JAIST Repository
https://dspace.jaist.ac.jp/

Title	元素置換戦略に基づく系統的触媒開発のための触媒インフォマティクス
Author(s)	中野渡, 淳
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19073
Rights	
Description	Superviser: 谷池 俊明, 先端科学技術研究科, 博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

博士論文

Catalyst Informatics Based on Element Substitution Strategies for Systematic Catalyst Development

中野渡　淳

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６年３月

Referee-in-chief: Professor Toshiaki Taniike
Japan Advanced Institute of Science and Technology

Referees: Professor Hieu-Chi Dam
Japan Advanced Institute of Science and Technology

Associate Professor Shun Nishimura
Japan Advanced Institute of Science and Technology

Associate Professor Jumpei Ueda
Japan Advanced Institute of Science and Technology

Professor Keisuke Takahashi
Hokkaido University

Catalyst Informatics Based on Element Substitution Strategies for Systematic Catalyst Development

Sunao Nakanowatari

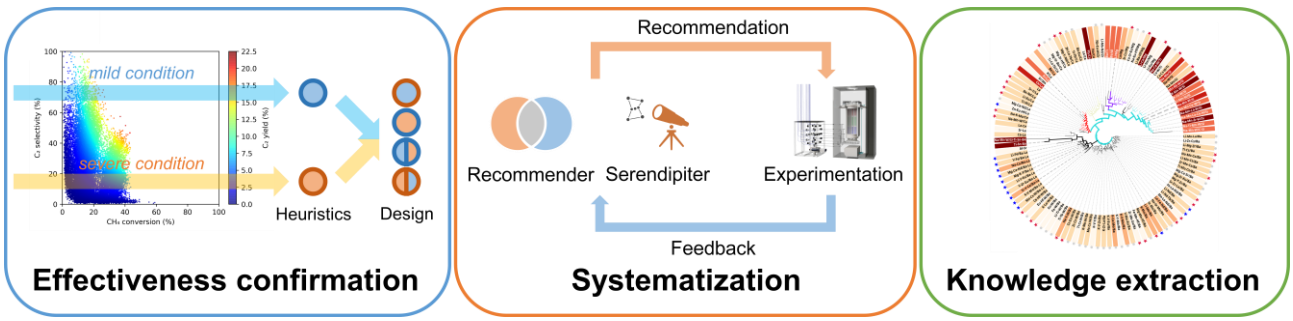
2120031

Solid catalysts are crucial in industry for producing a diverse range of chemical compounds. However, their working mechanisms are complex, influenced by many variables across different spatial and temporal scales. This has prevented rational catalyst design based on researchers' understanding; historically, most practical catalysts have been developed through empirical trials and errors, often accompanied by serendipitous discoveries. However, the traditional trial-and-error method of catalyst development is essentially ad hoc and cost intensive. In this thesis, I attempted to systematize catalyst discoveries based on a strategy of elemental substitution in a data-driven manner, aiming at establishing a reproducible and generalizable method for discovering high-performance catalysts. In doing so, a catalyst dataset featured with both high quality and high quantity was essential. A dataset for oxidative coupling of methane, previously acquired by us through high-throughput experimentation, was adopted as a training dataset [1]. The same experimental system was also used to validate concepts for individual chapters.

In **Chapter 2**, combinatorial rules for superior methane conversion under mild conditions and those for suppressing deep oxidation under severe conditions were extracted from the training data. These rules were successfully combined to design novel OCM catalysts. Key findings include temperature-dependent catalyst design guidelines, the importance of support properties, and the effectiveness of mixed supports like La_2O_3 and BaO. This chapter validated the effectiveness of catalyst design based on combinations of catalyst components.

In **Chapter 3**, to systematize catalyst discovery based on component combinations, an adaptive sampling loop has been designed with implementing a catalyst recommender system which can estimate catalyst performance through the substitution of component combinations. The system is based on the evidence theory, a mathematical framework to quantify the certainty of a hypothesis by combining multiple pieces of evidence and capable of managing uncertainties. It recommends catalysts according to the degree of belief that the catalyst is good or uncertain, which is equal to controlling the exploration/exploitation trade-off. The adaptive sampling led to the discovery of various high-performing catalysts. Moreover, by finding discernible patterns within serendipitous catalysts (unexpectedly high-performing catalysts), a catalyst serendipiter system was developed. The serendipiter system was designed to predict the occurrence of serendipity based on catalyst compositions and prediction outcomes from various classifiers, working in tandem with the recommender system. The system was exploited to induce serendipitous catalysts.

In **Chapter 4**, a method to visually extract the combinatorial knowledge for catalyst design from a set of data with different backgrounds was introduced based on a phylogenetic tree. The developed method was not only able to visualize the history of catalyst development in OCM, but also clarified a significant contribution of my research (Chapters 2 and 3) to widen the scope of high-performing catalysts. Moreover, these visualizations highlighted that each catalyst system follows different rules of elemental combinations to achieve good performance. The insights gained from this analysis were exploited to design promising mixed oxide-based catalysts, a category that has been relatively underexplored in the history of OCM catalyst development.



In this thesis, systematic and efficient methods for data-driven catalyst discovery based on combinatorial rules have been developed. Due to the un-necessity of specific descriptors, the methods are generally applicable to other catalysis, and any cases where materials are developed from combinations of components.

[1] T. N. Nguyen, S. Nakanowatari, T. P. N. Tran, A. Thakur, L. Takahashi, K. Takahashi, T. Taniike. *ACS Catal.*, **2021**, 11, 3, 1797–1809.

Keywords: Catalyst informatics, Catalyst discovery, Combinatorial rule, Substitution of component combination, Oxidative coupling of methane

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 “Catalyst Informatics Based on Element Substitution Strategies for Systematic Catalyst Development” under the supervision of Prof. Toshiaki Taniike during April 2021– March 2024 at Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology.

Chapter 1 describes a general introduction and the purpose of this thesis. **Chapter 2** focuses on the selection and combination of supported elements and supports to enhance catalyst selectivity at high temperatures and activity at low temperatures. In **Chapter 3**, a catalyst recommender system which can balance exploration/exploitation trade-off using evidence theory was developed for estimating the performance of catalysts through elemental substitution and a serendipiter system for unexpected discoveries, both working together to enhance catalyst development by combining performance prediction and serendipitous findings. In **Chapter 4**, I developed a method to visualize historical catalyst data using a phylogenetic tree to discern the history of catalyst development, highlighting underexplored combinations and aiding the design of new catalysts through horizontal propagation strategies. **Chapter 5** describes the summary and general conclusion of this thesis.

The work is original and no part of this thesis has been plagiarized.

Sunao Nakanowatari

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

November 2023

Acknowledgements

First of all, I would like to express the deepest sense of gratitude to my supervisor Assoc. Prof. Toshiaki Taniike, Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology for his continuous support, and enlightening suggestions throughout my Ph.D course. I am thankful to him for his never ending patience, motivation and immense knowledge.

I would also take an opportunity to thank Senior Lecturer Patchanee Chammingkwan, Asst. Prof. Toru Wada, Dr. Thanh Nhat Nguyen for their valuable inputs, cooperation and stimulating discussions.

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

I would like to appreciate to Prof. Hieu-Chi Dam, Assoc. Prof. Shun Nishimura, Assoc. Prof. Jumpei Ueda, and Prof. Keisuke Takahashi to revise this thesis.

I also would like to thank my second supervisor, Prof. Kazuaki Matsumura, and my advisor for minor research, Prof. Shintaro Ishiwata and Assoc. Prof. Shun Nishimura for the time they spent for me.

I also appreciate the Japan Society for the Promotion of Science (JSPS) for financial support in carrying out this research. (JSPS KAKENHI Grant Number JP22KJ1482)

I would like to thank my parents, grandmothers, and sister for always encouraging me and supporting my ambitions.

