

Electronic Supplementary Information

A Data-driven High-throughput Workflow Applied to Promoted In-oxide Catalysts for CO₂ Hydrogenation to Methanol

Mohammad Khatamirad^{‡}, Edvin Fako^{b‡}, Chiara Boscagli^c, Matthias Müller^c, Fabian Ebert^a, Raoul Naumann d'Alnoncourt^{*a}, Ansgar Schaefer^b, Stephan Andreas Schunk^{b,c,d}, Ivana Jevtovikj^c, Frank Rosowski^{a,b}, Sandip De^{*b}*

^a BasCat – UniCat BASF JointLab, Technische Universität Berlin, Berlin, Germany

^b BASF SE, Ludwigshafen, Germany

^c hte GmbH, Heidelberg, Germany

^d Universität Leipzig, Institute of Chemical Technology, Leipzig, Germany

[‡] Equal contribution

*Corresponding author. E-mail: khatamirad@tu-berlin.de, r.naumann@bascat.tu-berlin.de, sandip.de@basf.com

Supplementary Methods-1

Density Functional Theory

Computational Details Density functional theory (DFT) calculations were performed using the Vienna Ab initio Simulation Package (VASP) code ^{1,2}.

The functional of choice was Perdew-Becke-Ernzerhof (PBE) ³ Core electrons were described as projector augmented-waves (PAW), ^{4,5} with the valence mono-electronic states were expanded as plane waves with maximum kinetic energy of 450 eV. The (110) surface of cubic In₂O₃ was selected as it was identified to provide a good compromise between stability and activity ⁶. The slab was represented with 4 layers (**Figure 1**), and a single oxygen vacancy was constructed by removing a surface O atom ⁶. All intermediates were constructed on the undoped In₂O₃ slab, and relaxed. The relaxed In₂O₃ slabs were then used as a starting point for the construction of doped slabs. For each intermediate, 2 of the non-identical In atoms neighboring the O vacancy were replaced one-by-one, and the yielded system was relaxed once again (**Figure S1**). All slabs were separated by at least 10 Å of vacuum. The k-points sampling was a G-centered grid with a density of ~0.3 k-points per Å. A dipole correction was employed along the z-direction normal to the slab faces ⁷. The molecules were placed in a cubic box of 10 Å sides. In all cases the optimization thresholds were 10⁻⁶ and 0.01 eV/Å for electronic and ionic relaxations, respectively. In a few cases where the particular doping/adsorbate configuration caused slow ionic convergence, the criterion was relaxed to 0.05 eV/Å. The reference energies of and full traceback for constructing individual intermediates is provided in **Figure S2**.

Supplementary Methods-2

Subgroup-discovery (SGD) Algorithm

The SGD algorithm, is employed as a data-mining approach for detecting local patterns in data^{8,9}. This algorithm is specifically of interest when a global prediction model of target properties do not yield an optimal fit. In such scenarios, detecting interpretable local patterns has the potential of providing more knowledge about the underlying mechanisms of target properties, which might be different for each subpopulation of the data. The aim here, is to identify subgroups which depict a unique response of the target property, and then describe them based on values of candidate descriptors which are provided for the algorithm. To generate its output, the algorithm creates a pool of propositions that define only subpopulations of the data. These propositions are Boolean statements which put constraints on the value of most relevant candidate descriptors. Each subgroup is then defined as a combination of propositions.

There are different methods for evaluating the quality of detected subgroups, which might differ based on the number of target properties, as well as the type of patterns to be identified. These methods are usually similar in that they are product of two main terms, where the first term determines relative size of detected subgroup, and the second terms is a utility function. The utility function focuses on detecting subgroups whose distribution of target property is as far as possible, from the rest of the data. In our study, the quantitative comparison for each subgroup (SG) in terms of how unique it is from the rest of the dataset (P) is done *via* calculation of a quality function as follows¹⁰:

$$Q(P, SG) = \frac{s(SG)}{s(P)} \times D_{cJS}(P, SG).$$

The first term in equation above measures size ratio between subgroup and whole dataset, and ensures the detected subgroup is large enough for statistical significance. The second term is cumulative Jensen-Shanon divergence between distribution of target property in the subgroup and the rest of dataset. This term will be zero for identical distribution and increases when distributions are further apart.

Supplementary Methods-3

In the spirit of open-access research, and to promote reproducibility of our results, we have made all of the python scripts needed to generate the analysis results available on a Git repository, along with the performance and DFT dataset:

