Title	特徴量設計に基づく機械学習を用いたメタン酸化カップリング触媒の開発
Author(s)	藤原, 綾
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19939
Rights	
Description	Supervisor: 谷池 俊明, 先端科学技術研究科, 博士

博士論文

Development of methane oxidative coupling catalysts with feature engineering-based machine learning

藤原　綾

主指導教員　谷池　俊明

北陸先端科学技術大学院大学

先端科学技術専攻

[マテリアルサイエンス]

令和７年３月

Referee-in-chief: Professor Toshiaki Taniike

Japan Advanced Institute of Science and Technology

Referees: Professor Yuki Nagao

Japan Advanced Institute of Science and Technology

Associate Professor Shun Nishimura

Japan Advanced Institute of Science and Technology

Associate Professor Kenta Hongo

Japan Advanced Institute of Science and Technology

Professor Akira Nakayama

The University of Tokyo

Abstract

Aya Fujiwara (2220027)

Materials Informatics (MI) aims to accelerate the discovery and understanding of materials through data science, especially in cases where complexity exceeds human intuition. However, there are two main challenges to implementing MI in practical applications: one is the lack of data that meets the quality, scale, and consistency required for effective machine learning, and the other is the need for advanced domain knowledge to design descriptors (numerical representations of catalysts). To address the data challenge, Taniike and his research group have developed their own datasets using high-throughput experimentation (HTE). Additionally, they introduced an automatic feature engineering (AFE) method that enables the design of descriptors without domain-specific knowledge, thereby addressing the descriptor challenge. This technique generates numerous features (descriptor candidates) from the physicochemical properties of catalyst components and extracts those that are relevant for describing catalyst performance.

The oxidative coupling of methane (OCM) is a catalytic reaction that directly synthesizes ethylene (C_2H_4) from methane (CH_4). This process is significantly more energy-efficient than conventional synthesis routes via syngas. However, due to the higher stability of CH_4 compared to C_2H_4, achieving high yields remains challenging, necessitating the development of high-performance catalysts. Taniike and his group have accumulated a dataset containing performance data for over 600 catalysts in OCM using their HTE system. Using this data as the training dataset, they have expanded the scope of catalyst exploration with machine learning techniques, such as decision tree analysis and support vector regression (SVR). These approaches, however, adopted one-hot encoding to represent catalyst compositions, where a catalyst component is represented as either 0 (element x is absent) or 1 (element x is present). The limitation of this method is that it treats catalyst compositions as symbols without providing insight. While researchers may recognize that catalysts A and B share common physical properties, the machine learning model perceives only common symbols. For efficient catalyst exploration, it would be ideal to define some physical features as a descriptor, however, achieving this in such a complex catalytic reaction is highly challenging. This illustrates a core challenge in MI: the need for a machine learning framework capable of generating hypotheses without requiring prior knowledge of the target system, particularly for the vast exploration of candidate materials.

From the perspective that "researchers derive insights from multiple observations," when observational data is limited, it becomes difficult to reject competing hypotheses. Thus, researchers conduct repeated control experiments to refine their hypotheses and arrive at accurate conclusions. This hypothesis refinement process is heavily reliant on a researcher's intuition, impacting both the speed and quality of research outcomes. If this process could be integrated into a machine learning framework and streamlined with high-throughput experimentation, it could establish a versatile research framework applicable to a wide range of experimental research.

In this thesis, a refinement process is simulated using an adaptive catalyst design cycle that combines AFE, farthest point sampling (FPS), and HTE, targeting OCM catalysts. The study aims to develop and validate a robust machine learning model, demonstrating its utility in understanding catalyst design principles and supporting efficient catalyst development.

In **Chapter 2**, an efficient catalyst discovery process combining data-driven AFE, active learning, and HTE was demonstrated targeting BaO-supported catalysts for OCM. Through the refinement of feature space across diverse catalysts via AFE and HTE, predictive accuracy was enhanced, leading to the discovery of high-performance catalysts with C_2 yields exceeding 15%.

Building on this methodology, **Chapter 3** expands the scope to multiple catalyst supports, including BaO, CaO, La_2O_3, TiO_2, and ZrO_2. The analysis revealed two main patterns: one where a single element, such as La on BaO, dominated performance, achieving high yields without additional elements; and another, exemplified by CaO, where a combination of multiple elements, particularly alkaline earth metals with Cs, was essential. These findings highlighted distinctive design rules for each support.

In **Chapter 4**, active learning efficiency was explored to achieve high-accuracy learning with smaller datasets. Additionally, I explored the possibility of applying design hypotheses derived from known supports to new supports. We tested whether design hypotheses from the five supports in Chapter 3 could guide predictions for Y_2O_3 supports. This approach confirmed the transferability of design insights across supports.

The catalysts newly developed in Chapters 2, 3, and 4 are synthesized in **Chapter 5**, which highlights the high-performance catalysts created throughout the thesis.

In summary, this thesis presents an original adaptive catalyst design cycle with recursive feature engineering, proving the utility of the resulting machine learning models in understanding catalyst design heuristics and enhancing catalyst development. By exploring a broad range of compositions for OCM catalysts, this approach successfully identified high-performance catalysts and valuable design heuristics.

Keywords: Catalyst informatics, Machine learning, High-throughput experimentation, Oxidative coupling of methane, Descriptor design technology

Preface

The present thesis is submitted for the Degree of Doctor of Philosophy at Japan Advanced Institute of Science and Technology, Japan. The thesis is consolidation of results of the research work on the topic “Development of methane oxidative coupling catalysts with feature engineering-based machine learning” under the supervision of Prof. Toshiaki Taniike during April 2022– March 2025 at Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology.

Chapter 1 provides a general introduction and outlines the purpose of this thesis, focusing on the challenges in Materials Informatics (MI) for catalyst discovery. **Chapter 2** employs an adaptive design cycle incorporating AFE, HTE, and FPS for BaO-supported catalysts, successfully deriving a validated design hypothesis through four iterative cycles. The results also demonstrate the system's capability to accurately predict highly efficient catalysts. **Chapter 3** expands this approach to multiple supports (BaO, CaO, La_2O_3, TiO_2, ZrO_2), revealing characteristics of each supports. **Chapter 4** then active learning efficiency was explored to enable high-accuracy learning with smaller datasets. Design hypotheses from the five supports in Chapter 3 were tested on Y_2O_3 supports, confirming the transferability of design insights across supports. **Chapter 5** summarizes the performance of the catalysts measured in this thesis. **Chapter 6** describes the summary and general conclusion of this thesis. This work is entirely original, and no portion of this thesis has been plagiarized.

Aya Fujiwara

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

November 2023

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Professor Toshiaki Taniike, of the Graduate School of Advanced Science and Technology at the Japan Advanced Institute of Science and Technology. His support, insightful guidance, and invaluable expertise have been crucial throughout the course of my doctoral studies.

I would also like to take this opportunity to thank Assistant Professor Toru Wada, Doctor Patchanee Chammingkwan, and Doctor Sunao Nakanowatari. Their expertise and thoughtful input have greatly enriched my research and learning experience.

I am also heartily grateful to all other members of Taniike laboratory for their valuable suggestions, cooperation and support.

I would like to extend my gratitude to my second supervisor, Professor Masayuki Yamaguchi, for his invaluable instruction and guidance. I am also deeply thankful to Associate Professor Supareak and Lecturer Meena Rittiruam for their support and guidance in my minor research, as well as the other members of their lab, for the time and effort they generously dedicated to assisting me.

I would like to thank my parents and grandmother for always encouraging me and supporting my ambitions.

Table of contents

Chapter 1

General introduction

1.1. HETEROGENEOUS CATALYSTS

Heterogeneous catalysts, exist in a different phase from the reaction products, enabling easy separation and recycling after the reaction.[1] This characteristic makes heterogeneous catalysts indispensable, playing a role in over 70% of industrial chemical processes.[2] For example, heterogeneous catalysts are used in biomass conversion,[3] ammonia synthesis,[4] selective oxidation, fuel cells, renewable fuels, and renewable chemicals. As addressing energy issues has become an urgent task, catalysts with the ability to improve the efficiency of fossil fuel utilization are being actively researched.[5,6]

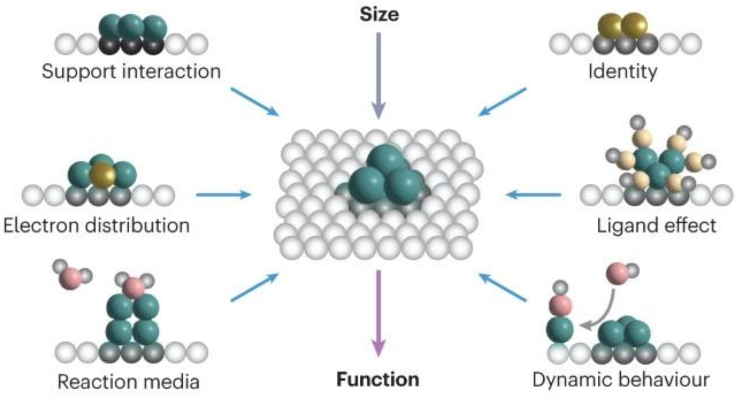


Figure 1.1. With solid catalysts, it is possible to achieve a hierarchical design that combines different properties by accumulating components with different roles on the solid surface.[7]

Despite the critical role of heterogeneous catalysts and ongoing research challenges, their development has largely relied on human ingenuity and serendipitous discoveries. For example, the Ziegler-Natta catalyst was discovered during an ethylene polymerization reaction, where the use of an apparatus containing metallic impurities unexpectedly accelerated the reaction. Likewise, in ammonia synthesis, the successful combination of reagents was found through a systematic exploration of various chemicals in the laboratory. The reason that research has been carried out in an ad hoc manner until now is the difficulty of predicting catalytic reactions. Catalytic reactions involve numerous potential reactions between the catalyst surface and reactants, each intricately interconnected, making it difficult to fully understand or predict their behavior.[8–11] In addition, catalytic reactions proceed near the surface, but the surface structure can easily change depending on the composition and synthesis process. In order to obtain the desired performance, it is necessary to control these elements that change so delicately.[12–14]

One of the primary functions of catalysts is to accelerate the rate of chemical reactions, thereby enhancing their "activity." In intricate systems involving multiple possible reactions, the "selectivity" of the process can be adjusted by preferentially accelerating specific reactions. As the reaction proceeds, however, by-products may accumulate on the catalyst's surface, potentially compromising its "durability." Additionally, certain catalytic reactions impose a substantial "environmental burden," resulting in adverse environmental impacts. Therefore, when evaluating catalyst performance, these four parameters—activity, selectivity, durability, and environmental burden—are especially critical.

1.2. OXIDATIVE COUPLING OF METHANE

1.2.1. Fundamentals

Oxidative coupling of methane (OCM) is one of the most challenging catalytic reactions to develop. This process converts methane (CH_4) directly into ethylene (C_2H_4) in a single step. As the primary component of natural gas and biogas, CH_4 has attracted significant attention amid the global expansion of natural gas exploration and extraction efforts.[15,16] Consequently, interest in utilizing CH_4 via OCM has grown. C_2H_4, the product of this reaction, is an essential feedstock for a wide range of petrochemical products.

Currently, C_2H_4 is mainly produced from naphtha, a petroleum derivative.[17,18] When using natural gas, production involves two steps: first, converting CH_4 into synthesis gas through reforming reactions, and second, converting the synthesis gas into synthetic fuel via Fischer-Tropsch synthesis (FTS).[19] However, the high cost and complexity of FTS have driven interest in the direct conversion of CH_4 through OCM.[20] One of the key challenges in the OCM reaction is the chemical inertness of CH_4 compared to its products, making high yields difficult to achieve. As a result, the maximum yield for OCM is generally considered to be around 35%.[9]

OCM involves many potential reactions, with the seven reactions listed in Table 1.1 representing the main pathways.

Table 1.1. Possible potential reactions that can occur during the oxidative coupling of methane process.

No.	Reaction
1	$2CH_4 + 1/2O_2 \rightarrow C_2H_6 + H_2O$
2	$CH_4 + 2/3O_2 \rightarrow CO + 2H_2O$
3	$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$
4	$C_2H_6 + 1/2O_2 \rightarrow C_2H_4 + H_2O$
5	$C_2H_4 + 2O_2 \rightarrow 2CO + 2H_2O$
6	$CO + H_2O \rightarrow CO_2 + H_2$
7	$CO_2 + H_2 \rightarrow CO + H_2O$

In all of the reaction stages shown in Figure 1.2, the OCM reaction forms products through catalytic reactions. In the process, CH_4 is converted through reactions 1−3. In the selective primary reaction, C_2H_4 is produced by OCM reaction 1. When OCM was first discovered, it was thought that the dimerization reaction of methyl radicals occurred on the catalyst surface, but subsequent research revealed that the reaction field was the gas phase near the catalyst.[21,22] Within 2, 3 steps of the reaction, gas-phase reactions occur, leading to the production of carbon oxides through the non-selective oxidation of methane. What this reaction means is that if the surface area of the catalyst is large, methyl radicals are likely to be generated, but at the same time, reactions that lower the yield, such reaction as 2 and 3, are also likely to progress due to the increase in the volume of the gas phase near the catalyst.[23] The reaction in which C_2 products are further oxidized to become CO_x becomes more pronounced as the conversion rate of CH_4 increases. Although the generated ethylene has the potential to react with oxygen to form C_2H_4 via reaction 4, the majority instead converts to carbon monoxide through reaction 5, which then acts as an intermediate, ultimately leading to carbon dioxide production in reaction 6. The amount of conversion from carbon monoxide to carbon dioxide is affected by the water-gas shift reaction, which is a reversible reaction 6 and 7.[24,25]

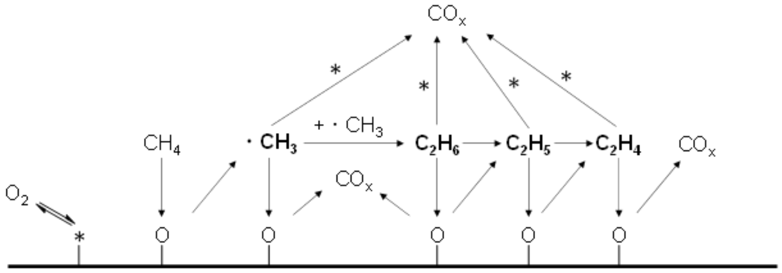


Figure 1.2 General reaction scheme for oxidative coupling of methane.

1.2.2. Catalysts for OCM

Mavlyankariev et al. have demonstrated that, in the absence of a catalyst, even with improvements to the reactor, the achievable selectivity and yield in the OCM reaction are limited to 25% and 7.3%, respectively.[26] Therefore, catalyst development is essential for improving the performance of the OCM reaction. Moreover, very few catalysts are known to achieve stable technological and economic performance in conventional fixed-bed flow reactors, such as C_2 yields above 30%, 80% selectivity, or C_2 yields of 28%. Catalyst development has been ongoing for over 40 years, starting in 1982, leading to the discovery of several high-performance catalyst families, including Li/MgO, Sr/La_2O_3, and Mn–Na_2WO_4/SiO_2.[27–29] These catalyst families have been developed independently, each following distinct design principles. Identifying and leveraging their commonalities and differences could lead to a deeper understanding and more effective catalyst design.

1.2.2.1. Li/MgO

The surface of pure MgO, acting as the support, predominantly consists of {100} planes, which are known to be the most stable and contain relatively few active oxygen species. Typically, the addition of a supported element is expected to enhance specific properties, such as increasing and tuning the number of basic sites, as well as affecting lattice oxygen. When Li is doped into MgO, {110} and {111} planes increase due to Li migration from the bulk phase. These planes are energetically less stable, leading to the formation of $Mg^{2+}O^{2-}$ sites or defective MgO surfaces, both of which contribute to CH_4 activation.[30–34] With sufficient Li doping, $Mg^{2+}O^{2-}$ sites on the MgO{110} planes become the dominant surface. These sites have been reported to efficiently and selectively activate CH_4, forming a methyl radical intermediate while minimizing further dissociation and combustion reactions.[30] In this way, Li acts as a structural modifier rather than an active center. Li/MgO catalysts provide high performance instantaneously, but because Li is easily volatilized at high temperatures and has poor durability, there is a need for improvements that will increase the usable time of this catalyst.

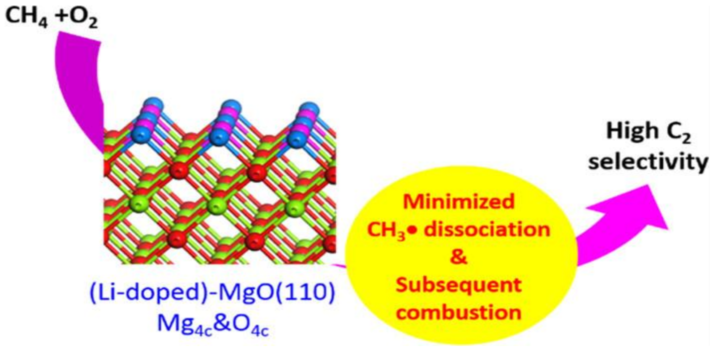


Figure 1.3. The role of (Li-doped) -MgO.[30]

1.2.2.2. Sr/La2O3

Regarding Sr/La_2O_3, La_2O_3 is a support with a highly basic and stable structure, but it has poor C_2 selectivity. Among the alkaline earth metal oxides, Sr^{2+} ions have been experimentally shown to be the most effective in improving the selectivity of La_2O_3catalysts.[35] The active species of Sr/La_2O_3 is lattice oxygen. When Sr is incorporated into the crystal lattice, distortions and changes in electron density occur, resulting in the formation of several types of active oxygen species, such as superoxides (O_2^-), peroxide ions (O_2^{2-}), hydroxide ions (OH^-), carbonates (CO_3^{2-}), and lattice oxygens (O^{2-}).[35,36] It is suggested that this active oxygen, particularly O_2^-, activates CH_4. The advantages of this catalyst are its long lifespan and activity at low temperatures. From an industrial perspective, achieving activity at even lower temperatures is a key goal.

1.2.2.3. Mn-Na_2WO_4/SiO_2

Regarding the support, while SiO_2 itself does not possess catalytic activity, it can retain molten active species by forming a cristobalite phase. The active species in Mn–Na_2WO_4/SiO_2 is WO_4^{2-} with a tetrahedral structure, which is generated when Na transforms W into an oxometalate anion. W can exist in multiple oxidation states, but it is rapidly regenerated to W^{6+} by Mn. The interaction of these elements results in a high C_2 yield. While Mn–Na_2WO_4/SiO_2 has excellent selectivity at high temperatures, the CH_4 conversion rate remains a challenge.

1.3. CATALYSTS INFORMATICS

Conventional catalyst research has relied on the researcher's intuition, but this method is time-consuming and expensive.[37–39] therefore, data-driven catalyst research (catalyst informatics) using machine learning and high-throughput experiments has been attracting attention due to advances in computational technology.[40] This is especially effective when used for calculations on a scale that exceeds human perception. The clarification of catalytic reactions is complex because reactions are involved. Furthermore, the reaction mechanism is affected by changes in reaction conditions and catalysts, so it is difficult to track the complete reaction mechanism experimentally.

The bottlenecks in catalyst informatics are the scarce availability of catalyst data suitable for data science and the difficulty of hand-crafting descriptors that capture the essence of intricate structure-function relationships. This thesis utilizes high-throughput experimentation (HTE) to prepare and evaluate a large number of solid catalysts, generating sizable, high-quality, and consistent datasets for various heterogeneous catalytic reactions, including OCM. Moreover, we recently introduced an automatic feature engineering (AFE) technique, which programmatically designs descriptors that can capture the essence of target catalysis, starting from general physical properties of elements such as atomic radii and electronegativity.

1.4. HIGH-THROUGHPUT EXPERIMENTATION

Numerous studies have reported the development and design of new materials using high-throughput experiments.[41–43] Furthermore, integrating this approach into catalyst informatics is expected to enable unprecedented unified data analysis, facilitating a deeper understanding of catalytic reactions. However, in order to process large datasets without losing critical information, optimal design and configuration of reactors is essential. The high-throughput equipment used in this study was custom-designed, and through repeated measurements and iterative optimization, a mechanism specifically tailored for OCM measurements was successfully established.[42,44,45]

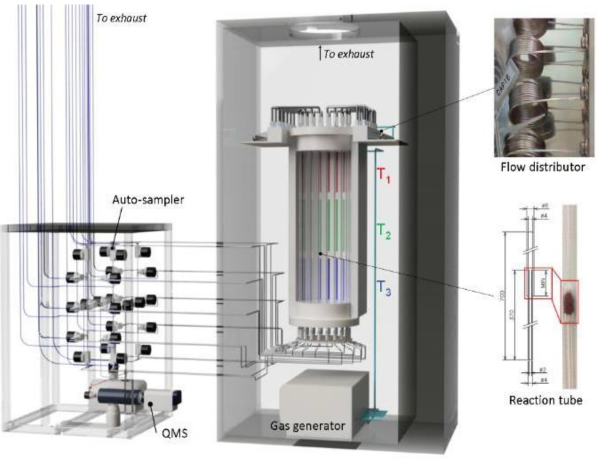


Figure 1.3. Illustration of the developed high-throughput screening (HTS) system. [42]

Figure 1.3 shows the high-throughput screening system (HTS device). This device is a combination of a gas mixer, flow rate distributor, quartz reaction tube, electric furnace, autosampler, and quadrupole mass spectrometer (QMS). After each gas leaves the gas cylinder, it is first sent to the gas mixer. The flow rate is controlled to create a mixture of $CH_4/O_2/Ar$, which is then evenly distributed to the 20 reaction tubes using a flow rate distributor. These reaction tubes are installed inside a electric furnace and 20 of them are heated simultaneously. The electric furnace is equipped with three temperature control mechanisms, with T_1 being the gas inflow side, T_2 being the catalyst bed section, and T_3 being the gas outflow side. The inner diameter of the reaction tube is 4 mm on the gas supply side and 2 mm on the gas outflow side. The aim of this is to increase the gas flow rate after the reaction to suppress secondary oxidation of the product. The catalyst is loaded to a height of 10 mm in the section where the inner diameter of the reaction tube transitions at the center. The gas after the reaction is sequentially sent to the autosampler and analyzed by the QMS.

1.5. AUTOMATIC FEATURE ENGINEERING

AFE automates the design of physically meaningful features for a given catalyst dataset within the framework of supervised ML.[46] This process involves a structured pipeline of feature assignment, synthesis, and selection. First, physical quantities of elements are assigned to the catalysts, with their elemental compositions represented through commutative operations. Higher-order features involving nonlinear and combinatorial effects are then synthesized from these assigned primitive features using mathematical operations. Finally, a specified number of features that optimize the score of the supervised ML are selected from the large synthesized array of features.

In this study, we utilized 58 parameters of elements from XenonPy, normalized according to the literature. These parameters were assigned to each catalyst using five commutative operations (maximum, minimum, average, product, and standard deviation), yielding 290 primary features. These primary features were further synthesized into 3,480 features using 12 functional forms (x, sqrt(x), x^2, x^3, ln(x), exp(x), and their reciprocals, where x represents each primary feature). A genetic algorithm-based approach was employed to select eight features that minimized the mean absolute error (MAE) in leave-one-out cross-validation (LOOCV) using Huber regression. This process involved assessing approximately 4,000,000 models per dataset with various feature combinations and selecting the combination of features (X) and model (f(X)) that yielded the lowest cross-validation (CV) score as the most plausible design hypothesis. Huber regression, which is a form of multiple linear regression, was employed to prevent overfitting owing to its reduced number of parameters and ensure robustness against outliers, such as experimental failures. The number of selected features was empirically determined to balance the CV score and the cost of feature selection.

Similar to how researchers cannot dismiss alternative hypotheses when evidence is limited, AFE cannot overlook alternative design hypotheses (ML models with differently tailored descriptors that fit the training data) when the diversity of catalysts in the training data is restricted. Consequently, we implemented an active learning strategy (Figure 1.4). This strategy employs farthest point sampling (FPS) within the descriptor space established by AFE to propose catalysts that are maximally dissimilar to those included in the training data.[47–49] These catalysts served as rigorous control experiments to validate the proposed design hypothesis. The performance of the proposed catalysts were assessed using HTE to reinforce the training data and update the design hypothesis via AFE.

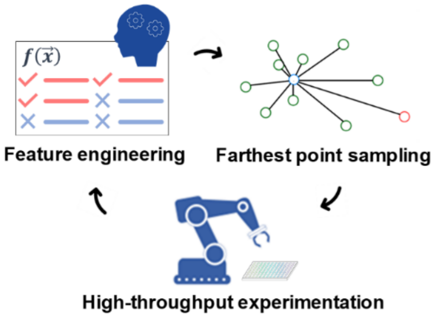


Figure 1.4. Schematic of the active learning cycle. Feature engineering was performed in several iterations, with data from 20 additional catalysts incorporated in each update.[46]

1.6. OBJECTIVE

When we take the perspective that "researchers gain some insight from multiple observations", if the number of observational data used for reference is small, it is not possible to reject the opposing hypothesis. For this reason, experiments are repeated to narrow down the truly correct hypothesis. This process is the part of the experimental process that relies most on the researcher's personal sense. The way hypotheses are created and the way experiments are set up will affect how quickly you reach a conclusion and the quality of the research. If we can incorporate this process into a machine learning framework and make it more efficient using high-throughput experiments, we can build a general-purpose research framework that can be applied to all kinds of experimental activities. In this paper, we aim to obtain a well-validated machine learning model by implementing an adaptive catalyst design loop that combines AFE, as shown in Figure 1.4, with farthest point sampling (FPS) and HTE, and to demonstrate the usefulness of the obtained model in understanding catalyst design rules and efficient catalyst development.

In Chapter 2, an adaptive catalyst design cycle that combines machine learning and high-throughput experiments was implemented. I demonstrated that a well-validated machine learning model (design hypothesis) is useful for understanding catalyst design rules and for efficient catalyst development.

In Chapter 3, we derived design hypotheses for 5 different supports (BaO, CaO, La_2O_3, TiO_2, ZrO_2) by executing an adaptive catalyst design cycle. By acquiring and comparing design hypotheses for these multiple supports, we clarified the characteristics of each support.

The research in Chapter 4 is based on the idea that if there is a similarity in the design rules between supports, then the design hypothesis acquired for another support may be useful for deriving the design hypothesis for an unknown support. We demonstrated that the design hypothesis for the 5 supports derived in Chapter 3 can be applied to Y_2O_3 supports that are not included in the training data at all.

The catalysts newly developed in Chapters 2, 3, and 4 are synthesized in Chapter 5, which highlights the high-performance catalysts created throughout the thesis.

Based on the above research and results, this paper presents a catalyst design technique that can be applied to all reactions. We believe that this research will contribute to the development of the catalyst.

REFERENCE

(1) Kibar, M. E.; Hilal, L.; Çapa, B. T.; Bahçıvanlar, B.; Abdeljelil, B. B. Assessment of Homogeneous and Heterogeneous Catalysts in Transesterification Reaction: A Mini Review. *ChemBioEng Reviews* **2023**, *10* (4), 412–422. https://doi.org/10.1002/cben.202200021.

(2) Boudart, M. Heterogeneous Catalysis by Metals. *Journal of Molecular Catalysis* **1985**, *30* (1–2), 27–38. https://doi.org/10.1016/0304-5102(85)80014-6.

(3) Lin, L.; Han, X.; Han, B.; Yang, S. Emerging Heterogeneous Catalysts for Biomass Conversion: Studies of the Reaction Mechanism. *Chem. Soc. Rev.* **2021**, *50* (20), 11270–11292. https://doi.org/10.1039/D1CS00039J.

(4) Marakatti, V. S.; Gaigneaux, E. M. Recent Advances in Heterogeneous Catalysis for Ammonia Synthesis. *ChemCatChem* **2020**, *12* (23), 5838–5857. https://doi.org/10.1002/cctc.202001141.

(5) Chee, S. W.; Lunkenbein, T.; Schlögl, R.; Roldán Cuenya, B. *Operando* Electron Microscopy of Catalysts: The Missing Cornerstone in Heterogeneous Catalysis Research? *Chem. Rev.* **2023**, *123* (23), 13374–13418. https://doi.org/10.1021/acs.chemrev.3c00352.

(6) Chatterjee, S.; Parsapur, R. K.; Huang, K.-W. Limitations of Ammonia as a Hydrogen Energy Carrier for the Transportation Sector. *ACS Energy Lett.* **2021**, *6* (12), 4390–4394. https://doi.org/10.1021/acsenergylett.1c02189.

(7) Xinzhe, L.; Sharon M.; Yiyun F.; Jun L.; Javier P. R.; Jiong L. L.; X. Advances in heterogeneous single-cluster catalysis. *Nat. Rev. Chem.* **2023**, 7, 754-767. 10.1038/s41570-023-00540-8

(8) Sun, J.; Thybaut, J.; Marin, G. Microkinetics of Methane Oxidative Coupling. *Catalysis Today* **2008**, *137* (1), 90–102. https://doi.org/10.1016/j.cattod.2008.02.026.

(9) Thybaut, J. W.; Sun, J.; Olivier, L.; Van Veen, A. C.; Mirodatos, C.; Marin, G. B. Catalyst Design Based on Microkinetic Models: Oxidative Coupling of Methane. *Catalysis Today* **2011**, *159* (1), 29–36. https://doi.org/10.1016/j.cattod.2010.09.002.

(10) Alexiadis, V. I.; Thybaut, J. W.; Kechagiopoulos, P. N.; Chaar, M.; Van Veen, A. C.; Muhler, M.; Marin, G. B. Oxidative Coupling of Methane: Catalytic Behaviour Assessment via Comprehensive Microkinetic Modelling. *Applied Catalysis B: Environmental* **2014**, *150–151*, 496–505. https://doi.org/10.1016/j.apcatb.2013.12.043.

(11) Ishikawa, A.; Tateyama, Y. A First-Principles Microkinetics for Homogeneous–Heterogeneous Reactions: Application to Oxidative Coupling of Methane Catalyzed by Magnesium Oxide. *ACS Catal.* **2021**, *11* (5), 2691–2700. https://doi.org/10.1021/acscatal.0c04104.

(12) Takasao, G.; Wada, T.; Thakur, A.; Chammingkwan, P.; Terano, M.; Taniike, T. Machine Learning-Aided Structure Determination for $TiCl_4$–Capped $MgCl_2$ Nanoplate of Heterogeneous Ziegler–Natta Catalyst. *ACS Catal.* **2019**, *9* (3), 2599–2609. https://doi.org/10.1021/acscatal.8b05080.

(13) Chikuma, H.; Takasao, G.; Wada, T.; Chammingkwan, P.; Behler, J.; Taniike, T. Accelerating Non-Empirical Structure Determination of Ziegler–Natta Catalysts with a High-Dimensional Neural Network Potential. *J. Phys. Chem. C* **2023**, *127* (24), 11683–11691. https://doi.org/10.1021/acs.jpcc.3c01511.

(14) Da Silveira, J. M.; Chikuma, H.; Takasao, G.; Wada, T.; Chammingkwan, P.; Taniike, T. Deciphering the Role of Internal Donors in Shaping Heterogeneous Ziegler–Natta Catalysts Based on Nonempirical Structural Determination. *ACS Catal.* **2024**, *14* (4), 2300–2312. https://doi.org/10.1021/acscatal.3c05480.

(15) Balat, M. Global Trends on Production and Utilization of Natural Gas. *Energy Sources, Part B: Economics, Planning, and Policy* **2009**, *4* (4), 333–346.

https://doi.org/10.1080/15567240701621125.

(16) Shaibu, R.; Sambo, C.; Guo, B.; Dudun, A. An Assessment of Methane Gas Production from Natural Gas Hydrates: Challenges, Technology and Market Outlook. *Adv. Geo-Energy Res.* **2021**, *5* (3), 318–332. https://doi.org/10.46690/ager.2021.03.07.

(17) Zaffran, J.; Yang, B. Theoretical Insights into the Formation Mechanism of Methane, Ethylene and Methanol in Fischer-Tropsch Synthesis at Co $_2$ C Surfaces. *ChemCatChem* **2021**, *13* (11), 2674–2682. https://doi.org/10.1002/cctc.202100216.