Table of contents

Chapter 1 ..... **5**

General introduction **5**

1.1. Development of solid catalysts 6

1.2. Oxidative coupling of methane 8

1.2.1. Fundamentals 8

1.2.2. Catalysts for OCM 11

1.3. Machine learning 13

1.3.1. Fundamentals 13

1.3.2. Evidence theory 14

1.3.3. Logistic regression 14

1.3.4. Support vector machine 15

1.3.5. Random forest 15

1.3.6. Naïve Bayes 16

1.4. High-throughput experimentation 16

1.5. Catalysts informatics 18

1.5.1. Fundamentals 18

1.5.2. Random catalyst dataset for OCM 19

1.5.3. Catalyst informatics based on this dataset 22

1.6. Adaptive sampling 23

1.7. Objective 25

Chapter 2 ..... **37**

Extraction of catalyst design heuristics **37**

from random catalyst dataset and their utilization **37**

in catalyst development for oxidative coupling of methane **37**

2.1. Introduction 39

2.2. Method 41

2.2.1. Catalyst dataset 41

2.2.2. Validation experiments 42

2.3. Results 44

2.3.1. Data analysis for mild conditions and severe conditions 44

2.3.2. Catalyst development based on extracted heuristics 48

2.4. Conclusions 63

Chapter 3 ..... **69**

Adaptive sampling for catalyst discovery **69**

incorporating artificial serendipity **69**
3.1. Introduction 71
3.2. Methods 74
3.2.1. Catalyst preparation and evaluation 74
3.2.2. Prediction accuracy metric calculation 76
3.2.3. Generating phylogenetic tree using the neighbor-joining method 77
3.2.4. Implementation of a catalyst serendipiter system 77
3.3. Results 79
3.4. Conclusion 98
Chapter 4 ..... **105**
Visualization for Evolutionary History of Catalysis **105**
4.1. Introduction 107
4.2. Methods 109
4.2.1. Data pretreatment 109
4.2.2. Distance definition 110
4.2.3. Construction of phylogenetic trees 112
4.2.4. Catalyst preparation and evaluation 113
4.2.5. Synthesis of Ti-Sn-Ba Support 114
4.3. Results 115
4.3.1. Visualizing history using phylogenetic trees 115
4.3.2. Phylogenetic tree reconstructed with the addition of our catalysts 119
4.3.3. Catalyst discoveries based on phylogenetic tree 124
4.4. Conclusions 129
Chapter 5 ..... **137**
General conclusion **137**

Chapter 1

General introduction

1.1. Development of solid catalysts

Solid catalysts are substances that control rates of chemical reactions without their being explicitly consumed. Their ability to efficiently transform large amounts of compounds with minimal quantities makes them vital to modern industry[1]. In many catalytic reactions, multiple products are formed. Usually, the most thermodynamically stable product is favored when the reaction is hastened. However, there are instances where the desired product is not the most stable one[2,3]. In these cases, catalysts are crucial for quickly activating the starting material while also slowing down the formation of the most stable product, allowing the less stable product to form[1].

Developing solid catalysts involves numerous challenges. These catalysts work through complex interactions between solid surfaces and reactants, making it difficult to fully understand and predict their behavior[4–7]. The surface properties, defects, and morphology of the catalyst have a significant impact on the efficiency of the reaction, requiring precise design for optimal performance[8,9]. Choosing the right materials and creating them is another major obstacle. With a wide range of potential materials like metals, metal oxides, zeolites, and metal-organic frameworks, finding the perfect match for specific requirements is challenging. Solid catalysts are further complicated by combinatorial or synergistic effects between their components[10–12]. The impact of combining multiple elements on catalysts is varied and largely unpredictable. This includes instances where specific structures are formed through the combination of certain elements, or cases where certain elements play a key role in stabilizing specific structures. Additionally, solid catalysts can undergo surface changes over time, which can degrade their performance. This degradation can be due to factors like carbide formation, contamination by harmful substances, or damage from thermal cycles. This dynamic property is a major obstacle to understanding catalysis[13]. Another key

challenge is enhancing the selectivity of reactions. In cases where multiple products come from a single starting material, it is crucial to precisely control the reaction pathways to obtain the desired product. This requires sophisticated technology to accurately adjust the catalyst's surface chemistry. Furthermore, optimizing the conditions of the reaction, such as temperature, pressure, and the concentration of reactants, is essential for activating the catalyst. Finding these optimal conditions often requires extensive experimentation and time[14,15].

Catalyst development can be divided into two main approaches: the knowledge-based approach and the Edisonian approach. The knowledge-based approach aims to deeply understand catalysis and design catalysts based on the understanding[16–18]. It involves extensive studies of simplified model catalysts through in situ and kinetic analyses, and the obtained insights are used to design catalysts for specific performance targets. However, this approach has limitations, as insights from simpler systems may not always apply accurately to more complex ones. The complexity in real-world catalytic systems, often intensified by the synergistic effect and other unexpected things, can significantly diverge from these simpler models[10–12]. On the other hand, the Edisonian approach relies on trials and errors, focusing on practical optimization by testing different compositions and combinations of elements. This method is widely used in the industry today. Its advantage is that complex solid catalysts can be processed in their complexity. However, unlike the knowledge-based approach, the insights gained during the catalyst development process are not readily shared among researchers. The accumulation of such tacit knowledge is leading to a lack of collective knowledge building in the field. As a result, the Edisonian approach does not establish a foundation for future development with leaving no unified framework for catalyst

design. Therefore, there is an increasing need for a new catalyst development approach that combines the advantages of both the knowledge-based and Edisonian approaches.

1.2. Oxidative coupling of methane

1.2.1. Fundamentals

The shale gas revolution has substantially increased global natural gas reserves[19]. Methane, the main component of natural gas, is a crucial source for producing fuels and chemicals[20]. However, these natural gas reserves are often found in remote areas, making methane transportation economically unfeasible. As a result, methane that cannot be converted on-site into more easily transportable hydrocarbons and chemicals is frequently burned off as the fuel. This process releases greenhouse gases, primarily CO and CO_2, into the atmosphere[2]. Therefore, there is an urgent need to develop efficient technologies for upgrading methane. These technologies would enable the economical use of natural gas and help make the petroleum industry more environmentally friendly. This necessity has reignited interest in exploring catalytic methods and new reactor designs for methane upgrading.

Industrial conversion of methane usually follows an indirect route, where methane is first converted into syngas which is a mixture of carbon monoxide and hydrogen[2]. This process begins with methane reforming into syngas, which is subsequently used to produce olefins, petrol, diesel, or oxygenates through Fischer-Tropsch synthesis. However, this indirect route has several drawbacks. Firstly, methane reforming is highly energy-intensive, requiring high temperatures and pressures (900-1200 K; 15-40 atm)[21]. Additionally, the hydrogen needed for this process is often supplied from naphtha cracking, which releases a significant amount of CO_2. There are

also operational challenges, such as the formation of carbonaceous deposits and sintering, that lead to the deactivation of catalysts[21].

Direct methods for converting methane present more cost-effective and simpler alternatives compared to the indirect syngas route[22]. These involves non-oxidative processes such as methane to olefins, aromatics, and hydrogen (MTOAH), and methane dehydrogenation (MDA), and oxidative processes such as oxidative coupling of methane (OCM). In particular, OCM directly converts methane into ethane and ethylene, the latter being an essential precursor for a variety of chemicals and polymers[23,24]. Although the potential of the OCM process has been recognized since its discovery in 1982 as a way to supply raw materials to the chemical industry and reduce dependence on crude oil, its industrial use is still limited. This limitation is due to challenges such as low selectivity and the formation of unwanted deep oxidation byproducts (CO_x), which are thermodynamically more favored than C_2 hydrocarbons[25,26]. Possible reactions that can occur during the OCM process are outlined in Table 1.1[27].

Table 1.1. Possible reactions that can occur during the oxidative coupling of methane process [27].

No.	Reaction	ΔH_{298}^{Θ} (kJ·mol^{-1})
1	$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$	−802
2	$2CH_4 + 1/2O_2 \rightarrow C_2H_6 + H_2O$	−177
3	$CH_4 + O_2 \rightarrow CO + H_2 + H_2O$	−278
4	$CH_4 + 1/2O_2 \rightarrow CO + 2H_2$	−36
5	$2CH_4 + O_2 \rightarrow C_2H_4 + 2H_2O$	−282
6	$C_2H_6 + 1/2O_2 \rightarrow C_2H_4 + H_2O$	−105
7	$C_2H_6 \rightarrow C_2H_4 + H_2$	137
8	$C_2H_4 + 2O_2 \rightarrow 2CO + 2H_2O$	−757
9	$C_2H_4 + 2H_2O \rightarrow 2CO + 4H_2$	210
10	$CO_2 + H_2 \rightleftharpoons CO + H_2O$	41

Among the potential reactions, reactions (2) and (5) are the targeted, desired reactions. These reactions are characterized by negative ΔH values, indicating that they are exothermic and can proceed easily under mild reaction conditions. However, there is a challenge posed by competing reactions, specifically reactions (1) and (8). The ΔH values for these competing reactions are three to four times more negative than those for reactions (2) and (5), making them more thermodynamically favorable, thus leading to the formation of deep oxidation products like CO and CO_2. This tendency towards the formation of CO_x reduces the selectivity for the desired C_2 hydrocarbons in the process.

The initial step in the relevant reactions, methane activation, involves breaking C-H bonds to remove one hydrogen atom. The methane molecule is highly stable due to its strong (with a first bond dissociation energy of 439.3 kJ/mol) and weakly polarized (2.84×10^{-40} C^2 m^2 J^{-1}) C-H bonds[28]. This stability is attributed to the molecule's geometry, where the central carbon atom is tetrahedrally coordinated to four hydrogen atoms with the C-H bond length of 1.090 Å, and the H-C-H angle of

109.47 °[29]. With this symmetry methane has no dipole moment, making it resistant to nucleophilic and electrophilic attacks [29].

