Accelerated discovery of multi-elemental reverse water-gas shift catalysts using extrapolative machine learning approach

Gang Wang,[1†] Shinya Mine,[1†] Duotian Chen,[1†] Yuan Jing,[1] Kah Wei Ting,[1] Taichi Yamaguchi,[1] Motoshi Takao,[1] Zen Maeno,[2] Ichigaku Takigawa,*[3,4] Koichi Matsushita,[5] Ken-ichi Shimizu,*[1] Takashi Toyao*[1]

[1] Institute for Catalysis, Hokkaido University, N-21, W-10, Sapporo 001-0021, Japan

[2] School of Advanced Engineering, Kogakuin University, 2665-1, Nakano-cho, Hachioji 192-0015, Japan

[3] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

[4] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, N-21, W-10, Sapporo 001-0021, Japan

[5] Central Technical Research Laboratory, ENEOS Corporation, 8, Chidori-cho, Naka-ku, Yokohama 231-0815, Japan

†These authors contributed equally to this work.

*Corresponding author Email: ichigaku.takigawa@riken.jp (I.T.), kshimizu@cat.hokudai.ac.jp (K.S.), toyao@cat.hokudai.ac.jp (T.T.)

Abstract

Designing novel catalysts is key to solving many energy and environmental challenges. Despite the promise that data science approaches, including machine learning (ML), can accelerate the development of catalysts, truly novel catalysts have rarely been discovered by ML because of one of its most common limitations and criticisms—the assumed inability of the models to extrapolate and identify extraordinary materials beyond those present in the training data set. Herein, we demonstrate an extrapolative ML approach to develop new multi-elemental catalysts based on supported Pt as an active metal and TiO_2 as a support for the low-temperature reverse water-gas shift (RWGS) reaction. Using 45 catalysts as the initial data points and performing 44 cycles of the closed loop discovery system (ML prediction + experiment), we experimentally tested a total of 300 catalysts and identified more than 100 catalysts with superior activity as compared to the previously reported high-performance catalysts. The composition of the optimal catalyst discovered by this approach was Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2. Notably, Nb was not included in the original dataset, and the catalyst composition identified was unpredictable even by human experts.

One-Sentence Summary: Multi-element catalysts for the reverse water-gas shift (RWGS) reaction have been developed using a machine learning (ML) technique.

Short title: ML-assisted discovery of multi-elemental catalysts

Discovery of novel catalysts is essential for accelerating the transition to a sustainable future(*1*). Despite the significant progress in the development of highly efficient catalysts, heterogeneous catalysis remains largely an empirical science owing to the complexity of the underlying surface chemistry(*2–4*). Currently, there is a lack of data and design guidelines for heterogeneous catalysis because the computational cost of obtaining accurate theoretical models for such complex systems is currently prohibitively high while high-throughput experimental methods that have been applied successfully in related fields have not yet been thoroughly explored(*5–7*). Most of the important catalysts were discovered by chance or through trial-and-error processes extending over several years; the discovery of truly novel catalysts is still challenging(*8*).

The recent revolution in data science is expected to accelerate the development of new catalysts significantly, and hence, impact catalysis research(*9–12*). Machine learning (ML) will play a central role in this paradigm shift. The application of ML-based approaches to catalysis(*13–17*) and broader fields of chemistry and materials science has attracted considerable attention(*18–22*). Although proof-of-concept examples of reduction in time and cost of catalyst development have been demonstrated using ML-based approaches, most of the ML-based research is directed toward the resolution of benchmark problems, while truly novel compounds and materials have rarely been discovered(*23*, *24*). This is due to one of the most common limitations of ML—the assumed inability of the models to extrapolate and identify extraordinary materials beyond those present in the training data set(*25*). In materials and catalysis informatics, we often desire to use ML models to discover an entirely new class of materials and catalysts with unprecedented combinations of elements. In this context, our group has developed a new ML approach wherein elemental features are used as input representations rather than inputting the catalyst compositions directly(*26*, *27*). Namely, each catalyst is represented as a set of elemental descriptors such as electronegativities and melting points, which are scaled by the element content, followed by aggregation into a single feature vector by a permutation-invariant readout operation (elementwise sort pooling, referred to as sorted weighted elemental descriptor (SWED))(*26*, *27*). This ML method can guide catalyst design and discovery in areas where there is limited overlap of catalyst compositions and even for elements that were previously never included in a given dataset, thereby enabling extrapolative and ambitious prediction beyond the training data. Other studies have also validated the possibility of such extrapolative prediction using relevant feature engineering/selection approaches(*28*). Despite the theoretical evidences on the possibilities of finding novel catalysts and exceptional materials through extrapolative prediction, the use of ML to identify truly new and exceptional materials has remained elusive(*29*).

In this study, we have applied the extrapolative ML approach to develop new multi-elemental catalysts based on supported Pt as an active metal and TiO_2 (P25) as a support for the low-temperature

