ARTICLE

High-Throughput Screening and Literature Data Driven Machine Learning Assisting Investigation of Multi-component La_2O_3-based Catalysts for Oxidative Coupling of Methane

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Shun Nishimura,*[a] Son Dinh Le,[a] Itsuki Miyazato,[b] Jun Fujima,[b] Toshiaki Taniike,[a] Junya Ohyama,[c] and Keisuke Takahashi*[b]

Multi-component La_2O_3-based catalysts for oxidative coupling of methane (OCM) were designed based on high-throughput screening (HTS) and literature datasets with multi-output machine learning (ML) approaches including random forest regression (RFR), support vector regression (SVR), Gaussian process regression (Bayesian), and Item set mining (LCM). Combined use of HTS data and SVR successively assisted the finding of multi-component La_2O_3-based OCM catalysts of 11 types in 20 validations with C_2 yields appearing at 450°C based on indirect ML assistance. The appropriate multi-component predicted from ML contributes to determination of a characteristic feature of the lower onset temperature for a La_2O_3-based OCM catalyst. The LCM application on the SVR extended HTS data area supports spotting of the effective elements in the HTS area. However, a challenging subject remains: multi-component La_2O_3-based catalysts of two types afford effective C_2 yield (>5.0%) at 450°C, as inferred from the 20 selected types of catalyst validation. To predict unique multi-component La_2O_3-based OCM catalysts further, a combination of HTS and literature data was applied for four ML approaches. These were helpful to discover 17 additional combinations of multi-component La_2O_3-based catalysts affording effective C_2 yield (>5.0%) at 450°C in the 38 selected types of predictions. Completely, 30 multi-component La_2O_3-based catalysts of new types with C_2 yield greater than 5.0% at 450°C in CH_4/O_2 = 2.0 condition were found based on the indirect ML assistance driven by HTS and literature data.

1. Introduction

The recently increased availability of shale gas and methane hydrate have cast strong attention on the industrial transformation expected to derive from natural gas conversion to value-added petrochemicals such as ethylene. Oxidative coupling of methane (OCM) discovered in the 1980s[1–2] has received much attention because of its high potential to afford C_2 hydrocarbons of C_2H_4 and C_2H_6 directly from CH_4.[3–4] Several assumptions must be made of the demand of OCM catalyst performance for cost-effective design of industrial plant instead of naphtha cracking: these assumed C_2 yields are higher than 18% with C_2 selectivity higher than 80%.[5–7] Other research groups have used modeling and kinetic studies to estimate the limit of C_2 yield in the OCM reaction as 28–35% at maximum, in accordance with thermodynamic investigations.[8–10] According to these proposals, research targets for OCM catalyst development have been carried out to achieve a higher C_2 yield than 30% and/or significant C_2 selectivity of 80% with appropriate C_2 yield for the past 40 years. Aiming at this target, not only catalyst investigation but also reactor design concepts[11–12] and optimization of reaction conditions[13–14] have been widely reported.[15–18]

With the intense growth of data analysis and predictions intended to reveal hidden trends and undiscovered catalysts, machine learning (ML) assisted catalyst chemistry has received much attention.[19–21] The pursuit has been designated as catalyst informatics. Because conventional approaches to overcoming industrial demands for OCM have remained challenging,[15–18] the authors believe that an informatics approach has great potential as a game changer. Indeed, in just the last a few years, the combined uses of tailor-made experimental datasets obtained with a high throughput screening (HTS) machine and data analytic techniques such as multi-output ML and network profiling, which have presented new avenues for the design and elucidation of catalysts' nature and catalysis, in particular OCM.[22–32]

Several parallel reactor systems for effective data collection in a fixed bed catalytic reaction have become commercially available[33–34] and others have been proposed in research reports[22, 35–37]. Nevertheless, these still require special skills and entail high costs for operation and construction similarly to conventional modes of catalyst investigation. It will take some

[a] Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi 923-1292, Japan
[b] Department of Chemistry, Hokkaido University, N-10 W-8, Sapporo 060-0810, Japan
[c] Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: Sequence of reaction, OCM performances as the plots of C_2 yield and C_2 selectivity, True and predicted C_2 yield, List of chemicals, ML prediction results, and All data generated during this study. See DOI: 10.1039/x0xx00000x

time for tailor-made HTS dataset-based ML to become popular. Those situations engender common interests among experimental scientists to ascertain how those HTS datasets obtained by different researchers can assist their own discoveries in a laboratory with different reactors. Regarding ML engineering in pursuit of the prediction of heterogeneous catalysts, selecting adaptive descriptors to follow the target catalysis, to develop adequate algorithm refinement, and to expand the prediction area to a range beyond the area of collected data is an area that is still being investigated.[38–40] In other words, without ascertaining the hidden layers associated to the catalyst performance between the input layer (e.g. elemental number) and the output layer (C_2 yield, in this study), the prediction of catalyst activity derived only from ML regression is applicable only in a limited area at present. Barring development of a new mode of insight for ML methodology or confirmation of the contents of hidden layers, a common interest for data scientists and ML engineers is how experimental scientists should apply ML methodology to advance catalyst chemistry.

Reportedly, La_2O_3-based OCM catalysts possess the interesting character of OCM activity at lower temperatures. These catalysts exhibit a lower onset temperature in OCM. For instance, $La_2O_3CO_3$ nanorods serve OCM activity starting from 410–420°C.[41–42] The crystal shape of La_2O_3 plays a crucially important role on such low onset performance.[43] Nanocomposites of La_2O_3 and CeO_2 exhibit markedly good performance starting from 450–600°C.[44–47] Since the reaction at high temperature lead to opportunities not only to produce C_2 but also to decompose to CO/CO_2 and H_2O suppressing the selectivity of C_2 in OCM, such a unique nature derived from La_2O_3-based catalysts has motivated the design of new categories of OCM catalysts other than $NaMnW/SiO_2$, which is well known as a higher-temperature OCM catalyst operating at around 800°C. Various La_2O_3-based OCM catalysts were investigated in earlier studies, but the study on multi-component La_2O_3-based catalysts is rare.[36] Therefore, in this study, the authors specifically examine the discovery of multi-component La_2O_3-based OCM catalysts using HTS and literature data, combined with the use of multi-output ML approaches including random forest regression (RFR), support vector regression (SVR), Bayesian inference (Bayes), and itemset miner (LCM). This study was conducted based on the assumption that the M1-M2-M3 combinations serving high C_2 yield predicted by ML would be helpful for ascertaining the appropriate combination enhancing unique lower-temperature OCM performance of La_2O_3 itself. Such indirect ML-assisted investigations (**Table 1**) have supported research motivations particularly addressing multi-component La_2O_3-based OCM catalysts of 75 types for preparation and validation, and then multi-component La_2O_3-based OCM catalysts of 30 types acting at lower temperatures. Particularly, C_2 yields higher than 5.0%, detected at 450°C in CH_4/O_2 =2.0 condition, were obtained from this study.