This high stability of methane's C-H bond, combined with the energetic favorability of forming CO_x compounds, is a significant barrier to adopting the OCM process in industry. Techno-economic analyses indicate that for an OCM process targeting C_2 hydrocarbons to be economically viable, it would need a single-pass yield of 25-30%[30]. Despite extensive research since its discovery in 1982, this techno-economic target is not achieved[22].

1.2.2. Catalysts for OCM

OCM requires high temperatures ranging from 973 to 1123 K and oxidizing agents to break the strong C-H bonds in methane. However, CO_x compounds are thermodynamically favored and form more rapidly than ethylene and other C_2 hydrocarbons. Therefore, the principal challenge in developing OCM catalysts lies in effectively activating methane while minimizing the subsequent combustion reaction that produces CO_x. There are several well-studied OCM catalysts in the 40-year history of OCM. They are briefly discussed hereafter.

1.2.2.1.Alkaline earth oxide-based catalysts

In the early 1990s, alkaline earth metal oxides were evaluated as potential catalysts for the OCM reaction. It was discovered that the C_{2+} selectivity of these oxides increased with their basicity, in the order of BeO < MgO < CaO < SrO. Among them, MgO-based catalysts, particularly Li/MgO (MgO doped with Li), have received great attention. Catalysts with 3 wt.% Li added to MgO exhibited an initial methane conversion and C_{2+} selectivity of 30% and 62%, respectively, at 780 °C[31], as compared to 23% conversion and 43% selectivity achieved by pure MgO at the same

temperature. However, Li/MgO catalysts exhibited a drawback in maintaining catalytic stability during the reaction, as Li atoms tended to sublimate under the operational conditions[32,33].

1.2.2.2.Rare earth oxide-based catalysts

La_2O_3-based catalysts demonstrate superior catalytic activity and stability in OCM reactions at low temperatures[34–37]. Recent studies involving doped and undoped lanthanide catalysts have indicated that Sr-doped La2O3 is the most promising rare earth oxide-based catalyst for OCM[7,38,39]. The catalytic performance of rare-earth oxide-based catalysts is linked to the strength of the surface basicity, similar to that observed with alkaline earth oxide-based catalysts, where lanthanum exhibits the highest basicity[40].

1.2.2.3.Mn-Na_2WO_4/SiO_2 catalyst

Mn-Na_2WO_4/SiO_2 is known for its high C_2 production and high thermal stability, being regarded as a kind of benchmarking catalysts in the community. The structure of this catalyst remains a subject of debate due to the co-presence of Mn_2O_3, $MnWO_4$, Na_2WO_4, MnO_2, and various Na-W oxides[27,41]. A variety of species such as distorted WO_4^{2-}, α-cristobalite SiO_2-stabilized WO_4^{2-}, and so on were previously posited as a potential active site of this catalyst. But these proposals were based on characterization conducted at room temperature[42–45]. Recent advancements in in-situ and operando techniques have revealed that most of structures previously believed actually melt at the high temperatures of the reaction, allowing only the presence of a β-cristobalite structure [46]. What remains clear is the catalyst's synergistic effect: the removal of any component significantly reduces its performance[41,47]. This observation underscores the complexity and fascination of solid-state catalysis. The

$Mn\text{-}Na_2WO_4/SiO_2$ catalyst, with its intricate structure and performance characteristics, continues to be a subject of intense interest and study in the field of OCM.

1.3. Machine learning

1.3.1. Fundamentals

Machine learning (ML) is a field of artificial intelligence that focuses on creating systems that can learn from and make decisions based on data. It includes various types of learning, such as supervised, unsupervised, and reinforcement learning, each suitable for different kinds of tasks. ML is used in numerous applications like speech recognition, medical diagnosis, and self-driving cars, employing algorithms like decision trees and neural networks. Assessing the accuracy and reliability of the obtained models is crucial, where accuracy validation methods come into play.

Accuracy validation methods in ML evaluate a model's performance, ensuring its reliability and effectiveness on unseen data. Popular methods include cross-validation, where the data is divided into subsets to train and test the model multiple times; train/test split, which separates the data into training and testing sets; and the use of a confusion matrix to understand classification performance. Metrics like precision, recall, and the receiver operating characteristic (ROC) curve help in assessing aspects beyond mere accuracy, especially in contexts where data imbalance might skew simple accuracy metrics. These methods are essential in guiding the development of robust ML models, addressing challenges like overfitting and ensuring that models perform well across diverse scenarios. Here, several machine learning methods and algorithms are explained.

1.3.2. Evidence theory

Dempster-Shafer Theory, also known as the Theory of Evidence, is a mathematical framework for modeling uncertainty and combining evidence from different sources[48]. It has evolved around the concept of Basic Belief Assignments (BBAs), which allocate belief mass to subsets of a frame of discernment (the set of all possible outcomes). The theory distinguishes between belief and plausibility functions: belief measures the minimum belief committed to a set, while plausibility reflects the maximum possible belief given current knowledge. A key feature of this theory is Dempster's Rule of Combination, allowing the integration of independent pieces of evidence to update beliefs. Unlike Bayesian probability, Dempster-Shafer theory does not require prior probabilities for all hypotheses and allows for expressing uncertainty and ambiguity by assigning belief to sets of possibilities[49]. This makes it particularly useful in fields like artificial intelligence, decision making, and statistics, especially for decision support systems and applications involving incomplete or uncertain information.

1.3.3. Logistic regression

Logistic regression is a statistical method used primarily for binary classification tasks. It predicts the probability of an outcome that is categorically dichotomous – such as 'yes' or 'no', 'spam' or 'not spam'. The core of logistic regression is the logistic or sigmoid function, which maps any real-valued number into a value between 0 and 1, representing the probability of the dependent variable belonging to a certain class. This model computes the log odds of the probability as a linear combination of the independent variables and is particularly noted for its ease of interpretation and efficiency in classification tasks. While logistic regression is simple and fast, making it widely used in fields like medicine, finance, and marketing, it

assumes a linear response for the explanatory variables, which can be a limitation in complex scenarios where this linearity does not hold.

1.3.4. Support vector machine

Support vector machine (SVM) is a versatile supervised machine learning algorithm mainly used for classification and regression tasks. It excels in high-dimensional spaces, identifying a hyperplane that maximally separates classes by maximizing the margin, the distance between the hyperplane and the nearest data points (support vectors) from each class. SVM is effective for both linear classification and, using the kernel trick, nonlinear classification, where it transforms data into a higher dimension for separability. This algorithm balances margin maximization and classification error through a regularization parameter, allowing for some misclassifications (soft margin). While SVMs are powerful in handling complex datasets and are memory efficient, they require careful parameter tuning and can be computationally demanding for very large datasets. They are widely used in applications like image and text classification, bioinformatics, and more.

1.3.5. Random forest

Random forest is a prime example of ensemble learning, a machine learning technique that combines the predictions of multiple models to achieve greater accuracy and robustness than separate models. In the case of random forest, it constructs a "forest" of decision trees, each trained on random subsets of the data. Ensemble learning in random forest is realized through a method called bootstrap aggregating or bagging, where each decision tree is built on a different subset of the data and considers a random subset of features at each split. This approach ensures diversity among the trees, enhancing the model's overall performance and reducing the risk of overfitting, a common problem in individual decision trees. Random forest excels in both

classification and regression tasks, handling large datasets with high dimensionality efficiently. It also provides insights into feature importance but can be computationally demanding and less interpretable due to its complexity. Ensemble learning, as demonstrated by random forest, is particularly effective in situations where a single model is insufficient to capture complex patterns in data, making it a widely used technique in various applications.

1.3.6. Naïve Bayes

Naive Bayes is a straightforward and efficient probabilistic machine learning algorithm based on Bayes' Theorem, with the key assumption that features in a dataset are independent of each other given the class label. This "naive" assumption of feature independence simplifies calculations, enabling the algorithm to perform well even when this condition is not strictly met. It calculates the conditional probability of each class given each feature and is commonly used in applications like spam filtering, text classification, and sentiment analysis. There are different variants of Naive Bayes, such as Gaussian, multinomial, and Bernoulli, each tailored to different data distributions. Despite its simplicity, Naive Bayes is known for its efficiency, requiring relatively little training data to estimate parameters. However, its reliance on the independence assumption can be a limitation in complex real-world scenarios where features are often interdependent. Overall, Naive Bayes is valued for its ease of implementation and effectiveness, particularly in scenarios with high-dimensional data.

1.4. High-throughput experimentation

The development of new materials, including catalysts, involves exploring unknown materials and improving known ones[50,51]. This process often requires a

balance between trial and error, with more complex materials needing more time for development. High-throughput experimentation (HTE) has emerged as an effective method to expedite this process. HTE integrates the parallel production of diverse material sets with advanced measurement techniques to evaluate their intrinsic and performance characteristics[52]. This ability to simultaneously test multiple parameters allows for easy comparison of results. However, having numerous experiments alone doesn't speed up research unless these experiments are thoughtfully chosen. This is where the Design of Experiments (DoE) plays a crucial role[50].