https://github.com/khatamirad/manuscript-RSC_In2O3_14

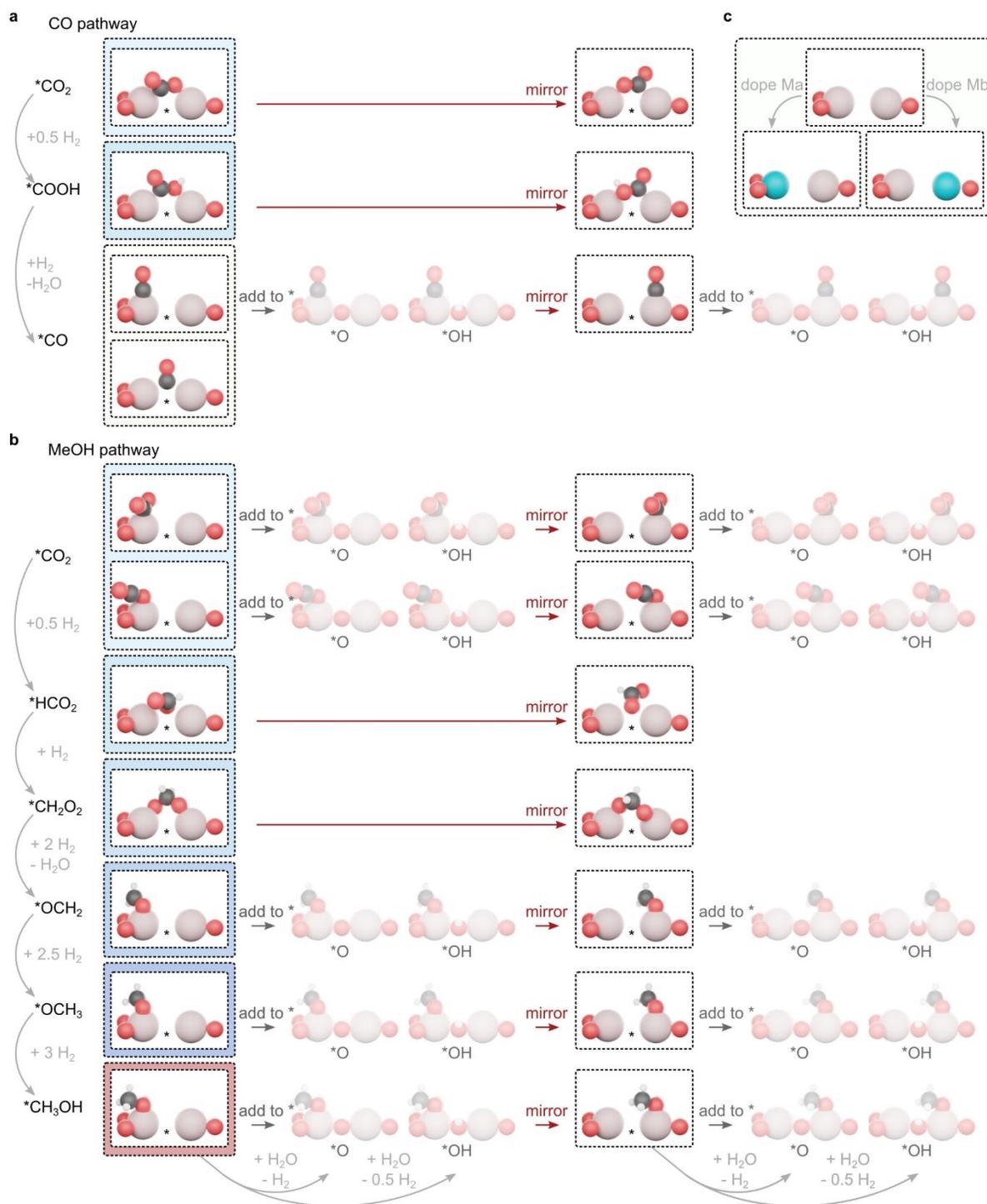


Figure S2. Intermediates along the CO (a) and MeOH (b) reaction pathway were selected maximize chemical diversity within the scope of intermediates in CO_2 reduction [Dang2021: 10.1126/sciadv.aaz2060]. Out of the three possible binding sites (m1 – top, m2 – top, O - vacancy), each top – site intermediate was placed on m1 and m1 (c), and the Ovac. Intermediates were rotated 180 degrees around the surface normal. In addition to repositioning the intermediates, the O vacancy site was populated with O (effectively completing the pristine surface), and OH for all top bonded structures.

