Science Advances
AAAS

Supplementary Note 8
(Details of Broad Candidate Space Exploration)
for

Leveraging Data Mining, Active Learning, and Domain Adaptation in a Multi-Stage, Machine Learning-Driven Approach for the Efficient Discovery of Advanced Acidic Oxygen Evolution Electrocatalysts

Rui Ding *et al.*

*Corresponding author. Email: junhongchen@uchicago.edu, chenyuxin@uchicago.edu

Supplementary Note Discussion SND 8-1
Broad Candidate Space Exploration

In the main text, we have presented how we based on domain adaptation method to obtain Committee S-T as the ML surrogate for determining the doping configuration considered in the commonly studied RuO_2 as the basic matrix.
To further utilize the domain-adapted committee S-T, we expanded our search to a broader candidate space for lower crystal energy. This space namely allowed for the inclusion of all 49 element candidates, with the possibility of randomly doping up to three elements at random sites, each at a maximum of 16.67% atomic percentage. Through the similar method we have been using on the ML committee, we repeated the GA search for 240 times with larger population size. As illustrated in **Figs. SN 8-1** and **8-2**, we found that the most frequently doped sites, according to their serial numbers, were located in the second and third-to-last layer from the top in our five-layer slab model. Notably, Zr, Nb, and Ta emerged as superior elements in achieving lower energy configurations among the vast candidate spaces. This suggests that these elements could be easily incorporated into the RuO_2 matrix, potentially indicating feasibility in further tuning the electronic structure. Interestingly, these elements were rarely reported as dopants in RuO_2 or IrO_2, but in support materials like TiO_2[1-3]. Our results, therefore, not only provide insights into the doping mechanism but also point towards a potential new direction for materials discovery.

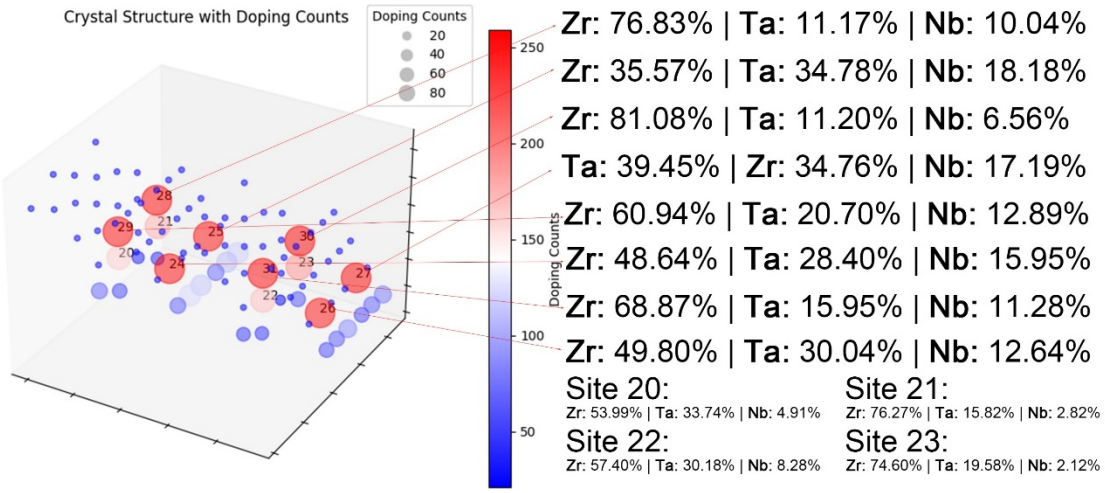

Figure SN 8-1 The statistical counts of the doped sites among the 85-time repeated broad candidate space search GA prediction with all 49 elements available and maximum doping amount of 16 atoms. The size and color of the serial number represents the total number of counts that the corresponding site has been replaced. On the right-hand side is the statistics of percentage of the site-replaced element.

