu.cn (D.X), Tel: 86-28-85406156.

1. **The number of structures for each doping concentration in the initial dataset**
2. **The number of the structures by the three substitution mechanisms in the lowest energy range.**
3. **The two leading principal components of the feature extraction vector learned by 5 Zn atoms doped HAP.**
4. **Detailed display of three Zn^{2+} doped structures.**
5. **Substitution mechanisms when two Zn are doped in the same Ca2-V channel.**
6. **The volumes between the most stable and the most unstable structures.**

Table S1. The number of structures for 5 doping concentrations in the initial dataset.

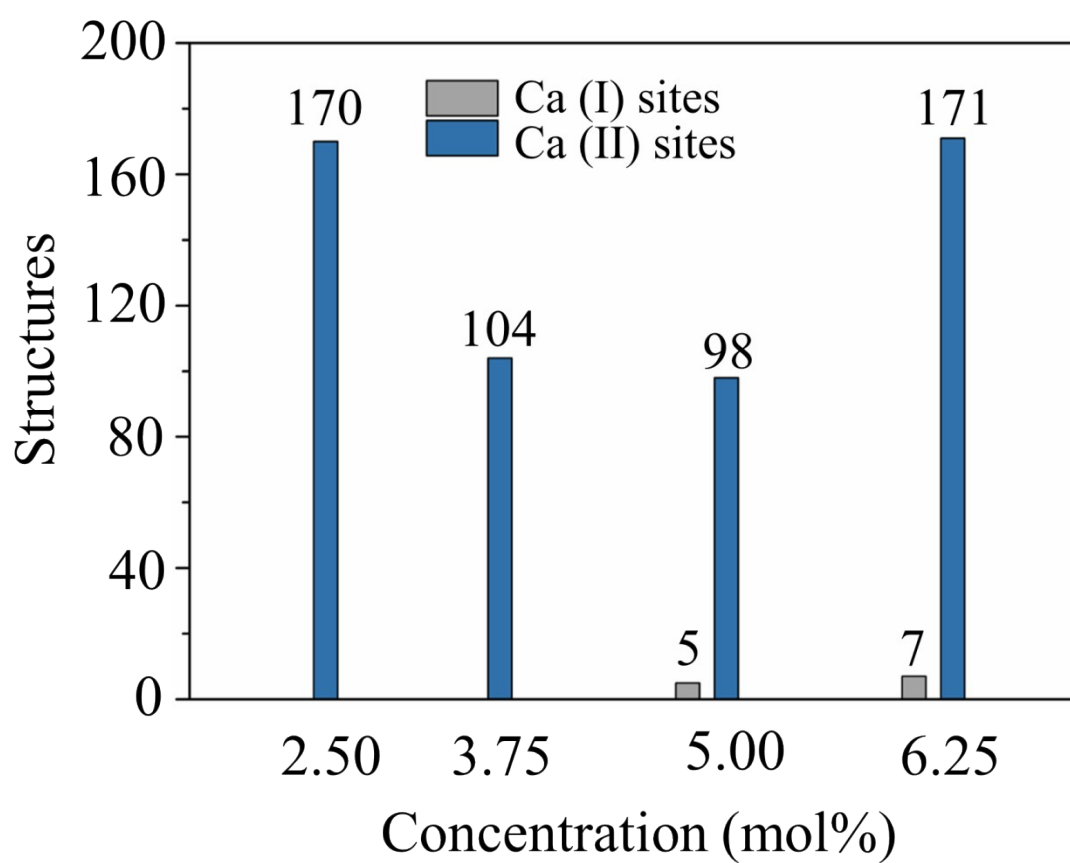| Concentration (mol%)	1.25	2.5	3.75	5	6.25
all the structures	80	2800	56000	700000	5600000
Structures in the initial dataset	56	130	300	604	1370