(18) Thyssen, V. V.; Vilela, V. B.; De Florio, D. Z.; Ferlauto, A. S.; Fonseca, F. C. Direct Conversion of Methane to C $_2$ Hydrocarbons in Solid-State Membrane Reactors at High Temperatures. *Chem. Rev.* **2022**, *122* (3), 3966–3995. https://doi.org/10.1021/acs.chemrev.1c00447.

(19) Kim, Y. H.; Jun, K.-W.; Joo, H.; Han, C.; Song, I. K. A Simulation Study on Gas-to-Liquid (Natural Gas to Fischer–Tropsch Synthetic Fuel) Process Optimization. *Chemical Engineering Journal* **2009**, *155* (1–2), 427–432. https://doi.org/10.1016/j.cej.2009.08.018.

(20) Damasceno, S.; Trindade, F. J.; Fonseca, F. C.; Florio, D. Z. D.; Ferlauto, A. S. Oxidative Coupling of Methane in Chemical Looping Design. *Fuel Processing Technology* **2022**, *231*, 107255. https://doi.org/10.1016/j.fuproc.2022.107255.

(21) Ito, T.; Wang, J.; Lin, C. H.; Lunsford, J. H. Oxidative Dimerization of Methane over a Lithium-Promoted Magnesium Oxide Catalyst. *J. Am. Chem. Soc.* **1985**, *107* (18), 5062–5068. https://doi.org/10.1021/ja00304a008.

(22) Campbell, K. D.; Morales, E.; Lunsford, J. H. Gas-Phase Coupling of Methyl Radicals during the Catalytic Partial Oxidation of Methane. *J. Am. Chem. Soc.* **1987**, *109* (25), 7900–7901. https://doi.org/10.1021/ja00259a059.

(23) Iwamatsu, E.; Moriyama, T.; Takasaki, N.; Aika, K. Importance of the Specific Surface Area of the Catalyst in Oxidative Dimerization of Methane over Promoted Magnesium Oxide. *J. Chem. Soc., Chem. Commun.* **1987**, No. 1, 19. https://doi.org/10.1039/c39870000019.

(24) Xu, J.; Liu, Y.; Xu, X.; Zhang, Y.; Xi, R.; Zhang, Z.; Fang, X.; Wang, X. Influence of Cesium Loading on Oxidative Coupling of Methane (OCM) over Cs/SnO2 Catalysts. *Chemistry Africa* **2020**, *3* (3), 687–694. https://doi.org/10.1007/s42250-020-00143-6.

(25) Stansch, Z.; Mleczko, L.; Baerns, M. Comprehensive Kinetics of Oxidative Coupling of Methane over the La $_2$ O $_3$ /CaO Catalyst. *Ind. Eng. Chem. Res.* **1997**, *36* (7), 2568–2579. https://doi.org/10.1021/ie960562k.

(26) Mavlyankariev, S. Investigation of Gas-Phase Methane Oxidation by Reactor Profile Measurements and Microkinetic Modeling. Technische Universität Berlin July 14, 2013. https://doi.org/10.14279/DEPOSITONCE-3731.

(27) Lin, C. H.; Ito, T.; Wang, J.; Lunsford, J. H. Oxidative Dimerization of Methane over Magnesium and Calcium Oxide Catalysts Promoted with Group IA Ions: The Role of [M^+O^-] Centers. *Journal of the American Chemical Society* **1987**, *109*, 4808–4810.

(28) DeBoy, J. M.; Hicks, R. F. The Oxidative Coupling of Methane over Alkali, Alkaline Earth, and Rare Earth Oxides. *Ind. Eng. Chem. Res.* **1988**, *27* (9), 1577–1582. https://doi.org/10.1021/ie00081a004.

(29) Shahri, S. M. K.; Alavi, S. M. Kinetic Studies of the Oxidative Coupling of Methane over the $Mn/Na_2WO_4/SiO_2$ Catalyst. *Journal of Natural Gas Chemistry* **2009**, *18* (1), 25–34. https://doi.org/10.1016/S1003-9953(08)60079-1.

(30) Qian, K.; You, R.; Guan, Y.; Wen, W.; Tian, Y.; Pan, Y.; Huang, W. Single-Site Catalysis of Li-MgO Catalysts for Oxidative Coupling of Methane Reaction. *ACS Catal.* **2020**, *10* (24), 15142–15148. https://doi.org/10.1021/acscatal.0c03896.

(31) Luo, L.; Jin, Y.; Pan, H.; Zheng, X.; Wu, L.; You, R.; Huang, W. Distribution and Role of Li in Li-Doped MgO Catalysts for Oxidative Coupling of Methane. *Journal of Catalysis* **2017**, *346*, 57–61. https://doi.org/10.1016/j.jcat.2016.11.034.

(32) Myrach, P.; Nilius, N.; Levchenko, S. V.; Gonchar, A.; Risse, T.; Dinse, K.; Boatner, L. A.; Frandsen, W.; Horn, R.; Freund, H.; Schlögl, R.; Scheffler, M. Temperature-Dependent Morphology, Magnetic and Optical Properties of Li-Doped MgO. *ChemCatChem* **2010**, *2* (7), 854–862. https://doi.org/10.1002/cctc.201000083.

(33) Zavyalova, U.; Geske, M.; Horn, R.; Weinberg, G.; Frandsen, W.; Schuster, M.; Schlögl, R. Morphology and Microstructure of Li/MgO Catalysts for the Oxidative Coupling of Methane. *ChemCatChem* **2011**, *3* (6), 949–959. https://doi.org/10.1002/cctc.201000098.

(34) Richter, N. A.; Stavale, F.; Levchenko, S. V.; Nilius, N.; Freund, H.-J.; Scheffler, M. Defect Complexes in Li-Doped MgO. *Phys. Rev. B* **2015**, *91* (19), 195305. https://doi.org/10.1103/PhysRevB.91.195305.

(35) Kiatsaengthong, D.; Jaroenpanon, K.; Somchuea, P.; Chukeaw, T.; Chareonpanich, M.; Faungnawakij, K.; Sohn, H.; Rupprechter, G.; Seubsai, A. Effects of Mg, Ca, Sr, and Ba Dopants on the Performance of La_2O_3 Catalysts for the Oxidative Coupling of Methane. *ACS Omega* **2022**, *7* (2), 1785–1793. https://doi.org/10.1021/acsomega.1c04738.

(36) Schucker, R. C.; J. Derrickson, K.; K. Ali, A.; J. Caton, N. The Effect of Strontium Content on the Activity and Selectivity of Sr-Doped La_2O_3 Catalysts in Oxidative Coupling of Methane. *Applied Catalysis A: General* **2020**, *607*, 117827. https://doi.org/10.1016/j.apcata.2020.117827.

(37) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8* (8), 7403–7429. https://doi.org/10.1021/acscatal.8b01708.

(38) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nat Commun* **2016**, *7* (1), 11241. https://doi.org/10.1038/ncomms11241.

(39) Cao, B.; Adutwum, L. A.; Oliynyk, A. O.; Luber, E. J.; Olsen, B. C.; Mar, A.; Buriak, J. M. How To Optimize Materials and Devices *via* Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics. *ACS Nano* **2018**, *12* (8), 7434–7444. https://doi.org/10.1021/acsnano.8b04726.

(40) Hendershot, R. J.; Snively, C. M.; Lauterbach, J. High-Throughput Heterogeneous Catalytic Science. *Chemistry A European J* **2005**, *11* (3), 806–814. https://doi.org/10.1002/chem.200400613.

(41) Farrusseng, D. High-Throughput Heterogeneous Catalysis. *Surface Science Reports* **2008**, *63* (11), 487–513. https://doi.org/10.1016/j.surfrep.2008.09.001.

(42) Nguyen, T. N.; Nhat, T. T. P.; Takimoto, K.; Thakur, A.; Nishimura, S.; Ohyama, J.; Miyazato, I.; Takahashi, L.; Fujima, J.; Takahashi, K.; Taniike, T. High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. *ACS Catal.* **2020**, *10* (2), 921–932. https://doi.org/10.1021/acscatal.9b04293.

(43) Yanagiyama, K.; Takimoto, K.; Dinh Le, S.; Nu Thanh Ton, N.; Taniike, T. High-Throughput Experimentation for Photocatalytic Water Purification in Practical Environments. *Environmental Pollution* **2024**, *342*, 122974. https://doi.org/10.1016/j.envpol.2023.122974.

(44) Nguyen, T. N.; Nakanowatari, S.; Nhat Tran, T. P.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane. *ACS Catal.* **2021**, *11* (3), 1797–1809. https://doi.org/10.1021/acscatal.0c04629.

(45) Nakanowatari, S.; Nguyen, T. N.; Chikuma, H.; Fujiwara, A.; Seenivasan, K.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Extraction of Catalyst Design Heuristics from Random Catalyst Dataset and Their Utilization in Catalyst Development for Oxidative Coupling of Methane. *ChemCatChem* **2021**, *13* (14), 3262–3269. https://doi.org/10.1002/cctc.202100460.

(46) Taniike, T.; Fujiwara, A.; Nakanowatari, S.; García-Escobar, F.; Takahashi, K. Automatic

Feature Engineering for Catalyst Design Using Small Data without Prior Knowledge of Target Catalysis. *Commun Chem* **2024**, *7* (1), 11. https://doi.org/10.1038/s42004-023-01086-y.
(47) Qi, C. R.; Yi, Li; Su, Hao; Guibas, J. Leonidas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.
(48) Shi, S.; Wang, X.; Li, H. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. arXiv 2018. https://doi.org/10.48550/ARXIV.1812.04244.
(49) Guo, M.-H.; Cai, J.-X.; Liu, Z.-N.; Mu, T.-J.; Martin, R. R.; Hu, S.-M. PCT: Point Cloud Transformer. **2020**. https://doi.org/10.48550/ARXIV.2012.09688.

Chapter 2

Application of automatic feature engineering to development of catalysts for oxidative methane coupling

KEYWORDS: Oxidative coupling of methane, computational methods, high-throughput experimentation, Structure prediction

ABSTRACT

Traditional catalyst discovery has been driven largely by researchers intuition, with hypotheses constructed and validated through experimental testing. The rise of data-driven approaches, however, has enabled a transformation in catalyst research, offering significant advancements. Nonetheless, data scarcity and challenges in designing suitable descriptors for machine learning (ML) remain major obstacles, as catalyst datasets are typically limited in quantity and complexity. To address these limitations, this study leverages automatic feature engineering (AFE) integrated with active learning and high-throughput experimentation (HTE) to systematically refine design hypotheses for BaO-supported catalysts in oxidative coupling of methane (OCM).

AFE automates the selection of meaningful features, enhancing model interpretability and generalizability across different catalyst compositions. Through iterative active learning cycles, catalysts were selected using farthest point sampling (FPS) within the feature space, maximizing diversity to filter out locally fit models and reinforce robust, globally applicable hypotheses. This approach enabled the identification of a globally accurate design hypothesis, significantly reducing discrepancies between predicted and observed catalyst performance. Performance evolution was visualized using t-SNE, showing a gradual refinement of the feature space as diverse catalysts were added. The study demonstrated that AFE combined with HTE and active learning not only improved predictive accuracy but also accelerated high-yield catalyst discovery, revealing 30 catalysts with C_2 yields ≥15%. This approach offers a scalable framework for advancing data-driven catalyst design in complex catalytic reactions like OCM.

2.1. INTRODUCTION

The intricacy of structure-function relationships has made trial and error a major driver in the development of solid catalysts.[1,2] Individual researchers propose their own hypotheses and test various materials as potential catalysts. The research focus then evolves along with improved hypotheses and catalysts found in prior research for further refinement. This situation is exemplified by oxidative coupling of methane (OCM)—a reaction that converts methane (CH_4), the main component of natural gas and biogas, into C_2 compounds in a single step.[3] OCM is considered potentially more efficient than the current two-step route involving steam reforming and the Fischer–Tropsch process.[4] However, due to the chemical inertness of CH_4 when compared to the C_2 products, achieving high yields is challenging. Currently, few catalysts are known to stably achieve techno-economical performance (e.g., C_2 yields over 30% otherwise 28% with 80% selectivity) in a conventional fixed-bed flow reactor.[5–7] Efforts have been directed at optimizing the conversion-selectivity tradeoff through various methods, including catalyst preparation techniques, the addition of extra elements, and so on.

Traditional catalyst research has relied on the intuition of researchers, with hypotheses constructed and validated through experiments. However, the rise of data-driven approaches has transformed this process, yielding significant advancements across many scientific fields, particularly in catalysis.[8–11] Data-driven catalysis research, also known as catalyst informatics, aims to accelerate the development and understanding of catalysts by discovering applicable trends and patterns hidden in catalyst data with the aid of data science techniques such as machine learning (ML) and visualization. The bottlenecks in catalyst informatics are the scarce availability of catalyst data suitable for data science and the difficulty of hand-crafting descriptors that capture the essence of intricate structure-function relationships.[12–15] Catalyst datasets typically consist of fewer than 1000 observations, making the application of complex ML models, especially deep learning, difficult. As a result, researchers commonly select key descriptors and use simpler ML models to efficiently process the data. However, this process requires deep expertise to identify system-specific factors. Moreover, the complexity of catalytic mechanisms makes descriptor design particularly challenging compared to other fields.

To address these challenges, this study utilized automatic feature engineering (AFE) without prior system knowledge to explore design principles for BaO-supported catalysts in OCM reactions. BaO is the catalyst support with the highest average performance in the dataset of Taniike group. Additionally, I utilize a new approach by integrating AFE with active learning and high-throughput experimentation (HTE) to accelerate catalyst discovery. This is the first attempt to simulate the process of hypothesis refinement in data science using this method.

2.2. METHOD

Just as researchers cannot dismiss alternative hypotheses when evidence is limited, AFE must also consider alternative design hypotheses, such as ML models with differently tailored descriptors that fit the training data, especially when catalyst diversity within the training set is limited. To address this, we implemented an active learning strategy (Figure 2.1) that uses farthest point sampling (FPS) within the descriptor space defined by AFE to suggest catalysts that are as dissimilar as possible to those in the training set.[16–18] These catalysts serve as rigorous control experiments to test and validate the proposed design hypothesis.

The performance of each proposed catalyst was evaluated using HTE, reinforcing the training data and allowing AFE to update and refine the design hypothesis iteratively. This process aims to discard design hypotheses that lack generalizability across diverse catalysts, ultimately leading to a robust, experimentally validated design hypothesis. Further details are provided below.

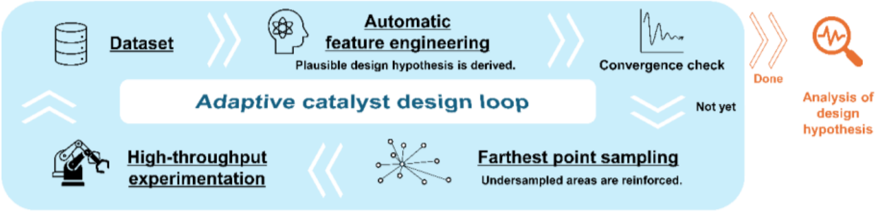


Figure 2.1. Active learning cycle employed in this study. Automatic feature engineering was applied to a given catalyst dataset to derive a design hypothesis (an ML model with tailored descriptors). Catalysts recommended by FPS were assessed using HTE. The resulting data were integrated back into the dataset, and this iterative process continued until a robust design hypothesis was established to elucidate the relationship between catalyst compositions and performances.

2.2.1. Dataset

Taniike group has accumulated OCM data for quaternary catalysts represented as M1–M2–M3/Support using a consistent experimental protocol through HTE. M1–3 represent the supported elements, which can be selected, with duplication allowed, from Li, Na, Mg, K, Ca, Ti, V, Mn, Fe, Co, Ni, Cu, Zn, Sr, Y, Zr, Mo, Pd, Cs, Ba, La, Ce, Nd, Eu, Tb, Hf, W, and none (none indicates no addition of elements). The support is selected from MgO, Al_2O_3, SiO_2, CaO, TiO_2, ZrO_2, BaO, La_2O_3, and CeO_2. The loading amount of supported elements, except for none, is 0.37 mmol per gram of support per selection. This constitutes a parameter space containing 4,060 catalysts per support and a total of 36,540 catalysts.

In this chapter, out of 636 quaternary catalysts that we have reported, we extracted 95 catalysts relating to BaO supports, for establishing design hypotheses through active learning. BaO was selected as a reference to these supports but with higher C_2 yields and selectivity. Among the 95 catalysts, 29 were obtained via random sampling from the entire space, while the remaining 66 catalysts were obtained to validate various ML techniques.

Additionally, 20 new catalysts were evaluated using HTE. Of these, 18 catalysts were selected based on FPS in the feature space, while the remaining 2 were chosen based on the largest absolute error between observed and predicted yields from the previous regression model. This process was repeated 4 cycles, updating the AFE feature space with the obtained data, leading to the evaluation of a total of 80 catalysts.

2.2.2. Automatic feature engineering

AFE automates the design of physically meaningful features for a given catalyst dataset within the framework of supervised ML.[19] This process involves a structured pipeline of feature assignment, synthesis, and selection. First, physical quantities of elements are assigned to the catalysts, with their elemental compositions represented through commutative operations.[20] Higher-order features involving nonlinear and combinatorial effects are then synthesized from these assigned primitive features using mathematical operations.[21–23] Finally, a specified number of features (descriptors) that optimize the score of the supervised ML are selected from the large synthesized array of features.

In this study, we utilized 58 parameters of elements from XenonPy, normalized according to the literature.[24] These parameters were assigned to each catalyst using five commutative operations (maximum, minimum, average, product, and standard deviation), yielding 290 primary features. These primary features were further synthesized into 3,480 features using 12 functional forms (x, sqrt(x), x^2, x^3, ln(x), exp(x), and their reciprocals, where x represents each primary feature). A genetic algorithm-based approach was employed to select eight features that minimized the mean absolute error (MAE) in leave-one-out cross-validation (LOOCV) using Huber regression.[25] This process involved assessing approximately 4,000,000 models per dataset with various feature combinations and selecting the combination of features (X) and model (f(X)) that yielded the lowest cross-validation (CV) score as the most plausible design hypothesis. Huber regression, which is a form of multiple linear regression, was employed to prevent overfitting owing to its reduced number of parameters and ensure robustness against outliers, such as experimental failures. The number of selected features was empirically determined to balance the CV score and the cost of feature selection. Further details on AFE and parameter selection are provided in previous study.[19]

2.2.3. Farthest point sampling

To validate and refine the design hypothesis presented in Section 2.2, we added either 10 or 20 catalysts to each active learning cycle for experimental testing. Among these, 90% were selected using FPS within the normalized eight-dimensional feature space established by AFE. The remaining 10% corresponded to the re-evaluation of the catalysts that exhibited the largest deviations between observed and predicted values in the last cycle.

2.2.4. High-throughput experimentation

The catalysts proposed in Section 2.3 were prepared and evaluated using the same experimental methods and conditions as those employed when acquiring the original training data, which are briefly described as follows.[26]

2.2.4.1. Materials

The metal precursors used were $LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, KNO_3, $Ca(NO_3)_2\cdot4H_2O$, $Ti(OiPr)_4$, $VOSO_4\cdot xH_2O$ (x = 3–5), $Mn(NO_3)_2\cdot6H_2O$, $Fe(NO_3)_3\cdot9H_2O$, $Co(NO_3)_2\cdot6H_2O$, $Ni(NO_3)_2\cdot6H_2O$, $Cu(NO_3)_2\cdot3H_2O$, $Zn(NO_3)_2\cdot6H_2O$, $Sr(NO_3)_2$, $Y(NO_3)_3\cdot6H_2O$, $ZrO(NO_3)_2\cdot xH_2O$ (x = 2), $(NH_4)_6Mo_7O_{24}\cdot4H_2O$, $Pd(OAc)_2$, $CsNO_3$, $Ba(NO_3)_2$, $La(NO_3)_3\cdot6H_2O$, $Ce(NO_3)_3\cdot6H_2O$, $Nd(NO_3)_3\cdot6H_2O$, $Eu(NO_3)_3\cdot5H_2O$, $Tb(NO_3)_3\cdot5H_2O$, $Hf(OEt)_4$, and $(NH_4)_{10}H_2(W_2O_7)_6$. These materials were purchased from one of the following suppliers: Sigma-Aldrich, Kanto Chemical, Wako Pure Chemical Industries, Alfa-Aesar, or Sumitomo Chemical. The oxide support and their precursor included $Ba(OH)_2\cdot8H_2O$ (1.1 m^2/g, Wako Pure Chemical Industries).

2.2.4.2. Preparation of catalysts

The catalyst preparation was conducted using a parallelized wet impregnation method, where support powder was impregnated with an aqueous solution of specified metal precursors at 50 °C for 6 hours. The loading amount of elements was fixed at 0.37 mmol/g-support per selection within M1–3. Following impregnation, the powder was vacuum-dried and then calcined in air at 1000 °C for a duration of 3 hours. When the precursors involved metal alkoxides, the impregnation was performed in two steps: First with an aqueous solution of the other precursors, followed by impregnation with an ethanol solution of metal alkoxides.[26]

2.2.4.2. Evaluation of catalysts

The catalysts were evaluated for OCM using a custom-built HTE system. The system functions through a combination of a gas mixer, flow distributor, electric furnace bearing reactors, autosampler, and quadrupole mass spectrometer (QMS), enabling the automated evaluation of the performance of 20 catalysts under a programmed sequence of reaction conditions. The catalysts were fixed into 1 cm-high beds in reaction tubes (made of fused quartz tubes with 4 mm and 2 mm inner diameters), with the aid of quartz wool. After inline activation at 1000 °C for 3 hours in an oxygen stream, the 20 catalysts were tested under 135 conditions varying in temperatures (700, 750, 800, 850, and 900 °C), total gas flow rates (10, 15, 20 mL/min/channel), CH_4/O_2 ratios (2, 4, 6 mol/mol), and Ar partial pressures as a balance gas (0.15, 0.40, 0.70 atm). Each catalyst was labeled based on its highest C_2 yield among the 135 conditions.

2.3. RESULTS AND DISCUSSION

2.2.5. Active learning for obtaining robust design hypotheses

In conventional hypothesis validation, researchers add data not only to verify the primary hypothesis but also to eliminate alternative hypotheses that could explain the original data. Similarly, when the training data are limited, the AFE may generate multiple design hypotheses with similar scores, which are considered alternative design hypotheses. These arise owing to multicollinearity among features within the given data. Eliminating incorrect or less robust hypotheses requires active learning. In this study, catalysts were selected using FPS to enhance the diversity of the data and assess the robustness of the proposed design hypotheses. Design hypotheses that were not applicable to diverse data were filtered out through repeated active learning cycles, resulting in more robust and reliable design hypotheses.

Although these models demonstrate similar performance in explaining the training data, their predictive behaviors for unknown catalysts can vary significantly. In other words, many of these models are only locally fit, lacking the global characteristics necessary to explain the entire composition. An active learning strategy enables AFE to exclude locally fit models and identify a globally fit model, i.e., the true hypothesis set. Figure 2.2 is a graph showing the CH_4 conversion and C_2 selectivity of the catalysts used in this study, and you can see the performance of each catalyst along with its name. Figure 2.3 is a graph showing only the measured and predicted values of the catalysts from the dataset used in this study. In this graph, the measured values of the original and added BaO-based catalysts in each cycle are plotted on the x-axis, while the predicted values from the each design hypothesis are shown on the y-axis. As the loop progresses, you can see that the predictions become more accurate.

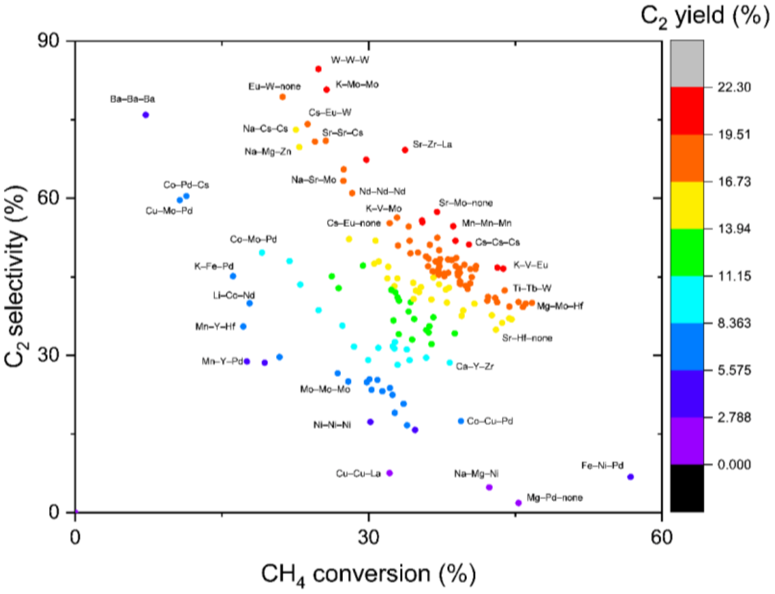


Figure 2.2. Scatter plot of 175 catalysts experimentally tested in previous and current studies. The x-axis represents CH_4 conversion (%), the y-axis represents predicted C_2 selectivity (%), and the color of the dot indicates the C_2 yield (%).

Table 2.1 shows the evolution of design hypotheses and active learning cycles of BaO-based catalysts. Notably, the MAE_{CV} values were similar to the MAE_{train} values across all the cycles, indicating an absence of overfitting. In the first cycle, the exploration of the catalyst space using FPS led to a slight increase in the MAE values for both training and cross-validation. However, in subsequent cycles, these values remained relatively unchanged. The final MAE values (2.2–2.3%) were slightly higher than the typical experimental error (1.5–2.0%). This is likely because the linear model failed to capture sudden deactivation phenomena in some catalysts, such as carbonization or clogging of the reaction tube due to dissolution. When the data points with 0% C_2 yield were excluded, the MAE decreased to approximately 1.9%. However, the elimination of less robust design hypotheses that did not apply to the newly added data resulted in variations in the selected features. The selected features are ordered in Table 2.1 according to their importance. gs_mag_moment_std appears frequently, but it is not the most important feature.

Figure 2.4a and b provide a summary of the relevant scores and individual test results, respectively. In the first cycle, the largest diversification of catalyst composition driven by FPS moderately increased the $\mathrm{MAE}_{\mathrm{train,CV}}$ values, but subsequent cycles did not largely change these values. The final $\mathrm{MAE}_{\mathrm{train,CV}}$ values (2.2–2.3%) were higher than the typical experimental error (1.0–2.0%), partly because the linear model failed to capture various 0% C_2 yield data (any observed inactivity may be attributed to several reasons). Excluding these data points reduced the $\mathrm{MAE}_{\mathrm{CV}}$ to ~1.9%. The changes in the test score were larger than those in the training and CV scores. Several extrapolations occurred during the first cycle, where the predicted yield was >30% or <0%, resulting in an extremely large $\mathrm{MAE}_{\mathrm{test}}$. These extrapolations correspond to the model attempting to explain catalysts entirely beyond its original consideration. As the cycle progressed and the catalysts in the training dataset diversified sufficiently, these extrapolations disappeared, and the difference between the observations and predictions decreased monotonically. Pearson's correlation coefficient between the regression models increased from 0.6 in Cycles 0 and 1 to 0.9 in Cycles 3 and 4, indicating the convergence of feature engineering toward a global model.

Table 2.1. Development of design hypotheses for BaO-based catalysts throughout the active learning cycles. This table shows the scores and selected features of the design hypotheses obtained in each individual active learning cycle.

Cycle	MAE$_{\text{CV (train)}}$[a]	Selected features[b]
0	1.73% (1.69%)	1. 1/ln(*en_allen_max*) 2. ln(*num_s_valence_min*) 3. 1/ln(*atomic_radius_rahm_min*) 4. (*en_pauling_ave*)$^{1/2}$ 5. (*gs_mag_moment_max*)^{3} 6. 1/ln(*vdw_radius_ave*) 7. 1/ln(*dipole_polarizability_min*) 8. (*atomic_number_std*)$^{1/2}$
1	2.46% (2.43%)	1. ln(*covalent_radius_pyykko_min*) 2. 1/(*sound_velocity_max*) 3. 1/ln(*fusion_enthalpy_min*) 4. (*en_pauling_ave*)^{2} 5. exp(*gs_mag_moment_std*) 6. (*num_s_valence_min*)$^{1/2}$ 7. ln(*electron_affinity_ave*) 8. 1/exp(*lattice_constant_std*)
2	2.38% (2.28%)	1. 1/(*gs_est_bcc_latcnt_min*) 2. (*gs_mag_moment_std*)$^{1/2}$ 3. 1/ln(*electron_affinity_ave*) 4. 1/exp(*covalent_radius_pyykko_triple_min*) 5. 1/(*sound_velocity_max*)$^{1/2}$ 6. ln(*num_s_valence_min*) 7. *gs_energy_ave* 8. (*covalent_radius_slater_ave*)^{3}
3	2.36% (2.28%)	1. (*num_s_valence_min*)$^{1/2}$ 2. 1/(*gs_est_fcc_latcnt_min*) 3. *gs_mag_moment_std* 4. 1/ln(*electron_affinity_ave*) 5. 1/(*covalent_radius_pyykko_min*)$^{1/2}$ 6. (*hhi_r_max*)$^{1/2}$ 7. 1/ln(*sound_velocity_max*) 8. (*atomic_radius_rahm_pro*)^{3}
4	2.29% (2.24%)	1. 1/(*Polarizability_min*)$^{1/2}$ 2. 1/(*dipole_polarizability_min*)$^{1/2}$ 3. (*first_ion_en_max*)^{3} 4. *hhi_r_max* 5. 1/ln(*gs_mag_moment_min*) 6. (*gs_mag_moment_std*)^{2} 7. 1/exp(*electron_affinity_pro*) 8. *lattice_constant_min*

[a] The MAE of C_2 yields during CV and training is shown in parentheses.

[b] Eight features were selected to minimize the MAE_{CV} value and are listed in order of permutation feature importance. The features are described in the same manner as XenonPy, and their meanings can be found in the Table S1.

According to Figure 2.4a, the model's convergence is evident in the reduction of extreme predicted values (over 30% or below 0%) that initially appeared. These outliers were due to the model's limited early scope in the catalytic space, which caused local fitting errors. As training cycles advanced with more diverse catalysts, these errors gradually reduced, and the gap between observed and predicted yields diminished.

Model performance evolution is visually tracked using t-distributed stochastic neighbor embedding（t-SNE）,[27] as shown in Figure 2.4b This technique compresses the features selected in each active learning cycle to two dimensions while maintaining catalyst similarities, illustrating the model's developing "recognition" of diverse catalyst designs. The plot includes all 4,060 catalysts, both tested and untested, with colors representing the predicted C_2 yields. With each cycle, active learning clarifies data clusters, refining the feature space and enabling more distinct composition-performance relationships. The expanding catalyst space in the t-SNE visualization highlights increasingly precise yield predictions as the model progresses.

Figure 2.3. Active learning implemented for OCM catalyst design. The model scores and test results for active learning are shown in (a) and (b), respectively. (a) displays the model scores alongside test results. According (b), over the course of the active learning cycles, the discrepancy between predicted and observed C_2 yields steadily decreased. Evolution

in the feature space and prediction is plotted based on t-SNE in (c), where circled data points correspond to the test results except for the last cycle adopting the training data. The color reflects either the predicted or observed C_2 yield, and the counters indicate the Gaussian kernel density estimation for the C_2 yield above 18%.

Figure 2.4c visualizes the progress of feature engineering using t-distributed stochastic neighbor embedding (t-SNE), where the eight features selected during each active learning cycle were reduced in two dimensions, maintaining the pairwise similarities of the catalysts. This approach allowed us to monitor the evolution of the machine's ability to perceive individual catalysts. The plot shows all 4060 catalysts in the library (including both tested and untested ones), with the color indicating the predicted C_2 yield and circled data points representing the test results. Leveraging the advancements in active learning, the data were divided into a larger number of clusters, representing the machine's process of refining a feature space to distinguish the catalysts better through distinct composition–performance relationships. Then, the question is how does the machine perceive the composition-performance relationships? This was addressed in two steps. First, the dataset was subjected to manual statistical analysis, as shown in Figure 2.5. Early transition metals such as Mo and Zr and heavy alkali metals such as K and Cs are attributed high performance (Figure 2.5a, b). This is because early transition metals can form oxometalate anions active for OCM when they are combined with Ba in the support or other supported elements with low electron affinity. Alkali metals can enhance the C_2 selectivity by strengthening the basicity of alkali earth metal oxides. By contrast, late transition metals (excluding Zn with completely filled 3d orbitals) tend to decrease the C_2 yield with increasing group number (Figure 2.5a, c), as they act as combustion catalysts.