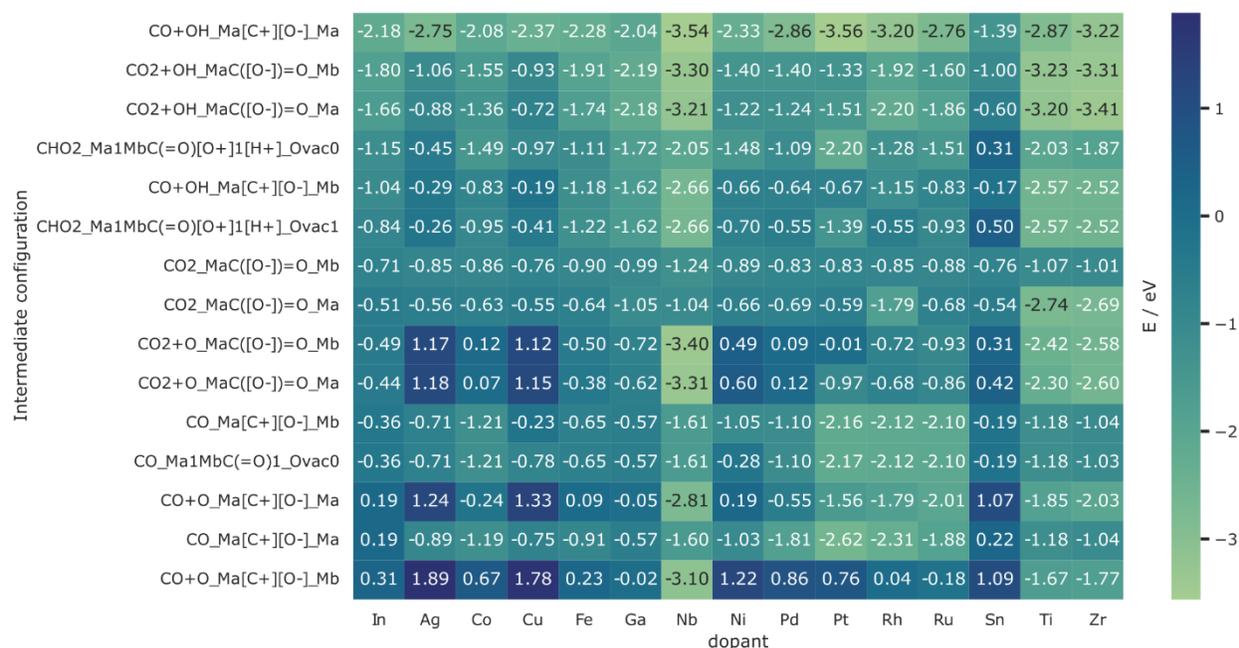


Figure S3. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

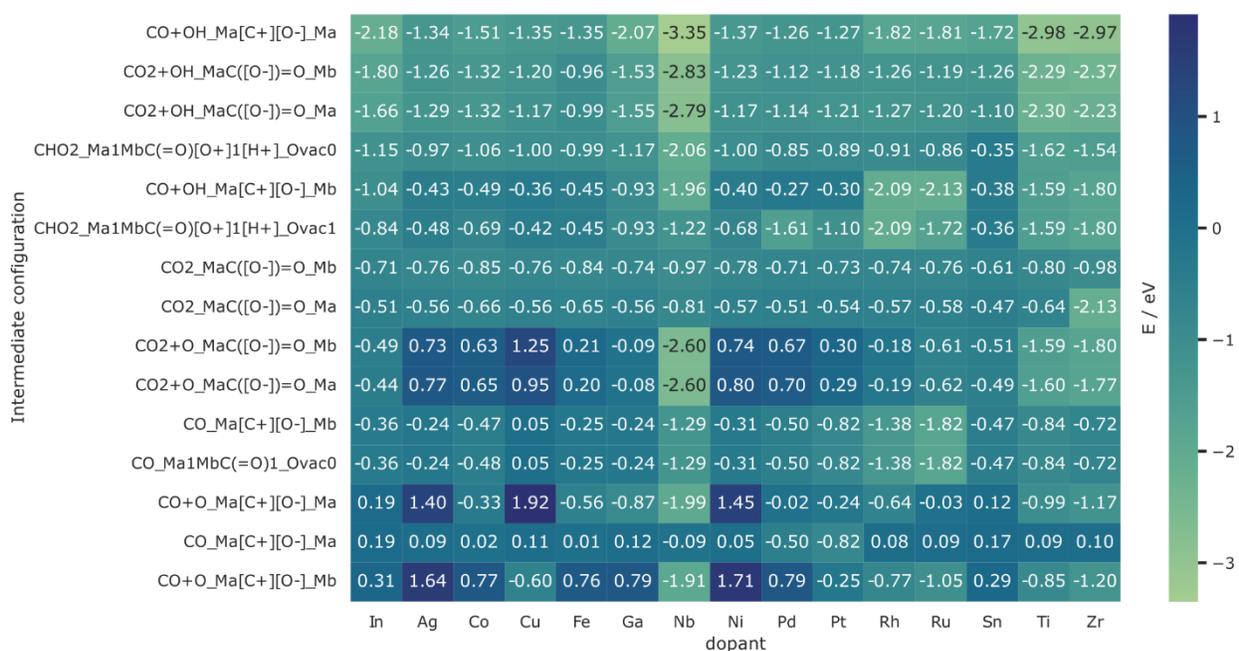


Figure S4. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

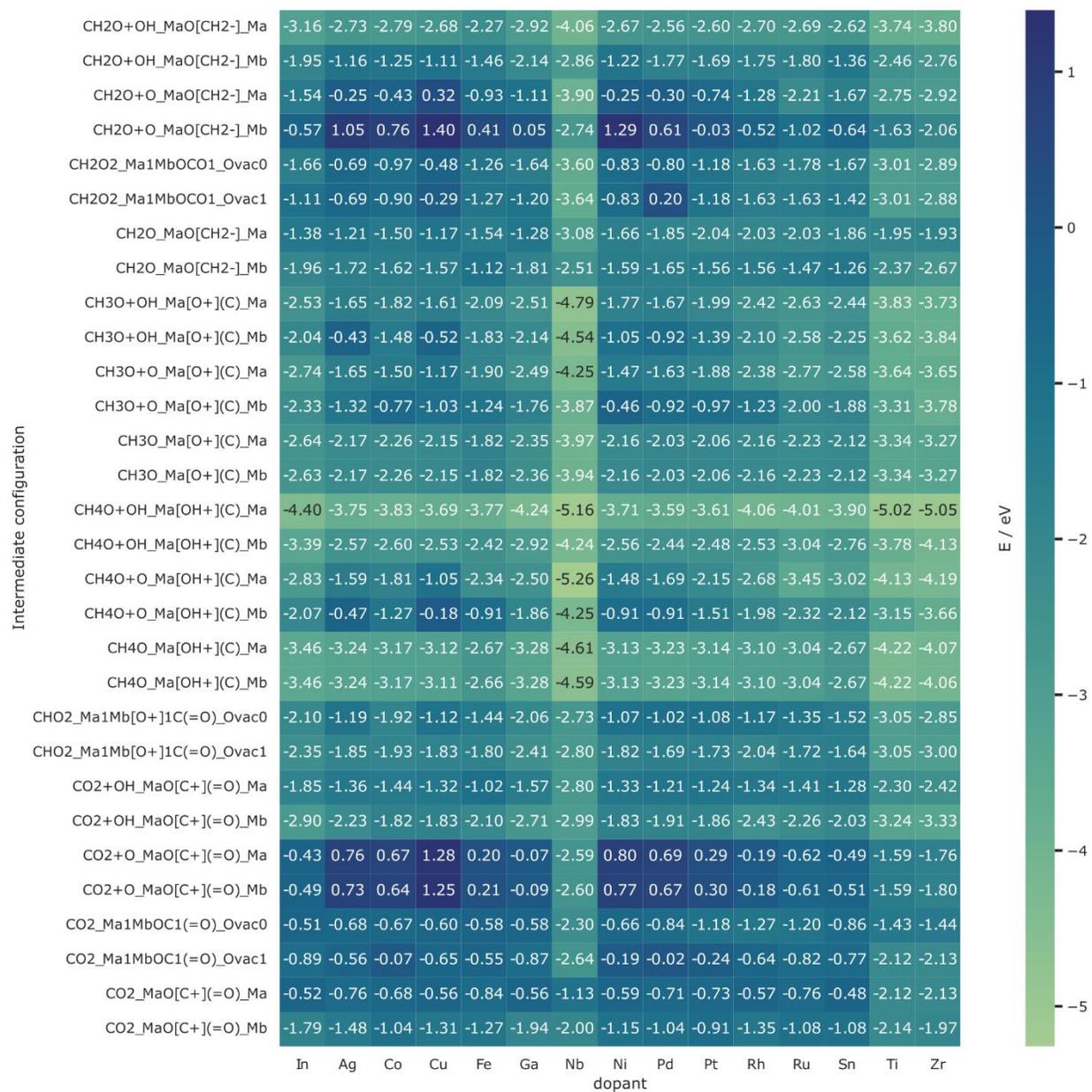


Figure S5. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

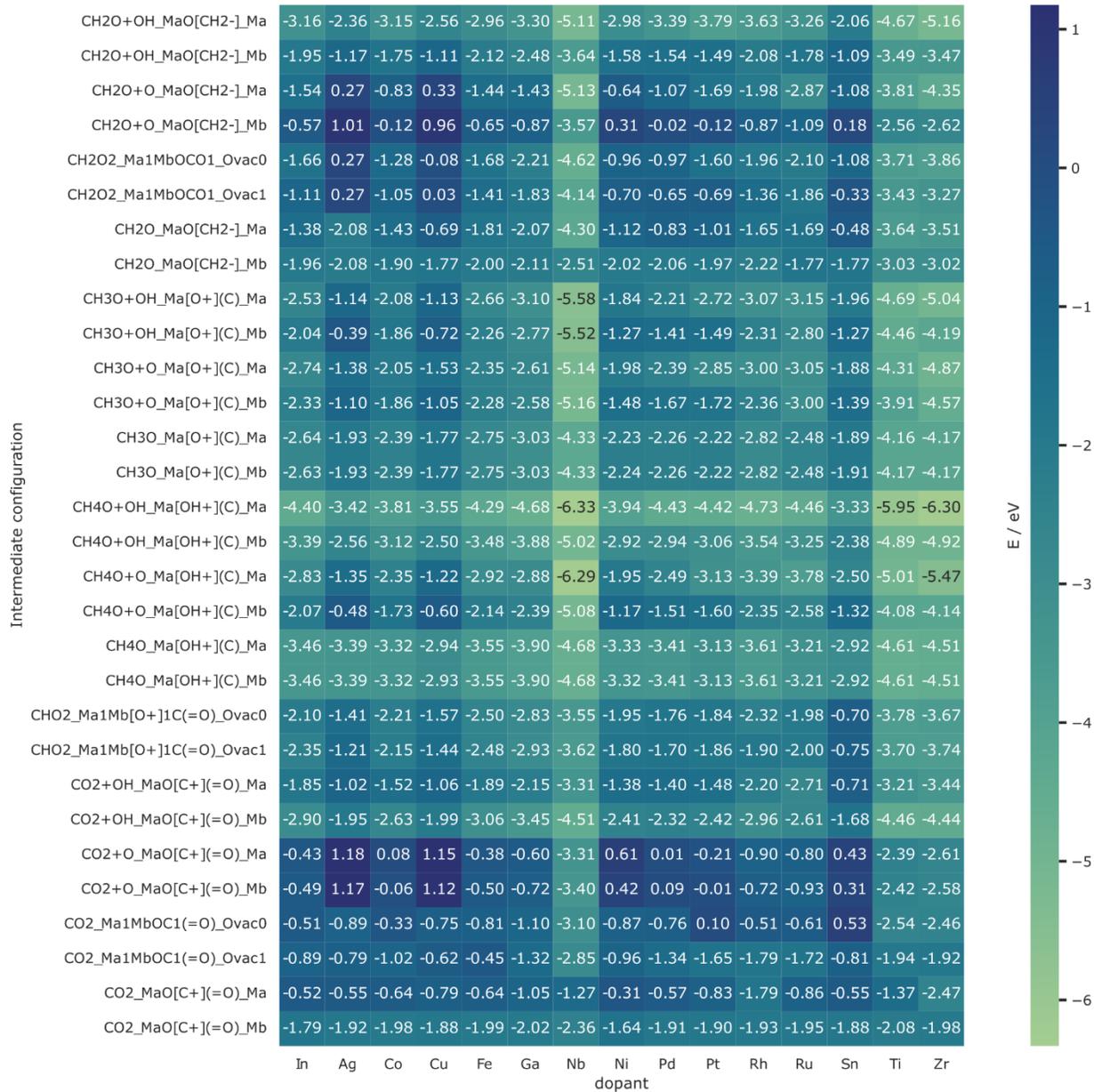


Figure S6. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