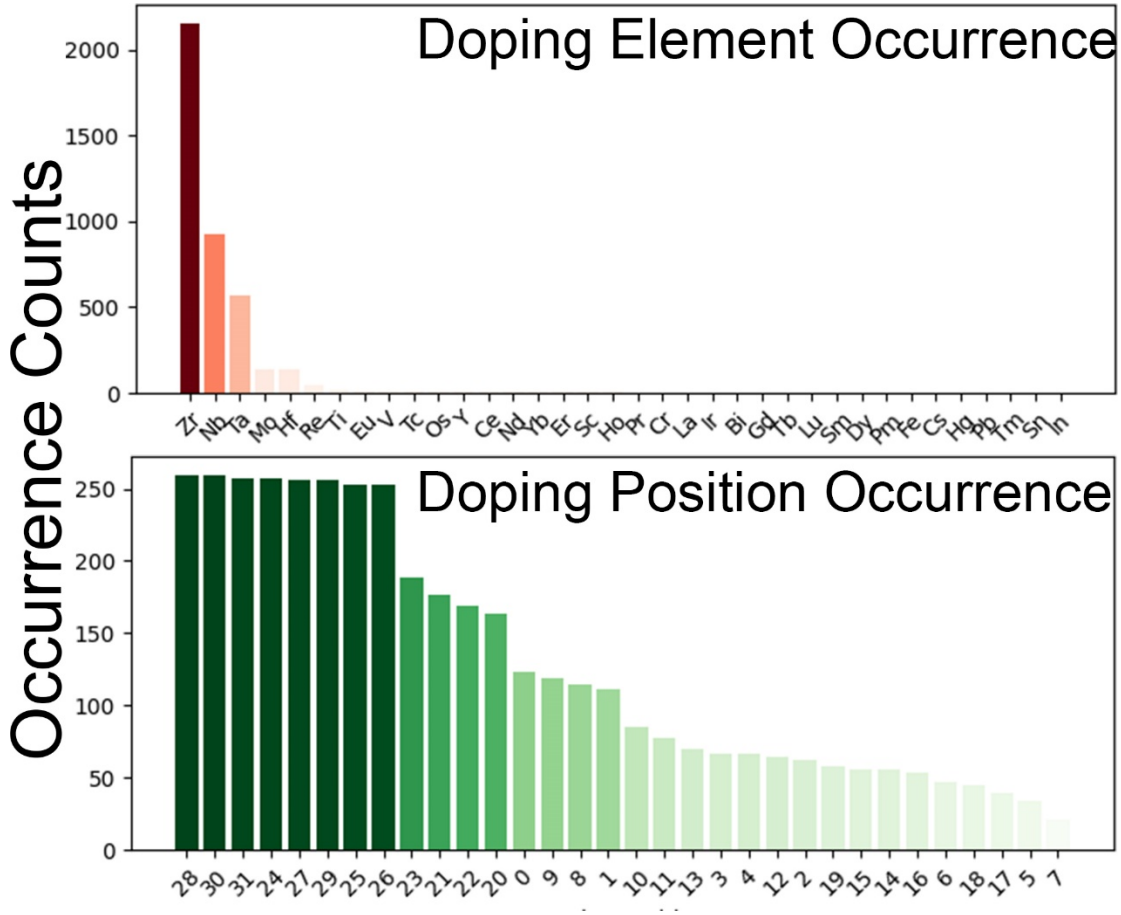

Figure SN 8-2 A histogram representing a broad candidate space search, conducted 85 times, detailing the occurrences of doped elements and their positions within the studied slab (with serial numbers indicating the atom's position in the slab per DFT calculations).

Supplementary Note 8 References:

1. N. Todoroki, R. Kudo, K. Hayashi, M. Yokoi, N. Naraki and T. Wadayama, *ACS Catalysis*, 2023, **13**, 11433-11440.
2. B. Liu, S. Wang, C. Wang, Y. Chen, B. Ma and J. Zhang, *Journal of Alloys and Compounds*, 2019, **778**, 593-602.
3. A. Nakada, S. Nishioka, J. J. M. Vequizo, K. Muraoka, T. Kanazawa, A. Yamakata, S. Nozawa, H. Kumagai, S.-i. Adachi, O. Ishitani and K. Maeda, *Journal of Materials Chemistry A*, 2017, **5**, 11710-11719.