The high-throughput or combinatorial approach in materials research was first demonstrated easily by Thomas A. Edison in 1878[50]. Edison's use of parallel and combinatorial methods led to the discovery that carbonized cotton yarn was an ideal material for light bulb filaments, a finding that revolutionized electric lighting. Later, massive high-throughput methods were primarily used in medicinal research, where their success was attributed to the relative simplicity of experimental manipulations required[53].

In the late 1990s, the method was adopted in catalysis research, particularly for homogeneous catalysts, which are generally used in liquid-phase reactions, allowing for smaller-scale experiments[54,55]. On the other hand, for the heterogeneous catalysts, developments in heterogeneous catalysis have mainly occurred in industrial settings over the past 15 years. The preparation of the catalyst is a critical step in high-throughput screening (HTS). Preparation time is essential for efficient HTS, with impregnation and precipitation being key methods for accelerating catalyst synthesis. These methods can be scaled up for miniaturization to handle numerous samples. Additionally, the use of synthetic robots has significantly enhanced the catalyst preparation process by automating it and reducing human error. Another important

aspect of high-throughput experimentation in catalysis is evaluating the catalysts, with spectroscopy now being the primary method for assessing HTS catalysts[56,57].

1.5. Catalysts informatics

1.5.1. Fundamentals

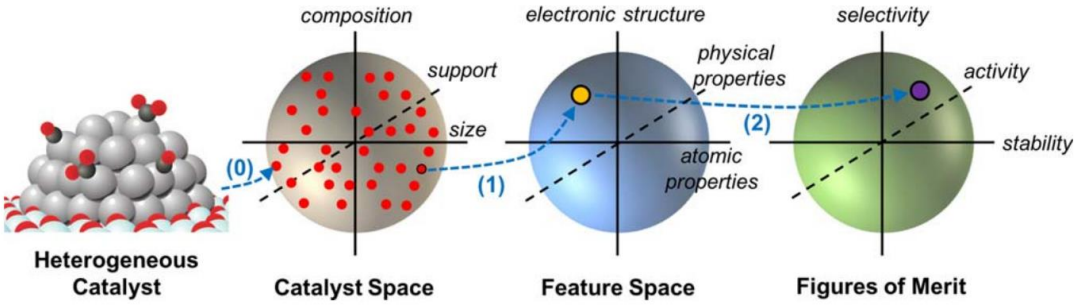


Figure 1.1. (0) A heterogeneous catalyst sample within catalyst space – containing catalysts with different composition, support type, and particle size – can be described by its (1) features within feature space, which is made up of electronic-structure properties, physical properties, and atomic properties. Machine learning can (2) build models or find descriptors that map the features describing the catalysts to their figures of merit. Reproduced from Ref.[58].

As shown in Figure 1.1, the final outputs of a catalyst, such as yield and selectivity, are determined by its physical properties, atomic properties, and electronic structure[58]. Additionally, these properties are shaped by various interacting factors, including the type and amount of the active metal, the promoter's type and quantity, the support type, and the methods used for preparation and pretreatment[59]. Subsequently, another dimension of variables, the process conditions such as temperature and pressure for the catalytic reaction to take place, is also added. Moreover, the non-linear nature

of catalytic performance, owing to unexpected synergistic effects, complicates the ability to directly influence outcomes during the catalyst preparation stage. As a result, achieving precise control and intentional design of catalysts is extremely challenging[11,12,60]. This complexity has led to the current reliance on empirical trial and error to develop most practical catalysts in industrial use[61,62].

Catalyst informatics is an initiative that aims to expedite catalyst development by applying data science and statistical methods to catalyst big data, which includes experimental, computational, and accumulated literature data[63,64]. In the initial stages of catalyst informatics, literature and computational data are often utilized[8,9,65–69]. However, machine learning analyses of literature data have indicated that such data exhibit low homogeneity, attributed to inconsistencies in data acquisition methods among different researchers and the reflection of historical trends in the field[70,71]. If possible, generating experimental data in-house is considered the most beneficial approach for catalyst informatics.

1.5.2. Random catalyst dataset for OCM

Catalyst informatics in OCM was pioneered in 2011 by Zavyalova et al. when catalytic reaction data appearing in the literature published from the discovery of OCM in 1982 to that point were compiled and statistical analyses performed on them[65]. A unified HTE framework was then developed by us in 2020[70]. We obtained bias-free catalyst big data of sufficient quality and quantity as training data for machine learning by this HTS system and random sampling from a predefined catalyst composition space[71]. Various catalysts have been developed by machine learning trained on this data[72–75], and the learners constructed in this thesis have all been trained on that data. This section describes in detail how it was obtained.

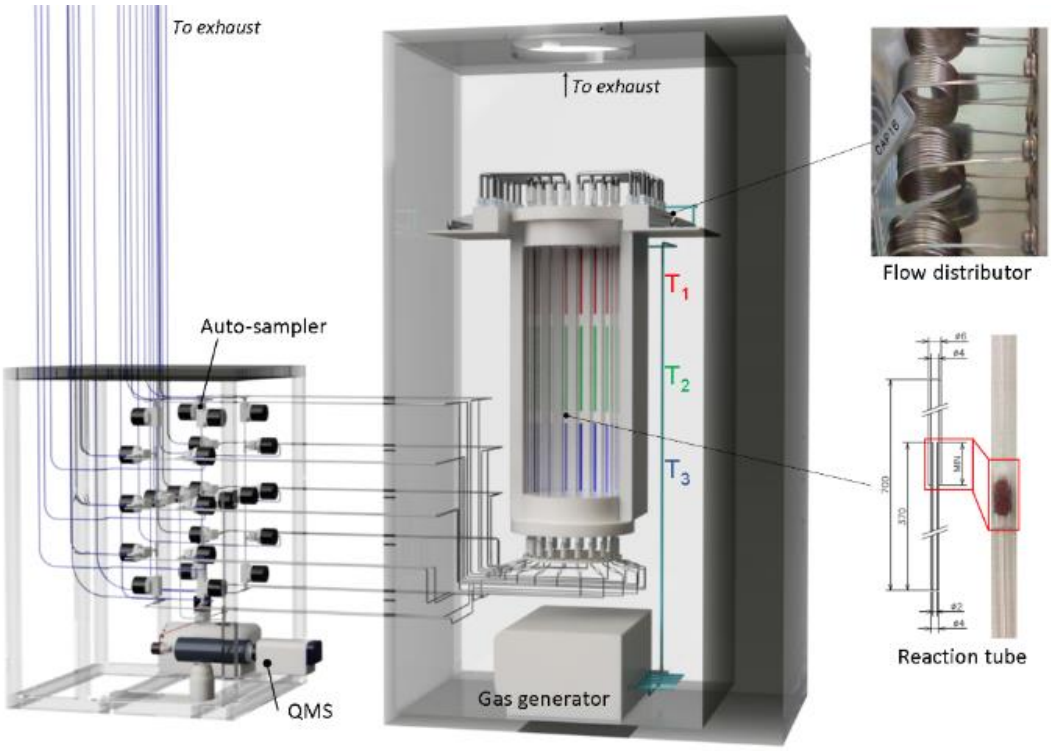


Figure 1.2. Illustration of the developed HTS system. Reproduced from ref.[71].

Figure 1.2 depicts the High-Throughput Screening (HTS) system developed in our previous study[70]. This system comprises several components: a mixed gas generator, a flow distributor, reaction tubes, an electric furnace, an auto-sampler, a quadrupole mass spectrometer (QMS), and an exhaust system. The gas generator mixes methane, oxygen, and argon in varying volumes, maintaining a total flow volume typically between 200 and 400 mL/min. The flow distributor which distribute the gas mixture into 20 reaction tubes is constructing with 20 capillaries. The reaction tubes, made of quartz with stepwise varying internal diameters, are designed to prevent gas-phase reactions on the effluent side and are filled with a specific amount of catalyst powder on the part changing the diameter. These tubes are placed symmetrically in a

hollow electric furnace with three temperature zones, each controlled by a thermocouple and ceramic heater. The middle zone, where the catalyst beds are located, is stabilized by the other two zones. The effluent gas from the reaction tubes is directed to the auto-sampler connected to the QMS for analysis. The auto-sampling involves a programmed sequence of valve operations, allowing for precise gas sampling and analysis. The QMS records mass spectra of the effluent gas, identifying various gas species. The mass signal intensities are converted to relative pressures for individual gas species using external calibration. This allows for dealing with overlapping fragments by obtaining scaling factors for major fragments. The system's design enables full automation in evaluating the performance of 20 catalysts under pre-set reaction conditions. The consistency of results across the 20 channels was verified using a Mn-Na_2WO_4/SiO_2 catalyst.