reverse water-gas shift (RWGS) reaction. This reaction was chosen because its product, CO, is an important intermediate in various well-established catalytic processes for manufacturing value-added chemicals; that is, the RWGS reaction enables highly flexible utilization of CO_2(*30*, *31*). We explored M elements of up to five types for Pt(3)/$M_1(X_1)$-$M_2(X_2)$-$M_3(X_3)$-$M_4(X_4)$-$M_5(X_5)$/TiO_2 RWGS catalysts (3 wt% Pt, TiO_2 = P25). For M, elements with atomic number 3 (Li) through 83 (Bi), except for Be, B, C, N, O, P, S, As, Se, Tc, Te, Pm, Ta, Hg, Tl, halogens, noble gasses, and platinum group metals, were used as catalyst components (50 elements in total). Each M element had a unique loading amount (X) for each catalyst. Thus, the total number of catalyst candidates easily exceeded 10^{11} even if only integer values up to 5 wt% were considered as the loading amount of M ($_{50}C_5 \times 5^5 \approx$ 800 billion). We have tested three types of ML approaches, each of which differs in the input representations of the catalysts: (i) a *conventional* ML model, which uses only elemental compositions; (ii) an *exploitative* ML model, which uses both elemental compositions and elemental properties; and (iii) an *explorative* ML model, which uses only elemental properties. For the input representation of the elemental compositions, each catalyst was represented as a vector of the compositional fractions for all the 50 elements under consideration. On the other hand, for the input representation of the elemental properties, vectors of 8 selected elemental descriptors for each element, scaled by its composition fraction, were aggregated into a single feature vector by sum pooling. The initial dataset consisting of 45 data points was constructed using the catalysts reported in our previous experimental study(*32*) and some catalysts fabricated in the present study (**Table S4**); this dataset was set as “Iteration” = 0. We then trained the explorative ML model based on Extra-Trees regression (ETR)(*33*) with the initial dataset (45 data points), calculated the expected improvement (EI) for all the test points in the catalyst composition grid, selected several prominent catalyst candidates considering the EI values and catalyst variety, synthesized the catalysts using the sequential impregnation method, performed the RWGS reaction, and updated the dataset to close the loop (**Scheme S1**). We continued this process for 44 loops to test 300 catalysts. The explorative ML model was used in the initial effort to explore many elements and because its prediction accuracy was the highest among the three ML models while the exploitative ML model was used after the prediction accuracy reached a certain level (after 30 iterations).

Through experimental testing of 255 ML-predicted new catalysts corresponding to 44 cycles of the closed loop discovery system (ML prediction + experiment), we found more than 100 catalysts that showed higher activity than the previously reported high-performance catalyst (Pt(3)/Mo(10)/TiO_2)(*32*), (**Figure 1)**. The composition of the best catalyst discovered by this approach was Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2, and it exhibited the highest activity (CO formation rate per catalyst mass, mmol min^{-1} g_{cat}^{-1}) at temperatures below 250 °C compared to the previously reported catalysts, while retaining

100% CO selectivity (**Table S3**). Notably, Nb was not included in the original dataset (**Figure 2**), and the identified catalyst composition could hardly be predicted even by human experts.

Although ML is often employed as a black box without any prior insight into what the model has actually learned, supervised ML models can be used to identify important chemical moieties influencing the prediction, even without any explicit knowledge of its underlying principles(*34*). Extrapolative ML can reveal not only the effective catalyst compositions but also the required elemental features and electronic properties for the precise designing of ideal catalysts. SHapley Additive exPlanations (SHAP)(*35*, *36*) was used to understand the importance of the descriptors for ML prediction, as shown in **Figure 2D**. SHAP can be used to explain the contribution of a given input feature to the model output (e.g. CO formation rate) by identifying and ranking the descriptors(*26*). Elemental properties such as group, electronegativity (EN), and density were identified as important factors, with smaller values tending to positively affect the model output. The SHAP values were also calculated using the exploitative elemental descriptor representation, because this method considers the elemental composition directly and facilitates the understanding of the contribution of the elements in the given data. For the catalyst composition, Mo, Tb, Na, and Ba were identified as important descriptors. The SHAP values were analyzed using waterfall plots for the two representative catalysts (Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 and Pt(3)/Mo(10)/TiO_2), as shown in **Figure 2C,D**. The analysis reveals the descriptors that are responsible for the increase or decrease from the average value of the dataset (2.28) relative to the predicted value. EN, group, and oxide band gap (BG) values were found to strongly contribute to the high activity of the catalyst. The waterfall plots for some additional catalysts are also included in the Supplementary Material (**Figures S9, 10, 13 and 14**).