Next, keeping the abovementioned researcher's observations in mind, the machine's perception was interpreted by analyzing the distribution of individual elements in the feature space (Figure 2.7). Figure 2.6. summarizes the regions where individual elements are concentrated after active learning, which decodes the machine perception. Late transition metals form separate clusters, whereas Mo and W are concentrated in narrow regions, indicating that the machine recognizes these elements as having differently significant impacts on the performance. By contrast, elements with a wide spatial distribution either have limited data points (e.g., La) or exhibit significantly different performance depending on their combination (e.g., Mg and Mn). Elements with overlapping distributions are not only similar in their physicochemical properties but also in their impact on the catalytic performance. For example, high-performing K and Cs have overlapping distributions, whereas the less-effective Li and Na are separated. These observations align with the researchers' understanding acquired from Figure 2.7. An application of the same analysis to the unselected feature set and the feature set selected before active learning (Figure 2.7) revealed the essentiality of both feature engineering and active learning in achieving such level of discrimination. Eventually, AFE transformed general physicochemical knowledge of elements into an OCM-specific one, while active learning enhanced the machine's accuracy in discriminating elements. The visualization of the feature space is also valuable for uncovering combinatorial rules (Figure.2.8). For example, catalysts containing both high-performing Mo and low-performing Pd are found within the cluster of Pd-based catalysts, suggesting that Pd has a more dominant influence than Mo in OCM. Strongly interacting combinations, such as those of Cs with Ti, Zr, and Mo, that are frequently observed in high-performing catalysts, are distributed in small clusters separated from the main cluster for Cs-based catalysts.

Additionally, Fe-Zn, while not prominently featured in the training data, is isolated in a very narrow region with relatively high predicted C_2 yields, an aspect to be explored further.

Figure 2.4. Relationship between the supported elements and the performance of catalysts containing them. In (**a**), the distribution of C_2 yield for catalysts containing each element is shown in the box and swarm plots. Elements that are frequently contained in the catalysts with the C_2 yield above 18% (**b**) and below 7% (**c**).

Figure 2.5. Machine recognition for OCM catalyst design. The feature space of the latest model is visualized by t-SNE, along with the Gaussian kernel density estimation for the C_2 yield above 18%. The dotted lines indicate the region where catalysts containing each element concentrate.

Figure 2.6. Element-wise machine recognition for the OCM catalyst design. The feature space of the latest model is visualized by t-SNE. The black dots represent catalysts containing a specified element with their estimated Gaussian kernel density. The color corresponds to the predicted C_2 yield in the range of 25% (red) to 0% (blue).

Figure 2.7. Development of the machine recognition for OCM catalyst design. In **a**–**c**, all the features in the library, the eight features engineered for the initial training data, and the eight engineered features in the end of the active learning are used as the input of t-SNE visualization, respectively. The distribution of a specified element in the respective t-SNE spaces is presented with black dots and Gaussian kernel density estimation.

Figure 2.8. Visualization of combinatorial rules using t-SNE. The black dots represent catalysts containing a specified binary combination with the estimated Gaussian kernel density. The color corresponds to the predicted C_2 yield in the range of 25% (red) to 0% (blue).

2.2.6. Design of high-yield catalysts

The primary advantage of AFE, particularly when combined with active learning, lies in its high predictive accuracy and applicability across a wide range of catalysts. To showcase this, we applied FPS to a subset of catalysts with predicted C_2 yields $\geq$ 15% using the model obtained after active learning; this resulted in the recommendation of 36 catalysts. Subsequent experimental evaluation revealed that 30 out of the 36 catalysts actually exhibited C_2 yields $\geq$ 15%, with 16 of them surpassing a yield of 18% (Figure 2.9). This is compared to only 37 cases exceeding a yield of 18% among 175 catalysts in the training data. These catalysts predominantly comprise elements whose oxides possess high basicity, such as alkaline, alkaline earth, and rare earth metal elements, along with early transition metal elements from groups 4 to 6. By contrast, many of the high-performing catalysts identified in Figure 2.9 do not conform to this pattern, with a notable presence of elements like Fe and Zn. These elements are largely underexplored in the history of OCM research. A unique advantage of this methodology lies in utilizing the integration of AFE and HTE to systematize the model's education, rather than solely focusing on catalyst discoveries. As a result, the model, enhanced through active learning, significantly streamlined the discovery of high-performing catalysts.

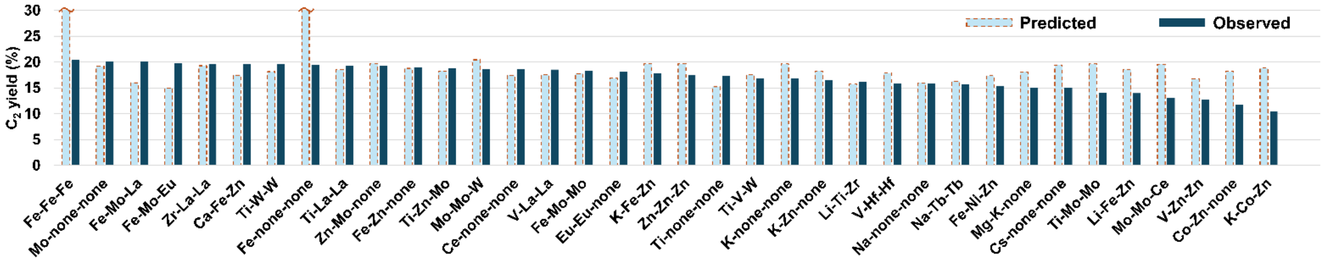


Figure 2.9. 18 catalysts with predicted C_2 yields of 15% or more were tested.

The preceding discussion has highlighted the value of the model using engineered features for understanding catalyst design principles and identifying various high-performing catalysts. However, directly extracting physical insights from these engineered features remains challenging. While the engineered features, whether individually or in combination, exhibit statistical correlations with catalytic performance, statistical correlation does not imply causation in catalysis. Moreover, the physical properties of individual elements used to construct catalyst features are often too far removed from the causal factors in catalytic behavior.

For example, the model obtained after active learning represents a combination of features: 22.0 (first_ion_en_max)^{3} + 3.32 ln(gs_mag_moment_min)$^{-1}$ − 8.63 (Polarizability_min)$^{-0.5}$ + 4.59 (dipole_polarizability_min)$^{-0.5}$ − 4.22 lattice_constant_min − 6.44 exp(electron_affinity_pro)$^{-1}$ + 10.0 (gs_mag_moment_std)^{2} + 3.26 hhi_r_max +27.8, with features such as *Polarizability_min*, *dipole_polarizability_min*, and *first_ion_en_max* identified as particularly impactful. These features help differentiate between elements with strongly basic oxides, those that aid O_2 activation, and late transition metals that often catalyze unselective combustion. However, these interpretations arise not directly from the engineered features but are instead post hoc explanations based on existing knowledge. Therefore, in order to develop a catalyst using this method, it is necessary to either gather sufficient data on the specific composition of the target, borrow design hypotheses on other catalysts, or develop a more efficient learning method.

2.4. CONCLUSION

In conclusion, AFE was validated using BaO-supported catalysts for OCM, demonstrating its effectiveness in applying machine learning (ML) to small, diverse datasets of solid catalysts. AFE excelled in automatically generating system-specific features without prior knowledge and effectively identified optimal feature combinations. The success of active learning with AFE was largely attributed to the consistent datasets produced through high-throughput experimentation (HTE) and the expansion of the search space via farthest point sampling (FPS). By integrating these techniques into an iterative learning cycle, the efficiency of ML model training was significantly enhanced, allowing for the exclusion of alternative hypotheses and the identification of robust hypotheses applicable to a wide range of catalyst compositions.

REFERENCE

(1) Martín, A. J.; Mitchell, S.; Mondelli, C.; Jaydev, S.; Pérez-Ramírez, J. Unifying Views on Catalyst Deactivation. *Nat Catal* **2022**, *5* (10), 854–866. https://doi.org/10.1038/s41929-022-00842-y.

(2) Heard, A. W.; Suárez, J. M.; Goldup, S. M. Controlling Catalyst Activity, Chemoselectivity and Stereoselectivity with the Mechanical Bond. *Nat Rev Chem* **2022**, *6* (3), 182–196. https://doi.org/10.1038/s41570-021-00348-4.

(3) Keller, G. Synthesis of Ethylene via Oxidative Coupling of Methane I. Determination of Active Catalysts. *Journal of Catalysis* **1982**, *73* (1), 9–19. https://doi.org/10.1016/0021-9517(82)90075-6.

(4) Fischer, F.; Tropsch, H. The Preparation of Synthetic Oil Mixtures (Synthol) from Carbon Monoxide and Hydrogen. *Brennstoff-Chemie* **1923**, *4*, 276–285.

(5) Cruellas, A.; Bakker, J. J.; Van Sint Annaland, M.; Medrano, J. A.; Gallucci, F. Techno-Economic Analysis of Oxidative Coupling of Methane: Current State of the Art and Future Perspectives. *Energy Conversion and Management* **2019**, *198*, 111789. https://doi.org/10.1016/j.enconman.2019.111789.

(6) Reyes, S. C.; Iglesia, E.; Kelkar, C. P. Kinetic-Transport Models of Bimodal Reaction Sequences—I. Homogeneous and Heterogeneous Pathways in Oxidative Coupling of Methane. *Chemical Engineering Science* **1993**, *48* (14), 2643–2661. https://doi.org/10.1016/0009-2509(93)80274-T.

(7) Su, Y. Upper Bound on the Yield for Oxidative Coupling of Methane. *Journal of Catalysis* **2003**, *218* (2), 321–333. https://doi.org/10.1016/S0021-9517(03)00043-5.

(8) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Comput Mater* **2017**, *3* (1), 54. https://doi.org/10.1038/s41524-017-0056-5.

(9) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2.

(10) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10* (3), 2260–2297. https://doi.org/10.1021/acscatal.9b04186.

(11) Takahashi, K.; Ohyama, J.; Nishimura, S.; Fujima, J.; Takahashi, L.; Uno, T.; Taniike, T. Catalysts Informatics: Paradigm Shift towards Data-Driven Catalyst Design. *Chem. Commun.* **2023**, *59* (16), 2222–2238. https://doi.org/10.1039/D2CC05938J.

(12) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144* (11), 4819–4827. https://doi.org/10.1021/jacs.1c12005.

(13) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Comput Mater* **2019**, *5* (1), 83. https://doi.org/10.1038/s41524-019-0221-0.

(14) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments.

Angew Chem Int Ed **2022**, *61* (29), e202204647. https://doi.org/10.1002/anie.202204647.

(15) Taniike, T.; Takahashi, K. The Value of Negative Results in Data-Driven Catalysis Research. *Nat Catal* **2023**, *6* (2), 108–111. https://doi.org/10.1038/s41929-023-00920-9.

(16) Qi, C. R.; Yi, Li; Su, Hao; Guibas, J. Leonidas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.

(17) Shi, S.; Wang, X.; Li, H. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. arXiv 2018. https://doi.org/10.48550/ARXIV.1812.04244.

(18) Guo, M.-H.; Cai, J.-X.; Liu, Z.-N.; Mu, T.-J.; Martin, R. R.; Hu, S.-M. PCT: Point Cloud Transformer. **2020**. https://doi.org/10.48550/ARXIV.2012.09688.

(19) Taniike, T.; Fujiwara, A.; Nakanowatari, S.; García-Escobar, F.; Takahashi, K. Automatic Feature Engineering for Catalyst Design Using Small Data without Prior Knowledge of Target Catalysis. *Commun Chem* **2024**, *7* (1), 11. https://doi.org/10.1038/s42004-023-01086-y.

(20) Liu, C.; Fujita, E.; Katsura, Y.; Inada, Y.; Ishikawa, A.; Tamura, R.; Kimura, K.; Yoshida, R. Machine Learning to Predict Quasicrystals from Chemical Compositions. *Advanced Materials* **2021**, *33* (36), 2102507. https://doi.org/10.1002/adma.202102507.

(21) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114* (10), 105503. https://doi.org/10.1103/PhysRevLett.114.105503.

(22) Kim, C.; Pilania, G.; Ramprasad, R. From Organized High-Throughput Data to Phenomenological Theory Using Machine Learning: The Example of Dielectric Breakdown. *Chem. Mater.* **2016**, *28* (5), 1304–1311. https://doi.org/10.1021/acs.chemmater.5b04109.

(23) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B. P.; Ramprasad, R.; Gubernatis, J. E.; Lookman, T. Machine Learning Bandgaps of Double Perovskites. *Sci Rep* **2016**, *6* (1), 19375. https://doi.org/10.1038/srep19375.

(24) Yoshida, R. XenonPy Is a Python Software for Materials Informatics. **2018**.

(25) Huber, P. J. Robust Estimation of a Location Parameter. *Ann. Math. Statist.* **1964**, *35* (1), 73–101. https://doi.org/10.1214/aoms/1177703732.

(26) Nguyen, T. N.; Nhat, T. T. P.; Takimoto, K.; Thakur, A.; Nishimura, S.; Ohyama, J.; Miyazato, I.; Takahashi, L.; Fujima, J.; Takahashi, K.; Taniike, T. High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. *ACS Catal.* **2020**, *10* (2), 921–932. https://doi.org/10.1021/acscatal.9b04293.

Chapter 3

Comparative analysis of design hypotheses for oxidative coupling of methane using automatic feature engineering

KEYWORDS: Catalyst informatics, machine learning, high-throughput experimentation, descriptor, oxidative coupling of methane

ABSTRACT

Solid catalyst development has traditionally relied on trial-and-error approaches, limiting the broader application of valuable insights across different catalyst families. To overcome this fragmentation, we introduce a framework that integrates high-throughput experimentation (HTE) and automatic feature engineering (AFE) with active learning to acquire comprehensive catalyst knowledge. The framework is demonstrated for oxidative coupling of methane (OCM), where active learning is continued until the machine learning model achieves robustness for each of the BaO-, CaO-, La_2O_3-, TiO_2-, and ZrO_2-supported catalysts, with 333 catalysts newly tested. The resulting models are utilized to extract catalyst design rules, revealing key synergistic combinations in high-performing catalysts. Moreover, we propose a method for transferring knowledge between supports, showing that features refined on one support can improve predictions on others. This framework advances the understanding of catalyst design and promotes reliable machine learning.

3.1. INTRODUCTION

The complexity of structure–function relationships has made trial-and-error approaches crucial for developing solid catalysts.[1,2] Studies have proposed several hypotheses and evaluated various materials as potential catalysts. As improved hypotheses and catalysts emerge from previous studies, the focus of catalyst research evolves, leading to further refinement. This development typically occurs concurrently, resulting in the discovery of multiple catalyst families that differ in design concepts. An outstanding example is the oxidative coupling of methane (OCM). This reaction converts methane, which is the primary component of natural gas and biogas, into C_2 compounds in a single step.[3] OCM is considered potentially more efficient than the current two-step route, which involves steam reforming and the Fischer–Tropsch process.[4] However, achieving high yields remains challenging owing to the chemical inertness of CH_4 relative to the C_2 products. Only a few catalysts are known to achieve consistent techno-economic performance (e.g., C_2 yields exceeding 30%; otherwise, 28% with 80% selectivity) in conventional fixed-bed flow reactors.[5–7]

Despite these challenges, catalyst development has progressed since 1982,[3] leading to the identification of several high-performance catalyst families, including Li/MgO, Sr/La_2O_3, and $Mn–Na_2WO_4/SiO_2$.[8–10] In Li/MgO, the incorporation of Li increases the number of active sites on the MgO surface.[11–15] However, strategies are needed to prevent Li sublimation at high reaction temperatures. In Sr/La_2O_3, the introduction of Sr^{2+} into La_2O_3 causes lattice distortion and/or electronic modulation, which enhances the C_2 selectivity of La_2O_3, likely owing to the formation of superoxide species (O_2^-).[16,17] This development has continued, with the primary aim of strengthening the advantages of low-temperature activity. The $Mn–Na_2WO_4/SiO_2$ catalyst family is considered one of the most promising catalysts for OCM because of its high C_2 selectivity and durability resulting from synergistic interactions among its components. Na^+ stabilizes tetrahedral WO_4^{2-} species, which are active in OCM, by interacting with them and promoting the formation of the cristobalite phase of SiO_2, which further stabilizes these species. Mn^{2+} enhances the OCM process by facilitating the recovery of W^{6+}.[18,19] Several studies have focused on optimizing the conversion–selectivity trade-off through various methods, including catalyst preparation techniques and the incorporation of additional elements. However, these catalyst families are based on distinct design guidelines and have been developed largely independently, without an explicit exchange of design guidelines across families [20] A more comprehensive understanding or rational design of catalysts could be achieved by elucidating and applying the commonalities and differences among various catalyst families.

Data-driven catalysis research, also known as catalyst informatics, leverages data science techniques such as machine learning (ML) and visualization to accelerate the development and understanding of catalysts by revealing applicable trends and patterns hidden within catalyst data. However, catalyst informatics is constrained by the scarcity of catalyst data suitable for data science and the difficulty in handcrafting descriptors that capture the essence of intricate structure–function relationships.[21] I used high-throughput experimentation (HTE) to prepare and evaluate numerous solid catalysts, generating sized, qualified, and consistent datasets for various heterogeneous catalytic processes, including OCM.[22–26] Furthermore, we recently introduced an automatic feature engineering (AFE) technique that programmatically designs descriptors to capture the essence of the target catalysis, beginning with the general physical properties of the elements, such as atomic radii and electronegativity.[27] AFE generates predictive ML models with tailored descriptors without requiring researchers to make assumptions or hypotheses about the target system. Using BaO-supported catalysts for OCM as an example, a robust design hypothesis was developed by recursively refining the ML model and descriptors through active learning with AFE. This refined hypothesis can be applied across a broad range of catalysts and is effective for identifying various high-performance catalysts and understanding the underlying design guidelines.[27]

In this study, the same active learning approach integrated with HTE and AFE was applied to five OCM catalyst families. We began with previously acquired OCM catalyst data for five supports (BaO, CaO, La_2O_3, TiO_2, and ZrO_2) and conducted large-scale active learning by adding 333 catalysts to establish a robust design hypothesis for each support. These design hypotheses were then analyzed to elucidate the commonalities and differences among the five catalyst families. Notably, a design hypothesis established for one catalyst family can aid in constructing an ML model for another family based on their commonalities, thus demonstrating that knowledge gained from one catalyst family can be transferred to facilitate the design of another.

3.2. METHOD

This study used previously acquired data to establish a design hypothesis for each of the five catalyst families corresponding to different support materials. However, similar to how researchers cannot dismiss alternative hypotheses when evidence is limited, AFE cannot overlook alternative design hypotheses (ML models with differently tailored descriptors that fit the training data) when the diversity of catalysts in the training data is restricted. Consequently, we implemented an active learning strategy (Figure 3.1). This strategy employs farthest point sampling (FPS) within the descriptor space established by AFE to propose catalysts that are maximally dissimilar to those included in the training data.[28] These catalysts served as rigorous control experiments to validate the proposed design hypothesis. The performances of the proposed catalysts were assessed using HTE to reinforce the training data and update the design hypothesis via AFE. This iterative process aims to eliminate design hypotheses that do not generalize well across catalysts, resulting in a robust and experimentally validated design hypothesis. Further details are provided below.

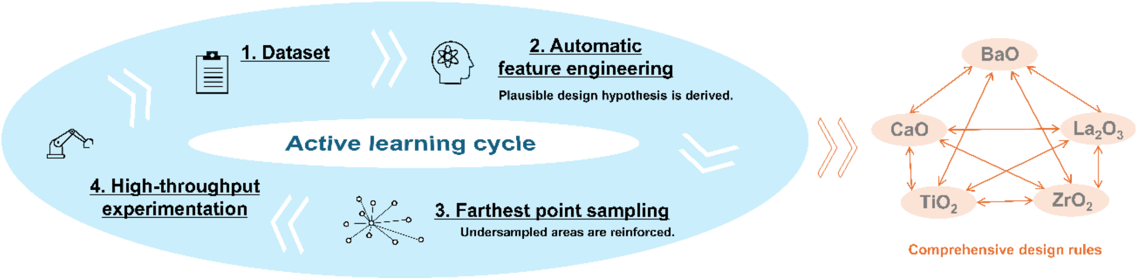


Figure 3.1. Active learning cycle employed in this study. Automatic feature engineering (AFE) was applied to a given catalyst dataset to derive a design hypothesis (an machine learning (ML) model with tailored descriptors). Catalysts recommended by farthest point

sampling (FPS) were assessed using high-throughput experimentation (HTE). The resulting data were integrated back into the dataset, and this iterative process continued until a robust design hypothesis was established to elucidate the relationship between catalyst compositions and performances.

3.2.1. Dataset

Taniike group accumulated OCM data for quaternary catalysts represented as M1–M2–M3/Support using a consistent experimental protocol involving HTE.[22,27,29,30] M1–M3 correspond to the supported elements, which can be selected (with duplication allowed) from the following: Li, Na, Mg, K, Ca, Ti, V, Mn, Fe, Co, Ni, Cu, Zn, Sr, Y, Zr, Mo, Pd, Cs, Ba, La, Ce, Nd, Eu, Tb, Hf, W, and none (where "none" indicates no addition of elements). The selected supports were MgO, Al_2O_3, SiO_2, CaO, TiO_2, ZrO_2, BaO, La_2O_3, and CeO_2. The loading amount of supported elements, except for the "none" option, was set at 0.37 mmol per gram of support for each selection. This resulted in a parameter space containing 4,060 catalysts per support, amounting to a total of 36,540 catalysts.

In this study, we extracted 381 catalysts related to CaO, BaO, La_2O_3, TiO_2, and ZrO_2 supports from the 636 quaternary catalysts we previously reported to establish design hypotheses through active learning.[30–32] CaO and La_2O_3 were selected because they are the most extensively studied basic oxides in OCM and are likely to exhibit distinct catalyst designs.[33] BaO was selected as the reference for these supports because it offers higher C_2 yields and selectivities. In contrast, the redox-active supports TiO_2 and ZrO_2 were included to investigate potential relationships between the general physical properties and design hypotheses. Among the 381 catalysts, 175 were obtained via random sampling from the entire space, whereas the remaining 206 were selected to validate various ML techniques.[30–32]

3.2.2. Automatic feature engineering

AFE automates the design of physically meaningful features for a given catalyst dataset within the framework of supervised ML. This process involves a structured pipeline of feature assignment, synthesis, and selection. First, physical quantities of elements are assigned to the catalysts, with their elemental compositions represented through commutative operations. Higher-order features involving nonlinear and combinatorial effects are then synthesized from these assigned primitive features using mathematical operations. Finally, a specified number of features (descriptors) that optimize the score of the supervised ML are selected from the large synthesized array of features.

In this study, we utilized 58 parameters of elements from XenonPy, normalized according to the literature[34]These parameters were assigned to each catalyst using five commutative operations (maximum, minimum, average, product, and standard deviation), yielding 290 primary features. These primary features were further synthesized into 3,480 features using 12 functional forms (x, sqrt(x), x^2, x^3, ln(x), exp(x), and their reciprocals, where x represents each primary feature). A genetic algorithm-based approach was employed to select eight features that minimized the mean absolute error (MAE) in leave-one-out cross-validation (LOOCV) using Huber regression.[35] This process involved assessing approximately 4,000,000 models per dataset with various feature combinations and selecting the combination of features (X) and model (f(X)) that yielded the lowest cross-validation (CV) score as the most plausible design hypothesis. Huber regression, which is a form of multiple linear regression, was employed to prevent overfitting owing to its reduced number of parameters and ensure robustness against outliers, such as experimental failures. The number of selected features was empirically determined to balance the CV score and the cost of feature selection. Further details on AFE and parameter selection are provided in previous study.[27]

3.2.3. Farthest point sampling

To validate and refine the design hypothesis presented in Section 2.2, we added either 10 or 20 catalysts to each active learning cycle for experimental testing. Among these, 90% were selected using FPS within the normalized eight-dimensional feature space established by AFE. The remaining 10% corresponded to the re-evaluation of the catalysts that exhibited the largest deviations between observed and predicted values in the last cycle.

3.2.4. High-throughput experimentation

The catalysts proposed in Section 2.3 were prepared and evaluated using the same experimental methods and conditions as those employed when acquiring the original training data, which are briefly described as follows.

3.2.4.1. Materials

The metal precursors used were $LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, KNO_3, $Ca(NO_3)_2 \cdot 4H_2O$, $Ti(OiPr)_4$, $VOSO_4 \cdot xH_2O$ ($x = 3–5$), $Mn(NO_3)_2 \cdot 6H_2O$, $Fe(NO_3)_3 \cdot 9H_2O$, $Co(NO_3)_2 \cdot 6H_2O$, $Ni(NO_3)_2 \cdot 6H_2O$, $Cu(NO_3)_2 \cdot 3H_2O$, $Zn(NO_3)_2 \cdot 6H_2O$, $Sr(NO_3)_2$, $Y(NO_3)_3 \cdot 6H_2O$, $ZrO(NO_3)_2 \cdot xH_2O$ ($x = 2$), $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, $Pd(OAc)_2$, $CsNO_3$, $Ba(NO_3)_2$,

$La(NO_3)_3 \cdot 6H_2O$, $Ce(NO_3)_3 \cdot 6H_2O$, $Nd(NO_3)_3 \cdot 6H_2O$, $Eu(NO_3)_3 \cdot 5H_2O$, $Tb(NO_3)_3 \cdot 5H_2O$, $Hf(OEt)_4$, and $(NH_4)_{10}H_2(W_2O_7)_6$. These materials were purchased from one of the following suppliers: Sigma-Aldrich, Kanto Chemical, Wako Pure Chemical Industries, Alfa-Aesar, or Sumitomo Chemical.

The oxide supports and their precursors included $Ca(OH)_2$ (3.0 m^2/g, Wako Pure Chemical Industries), $Ba(OH)_2 \cdot 8H_2O$ (1.1 m^2/g, Wako Pure Chemical Industries), La_2O_3 (8.3 m^2/g, Wako Pure Chemical Industries), TiO_2 (17.4 m^2/g, anatase type, Kanto Chemical), and ZrO_2 (3.2 m^2/g, Kanto Chemical).

3.2.4.2. Preparation of catalysts

The catalysts were prepared using a parallelized wet impregnation method, in which support powder was impregnated with an aqueous solution of specified metal precursors at 50 °C for 6 h. The loading amount of elements was fixed at 0.37 mmol/g-support per selection within M1–3. Following impregnation, the powder was vacuum-dried and calcined in air at 1000 °C for 3 h. For precursors that contained metal alkoxides, impregnation was performed in two steps: first, with an aqueous solution of the other precursors, followed by impregnation with an ethanol solution of the metal alkoxides.

3.2.4.3. Evaluation of catalysts

The catalysts were evaluated for OCM using a custom-built HTE system.[22] This system functions by integrating a gas mixer, flow distributor, reactors housed within an electric furnace, an autosampler, and a quadrupole mass spectrometer (QMS). This setup allows the automated evaluation of 20 catalysts under a predefined sequence of reaction conditions. The catalysts were secured in beds 1 cm in height within reaction tubes made of fused quartz, which had inner diameters of 4 and 2 mm, using quartz wool. After inline activation at 1000 °C for 3 h in an oxygen stream, the 20 catalysts were tested under 135 conditions varying in temperature (700, 750, 800, 850, and 900 °C), total gas flow rates (10, 15, and 20 mL/min/channel), CH_4/O_2 ratios (2, 4, and 6 mol/mol), and Ar partial pressures as a balancing gas (0.15, 0.40, and 0.70 atm). Each catalyst was labeled according to the highest C_2 yield achieved among all 135 conditions.

3.3. RESULTS AND DISCUSSION

3.3.1. Active learning for obtaining robust design hypotheses

In conventional hypothesis validation, researchers add data not only to verify the primary hypothesis but also to eliminate alternative hypotheses that could explain the original data. Similarly, when the training data are limited, the AFE may generate multiple design hypotheses with similar scores, which are considered alternative design hypotheses. These arise owing to multicollinearity among features within the given data[23] Eliminating incorrect or less robust hypotheses requires active learning. In this study, catalysts were selected using FPS to enhance the diversity of the data and assess the robustness of the proposed design hypotheses. Design hypotheses that were not applicable to diverse data were filtered out through repeated active learning cycles, resulting in more robust and reliable design hypotheses. According to chapter 2, the active learning cycle was conducted five times and, in some cases, six times.[27]

Table 3.1 shows the evolution of design hypotheses and active learning cycles of CaO-based catalysts. Notably, the MAE_{CV} values were similar to the MAE_{train} values across all the cycles, indicating an absence of overfitting. Moreover, both values remained within a narrow range of 1.6–1.9%, which is comparable to the experimental error (1.5–2% in C_2 yields). This consistency suggests that the AFE process identified design hypotheses that fit the training data similarly, despite its diversity increasing with the number of cycles. However, the elimination of less robust design hypotheses that did not apply to the newly added data resulted in variations in the selected features. The selected features are ordered in Table 3.1 according to their importance. Three features (covalent_radius_pyykko_double_min, atomic_radius_rahm_min, and covalent_radius_cordero_min) consistently appeared in most cycles with high importance, despite slight differences in functional forms. Therefore, these features are irreplaceable descriptors of CaO-based catalyst performance. Other less important features varied across cycles, reflecting the elimination of less robust design hypotheses.

Less robust design hypotheses are eliminated through active learning; however, this does not necessarily lead to convergence to a single design hypothesis. Certain feature combinations may exhibit significant collinearity across the entire space (i.e., 4059 catalysts), rendering them indistinguishable regardless of additional data. Such global collinearity suggests that the design hypotheses, despite possessing differing features, exhibit similar predictive behavior throughout the entire space and are therefore considered similarly robust. To address this, we compared the predictive behavior of the design hypothesis obtained in one cycle with that obtained in the previous cycle. Specifically, we predicted the C_2 yields for 4059 catalysts using two design hypotheses and calculated the MAE between the two prediction sets. Figure 3.2 shows the development of MAE values across the active learning cycles. For all the supports, the MAE value significantly decreased during the early stages of active learning. This indicates that the less robust design hypotheses proposed by AFE, owing to limited data, were rapidly corrected in these early stages. In the later stages, the changes in MAE became considerably smaller, indicating that the predictive behavior of the design hypotheses stabilized with the addition of data. In the final cycle, the MAE values for all the supports were below 2%, which is comparable to or lower than the experimental error (further reductions in the MAE are considered negligible owing to limitations in experimental precision). Based on this, we concluded that the design hypotheses were sufficiently robust for use in subsequent analyses.

The active learning process involving the five supports added 333 catalysts to the original dataset of 381 catalysts. Their measured OCM performances are summarized in Figure 3.3 and are digitally accessible on a database platform (https://cads.eng.hokudai.ac.jp/).[24]

Table 3.1. Development of design hypotheses for CaO-based catalysts throughout the active learning cycles. This table shows the scores and selected features of the design hypotheses obtained in each individual active learning cycle.