The catalyst library was created by preparing 300 catalysts. The catalysts were expressed in the form of M1–M2–M3/Support. The three active elements (M1–M3) were chosen randomly from a group consisting of 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 an option for "none," with the possibility of selecting the same element more than once. These elements were paired with a support material, randomly picked from MgO, Al_2O_3, SiO_2, CaO, TiO_2, ZrO_2, BaO, La_2O_3, and CeO_2. It's important to note that CaO and BaO were derived from their respective hydroxides through calcination. The potential combinations of three elements (including "none") and nine supports lead to a total of 36,540 possible catalysts. While the 300 chosen catalysts may not cover all significant trends in the entire parameter space, they ensure that each element and support is sampled at least 20 times.

The catalysts were prepared using a co-impregnation method. Each support (1.0 g) was impregnated with 4–5 mL of a metal precursor solution at 50 °C for 6 hours. Following vacuum drying at 110 °C, the product was calcined at 1000 °C in the air for 3 hours to yield the catalyst. When using water-sensitive metal alkoxides, impregnation was done sequentially, starting with an aqueous tungstate solution followed by an ethanol solution of the metal alkoxide. The catalysts were ground thoroughly before use. The loading of each element was standardized at 0.37 mmol per gram of support, with double the loading for repeated elements and no addition for "none." The catalysts were impregnated and processed as described, with twenty catalysts produced in a single experiment using a parallel hot stirrer and a centrifugal evaporator.

The height of the catalyst bed was set at 10 mm, considering the gas hourly space velocity (GHSV) as a key factor in OCM[65]. Catalysts were first activated at 1000 °C for 160 minutes under O_2, then the temperature was gradually reduced from 900 to 700 °C in steps. Catalysts were first activated at 1000 °C for 160 min under O_2. Then, the temperature was stepwise decreased from 900 to 850, 800, 750, and 700 °C. At each temperature, the total flow volume (10, 15, 20 mL/min/channel, corresponding to the contact time of 0.75, 0.50, 0.38 s, respectively), the CH_4/O_2 ratio (2, 4, 6 mol/mol), and the Ar concentration (PAr = 0.15, 0.40, 0.70 atm) were stepwise varied. This resulted in 135 conditions per catalyst and a total of 2700 observations for 20 catalysts in a single automated operation.

1.5.3. Catalyst informatics based on this dataset

Several studies have utilized the random catalyst dataset for catalyst informatics research. They are outlined here. K. Takahashi et al. introduced a novel method for representing catalysts by defining 'catalytic genes' using data that includes by-products from catalytic reactions[73]. Utilizing 300 quaternary catalyst data, they

employed hierarchical clustering and natural language processing to identify similar catalysts based on edit distance. They also demonstrated the potential for designing new catalysts by altering these catalytic genes, with validation experiments conducted via high-throughput methods. L. Takahashi et al proposed a technique to leverage this extensive catalytic dataset for constructing an ontology-based catalytic knowledge network[74]. This network aims to represent the performance and properties of catalysts for OCM and to elucidate their interactions. By clarifying structural relationships within the catalyst data, this network provides valuable direction and insights for the design of new catalysts. S. Nishimura et al.'s research focused on enhancing catalytic performance by adding Mn as a fourth element to the top ten three-element catalysts, selected based on their high C_2 yield and selectivity [76]. Experimental results verified that adding Mn improved the C_2 yield for several catalysts. Notably, the LiFeBa/La_2O_3 and LiBaLa/La_2O_3 catalysts with 5 wt% Mn showed significant performance enhancements. This study underscores the efficacy of machine learning in developing OCM catalysts.

1.6. Adaptive sampling

In the creation of new materials, the challenge of efficiently obtaining materials with desired properties through experimentation becomes more pronounced as the exploration space expands. This increase in complexity arises because, as previously mentioned, a larger search space not only multiplies the number of potential complementary materials but also compounds the intricacies of the materials themselves, complicating the establishment of reliable experimental guidelines. Conventionally, researchers have successfully developed desired materials through trial and error, based on accumulated experience in material development that enables them

to classify or regress materials against target properties, or occasionally by serendipity. Techniques such as adaptive sampling and black-box optimization are employed to codify the implicit knowledge about the input-output relationships of the target material—knowledge that is typically gained through iterative processes—using machine learning and statistical methods for an efficient and systematic exploration[64], [69]. This paper standardizes the terminology to "adaptive sampling." This approach facilitates research across vast material spaces, a feat that is challenging with traditional methods. However, machine learning, akin to human researchers, can exhibit prediction instability in data-sparse material spaces. Therefore, this instability in predictions is quantified as uncertainty, which reflects the extent to which a material is unknown and guides the selection of materials for further exploration. This is because, as mentioned in the section on HTE, material development has traditionally involved exploring uncharted materials and refining those already known. It is also recognized from a machine learning standpoint that balancing the predictive value with uncertainty—a concept known as the trade-off between exploration and exploitation—tends to yield optimal results[70]. The main algorithms employed in adaptive sampling include Bayesian optimization, best-arm identification, Monte Carlo tree search, and active learning[71–74].

If HTE is a strategy that increases the volume of hypothesis testing in iterative trials within the materials development process, then adaptive sampling enhances the quality of those tests. Utilizing both strategies in tandem can synergistically accelerate the pace of material discovery.

1.7. Objective

Solid catalysts with complicated compositions embody a complexity due to the numerous variables that constitute their makeup. Given this complexity, theoretical or computational approaches are hardly applicable, leading to most practical catalysts having being developed through trials and errors. However, with the increasing demands on catalysts, it has become necessary to navigate a vast material space that the traditional trial-and-error method cannot sufficiently address due to resource constraints. This thesis aims at the extraction of catalyst design rules and the systematization of the catalyst discoveries within the framework of catalyst informatics. For this, the OCM process, a catalytic reaction that has been a target of extensive trials and errors for more than 40 years with the goal of industrialization yet remains unrealized, was chosen as a case study.

Chapter 2 examines a dataset from high-throughput experimentation to derive heuristics for catalyst design in OCM, balancing activation and selectivity. It finds a mixed La_2O_3 and BaO support enhances low-temperature activity and high-temperature selectivity for C_2 compounds, with La_2O_3 boosting BaO's low-temperature performance.

In Chapter 3, a pioneering catalyst recommender system employing elemental substitution and adaptive sampling predicts catalyst performance, with a novel 'serendipiter' feature that strategically increases the discovery of serendipitous catalysts. Validated by high-throughput experiments, this approach led to identifying unique catalysts, showcasing the system's potential to harness serendipity systematically in catalyst development.

In Chapter 4, I tackled solid catalyst design, employing phylogenetic trees to visualize catalyst evolution and guide design. This approach successfully identifies

high-performance catalysts, demonstrating the technique's value in catalyst development.

Based on all of the above-explained researches and achievements, I believe the thesis made a substantial progress in the field of catalyst informatics.

Reference

[1] R. Schlögl, "Heterogeneous catalysis," *Angew. Chem., Int. Ed.*, **2015**, 54, 11, 3465–3520. doi: 10.1002/anie.201410738.

[2] P. Schwach, X. Pan, X. Bao, "Direct Conversion of Methane to Value-Added Chemicals over Heterogeneous Catalysts: Challenges and Prospects," *Chem. Rev.*, **2017**, 117, 13, 8497–8520. doi: 10.1021/acs.chemrev.6b00715.

[3] E. V. Kondratenko et al., "Methane conversion into different hydrocarbons or oxygenates: Current status and future perspectives in catalyst development and reactor operation," *Catal. Sci. Technol.*, **2017**, 7, 2, 366–381. doi: 10.1039/c6cy01879c.

[4] J. Sun, J. W. Thybaut, G. B. Marin, "Microkinetics of methane oxidative coupling," *Catal. Today*, **2008**, 137, 1, 90–102. doi: 10.1016/j.cattod.2008.02.026.

[5] J. W. Thybaut, J. Sun, L. Olivier, A. C. Van Veen, C. Mirodatos, G. B. Marin, "Catalyst design based on microkinetic models: Oxidative coupling of methane," *Catal. Today*, **2011**, 29–36. doi: 10.1016/j.cattod.2010.09.002.

[6] V. I. Alexiadis et al., "Oxidative coupling of methane: Catalytic behaviour assessment via comprehensive microkinetic modelling," *Appl. Catal. B*, **2014**, 150–151, 496–505. doi: 10.1016/j.apcatb.2013.12.043.

[7] A. Ishikawa, Y. Tateyama, "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. doi: 10.1021/acscatal.0c04104.

[8] G. Takasao, T. Wada, A. Thakur, P. Chammingkwan, M. Terano, T. Taniike, "Machine Learning-Aided Structure Determination for TiCl4-Capped $MgCl_2$ Nanoplate of Heterogeneous Ziegler-Natta Catalyst," *ACS Catal.*, **2019**, 9, 3, 2599–2609. doi: 10.1021/acscatal.8b05080.

[9] H. Chikuma, G. Takasao, T. Wada, P. Chammingkwan, J. Behler, T. Taniike, “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. doi: 10.1021/acs.jpcc.3c01511.