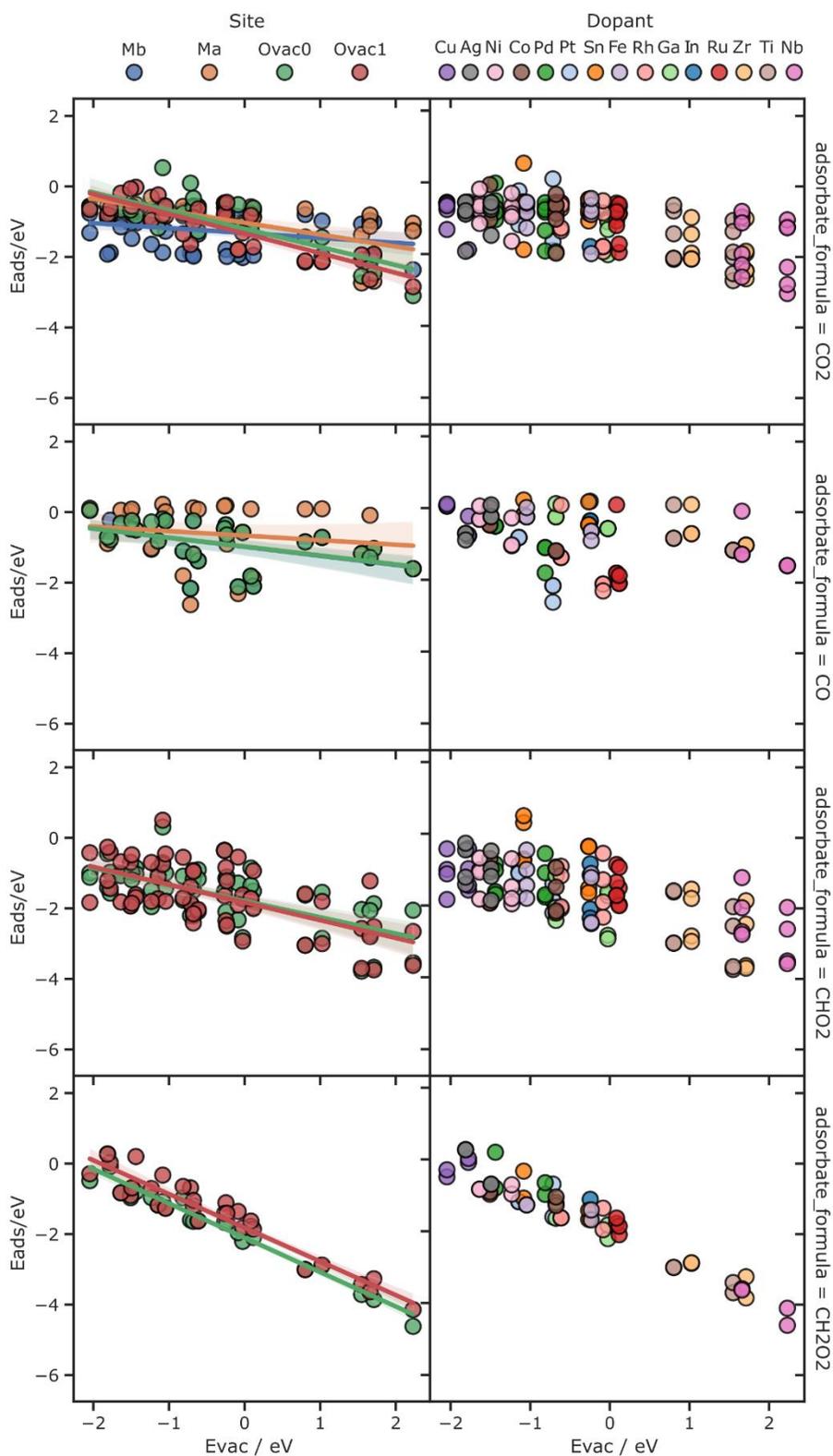


Figure S7. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

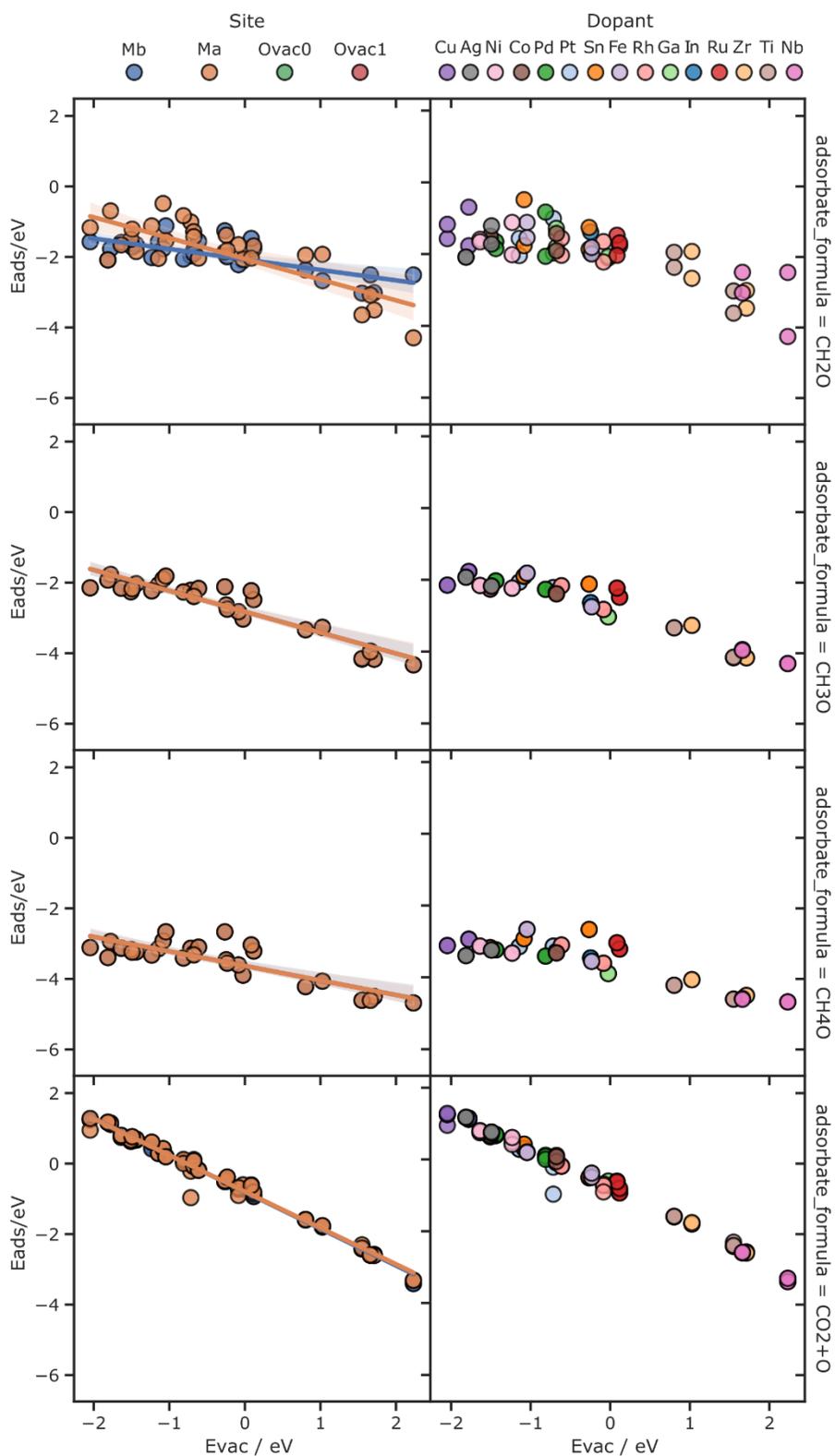


Figure S8. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

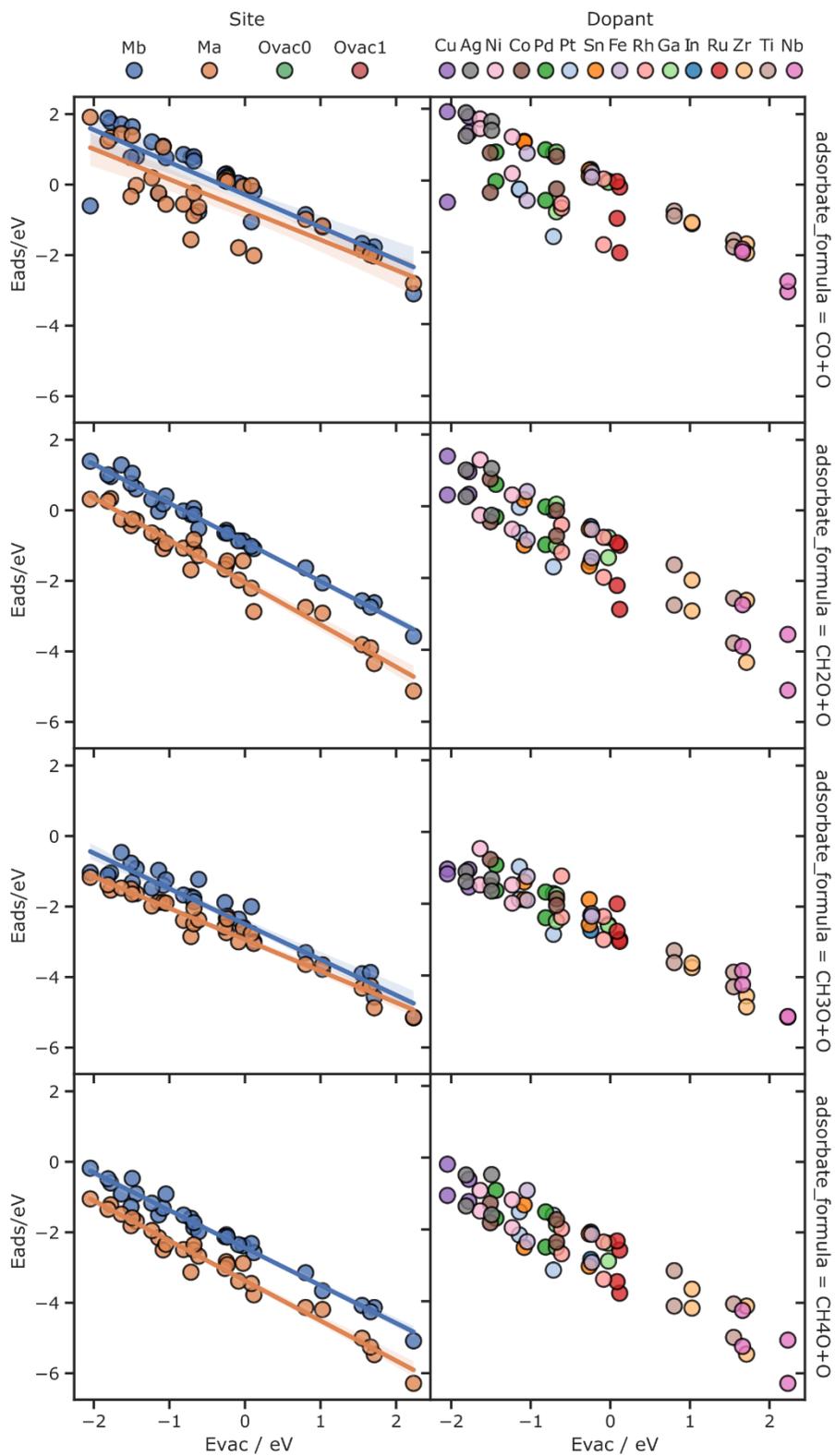


Figure S9. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

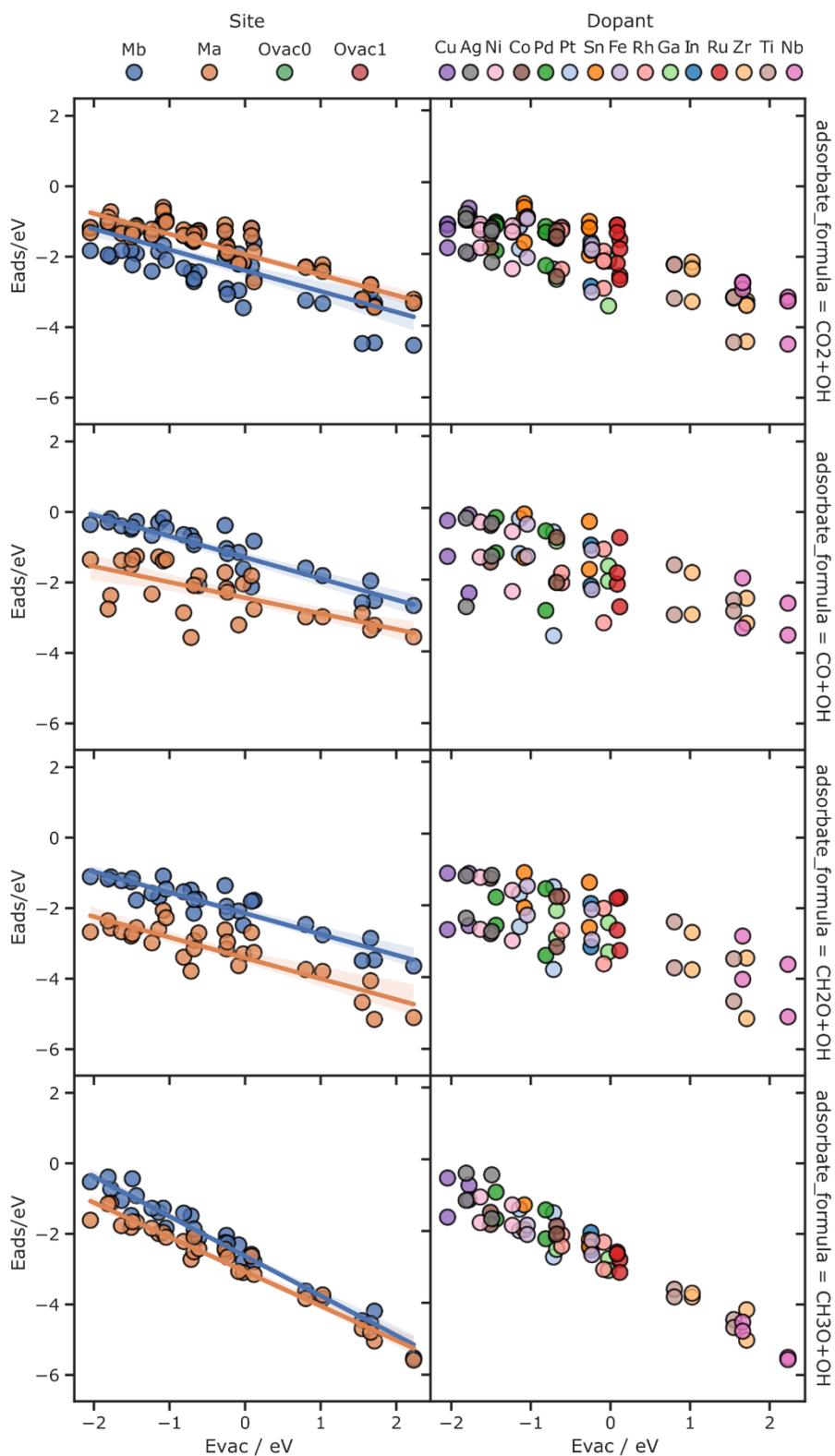


Figure S10. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

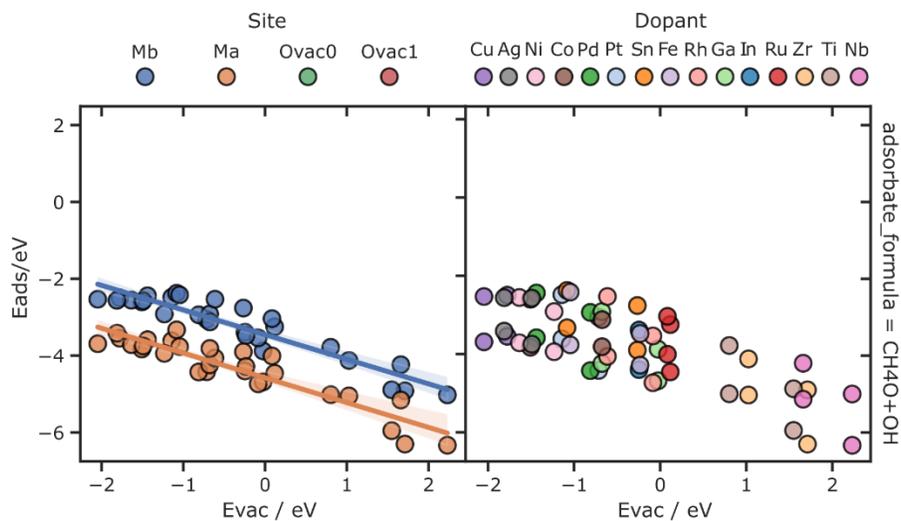