2. Experiment

2.1. Catalyst preparation

All chemical information is listed in **Table S1** of Electronic Supplementary Information (ESI). Multi-component La_2O_3-based catalysts ($M1\text{-}M2\text{-}M3/La_2O_3$) were prepared using a co-impregnation method, which was dominant in the resources of HTS and literature datasets. First, 1.0 g of La_2O_3 is dispersed into an aqueous solution of mixed element resources (6 mL). They were then stirred at 50°C for 6 h (NHP-B077; Nissin Rika) under vigorous stirring (AMG-H; ASH Co., Ltd.). The nominal loading amount for each metal was fixed as 0.3 wt% in theory according to our earlier report.[48] Thereafter, the resulting slurry was centrifuged and evaporated at 80°C (CVE-3110; EYELA – Tokyo Rika Kikai Ltd.), and dried at 110°C for 2 h (ON-300S, ETTAS; As One Corp.). After grinding the powder, it is calcined at 600°C for 3 h at the ramping rate of 10°C min^{-1} in a furnace (3000 plus; KDF – Denken Highdental Co. Ltd.). Finally, the obtained powder was ground further and kept in a glass bottle. The bare La_2O_3 was also treated using the same protocol without metal resources. It was designated as $none/La_2O_3$.

2.3. Evaluating OCM reactivity

The reaction was conducted with a step-jointed quartz tube (4 mm ID, 235 mm L (top); 2 mm ID, 150 mm L (bottom)). The powder catalyst (50 mg) was stabilized at the jointed position of the reactor tube using quartz glass wool (<10 mg). It was then heated in a furnace (L = 270 mm, ARF-30KC; Asahirika Co. Ltd., Japan). The furnace temperature was monitored and controlled at the outer space close to the jointed portion using a type R

Table 1 Schematic illustration of approaches used for this study

	Validation 1	Validation 2	Validation 3-6
Data source	HTS data of 59 + 300 catalysts	HTS data of 300 catalysts	HTS data of 300 catalysts and Literature data
Number of data points	40,330 (350 cat.)[a]	27,622 (291 cat.)[a]	27,622 (291 cat.)[a] + 1,802 (1,286 cat)
ML	SVR	LCM in SVR	RFR, SVR, Bayesian-1 and Bayesian-2
Number of validations[b]	top 20 cat.	selected 20 cat.	top 12, top 12, top 7, and top 7 cat.
Number of novel catalysts[c]	11	2	9, 2, 1, and 6[d]

[a] Data points derived from 9 cat. are removed owing to extraordinary trends in selectivity. [b] When the same component appears in top C_2 catalyst, it is not counted in each validation. [c] Definition: C_2 yield is higher than 5.0% at 450°C in the CH_4/O_2 = 2.0 condition. [d] One catalyst of GaSrYb is double-counted between Validation 4 and 6.

thermal sensor connected to the control unit (Model SU; Chino, Japan). The catalyst was pre-treated at 400°C for 1 h under O_2 flow (2.0 × 10 ml min^{-1}). It was then purged using N_2 (2.0 × 10 ml min^{-1}) for 3 min. After introduction of reaction gas composed with $CH_4/O_2/N_2$ (Total flow: 2.1 × 10 ml min^{-1} including N_2 (3.0 ml min^{-1}, fixed) as an internal standard) for 5 min, the temperature program was started from 400–850°C in 50°C intervals. It was held at each temperature for 17 min to analyze the product by microGC, followed by 5 min for ramping toward the next temperature (**Figure S1**, in ESI). After passing through the catalyst bed (downstream) and a successive ice trap of water, the reaction mixture was analyzed for 3 min (1st sampling) and 10 min (2nd sampling) at each temperature (Micro GC FusionTM; INFICON) with a Rt-Molsieve 5A column (0.25 mm × 10 m, Ar carrier gas) and an Rt-U-Bond column (0.25 mm × 8 m, He carrier gas). Under the present conditions, standard Na-Mn-W/SiO_2, which is preferred as a higher-temperature OCM catalyst, gives C_2 yields of 24.8%, 17.7%, and 18.1% at 800°C under CH_4/O_2 = 2.0, 3.5 and 5.0 conditions, respectively, as shown in **Figure S2** in ESI. It is noteworthy that almost all catalysts show similar C_2 yield values between first and second samplings (standard deviation is <0.3 as usual, as explained in ESI); the discussed OCM performance here can be maintained for at least 10 min. The second sampling data are discussed mainly in the manuscript if there are no annotations.

2.3. Machine Learning

The HTS and literature data were collected respectively from our earliest reports on 59 + 300 catalysts[22,26] and a commonly used OCM review[15]. It is noteworthy that HTS datasets collected with our tailor-made machine are distinguished from the data obtained by different teams, denoted as "HTS data" in this paper. From these datasets, catalyst composition, reaction temperature, gas flow, and the corresponding C_2 yield value were selected for in-put and the target in the ML. ML using scikit-learn was implemented.[49] Particularly, SVR, RFR, and Gaussian process regression were used. Three datasets were prepared for machine learning. The first includes 291 and 59 high throughput catalysts data with various experimental conditions resulting total 40,330 HTS data points.[22,26] In this dataset, SVR is implemented where the radial basis functional kernel has C = 50 and gamma = 0.05 for which a cross-validation study is shown in **Figure S3(a)** in ESI.