Cycle	$\text{MAE}_{\text{CV (train)}}$[a]	Selected features[b]
0	1.65% (1.70%)	1. 1/(*covalent_radius_pyykko_double_min*) 2. ln(*covalent_radius_cordero_min*) 3. (*atomic_radius_rahm_min*)$^{1/2}$ 4. (*density_max*)^{2} 5. 1/ln(*thermal_conductivity_ave*) 6. ln(*density_ave*) 7. (*atomic_weight_min*)^{3} 8. 1/exp(*gs_est_fcc_latcnt_max*)
1	1.75% (1.67%)	1. 1/(*covalent_radius_cordero_min*) 2. 1/(*covalent_radius_pyykko_min*) 3. (*atomic_radius_rahm_min*)$^{1/2}$ 4. 1/(*covalent_radius_pyykko_double_min*)$^{1/2}$ 5. ln(*atomic_radius_rahm_min*) 6. exp(*num_f_unfilled_max*) 7. (*gs_mag_moment_max*)^{3} 8. ln(*period_ave*)

2	1.85% (1.79%)	1. (*first_ion_en_max*)^2 2. 1/(*polarizability_min*)$^{1/2}$ 3. (*atomic_radius_rahm_min*)$^{1/2}$ 4. 1/(*covalent_radius_pyykko_double_min*)$^{1/2}$ 5. (*num_f_unfilled_std*)$^{1/2}$ 6. (*num_d_valence_max*)^2 7. (*hhi_p_max*)^3 8. *num_f_valence_ave*
3	1.63% (1.56%)	1. 1/(*covalent_radius_pyykko_double_min*) 2. 1/(*covalent_radius_cordero_min*) 3. (*atomic_radius_rahm_min*)$^{1/2}$ 4. (*num_d_valence_max*)^2 5. 1/(*covalent_radius_pyykko_triple_min*) 6. ln(*vdw_radius_uff_max*) 7. 1/(*hhi_p_ave*) 8. 1/exp(*num_f_unfilled_std*)
4	1.67% (1.62%)	1. 1/(*covalent_radius_pyykko_double_min*)$^{1/2}$ 2. (*atomic_radius_rahm_min*)$^{1/2}$ 3. (*covalent_radius_cordero_min*)$^{1/2}$ 4. (*covalent_radius_pyykko_triple_min*)$^{1/2}$ 5. *covalent_radius_cordero_ave* 6. (*num_d_valence_std*)^2 7. exp(*fusion_enthalpy_max*) 8. *heat_capacity_mass_std*
5	1.67% (1.65%)	1. (*atomic_radius_rahm_min*)$^{1/2}$ 2. 1/(*covalent_radius_pyykko_double_min*)$^{1/2}$ 3. ln(*covalent_radius_cordero_min*) 4. (*bulk_modulus_max*)^2 5. 1/(*sound_velocity_max*)$^{1/2}$ 6. 1/ln(*covalent_radius_pyykko_double_std*) 7. 1/exp(*heat_capacity_mass_min*) 8. exp(*gs_mag_moment_max*)

[a] The MAE of C_2 yields during CV and training is shown in parentheses.

[b] Eight features were selected to minimize the MAE_{CV} value and are listed in order of permutation feature importance. The features are described in the same manner as XenonPy, and their meanings can be found in the Table S1.

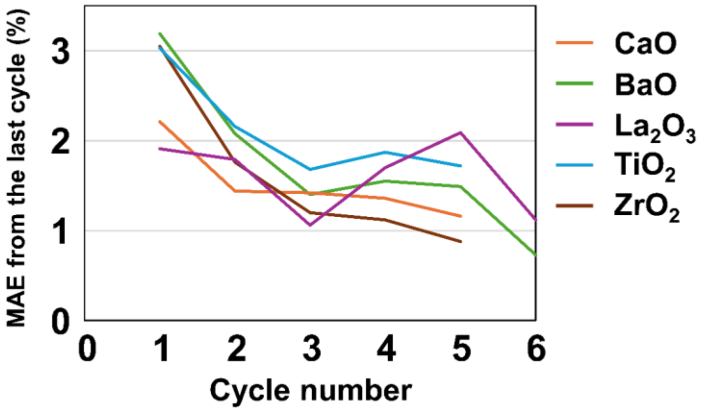


Figure 3.2. Development of the predictive behavior of design hypotheses through active learning. The MAE was calculated by comparing the predicted C_2 yields of 4059 catalysts between the design hypothesis obtained in one cycle and that from the previous cycle. Lower MAE values indicate greater similarity between the predictive behaviors of the two design hypotheses. In the early stages of active learning, MAE values decreased significantly, reflecting the elimination of less robust design hypotheses as data were added. In the later stages, additional data led to minor changes in MAE, suggesting that the design hypotheses became similarly robust. Note that any accidental increases in MAE values during certain cycles were attributed to the random nature of the genetic algorithm during feature selection; in such cases, the sixth cycle was included.

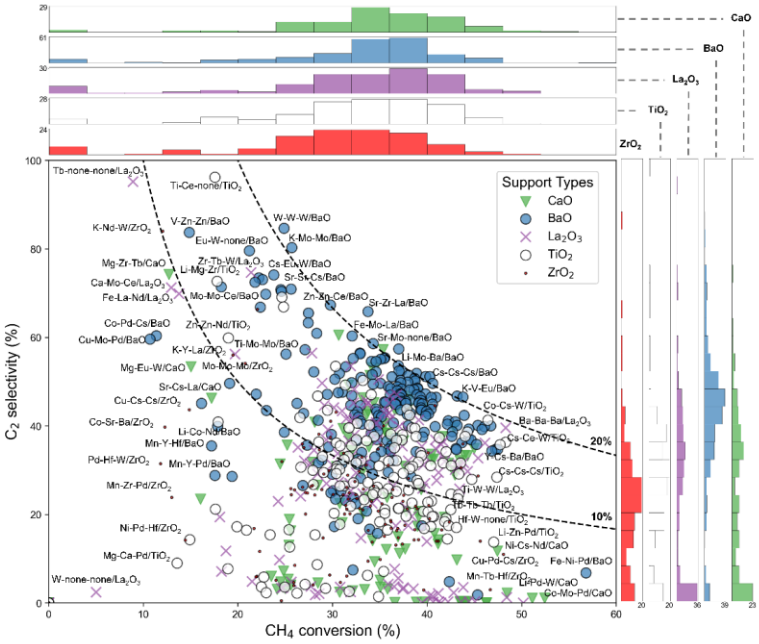


Figure 3.3. Scatter plot of 726 catalysts experimentally tested in previous and current studies. The x-axis represents CH_4 conversion, and the y-axis represents C_2 selectivity. Different supports are indicated by symbols of varying colors and shapes. Histograms outside the plot area show the data distribution for each support.

3.3.2. Extracting catalyst design guidelines

The design hypotheses obtained in Section 3.1 for the five catalyst families were applied to extract catalyst design guidelines. Specifically, the relationships between the elemental compositions of the catalysts and their predicted C_2 yields were analyzed to identify design rules, including high-performance designs for each support, along with similarities and differences across supports.

First, the distribution of predicted C_2 yields for the 4059 catalysts on each support was analyzed (Figure 3.4). The distribution varied according to the support, with the medians ranked as BaO > La_2O_3 > CaO > TiO_2 > ZrO_2. This pattern suggests that basic oxides generally achieved higher yields, which is consistent with common OCM practices. To assess the overall similarity between the supports, irrespective of absolute yield differences, Pearson correlation coefficients were calculated for the predicted C_2 yields of the 4059 catalysts across different supports (Figure 3.5). High positive correlation coefficients among the supports indicate that the design hypotheses were largely similar. The correlation coefficients among CaO, BaO, and La_2O_3 all exceeded +0.8, indicating that these basic supports not only achieve higher yields but also share similar design hypotheses. In contrast, the design hypotheses for the Group 4 oxides, TiO_2 and ZrO_2, exhibited lower similarity, which can likely be attributed to differences in their redox behaviors. Notably, La_2O_3 exhibited the highest correlation with the other supports, suggesting that its design hypothesis is generally applicable.

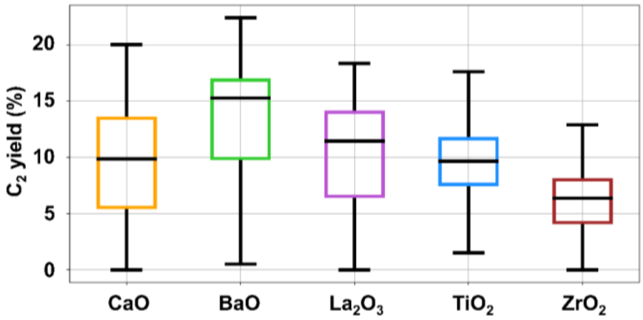
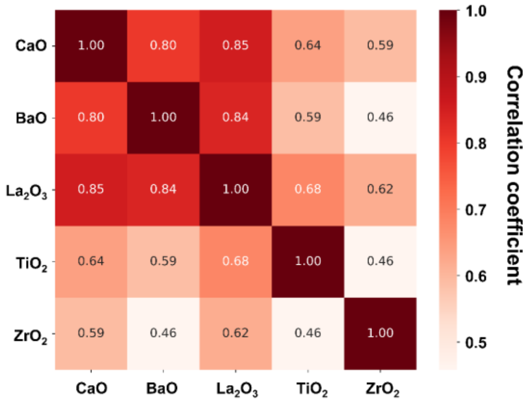


Figure 3.4. Box plot of predicted C_2 yields for each support. The line represents the median value, the box indicates the interquartile range (IQR), and the whiskers denote 1.5 times the IQR, with any data points beyond the whiskers considered outliers.



	CaO	BaO	La_2O_3	TiO_2	ZrO_2
CaO	1.00	0.80	0.85	0.64	0.59
BaO	0.80	1.00	0.84	0.59	0.46
La_2O_3	0.85	0.84	1.00	0.68	0.62
TiO_2	0.64	0.59	0.68	1.00	0.46
ZrO_2	0.59	0.46	0.62	0.46	1.00

Figure 3.5. Similarity of design hypotheses across supports. The Pearson correlation coefficients for the predicted C_2 yields of the 4059 catalysts across different supports were calculated and are presented as a heatmap. A correlation coefficient closer to 1 indicates greater similarity between the design hypotheses. Basic oxides such as CaO, BaO, and La_2O_3 exhibit similar design hypotheses, which differ from those of TiO_2 and ZrO_2.

Next, to analyze the design rules for high-performance catalysts, a subset of catalysts with predicted C_2 yields in the top 10% for each support was created, and the frequency of individual elements within this subset was examined. The results are shown in Figure 3.6a in a periodic table format, where elements with higher frequencies indicate more effective pairing with the corresponding supports. Moreover, the frequency of secondary elements associated with the three most frequent elements for each support was analyzed. Figure 3.6b shows a pie chart of the results. The insights derived from Figures 3.6a and 3.6b for each support are summarized below:

CaO: The elements most commonly found in high-performance catalysts, ranked by frequency, are Sr > Mg > Ca > Cs. Alkaline earth metals are the predominant elements, followed by rare earth and Group 4 elements. Notably, Cs is the only alkali metal that appears frequently. The effectiveness of pairing alkaline earth metals, particularly Sr, with CaO is well known, supporting the validity of the established design hypothesis.[25] The secondary elements frequently associated with the top three elements (Sr, Mg, and Ca) are shown in Figure 3.6b. The patterns of secondary elements are relatively similar among these primary elements and mainly include alkali metals, alkaline earth metals, rare earth elements, and Group 4 elements (excluding Ti). In particular, Cs is the most frequent secondary element, despite alkaline earth metal elements being expected to be more prevalent based on Figure 3.6a. This observation suggests a synergistic combination of alkaline earth metal elements and Cs on CaO, as confirmed in Figure 3.7.

BaO: The trends of the elements frequently observed in BaO differ considerably from those in CaO, which is another alkaline earth metal oxide (Figure 3.6a). No overall preference exists for alkaline earth or rare earth elements; rather, specific elements are frequently found in BaO. In particular, La exhibits a significantly higher frequency, followed by Mo, Cs, Zn, and Sr. As shown in Figure 3.6b, the elements associated with La are diverse, suggesting that the La–BaO combination alone can achieve high performance. Furthermore, Mo is frequently paired with Zn, W, and La. Notably, the Zn–La combination effectively enhances catalyst stability and prevents carbon deposition in OCM.[26] The association of Mo and W suggests that increasing the amount of Group 6 elements can improve performance, as evidenced by the observed high yield of W–W–W/BaO. Cs is frequently associated with the primary elements, La, Mo, and Sr, and a similar specificity of Cs among alkali metal elements is also observed for CaO.

La_2O_3: The elements frequently found in this basic oxide resemble those in CaO, with alkaline earth metals, rare earth elements (excluding Ce), and Group 4 elements (excluding Ti) being most prevalent. Doping La_2O_3 with elements that possess an oxidation state equal to or lower than that of La (3+) enhances its activity.[16] This aligns with the frequent occurrence of alkaline earth metals (2+) and rare earth elements (3+), excluding Ce, which typically has a 4+ oxidation state. Notably, Ba is highly prevalent, similar to the La–BaO combination observed in BaO. The diversity of the elements associated with Ba further substantiates the strong synergy between La and Ba. Previous study demonstrated that using BaO and La_2O_3 as mixed supports results in a complementary effect, where the high C_2 selectivity of BaO at high temperatures and the high activity of La_2O_3 at low temperatures collectively enhance C_2 yields.[10]

TiO_2: Alkaline metals, particularly Cs, are significantly more prevalent in this support (Cs >> K > Li). Although a Cs–TiO_2 combination has not been reported for OCM, Cs is known to enhance the photocatalytic activity of TiO_2 by promoting its reduction.[27, 28] This suggests that Cs may also facilitate the redox activity of TiO_2 in OCM. Cs is frequently associated with elements that prefer a 4+ oxidation state, such as Hf and Ce, followed by other rare earth elements.[29,30]

ZrO_2: Similar to TiO_2, alkaline metals are frequently observed in ZrO_2; however, Cs is considerably less prevalent than in TiO_2. In addition, alkaline earth metals and early transition metal elements are more prevalent. Figure 3.6b shows that alkaline metals and alkaline earth metals frequently appear alongside early transition metal elements. This suggests that ZrO_2, which exhibits minimal intrinsic OCM activity on its own, benefits from elemental combinations capable of forming oxometalate anions that are active in OCM.[30,31]

We also analyzed the design rules for the catalysts in the bottom 10% based on the predicted C_2 yields (Figure 3.8). In contrast to high-performance catalysts, the design rules for low-performance catalysts show less dependence on the support, with late transition metals being predominant. These elements are known to catalyze nonselective combustion, suggesting that they are generally effective in methane combustion, irrespective of the support.

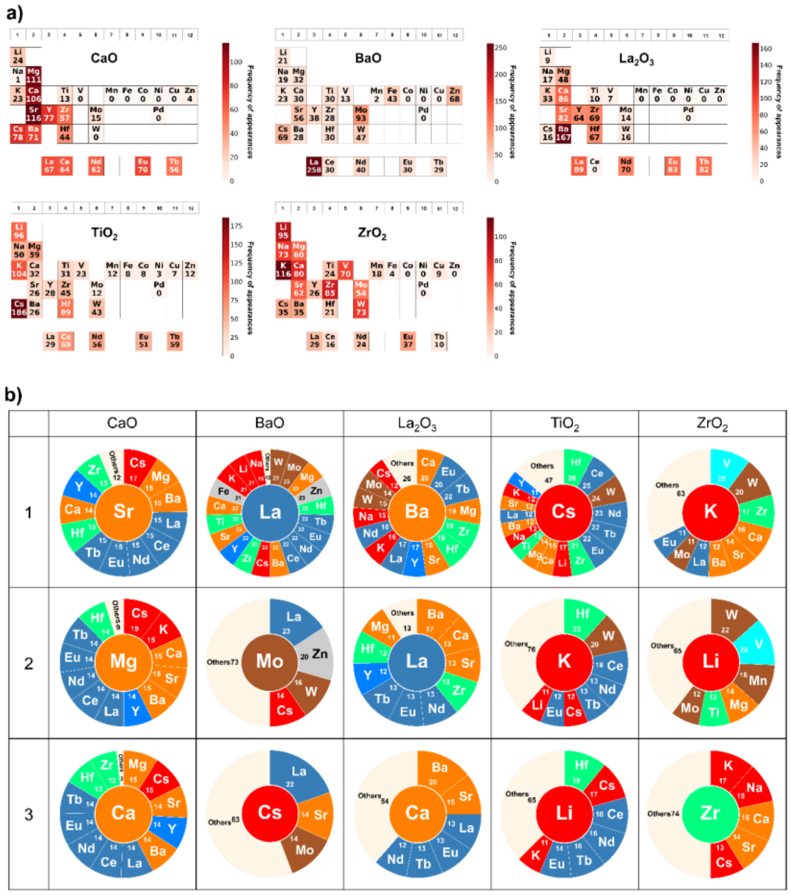


Figure 3.6. Design rules in high-performance catalysts. The characteristics of catalysts with predicted C_2 yields in the top 10% for each support are summarized as follows: (a) Heatmaps presented in a periodic table format show the frequency of individual elements in the high-performance catalysts. Elements with higher frequencies (indicated by darker red) are more likely to contribute to high performance when combined with the corresponding support. (b) Pie charts for the top three most frequent elements in (a), showing the frequency of appearances of secondary elements associated with these main

elements. A selective association between a main element and a specific secondary element suggests synergy between them.

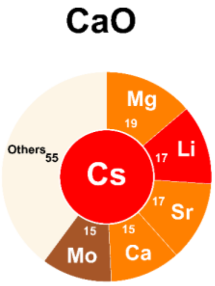


Figure 3.7. Pie chart illustrating the frequency of secondary elements associated with Cs–CaO.

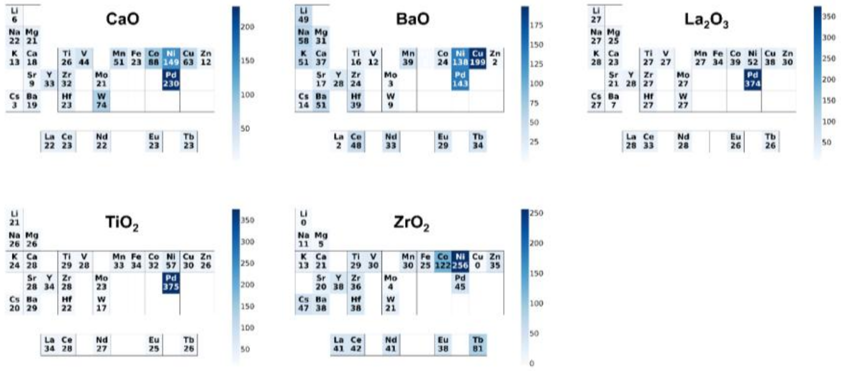


Figure 3.8. Design rules for low-performance catalysts. The characteristics of catalysts with predicted C_2 yields in the bottom 10% for each support are summarized. Heatmaps presented in a periodic table format show the frequency of individual elements in the low-performance catalysts. Elements with higher frequencies (shown in darker blue) are more likely to contribute to low performance when paired with the corresponding support.

3.3.3. Transferability of design hypotheses

Experienced catalyst researchers often apply the knowledge and insights gained from existing catalytic systems when developing catalysts for new systems. This transfer requires a certain degree of intuition, and replicating this process through data science is a key aspect of catalyst informatics. To demonstrate this concept, we aimed to transfer the design hypotheses across various catalyst families.

The underlying concept was to leverage the similarities in design hypotheses across different catalyst supports. Specifically, the design hypothesis obtained for one support was transformed into features to facilitate the development of a design hypothesis for another support through AFE. The most straightforward approach is to directly use the predicted C_2 yields for one support as a feature, as indicated by the correspondence of predicted yields across the various supports (cf. Figure 3.5). In addition, we explored the use of descriptors obtained for one support as features through dimensional reduction. This approach is based on the assumption that the similarities observed between elements or elemental combinations on one support are applicable to another support, irrespective of the C_2 yields. We reduced the original eight features to two using two distinct methods. The first method employs principal component analysis (PCA), which maximizes the variance in a higher-dimensional space to effectively represent the distribution of elemental combinations within the original feature space. The second method uses t-distributed stochastic neighbor embedding (t-SNE),[32] which prioritizes preserving the local relationships between elemental combinations in the original space to effectively capture their similarities. The features derived from design hypotheses across different catalyst families are referred to as "design hypothesis features" to distinguish them from the standard features derived from the XenonPy elemental physical properties.

Figure 3.9 compares the MAE_{CV} scores of the ML models that incorporate at least one design hypothesis feature from other supports with those that do not use such features. This comparison includes three catalyst families: CaO, La_2O_3, and ZrO_2. The design hypothesis features selected through AFE, along with their ranking based on permutation feature importance, are also shown. For CaO and La_2O_3, incorporating design hypotheses from other supports significantly reduced the MAE_{CV} scores, surpassing the fluctuations (indicated by the error bars in Figure 3.9) observed with genetic algorithm-based feature selection. In both cases, the predicted C_2 yield of BaO was the most important descriptor, which aligns with the high correlation between the predicted C_2 yields of these supports, as shown in Figure 3.5. In addition, although less critical, the t-SNE feature of TiO_2 was selected for both supports. For ZrO_2, transferring design hypotheses from other supports did not result in lower MAE_{CV} scores, which is consistent with its distinct design hypothesis compared with those of other supports, as discussed in Section 3.2. Notably, only t-SNE features were selected from the dimensionally reduced features, and none of the PCA features were selected. This suggests that the local relationships or similarities between elemental combinations are more effective for transferring knowledge than their distribution in the original feature space. This finding is promising for catalyst development because local similarities are more accessible than distributions in feature space.

In summary, we demonstrated that, within the context of AFE, knowledge can be effectively transferred between different catalyst families. This indicates that when a catalytic system exhibits similarities with a target system, insights gained from the former can facilitate the acquisition of knowledge for the latter, thereby potentially accelerating active learning to develop a robust design hypothesis.

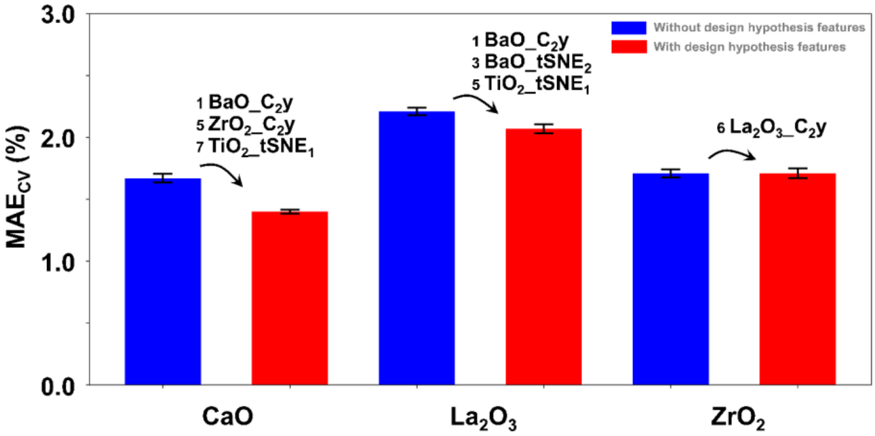


Figure 3.9. Results of knowledge transfer between catalyst families within the context of data science. Design hypotheses obtained from other supports were transformed into various features (referred to as design hypothesis features) and used in ML model development with AFE. For CaO and La_2O_3, models incorporating design hypothesis features (red bars) exhibit lower MAE_{CV} scores than models without these features (blue bars), indicating effective knowledge transfer. In contrast, ZrO_2, which lacks similarity with other supports, does not exhibit a reduction in MAE_{CV} scores. Error bars represent the standard deviation of MAE_{CV} values from three independent feature selection runs. The selected design hypothesis features and their permutation feature importance rankings are also shown.

3.4. CONCLUSION

The empirical aspects of catalyst development often lead to a focus or bias toward specific compositions, resulting in fragmented knowledge that is difficult to integrate. However, in catalyst informatics, the iterative updating of descriptors and ML models, combined with strategic data addition, can generate more comprehensive and applicable catalyst knowledge. In this study, active learning was implemented using HTE and AFE to acquire comprehensive knowledge on BaO-, CaO-, La_2O_3-, TiO_2-, and ZrO_2-supported catalysts for OCM. The knowledge acquired was then used to extract catalyst design rules and evaluate the potential for knowledge transfer across different catalyst families.

Active learning was conducted for each catalyst family, beginning with the previously acquired training data. A design hypothesis comprising an ML model and descriptors was generated using AFE, which produced numerous catalyst features and selected the most plausible ones within the context of supervised ML. Challenging catalysts were identified using FPS and experimentally tested to enhance the training data and refine the design hypothesis. This process was repeated five to six times, resulting in a robust design hypothesis applicable to all compositions and yielding comprehensive catalyst knowledge for each support.

The established design hypotheses were employed to identify catalyst design rules and assess their similarities and differences across the supports. Synergistic combinations commonly observed in high-performance catalysts were identified, including alkaline earth metals for CaO, La for BaO, Ba for La_2O_3, Cs for TiO_2, and alkali metals for ZrO_2, along with some novel findings.

Finally, a method for transferring catalyst knowledge between different supports using AFE was introduced, leveraging the similarities between the design hypotheses. Notably, the descriptors or model outputs refined through active learning on one support were adapted into features that enhanced the predictive accuracy of the ML models for other supports.

In summary, this study established a scientific data framework for acquiring comprehensive catalyst knowledge to facilitate the design of novel and diverse high-performance catalysts. Transferring validated design hypotheses across different catalytic systems can enhance model reliability and reduce the need for extensive active learning.

REFERENCE

(1) Martín, A. J.; Mitchell, S.; Mondelli, C.; Jaydev, S.; Pérez-Ramírez, J. Unifying Views on Catalyst Deactivation. *Nat Catal* **2022**, *5* (10), 854–866. https://doi.org/10.1038/s41929-022-00842-y.

(2) Heard, A. W.; Suárez, J. M.; Goldup, S. M. Controlling Catalyst Activity, Chemoselectivity and Stereoselectivity with the Mechanical Bond. *Nat Rev Chem* **2022**, *6* (3), 182–196. https://doi.org/10.1038/s41570-021-00348-4.

(3) Keller, G. Synthesis of Ethylene via Oxidative Coupling of Methane I. Determination of Active Catalysts. *Journal of Catalysis* **1982**, *73* (1), 9–19. https://doi.org/10.1016/0021-9517(82)90075-6.

(4) Fischer, F.; Tropsch, H. The Preparation of Synthetic Oil Mixtures (Synthol) from Carbon Monoxide and Hydrogen. *Brennstoff-Chemie* **1923**, *4*, 276–285.

(5) Cruellas, A.; Bakker, J. J.; Van Sint Annaland, M.; Medrano, J. A.; Gallucci, F. Techno-Economic Analysis of Oxidative Coupling of Methane: Current State of the Art and Future Perspectives. *Energy Conversion and Management* **2019**, *198*, 111789. https://doi.org/10.1016/j.enconman.2019.111789.

(6) Reyes, S. C.; Iglesia, E.; Kelkar, C. P. Kinetic-Transport Models of Bimodal Reaction Sequences—I. Homogeneous and Heterogeneous Pathways in Oxidative Coupling of Methane. *Chemical Engineering Science* **1993**, *48* (14), 2643–2661. https://doi.org/10.1016/0009-2509(93)80274-T.

(7) Su, Y. Upper Bound on the Yield for Oxidative Coupling of Methane. *Journal of Catalysis* **2003**, *218* (2), 321–333. https://doi.org/10.1016/S0021-9517(03)00043-5.

(8) Lin, C. H.; Ito, T.; Wang, J.; Lunsford, J. H. Oxidative Dimerization of Methane over Magnesium and Calcium Oxide Catalysts Promoted with Group IA Ions: The Role of [M^+O^-] Centers. *Journal of the American Chemical Society* **1987**, *109*, 4808–4810.

(9) DeBoy, J. M.; Hicks, R. F. The Oxidative Coupling of Methane over Alkali, Alkaline Earth, and Rare Earth Oxides. *Ind. Eng. Chem. Res.* **1988**, *27* (9), 1577–1582. https://doi.org/10.1021/ie00081a004.

(10) Shahri, S. M. K.; Alavi, S. M. Kinetic Studies of the Oxidative Coupling of Methane over the Mn/Na_2WO_4/SiO_2 Catalyst. *Journal of Natural Gas Chemistry* **2009**, *18* (1), 25–34. https://doi.org/10.1016/S1003-9953(08)60079-1.

(11) Qian, K.; You, R.; Guan, Y.; Wen, W.; Tian, Y.; Pan, Y.; Huang, W. Single-Site Catalysis of Li-MgO Catalysts for Oxidative Coupling of Methane Reaction. *ACS Catal.* **2020**, *10* (24), 15142–15148. https://doi.org/10.1021/acscatal.0c03896.

(12) Luo, L.; Jin, Y.; Pan, H.; Zheng, X.; Wu, L.; You, R.; Huang, W. Distribution and Role of Li in Li-Doped MgO Catalysts for Oxidative Coupling of Methane. *Journal of Catalysis* **2017**, *346*, 57–61. https://doi.org/10.1016/j.jcat.2016.11.034.

(13) Myrach, P.; Nilius, N.; Levchenko, S. V.; Gonchar, A.; Risse, T.; Dinse, K.; Boatner, L. A.; Frandsen, W.; Horn, R.; Freund, H.; Schlögl, R.; Scheffler, M. Temperature-Dependent Morphology, Magnetic and Optical Properties of Li-Doped MgO. *ChemCatChem* **2010**, *2* (7), 854–862. https://doi.org/10.1002/cctc.201000083.

(14) Zavyalova, U.; Geske, M.; Horn, R.; Weinberg, G.; Frandsen, W.; Schuster, M.; Schlögl, R. Morphology and Microstructure of Li/MgO Catalysts for the Oxidative Coupling of Methane. *ChemCatChem* **2011**, *3* (6), 949–959. https://doi.org/10.1002/cctc.201000098.

(15) Richter, N. A.; Stavale, F.; Levchenko, S. V.; Nilius, N.; Freund, H.-J.; Scheffler, M. Defect Complexes in Li-Doped MgO. *Phys. Rev. B* **2015**, *91* (19), 195305. https://doi.org/10.1103/PhysRevB.91.195305.

(16) Kiatsaengthong, D.; Jaroenpanon, K.; Somchuea, P.; Chukeaw, T.; Chareonpanich, M.; Faungnawakij, K.; Sohn, H.; Rupprechter, G.; Seubsai, A. Effects of Mg, Ca, Sr, and Ba Dopants on the Performance of La_2O_3 Catalysts for the Oxidative Coupling of Methane.

ACS Omega **2022**, *7* (2), 1785–1793. https://doi.org/10.1021/acsomega.1c04738.

(17) Schucker, R. C.; J. Derrickson, K.; K. Ali, A.; J. Caton, N. The Effect of Strontium Content on the Activity and Selectivity of Sr-Doped La_2O_3 Catalysts in Oxidative Coupling of Methane. *Applied Catalysis A: General* **2020**, *607*, 117827. https://doi.org/10.1016/j.apcata.2020.117827.

(18) Kidamorn, P.; Tiyatha, W.; Chukeaw, T.; Niamnuy, C.; Chareonpanich, M.; Sohn, H.; Seubsai, A. Synthesis of Value-Added Chemicals via Oxidative Coupling of Methanes over $Na_2WO_4TiO_2$ –MnO_x/SiO_2 Catalysts with Alkali or Alkali Earth Oxide Additives. *ACS Omega* **2020**, *5* (23), 13612–13620. https://doi.org/10.1021/acsomega.0c00537.

(19) Farrell, B. L.; Igenegbai, V. O.; Linic, S. A Viewpoint on Direct Methane Conversion to Ethane and Ethylene Using Oxidative Coupling on Solid Catalysts. *ACS Catal.* **2016**, *6* (7), 4340–4346. https://doi.org/10.1021/acscatal.6b01087.

(20) Kiani, D.; Sourav, S.; Baltrusaitis, J.; Wachs, I. E. Oxidative Coupling of Methane (OCM) by SiO_2-Supported Tungsten Oxide Catalysts Promoted with Mn and Na. *ACS Catal.* **2019**, *9* (7), 5912–5928. https://doi.org/10.1021/acscatal.9b01585.

(21) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Comput Mater* **2017**, *3* (1), 54. https://doi.org/10.1038/s41524-017-0056-5.