Figure S11. Correlation between the energy of the intermediate (E_{ads}) and oxygen vacancy formation energy (E_{vac}) for intermediates along the CO on MeOH pathway, colored by binding site (left) and dopant (right).

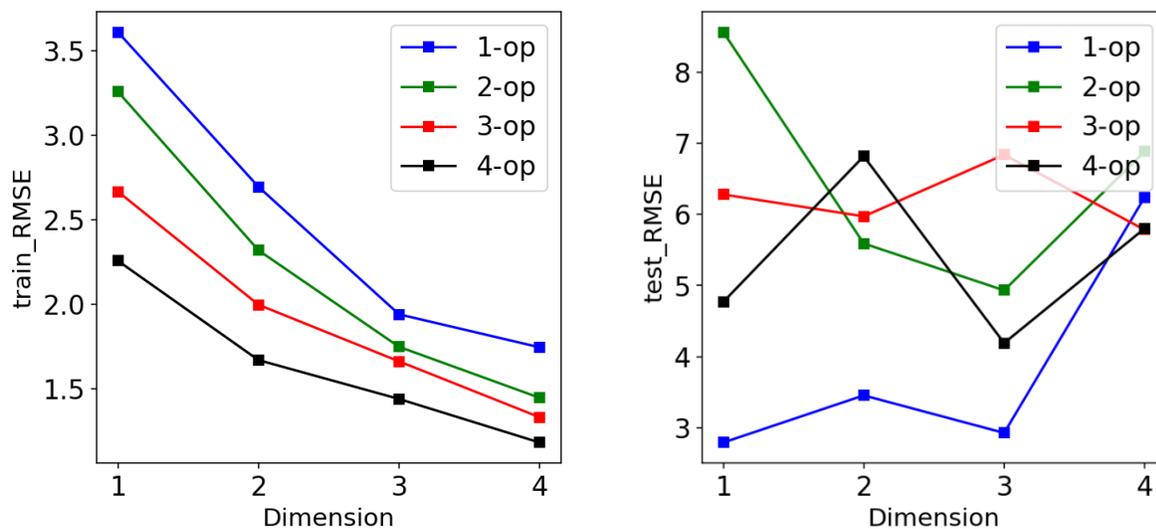


Figure S12. Obtained RMSE for train set (left) and test set (right) at different model complexities. Each RMSE value depicts the average RMSE from 14 cross validations.