[10] C. Yu, J. He, “Synergic catalytic effects in confined spaces,” *Chem. Commun.,* **2012**, 48, 41, 4933–4940. doi: 10.1039/c2cc31585h.

[11] Y. Ying, X. Luo, J. Qiao, H. Huang, “‘More is Different:’ Synergistic Effect and Structural Engineering in Double-Atom Catalysts,” *Adv. Funct. Mater.*, **2021**, 31, 3. doi: 10.1002/adfm.202007423.

[12] A. G. Dedov, A. S. Loktev, I. I. Moiseev, A. Aboukais, J. F. Lamonier, I. N. Filimonov, “Oxidative coupling of methane catalyzed by rare earth oxides: Unexpected synergistic effect of the oxide mixtures,” *Appl. Catal. A Gen.*, **2003**, 245, 2, 209–220. doi: 10.1016/S0926-860X(02)00641-5.

[13] M. D. Argyle, C. H. Bartholomew, “Heterogeneous catalyst deactivation and regeneration: A review,” *Catalysts*, **2015**, 5, 1, 145–269. doi: 10.3390/catal5010145.

[14] E. J. Ras, G. Rothenberg, “Heterogeneous catalyst discovery using 21st century tools: A tutorial,” *RSC Adv.*, **2014**, 4, 12, 5963–5974. doi: 10.1039/c3ra45852k.

[15] J. S. Kim, B. Kim, H. Kim, K. Kang, “Recent Progress on Multimetal Oxide Catalysts for the Oxygen Evolution Reaction,” *Adv. Energy Mater.*, **2018**, 8, 11. doi: 10.1002/aenm.201702774.

[16] J. Timoshenko, B. Roldan Cuenya, “In Situ/Operando Electrocatalyst Characterization by X-ray Absorption Spectroscopy,” *Chem. Rev.*, **2021**, 121, 2, 882–961. doi: 10.1021/acs.chemrev.0c00396.

[17] H. Topsøe, “Developments in operando studies and in situ characterization of heterogeneous catalysts,” *J. Catal.*, **2003**, 155–164. doi: 10.1016/S0021-9517(02)00133-1.

[18] J. M. Thomas, “Design, synthesis, and in situ characterization of new solid catalysts,” *Angew. Chem., Int. Ed.*, **1999**, 38, 24, 3588–3628. doi: 10.1002/(SICI)1521-3773(19991216)38:24<3588::AID-ANIE3588>3.0.CO;2-4.

[19] N. Koike et al., “Transcriptional architecture and chromatin landscape of the core circadian clock in mammals,” *Science*, **2012**, 338, 6105, 349–354. doi: 10.1126/science.1226339.

[20] W. Taifan, J. Baltrusaitis, “CH4 conversion to value added products: Potential, limitations and extensions of a single step heterogeneous catalysis,” *Appl. Catal. B: Environ.*, **2016**, 198, 525–547. doi: 10.1016/j.apcatb.2016.05.081.

[21] W. J. Jang, J. O. Shim, H. M. Kim, S. Y. Yoo, H. S. Roh, “A review on dry reforming of methane in aspect of catalytic properties,” *Catal. Today*, **2019**, 15–26. doi: 10.1016/j.cattod.2018.07.032.

[22] G. E. Keller, M. M. Bhasin, “Synthesis of Ethylene via Oxidative Coupling of Methane. I. Determination of Active Catalysts.”

[23] D. Fan, D. J. Dai, H. S. Wu, “Ethylene formation by catalytic dehydration of ethanol with industrial considerations,” *Materials*, **2013**, 6, 1, 101–115. doi: 10.3390/ma6010101.

[24] Y. Gao et al., “Recent Advances in Intensified Ethylene Production - A Review,” *ACS Catal.*, **2019**, 9, 9, 8592–8621. doi: 10.1021/acscatal.9b02922.

[25] B. L. Farrell, V. O. Igenegbai, S. Linic, “A Viewpoint on Direct Methane Conversion to Ethane and Ethylene Using Oxidative Coupling on Solid Catalysts,” *ACS Catal.*, **2016**, 6, 7, 4340–4346. doi: 10.1021/acscatal.6b01087.

[26] A. P. Ortiz-Espinoza, M. M. B. Noureldin, M. M. El-Halwagi, A. Jiménez-Gutiérrez, “Design, simulation and techno-economic analysis of two processes for the conversion of shale gas to ethylene,” *Comput. Chem. Eng.*, **2017**, 107, 237–246. doi: 10.1016/j.compchemeng.2017.05.023.

[27] J. Liu et al., “From fundamentals to chemical engineering on oxidative coupling of methane for ethylene production: A review,” *Carbon Resour. Convers.*, **2022**, 5, 1, 1–14. doi: 10.1016/j.crcon.2021.11.001.

[28] R. D. Amos, “An accurate ab initio study of the multipole moments and polarizabilities of methane,” *Mol. Phys.*, **1979**, 38, 1, 33–45. doi: 10.1080/00268977900101511.

[29] R. Horn, R. Schlögl, “Methane Activation by Heterogeneous Catalysis,” *Catal. Lett.*, **2015**, 145, 1, 23–39. doi: 10.1007/s10562-014-1417-z.

[30] A. Cruellas, J. J. Bakker, M. van Sint Annaland, J. A. Medrano, F. Gallucci, “Techno-economic analysis of oxidative coupling of methane: Current state of the art and future perspectives,” *Energy Convers. Manag.*, **2019**, 198. doi: 10.1016/j.enconman.2019.111789.

[31] Z. Gao, J. Zhang, R. Wang, “Formation of hydrogen in oxidative coupling of methane over BaCO3 and MgO catalysts,” *J. Nat. Gas Chem.*, **2008**, 17, 3, 238–241. doi: 10.1016/S1003-9953(08)60057-2.

[32] S. Arndt et al., “A critical assessment of Li/MgO-Based catalysts for the oxidative coupling of methane,” *Catal. Rev. Sci. Eng.*, **2011**, 53, 4, 424–514. doi: 10.1080/01614940.2011.613330.

[33] P. Myrach et al., “Temperature-Dependent Morphology, Magnetic and Optical Properties of Li-Doped MgO,” *ChemCatChem*, **2010**, 2, 7, 854–862. doi: 10.1002/cctc.201000083.

[34] V. R. Choudhary, S. A. R. Mulla, B. S. Uphade, “Oxidative Coupling of Methane over Supported La2O3 and La-Promoted MgO Catalysts: Influence of Catalyst-Support Interactions,” *Fuel* **1999**, 78(4), 427-437.

[35] P. Huang, Y. Zhao, J. Zhang, Y. Zhu, Y. Sun, “Exploiting shape effects of La2O3 nanocatalysts for oxidative coupling of methane reaction,” *Nanoscale*, **2013**, 5, 22, 10844–10848. doi: 10.1039/c3nr03617k.

[36] T. Jiang et al., "La_2O_3 catalysts with diverse spatial dimensionality for oxidative coupling of methane to produce ethylene and ethane," *RSC Adv.*, **2016**, 6, 41, 34872–34876. doi: 10.1039/c6ra01805j.

[37] J. Ohyama et al., "Direct design of active catalysts for low temperature oxidative coupling of methane via machine learning and data mining," *Catal. Sci. Technol.*, **2021**, 11, 2, 524–530. doi: 10.1039/d0cy01751e.

[38] J. M. Deboy, R. F. Hicks, "The Oxidative Coupling of Methane over Alkali, Alkaline Earth, and Rare Earth Oxides," *Ind. Eng. Chem. Res.* **1988**. 27, 9, 1577–1582, doi: 10.1021/ie00081a004

[39] L. Cong, Y. Zhao, S. Li, Y. Sun, "Sr-doping effects on La2O3 catalyst for oxidative coupling of methane," *Chin. J. Catal.*, **2017**, 38, 5, 899–907. doi: 10.1016/S1872-2067(17)62823-7.

[40] S. L. S. Kuś, M. Otremba, M. Taniewski, "The catalytic performance in oxidative coupling of methane and the surface basicity of La_2O_3, Nd_2O_3, ZrO_2, and Nb_2O_5," *Fuel*, **2003**, 82, 11, 1331–1338. doi: 10.1016/S0016-2361(03)00030-9.

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

[42] J. Wu, S. Li, "The Role of Distorted WO_4 in the Oxidative Coupling of Methane on Tungsten Oxide Supported Catalysts," *J. Phys. Chem.* **1995**, 99, 13, 4566–4568, doi: 10.1021/j100013a030.

[43] A. Palermo, J. Pedro, H. Vazquez, R. M. Lambert, "New efficient catalysts for the oxidative coupling of methane," **2000**, 68, 191–196, doi: 10.1023/A:1019072512423.