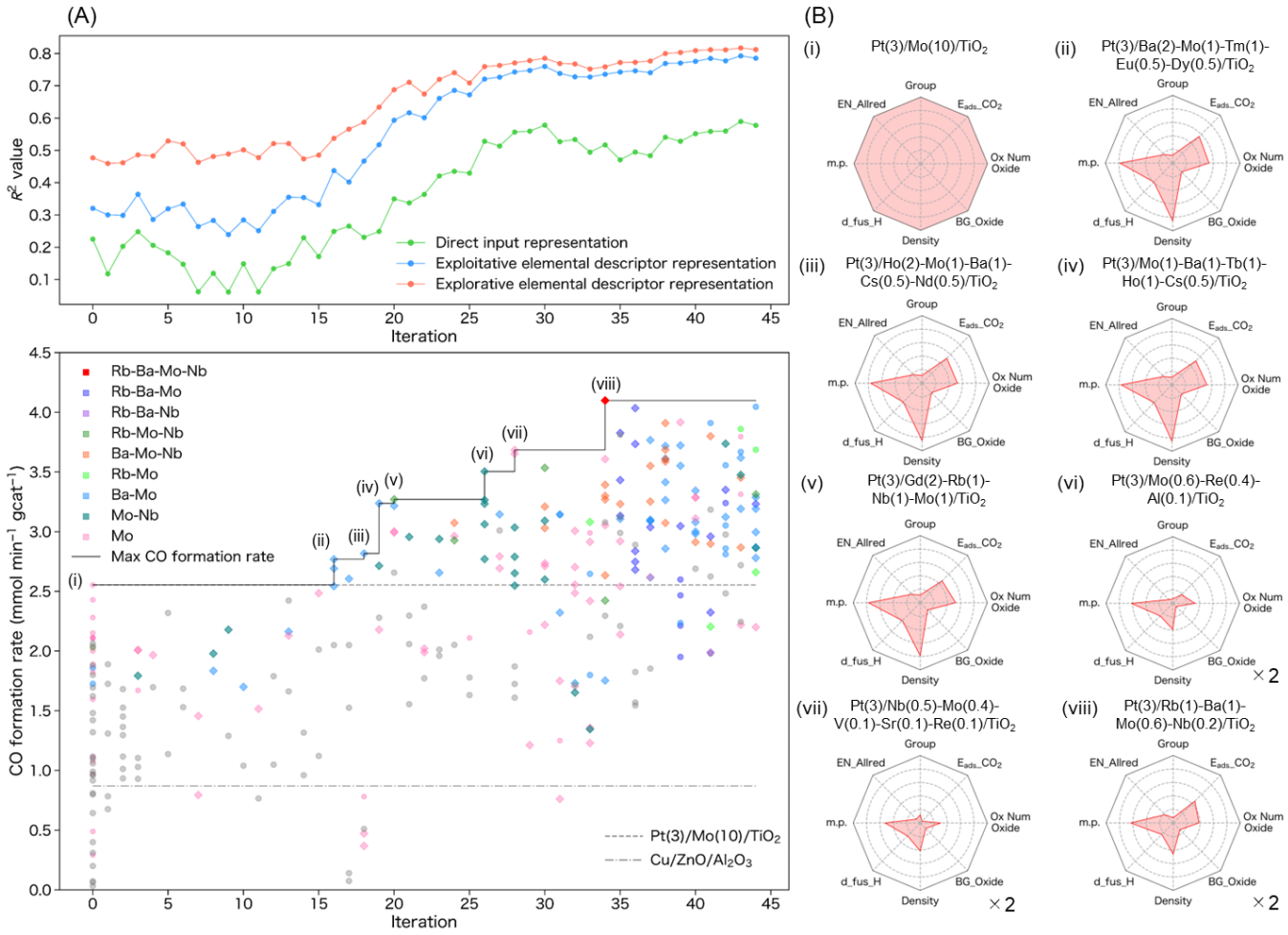

Fig. 1. ML-assisted exploration of RWGS catalysts. (A) ML-assisted exploration of RWGS catalysts using the explorative and exploitative ML methods based on ETR. (B) Radar charts of the elemental descriptors for the best catalysts at each iteration. Descriptor values relative to the (i) Pt(3)/Mo(10)/TiO_2 catalyst are shown.