(22) Nguyen, T. N.; Nhat, T. T. P.; Takimoto, K.; Thakur, A.; Nishimura, S.; Ohyama, J.; Miyazato, I.; Takahashi, L.; Fujima, J.; Takahashi, K.; Taniike, T. High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. *ACS Catal.* **2020**, *10* (2), 921–932. https://doi.org/10.1021/acscatal.9b04293.

(23) Schucker, R. C.; J. Derrickson, K.; K. Ali, A.; J. Caton, N. The Effect of Strontium Content on the Activity and Selectivity of Sr-Doped La_2O_3 Catalysts in Oxidative Coupling of Methane. *Applied Catalysis A: General* **2020**, *607*, 117827.

(24) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8* (8), 7403–7429. https://doi.org/10.1021/acscatal.8b01708.

(25) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nat Commun* **2016**, *7* (1), 11241. https://doi.org/10.1038/ncomms11241.

(26) Cao, B.; Adutwum, L. A.; Oliynyk, A. O.; Luber, E. J.; Olsen, B. C.; Mar, A.; Buriak, J. M. How To Optimize Materials and Devices *via* Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics. *ACS Nano* **2018**, *12* (8), 7434–7444. https://doi.org/10.1021/acsnano.8b04726.

(27) Javed, H. M. A.; Ahmad, M. I.; Que, W.; Qureshi, A. A.; Sarfaraz, M.; Hussain, S.; Iqbal, M. Z.; Zubair Nisar, M.; Shahid, M.; AlGarni, T. S. Encapsulation of TiO_2 Nanotubes with Cs Nanoparticles to Enhance Electron Injection and Thermal Stability of Perovskite Solar Cells. *Surfaces and Interfaces* **2021**, *23*, 101033. https://doi.org/10.1016/j.surfin.2021.101033.

(28) Brause, M.; Skordas, S.; Kempter, V. Study of the Electronic Structure of TiO_2(110) and Cs/TiO_2(110) with Metastable Impact Electron Spectroscopy and Ultraviolet Photoemission Spectroscopy (HeI). *Surface Science* **2000**, *445* (2–3), 224–234. https://doi.org/10.1016/s0039-6028(99)01052-3.

(29) Elkins, T. W.; Roberts, S. J.; Hagelin-Weaver, H. E. Effects of Alkali and Alkaline-Earth Metal Dopants on Magnesium Oxide Supported Rare-Earth Oxide Catalysts in the Oxidative Coupling of Methane. *Applied Catalysis A: General* **2016**, *528*, 175–190. https://doi.org/10.1016/j.apcata.2016.09.011.

(30) Ji, S. Surface WO_4 Tetrahedron: The Essence of the Oxidative Coupling of Methane over M_W_Mn/SiO_2 Catalysts. *Journal of Catalysis* **2003**, *220* (1), 47–56. https://doi.org/10.1016/S0021-9517(03)00248-3.

(31) Kiani, D.; Sourav, S.; Baltrusaitis, J.; Wachs, I. E. Elucidating the Effects of Mn Promotion on SiO_2 -Supported Na-Promoted Tungsten Oxide Catalysts for Oxidative Coupling of Methane (OCM). *ACS Catal.* **2021**, *11* (16), 10131–10137. https://doi.org/10.1021/acscatal.1c01392.

(32) van der Maaten, L.; Hinton, G. E. Visualizing Data Using t=SNE. *Journal of Machine Learning Research* **2008**, *9* (11), 2579–2605.

Chapter 4

Accelerating acquisition of design hypothesis via knowledge transfer in automatic feature engineering

Keywords: Catalyst informatics, machine learning, high-throughput experimentation, descriptor, oxidative coupling of methane

ABSTRACT

Experienced catalyst designers can identify high-performance catalysts by applying their knowledge of correlations between catalyst composition, structure, and performance—gained through prior development experience—and by leveraging insights into similarities between elements to extrapolate to unknown catalyst systems. This transfer of experience effectively shortens the knowledge-gathering process, embodying the essence of "intuition." Reproducing this intuitive process through data science is the ultimate goal of catalyst informatics.

In this chapter, we explored the potential for transferring design hypotheses from established catalyst supports to novel supports, specifically Y_2O_3. We evaluated the effectiveness of several approaches within active learning cycles: integrating feature values from other supports, applying dimensionality reduction techniques like principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), and using a physical filter to refine the data. These strategies were implemented to derive robust design hypotheses with limited data, demonstrating that design principles from known supports can guide predictions for new supports. The findings highlight that incorporating diverse knowledge sources can extend the predictive power of machine learning models, enhancing the adaptability and efficiency of catalyst design frameworks. This chapter thus contributes to a broader understanding of knowledge transfer and its role in developing data-driven, flexible approaches for catalyst discovery and optimization.

4.1. INTRODUCTION

In traditional catalyst design, manually evaluating a vast number of combinations requires considerable time and resources.[1,2] The trial-and-error approach makes identifying optimal catalyst designs challenging and often reduces efficiency. To address these challenges, data-driven catalysis research, or catalyst informatics, uses data science techniques like machine learning (ML) and visualization to accelerate catalyst development and uncover critical trends and patterns within catalyst data.[3] However, catalyst informatics is constrained by the scarcity of suitable data and the complexity of designing descriptors that accurately capture intricate structure–function relationships.[4]

To address data scarcity, the Taniike group has implemented high-throughput experimentation (HTE) to prepare and evaluate catalysts, generating datasets under consistent conditions suitable for machine learning across various catalytic processes.[5–7] Additionally, to overcome the challenges in descriptor design, automatic feature engineering (AFE) and active learning have been explored to streamline catalyst discovery.[8] Active learning is especially promising for its efficient use of limited datasets, enabling effective catalyst optimization while reducing the number of experimental trials. farthest point sampling (FPS) is used here to obtain catalytic data that more efficiently expresses the target catalytic space than random selection.[9–11] FPS adds catalysts that are least similar to those in the training data within the selected feature space, which aids in efficiently excluding models lacking global characteristics. While AFE has shown promise in enhancing the efficiency and sustainability of chemical processes, integrating new data over five to six cycles using FPS remains both time- and cost-intensive. This challenge partly stems from the difficulty in determining the nature of the correlations between the features and catalyst performance. Since AFE-generated features are typically derived from basic elemental properties, they may lack a direct link to observable phenomena in catalytic reactions. Even when features are closely related to catalytic reactions, their creation involves mathematical manipulation to form multi-dimensional features, making interpretation subjective and preventing them from serving as definitive evidence. Furthermore, AFE-generated features often diverge from established catalyst design principles, making it unclear how these features contribute specifically to reactions. A drawback of using features that are not human-guided is that it becomes difficult to understand the rationale behind the selection of each feature, leaving the model's progression as the main method for assessing feature reliability. Additionally, about five cycles of active learning are generally required to develop a robust design hypothesis. A faster approach to select effective catalysts from a large pool of candidates is essential for achieving optimal catalyst design. Adjustments in feature selection criteria and the integration of specific support features hold promising potential for improving catalyst performance. However, determining the most effective approaches remains challenging and underscores the need for comprehensive comparison and evaluation.

This chapter focuses on optimizing the catalyst design process through active learning by comparing the effectiveness of four approaches. One approach, used as a baseline, involves applying active learning without any particular adjustments. Another approach, the support-feature-based approach, incorporates key aspects of design hypotheses from other supports as features to create a new design hypothesis. The hypothesis here is that a robust design hypothesis can be constructed by drawing on key insights from well-established design hypotheses of other supports. In this study, values from eight features in the design hypotheses of other supports were reduced to two dimensions using t-distributed stochastic neighbor embedding (t-SNE) and principal component analysis (PCA), with predicted values analyzed to refine the hypotheses further.[12]

The third approach involves using a genetic algorithm (GA) to select feature combinations, where thresholds on scores and prediction performance metrics are set to identify feature combinations that can yield realistic predictions early in the process. The fourth approach combines both the second and third approaches to maximize effectiveness.

As data for the new support, Y_2O_3 was selected as the target catalyst. This choice was based on its proximity to the previously evaluated supports in the periodic table, as well as the high likelihood of it exhibiting new characteristics. These characteristics are expected to allow effective utilization of the design hypotheses developed for other supports. A total of 140 catalyst candidates were randomly selected and divided into seven groups. One group was used as the original data, five groups were used for conducting five cycles of active learning, and the final group was reserved for testing. Using different approaches, hypothesis accuracy was evaluated in each cycle. These approaches aim to derive the most robust hypothesis with minimal data. By comparing test scores and results from active learning simulations, the most effective approach for establishing robust hypotheses with the smallest dataset can be identified.

4.2. METHOD

In active learning, data is added not only to validate the primary hypothesis but also to eliminate alternative hypotheses that could explain the original data and to expand the exploration space.[13] When training data is limited, AFE may generate multiple design hypotheses with similar scores, considered as alternative hypotheses. While a larger dataset increases the likelihood of obtaining a global design hypothesis, getting sufficient data can be challenging. In this study, we proposed a more efficient active learning process by reusing design hypotheses from other supports and restricting the combinations of features selected by the GA. To evaluate the performance of these modified approaches, I used a dataset of 140 randomly selected catalysts. While active learning is necessary to eliminate inaccurate or less robust hypotheses, we examined whether the enhanced GA could derive a global design hypothesis even with a smaller dataset. The 140 randomly selected catalysts were divided into seven parts of 20 each: one part served as the original data, five parts were used to vary data volume as additional catalysts, and the remaining part was used as test data.

4.2.1. Automatic feature engineering

AFE automates the design of physically meaningful features for a given catalyst dataset within the framework of supervised ML.[8] This process involves a structured pipeline of feature assignment, synthesis, and selection. First, physical quantities of elements are assigned to the catalysts, with their elemental compositions represented through commutative operations. Higher-order features involving nonlinear and combinatorial effects are then synthesized from these assigned primitive features using mathematical operations. Finally, a specified number of features (descriptors) that optimize the score of the supervised ML are selected from the large synthesized array of features.

In this study, we utilized 58 parameters of elements from XenonPy, normalized according to the literature.[14] These parameters were assigned to each catalyst using five commutative operations (maximum, minimum, average, product, and standard deviation), yielding 290 primary features. These primary features were further synthesized into 3,480 features using 12 functional forms (x, sqrt(x), x^2, x^3, ln(x), exp(x), and their reciprocals, where x represents each primary feature). A genetic algorithm-based approach was employed to select eight features that minimized the mean absolute error (MAE) in leave-one-out cross-validation (LOOCV) using Huber regression.[15] This process involved assessing approximately 4,000,000 models per dataset with various feature combinations and selecting the combination of features (X) and model (f(X)) that yielded the lowest cross-validation (CV) score as the most plausible design hypothesis. Huber regression, which is a form of multiple linear regression, was employed to prevent overfitting owing to its reduced number of parameters and ensure robustness against outliers, such as experimental failures. The number of selected features was empirically determined to balance the CV score and the cost of feature selection. Further details on AFE and parameter selection are provided in previous study.

4.2.2. High-throughput experimentation

The catalysts proposed in Section 2.3 were prepared and evaluated using the same experimental methods and conditions, which are briefly described as follows.

4.2.2.1. Materials

The metal precursors used were $LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, KNO_3, $Ca(NO_3)_2 \cdot 4H_2O$, $Ti(OiPr)_4$, $VOSO_4 \cdot xH_2O$ ($x = 3$–5), $Mn(NO_3)_2 \cdot 6H_2O$, $Fe(NO_3)_3 \cdot 9H_2O$, $Co(NO_3)_2 \cdot 6H_2O$, $Ni(NO_3)_2 \cdot 6H_2O$, $Cu(NO_3)_2 \cdot 3H_2O$, $Zn(NO_3)_2 \cdot 6H_2O$, $Sr(NO_3)_2$, $Y(NO_3)_3 \cdot 6H_2O$, $ZrO(NO_3)_2 \cdot xH_2O$ ($x = 2$), $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, $Pd(OAc)_2$, $CsNO_3$, $Ba(NO_3)_2$, $La(NO_3)_3 \cdot 6H_2O$, $Ce(NO_3)_3 \cdot 6H_2O$, $Nd(NO_3)_3 \cdot 6H_2O$, $Eu(NO_3)_3 \cdot 5H_2O$, $Tb(NO_3)_3 \cdot 5H_2O$, $Hf(OEt)_4$, and $(NH_4)_{10}H_2(W_2O_7)_6$. These materials were purchased from one of the following suppliers: Sigma-Aldrich, Kanto Chemical, Wako Pure Chemical Industries, Alfa-Aesar, or Sumitomo Chemical.

To investigate new support materials, I selected Y_2O_3 as the subject of this research. Y_2O_3 has high thermal stability, making it suitable for reactions under high-temperature conditions in OCM. Due to its resistance to decomposition and activity degradation, it is expected to have a long lifespan as a catalyst.[16] Y_2O_3 (5.01 g/mL, Sigma-Aldrich) was used as the reagent for Y_2O_3 in this study.

4.2.2.2. Preparation of catalysts

The catalysts were prepared using a parallelized wet impregnation method, in which support powder was impregnated with an aqueous solution of specified metal precursors at 50 °C for 6 h. The loading amount of elements was fixed at 0.37 mmol/g-support per selection within M1–3. Following impregnation, the powder was vacuum-dried and calcined in air at 1000 °C for 3 h. For precursors that contained metal alkoxides, impregnation was performed in two steps: first, with an aqueous solution of the other precursors, followed by impregnation with an ethanol solution of the metal alkoxides.

4.2.2.3. Evaluation of catalysts

The catalysts were evaluated for OCM using a custom-built HTE system. This system operates through a combination of a gas mixer, flow distributor, electric furnace-bearing reactors, autosampler, and quadrupole mass spectrometer (QMS), enabling the performance of 20 catalysts to be automatically assessed under a programmed sequence of reaction conditions. The catalysts were secured in beds 1 cm in height within reaction tubes made of fused quartz, which had inner diameters of 4 and 2 mm, using quartz wool.

After inline activation at 1000 °C for 3 h in an oxygen stream, the 20 catalysts were tested under 135 conditions varying in temperature (700, 750, 800, 850, and 900 °C), total gas flow rates (10, 15, and 20 mL/min/channel), CH_4/O_2 ratios (2, 4, and 6 mol/mol), and Ar partial pressures as a balancing gas (0.15, 0.40, and 0.70 atm). Each catalyst was labeled according to the highest C_2 yield achieved among all 135 conditions.

4.2.3. Improvements to the catalyst search process

We tested whether adding feature values derived from predictions on other supports to the original dataset and performing feature selection could lead to earlier identification of robust design hypotheses compared to normal approaches. The additional features from other supports included predicted C_2 yield, as well as PCA and t-SNE results based on the five final design hypothesis equations from BaO, CaO, La_2O_3, TiO_2, and ZrO_2 families. PCA and t-SNE were used to reduce the dimensionality of each support's final hypothesis equations, allowing key information to be retained and incorporated into the feature set.

PCA achieves dimensionality reduction by creating a new set of orthogonal axes that maximize variance within the original dataset, projecting the data onto these axes to reduce its dimensions. It defines new variables, or principal components, in order of decreasing variance, capturing as much of the original data's variability as possible with the fewest components. The second technique, t-distributed stochastic neighbor embedding (t-SNE), focuses on preserving the local relationships between elemental combinations in the original space to capture their similarities effectively.46

In this study, high-dimensional data from each support’s final hypothesis equation was transformed into lower dimensions using PCA and t-SNE, retaining essential features and adding them as new features to the dataset. The features derived from design hypotheses across different catalyst families are referred to as “design hypothesis features” to distinguish them from the standard features derived from the XenonPy elemental physical properties.

As an approach to restrict the models being selected, two conditions were applied after GA selected eight features. The first condition limited the score value, with the MAE (mean absolute error) required to be 10% or less or no more than 1.5 times the highest elite score. The second condition excluded feature combinations that produced predicted C_2 yield values outside the acceptable range (-20% to 50%) for the 4,059 catalysts during the GA process, with MAE used as the scoring metric. By adding these conditions, we aim to narrow down effective features even with limited data, avoid local overfitting, and derive design hypotheses suitable for catalyst performance prediction. This approach is expected to generate design hypotheses applicable to catalyst development from the outset.

4.3. RESULTS AND DISCUSSION

4.3.1. Outcomes of feature engineering for Y_2O_3 support

Figure 4.1 is a graph showing the CH_4 conversion and C_2 selectivity of the catalysts used in this study, and you can see the performance of each catalyst along with its name. The Y_2O_3 support demonstrates strong CH_4 conversion performance. The figure also indicates that a composition capable of enhancing selectivity is necessary.

It is anticipated from previous reports that repeating the active learning loop five times would exhibit behavior suggesting convergence of the design hypothesis. I derived design hypothesis for the Y_2O_3 support by randomly selecting catalysts across five cycles without adding feature values or applying physical filter. I compiled the MAE during cross-validation (MAE_{CV}), selected features, and MAE during training (MAE_{train}) for Y_2O_3-supported catalysts across these active learning cycles, as shown in Table 4.1. The MAE_{CV} values were closely aligned with MAE_{train} across all cycles, indicating minimal overfitting. Both metrics remained within a narrow range of 0.4–1.8%, comparable to the experimental error (1.5–2% for C_2 yields).

When examining the features selected in each cycle, we found that while some features, such as vdw_radius_std, reappeared in different combinations, no single feature was consistently selected across all cycles. Since the order of features follows their importance, it is clear that vdw_radius_std does not dominate the design hypothesis. This suggests that, at this stage, no dominant feature drives robust design hypotheses for Y_2O_3; instead, effective design relies on a balanced mix of various feature combinations. The diversity of features selected across cycles also indicates that a single, robust feature set has not yet emerged, showing ongoing competition among different feature combinations.

Table 4.1 presents these observations by tracking the development of design hypotheses for Y_2O_3-supported catalysts through each active learning cycle. The consistent alignment of MAE_{CV} and MAE_{train} values across cycles suggests that overfitting is effectively managed. However, the variation in MAE_{test} highlights gaps in the coverage of the catalyst search space, likely due to the random selection of catalysts.

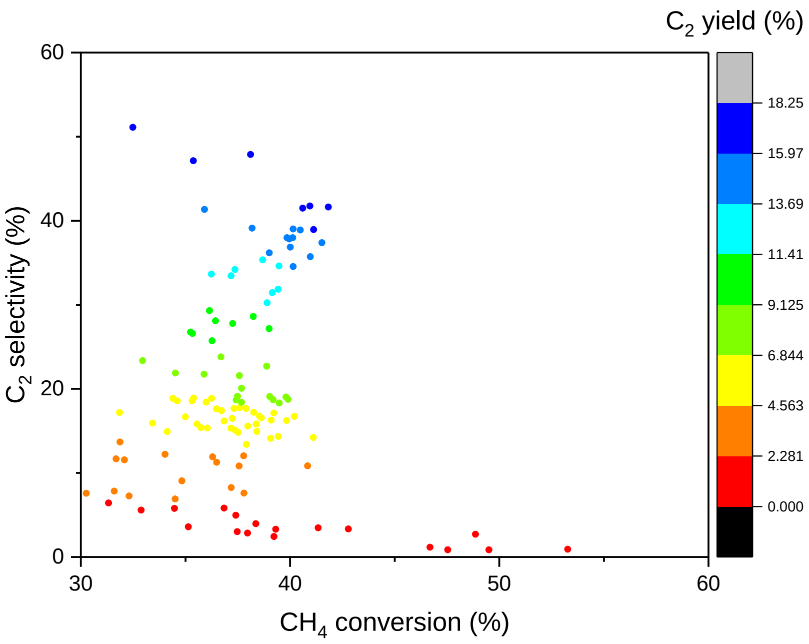


Figure 4.1. Scatter plot of 140 catalysts experimentally tested in previous and current studies. The x-axis represents CH_4 conversion (%), the y-axis represents predicted C_2 selectivity (%), and the color of the dot indicates the C_2 yield (%).

Table 4.1. Development of design hypotheses for Y_2O_3-based catalysts throughout the active learning cycles. This table shows the scores and selected features of the design hypotheses obtained in each individual active learning cycle.

Cycle	MAE_{CV}^{a}	Selected features^b
0	0.65%	1. (*gs_mag_moment_max*)^3 2. 1/(*dipole_polarizability_ave*) 3. 1/ln(*atomic_radius_rahm_ave*) 4. 1/ln(*gs_est_bcc_latcnt_ave*) 5. 1/ln(*mendeleev_number_min*) 6. 1/(*Polarizability_min*)$^{1/2}$ 7. exp(*thermal_conductivity_ave*) 8. ln(*hhi_r_max*)
1	0.83%	1. (*atomic_radius_rahm_min*)^3 2. 1/exp(*num_d_valence_max*)

		3. 1/ln(*en_pauling_pro*) 4. 1/ln(*first_ion_en_pro*) 5. 1/ln(*thermal_conductivity_pro*) 6. 1/(*hhi_r_ave*)$^{1/2}$ 7. exp(*num_f_valence_min*) 8. (*gs_mag_moment_ave*)$^{1/2}$
2	1.48%	1. (*electron_negativity_min*)^{2} 2. (*hhi_r_std*)^{2} 3. 1/exp(*num_valance_pro*) 4. *heat_capacity_molar_min* 5. (*atomic_radius_rahm_min*)$^{1/2}$ 6. (*bulk_modulus_max*)$^{1/2}$ 7. (*gs_energy_pro*)$^{1/2}$ 8. (*thermal_conductivity_pro*)$^{1/2}$
3	1.62%	1. 1/exp(*atomic_radius_rahm_min*) 2. 1/ln(*heat_capacity_molar_min*) 3. 1/ln(*mendeleev_number_pro*) 4. (*boiling_point_pro*)$^{1/2}$ 5. (*bulk_modulus_max*)$^{1/2}$ 6. (*gs_energy_pro*)$^{1/2}$ 7. (*thermal_conductivity_min*)$^{1/2}$ 8. (*thermal_conductivity_pro*)$^{1/2}$
4	1.82%	1. (*boiling_point_ave*)^{3} 2. (*num_s_unfilled_ave*)^{3} 3. (*thermal_conductivity_max*)^{2} 4. 1/(*atomic_radius_min*) 5. 1/ (*atomic_volume_min*)$^{1/2}$ 6. 1/ (*gs_volume_per_min*)$^{1/2}$ 7. ln(*melting_point_ave*) 8. (*Polarizability_min*)$^{1/2}$
5	2.12%	1. 1/(*dipole_polarizability_max*) 2. 1/(*lattice_constant_std*)$^{1/2}$ 3. exp(*atomic_volume_max*) 4. exp(*num_d_valence_ave*) 5. ln(*Polarizability_min*) 6. ln(*num_f_unfilled_ave*) 7. ln(*vdw_radius_min*) 8. ln(*vdw_radius_mm3_pro*)

[a] The MAE of C_2 yields during CV is shown in parentheses.

[b] Eight features were selected to minimize the MAE_{CV} value and are listed in order of permutation feature importance. The features are described in the same manner as XenonPy, and their meanings can be found in the Table S1.

4.3.1.1. Other support features

Next, feature selection was performed by adding feature values from other supports (Table 4.2). Compared to the normal approach described above, adding feature values from other supports resulted in a decrease of about 0.1 in the final MAE_{CV} value.

Table 4.2. Development of design hypotheses for Y_2O_3-based catalysts throughout the active learning cycles. Additional feature values from other supports have been added.

Cycle	MAE_{CV}[a]	Selected features[b]
0	0.37%	1. 1/(*first_ion_en_max*) 2. 1/exp(*en_allen_min*) 3. 1/(BaO_PCA_1)$^{1/2}$ 4. 1/(*num_s_valence_ave*)$^{1/2}$ 5. *hhi_p_min* 6. *hhi_r_max* 7. ln(*c6_gb_ave*) 8. (*thermal_conductivity_ave*)$^{1/2}$
1	0.86%	1. ($La_2O_3\_tSNE_2$)^3 2. ($TiO_2\_C_2y$)^2 3. 1/(*hhi_r_ave*) 4. 1/ln(*gs_mag_moment_std*) 5. 1/ln(*sound_velocity_max*) 6. *dipole_polarizability_std* 7. exp(*hhi_p_min*) 8. (*Polarizability_pro*)$^{1/2}$
2	1.47%	1. (CaO_C_2y)^3 2. (*thermal_conductivity_ave*)^3 3. 1/ln(*heat_of_formation_std*) 4. exp(*density_ave*) 5. (BaO_tSNE_2)$^{1/2}$ 6. (*bulk_modulus_max*)$^{1/2}$ 7. (*gs_energy_pro*)$^{1/2}$ 8. (*vdw_radius_pro*)$^{1/2}$
3	1.59%	1. (CaO_C_2y)^3 2. (*heat_of_formation_min*)^3 3. (*thermal_conductivity_ave*)^3 4. 1/exp(BaO_tSNE_2) 5. 1/exp(*bulk_modulus_max*) 6. exp($ZrO_2\_PCA_1$) 7. exp(*num_f_valence_ave*) 8. (*gs_energy_pro*)$^{1/2}$

4	1.83%	1. (*thermal_conductivity_ave*)^2 2. 1/(*covalent_radius_cordero_ave*) 3. 1/exp(*bulk_modulus_max*) 4. 1/exp(*gs_energy_pro*) 5. 1/ln(*electron_negativity_std*) 6. *CaO_C*$_2$*y* 7. ln(*covalent_radius_pyykko_triple_ave*) 8. (*Polarizability_min*)$^{1/2}$
5	2.17%	1. (*BaO_tSNE*$_2$)^3 2. 1/ln(*CaO_PCA*$_2$) 3. 1/ln(*thermal_conductivity_ave*) 4. exp(*fusion_enthalpy_pro*) 5. ln(*ZrO*$_2$*_PCA*$_2$) 6. *num_f_valence_std* 7. (*ZrO*$_2$*_C*$_2$*y*)$^{1/2}$ 8. (*vdw_radius_max*)$^{1/2}$

[a] The MAE of C_2 yields during CV is shown in parentheses.

[b] Eight features were selected to minimize the MAE_{CV} value and are listed in order of permutation feature importance. The features are described in the same manner as XenonPy, and their meanings can be found in the Table S1.

4.3.1.2. Introducing filter for model selection

Furthermore, feature selection was performed by using GA to select only feature combinations with a certain level of accuracy (Table 4.3). Compared to the normal approach in Table 4.1, The MAE_{CV} value for 0 cycle has decreased. Since only feature combinations that yield realistic predicted values for all 4059 M1-M2-M3 combinations are selected, the MAE_{CV} value of the final design hypothesis is relatively high.

Table 4.3. Development of design hypotheses for Y_2O_3-based catalysts throughout the active learning cycles. Additional feature values from filter added.

Cycle	MAE_{CV}[a]	Selected features[b]
0	1.15%	1. (*Polarizability_min*)^2 2. (*boiling_point_max*)^2 3. 1/exp(*electron_negativity_max*) 4. 1/exp(*hhi_p_min*)

		5. *bulk_modulus_std* 6. *first_ion_en_ave* 7. (*first_ion_en_ave*)$^{1/2}$ 8. (*icsd_volume_ave*)$^{1/2}$
1	1.08%	1. (*hhi_p_min*)^{2} 2. (*mendeleev_number_max*)^{3} 3. (*vdw_radius_ave*)^{3} 4. exp(*boiling_point_max*) 5. exp(*gs_energy_min*) 6. ln(*Polarizability_min*) 7. (*covalent_radius_pyykko_double_std*)$^{1/2}$ 8. (*first_ion_en_pro*) $^{1/2}$
2	1.63%	1. (*first_ion_en_max*)^{3} 2. (*heat_of_formation_std*)^{3} 3. (*num_d_valence_max*)^{2} 4. (*thermal_conductivity_ave*)^{2} 5. ln(*heat_capacity_molar_min*) 6. (*bulk_modulus_max*)$^{1/2}$ 7. (*covalent_radius_slater_std*)$^{1/2}$ 8. (*thermal_conductivity_ave*)$^{1/2}$
3	1.61%	1. (*num_f_valence_ave*)^{3} 2. 1/exp(*bulk_modulus_max*) 3. ln(*atomic_radius_max*) 4. (*atomic_radius_rahm_min*)$^{1/2}$ 5. (*gs_energy_pro*)$^{1/2}$ 6. (*heat_capacity_molar_min*)$^{1/2}$ 7. (*num_d_unfilled_ave*)$^{1/2}$ 8. (*thermal_conductivity_pro*)$^{1/2}$
4	1.87%	1. (*melting_point_std*)^{2} 2. (*num_f_valence_ave*)^{3} 3. 1/exp(*Polarizability_min*) 4. 1/exp(*bulk_modulus_max*) 5. 1/exp(*covalent_radius_cordero_ave*) 6. *Polarizability_pro* 7. *en_allen_ave* 8. ln(*heat_capacity_molar_min*)
5	2.12%	1. (*num_f_valence_std*)^{3} 2. 1/exp(*num_valance_std*) 3. 1/ (*covalent_radius_pyykko_double_min*)$^{1/2}$ 4. 1/ (*gs_est_fcc_latcnt_min*)$^{1/2}$ 5. *en_allen_ave* 6. (*Polarizability_min*)$^{1/2}$ 7. (*covalent_radius_pyykko_double_std*)$^{1/2}$ 8. (*heat_capacity_molar_ave*)$^{1/2}$

[a] The MAE of C_2 yields during CV is shown in parentheses.

[b] Eight features were selected to minimize the MAE$_{CV}$ value and are listed in order of permutation feature importance. The features are described in the same manner as XenonPy, and their meanings can be found in the Table S1.

4.3.1.3. Other support features and filter

Next, feature selection was performed using the filter approach after adding features derived from other supports (Table 4.4). Compared to other approaches, the MAE$_{CV}$ value at 0 cycles was the highest, suggesting that a more robust model may have been obtained from the beginning.

Table 4.4. Development of design hypotheses for Y_2O_3-based catalysts throughout the active learning cycles. Other supports feature approach and filter approach have been added.

Cycle	MAE$_{CV}$[a]	Selected features[b]
0	0.85%	1. (*hhi_p_ave*)^3 2. (*hhi_r_max*)^3 3. 1/exp(*icsd_volume_ave*) 4. 1/ln(*gs_mag_moment_std*) 5. *covalent_radius_cordero_min* 6. exp(*atomic_radius_rahm_ave*) 7. ln(*fusion_enthalpy_min*) 8. (*first_ion_en_pro*)$^{1/2}$
1	1.13%	1. (*gs_energy_min*)^2 2. 1/ln(*hhi_r_ave*) 3. 1/(*Polarizability_min*)$^{1/2}$ 4. exp(*hhi_p_pro*) 5. *hhi_r_pro* 6. (*boiling_point_min*)$^{1/2}$ 7. (*covalent_radius_pyykko_double_max*)$^{1/2}$ 8. (*gs_mag_moment_max*)$^{1/2}$
2	1.53%	1. (*CaO_C*$_2$*y*)^2 2. (*thermal_conductivity_ave*)^3 3. 1/exp(*BaO_tSNE*$_2$) 4. 1/exp(*covalent_radius_cordero_std*) 5. 1/(*thermal_conductivity_ave*)$^{1/2}$ 6. ln(*bulk_modulus_max*) 7. ln(*covalent_radius_pyykko_double_std*) 8. ln(*melting_point_max*)
3	1.63%	1. (*CaO_C*$_2$*y*)^2 2. (*thermal_conductivity_max*)^3 3. 1/exp(*thermal_conductivity_pro*) 4. exp(*BaO_tSNE*$_2$)

		5. ln(*c6_gb_std*) 6. ln(*molar_volume_std*) 7. (*bulk_modulus_max*)$^{1/2}$ 8. (*gs_energy_pro*)$^{1/2}$
4	1.81% (%)	1. $(ZrO_2\_PCA_2)^3$ 2. 1/(*gs_volume_per_min*) 3. 1/exp(*Polarizability_min*) 4. 1/exp(*en_allen_ave*) 5. exp(*dipole_polarizability_min*) 6. ln(*c6_gb_min*) 7. ln(*vdw_radius_mm3_pro*) 8. (*atomic_number_max*)$^{1/2}$

[a] The MAE of C_2 yields during CV is shown in parentheses.

[b] Eight features were selected to minimize the MAE$_{CV}$ value and are listed in order of permutation feature importance. The features are described in the same manner as XenonPy, and their meanings can be found in the Table S1.