Figure S1. For the doping concentration of 2.5-6.25mol%, the number of the structures by the three substitution mechanisms in the lowest 5kcal/mol energy range.
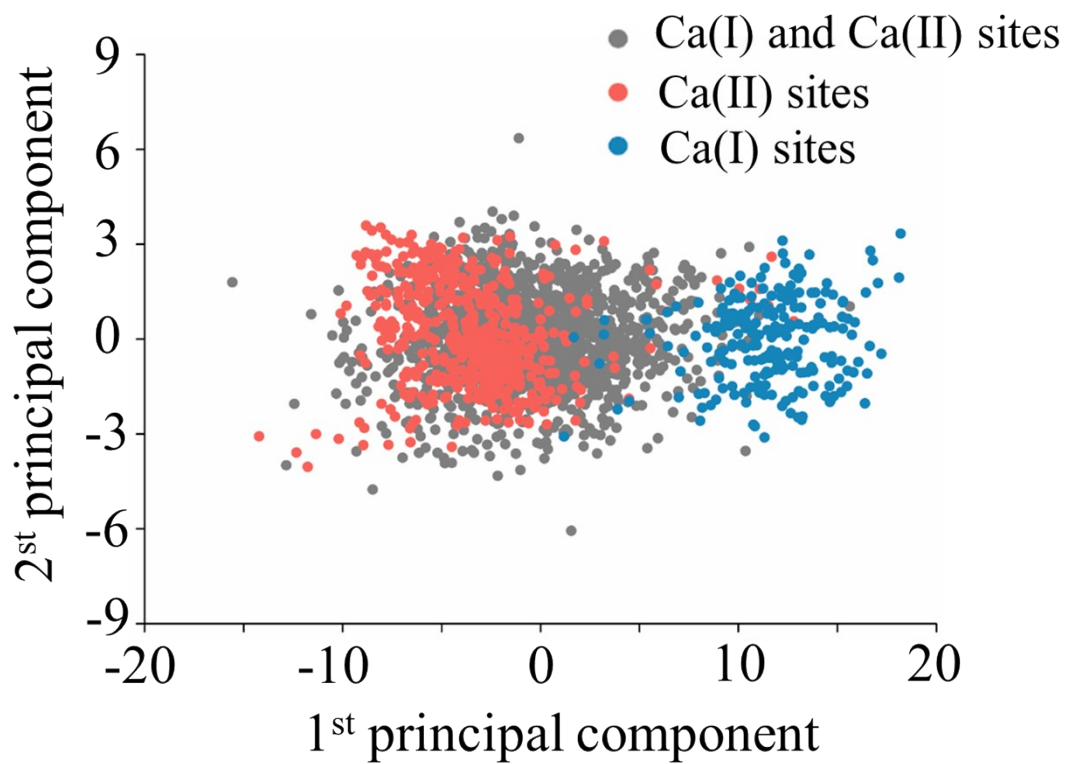
Figure S2. The two leading principal components of the feature extraction vector learned by Zn doped HAP with the concentration of 6.25 at.%.
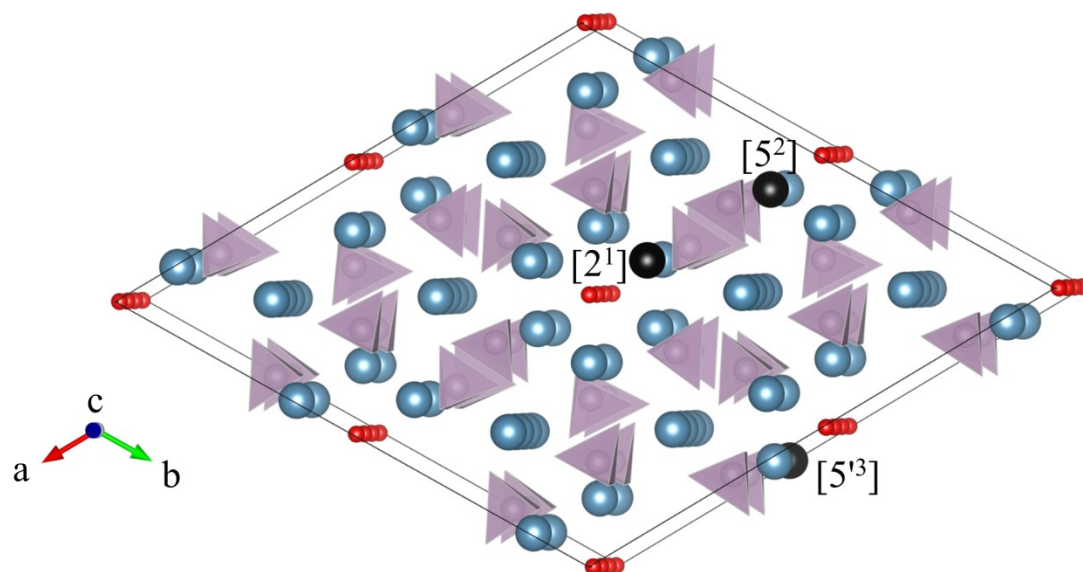
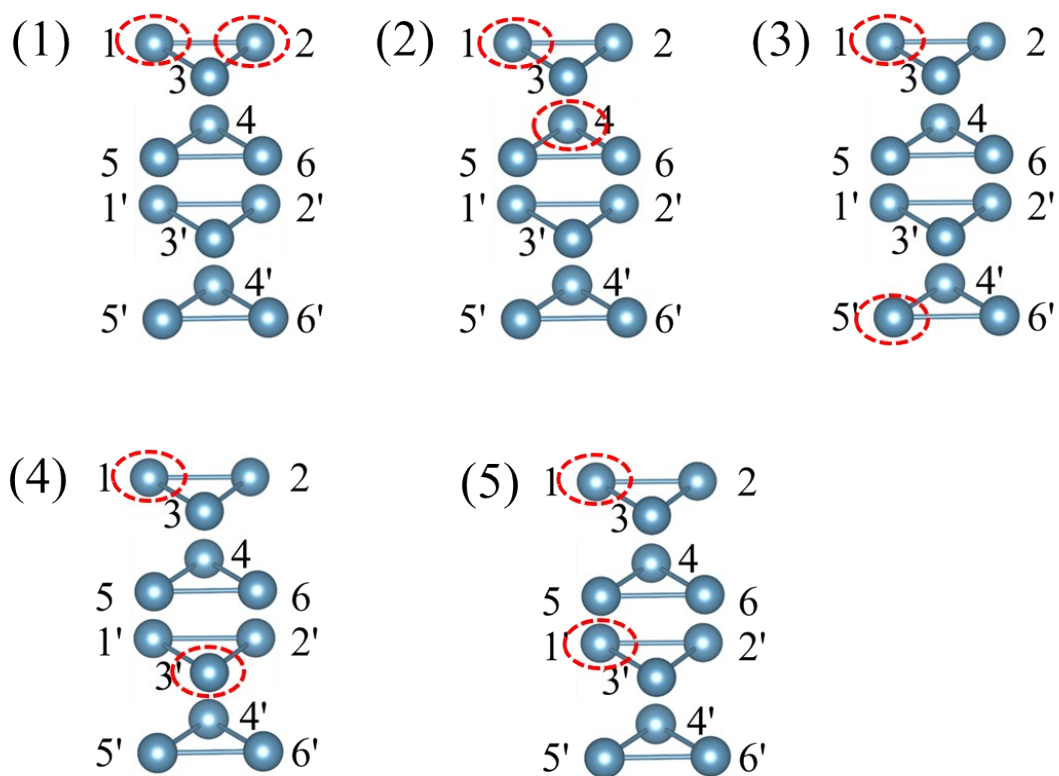**Figure S3**. Detailed display of three Zn^{3+} doped structures.