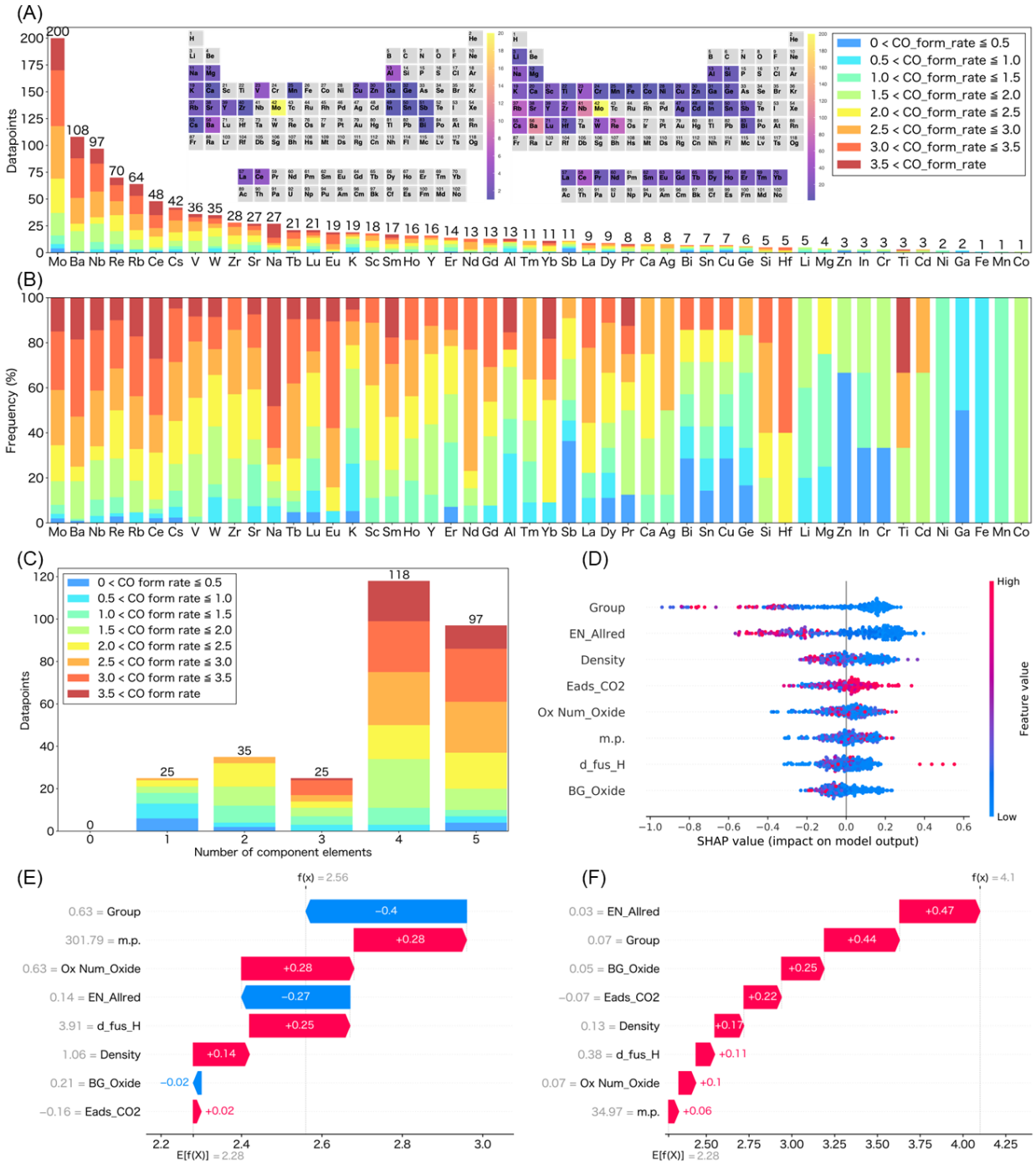

Fig. 2. Visualization of RWGS catalyst datasets and ML-assisted statistical analysis. (A,B) Histograms for each additive oxide component categorized by the RWGS activity; the elements in the original (left) and final (right) datasets are shown in the periodic tables. The maximum values on the Y axis for (A) represent the sum of the number of data points while that for (B) represent percentage of the RWGS activity category. (C) Number of component elements as additive oxides. (D) SHAP values of the descriptors used to predict CO formation rates (red and blue correspond to high and low features, respectively). Breakdown of SHAP values as waterfall plots for (E) the original best catalyst Pt(3)/Mo(10)/TiO_2 and (F) the current best Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 to determine the feature values that are responsible for the increase or decrease from the base. Positive and negative contributions of each feature are shown in red and blue, respectively. Explorative elemental descriptor representation was used.

With the best catalyst composition in hand, we then performed structural analysis (**Figures 3 and S15-23 and Tables S6 and S7**) and mechanistic studies (**Figures 4 and S24-28 and Table S8**). This is important because investigations of extraordinary materials can provide new scientific insights. The X-ray diffraction (XRD) pattern of Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 was essentially the same as that of pristine TiO_2 (P25) (**Figure S15**). High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) was performed for TiO_2 (P25), Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2, and Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 (**Figure 3A**). The oxide additive species was found to be highly dispersed over the TiO_2 surface. In addition, the Pt nanoparticles in Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 were highly dispersed, with an average Pt particle diameter of 1.8 nm (**Figure S18**). Comparison with the previously identified Pt(3)/Mo(10)/TiO_2 active catalyst (particle size of 2.6 nm)(*32*) revealed that the average particle size of the supported Pt was smaller in Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2.

X-ray absorption spectroscopy (XAS) was conducted to identify the chemical states of the introduced species in the RWGS catalyst (**Figures 3B and S20**). The X-ray absorption near-edge structure (XANES) of the reduced Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 catalyst was identical to that of the Pt foil used as the reference. Mo K-edge XANES showed that the shape and edge position of the unreduced Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 catalyst were identical to those of the reference MoO_3. For the reduced Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 sample, the absorption edge shifted toward lower energies, indicating the reduction of the Mo species upon pretreatment with H_2. X-ray photoelectron spectroscopy (XPS) measurements were conducted to identify the oxidation states of Mo (**Figure 3C**). Peaks corresponding to Mo^{4+} were predominantly observed, in addition to small peaks of Mo^{6+} and Mo^{2+}. The other additives, including Rb, Ba, and Nb, did not change their oxidation states and existed in the form of Rb_2O, BaO, and Nb_2O_5, respectively, after the reduction pretreatment with H_2 (**Figure S22**).

In situ CO adsorption IR spectroscopy experiments were conducted to examine the electronic state of the Pt species on a series of supported Pt catalysts in order to understand the effects of the introduced additives (**Figure 3D**). All the spectra showed a peak at 2071–2077 cm^{-1}, corresponding to the CO bound to the on-top sites of the metallic Pt surface. The center of the CO adsorption peak shifted to higher wavenumbers, following the order Pt(3)/TiO_2, Pt(3)/Rb(1)-Ba(1)-Mo(0.6)/TiO_2, Pt(3)/Mo(0.6)/TiO_2 and Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2. Therefore, the introduction of additives favors the formation of more electron-deficient metallic Pt^0 species, compared to pristine Pt(3)/TiO_2, and is expected to improve the resistance to CO poisoning during the RWGS reaction. The same trend was also observed by XPS (**Figure S23**).