The second dataset consists of 291 high-throughput catalysts data with various experimental condition resulting total 27,622 data points.[26] Common physical rule in predicted top C_2 yield catalysts is explored by item set mining. Item set mining is implemented using LCM47 to find common rules for elements of catalysts compositions found in SVR analysis for which a cross-validation study is shown in **Figure S3(b)** in ESI. Then, the element calcification (actinoid, alkali, etc.) and the group of each element are regarded as properties for determining items of catalysts. First, items representing expressions of inequality are defined for each property p. Here the following items are defined: -0.25<=p, p<0.0, 0.0<=p, p<0.25, 0.25<=p, p<0.5, 0.5<=p, p<0.75, 0.75<=p, p<1.0, 1.0<=p. Next, the atomic fractions of classified categories are calculated for the respective elements of the respective catalysts. For example, Ba-Ca-Y is represented as a list of property values (actinoid: 0, alkali: 0, alkaine: 0.666667, ..., metal: 1.0, ..., g_2: 0.666667, g_3: 0.333333, ...). Finally, items representing the target catalyst are defined. For the case in which Ba-Ca-Y is (actinoid<0.25, actinoid>=0.0, chalcogen<0.25, chalcogen>=0.0, g_11<0.25, g_11>=0.0, g_13<0.25, g_13>=0.0, g_14<0.25, g_14>=0.0, g_15<0.25, g_15>=0.0, g_16<0.25, g_16>=0.0, g_17<0.25, g_17>=0.0, g_18<0.25, g_18>=0.0, g_5<0.25, g_5>=0.0, g_7<0.25, g_7>=0.0, g_8<0.25, g_8>=0.0, g_9<0.25, g_9>=0.0, halogen<0.25, halogen>=0.0, metal<1.25, metal>=1.0, metalloid<0.25, metalloid>=0.0, noble_gas<0.25, noble_gas>=0.0, post_transition_metal<0.25, post_transition_metal>=0.0). In this way, each catalyst is converted to list of items. Then LCM is used for extracting common rules among catalysts.

The third dataset is prepared by combining HTP data with data obtained from the literature,[15] which comprises 1,802 data points (1,286 catalysts) obtained under various experiment conditions. Also, ML of three types are used: RFR, SVR, and Gaussian process regression. The number of trees is set to 100 within RFR. A radial basis functional kernel with C = 50 and gamma = 0.05 is used for SVR. A cross-validation study is shown in **Figure S3(c-d)** in ESI. Constant, radial basis functional and white kernels were used for Gaussian process regression. Here, Gaussian process regression was performed with two datasets where the C_2 yields are divided by 100 and 10, respectively, to standardize the data.

Please note that one-hot encoding is applied to represent the catalyst composition and support where binary number 1 and 0 are assigned. Descriptors for all ML are set to catalyst composition via one-hot encoding, temperature, P_{CH4}, P_{O2}, and Pinert (inert gas), where the total of P_{CH4}, P_{O2}, and Pinert is set as 1. Also, in accordance with the experimental scale in a conventional reactor for validations, the predicted top C_2 yield catalyst is selected under the following conditions: the flow rate is higher than 20 ml min^{-1}, $CH_4 + O_2$ conc. is over 85 vol%, and CH_4/O_2 = 2.0.

3. Results and Discussions

3.1. HTS dataset driven ML assistance for investigation of multi-component La_2O_3-based OCM catalyst

Based on SVR of the 40,330 HTS data points, 41 multi-component La_2O_3-based catalysts affording C_2 yield higher than 16.01% were predicted, as presented in the **Data 1.csv** file of ESI. The maximum predicted C_2 yield was 18.28% in the SVR-predicted La_2O_3-based catalyst area. According to our assumption, the predicted La_2O_3-based catalyst possessing high score in the predicted C_2 yield value is associated to adequate matching of multi-component or La_2O_3 support. That feature is expected to enhance the unique lower-temperature OCM feature. Then, the top 20 C_2 yield scoring catalysts are prepared from the list of **Data-1.csv**, in ESI. Their reactivities were validated in CH_4/O_2 = 2.0. It is better to be noted that this

assumption would be contributed to escape the interpolation issue on the unvalidated catalyst because the catalyst library for 300 HTS was constructed based on random selection from either Ba, Ca, Ce, Co, Cs, Cu, Eu, Fe, Hf, K, La, Li, Ni, Mg, Mn, Mo, Na, Nd, Pd, Sr, Tb, Ti, V, W, Y, Zn, Zr or "none", with support picked up from either γ-Al_2O_3, BaO, CaO, CeO_2, La_2O_3, MgO, SiO_2, anatase-TiO_2 or ZrO_2. The results of validations are presented in **Table S8** of ESI. As shown in **Figure 1A**, the trend of lower-temperature OCM performance presents some interesting features. Unfortunately, no catalyst possessing C_2 yield appeared at 400°C. The bare La_2O_3 denoted as "none" provides a C_2 yield (11.2%) from 500°C under this condition, but it is inactive at 400°C and 450°C. Among these 20 catalysts, 11 catalysts provided C_2 yield at 450°C: YEuHf (13.1%), CaYHf (10.1%), CaYEu (15.1%), CaYLa (15.0%), CaYBa (13.6%), MgNdHf (9.5%), CaEuHf (13.2%), CaYNd (10.2%), SrNdHf (9.7%), CaNiY (6.1%), and MgEuHf (11.4%). The numbers in parentheses are C_2 yields at 450°C. Accordingly, the relationships between a high C_2 yield scoring catalyst in the predicted area and a lower onset feature catalyst are non-liner. However, the indirect ML-assisted discovery of lower-temperature multi-component La2O3 is apparently a useful idea. It is particularly interesting that the performance of CaNiY/La_2O_3 gives unique features in terms of C_2 composition under the present conditions. The C_2H_6 is the mother product in the C_2 component. The C_2H_4 yield is 0.3% at maximum, only in this case, as shown in **Figure S4(a)** of ESI. Moreover, the predicted C_2 yield higher than 16% in ML (**Data 1.csv**, in ESI) is scarcely obtained, even for data surveyed at high temperatures, as shown in **Table S8** of ESI. This point is unimportant for the present study because the metal loading, calcination temperature, and reactor design differ from those of the HTS machine.[22,26]

Because the HTS data-derived indirect ML nicely assisted the prediction of lower-temperature La_2O_3-based OCM catalysts (11 appearances/20 validations), item set mining with LCM[48,50] is further implemented to ascertain common physical rules within highly active catalysts in the SVR space of HTS datasets toward new ideas. Using this approach, 300 random catalyst-based HTS datasets[26,‡] were used, excluding 59 catalyst-based HTS datasets. The 59 catalyst HTS datasets mainly comprised data for Na-Mn-W-related catalysts. Therefore, utilization of 59 + 300 catalyst-based HTS data package would provide worse results: such bias in the 59 catalyst library guides the mining focusing on those areas. Furthermore, the LCM suggested multi-component catalyst variations of 165 types in accordance with the common role, as presented in **Tables S2–S3**, in ESI. This approach can choose the 11 elements of Cr, Ni, Zn, Y, Zr, Pd, Cs, Ce, Sm, Eu, and Hf in the periodic table as the effective component for high C_2 yield in OCM. However, combinations of these M1-M2-M3 are apparently nonsense in their present form: i.e., combinations of 3 element selections from 11 varieties (denoted as $_{11}C_3$ mathematically) are to be 165 variations at maximum. This finding might be attributable to the fact that the numbers of catalyst variations (300) for construction of SVR space is insufficient for LCM identification in multi-component M1-M2-M3 categories. From the viewpoint of conventional knowledge of catalyst scientists, this proposed