[44] Z.-C. Jiang, C.-J. Yu, X.-P. Fang, S.-B. Li, H.-L. Wang, "Oxide/Support Interaction and Surface Reconstruction in the Na_2WO_4/SiO_2 System," *J. Phys. Chem.* **1993**, 97, 12870-12875.

[45] A. Palermo, J. Pedro, H. Vazquez, A. F. Lee, M. S. Tikhov, R. M. Lambert, “Critical Influence of the Amorphous Silica-to-Cristobalite Phase Transition on the Performance of Mn/Na2WO4/SiO2 Catalysts for the Oxidative Coupling of Methane,” *J. Catal.* **1998**, 177, 2, 259-266, doi: 10.1006/jcat.1998.2109.

[46] D. Matras et al., “Effect of thermal treatment on the stability of Na-Mn-W/SiO2 catalyst for the oxidative coupling of methane,” *Faraday Discuss.*, **2021**, 229, 176–196. doi: 10.1039/c9fd00142e.

[47] S. Arndt, T. Otremba, U. Simon, M. Yildiz, H. Schubert, R. Schomäcker, “Mn-Na_2WO_4/SiO_2 as catalyst for the oxidative coupling of methane. What is really known?,” *Appl. Catal. A Gen.*, **2012**, 425–426, 53–61. doi: 10.1016/j.apcata.2012.02.046.

[48] A. P. Dempster, “A Generalization of Bayesian Inference,” *J. R. Stat. Soc. Ser. B Methodol.* **1968**, 30, 205–232.

[49] M. Q. Ha et al., “Evidence-based recommender system for high-entropy alloys,” *Nat. Comput. Sci.*, **2021**, 1, 7, 470–478. doi: 10.1038/s43588-021-00097-w.

[50] R. Hoogenboom, M. A. R. Meier, U. S. Schubert, “Combinatorial methods, automated synthesis and high-throughput screening in polymer research: Past and present,” *Macromol. Rapid Commun.*, **2003**, 24, 1, 15–32. doi: 10.1002/marc.200390013.

[51] K. Terayama, M. Sumita, R. Tamura, K. Tsuda, “Black-Box Optimization for Automated Discovery,” *Acc. Chem. Res.*, **2021**, 54, 6, 1334–1346. doi: 10.1021/acs.accounts.0c00713.

[52] R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, H. Lam, “Combinatorial and high-throughput screening of materials libraries: Review of state of the art,” *ACS Comb. Sci.*, **2011**, 13, 6, 579–633. doi: 10.1021/co200007w.

[53] H. E. Tuinstra, C. H. Cummins, “Combinatorial Materials and Catalyst Research,” *Advanced Materials* **2000**, doi: 10.1002/1521-4095(200012)12:23<1819::AID-ADMA1819>3.0.CO;2-9

[54] S. M. Selim, “High-throughput screening of solid-state catalyst libraries,” *Nature*, **1998**, 394, 350–353, doi: 10.1038/28575.

[55] E. S. Isbrandt, R. J. Sullivan, S. G. Newman, “High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions,” *Angew. Chem.*, **2019**, 131, 22, 7254–7267. doi: 10.1002/ange.201812534.

[56] L. Olivier, S. Haag, H. Pennemann, C. Hofmann, C. Mirodatos, A. C. van Veen, “High-temperature parallel screening of catalysts for the oxidative coupling of methane,” *Catal. Today*, **2008**, 137, 1, 80–89. doi: 10.1016/j.cattod.2008.02.024.

[57] S. Moehmel et al., “New catalytic materials for the high-temperature synthesis of hydrocyanic acid from methane and ammonia by high-throughput approach,” *Appl. Catal. A Gen.*, **2008**, 334, 1–2, 73–83. doi: 10.1016/j.apcata.2007.09.035.

[58] B. R. Goldsmith, J. Esterhuizen, J. X. Liu, C. J. Bartel, C. Sutton, “Machine learning for heterogeneous catalyst design and discovery,” *AIChE J.*, **2018**, 64, 7, 2311–2323. doi: 10.1002/aic.16198.

[59] M. E. Günay, R. Yıldırım, “Recent advances in knowledge discovery for heterogeneous catalysis using machine learning,” *Catal. Rev. Sci. Eng.*, **2021**, 63, 1, 120–164. doi: 10.1080/01614940.2020.1770402.

[60] J. Yu et al., “Mining the synergistic effect in hydrothermal co-liquefaction of real feedstocks through machine learning approaches,” *Fuel*, **2023**, 334. doi: 10.1016/j.fuel.2022.126715.

[61] R. Schlögl, “Catalytic synthesis of ammonia - A ‘never-ending story’?,” *Angew. Chem., Int. Ed.*, **2003**, 42, 18, 2004–2008. doi: 10.1002/anie.200301553.

[62] J. M. Venegas, W. P. McDermott, I. Hermans, “Serendipity in Catalysis Research: Boron-Based Materials for Alkane Oxidative Dehydrogenation,” *Acc. Chem. Res.*, **2018**, 51, 10, 2556–2564. doi: 10.1021/acs.accounts.8b00330.

[63] K. Takahashi et al., “The Rise of Catalyst Informatics: Towards Catalyst Genomics,” *ChemCatChem*, **2019**, 11, 4, 1146–1152. doi: 10.1002/cctc.201801956.

[64] K. Takahashi et al., “Catalysts informatics: paradigm shift towards data-driven catalyst design,” *Chem. Commun.*, **2023**, 59, 16, 2222–2238. doi: 10.1039/d2cc05938j.

[65] U. Zavyalova, M. Holena, R. Schlögl, M. Baerns, “Statistical analysis of past catalytic data on oxidative methane coupling for new insights into the composition of high-performance catalysts,” *ChemCatChem*, **2011**, 3, 12, 1935–1947. doi: 10.1002/cctc.201100186.

[66] R. Schmack, A. Friedrich, E. V. Kondratenko, J. Polte, A. Werwatz, R. Kraehnert, “A meta-analysis of catalytic literature data reveals property-performance correlations for the OCM reaction,” *Nat. Commun.*, **2019**, 10, 1. doi: 10.1038/s41467-019-08325-8.

[67] P. Schlexer Lamoureux et al., “Machine Learning for Computational Heterogeneous Catalysis,” *ChemCatChem*, **2019**, 11, 16, 3581–3601. doi: 10.1002/cctc.201900595.

[68] S. Mine et al., “Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts,” *ChemCatChem*, **2021**, 13, 16, 3636–3655. doi: 10.1002/cctc.202100495.

[69] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. I. Shimizu, “Machine Learning for Catalysis Informatics: Recent Applications and Prospects,” *ACS Catal.*, **2020**, 10, 3, 2260–2297. doi: 10.1021/acscatal.9b04186.

[70] T. N. Nguyen et al., “High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane,” *ACS Catal.*, **2020**, 10, 2, 921–932. doi: 10.1021/acscatal.9b04293.

[71] T. N. Nguyen et al., “Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane,” *ACS Catal.*, **2021**, 11, 3, 1797–1809. doi: 10.1021/acscatal.0c04629.

[72] S. Nakanowatari et al., “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. doi: 10.1002/cctc.202100460.

[73] K. Takahashi et al., “Catalysis Gene Expression Profiling: Sequencing and Designing Catalysts,” *J. Phys. Chem. Lett.*, **2021**, 12, 30, 7335–7341. doi: 10.1021/acs.jpclett.1c02111.

[74] L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike, K. Takahashi, “Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts,” *Chem. Sci.*, **2021**, 12, 38, 12546–12555. doi: 10.1039/d1sc04390k.

[75] S. Ishioka, A. Fujiwara, S. Nakanowatari, L. Takahashi, T. Taniike, K. Takahashi, “Designing Catalyst Descriptors for Machine Learning in Oxidative Coupling of Methane,” *ACS Catal.*, **2022**, 12, 19, 11541–11546. doi: 10.1021/acscatal.2c03142.

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

[77] T. Lookman, P. V. Balachandran, D. Xue, R. Yuan, “Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design,” *npj Comput. Mater.*, **2019**, 5, 1. doi: 10.1038/s41524-019-0153-8.

[78] C. B. Browne et al., “A survey of Monte Carlo tree search methods,” *IEEE Trans. Comput. Intell. AI Games*, **2012**, 4, 1, 1–43. doi: 10.1109/TCIAIG.2012.2186810.

[79] K. Terayama et al., “Efficient construction method for phase diagrams using uncertainty sampling,” *Phys. Rev. Mater.*, **2019**, 3, 3. doi: 10.1103/PhysRevMaterials.3.033802.

[80] K. Shin, D. P. Tran, K. Takemura, A. Kitao, K. Terayama, K. Tsuda, “Enhancing Biomolecular Sampling with Reinforcement Learning: A Tree Search Molecular Dynamics Simulation Method,” *ACS Omega*, **2019**, 4, 9, 13853–13862. doi: 10.1021/acsomega.9b01480.