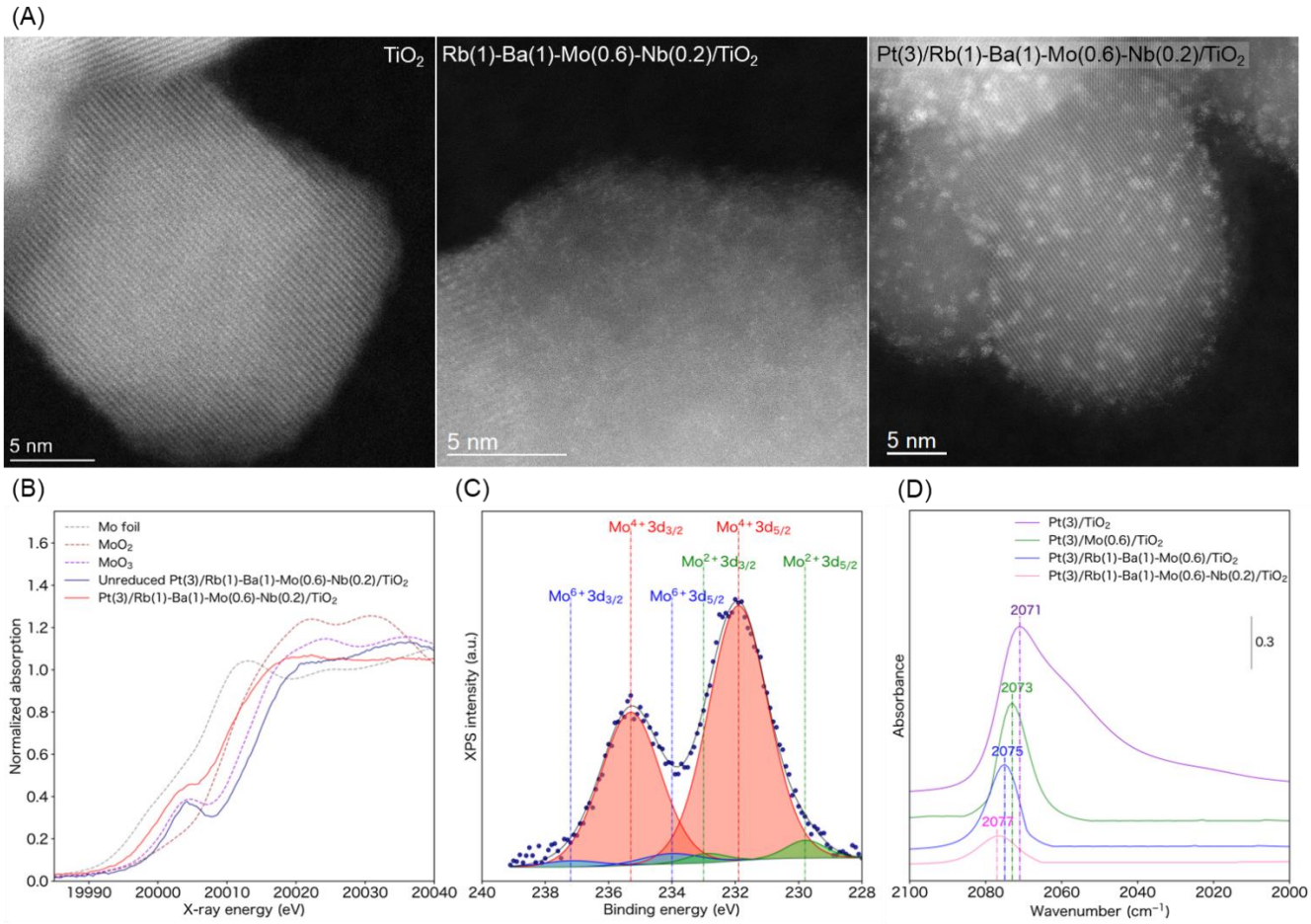

Fig. 3. Structural analysis of the ML-identified RWGS catalyst. (A) HAADF-STEM images of TiO_2, Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2, and Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2. (B) Mo K-edge XANES of unreduced and reduced Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 and reference compounds. (C) XPS spectra of the supported Pt catalysts after the H_2 reduction pretreatment at 300 °C without exposure to air. (D) IR spectra of the CO adsorbed on the supported Pt catalysts recorded at 250 °C after the H_2 reduction pretreatment at 300 °C. The sample was exposed to a flow of 1% CO/He (100 mL min^{-1}) for 5 min and purged with He for 5 min.

Kinetic studies were conducted on the optimal catalyst (Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2). The apparent activation energy (E_a), as calculated from the Arrhenius plot, was 45.6 kJ mol^{-1} (**Figure 4A**). Similarly, the E_a values of Pt(3)/Rb(1)-Ba(1)-Mo(0.6)/TiO_2, Pt(3)/Mo(0.6)/TiO_2, and Pt(3)/TiO_2 were 48.7, 52.8, and 58.4 kJ mol^{-1}, respectively. The apparent reaction orders with respect to H_2, CO_2, and CO were calculated to understand the effect of the introduced additives (**Table S8**).The apparent reaction orders for both CO_2 and H_2 in the case of the catalyst with oxide additives decreased as compared to those for pristine Pt(3)/TiO_2, indicating weaker dependence on their concentrations. In addition, the reaction order with respect to CO was the smallest for Pt(3)/Rb(1)-Ba(1)-Mo(0.6)/TiO_2, indicating the less inhibitory effect of CO for the best catalyst. This result is consistent with results of the *in situ* IR and XPS experiments. These combined results indicate that the introduction of Nb renders Pt more electron-

deficient and induces high tolerance to CO poisoning, leading to a high catalytic activity. The CO_2-TPD analysis of the catalysts without Pt (**Figure S24**) suggested that the introduced additives could facilitate the adsorption of CO_2 owing to the introduced base metal oxides, particularly Rb and Ba, thereby promoting the reaction efficiently.

The RWGS reaction is known to proceed mainly via the (i) redox mechanism and (ii) associative mechanism (*37*). In the former, oxygen vacancies are formed on the surface of the support oxide by H_2, while CO_2 reoxidizes the partially reduced oxide to fill the formed oxygen vacancies(*38*), resulting in the formation of CO. In the latter mechanism, CO is produced through decomposition of the surface-reactive intermediates such as formate and carbonates(*37*).

To elucidate the reaction mechanism, *operando* XANES measurements were conducted under CO_2, H_2, and $CO_2 + H_2$ flow at 250 °C (**Figure 4B**). The Mo K-edge XANES of Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 shows that the absorption edge shifts to higher energies after the introduction of CO_2, while CO was simultaneously detected by GC. The results clearly demonstrated that CO_2 acted as an oxidant to oxidize the Mo species. Notably, CO was formed even upon the introduction of H_2, suggesting that the reaction also proceeded through the associative mechanism. For the Pt L_3-edge (**Figure S25**), the white line intensity became slightly stronger under CO_2 flow, suggesting that metallic Pt was also oxidized by CO_2. Note that this change can be solely because of the adsorption of the CO formed, as it is well-known that the Pt L_3-edge XANES intensity and shape is altered by the adsorption of CO(*39*). The K-edge XANES spectra of Ti, Ba, Rb, and Nb were also obtained employing a protocol similar to that described above (**Figure S25**). The edge positions in all these XANES spectra hardly changed following the introduction of CO_2, indicating that no redox reactions of TiO_2, BaO, Rb_2O, and Nb_2O_5 occurred during the RWGS reaction.