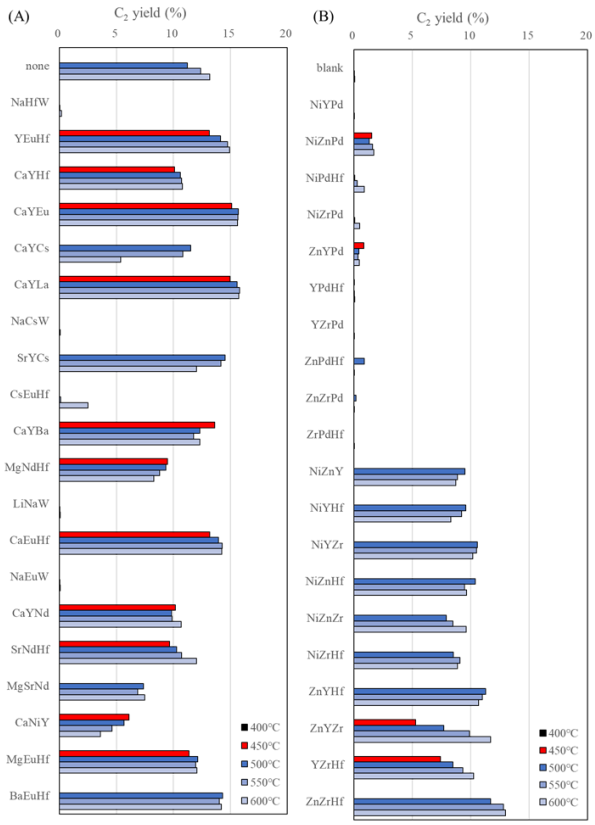

Figure 1 Plots of C_2 yields over multi-component La_2O_3 of 20 types prepared based on (A) SVR of HTS data and (B) ICM approach, together with none/La_2O_3 and blank test in CH_4/O_2 = 2.0. All raw data are presented in **Tables S7 and S8** (in ESI).

list produces some strange feelings in some cases. Fore examples, the LCM-proposed multi-component OCM catalysts for higher C_2 yields included Pd and Ni. Actually, combinational reports describe that Pd prefers CH_4 combustion,[52] whereas Ni is familiar for CH4 reforming and carbon deposition.[53] Indeed, in our earlier classifications,[22,26,48] these two elements were placed in the "worse" element set for OCM. Regarding the interesting feeling for such mismatching, the LCM tentatively suggested catalysts of 20 kinds including Pd, Ni, and some other elements, which were validated for reactivity as listed in **Table S9**, in ESI. Unfortunately, **Figure 1B** shows that the probability of hitting the target of lower-temperature OCM with LCM is lower than that with SVR approach, as shown in **Figure 1A**, however multi-component catalysts of two kinds, ZnYZr and YZrHf, serve C2 yield at 450°C higher than 5.0%: 5.3% and 7.4% are respectively obtained (2 appearances/tentatively selected 20 validations).

Among the La_2O_3-based catalysts with the CH_4/O_2 = 2.0 condition, the C_2 selectivity is limited to a maximum of approx. 40% (**Figure S5**, in ESI), which is much lower than the 70% limit of Na-Mn-W/SiO_2 (**Figure S2(a)**, in ESI) under the same conditions. The bare La_2O_3 denoted as none/La_2O_3 also shows 40% selectivity as maxima (**Table S5**, in ESI). Therefore, this potential is expected to be attributable to the nature of La_2O_3 support itself, which serves a high O_2 conversion value, even at lower temperatures. To overcome this difficulty, the CH_4/O_2

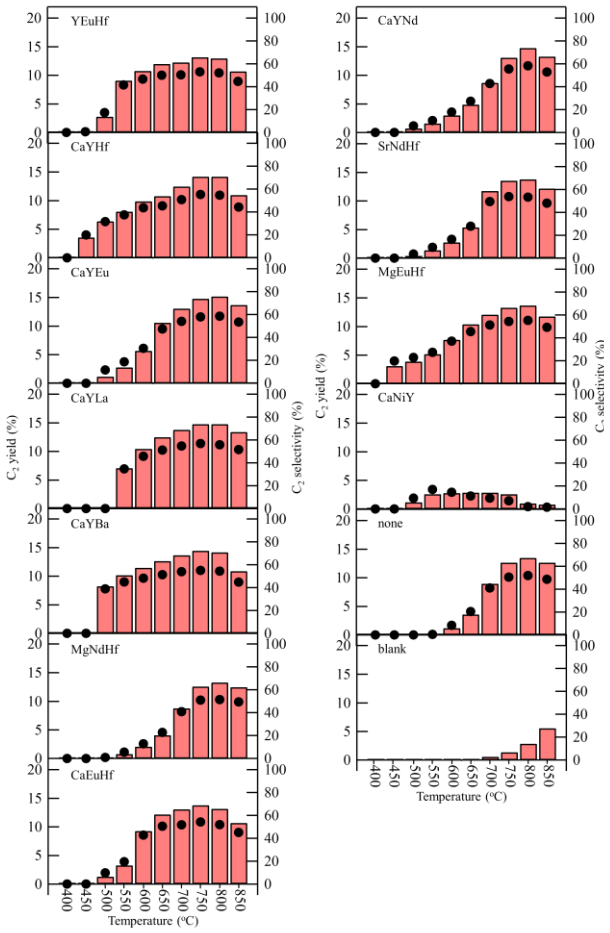

Figure 2 Plots of C_2 yield (bar) and C_2 selectivity (closed sphere) of selected catalysts predicted from HTS data-driven ML at CH_4/O_2 = 3.5. Raw data are listed in **Table S9** (in ESI).