References

1. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B - Condens. Matter Mater. Phys.* **54**, 11169–11186 (1996).
2. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
3. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
4. Blöchl, P. E., Först, C. J. & Schimpl, J. Projector augmented wave method: Ab initio molecular dynamics with full wave functions. *Bull. Mater. Sci.* **26**, 33–41 (2003).
5. Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B - Condens. Matter Mater. Phys.* **59**, 1758–1775 (1999).
6. Dang, S. *et al.* Rationally designed indium oxide catalysts for CO₂ hydrogenation to methanol with high activity and selectivity. *Sci. Adv.* **6**, (2020).
7. Makov, G. & Payne, M. C. Periodic boundary conditions in ab initio calculations. *Phys. Rev. B* **51**, 4014–4022 (1995).
8. Goldsmith, B. R., Esterhuizen, J., Liu, J. X., Bartel, C. J. & Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **64**, 2311–2323 (2018).
9. Goldsmith, B. R., Boley, M., Vreeken, J., Scheffler, M. & Ghiringhelli, L. M. Uncovering structure-property relationships of materials by subgroup discovery. *New J. Phys.* **19**, (2017).
10. Foppa, L. *et al.* Learning Design Rules for Selective Oxidation Catalysts from High-Throughput Experimentation and Artificial Intelligence. *ACS Catal.* **12**, 2223–2232 (2022).