4.3.2. Verification of the effectiveness of each approach

To evaluate the effectiveness of each approach, the predictive behavior of design hypotheses through active learning cycle was analyzed. First, the 120 pre-evaluated catalysts were randomly split for training, while 20 catalysts were randomly selected for testing. By comparing the predicted values before and after each cycle, the stability of predictive accuracy in these approaches can be evaluated. Figure 4.2 displays the cycle number on the horizontal axis and the MAE value on the vertical axis. The MAE was calculated by comparing the predicted C_2 yields of 4,059 catalysts between the design hypothesis obtained in one cycle and that from the previous cycle. Lower MAE values indicate greater similarity between the predictive behaviors of the two design hypotheses. In the early stages of active learning, MAE values decreased significantly, indicating the elimination of less robust design hypotheses as more data were incorporated. In the later stages, additional data led to minor changes in MAE, suggesting that the design hypotheses became similarly robust. Focusing on the differences between the approaches, it is evident that using a filter results in significantly greater stability compared to either of the two approaches that do not use a filter. First, at cycle 0, the MAE value is 1% lower when using a filter compared to the non-filter approaches. As the number of cycles increases, this gap remains. Additionally, considering that the experimental error is 2%, it can be observed that when using a filter, stability improves to approximately the level of experimental error by cycle 2. On the other hand, when a filter is not used, the difference in predicted values does not fall below the experimental error threshold.

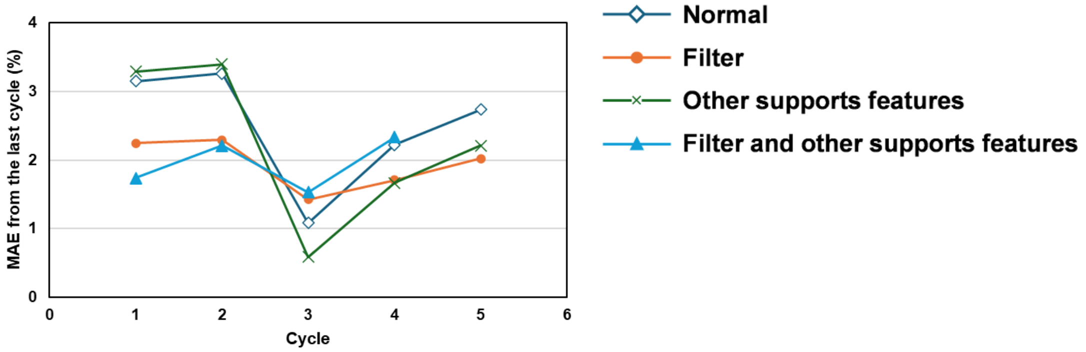


Figure 4.2. The evolution of predictive behavior in design hypotheses through active learning is observed by calculating MAE values between the predicted C_2 yields of 4,059 catalysts across successive cycles. Lower MAE values indicate higher consistency in predictive behavior. In the initial stages, MAE values decreased as weaker hypotheses were eliminated with added data, while later cycles showed minimal MAE changes, indicating that the hypotheses had stabilized in robustness. It has also been shown that the accuracy of predictions can be significantly stabilized by filter.

Next, Figure 4.3 presents the prediction accuracy for the test data, consists of 20 catalysts, across different approaches and cycles. This figure illustrates how closely the predicted C_2 yield with the actual value, as represented by the MAE on the vertical axis. At cycle 0, the approach using other supports features performs worse than the normal approach. This suggests that adding additional features leads to a higher likelihood of obtaining local solutions that fit only specific catalysts, especially when training data is limited. On the other hand, the two approaches using the filter performed slightly better than the normal approach. From cycle 1, the approach using other supports features also outperformed the normal approach, achieving an accuracy level comparable to the filter-based approaches. In addition, the two approaches utilizing the filter and other support features clearly have better initial predictions. This shows that combining a method that increases the number of features and a method that filters features is useful for feature selection.

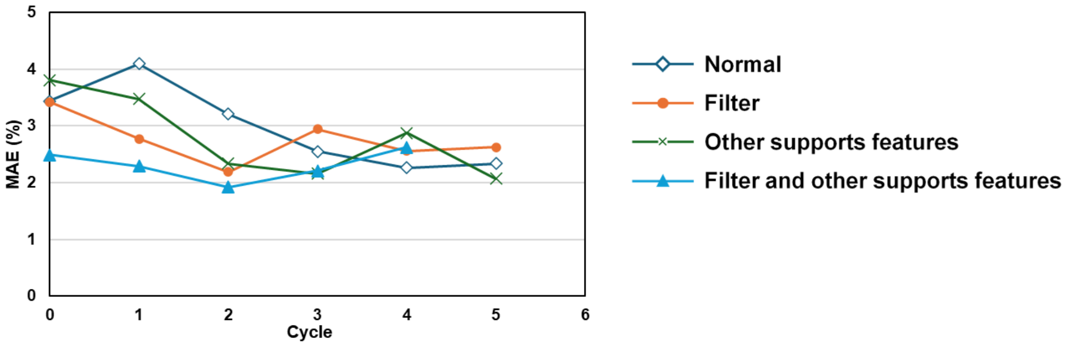


Figure 4.3. This graph shows the predictive performance of the test data. The vertical axis represents the MAE values for the observed and predicted yields of the 20 catalysts in the test data, while the horizontal axis indicates the number of cycles for each design hypothesis. Compared to the normal approach, other approaches show effectiveness from the cycle 1. Furthermore, the approach that combines both the filter and other support features proves to be more effective.

4.4. CONCLUSION

In this chapter, different approaches that incorporate physical filter and other support features were evaluated through active learning to determine the most effective approach. First, the transferability of design hypotheses developed for known supports (BaO, CaO, La_2O_3, TiO_2, ZrO_2) to a new support (Y_2O_3) was introduced. Additionally, applying a physical filter allowed for the generation of a generalized design hypothesis that remains applicable even with limited data. The filter approach consistently had the effect of improving the stability of predictions. However, the combined approach proved to be more effective than either approach alone. Since one approach increases the number of features while the other reduces the number of models derived from these features, their combination can be considered highly effective. Utilizing these approaches together successfully reduced the amount of data required for active learning.

REFERENCE

(1) Martín, A. J.; Mitchell, S.; Mondelli, C.; Jaydev, S.; Pérez-Ramírez, J. Unifying Views on Catalyst Deactivation. *Nat Catal* **2022**, *5* (10), 854–866. https://doi.org/10.1038/s41929-022-00842-y.

(2) Heard, A. W.; Suárez, J. M.; Goldup, S. M. Controlling Catalyst Activity, Chemoselectivity and Stereoselectivity with the Mechanical Bond. *Nat Rev Chem* **2022**, *6* (3), 182–196. https://doi.org/10.1038/s41570-021-00348-4.

(3) Benavides-Hernández, J.; Dumeignil, F. From Characterization to Discovery: Artificial Intelligence, Machine Learning and High-Throughput Experiments for Heterogeneous Catalyst Design. *ACS Catal.* **2024**, *14* (15), 11749–11779. https://doi.org/10.1021/acscatal.3c06293.

(4) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Comput Mater* **2017**, *3* (1), 54. https://doi.org/10.1038/s41524-017-0056-5.

(5) Nguyen, T. N.; Nhat, T. T. P.; Takimoto, K.; Thakur, A.; Nishimura, S.; Ohyama, J.; Miyazato, I.; Takahashi, L.; Fujima, J.; Takahashi, K.; Taniike, T. High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. *ACS Catal.* **2020**, *10* (2), 921–932. https://doi.org/10.1021/acscatal.9b04293.

(6) Nakanowatari, S.; Nguyen, T. N.; Chikuma, H.; Fujiwara, A.; Seenivasan, K.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Extraction of Catalyst Design Heuristics from Random Catalyst Dataset and Their Utilization in Catalyst Development for Oxidative Coupling of Methane. *ChemCatChem* **2021**, *13* (14), 3262–3269. https://doi.org/10.1002/cctc.202100460.

(7) Nguyen, T. N.; Nakanowatari, S.; Nhat Tran, T. P.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane. *ACS Catal.* **2021**, *11* (3), 1797–1809. https://doi.org/10.1021/acscatal.0c04629.

(8) Taniike, T.; Fujiwara, A.; Nakanowatari, S.; García-Escobar, F.; Takahashi, K. Automatic Feature Engineering for Catalyst Design Using Small Data without Prior Knowledge of Target Catalysis. *Commun Chem* **2024**, *7* (1), 11. https://doi.org/10.1038/s42004-023-01086-y.

(9) Qi, C. R.; Yi, Li; Su, Hao; Guibas, J. Leonidas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.

(10) Shi, S.; Wang, X.; Li, H. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. arXiv 2018. https://doi.org/10.48550/ARXIV.1812.04244.

(11) Guo, M.-H.; Cai, J.-X.; Liu, Z.-N.; Mu, T.-J.; Martin, R. R.; Hu, S.-M. PCT: Point Cloud Transformer. **2020**. https://doi.org/10.48550/ARXIV.2012.09688.

(12) van der Maaten, L.; Hinton, G. E. Visualizing Data Using t=SNE. *Journal of Machine Learning Research* **2008**, *9* (11), 2579–2605.

(13) Prince, M. Does Active Learning Work? A Review of the Research. *J of Engineering Edu* **2004**, *93* (3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x.

(14) Yoshida, R. XenonPy Is a Python Software for Materials Informatics. **2018**.

(15) Huber, P. J. Robust Estimation of a Location Parameter. *Ann. Math. Statist.* **1964**, *35* (1), 73–101. https://doi.org/10.1214/aoms/1177703732.

(16) Feng, J.; Liu, L.; Ju, X.; Wang, J.; Zhang, X.; He, T.; Chen, P. Highly Dispersed Ruthenium Nanoparticles on $Y_2 O_3$ as Superior Catalyst for Ammonia Decomposition. *ChemCatChem* **2021**, *13* (6), 1552–1558. https://doi.org/10.1002/cctc.202001930.

Chapter 5

Catalyst development for oxidative coupling of methane

KEYWORDS: Catalyst informatics, high-throughput experimentation, descriptor, oxidative coupling of methane

ABSTRACT

Solid catalyst development has traditionally relied on trial-and-error approaches, limiting the broader application of valuable insights across different catalyst families. To overcome this fragmentation, we introduce a framework that integrates high-throughput experimentation (HTE) and automatic feature engineering (AFE) with active learning to acquire comprehensive catalyst knowledge. The framework is demonstrated for oxidative coupling of methane (OCM), where active learning is continued until the machine learning model achieves robustness for each of the BaO-, CaO-, La_2O_3-, TiO_2-, Y_2O_3, and ZrO_2-supported catalysts. The catalysts obtained were recommended to maximize the exploration range of OCM catalysts, providing a broad perspective on the synergistic combinations of these supports. This chapter includes novel catalysts, enabling the evaluation of the performance of a diverse range of compositional combinations.

5.1. INTRODUCTION

Oxidative coupling of methane (OCM) is one of the most challenging catalytic reactions to develop. This process converts methane (CH_4) directly into ethylene (C_2H_4) in a single step.[1] As the primary component of natural gas and biogas, CH_4 has garnered increasing attention amid the global expansion of natural gas exploration and extraction efforts.[2,3] Consequently, interest in methane utilization via the OCM has grown.[4] C_2H_4, the product of this reaction, serves as a essential feedstock for a wide range of petrochemical products. Currently, it is primarily produced from naphtha, a substance derived from petroleum. The industrial method for obtaining C_2H_4 from natural gas involves two steps.[5,6] The first step involves converting natural gas into synthesis gas through reforming reactions. The second step is the conversion of synthesis gas into synthetic fuel through Fischer-Tropsch synthesis (FTS).[7] However, due to the high cost of the FTS method, there is growing interest in the methane OCM reaction.[8] One of the key challenges in the OCM reaction is the chemical inertness of CH_4 compared to its products, making high yields difficult to achieve. As a result, the maximum yield for OCM is generally considered to be around 35%.[9]

The reason why the development of solid catalysts is not easy is that the details of individual potential reactions are not clear,[10,11] the roles of the constituent elements are often unclear,[12] and the structure of the substance is indefinite,[13] and on top of that, it is a typically complex substance with a large number of constituent components.[14–16] It goes without saying that we are trying to find new and better ways to make catalysts. However, society's demand for catalysts is strong.[17–19] Researchers want methods that go beyond the traditional framework and give them a new perspective on catalyst development. In this context, catalytic informatics is attracting a lot of attention.[20–22]

Data-driven catalysis research, also known as catalyst informatics, leverages data science techniques such as machine learning (ML) and visualization to accelerate the development and understanding of catalysts by revealing applicable trends and patterns hidden within catalyst data. However, catalyst informatics is constrained by the scarcity of catalyst data suitable for data science and the difficulty in handcrafting descriptors that capture the essence of intricate structure–function relationships.[23,24] I used high-throughput experimentation (HTE) to prepare and evaluate numerous solid catalysts, generating sized, qualified, and consistent datasets for various heterogeneous catalytic processes, including OCM. Furthermore, we recently introduced an automatic feature engineering (AFE) technique that programmatically designs descriptors to capture the essence of the target catalysis , beginning with the general physical properties of the elements, such as atomic radii and electronegativity. AFE generates predictive ML models with tailored descriptors without requiring researchers to make assumptions or hypotheses about the target system.

In this study, the same active learning approach integrated with HTE and AFE was applied to six OCM catalyst families. We began with previously acquired OCM catalyst data for six supports (BaO, CaO, La_2O_3, TiO_2, Y_2O_3, ZrO_2) were evaluated for OCM. CaO and La_2O_3 were selected because they are the most extensively studied basic oxides in OCM and are likely to exhibit distinct catalyst designs. BaO was selected as the reference for these supports because it offers higher C_2 yields and selectivities. In contrast, the redox-active supports TiO_2 and ZrO_2 were included to investigate potential relationships between the general physical properties and design hypotheses.

5.2. METHOD

This study used previously acquired data to establish a design hypothesis for each of the six catalyst families corresponding to different support materials. However, similar to how researchers cannot dismiss alternative hypotheses when evidence is limited, AFE cannot overlook alternative design hypotheses (ML models with differently tailored descriptors that fit the training data) when the diversity of catalysts in the training data is restricted. Consequently, we implemented an active learning strategy (Figure 1). This strategy employs farthest point sampling (FPS) within the descriptor space established by AFE to propose catalysts that are maximally dissimilar to those included in the training data. These catalysts served as rigorous control experiments to validate the proposed design hypothesis. The performances of the proposed catalysts were assessed using HTE to reinforce the training data and update the design hypothesis via AFE. This iterative process aims to eliminate design hypotheses that do not generalize well across catalysts, resulting in a robust and experimentally validated design hypothesis. Further details are provided below.

The catalyst used in this study was selected either by random sampling or through farthest point sampling (FPS) based on feature combinations obtained from automatic feature engineering (AFE).

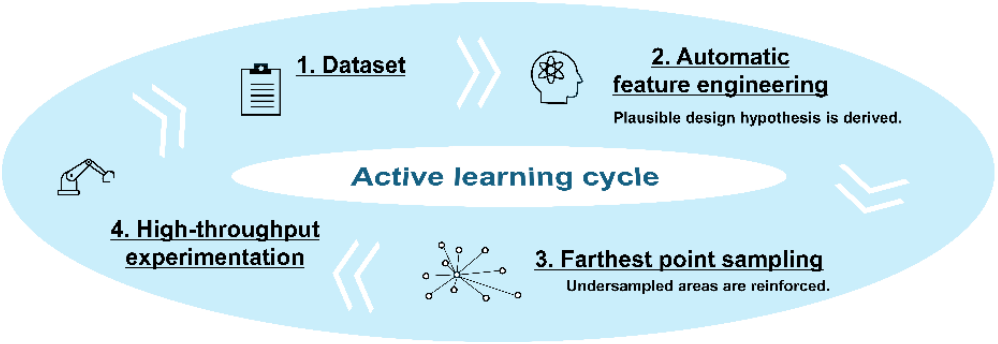


Figure 5.1. Active learning cycle employed in this study. Automatic feature engineering (AFE) was applied to a given catalyst dataset to derive a design hypothesis (an machine learning model with tailored descriptors). Catalysts recommended by farthest point sampling (FPS) were assessed using high-throughput experimentation (HTE). The resulting data were integrated back into the dataset, and this iterative process continued until a robust design hypothesis was established to elucidate the relationship between catalyst compositions and performances.

5.2.1. Dataset

Taniike and his group accumulated OCM data for quaternary catalysts represented as M1–M2–M3/Support using a consistent experimental protocol involving HTE.[25–28] M1–M3 correspond to the supported elements, which can be selected (with duplication allowed) from the following: Li, Na, Mg, K, Ca, Ti, V, Mn, Fe, Co, Ni, Cu, Zn, Sr, Y, Zr, Mo, Pd, Cs, Ba, La, Ce, Nd, Eu, Tb, Hf, W, and none (where "none" indicates no addition of elements). The selected supports were MgO, Al_2O_3, SiO_2, CaO, TiO_2, ZrO_2, BaO, La_2O_3, and CeO_2. The loading amount of supported elements, except for the "none" option, was set at 0.37 mmol per gram of support for each selection. This resulted in a parameter space containing 4,060 catalysts per support, amounting to a total of 36,540 catalysts.

In this study, we extracted 381 catalysts related to CaO, BaO, La_2O_3, TiO_2, and ZrO_2 supports from the 636 quaternary catalysts we previously reported to establish design hypotheses through active learning.[28–30] CaO and La_2O_3 were selected because they are the most extensively studied basic oxides in OCM and are likely to exhibit distinct catalyst designs.[31] BaO was selected as the reference for these supports because it offers higher C_2 yields and selectivities. In contrast, the redox-active supports TiO_2 and ZrO_2 were included to investigate potential relationships between the general physical properties and design hypotheses. Among the 381 catalysts, 175 were obtained via random sampling from the entire space, whereas the remaining 206 were selected to validate various ML techniques.

5.2.2. Automatic feature engineering

AFE automates the design of physically meaningful features for a given catalyst dataset within the framework of supervised ML. This process involves a structured pipeline of feature assignment, synthesis, and selection. First, physical quantities of elements are assigned to the catalysts, with their elemental compositions represented through commutative operations. Higher-order features involving nonlinear and combinatorial effects are then synthesized from these assigned primitive features using mathematical operations. Finally, a specified number of features (descriptors) that optimize the score of the supervised ML are selected from the large synthesized array of features.

In this study, we utilized 58 parameters of elements from XenonPy, normalized according to the literature.[32] These parameters were assigned to each catalyst using five commutative operations (maximum, minimum, average, product, and standard deviation), yielding 290 primary features. These primary features were further synthesized into 3,480 features using 12 functional forms (x, sqrt(x), x^2, x^3, ln(x), exp(x), and their reciprocals, where x represents each primary feature). A genetic algorithm-based approach was employed to select eight features that minimized the mean absolute error (MAE) in leave-one-out cross-validation (LOOCV) using Huber regression.[33] This process involved assessing approximately 4,000,000 models per dataset with various feature combinations and selecting the combination of features (X) and model (f(X)) that yielded the lowest cross-validation (CV) score as the most plausible design hypothesis. Huber regression, which is a form of multiple linear regression, was employed to prevent overfitting owing to its reduced number of parameters and ensure robustness against outliers, such as experimental failures. The number of selected features was empirically determined to balance the CV score and the cost of feature selection. Further details on AFE and parameter selection are provided in previous study.[26]

5.2.3. Farthest point sampling

To validate and refine the design hypothesis presented in Section 5.2.2, we added either 10 or 20 catalysts to each active learning cycle for experimental testing. Among these, 90% were selected using FPS within the normalized eight-dimensional feature space established by AFE. The remaining 10% corresponded to the re-evaluation of the catalysts that exhibited the largest deviations between observed and predicted values in the last cycle.

5.2.4. High-throughput experimentation

The catalysts proposed in Section 5.2.3 were prepared and evaluated using the same experimental methods and conditions as those employed when acquiring the original training data, which are briefly described as follows.

5.2.4.1. Materials

The metal precursors used were $LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, KNO_3, $Ca(NO_3)_2{\cdot}4H_2O$, $Ti(OiPr)_4$, $VOSO_4{\cdot}xH_2O$ ($x = 3$–5), $Mn(NO_3)_2{\cdot}6H_2O$, $Fe(NO_3)_3{\cdot}9H_2O$, $Co(NO_3)_2{\cdot}6H_2O$, $Ni(NO_3)_2{\cdot}6H_2O$, $Cu(NO_3)_2{\cdot}3H_2O$, $Zn(NO_3)_2{\cdot}6H_2O$, $Sr(NO_3)_2$, $Y(NO_3)_3{\cdot}6H_2O$, $ZrO(NO_3)_2{\cdot}xH_2O$ ($x = 2$), $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, $Pd(OAc)_2$, $CsNO_3$, $Ba(NO_3)_2$, $La(NO_3)_3{\cdot}6H_2O$, $Ce(NO_3)_3{\cdot}6H_2O$, $Nd(NO_3)_3{\cdot}6H_2O$, $Eu(NO_3)_3{\cdot}5H_2O$, $Tb(NO_3)_3{\cdot}5H_2O$, $Hf(OEt)_4$, and $(NH_4)_{10}H_2(W_2O_7)_6$. These materials were purchased from one of the following suppliers: Sigma-Aldrich, Kanto Chemical, Wako Pure Chemical Industries, Alfa-Aesar, or Sumitomo Chemical.

The oxide supports and their precursors included $Ca(OH)_2$ (3.0 m^2/g, Wako Pure Chemical Industries), $Ba(OH)_2{\cdot}8H_2O$ (1.1 m^2/g, Wako Pure Chemical Industries), La_2O_3 (8.3 m^2/g, Wako Pure Chemical Industries), Y_2O_3 (5.01 g/mL, Sigma-Aldrich) , TiO_2 (17.4 m^2/g, anatase type, Kanto Chemical), and ZrO_2 (3.2 m^2/g, Kanto Chemical).

5.2.4.2. Preparation of catalysts

The catalysts were prepared using a parallelized wet impregnation method, in which support powder was impregnated with an aqueous solution of specified metal precursors at 50 °C for 6 h. The loading amount of elements was fixed at 0.37 mmol/g-support per selection within M1–3. Following impregnation, the powder was vacuum-dried and calcined in air at 1000 °C for 3 h. For precursors that contained metal alkoxides, impregnation was performed in two steps: first, with an aqueous solution of the other precursors, followed by impregnation with an ethanol solution of the metal alkoxides.

5.2.4.3. Evaluation of catalysts

The catalysts were evaluated for OCM using a HTE system. This system operates through a combination of a gas mixer, flow distributor, electric furnace-bearing reactors, autosampler, and quadrupole mass spectrometer (QMS), enabling the performance of 20 catalysts to be automatically assessed under a programmed sequence of reaction conditions. The catalysts were secured in beds 1 cm in height within reaction tubes made of fused quartz, which had inner diameters of 4 and 2 mm, using quartz wool. After inline activation at 1000 °C for 3 h in an oxygen stream, the 20 catalysts were tested under 135 conditions varying in temperature (700, 750, 800, 850, and 900 °C), total gas flow rates (10, 15, and 20 mL/min/channel), CH_4/O_2 ratios (2, 4, and 6 mol/mol), and Ar partial pressures as a balancing gas (0.15, 0.40, and 0.70 atm). Each catalyst was labeled according to the highest C_2 yield achieved among all 135 conditions.

5.3. RESULTS AND DISCUSSION

5.3.1. Evaluated catalysts

In this study, the catalysts evaluated from the BaO, CaO, La_2O_3, TiO_2, Y_2O_3, and ZrO_2 families were proposed by FPS, based on AFE design hypotheses using past data as the original dataset. In contrast, the Y_2O_3 family was selected randomly. Additionally, among the BaO catalysts, 40 catalysts were evaluated, selected in the order proposed by FPS, with predicted yields of 15% or higher. Randomly selected catalysts were chosen without human bias, while catalysts chosen by FPS are considered to comprehensively represent the entire catalysts space. Therefore, these catalysts can be seen as representative of the overall characteristics of their respective families.

The top C_2 yield and the corresponding conditions for all catalysts measured in this study are shown in Tables 5.1 to 5.6. Additionally, the distributions of C_2 yield, CH_4 conversion, and C_2 selectivity for each support are represented in histograms (Figures 5.1 to 5.3).

From these graph, BaO demonstrates balanced performance across all metrics, including C_2 yield, CH_4 conversion, and C_2 selectivity, indicating its effectiveness in achieving both high yield and selectivity for C_2 products. In contrast, Y_2O_3 exhibits efficient CH_4 conversion but suffers from low C_2 selectivity, suggesting it may not be suitable for efficient C_2 product generation.

These findings suggest that BaO outperforms other supports in C_2 production, showing superior overall performance. In contrast, Y_2O_3 is characterized by high CH_4 conversion yet limited C_2 selectivity, highlighting its limitations in selective C_2 production.

For another catalysts, CaO and La_2O_3 show peaks at low C_2 selectivity, likely due to their excessively high basicity. This high basicity may lead to decreased selectivity through several mechanisms: (1) promoting side reactions, such as complete combustion of methane (CH_4) and carbon deposition, which lowers the C_2 yield; (2) inhibiting C–C bond formation, which is essential for generating C_2 products like ethylene and ethane; and (3) causing excessive catalytic activity on the surface, which accelerates CH_4 conversion but disrupts selective conversion to C_2 products. These observations suggest that the high basicity of CaO and La_2O_3 contributes to their low C_2 selectivity, and optimizing C_2 production may require adjusting the basicity or selecting a more suitable support material.

Table 5.1. Development of design hypotheses for BaO-based catalysts throughout the active learning cycles. This table shows the scores and selected features of the design hypotheses obtained in each individual active learning cycle. Catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp. (°C)
1	Cs–Cs–Cs/BaO	20.61	40.28	51.16	850
2	Ba–Ba–Ba/BaO	5.48	7.22	75.88	800
3	Fe–Ni–Pd/BaO	3.87	56.80	6.81	900
4	Pd–Pd–Pd/BaO	7.14	26.85	26.58	850
5	Co–Cu–Pd/BaO	6.89	39.45	17.45	750
6	Li–Ni–Pd/BaO	19.29	41.00	47.06	850
7	K–K–K/BaO	9.75	27.33	35.67	900
8	Co–Co–Co/BaO	n.d.	n.d.	n.d.	n.d.
9	Y–Zr–Pd/BaO	14.10	32.64	43.20	800
10	Mn–Mn–Mn/BaO	21.12	38.66	54.64	750
11	Nd–Nd–Nd/BaO	17.27	28.33	60.95	800
12	K–Cs–Cs/BaO	18.05	37.32	48.38	900
13	Fe–Ni–Mo/BaO	19.68	35.52	55.40	850
14	Mo–Mo–Mo/BaO	6.99	27.93	25.02	800
15	Fe–Pd–W/BaO	11.54	26.94	42.83	850
16	Cu–Mo–Pd/BaO	6.39	10.71	59.60	750
17	Ni–Ni–Ni/BaO	5.22	30.20	17.30	850
18	Mn–Y–Pd/BaO	18.56	38.50	48.20	800
19	K–Mo–La/BaO	18.78	40.43	46.46	850
20	Sr–Zr–La/BaO	20.05	29.78	67.33	800
21	W–W–W/BaO	21.06	24.88	84.65	800

22	Mn–Fe–Pd/BaO	13.74	32.33	42.48	850
23	Mn–Fe–Mo/BaO	17.03	42.14	40.42	850
24	Mn–Mn–Zn/BaO	16.83	36.54	46.06	850
25	Mn–Co–Hf/BaO	13.39	33.12	40.43	850
26	Co–Pd–Cs/BaO	6.87	11.37	60.39	750
27	Ni–Mo–Mo/BaO	18.31	38.91	47.05	850
28	Mg–Zn–Zn/BaO	18.62	34.11	54.58	800
29	Mn–Ni–Pd/BaO	9.75	31.00	31.44	850
30	Ni–Zn–Sr/BaO	10.01	23.02	43.51	850
31	Mn–Fe–Cs/BaO	17.55	35.89	48.89	850
32	Mn–Mn–Fe/BaO	14.62	28.01	52.20	800
33	Mg–Mg–none/BaO	16.96	36.09	47.00	850
34	Zn–Zn–Ce/BaO	17.99	27.46	65.50	750
35	K–Fe–Pd/BaO	7.28	16.15	45.12	850
36	Mn–Co–Cu/BaO	7.83	30.91	25.33	850
37	Sr–Sr–Cs/BaO	18.18	25.63	70.95	850
38	Ca–Co–Ni/BaO	10.56	33.92	31.13	900
39	Mn–Y–Pd/BaO	5.06	17.56	28.83	800
40	Li–Ni–Pd/BaO	5.54	19.38	28.58	750
41	Pd–Cs–Cs/BaO	n.d.	n.d.	n.d.	n.d.
42	Cu–Pd–Cs/BaO	5.48	34.76	15.76	850
43	Ni–Cs–Cs/BaO	7.65	30.09	25.42	850
44	Cu–Zn–Zn/BaO	7.29	32.46	22.46	750
45	Mn–Cs–Cs/BaO	10.63	32.70	32.52	800
46	Co–Cs–Cs/BaO	13.08	34.09	38.37	850
47	Mg–Mg–Mn/BaO	17.22	37.14	46.36	800
48	Pd–W–W/BaO	n.d.	n.d.	n.d.	n.d.
49	Ba–Ce–Ce/BaO	14.53	30.60	47.50	750
50	Cu–W–W/BaO	7.43	29.83	24.90	850
51	Na–K–none/BaO	17.40	35.07	49.63	800
52	Na–Cs–Cs/BaO	16.50	22.59	73.06	800
53	Mn–Mn–W/BaO	16.41	38.24	42.91	800
54	K–Zn–Cs/BaO	18.53	36.27	51.09	800
55	Hf–Hf–Hf/BaO	17.30	39.95	43.30	800
56	Zn–Mo–Pd/BaO	11.72	36.44	32.16	850

57	Eu–Eu–Eu/BaO	19.43	37.03	52.46	750
58	Zr–Cs–Cs/BaO	17.18	39.41	43.61	850
59	Mo–Mo–Mo/BaO	18.94	38.98	48.59	850
60	Mg–Pd–Ce/BaO	n.d.	n.d.	n.d.	n.d.
61	Na–Na–Na/BaO	16.14	37.91	42.57	850
62	Sr–Ba–Ba/BaO	15.24	38.02	40.08	800
63	Co–Mo–Pd/BaO	9.48	19.11	49.61	800
64	Na–Mg–Zn/BaO	15.99	22.93	69.76	800
65	Na–Ca–Ti/BaO	14.81	36.43	40.66	800
66	Nd–none–none/BaO	18.30	39.28	46.60	750
67	Na–Mg–Ni/BaO	2.04	42.35	4.82	750
68	K–Cu–Mo/BaO	7.11	30.31	23.44	750
69	Na–Mg–Mg/BaO	14.75	34.85	42.33	800
70	Zn–Nd–W/BaO	17.17	39.36	43.63	800
71	Mn–Mn–Cu/BaO	6.97	33.59	20.75	850
72	Ca–none–none/BaO	14.84	39.52	37.54	850
73	Mg–Ca–Cu/BaO	5.66	33.94	16.66	850
74	Mg–Mg–Nd/BaO	17.04	37.76	45.12	800
75	Mg–Nd–none/BaO	17.46	38.16	45.75	750
76	Mg–Nd–Nd/BaO	17.37	39.36	44.13	700
77	Mn–Mn–none/BaO	11.26	33.09	34.03	750
78	Mn–none–none/BaO	11.38	34.46	33.02	800
79	Ca–W–none/BaO	17.70	39.16	45.21	850
80	Pd–W–W/BaO	n.d.	n.d.	n.d.	n.d.