ratio rises to 3.5 to a CH_4-rich condition with same CH_4 + O_2 conc., which is more difficult situation for CH_4 and/or C_2 combustion. The 11 catalysts, including the Top 10 C_2 yield value (>9.0%) at 450°C in CH_4/O_2 = 2.0 condition were further investigated for their potential usage as an OCM catalyst: YEuHf, CaYHf, CaYEu, CaYLa, CaYBa, MgNdHf, CaEuHf, CaYNd, SrNdHf, and MgEuHf, and a unique CaNiY. These results are shown in **Figure 2 and Tables S10**, in ESI. Bare La_2O_3 and blank data are presented in the same figure. The onset temperatures of La_2O_3 itself rise to 600°C in the CH_4/O_2 = 3.5 condition. The two catalysts of CaYHf (C_2 yield: 3.6%) and MgEuHf (3.1%) show OCM activity at 450°C even in the CH_4/O_2 = 3.5 condition. In fact, five catalysts are active at 500°C: YEuHf (2.8%), CaYEu (1.2%), CaYBa (8.3%), CaEuHf (1.3%), and CaNiY (1.2%). Only one catalyst of CaYLa (7.1%) achieves nice performance at 550°C. Even though C_2 yield becomes very low in the cases of MgNdHf (0.1–0.8%), CaYNd (0.7–1.6%), and SrNdHf (0.4–1.4%), these have only slight potential at temperatures below 550°C. Accordingly, the nature of multi-component catalysts serving the lower onset OCM temperatures than bare La_2O_3 remains in all 11 cases in the CH_4/O_2 = 3.5 condition. The temperature dependence of C_2 yield becomes greater in the condition of CH_4/O_2 = 3.5. It is expected that because methane is a chemically inert compound with high C–H bond strengths, low polarizability, and high ionization energy, its activation frequently requires higher operating temperatures at such a CH_4-rich condition. Indeed, increasing the reaction temperature increases the C_2 yield. The best performance in C_2 yield values observed at 800°C in almost all cases, except for CaNiY. The best C_2 yield was 15.2%, with 58.6% selectivity obtained over CaYEu at 800°C. Moreover, a further increase of CH_4/O_2 ratio to 5.0 strongly influences the increase of C_2 selectivity to be approx. 70%; it also boosts the onset temperature to a higher value, close to that of the bare one (**Figure S6**, in ESI). Finally, the catalyst performance of CaYBa affording C_2 yield of 13.9% with C_2 selectivity of 70.3% at 800°C in CH_4/O_2 = 5.0 condition is achieved in the La_2O_3-based OCM catalyst. Such performance discovered using the indirect ML approach is important in comparison with results of an earlier study of multi-component La_2O_3-based OCM catalysts.[36] The unique performance of CaNiY remains in both conditions of CH_4/O_2 = 3.5 and 5.0. Indeed, C_2H_6 is the main C_2 product (C_2H_4 yield <0.6%). However, the C_2 yield and C_2 selectivity do not improve at all (**Figures S4(b) and S4(c)**, in ESI).

3.2. HTS and literature dataset driven ML assistance for investigation of multi-component La_2O_3-based OCM catalyst

To investigate undiscovered multi-component La_2O_3-based OCM catalysts further, datasets including not only the HTS data but also data reported from the literature are applied to reveal the nice M1-M2-M3 combinations in the case of La_2O_3 support. According to our earlier work,[37] literature data might misguide the data-driven prediction because of bias derived from differences in reactor systems or specific methods, yet the authors hold that long historical examinations are helpful to inspire new targets for OCM catalyst design.[54–55] At the next step, two ML regressions of RFR and SVR are examined for ascertaining valuable multi-component combinations in the HTS and literature datasets. Scores of ML are not so high (**Figure S3(c–d)**, in ESI), however, it is noteworthy that, after addition of literature data to our HTS datasets, the area of element survey becomes much wider, comprising 63 elements of Ag, Al, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Ho, In, K, La, Li, Lu, Mg, Mn, Mo, Na, Nb, Nd, Ni, P. Pb. Pd, Pr, Pt, Rb, Re, Rh, Ru, Sb, Sc, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tm, V, W, Y, Yb, Zn, Zr, and "none", and 12 supports of Al_2O_3, SiO_2-Al_2O_3, BaO, CaO, Ca-Si-O_x, CeO_2, La_2O_3, MgO-Al_2O_3, MgO, SiO_2, TiO_2, and ZrO_2. The results of La_2O_3-based prediction are shown respectively in the files of **Data-2.csv and Data-3.csv** (in ESI). The RFR includes 44 variations of C_2 yield in the range of 21.01–21.53%, but the predicted multi-component La_2O_3 are composed mainly of the M-GaHf/La_2O_3 group. The numbers of appearances of Ga and Hf are, respectively, 44 and 37 variations. It is apparently in the nature of RFR, which can follow trends obediently in original datasets. Indeed, revision of HTS data implies that Hf is included in multiple candidates such as HfNaW/SiO_2 (16.0%), TbHfW/La_2O_3 (15.2%), TbHfW/La_2O_3 (15.2%), MgZrHf/CaO (15.9%), CeNdHf/BaO (15.2%), and NaEuHf/MgO (16.6%). All were found to be nice OCM catalysts

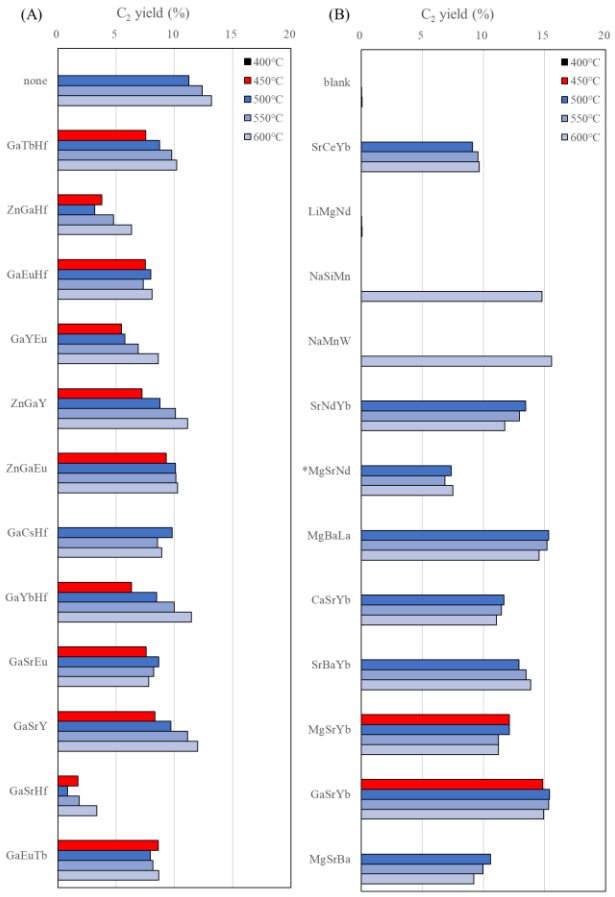

Figure 3 Plots of C_2 yields over multi-component La_2O_3 of 12 types prepared based on (A) RFR and (b) SVR of HTS and literature data, together with none/La_2O_3 and blank yield in CH_4/O_2 = 2.0. All raw data are listed in **Tables S11–S12** (in ESI).