Operando IR spectroscopy was also performed to investigate the adsorbed surface species that are likely to be involved in the RWGS reaction (**Figure 4C**). Bands in the range of 1700–1200 cm^{-1}, which can be assigned to the surface-adsorbed species such as carbonate and formate(*37*), appeared immediately after the introduction of CO_2. Simultaneous formation of CO in the gas phase was also observed using an IR gas cell at the outlet. Bands at 2100–1950 cm^{-1}, which can be assigned to adsorbed CO, were also observed. The amount of these surface species over the best catalyst was higher than those over Pt(3)/Mo(0.6)/TiO_2 and Pt(3)/TiO_2, yet lower than that over Pt(3)/Rb(1)-Ba(1)-Mo(0.6)/TiO_2 without Nb (**Figure S26**). The evolution of the bands in the ν_{CH} region (2800–2960 cm^{-1}) also supports the formation of formate species under the flow of CO_2 and H_2. These results indicate that the Ba and Rb species act as base components to generate the surface adsorbed species that lead to the formation of CO. To confirm this, H_2 was introduced to the Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 catalyst with such adsorbed species, as shown in **Figures 4C and S27**. Note that for this purpose, a lower temperature

(200 °C) was also employed to clearly observe the adsorbate peaks. Intensities of the bands between 1700 and 1200 cm^{-1} decreased upon the introduction of H_2, and simultaneous formation of CO in the gas phase was observed. These *operando* XAS and IR results indicated that Mo acted as a redox species while Rb and Ba acted as bases to promote the RWGS reaction. Nb was not directly involved in the reaction; it rather modified the electronic structure of Pt, ensuring high CO tolerance. These multiple functions realized by the combination of the oxide additives identified are vital for achieving high catalytic performance.

Finally, the durability test (**Figure S29**) confirmed the high durability of the optimal catalyst compared to those of the other reference catalysts, including the previously identified Pt(3)/Mo(10)/TiO_2 catalyst and a commercial Cu/ZnO/Al_2O_3 catalyst. Therefore, Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2, which was predicted by the ML-assisted method, is an outstanding state-of-the-art catalyst for the low-temperature (250 °C) RWGS reaction.