Figure S4. There are five different representative substitution mechanisms when two zinc atoms are doped in the same Ca2-V channel.
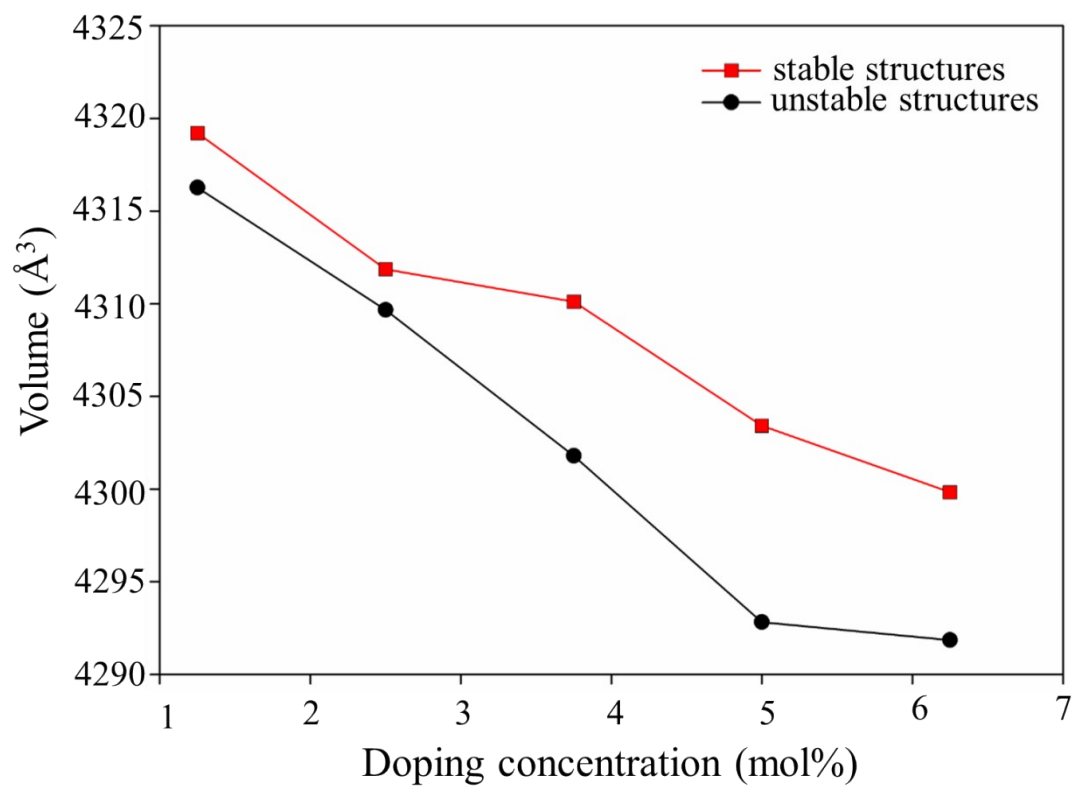
Figure S5. The comparison results of the volumes between the most stable and the most unstable structures.