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 5 catalysts was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

Table 5.2. Development of design hypotheses for BaO-based catalysts throughout the active learning cycles. This table shows the catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp. (°C)
81	Fe–Fe–Fe/BaO	20.38	39.83	51.17	850
82	K–Fe–Zn/BaO	17.77	39.49	45.00	850
83	Fe–Zn–none/BaO	18.94	37.96	49.89	850
84	Fe–none–none/BaO	19.42	38.73	50.14	850
85	K–Co–Zn/BaO	10.44	29.60	35.27	850
86	Li–Fe–Zn/BaO	13.95	34.82	40.06	850
87	Cs–none–none/BaO	14.95	22.37	66.83	850
88	Co–Zn–none/BaO	11.76	28.49	41.28	850
89	Ca–Fe–Zn/BaO	19.63	38.68	50.75	850
90	Mg–K–none/BaO	14.99	34.67	43.24	850
91	Ti–none–none/BaO	17.32	24.60	70.41	850
92	K–Zn–none/BaO	16.49	36.68	44.96	850
93	Na–none–none/BaO	15.80	21.80	72.48	850
94	Mo–none–none/BaO	20.14	40.60	49.61	850
95	Ce–none–none/BaO	18.58	32.89	56.49	850
96	Fe–Ni–Zn/BaO	15.30	30.25	50.58	850
97	Eu–Eu–none/BaO	18.07	34.61	52.21	850
98	K–none–none/BaO	16.77	31.40	53.41	850
99	Li–Co–Zn/BaO	9.33	28.86	32.33	850
100	Ba–Ba–Ba/BaO	13.10	35.75	36.64	850
101	Zn–Zn–Zn/BaO	17.54	37.16	47.20	850
102	Mo–Mo–Ce/BaO	13.03	18.24	71.44	800
103	Zn–Mo–none/BaO	19.34	32.06	60.32	800
104	Fe–Mo–Eu/BaO	19.82	36.95	53.64	850
105	V–Zn–Zn/BaO	12.45	14.87	83.73	850
106	Ti–Mo–Mo/BaO	14.11	25.08	56.26	800
107	Ti–Zn–Mo/BaO	18.70	37.18	50.30	850
108	V–Hf–Hf/BaO	16.11	35.48	45.41	850
109	Fe–Mo–Mo/BaO	18.00	36.85	48.85	850

110	Mo–Mo–W/BaO	18.87	39.01	48.37	850
111	Zr–La–La/BaO	19.58	38.05	51.46	800
112	V–La–La/BaO	18.88	34.59	54.58	800
113	Ti–V–W/BaO	16.32	22.22	73.45	850
114	Ti–W–W/BaO	19.56	37.28	52.47	850
115	Ti–La–La/BaO	19.34	35.10	55.10	800
116	Na–Tb–Tb/BaO	15.95	32.02	49.81	800
117	Li–Ti–Zr/BaO	16.15	35.60	45.37	750
118	Fe–Mo–La/BaO	20.10	34.29	58.62	800
119	Mn–Mn–Mn/BaO	11.39	29.89	38.11	850
120	Mg–Pd–none/BaO	n.d.	n.d.	n.d.	n.d.

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 1 catalyst was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

Table 5.3. Development of design hypotheses for CaO-based catalysts throughout the active learning cycles. This table shows the catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp. (°C)
121	W–W–W/CaO	12.77	43.47	29.38	800
122	Cu–Cu–Cu/CaO	5.69	38.85	14.65	700
123	Ni–Cs–W/CaO	3.64	36.58	9.95	700
124	Ni–Zn–W/CaO	0.73	46.28	1.58	850
125	Cs–W–W/CaO	13.75	44.2	31.11	750
126	Hf–Hf–Hf/CaO	13.39	44.08	30.38	750

127	Mn–Cs–W/CaO	8.62	39.8	21.66	750
128	Zn–Cs–W/CaO	12.91	43.56	29.64	800
129	Eu–Eu–W/CaO	14.98	43.83	34.18	700
130	Mg–Ni–Y/CaO	0.71	46.25	1.54	850
131	Zn–Pd–Ce/CaO	4.19	43.15	9.71	800
132	Fe–Pd–Nd/CaO	1.36	25.32	5.37	800
133	Fe–Zn–Pd/CaO	5.86	41.16	14.24	850
134	Ni–Pd–Nd/CaO	6.26	38.09	16.43	800
135	Fe–Ni–Ce/CaO	0.88	32.12	2.74	800
136	Co–Ni–Pd/CaO	1.07	48.45	2.21	750
137	Mn–Zn–Ce/CaO	7.6	39.13	19.42	750
138	Mn–Pd–Nd/CaO	8.12	39.64	20.48	800
139	Co–Pd–Ce/CaO	0.82	25.01	3.28	850
140	Ca–Nd–Tb/CaO	15.68	44.97	34.87	700
141	Y–Hf–W/CaO	8.11	41.83	19.39	750
142	Cu–Y–Hf/CaO	6.38	40.77	15.65	750
143	Pd–Tb–Hf/CaO	n.d.	n.d.	n.d.	n.d.
144	Eu–Eu–Eu/CaO	14.64	43.95	33.31	700
145	Ni–Y–Pd/CaO	n.d.	n.d.	n.d.	n.d.
146	Ni–Tb–W/CaO	1.41	23.52	5.99	700
147	Ni–Zn–Y/CaO	1.03	25.6	4.02	800
148	Ni–Cu–Hf/CaO	0.75	24.16	3.10	800
149	Li–La–W/CaO	14.73	44.27	33.27	700
150	Li–Y–La/CaO	10.18	40.72	25.00	700
151	Ni–Ni–Ni/CaO	1.17	24.53	4.77	750
152	Co–Tb–none/CaO	4.31	35.68	12.08	700
153	Co–Zn–Tb/CaO	4.16	37.08	11.22	800
154	Li–Ce–none/CaO	14.01	43.49	32.21	700
155	Co–Pd–Cs/CaO	1.81	35.05	5.16	750
156	Cu–Ce–none/CaO	6.75	38.98	17.32	750
157	Mn–Cu–Tb/CaO	5.49	41.15	13.34	850
158	Co–Zn–Ce/CaO	5.52	36.92	14.95	750
159	Co–Pd–Tb/CaO	1.29	28.46	4.53	750
160	Mn–Pd–Nd/CaO	1.25	35.07	3.56	700
161	Li–Pd–W/CaO	0.54	51.11	1.06	800

162	Zn–Zn–Mo/CaO	12.41	42.93	28.91	800
163	Cs–Cs–Cs/CaO	10.09	39.37	25.63	750
164	Li–Zn–Mo/CaO	12.79	42.94	29.79	800
165	Na–Na–Na/CaO	8.07	39.05	20.67	750
166	Co–Mo–Pd/CaO	0.57	52.44	1.09	900
167	Na–Cs–Cs/CaO	7.8	37.81	20.63	750
168	Li–Zn–none/CaO	13.35	42.24	31.61	750
169	Li–Ni–W/CaO	1.79	24.82	7.21	700
170	Ni–Pd–Nd/CaO	1.14	35.81	3.18	800

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 2 catalysts was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

Table 5.4. Development of design hypotheses for La_2O_3-based catalysts throughout the active learning cycles. This table shows the catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp. (°C)
171	Y–none–none/La_2O_3	12.58	42.22	29.80	700
172	Cu–none–none/La_2O_3	3.73	39.49	9.45	850
173	Mg–none–none/La_2O_3	10.34	40.65	25.44	700
174	Co–W–W/La_2O_3	1.35	41.47	3.26	700
175	Pd–W–W/La_2O_3	0.39	50.17	0.78	900
176	Zn–none–none/La_2O_3	15.21	45.06	33.75	700
177	Mn–Mn–Mn/La_2O_3	10.62	41.23	25.76	750
178	Zr–none–none/La_2O_3	12.9	43.62	29.57	700

179	W–none–none/La_2O_3	12.14	4.99	243.29	750
180	Fe–Sr–Y/La_2O_3	15.01	44.4	33.81	750
181	Ce–Ce–Ce/La_2O_3	13.79	41.58	33.16	700
182	Zn–Cs–none/La_2O_3	14.84	45.38	32.70	700
183	Ni–Zn–Cs/La_2O_3	11.61	31.68	36.65	800
184	Ni–Pd–Ba/La_2O_3	1.22	25.64	4.76	800
185	Ni–Cs–W/La_2O_3	1.19	37.45	3.18	750
186	Cs–Ce–Ce/La_2O_3	12.96	42.96	30.17	700
187	Na–Ce–Ce/La_2O_3	13.17	44.27	29.75	700
188	Ba–Ba–Ba/La_2O_3	19.13	48.3	39.61	750
189	Na–W–none/La_2O_3	19.37	49.19	39.38	800
190	Ti–Zr–Ba/La_2O_3	8.28	28.4	29.15	800
191	Li–none–none/La_2O_3	13.55	43.24	31.34	700
192	Co–Ni–Zn/La_2O_3	0.6	28.59	2.10	700
193	Ni–none–none/La_2O_3	0.65	40.27	1.61	800
194	Co–Cs–Cs/La_2O_3	1.55	43.81	3.54	750
195	Mg–Zn–Pd/La_2O_3	1.04	21.5	4.84	750
196	Ca–Co–Zn/La_2O_3	6.66	38.01	17.52	750
197	Tb–none–none/La_2O_3	8.48	8.9	95.28	750
198	Co–Pd–Cs/La_2O_3	0.63	48.42	1.30	750
199	Pd–none–none/La_2O_3	n.d.	n.d.	n.d.	n.d.
200	La–Tb–none/La_2O_3	13.26	43.24	30.67	700
201	Cs–none–none/La_2O_3	11.37	41.01	27.72	700
202	Mn–Pd–Cs/La_2O_3	1	42.82	2.34	800
203	Zn–Pd–Ba/La_2O_3	1.04	27.72	3.75	850
204	Na–Co–Pd/La_2O_3	0.8	31.93	2.51	750
205	K–Co–Zn/La_2O_3	2.07	36.95	5.60	800
206	Cs–Cs–Cs/La_2O_3	13.91	43.87	31.71	700
207	Li–Ni–Pd/La_2O_3	0.81	24.59	3.29	850
208	Co–Zn–Pd/La_2O_3	0.87	24.78	3.51	850
209	Cs–Ba–Ba/La_2O_3	16.68	46.11	36.17	700
210	Ti–W–W/La_2O_3	11.2	43.43	25.79	700
211	Li–Li–Fe/La_2O_3	10.75	39.77	27.03	750
212	Fe–Pd–Eu/La_2O_3	2.81	36.57	7.68	750
213	Na–K–Ce/La_2O_3	8.69	27.02	32.16	750

214	Na–none–none/La_2O_3	12.15	41.22	29.48	700
215	Pd–Cs–Cs/La_2O_3	n.d.	n.d.	n.d.	n.d.
216	K–Fe–Cs/La_2O_3	8.19	34.69	23.61	700
217	Na–Na–Cs/La_2O_3	13.14	39.86	32.97	700
218	K–Fe–Pd/La_2O_3	1.36	36.5	3.73	750
219	Li–Li–Na/La_2O_3	12.35	39.59	31.19	700
220	Cu–Cs–Cs/La_2O_3	3.51	18.08	19.41	800
221	Co-Pd-Ce/La_2O_3	9.34	29.96	31.17	850
222	Mn-Mn-Pd/La_2O_3	1.52	35.88	4.24	750
223	Mn-Mn-Fe/La_2O_3	1.24	38.55	3.22	700
224	Mn-Ni-Ce/La_2O_3	n.d.	n.d.	n.d.	n.d.
225	Fe-Cu-Cu/La_2O_3	5.56	24.24	22.95	750
226	V-Mn-Cu/La_2O_3	6.4	36.69	17.43	800
227	Mn-Mn-Ni La_2O_3	5.56	39.97	13.9	850
228	Li-Cu-Cu/La_2O_3	2.2	18.74	11.74	750
229	Li-Li-Mn/La_2O_3	10.24	39.54	25.9	700
230	Na-W-none/La_2O_3	8.71	27.64	31.53	700

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 2 catalysts was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

Table 5.5. Development of design hypotheses for TiO_2-based catalysts throughout the active learning cycles. This table shows the catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp. (°C)

231	Li–Li–Li/TiO_2	13.21	42.55	31.05	850
232	Pd–none–none/TiO_2	0.53	75.9	0.70	900
233	W–none–none/TiO_2	7.91	37.01	21.37	800
234	Cs–Cs–Cs/TiO_2	13.46	47.37	28.41	900
235	Ni–Ni–none/TiO_2	3.61	23.39	15.43	750
236	Ni–Ni–Pd/TiO_2	0.82	31.76	2.58	700
237	Na–Na–Na/TiO_2	12.49	44.71	27.94	900
238	Pd–Pd–Pd/TiO_2	n.d.	n.d.	n.d.	n.d.
239	Pd–Cs–W/TiO_2	12.81	36.54	35.06	850
240	Ti–Y–Ce/TiO_2	12.64	43.36	29.15	750
241	K–none–none/TiO_2	17.03	24.65	69.09	850
242	K–Ni–W/TiO_2	16.6	24.81	66.91	800
243	K–K–Eu/TiO_2	13.3	28.44	46.77	800
244	Mg–none–none/TiO_2	10.99	37.47	29.33	900
245	Ba–Ba–Ba/TiO_2	10.63	33.02	32.19	850
246	Pd–Pd–Cs/TiO_2	3.22	25.46	12.65	900
247	Cu–none–none/TiO_2	8.89	30.4	29.24	900
248	K–Pd–W/TiO_2	12.83	29.47	43.54	850
249	Ni–Pd–W/TiO_2	4.34	28.14	15.42	850
250	Mg–Ca–Pd/TiO_2	1.22	13.56	9.00	700
251	Pd–Cs–Ce/TiO_2	1.14	27.1	4.21	900
252	Pd–Ce–W/TiO_2	2.46	28.38	8.67	900
253	Ca–Ca–Ca/TiO_2	10.8	38.94	27.73	850
254	Na–Pd–Ce/TiO_2	0.82	22.46	3.65	700
255	Na–Mo–Cs/TiO_2	11.86	38.46	30.84	850
256	Ca–Cu–none/TiO_2	2.52	35.39	7.12	700
257	Li–Li–Mg/TiO_2	7.78	33.33	23.34	750
258	Cs–Ce–W/TiO_2	17.53	48.31	36.29	850
259	Li–Cu–none/TiO_2	7.92	40.54	19.54	850
260	Na–K–Cu/TiO_2	7.96	40.69	19.56	850
261	Hf–W–none/TiO_2	7.36	43.1	17.08	850
262	Ti–none–none/TiO_2	10.55	39.09	26.99	850
263	Cu–Cu–Cu/TiO_2	7.7	41.86	18.39	850
264	Na–Pd–Hf/TiO_2	0.5	34.87	1.43	900
265	Li–Na–Sr/TiO_2	10.61	43.32	24.49	900

266	K–Pd–Tb/TiO_2	0.53	23	2.30	700
267	K–Ce–Hf/TiO_2	17.06	47.7	35.77	850
268	Mo–Pd–Cs/TiO_2	8.59	34.57	24.85	800
269	Li–Ti–Pd/TiO_2	0.97	29.27	3.31	900
270	Li–Zn–Pd/TiO_2	6.47	47.03	13.76	900
271	Ti–Ti–Hf/TiO_2	10.19	43.24	23.57	850
272	Ba–Hf–Hf/TiO_2	8.26	40.09	20.60	800
273	Pd–Hf–Hf/TiO_2	n.d.	n.d.	n.d.	n.d.
274	K–Cs–Cs/TiO_2	17.1	44.25	38.64	850
275	Mo–W–W/TiO_2	6.09	42.14	14.45	800
276	K–Mo–Pd/TiO_2	9.26	35.88	25.81	800
277	Mo–Mo–Pd/TiO_2	1.33	22.73	5.85	700
278	Hf–Hf–Hf/TiO_2	6.98	40.39	17.28	800
279	Pd–Ba–Hf/TiO_2	1.47	18.93	7.77	700
280	Pd–W–W/TiO_2	3.44	20.89	16.47	900
281	V–Co–Zr/TiO_2	7.2	35.53	20.26	800
282	Mo–Mo–Mo/TiO_2	8.03	41.97	19.13	850
283	Pd–Tb–Tb/TiO_2	0.59	26.24	2.25	900
284	Pd–Cs–Hf/TiO_2	0.89	24.95	3.57	700
285	Cs–Hf–Hf/TiO_2	13.52	36.78	36.76	800
286	Tb–Tb–Tb/TiO_2	8.96	42.5	21.08	850
287	Ba–W–W/TiO_2	7.41	42.26	17.53	850
288	Mo–Hf–Hf/TiO_2	6.41	35	18.31	750
289	Ni–Zn–La/TiO_2	0.83	28.1	2.95	850
290	Mn–Nd–W/TiO_2	9.33	40.47	23.05	850
291	Ca-none-none/TiO_2	9.7	38.03	25.51	800
292	K-W-none/TiO_2	11.13	26.56	41.9	800
293	Cs-W-none/TiO_2	13.77	43.54	31.63	850
294	Cs-Cs-La/TiO_2	14.36	42	34.19	850
295	Fe-Cs-none/TiO_2	13.19	41.69	31.65	850
296	K-Sr-Ba/TiO_2	10.47	42.52	24.62	800
297	Na-Cs-none/TiO_2	11.37	32.03	35.49	900
298	Zr-Cs-none/TiO_2	11.82	36.74	32.17	850
299	Na-Y-none/TiO_2	9.27	36.74	25.24	800
300	Ni-Zn-La/TiO_2	3.43	19.9	17.23	850

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 2 catalysts was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

Table 5.6. Development of design hypotheses for ZrO_2-based catalysts throughout the active learning cycles. This table shows the catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp. (°C)
301	Cs–Cs–Cs/ZrO_2	7.89	35.21	22.42	800
302	W–W–W/ZrO_2	6.63	36.18	18.31	750
303	Ba–Ba–Ba/ZrO_2	12.53	35.13	35.67	750
304	K–K–K/ZrO_2	9.47	26.97	35.10	750
305	Li–Li–Li/ZrO_2	8.65	36.28	23.84	800
306	Cu–none–none/ZrO_2	8.84	29.61	29.85	750
307	Li–Ti–none/ZrO_2	12.10	34.82	34.76	850
308	Mg–Zn–Cs/ZrO_2	7.30	27.70	26.35	750
309	Ce–Eu–none/ZrO_2	6.74	36.14	18.65	800
310	Hf–W–W/ZrO_2	6.77	27.14	24.95	800
311	Li–Cs–W/ZrO_2	14.64	22.07	66.34	750
312	Mg–Eu–none/ZrO_2	6.98	33.21	21.01	750
313	Mo–Pd–Pd/ZrO_2	6.61	37.65	17.55	800
314	Mn–Zn–none/ZrO_2	7.68	32.56	23.59	750
315	K–Cs–Cs/ZrO_2	n.d.	n.d.	n.d.	n.d.
316	Mg–Mo–Cs/ZrO_2	9.33	36.38	25.64	800
317	Li–Ti–W/ZrO_2	9.13	31.61	28.87	850

318	Cs–Cs–Eu/ZrO_2	9.07	34.35	26.40	750
319	K–Y–La/ZrO_2	10.91	19.48	56.01	750
320	Li–Mo–Nd/ZrO_2	11.17	35.88	31.13	850
321	Ni–Pd–W/ZrO_2	n.d.	n.d.	n.d.	n.d.
322	Cu–Zn–Pd/ZrO_2	6.64	25.97	25.60	800
323	Li–Ni–Pd/ZrO_2	4.34	74.92	5.73	700
324	Li–Pd–W/ZrO_2	7.67	40.26	18.99	900
325	Li–Li–Pd/ZrO_2	n.d.	n.d.	n.d.	n.d.
326	Pd–Hf–W/ZrO_2	3.71	11.81	31.73	900
327	Ni–Pd–Hf/ZrO_2	2.05	14.45	14.33	700
328	Hf–Hf–Hf/ZrO_2	5.45	36.95	14.85	800
329	Fe–Pd–none/ZrO_2	7.61	31.13	24.48	850
330	Zn–Zn–W/ZrO_2	6.42	31.85	20.63	750
331	Li–Li–Hf/ZrO_2	9.74	32.82	29.81	800
332	Li–Mo–Pd/ZrO_2	7.93	32.45	24.46	800
333	Mn–Zr–Pd/ZrO_2	3.10	13.01	24.91	800
334	Li–Li–Zn/ZrO_2	8.57	33.27	25.66	850
335	Ni–Cu–W/ZrO_2	1.87	29.70	6.40	900
336	Zn–Pd–Tb/ZrO_2	4.26	29.01	14.79	700
337	Li–Ni–W/ZrO_2	12.24	29.81	41.15	850
338	Cu–Zn–Mo/ZrO_2	0.00	0.00	0.00	0
339	Zn–Zn–Hf/ZrO_2	6.86	28.12	24.72	800
340	K–Cs–Cs/ZrO_2	6.69	31.51	21.19	750
341	Mg–none–none/ZrO_2	6.85	41.54	16.49	800
342	La–none–none/ZrO_2	5.73	40.98	13.98	750
343	Mn–Mn–Mn/ZrO_2	5.50	38.63	14.24	800
344	Na–Na–Mg/ZrO_2	9.91	28.62	34.63	800
345	Ca–none–none/ZrO_2	6.80	27.35	24.86	800
346	Pd–none–none/ZrO_2	n.d.	n.d.	n.d.	n.d.
347	Ti–W–W/ZrO_2	4.98	37.39	13.32	750
348	Ti–none–none/ZrO_2	5.35	35.48	15.08	750
349	Ce–Hf–none/ZrO_2	5.60	39.08	14.33	750
350	Li–Li–Pd/ZrO_2	0.13	24.58	0.53	700
351	Cu–Pd–Cs/ZrO_2	5.27	48.07	10.97	900
352	Ni–Ni–Ni/ZrO_2	1.08	42.18	2.56	750

353	Li–Cu–Cu/ZrO_2	7.89	24.65	31.98	700
354	Cu–Mo–Pd/ZrO_2	7.40	27.89	26.53	850
355	Cu–Cs–Cs/ZrO_2	6.49	14.86	43.68	800
356	Mo–Mo–Mo/ZrO_2	11.24	20.76	54.14	800
357	Cu–Pd–Eu/ZrO_2	7.39	28.87	25.60	750
358	Cu–Mo–Mo/ZrO_2	8.87	32.35	27.41	850
359	Ce–Ce–Ce/ZrO_2	6.94	31.40	22.09	800
360	Li–Ca–Ca/ZrO_2	10.88	28.55	38.12	750
361	Li–Cu–Cs/ZrO_2	3.66	21.71	16.84	900
362	Li–none–none/ZrO_2	5.54	33.77	16.41	750
363	Li–Cs–Cs/ZrO_2	6.37	45.63	13.96	800
364	Cs–none–none/ZrO_2	6.27	25.65	24.43	800
365	Li–Ni–Cs/ZrO_2	1.44	30.84	4.66	700
366	Ni–Zn–Zn/ZrO_2	1.24	24.01	5.16	700
367	Na–Na–Cu/ZrO_2	2.60	22.28	11.67	900
368	Fe–Cs–Cs/ZrO_2	5.76	41.17	13.99	750
369	Mg–Cs–Cs/ZrO_2	13.32	43.57	30.58	800
370	Na–Eu–W/ZrO_2	13.09	30.38	43.10	750

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 4 catalysts was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

Table 5.7. Development of design hypotheses for Y_2O_3-based catalysts throughout the active learning cycles. This table shows the scores and selected features of the design hypotheses obtained in each individual active learning cycle. Catalyst library. The composition, the best performance, and the corresponding temperature are shown.[a,b,c]

No.	Cat.	C_2 yield (%)	CH_4 conv. (%)	C_2 sel. (%)	Temp.(℃)
371	Fe-Cs-none/Y_2O_3	6.14	37.24	16.48	750
372	V-Mn-Hf/Y_2O_3	6.69	37.90	17.66	800
373	Zn-Sr-W/Y_2O_3	5.15	28.77	17.90	700
374	Mn-Ni-Mo/Y_2O_3	2.01	31.32	6.41	750
375	Li-Fe-Eu/Y_2O_3	4.35	31.87	13.66	750
376	Y-Nd-none/Y_2O_3	4.15	34.03	12.20	800
377	Mn-Y-none/Y_2O_3	7.48	39.91	18.75	800
378	Cs-Ce-none/Y_2O_3	12.44	37.19	33.45	750
379	Mo-Eu-W/Y_2O_3	5.52	39.07	14.12	800
380	Mg-Ni-La/Y_2O_3	15.13	39.86	37.97	750
381	Y-Pd-none/Y_2O_3	n.d.	n.d.	n.d.	n.d.
382	K-Zn-Zr/Y_2O_3	4.32	36.30	11.90	700
383	Sr-La-W/Y_2O_3	5.51	28.78	19.15	750
384	Co-Cu-Hf/Y_2O_3	6.46	39.83	16.22	800
385	Li-Sr-Mo/Y_2O_3	7.57	39.81	19.02	800
386	Mg-Y-Pd/Y_2O_3	n.d.	n.d.	n.d.	n.d.
387	Fe-Ni-Tb/Y_2O_3	3.71	32.09	11.55	700
388	Ti-Mo-none/Y_2O_3	6.49	34.40	18.88	750
389	K-Ca-Mo/Y_2O_3	9.33	36.28	25.71	800
390	Zn-Hf-none/Y_2O_3	5.70	37.17	15.33	750
391	Cu-Sr-Y/Y_2O_3	5.73	38.42	14.93	750
392	K-K-Cs/Y_2O_3	12.31	39.15	31.43	750
393	K-Pd-Cs/Y_2O_3	0.40	47.54	0.84	700
394	Zn-Sr-Ba/Y_2O_3	8.10	37.58	21.56	800
395	Ba-W-none/Y_2O_3	8.33	25.18	33.10	750
396	Ti-Eu-none/Y_2O_3	9.39	35.34	26.58	700
397	K-Mo-none/Y_2O_3	12.19	36.24	33.65	750
398	Na-Ti-Fe/Y_2O_3	15.75	40.50	38.89	750
399	V-Ba-La/Y_2O_3	6.46	38.54	16.77	800
400	Mn-Nd-Nd/Y_2O_3	6.43	36.49	17.61	750
401	Zr-Zr-Hf/Y_2O_3	5.91	37.98	15.55	800
402	Fe-Fe-Zn/Y_2O_3	4.54	37.79	12.02	800
403	V-Ba-Tb/Y_2O_3	7.80	35.89	21.74	750

404	Zr-Pd-Ba/Y_2O_3	1.13	37.48	3.00	700
405	Ca-Mn-Fe/Y_2O_3	6.59	37.33	17.67	800
406	Ni-Sr-La/Y_2O_3	0.48	53.28	0.91	900
407	Nd-Tb-W/Y_2O_3	2.29	30.26	7.58	700
408	Na-Co-Ce/Y_2O_3	6.68	37.60	17.76	750
409	Li-Ca-Cs/Y_2O_3	16.02	41.13	38.95	800
410	Mn-Zn-Zr/Y_2O_3	5.09	34.13	14.91	700
411	Ca-Mo-Ce/Y_2O_3	5.85	41.12	14.22	850
412	K-Hf-Hf/Y_2O_3	8.73	36.70	23.79	750
413	Ni-Ba-Ce/Y_2O_3	1.08	37.97	2.84	700
414	Zn-Ba-Tb/Y_2O_3	14.94	38.19	39.12	700
415	Ca-Cu-Mo/Y_2O_3	14.63	40.97	35.71	700
416	Ni-Pd-La/Y_2O_3	n.d.	n.d.	n.d.	n.d.
417	Na-Ti-Ba/Y_2O_3	6.39	36.72	17.41	700
418	Ni-Nd-none/Y_2O_3	1.32	48.87	2.70	750
419	Cu-Cs-La/Y_2O_3	5.57	37.50	14.86	700
420	Mn-Co-Zn/Y_2O_3	2.34	32.31	7.25	750
421	Y-Y-Tb/Y_2O_3	16.67	35.38	47.12	750
422	Pd-Cs-Nd/Y_2O_3	0.54	46.69	1.15	700
423	Na-Mg-V/Y_2O_3	7.45	39.03	19.09	800
424	Ni-Zr-La/Y_2O_3	n.d.	n.d.	n.d.	n.d.
425	Li-Co-Nd/Y_2O_3	1.83	32.89	5.57	750
426	V-V-Ce/Y_2O_3	10.23	36.44	28.08	750
427	Sr-Y-Cs/Y_2O_3	17.09	40.95	41.74	750
428	Mg-Ni-Cs/Y_2O_3	n.d.	n.d.	n.d.	n.d.
429	Cu-Cs-Ba/Y_2O_3	5.56	37.52	14.82	750
430	K-Ni-Pd/Y_2O_3	n.d.	n.d.	n.d.	n.d.
431	Mo-Cs-none/Y_2O_3	10.59	39.01	27.15	750
432	Pd-Nd-Eu/Y_2O_3	n.d.	n.d.	n.d.	n.d.
433	Fe-Sr-Y/Y_2O_3	6.69	35.39	18.90	750
434	Li-Zn-W/Y_2O_3	5.00	28.96	17.27	750
435	Mn-Co-Cs/Y_2O_3	7.23	39.49	18.30	800
436	Mn-Pd-Pd/Y_2O_3	3.15	34.83	9.05	700
437	K-Ni-La/Y_2O_3	1.26	35.14	3.59	700
438	Li-Y-Mo/Y_2O_3	10.94	38.24	28.60	700

439	Mg-Zr-Mo/Y_2O_3	6.72	40.20	16.72	800
440	Li-Eu-W/Y_2O_3	2.48	31.60	7.84	700
441	Na-V-Mn/Y_2O_3	7.16	37.49	19.10	800
442	Li-Mg-Fe/Y_2O_3	5.53	36.05	15.33	750
443	Fe-Zn-La/Y_2O_3	2.38	34.51	6.90	800
444	K-K-Hf/Y_2O_3	10.34	37.26	27.76	700
445	Cs-Nd-none/Y_2O_3	14.11	39.01	36.18	750
446	Cs-Nd-Nd/Y_2O_3	12.56	39.43	31.85	700
447	Mg-V-Y/Y_2O_3	17.41	41.83	41.62	750
448	Cu-Pd-Tb/Y_2O_3	5.97	14.96	39.87	750
449	V-La-none/Y_2O_3	5.62	35.56	15.80	750
450	Ca-Y-Y/Y_2O_3	15.66	40.15	39.02	750
451	Na-V-W/Y_2O_3	7.56	37.68	20.05	800
452	Li-Ti-Fe/Y_2O_3	15.12	39.97	37.83	750
453	Li-Na-Tb/Y_2O_3	16.85	40.61	41.49	750
454	Li-Mg-Cu/Y_2O_3	3.70	31.69	11.67	750
455	K-Mo-La/Y_2O_3	16.60	32.48	51.10	750
456	Pd-Ce-none/Y_2O_3	0.42	49.51	0.85	700
457	V-Cu-W/Y_2O_3	5.47	31.85	17.19	750
458	K-V-Cs/Y_2O_3	6.72	39.23	17.12	800
459	Na-V-Tb/Y_2O_3	7.33	39.19	18.70	800
460	Sr-La-Eu/Y_2O_3	15.52	41.53	37.37	700
461	Mg-Mn-Zn/Y_2O_3	2.87	37.80	7.59	850
462	Sr-Nd-none/Y_2O_3	8.82	38.88	22.69	800
463	Ce-Tb-W/Y_2O_3	5.32	33.43	15.92	750
464	Mg-Cu-Ba/Y_2O_3	5.66	39.44	14.34	800
465	Mg-Ba-Ba/Y_2O_3	6.38	38.63	16.53	750
466	Co-Y-none/Y_2O_3	6.99	37.43	18.66	750
467	Li-K-none/Y_2O_3	13.87	40.15	34.54	700
468	Ni-Ni-Y/Y_2O_3	2.14	36.85	5.81	700
469	Fe-Pd-Ba/Y_2O_3	5.64	37.37	15.10	800
470	Ni-Ce-Nd/Y_2O_3	n.d.	n.d.	n.d.	n.d.
471	Li-Mg-Co/Y_2O_3	6.43	34.61	18.56	750
472	Ca-Ba-W/Y_2O_3	12.77	37.37	34.18	750
473	Na-Ba-Nd/Y_2O_3	7.55	34.53	21.88	700

371	Cs-Ba-W/Y_2O_3	12.56	29.11	43.15	750
372	K-Cu-La/Y_2O_3	6.63	36.00	18.41	750
373	Ti-V-W/Y_2O_3	13.68	38.69	35.35	700
374	Y-Nd-Nd/Y_2O_3	3.07	37.19	8.26	850
375	K-Mo-Mo/Y_2O_3	18.25	38.11	47.87	800
376	Li-Tb-none/Y_2O_3	5.83	35.00	16.65	700
377	V-Pd-Hf/Y_2O_3	4.42	40.84	10.83	900
378	Mg-Zr-Ba/Y_2O_3	10.59	36.15	29.29	700
379	Mn-Pd-La/Y_2O_3	4.11	36.49	11.26	750
380	Sr-Cs-W/Y_2O_3	14.84	35.91	41.34	750
381	Mg-V-Cu/Y_2O_3	4.06	37.56	10.82	800
382	Na-Sr-Mo/Y_2O_3	9.42	35.24	26.74	700
383	Li-Co-none/Y_2O_3	1.99	34.48	5.77	700
384	Mn-Co-Mo/Y_2O_3	1.42	42.79	3.32	700
385	Cu-Mo-none/Y_2O_3	6.59	38.26	17.21	750
386	Cs-Ba-none/Y_2O_3	14.75	40.02	36.86	700
387	Mn-Fe-Zn/Y_2O_3	3.36	29.62	11.33	750
388	Zn-Ce-Tb/Y_2O_3	5.96	36.86	16.16	750
389	Ca-V-W/Y_2O_3	6.93	37.68	18.40	800
390	Mn-La-Tb/Y_2O_3	5.07	37.91	13.38	800
391	Mg-Mg-Ba/Y_2O_3	15.24	40.13	37.97	750
392	V-Zn-Hf/Y_2O_3	6.36	39.10	16.27	800
393	Li-Mn-Eu/Y_2O_3	6.07	38.39	15.81	800
394	Li-Y-W/Y_2O_3	7.70	32.95	23.36	750
395	Sr-Mo-Pd/Y_2O_3	1.52	38.37	3.96	750
396	Mo-Pd-Nd/Y_2O_3	0.96	39.24	2.43	700
397	Ti-Y-Hf/Y_2O_3	5.50	35.74	15.39	750
398	Na-Na-Co/Y_2O_3	1.86	37.41	4.97	700
399	Na-Ti-Sr/Y_2O_3	6.84	36.25	18.87	750
400	Ni-Pd-none/Y_2O_3	n.d.	n.d.	n.d.	n.d.
401	K-Eu-Tb/Y_2O_3	11.76	38.91	30.23	750
402	Mg-Mn-Co/Y_2O_3	1.43	41.35	3.46	700
403	K-Pd-Ba/Y_2O_3	1.30	39.32	3.30	700
404	Mn-Ni-Pd/Y_2O_3	n.d.	n.d.	n.d.	n.d.
405	Fe-Ni-Nd/Y_2O_3	n.d.	n.d.	n.d.	n.d.