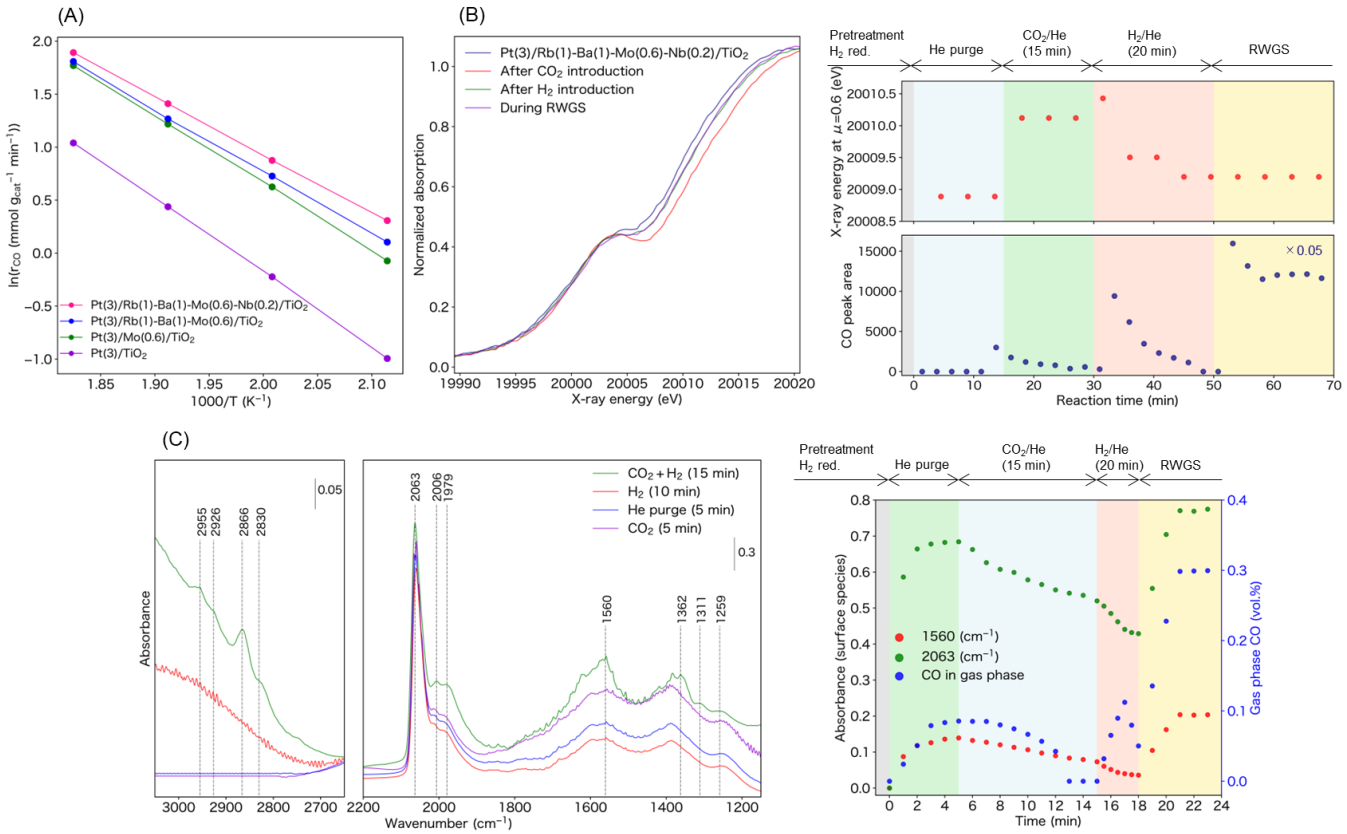

Fig. 4. Mechanistic studies. (A) Arrhenius plots for the RWGS reaction conducted over Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2, Pt(3)/Rb(1)-Ba(1)-Mo(0.6)/TiO_2, Pt(3)/Mo(0.6)/TiO_2, and Pt(3)/TiO_2 under the following reaction conditions: 10 mg catalyst, 20 mL min^{-1} CO_2, 60 mL min^{-1} H_2, 5 mL min^{-1} N_2, and 200–275 °C. (B) *Operando* Mo K-edge XANES of Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 measured under a sequential flow of 25% CO_2/He, 75% H_2/He, and 25% CO_2 + 75% H_2 at 250 °C (left). Changes in the X-ray energy at μ = 0.6 eV and CO concentration in gas phase (right). (C) *Operando* IR measurements for the Pt(3)/Rb(1)-Ba(1)-Mo(0.6)-Nb(0.2)/TiO_2 catalyst conducted under a sequential flow of CO_2, He, H_2, and 25% CO_2 + 75% H_2 at 250 °C (left). Variations in the intensities of the peaks related to the surface adsorbed species and concentration of CO in the effluent gas upon the introduction of CO_2 (right).

References and Notes

1. A. H. Motagamwala, R. Almallahi, J. Wortman, V. O. Igenegbai, S. Linic, Stable and selective catalysts for propane dehydrogenation operating at thermodynamic limit. *Science.* **373**, 217–222 (2021).
2. S. Pablo-García, A. Sabadell-Rendón, A. J. Saadun, S. Morandi, J. Pérez-Ramírez, N. López, Generalizing Performance Equations in Heterogeneous Catalysis from Hybrid Data and Statistical Learning. *ACS Catal.* **12**, 1581–1594 (2022).
3. Y. Wang, J. Kalscheur, Y. Q. Su, E. J. M. Hensen, D. G. Vlachos, Real-time dynamics and structures of supported subnanometer catalysts via multiscale simulations. *Nat. Commun.* **12**, 5430 (2021).
4. S. Hu, W.-X. Li, Sabatier principle of metal-support interaction for design of ultrastable metal nanocatalysts. *Science (80-.).* **374**, 1360–1365 (2021).
5. Z. W. Ulissi, A. J. Medford, T. Bligaard, J. K. Nørskov, To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **8**, 14621 (2017).
6. L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko, P. Nachtigall, Towards operando computational modeling in heterogeneous catalysis. *Chem. Soc. Rev.* **47**, 8307–8348 (2018).
7. K. McCullough, T. Williams, K. Mingle, P. Jamshidi, J. Lauterbach, High-throughput experimentation meets artificial intelligence: A new pathway to catalyst discovery. *Phys. Chem. Chem. Phys.* **22**, 11174–11196 (2020).
8. E. J. Ras, G. Rothenberg, Heterogeneous catalyst discovery using 21st century tools: a tutorial. *RSC Adv.* **4**, 5963–5974 (2014).
9. J. R. Kitchin, Machine learning in catalysis. *Nat. Catal.* **1**, 230–232 (2018).
10. L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, Z. Ulissi, Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.* **11**, 6059–6072 (2021).
11. M. Erdem Günay, R. Yıldırım, Recent advances in knowledge discovery for heterogeneous catalysis using machine learning. *Catal. Rev.* **63**, 120–164 (2021).
12. V. Fung, G. Hu, P. Ganesh, B. G. Sumpter, Machine learned features from density of states for accurate adsorption energy prediction. *Nat. Commun.* **12**, 88 (2021).
13. R. Schmack, A. Friedrich, E. V. Kondratenko, J. Polte, A. Werwatz, R. Kraehnert, A meta-analysis of catalytic literature data reveals property-performance correlations for the OCM reaction. *Nat. Commun.* **10**, 441 (2019).
14. J. A. Esterhuizen, B. R. Goldsmith, S. Linic, Uncovering electronic and geometric descriptors of chemical activity for metal alloys and oxides using unsupervised machine learning. *Chem Catal.* **1**, 923–940 (2021).
15. S. H. Wang, H. S. Pillai, S. Wang, L. E. K. Achenie, H. Xin, Infusing theory into deep learning for interpretable reactivity prediction. *Nat. Commun.* **12**, 5288 (2021).
16. C. Wulf, M. Beller, T. Boenisch, O. Deutschmann, S. Hanf, N. Kockmann, R. Kraehnert, M. Oezaslan, S. Palkovits, S. Schimmler, S. A. Schunk, K. Wagemann, D. Linke, A Unified Research Data Infrastructure for Catalysis Research – Challenges and Concepts. *ChemCatChem*. **13**, 3223–3236 (2021).
17. A. Mazheika, Y. G. Wang, R. Valero, F. Viñes, F. Illas, L. M. Ghiringhelli, S. V. Levchenko, M. Scheffler, Artificial-intelligence-driven discovery of catalyst genes with application to CO_2 activation on semiconductor oxides. *Nat. Commun.* **13**, 419 (2022).
18. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science. *Nature*. **559**, 547–555 (2018).
19. B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams, A. G. Doyle, Bayesian reaction optimization as a tool for chemical synthesis. *Nature*. **590**, 89–96 (2021).
20. N. I. Rinehart, A. F. Zahrt, J. J. Henle, S. E. Denmark, Dreams, False Starts, Dead Ends, and Redemption: A Chronicle of the Evolution of a Chemoinformatic Workflow for the Optimization of Enantioselective Catalysts. *Acc. Chem. Res.* **54**, 2041–2054 (2021).
21. A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, R. Q. Snurr, Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*. **4**, 1578–1597 (2021).
22. K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. WooPark, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong, C. Wolverton, Recent Advances and Applications of Deep Learning Methods in Materials Science. *npj Comput. Mater.* **8**, 59 (2022).