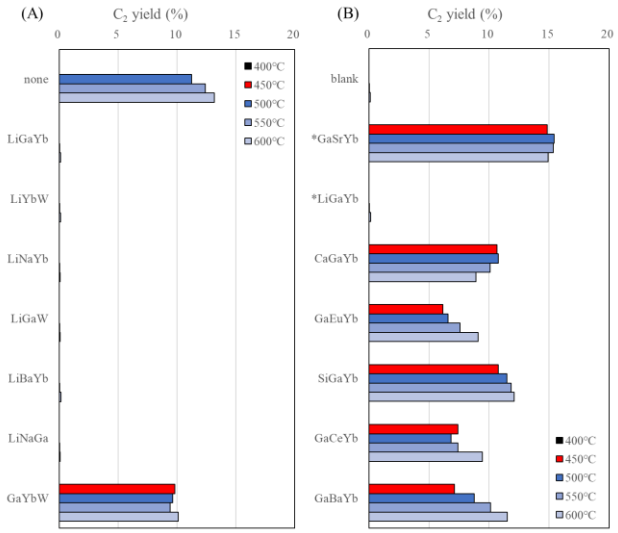

Figure 4 Plots of C_2 yield over each multi-component La_2O_3 of seven types prepared based on (A) Bayesian-1 and (b) Bayesian-2 of HTS and literature data, together with none/La_2O_3 and blank yield in CH_4/O_2 = 2.0. All raw data are listed in **Tables S13–S14** (in ESI).

from the HTS data, and all with 21.0% as the maximum C_2 yield. Moreover, in the data from the literature, extraordinarily high C2 yield (37.0%) achieved with BaGa catalyst was listed.15 These trends might guide the ML feature to these areas as highly frequent elements, leading to high scores in one-hot encoding identification. For SVR, 410 variations of C_2 yield in the range of 14.00–19.41% are predicted however same components with different C_2 values frequently appears. At this time, the most frequently appearing elements (Top 12) were the following: Sr (183), Yb (138), Nd (97), Na (95), Mg (86), Ba (84), Li (82), Ca (66), Zr(48), Ga (39), Eu (38), and La(34). Although Sr and/or Yb appearances are somewhat high, SVR is apparently constructed with a wider view than RFR. Multi-component La_2O_3-based catalysts of each Top 12 component in C2 yield value for RFR and SVR were demonstrated for preparation and evaluation for the OCM reactivity, respectively, from the list of **Data-2.csv and Data-3.csv**, in ESI. Actually, only the MgSrNd/La_2O_3 in the SVR-predicted catalyst is doubled with the catalyst predicted using SVR with HTS data. It is denoted with an asterisk in front (*). **Figures 3(A) and 3(B)** emphasize the C_2 yield below 600°C in the CH_4/O_2 = 2.0 condition. It is particularly interesting that these two ML assistances also serve OCM active multi-component La_2O_3-based catalysts of 13 types at 450°C. They are GaTbHf (C_2 yield: 7.5%), ZnGaHf (3.8%), GaEuHf (7.5%), GaYEu (5.5%), ZnGaY (7.2%), ZnGaEu (9.3%), GaYbHf (6.3%), GaSrEu (7.6%), GaSrY (8.3%), GaSrHf (1.8%), and GaEuTb (8.6%) in RFR, whereas MgSrYb (12.1%) and GaSrYb (14.9%) are in SVR assistance. Actually, RFR assistance provides larger quantities of La_2O_3-based catalysts (11 catalysts) than SVR (2 catalysts) in each of 17 validations. As discussed previously, the area in RFR is composed mainly from GaHf combinations. These elements are strongly attributable to the unique nature of lower-temperature OCM performance.

For challenging the ML-assisted discovery of La_2O_3-based OCM catalyst, Gaussian process regression (Bayesian) with different scales of compression view is further examined, denoted as Bayesian-1 and Bayesian-2. The Bayesian results of La_2O_3-based predictions are presented in the files of **Data-4.csv and Data-5.csv**, in ESI. Bayesian-1 gives 102 predictions at C_2 yields of 20.01–24.87%. This prediction includes particularly numerous appearances of Li (78 times), Ga (46), Yb (46), and Na (33), whereas the Bayesian-2 can suggest 85 predictions among 20.01%–24.10% with Yb (77 times) and Ga (58). It is apparent that major differences exist in the area of elements selected for multi-component prediction. The former is composed of 24 elements of B, Ba, Bi, Ca, Ce, Cs, Er, Eu, Ga, Gd, Hf, K, La, Li, Mg, Mn, Na, Nd, Si, Sm, Sr, Tb, W, and Yb, while the latter are from 49 elements of Ag, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Cr, Cs, Dy, Er, Eu, Ga, Gd, Ge, Hf, Ho, In, K, Li, Lu, Mg, Mn, Na, Nd, P, Pd, Pr, Rb, Rh, Ru, Sb, Sc, Si, Sm, Sr, Tb, Te, Th, Ti, Tm, V, W, Yb, Zn, and Zr. Accordingly, the Bayesian-2 approach possesses a wider view. Tentatively, the top 7 multi-component La_2O_3-based catalysts from each prediction are validated. It is noteworthy that LiGaYb and LaGaYb from Bayesian-2 are duplications with Bayesian-1 and SVR of the HTS and literature datasets, respectively. Overall results are presented in **Tables S14 and S15**, in ESI. As shown in **Figures 4**, differences in Bayesian-1 and Bayesian-2 were found

to be significant in terms of assistance to discover appropriate combinations for lower-temperature OCM activity. Only one catalyst of GaYbW provides C_2 yield of 9.8% at 450°C in the CH_4/O_2 = 2.0 condition, in Bayesian-1. By contrast six catalysts of GaSrYb, CaGaYb, CaEuYb, SiGaYb, GaCeYb, and GaBaYb were found to have C_2 yields of 14.9%, 10.7%, 6.1%, 10.8%, 7.4%, and 7.1% at 450°C, respectively, in Bayesian-2. Based on this trend in Bayesian-2, GaYb and Yb can be anticipated as positive factors, but Li is truly detrimental for lower-temperature OCM.