406	Zn-Hf-W/Y_2O_3	6.56	35.33	18.57	750
407	Co-Sr-Ce/Y_2O_3	13.66	39.47	34.62	750

[a] The catalyst composition is expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) are sorted along the atomic number.

[b] The best C_2 yield of individual catalysts is reported together with the corresponding parameters. The performance of 4 catalysts was not determined due to sintering at the calcination temperature. The corresponding cells are filled with n.d.

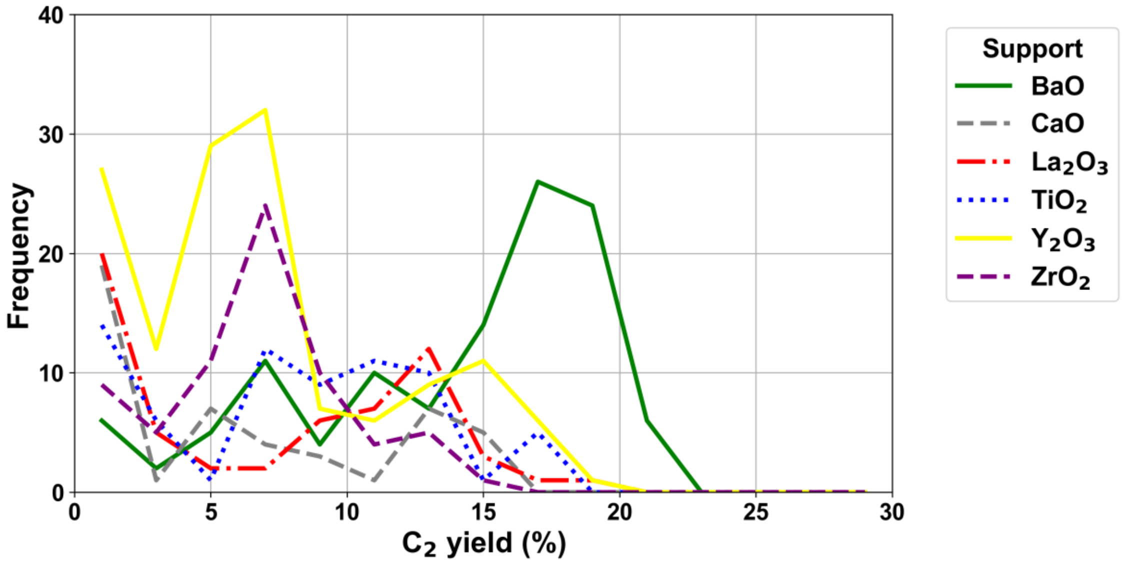


Figure 5.2. This figure presents a comparative histogram of C_2 yield (%) for various support types used in catalytic reactions. Each line represents a different support type, identified by color and line style as indicated in the legend. The x-axis shows the range of C_2 yield (%) values, while the y-axis displays the frequency, or the number of occurrences within each interval. The histogram reveals that Y_2O_3 has a prominent peak

around the 5% yield range, suggesting a higher frequency of occurrences at lower yields. In contrast, BaO shows a broader distribution with a peak around 15-20%, indicating a higher yield range compared to other supports. This comparison highlights the distinct yield profiles associated with each support type, illustrating their varying effectiveness in achieving C_2 yield in catalytic processes.

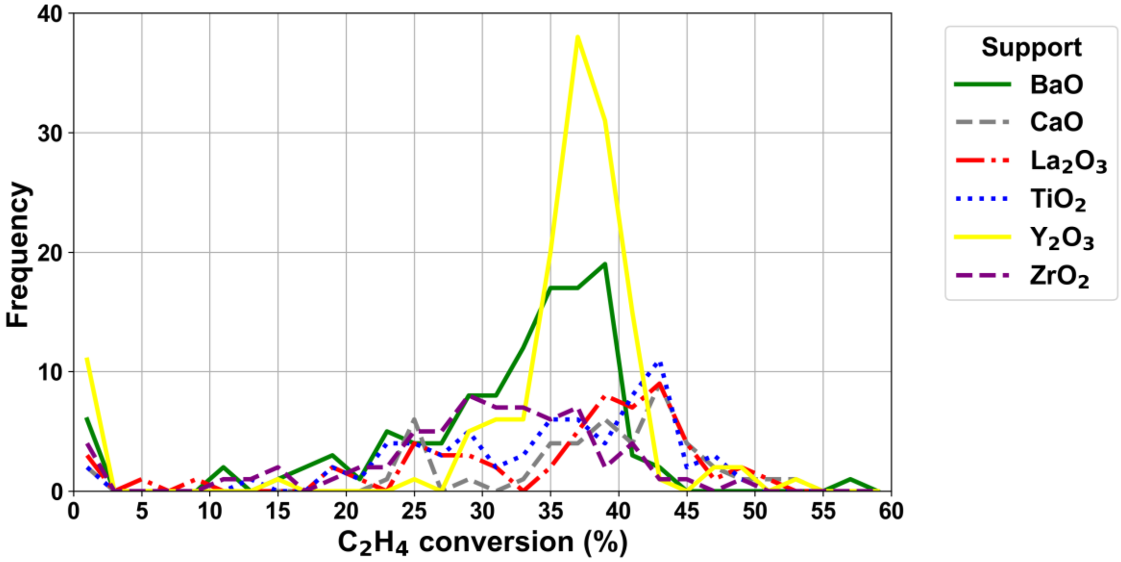


Figure 5.3. This figure presents a comparative histogram of CH_4 conversion (%) for various support used in catalytic reactions. Each line represents a distinct support, identified by color and line style as shown in the legend. The x-axis displays the range of CH_4 conversion (%) values, divided into specific intervals, while the y-axis represents the frequency (number of occurrences) within each interval. The histogram reveals that Y_2O_3 shows a significant peak around the 40-45% conversion range, indicating a higher frequency of CH_4 conversion within this interval. BaO exhibits a broader conversion range, with notable occurrences spanning from 30% to 40%. This comparison

underscores the unique conversion profiles of each support type, highlighting differences in their catalytic effectiveness for CH_4 conversion.

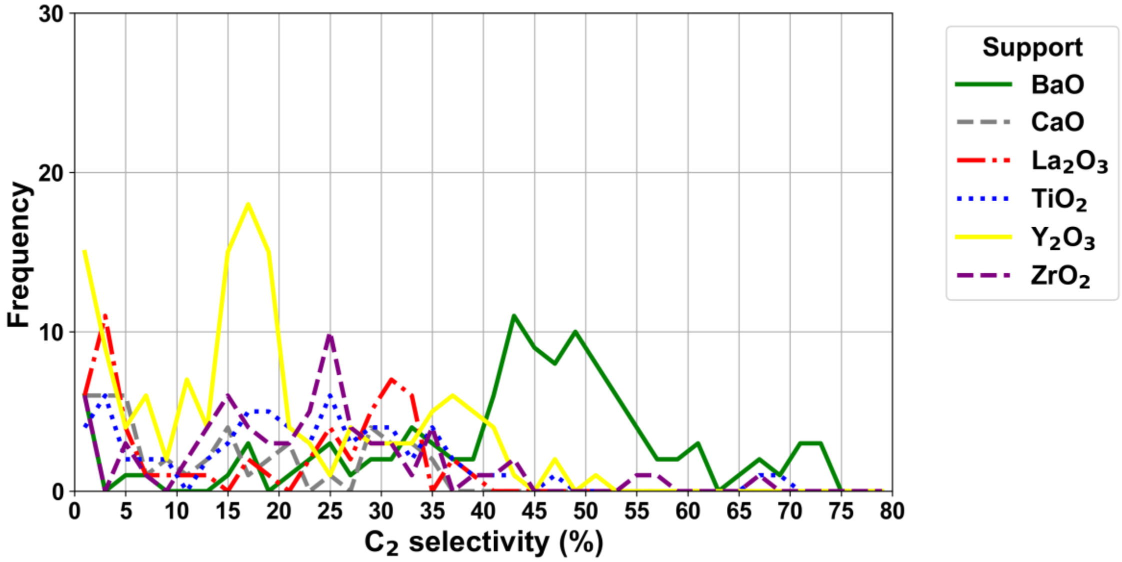


Figure 5.4. This figure presents a comparative histogram of C_2 selectivity (%) for various support types used in catalytic reactions. Each line represents a distinct support type, identified by color and line style as indicated in the legend. The x-axis displays the range of C_2 selectivity (%) values, divided into defined intervals, while the y-axis represents the frequency (number of occurrences) within each interval. The histogram reveals that Y_2O_3 exhibits a prominent peak around 15-20% selectivity, whereas BaO demonstrates a broader selectivity distribution, spanning the range of 35-55%. This comparative view highlights the distinct selectivity profiles associated with each support type.

5.4. CONCLUSION

In this chapter, the evaluation results of 490 catalysts were summarized across six supports: BaO, CaO, La_2O_3, TiO_2, Y_2O_3, and ZrO_2. These catalysts were selected either randomly or through Farthest Point Sampling (FPS) based on feature models generated by Automatic Feature Engineering (AFE). By analyzing the performance distributions and compositions of the top catalysts for each support, several insights into the design and optimization of OCM catalysts were obtained.

The findings highlight the distinct advantages of specific supports in the oxidative coupling of methane (OCM). Among them, BaO demonstrated a balanced performance across C_2 yield, CH_4 conversion, and C_2 selectivity, suggesting its suitability as a benchmark support material. Notably, the high C_2 yield observed in BaO-based catalysts underscores its strong potential for C_2 hydrocarbon production. In contrast, supports such as ZrO_2 exhibited efficient CH_4 conversion but lower C_2 selectivity, indicating that not all supports perform optimally across all metrics. For other supports, the trend was split into two.

The study also revealed that the choice of elements within catalyst compositions significantly affects performance. For instance, catalysts containing Ba and La displayed favorable activity, especially when paired with supports that provided an effective balance of basicity and redox properties. This balance proved essential for enhancing both methane conversion and C_2 product selectivity, with Ba-enhanced catalysts showing particularly promising results in these areas.

The use of high-throughput experimentation and catalyst informatics provided a systematic and data-driven approach to catalyst design, allowing for rapid and efficient exploration of catalyst compositions. This approach enabled the identification of promising design hypotheses for maximizing performance metrics in OCM. The ability to predict and select catalysts based on data-driven models, as demonstrated by the AFE and FPS approach, marks a significant advancement in optimizing OCM catalysts.

REFERENCE

(1) Ortiz-Bravo, C. A.; Chagas, C. A.; Toniolo, F. S. Oxidative Coupling of Methane (OCM): An Overview of the Challenges and Opportunities for Developing New Technologies. *Journal of Natural Gas Science and Engineering* **2021**, *96*, 104254. https://doi.org/10.1016/j.jngse.2021.104254.

(2) Balat, M. Global Trends on Production and Utilization of Natural Gas. *Energy Sources, Part B: Economics, Planning, and Policy* **2009**, *4* (4), 333–346. https://doi.org/10.1080/15567240701621125.

(3) Shaibu, R.; Sambo, C.; Guo, B.; Dudun, A. An Assessment of Methane Gas Production from Natural Gas Hydrates: Challenges, Technology and Market Outlook. *Adv. Geo-Energy Res.* **2021**, *5* (3), 318–332. https://doi.org/10.46690/ager.2021.03.07.

(4) Layritz, L. S.; Dolganova, I.; Finkbeiner, M.; Luderer, G.; Penteado, A. T.; Ueckerdt, F.; Repke, J.-U. The Potential of Direct Steam Cracker Electrification and Carbon Capture & Utilization via Oxidative Coupling of Methane as Decarbonization Strategies for Ethylene Production. *Applied Energy* **2021**, *296*, 117049. https://doi.org/10.1016/j.apenergy.2021.117049.

(5) Zaffran, J.; Yang, B. Theoretical Insights into the Formation Mechanism of Methane, Ethylene and Methanol in Fischer-Tropsch Synthesis at Co $_2$ C Surfaces. *ChemCatChem* **2021**, *13* (11), 2674–2682. https://doi.org/10.1002/cctc.202100216.

(6) Thyssen, V. V.; Vilela, V. B.; De Florio, D. Z.; Ferlauto, A. S.; Fonseca, F. C. Direct Conversion of Methane to C $_2$ Hydrocarbons in Solid-State Membrane Reactors at High Temperatures. *Chem. Rev.* **2022**, *122* (3), 3966–3995. https://doi.org/10.1021/acs.chemrev.1c00447.

(7) Kim, Y. H.; Jun, K.-W.; Joo, H.; Han, C.; Song, I. K. A Simulation Study on Gas-to-Liquid (Natural Gas to Fischer–Tropsch Synthetic Fuel) Process Optimization. *Chemical Engineering Journal* **2009**, *155* (1–2), 427–432. https://doi.org/10.1016/j.cej.2009.08.018.

(8) Damasceno, S.; Trindade, F. J.; Fonseca, F. C.; Florio, D. Z. D.; Ferlauto, A. S. Oxidative Coupling of Methane in Chemical Looping Design. *Fuel Processing Technology* **2022**, *231*, 107255. https://doi.org/10.1016/j.fuproc.2022.107255.

(9) Thybaut, J. W.; Sun, J.; Olivier, L.; Van Veen, A. C.; Mirodatos, C.; Marin, G. B. Catalyst Design Based on Microkinetic Models: Oxidative Coupling of Methane. *Catalysis Today* **2011**, *159* (1), 29–36. https://doi.org/10.1016/j.cattod.2010.09.002.

(10) Murthy, P. R.; Liu, Y.; Wu, G.; Diao, Y.; Shi, C. Oxidative Coupling of Methane: Perspective for High-Value C2 Chemicals. *Crystals* **2021**, *11* (9), 1011. https://doi.org/10.3390/cryst11091011.

(11) Kiani, D.; Sourav, S.; Wachs, I. E.; Baltrusaitis, J. A Combined Computational and Experimental Study of Methane Activation during Oxidative Coupling of Methane (OCM) by Surface Metal Oxide Catalysts. *Chem. Sci.* **2021**, *12* (42), 14143–14158. https://doi.org/10.1039/D1SC02174E.

(12) Yu, Y.; Obata, K.; Movick, W. J.; Yoshida, S.; Palomo, J.; Lundin, S.-T. B.; Urakawa, A.; Sarathy, S. M.; Takanabe, K. High-Pressure Oxidative Coupling of Methane on Alkali Metal Catalyst – Microkinetic Analysis and Operando Thermal Visualization. *Journal of Catalysis* **2024**, *432*, 115414. https://doi.org/10.1016/j.jcat.2024.115414.

(13) Oh, J.; Boucly, A.; Van Bokhoven, J. A.; Artiglia, L.; Cargnello, M. Palladium Catalysts for Methane Oxidation: Old Materials, New Challenges. *Acc. Chem. Res.* **2024**, *57* (1), 23–36. https://doi.org/10.1021/acs.accounts.3c00454.

(14) Kumar, G.; Nikolla, E.; Linic, S.; Medlin, J. W.; Janik, M. J. Multicomponent Catalysts: Limitations and Prospects. *ACS Catal.* **2018**, *8* (4), 3202–3208. https://doi.org/10.1021/acscatal.8b00145.

(15) Szijjártó, G. P.; Pászti, Z.; Sajó, I.; Erdőhelyi, A.; Radnóczi, G.; Tompos, A. Nature of the

Active Sites in Ni/MgAl2O4-Based Catalysts Designed for Steam Reforming of Ethanol. *Journal of Catalysis* **2013**, *305*, 290–306. https://doi.org/10.1016/j.jcat.2013.05.036.

(16) Védrine, J. C. Revisiting Active Sites in Heterogeneous Catalysis: Their Structure and Their Dynamic Behaviour. *Applied Catalysis A: General* **2014**, *474*, 40–50. https://doi.org/10.1016/j.apcata.2013.05.029.

(17) Naagar, M.; Chalia, S.; Thakur, P.; Sridhara, S. N.; Thakur, A.; Sharma, P. B. Nanoferrites Heterogeneous Catalysts for Biodiesel Production from Soybean and Canola Oil: A Review. *Environ Chem Lett* **2021**, *19* (5), 3727–3746. https://doi.org/10.1007/s10311-021-01247-2.

(18) Ciriminna, R.; Pagliaro, M.; Luque, R. Heterogeneous Catalysis under Flow for the 21st Century Fine Chemical Industry. *Green Energy & Environment* **2021**, *6* (2), 161–166. https://doi.org/10.1016/j.gee.2020.09.013.

(19) Hughes, A. E.; Haque, N.; Northey, S. A.; Giddey, S. Platinum Group Metals: A Review of Resources, Production and Usage with a Focus on Catalysts. *Resources* **2021**, *10* (9), 93. https://doi.org/10.3390/resources10090093.

(20) Takahashi, K.; Ohyama, J.; Nishimura, S.; Fujima, J.; Takahashi, L.; Uno, T.; Taniike, T. Catalysts Informatics: Paradigm Shift towards Data-Driven Catalyst Design. *Chem. Commun.* **2023**, *59* (16), 2222–2238. https://doi.org/10.1039/D2CC05938J.

(21) Wulf, C.; Beller, M.; Boenisch, T.; Deutschmann, O.; Hanf, S.; Kockmann, N.; Kraehnert, R.; Oezaslan, M.; Palkovits, S.; Schimmler, S.; Schunk, S. A.; Wagemann, K.; Linke, D. A Unified Research Data Infrastructure for Catalysis Research – Challenges and Concepts. *ChemCatChem* **2021**, *13* (14), 3223–3236. https://doi.org/10.1002/cctc.202001974.

(22) Suvarna, M.; Pérez-Ramírez, J. Embracing Data Science in Catalysis Research. *Nat Catal* **2024**, *7* (6), 624–635. https://doi.org/10.1038/s41929-024-01150-3.

(23) Zhang, D.; Smith, B.; Wu, H.; Nguyen, M.-T.; Rousseau, R.; Glezakou, V.-A. Data Analytics for Catalysis Predictions: Are We Ready Yet? *ACS Catal.* **2024**, *14* (10), 8073–8086. https://doi.org/10.1021/acscatal.3c05285.

(24) Benavides-Hernández, J.; Dumeignil, F. From Characterization to Discovery: Artificial Intelligence, Machine Learning and High-Throughput Experiments for Heterogeneous Catalyst Design. *ACS Catal.* **2024**, *14* (15), 11749–11779. https://doi.org/10.1021/acscatal.3c06293.

(25) Nguyen, T. N.; Nhat, T. T. P.; Takimoto, K.; Thakur, A.; Nishimura, S.; Ohyama, J.; Miyazato, I.; Takahashi, L.; Fujima, J.; Takahashi, K.; Taniike, T. High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. *ACS Catal.* **2020**, *10* (2), 921–932. https://doi.org/10.1021/acscatal.9b04293.

(26) Taniike, T.; Fujiwara, A.; Nakanowatari, S.; García-Escobar, F.; Takahashi, K. Automatic Feature Engineering for Catalyst Design Using Small Data without Prior Knowledge of Target Catalysis. *Commun Chem* **2024**, *7* (1), 11. https://doi.org/10.1038/s42004-023-01086-y.

(27) Nishimura, S.; Ohyama, J.; Li, X.; Miyazato, I.; Taniike, T.; Takahashi, K. Machine Learning-Aided Catalyst Modification in Oxidative Coupling of Methane via Manganese Promoter. *Ind. Eng. Chem. Res.* **2022**, *61* (24), 8462–8469. https://doi.org/10.1021/acs.iecr.1c05079.

(28) Nakanowatari, S.; Nguyen, T. N.; Chikuma, H.; Fujiwara, A.; Seenivasan, K.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Extraction of Catalyst Design Heuristics from Random Catalyst Dataset and Their Utilization in Catalyst Development for Oxidative Coupling of Methane. *ChemCatChem* **2021**, *13* (14), 3262–3269. https://doi.org/10.1002/cctc.202100460.

(29) Nguyen, T. N.; Nakanowatari, S.; Nhat Tran, T. P.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane. *ACS Catal.* **2021**, *11* (3), 1797–1809.

https://doi.org/10.1021/acscatal.0c04629.

(30) Takahashi, L.; Nguyen, T. N.; Nakanowatari, S.; Fujiwara, A.; Taniike, T.; Takahashi, K. Constructing Catalyst Knowledge Networks from Catalyst Big Data in Oxidative Coupling of Methane for Designing Catalysts. *Chem. Sci.* **2021**, *12* (38), 12546–12555. https://doi.org/10.1039/D1SC04390K.

(31) Yang, T. L.; Feng, L. B.; Shen, S. K. Oxygen Species on the Surface of La_2O_3/CaO and Its Role in the Oxidative Coupling of Methane. *Journal of Catalysis* **1994**, *145* (2), 384–389. https://doi.org/10.1006/jcat.1994.1048.

(32) Yoshida, R. XenonPy Is a Python Software for Materials Informatics. **2018**.

(33) Huber, P. J. Robust Estimation of a Location Parameter. *Ann. Math. Statist.* **1964**, *35* (1), 73–101. https://doi.org/10.1214/aoms/1177703732.

Chapter 6

General conclusion

The following is a summary of the work conducted in each chapter.

In **Chapter 2,** we applies an adaptive design loop to BaO-supported catalysts, generating a validated design hypothesis over four cycles. It was also demonstrated that the system has the ability to accurately predict highly efficient catalysts.

In **Chapter 3** this approach is expand to multiple supports (BaO, CaO, La_2O_3, TiO_2, ZrO_2), revealing distinct design patterns, with single-element dominance (e.g., La on BaO) and multi-element combinations (e.g., alkaline earth metals with Cs on CaO), highlighting unique design rules for each support.

In **Chapter 4**, active learning efficiency was explored to achieve high-accuracy learning with smaller datasets. It was tested whether design hypotheses from the five supports in Chapter 3 could guide predictions for Y_2O_3 supports. This approach confirmed the transferability of design insights across supports.

Chapter 5 summarizes the performance of the catalysts evaluated in this thesis.

I believe that the research conducted in this thesis highlights the advantages of applying catalyst informatics based on the elemental substitution strategy for catalyst development. This study provides a novel direction in catalyst research and development, with the potential to transform approaches. The proposed flamework has broad applicability across various catalyst systems. Further advancement of this research is expected to enhance the discovery of optimal compounds, enabling their rapid identification for a wide range of reactions within a short timeframe. To further realize this potential, it will be essential to develop methods for extracting information from a broader range of reactions and optimizing the transfer of relevant information between systems.

List of Publications and Other Achievements

Aya Fujiwara

A) PUBLICATION

1. Acquiring and Transferring Comprehensive Catalyst Knowledge through Integrated High-Throughput Experimentation and Automatic Feature Engineerin, **Aya Fujiwara**, Sunao Nakanowatari, Youhei Cho, Toshiaki Taniike, *Science and Technology of Advanced Material* **2025**, in press (DOI: https://doi.org/10.1080/14686996.2025.2454219).

2. Extraction of catalyst design heuristics from random catalyst dataset and their utilization in catalyst development for oxidative coupling of methane, Sunao Nakanowatari, Thanh Nhat Nguyen, Hiroki Chikuma, **Aya Fujiwara**, Kalaivani Seenivasan, Ashutosh Thakur, Lauren Takahashi, Keisuke Takahashi, Toshiaki Taniike, *ChemCatChem* **2021**, *13*, 3262–3269.

3. Catalysis Gene Expression Profiling: Sequencing and Designing Catalysts, Keisuke Takahashi, Jun Fujima, Itsuki Miyazato, Sunao Nakanowatari, **Aya Fujiwara**, Thanh Nhat Nguyen, Toshiaki Taniike, Lauren Takahashi, *The Journal of Physical Chemistry Letters* **2021**, *12*, 7335–7341.

4. Constructing Catalyst Knowledge Networks from Catalysts Big Data in Oxidative Coupling for Methane for Designing Catalysts, Lauren Takahashi, Thanh Nhat Nguyen, Sunao Nakanowatari, **Aya Fujiwara**, Toshiaki Taniike, Keisuke Takahashi, *Chemical Science* **2021**, *12*, 12546–12555.

5. Designing Catalysts Descriptors for Machine Learning in Oxidative Coupling of Methane, Sora Ishioka, **Aya Fujiwara**, Sunao Nakanowatari, Toshiaki Taniike, Keisuke Takahashi, *ACS Catalysis* **2022**, *12*, 11541–11546.

6. Automatic feature engineering for catalyst design using small data without prior knowledge of target catalysis, Toshiaki Taniike, **Aya Fujiwara**, Sunao Nakanowatari, Fernando García-Escobar, Keisuke Takahashi, *Commun. Chem.* 2024, 7 (11).

B) INTERNATIONAL CONFERENCE

1. Acquisition of catalyst design rules through feature engineering, **Aya Fujiwara**, Sunao Nakanowatari, Toshiaki Taniike, International Congress on Pure & Applied Chemistry (ICPAC) Mongolia 2024, Ulaanbaatar, Mongolia, Aug. 28–Sep. 1, 2024, poster.

C) DOMESTIC CONFERENCE

1. メタンの酸化的カップリングにおける適応的触媒設計, **藤原 綾**, 中野渡 淳, 谷池 俊明, 2022年度日本化学会近畿支部 北陸地区講演会と研究発表会, 富山大学五福キャンパス, 2022年11月11日, ポスター.

Supporting information

Table S1. XenonPy features used and their meanings.[1]

feature	description
period	Period in the periodic table
atomic_number	Number of protons found in the nucleus of an atom
mendeleev_number	Atom number in mendeleev's periodic table
atomic_radius	Atomic radius
atomic_radius_rahm	Atomic radius by Rahm et al
atomic_volume	Atomic volume
atomic_weight	The mass of an atom
icsd_volume	Atom volume in ICSD database
lattice_constant	Physical dimension of unit cells in a crystal lattice
vdw_radius	Van der Waals radius
vdw_radius_alvarez	Van der Waals radius according to Alvarez
vdw_radius_batsanov	Van der Waals radius according to Batsanov
vdw_radius_bondi	Van der Waals radius according to Bondi
vdw_radius_dreiding	Van der Waals radius from the DREIDING FF
vdw_radius_mm3	Van der Waals radius from the MM3 FF
vdw_radius_rt	Van der Waals radius according to Rowland and Taylor
vdw_radius_truhlar	Van der Waals radius according to Truhlar
vdw_radius_uff	Van der Waals radius from the UFF
covalent_radius_bragg	Covalent radius by Bragg
covalent_radius_cordero	Covalent radius by Cerdero et al
covalent_radius_pyykko	Single bond covalent radius by Pyykko et al
covalent_radius_pyykko_double	Double bond covalent radius by Pyykko et al
covalent_radius_pyykko_triple	Triple bond covalent radius by Pyykko et al
covalent_radius_slater	Covalent radius by Slater
c6	C_6 dispersion coefficient in a.u
c6_gb	C_6 dispersion coefficient in a.u
density	Density at 295K
proton_affinity	Proton affinity
dipole_polarizability	Dipole polarizability
electron_affinity	Electron affinity
electron_negativity	Tendency of an atom to attract a shared pair of electrons
en_allen	Allen's scale of electronegativity
en_ghosh	Ghosh's scale of electronegativity
en_pauling	Pauling's scale of electronegativity
gs_bandgap	DFT bandgap energy of T=0K ground state
gs_energy	DFT energy per atom (raw VASP value) of T=0K ground state
gs_est_bcc_latcnt	Estimated BCC lattice parameter based on the DFT volume
gs_est_fcc_latcnt	Estimated FCC lattice parameter based on the DFT volume
gs_mag_moment	DFT magnetic momenet of T=0K ground state
gs_volume_per	DFT volume per atom of T=0K ground state
hhi_p	Herfindahl−Hirschman Index (HHI) production values

hhi_r	Herfindahl−Hirschman Index (HHI) reserves values
specific_heat	Specific heat at 20°C
gas_basicity	Gas basicity
first_ion_en	First ionisation energy
fusion_enthalpy	Fusion heat
heat_of_formation	Heat of formation
heat_capacity_mass	Mass specific heat capacity
heat_capacity_molar	Molar specific heat capacity
evaporation_heat	Evaporation heat
linear_expansion_coefficient	Coefficient of linear expansion
boiling_point	Boiling temperature
brinell_hardness	Brinell Hardness Number
bulk_modulus	Bulk modulus
melting_point	Melting point
metallic_radius	Single-bond metallic radius
metallic_radius_c12	Metallic radius with 12 nearest neighbors
thermal_conductivity	Thermal conductivity at 25 °C
sound_velocity	Speed of sound
vickers_hardness	Value of Vickers hardness test
Polarizability	Ability to form instantaneous dipoles
youngs_modulus	Young’s modulus
poissons_ratio	Poisson’s ratio
molar_volume	Molar volume
num_unfilled	Total unfilled electron
num_valance	Total valance electron
num_d_unfilled	Unfilled electron in d shell
num_d_valence	Valance electron in d shell
num_f_unfilled	Unfilled electron in f shell
num_f_valence	Valance electron in f shell
num_p_unfilled	Unfilled electron in p shell
num_p_valence	Valance electron in p shell
num_s_unfilled	Unfilled electron in s shell
num_s_valence	Valance electron in s shell

Reference

(1) Yoshida, R. XenonPy Is a Python Software for Materials Informatics. **2018**.