23. T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **10**, 2260–2297 (2020).
24. R. J. Murdock, S. K. Kauwe, A. Y. T. Wang, T. D. Sparks, Is Domain Knowledge Necessary for Machine Learning Materials Properties? *Integr. Mater. Manuf. Innov.* **9**, 221–227 (2020).
25. S. K. Kauwe, J. Graser, R. Murdock, T. D. Sparks, Can machine learning find extraordinary materials? *Comput. Mater. Sci.* **174**, 109498 (2020).
26. S. Mine, M. Takao, T. Yamaguchi, T. Toyao, Z. Maeno, S. M. A. Hakim Siddiki, S. Takakusagi, K. Shimizu, I. Takigawa, Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts. *ChemCatChem*. **13**, 3636–3655 (2021).
27. S. Mine, Y. Jing, T. Mukaiyama, M. Takao, Z. Maeno, K. Shimizu, I. Takigawa, T. Toyao, Machine Learning Analysis of Literature Data on the Water Gas Shift Reaction toward Extrapolative Prediction of Novel Catalysts. *Chem. Lett.* **51**, 269–273 (2022).
28. A. Y. T. Wang, S. K. Kauwe, R. J. Murdock, T. D. Sparks, Compositionally restricted attention-based network for materials property predictions. *npj Comput. Mater.* **7**, 77 (2021).
29. A. R. Falkowski, S. K. Kauwe, T. D. Sparks, Optimizing Fractional Compositions to Achieve Extraordinary Properties. *Integr. Mater. Manuf. Innov.* **10**, 689–695 (2021).
30. M. D. Porosoff, B. Yan, J. G. Chen, Catalytic reduction of CO_2 by H_2 for synthesis of CO, methanol and hydrocarbons: Challenges and opportunities. *Energy Environ. Sci.* **9**, 62–73 (2016).
31. W. Zhang, D. Ma, J. Pérez-Ramírez, Z. Chen, Recent Progress in Materials Exploration for Thermocatalytic, Photocatalytic, and Integrated Photothermocatalytic CO_2 ‐ to ‐ Fuel Conversion. *Adv. Energy Sustain. Res.* **3**, 2100169 (2022).
32. S. Mine, T. Yamaguchi, K. W. Ting, Z. Maeno, S. M. A. H. Siddiki, K. Oshima, S. Satokawa, K. Shimizu, T. Toyao, Reverse water-gas shift reaction over $Pt/MoO_x/TiO_2$: reverse Mars-van Krevelen mechanism via redox of supported MoO_x. *Catal. Sci. Technol.* **11**, 4172–4180 (2021).
33. P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).
34. J. A. Esterhuizen, B. R. Goldsmith, S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis. *Nat. Catal.* **5**, 175–184 (2022).
35. S. M. Lundberg, S. I. Lee, in *Advances in Neural Information Processing Systems* (2017), pp. 4765–4774.
36. S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K. W. Low, S. F. Newman, J. Kim, S. I. Lee, Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).
37. L. F. Bobadilla, J. L. Santos, S. Ivanova, J. A. Odriozola, A. Urakawa, Unravelling the Role of Oxygen Vacancies in the Mechanism of the Reverse Water-Gas Shift Reaction by Operando DRIFTS and Ultraviolet-Visible Spectroscopy. *ACS Catal.* **8** 7455−7467 (2018).
38. A. V. Mironenko, D. G. Vlachos, Conjugation-Driven “reverse Mars-van Krevelen”-Type Radical Mechanism for Low-Temperature C-O Bond Activation. *J. Am. Chem. Soc.* **138**, 8104–8113 (2016).
39. O. V. Safonova, M. Tromp, J. A. Van Bokhoven, F. M. F. De Groot, J. Evans, P. Glatzel, Identification of CO adsorption sites in supported Pt catalysts using high-energy-resolution fluorescence detection X-ray spectroscopy. *J. Phys. Chem. B*. **110**, 16162–16164 (2006).

Acknowledgments:

Funding: This study was financially supported by KAKENHI (21K18185 and 22K14538) from the Japan Society for the Promotion of Science (JSPS), JST-CREST Program JPMJCR17J3, JST-FOREST Program JPMJFR211U, and the Joint Usage/Research Center for Catalysis. G.W. acknowledges the JSPS postdoctoral fellowship (P20345).

Author contributions:

Conceptualization: I.T., K.S., and T.T.

Methodology: S.M., M.T., and I.T.

Investigation: G.W., S.M., D.C., Y.J, K.W.T., T.Y., Z.M., and K.M.

Funding acquisition: K.S. and T.T.

Supervision: I.T., K.S., and T.T.

Writing – original draft: G.W., S.M., and D.C.

Writing – review & editing: Y.J, K.W.T., Z.M., I.T., K.M., K.S., and T.T.

Competing interests: The authors declare no competing interests.

Data and materials availability: All code and data can be found online at https://github.com/shinya-mine.

Supplementary Materials

Methods
Figs. S1 to S32
Tables S1 to S9
References