In addition to HTS data-driven ML-assisted investigation, low C_2 selectivity (<44%) at CH_4/O_2 = 2.0 was also observed in the investigation of La_2O_3-based catalysts with combined use of HTS and literature data (**Figures S7–S8**, in ESI). Therefore, selected multi-component La_2O_3 catalysts of six types with nice C_2 yield values (>9.0%) at 450°C were found: EuGaZn, MgSrYb, SrGaYb, GaWYb, CaGaYb, and SiGaYb. They were investigated further for their potential as OCM catalysts at CH_4/O_2 = 3.5 and 5.0 conditions as well. It is noteworthy that SrGaYb appeared in both cases of SVR and Bayesian-2. As shown in **Figure S9**, in ESI, an increased CH_4/O_2 ratio supports the improvement of C_2 selectivity while maintaining lower onset nature compared to the bare La_2O_3 were obtained. The C_2 yield and C_2 selectivity were found to depend strongly on the reactor temperature, and the best performance on C_2 production was observed in MgSrYb as approx. 13.0% yield, with approx. 55–60% selectivity at 800°C. Although this finding shows similar performance to that of none/La_2O_3, its unique lower onset nature is still beneficial for use as a multi-component La_2O_3-based catalyst.

Conclusions

In summary, beyond issues such as the direct ML prediction derived from the presence of hidden contributions (designated as the hidden layer) between the input layer (element component of catalyst) and output layer (OCM reactivity, C_2 yield) and different reactor design concept at each research groups, the investigation of multi-component La_2O_3-based OCM catalysts has been conducted at a conventional fixed-bed reactor with multi-output ML assistance using HTS and literature datasets. By specifically examining the unique features derived from La_2O_3-based OCM catalysts, such as the induction of lower onset temperature, a hypothesis is made and validated through this study: M1-M2-M3 combinations serving high C_2 yield predicted by direct ML prediction are expected to be helpful for finding familiar multi-components on La_2O_3 support, consequently enhancing the unique nature of La_2O_3 itself. To support this position, the authors selected multi-component La_2O_3-based OCM catalysts of 75 types with indirect ML assistance. Results indicated that multi-component La_2O_3-based OCM catalysts of more than 30 types work as lower-temperature OCM catalysts. Particularly, C_2 yields higher than 5.0% were detected at 450°C in a CH_4/O_2 = 2.0 condition. This finding can promote the next study on the role of each element and mechanism,§ and the optimization of catalyst preparation and reaction condition, as a lower-temperature OCM catalyst on these new catalysts. This study represents one successful style in the ML assistance of a new heterogeneous catalyst discovery.§§

Author Contributions

S.N. and K.T. conducted most of the experiments, discussed the concept of this study, analyzed the data, and wrote the manuscript. S.D.L. prepared the catalysts. T.T. examined HTS experiment and advised the approach of this study. J.O. and S.D.L. participated in discussions of catalyst performances and the data analysis. K.T., I.M. and J.F. performed the ML analysis including regressions and inferences. All authors contributed to the experimental design, discussed the results, and reviewed the manuscript.

Conflicts of interest

The authors declare that they have no competing financial interest.

Acknowledgements

This work is supported by Japan Science and Technology Agency (JST) CREST (grant no. JPMJCR17P2).

Notes and references

‡ The dataset is uploaded in a web platform Catalyst Acquisition by Data Science (CADS) for shared usage,[50] https://cads.eng.hokudai.ac.jp.

§ XRD and XPS were tentatively obtained in the catalysts examined at validation 1, however common trends for lower temperature OCM catalysis were not observed (**Figures S10 and S11**, and **Table S4**, in ESI)

§§ Data availability: All data generated during this study are included in this published article, the Electronic Supplementary Information (ESI). Additionally, all datasets are available free of charge in a web platform Catalyst Acquisition by Data Science (CADS) for shared usage, https://cads.eng.hokudai.ac.jp.

1 G. E. Keller and M. M. Bhasin, *J. Catal.*, 1982, **73**, 9-19.
2 W. Hinsen and M. Baerns, *Chem. Ztg.*, 1983, **107**, 223-226.
3 J. Haggin, J. *C&EN*, 1990, **68(1)**, 26-28.
4 A. H. Tullo, *C&EN*, 2014, 92(27), 20-21.
5 J. H. B. J. Hoebink, H. M. Venderbosch, P. C. van Geen, P. Y. van den Oosterkamp and G. B. Marin, *Chem. Eng. Technol.*, 1995, **18**, 12-16.
6 J. C. W. Kuo, C. T. Kresge, and R. E. Palermo, *Catal. Today*, 1989, **4**, 463-470.
7 M. J. Gradassi, and N. W. Green, *Fuel Proc. Technol.*, 1995, **42**, 65-83.
8 J. W. Thybaut, J. Sun, L. Olivier, A. C. van Veen, C. Mirodatos and G. B. Marin, *Catal. Today*, 2011, **159**, 29-36.
9 J. A. Labinger, *Catal. Lett.*, 1998, **1**, 371-376.
10 Y. S. Su, J. Y. Ying and W. H. Green, J. Catal., 2003, **218**, 321-333.
11 A. Cruellas, T. Melchiori, F. Gallucci and M. van Sint Annalang, *Catal. Rev.*, 2017, **59**, 234-294.
12 L. A. Vandewalle, R. V. de Vijver, K. M. A. Geen and G. B. Marin, *Chem. Eng. Sci.*, 2019, **198**, 268-289.
13 S. Sadjadi, S. Jašo, H. R. Godini, S. Arndt, M. Wollgarten, R. Blume, O. Görke, R. Schomäcker, G. Wozny and U. Simon, *Catal. Sci. Technol.*, 2015, **5**, 942-952.
14 J. Ohyama, S. Nishimura and K. Takahashi, *ChemCatChem*, 2019, **11**, 4307-4313.

15 U. Zavyalova, M. Holena, R. Schlogl and M. Baerns, *ChemCatChem*, 2011, **3**, 1935-1947.
16 E. V. Kondratenko, T. Peppel, D. Seeburg, V. A. Kondratenko, N. Kalevaru, A. Martin and S. Wohlrab, *Catal. Sci. Technol.*, 2017, **7**, 366-381.
17 R. Schmack, A. Friedrich, E. V. Kondratenko, J. Polte, A. Werwatz, and R. Kraehnert, *Nature Commun.*, 2019, **10**, 441.
18 S. Mine, M. Takao, T. Yamaguchi, T. Toyao, Z. Maeno, S, M. A. H. Siddiki, S. Takakusagi, K. Shimizu, and I. Takigawa, *ChemCatChem*, 2021, **13**, 3636-3655.
19 M. E. Gunay and R. Yildirim, *Catal. Rev.*, 2020, **63**, 120-164.
20 T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, and K. Shimizu, *ACS Catal.*, 2020, **10**. 2260-2297.
21 K. Takahashi, L. Takahashi, I. Miyazato, J. Fujima, Y. Tanaka, T. Uno, H. Satoh, K. Ohno, M. Nishida, K. Hirai, J. Ohyama, T. N. Nguyen, S. Nishimura, and T. Taniike, *ChemCatChem*, 2019, **11**, 1146-1152.
22 T. N. Nguyen, T. T. P. Nhat, K. Takimoto, A. Thakur, S. Nishimura, J. Ohyama, I. Miyazato, L. Takahashi, J. Fujima, K. Takahashi and T. Taniike, *ACS Catal.*, 2020, **10**, 921-932.
23 K. Takahashi, L. Takahashi, T. N. Nguyen, A. Thakur and T. Taniike, *J. Phys. Chem. Lett.*, 2020, **11**, 6819–6826.
24 K. Sugiyama, T. N. Nguyen, S. Nakanowatari, I. Miyazato, T. Taniike and K. Takahashi, *ChemCatChem*, 2021, **13**, 952-957.
25 I. Miyazato, T. N. Nguyen, L. Takahashi, T. Taniike and K. Takahashi, *J. Phys. Chem. Lett.*, 2021, **12**, 808-814.
26 T. N. Nguyen, S. Nakanowatari, T. P. N. Tran, A. Thakur, L. Takahashi, K. Takahashi and T. Taniike, *ACS Catal.*, 2021, **11**, 1797-1809.
27 S. Ishioka, I. Miyazato, L. Takahashi, T. N. Nguyen, T. Taniike, and K. Takahashi, *J. Comp. Chem.*, 2021, **42**, 1447-1451.
28 S. Nakanowatari, T. N. Nguyen, H. Chikuma, A. Fujiwara, K. Seenivasan, A. Thakur, L. Takahashi, K. Takahashi and T. Taniike, *ChemCatChem*, 2021, **13**, 3262-3269.
29 K. Takahashi, J. Fujima, I. Miyazato, S. Nakanowatari, A. Fujiwara, T. N. Nguyen, T. Taniike and K. Takahashi, *J. Phys. Chem. Lett.*, 2021, **12**, 7335–7341.
30 L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike and K. Takahashi, *Chem. Sci.*, 2021, **12**, 12546-12555.
31 T. N. Nguyen, K. Seenivasan, S. Nakanowatari, P. Mohan, T. P. N. Tran, S. Nishimura, K. Takahashi and T. Taniike, *Mol. Catal.*, 2021, **516**, 111976.
32 K. Chen, H. Tian, B. Li, S. Rangrajan, *AIChE J.*, *in press* (DOI: 10.1002/aic.17584).
33 C. Ortega, D. Otyuskya, E. Ras, L. D. Virla, G. S. Partience and H. Dathe, *Can J. Chem. Eng.*, 2021, **99**, 1288-1306.
34 Compact stand-alone fixed bed reactor Flowrence systems (Aventium) in REALCAT Webpage, https://www.realcat.fr/en/realcat-techniques-Equipments.html (accessible in 1st Dec., 2021); Avantium Web, https://www.catalysis.avantium.com (accessible in 1st Dec., 2021).
35 L. Olivier, S. Haag, H. Pennemann, C. Hogmann, C. Mirodatos and A. C. van Veen, *Catal. Today*, 2008, **137**, 80-89.
36 Z. Li, L. He, S. Wang, W. Yi, S. Zou, L. Xiao and J. Fan, *ACS Comb. Sci.*, 2017, **19**, 15-24.
37 S. Nishimura, J. Ohyama, T. Kinoshita, S. D. Le and K. Takahashi, *ChemCatChem*, 2020, **12**, 5888-5892.
38 T. Hattori and S. Kita, *Catal. Today*, 1995, **23**, 347-355.
39 A. Corma, J. M. Serra, P. Serna and M. Moliner, *J. Catal.*, 2005, **232**, 335-341.
40 R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi and C. Kim, *npj Comput. Mater.*, 2017, **3**, 54.
41 Y. Hou, W. Han, W. Xia and H. Wan, *ACS Catal.*, 2015, **5**, 1663-1674.
42 R. Feng, P. Niu, Q. Wang, B. Hou, L. Jia, M. Lin and D. Li, *Fuel*, 2022, **308**, 121848.
43 P. Huang, Y. Zhao, J. Zhang, Y. Zhu and Y. Sun, *Nanoscale*, 2013, **5**, 10844-10848.
44 Y. Zhang, J. Xu, X. Xu, R. Xi, Y. Liu, X. Fang and X. Wang, *Catal. Today*, 2020, **355**, 518-528.
45 D. Noon, A. Seubsai and S. Senkan, *ChemCatChem*, 2013, **5**, 146-149.
46 V. J. Ferreira, P. Tavares, J. L. Figueiredo and J. L. Faria, *Catal. Commun.*, 2013, **42**, 50-53.
47 Z. Zhang, Y. Gong, J. Xu, Y. Zhang, Q. Xiao, R. Xi, X. Xu, X. Fang and X. Wang, Catal. Today, *in press* (DOI: 10.1016/j.cattod.2021.11.012)
48 J. Ohyama, A. Hirayama, Y. Tsuchimura, N. Kondou, H. Yoshida, M. Machida, S. Nishimura, K. Kato, I. Miyazato and K. Takahashi, *Catal. Sci. Technol.*, 2021, **11**, 524-530.
49 F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Machine Learn. Res.*, 2011, **12**, 2825-2830.
50 T. Uno, M. Kiyomi and H. Arimura, LCM ver.3: collaboration of array, bitmap and prefix tree for frequent itemset mining. Proceedings of the 1st international workshop on opensource data mining: frequent pattern mining implementations, Chicago, Illinois, 2005, pp. 77-86.
51 J. Fujima, Y. Tanaka, I. Miyazato, L. Takahashi and K. Takahashi, *React. Chem. Eng.*, 2020, **5**, 903-911.
52 J. Chen, J. Zhong, Y. Wu, W. Hu, P. Qu, X. Xiao, G. Zhang, X. Liu, Y. Jiao, L. Zhong and Y. Chen, *ACS Catal.*, 2020, **10**, 10339-10349.
53 L. Xu, L. Xu, M. Chen, C. Wu, G. Cheng, N. Wang and X. Hu, *Catal. Sci. Technol.*, 2021, **11**, 6344-6368.
54 K. Takahashi, I. Miyazato, S. Nisimura and J. Ohyama, *ChemCatChem*, 2018, **10**, 3223-3228.
55 V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder and A. Jain, *Nature*, 2019, **571**, 95-98.