[81] K. Terayama, H. Iwata, M. Araki, Y. Okuno, K. Tsuda, “Machine learning accelerates MD-based binding pose prediction between ligands and proteins,” *Bioinformatics*, **2018**, 34, 5, 770–778. doi: 10.1093/bioinformatics/btx638.

[82] T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, K. Tsuda, “COMBO: An efficient Bayesian optimization library for materials science,” *Mater. Discov.*, **2016**, 4, 18–21. doi: 10.1016/j.md.2016.04.001.

Chapter 2

Extraction of catalyst design heuristics from random catalyst dataset and their utilization in catalyst development for oxidative coupling of methane

Abstract: In the oxidative coupling of methane (OCM), the activation of methane and the suppression of deep oxidation are in a persistent trade-off relationship, and a catalyst design strategy that balances the activity and the selectivity is desired. In this chapter, I analyzed a random catalyst dataset for OCM that was earlier obtained by high-throughput experimentation, and extracted heuristics such as elements, supports, and their combinations related to methane activation at a low temperature and selective formation of C_2 compounds at a high temperature. The obtained heuristics were used for catalyst development. The most effective was the use of a mixed support between La_2O_3 and BaO, which improved the low-temperature activity, the high-temperature selectivity, as well as the maximum C_2 yield. It was considered that La_2O_3 acted as a heater and helped low-temperature operation of BaO, which is highly selective but not active at a low temperature.

Keywords: Oxidative coupling of methane, catalyst informatics, high-throughput experimentation

2.1. Introduction

In recent years, catalyst informatics has received a great deal of attention for its use in accelerated catalyst discoveries and mechanistic elucidation[1–5]. However, its implementation to literature data has been greatly restricted by the nature of the data itself[6,7]. An irreplaceable role of the Edisonian approach, which comes from difficulties to generalize combinatorial design of solid catalysts, has more or less biased exploration of catalytic materials to past discoveries. Another problem is related to the lack of standard processes for the evaluation of catalyst performances. The biased sampling and the process inconsistency in literature data have made the outcome of machine learning much less fruitful than expected[8–10].

The oxidative coupling of methane (OCM) is a reaction that directly converts methane to ethane or ethylene without mediating syngas. Motivated by the expected shortage of petroleum and the abundance of natural gas, catalyst research has been continued over three decades[11–13]. However, few catalysts are known that can steadily achieve an C_2 yield over 30% or even 25% in a fixed-bed reactor configuration[8,14,15]. A grand challenge lies in the fact that methane is chemically more stable than the C_2 products against oxidation. Mild conditions (e.g. a lower temperature, a lower concentration of an oxidant) lead to insufficient activation of methane i.e. low conversion, while severe conditions promote mineralization of the C_2 products i.e. low selectivity. Thus, activation of methane at milder conditions and suppression of the secondary oxidation at severer conditions are important directions for catalyst development[16,17].

Recently, we have developed a high-throughput screening (HTS) instrument, which enables fully automated performance evaluation for 20 catalysts placed in

parallelized fixed-bed reactors at a predefined series of reaction conditions[18]. It allows an access to large, uniform, and process-consistent catalyst data in a realistic timeframe. Using this instrument, we created a bias-free catalyst dataset, where 300 quaternary catalysts were randomly sampled from a huge material space and their OCM performances were acquired at 135 reaction conditions[19]. In our previous research, the best C_2 yield data points were extracted for the individual catalysts from 135 conditions, and the extracted data were used to learn combinatorial rules for improving the C_2 yield as well as for predicting promising catalysts[19]. Here, the same dataset is further analyzed in terms of the activation of methane at milder conditions and the suppression of the secondary oxidation at severer conditions. The obtained heuristics are then employed to develop novel OCM catalysts.

2.2. Method

2.2.1. Catalyst dataset

The original dataset is a collection of OCM performance of 300 catalysts acquired at 135 conditions using the HTS instrument[18]. The catalysts are generally represented as M1–M2–M3/Support, consisting of three active elements (0.37 mmol for each) and a support (1.0 g). M1–M3 were randomly selected from either of 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 with repetitive selection allowed. The support was also randomly selected from either of MgO, Al_2O_3, SiO_2, CaO, TiO_2, ZrO_2, BaO, La_2O_3, and CeO_2. Each of the elements and supports were selected at least 20 times in the 300 catalysts. The 135 conditions consist of combinations of 5 temperatures (700, 750, 800, 850, 900 °C), 3 CH_4/O_2 ratios (2, 4, 6 mol/mol), 3 partial pressures of Ar as a carrier (0.15, 0.40, 0.70 atm), and 3 total flow volumes (10, 15, 20 mL/min/channel).

In the previous paper, the data points with the maximum C_2 yield were extracted for each of the catalysts, and catalyst design for increasing the C_2 yield regardless of the conditions was investigated. On the other hand, in this paper, four conditions were selected as representative conditions: (Temperature, CH_4/O_2 ratio) = (700 °C, 2 mol/mol), (700 °C, 6 mol/mol), (900 °C, 2 mol/mol), (900 °C, 2 mol/mol). No limits were set on the partial pressure of Ar (PAr) and total flow volume (Q). Note that the reaction temperature and the CH_4/O_2 ratio are the most impactful parameters in OCM[18],[41]. By analyzing composition-performance relationships at each of these conditions, catalyst design for the activation of methane at milder conditions and the suppression of the secondary oxidation at severer conditions would be derived.

2.2.2. Validation experiments

Based on heuristics obtained from the analysis, 20 catalysts were newly prepared and their OCM performances were evaluated. The methods and conditions for catalyst preparation and evaluation were exactly the same as those described in the previous paper[19].

Briefly, catalysts were prepared based on an impregnation method. Support powder was impregnated with an aqueous or ethanol solution of metal precursors (0.37 mmol for each of three active elements per unit gram of the support). The powder was vacuum dried and calcined in air at 1000 °C for 3 h.

The HTS instrument works with cooperative actions of a mixed gas generator, a flow distributor, reaction tubes, an electric furnace, an autosampler, and a quadrupole mass spectrometer (QMS). The gas mixer provides a gas mixture of $CH_4/O_2/Ar$ with controlled flow volumes. The generated gas mixture is equally split into 20 reaction tubes at the flow distributor. A reaction tube is made of quartz, whose internal diameter reduces from 4 mm in the feed side to 2 mm in the effluent side. A catalyst bed of 10 mm in height was fixed at the neck position of the tube with the aid of quartz wool. The 20 reaction tubes are placed in a hollow electric furnace. The furnace equips three temperature zones (T_1–T_3), where the catalyst beds are placed in the center zone (T_2). The effluent gas from the 20 reaction tubes is sequentially transferred to the autosampler and analyzed by the QMS so as to obtain the conversion of methane and the yields of products using external calibration. Similar to the previous paper, the catalytic performance was automatically acquired under 135 conditions[19]. The catalysts were first activated at 1000 °C under O_2. Then, the temperature of the catalyst beds (T_2) was stepwise decreased from 900 to 850, 800, 750, and 700 °C. T_1 and T_3

were set 100 °C lower than T_2 in order to minimize excessive oxidation in gas phase[18]. At each temperature, the total flow volume (Q = 10, 15, 20 mL/min/channel), the CH_4/O_2 ratio (2, 4, 6 mol/mol), and the Ar concentration (PAr = 0.15, 0.40, 0.70 atm) were stepwise varied. Each steady state is held for 6–7 min, which allows 2–3 sampling per catalyst.

2.3. Results

2.3.1. Data analysis for mild conditions and severe conditions

In this chapter catalyst design heuristics were extracted with focus on methane activation under mild conditions and high C_2 selectivity under severe conditions. Four subsets were newly created from the original dataset by extracting data points related to each of (temperature, CH_4/O_2 ratio) = (700 °C, 2 mol/mol), (700 °C, 6 mol/mol), (900 °C, 2 mol/mol), and (900 °C, 2 mol/mol). Figure 2.1a–d collects the best 20 catalysts at the four condition sets in terms of the C_2 yield. Note that one catalyst possesses nine data points at a specific set of the temperature and CH_4/O_2 ratio, which corresponds to the variation in the partial pressure of Ar (P_{Ar}) and total flow volume (Q). The performance of the catalyst corresponds to the highest C_2 yield among the nine data points.

Here, active elements and supports that frequently appear in the best 20 catalysts are analyzed for each condition set. At 700 °C, basic oxide-forming elements such as alkali metals, alkaline-earth metals, and lanthanides were found to be important (Figure 2.1a,b), while La_2O_3 was far the most selected support. La_2O_3 is known to be OCM-active by alone at low temperatures[20–22]. It was reported that La_2O_3 surfaces allow oxygen to be present as various anion species, (some of) which are highly active for methane activation[22,23]. Moreover, La_2O_3 has been frequently coupled with Sr and Ce[24–27], which is likely consistent with the observed importance of basic oxide-forming elements. At 900 °C, high performing catalysts often contained both basic oxide-forming elements and early transition metals as active elements (Figure 2.1c,d).

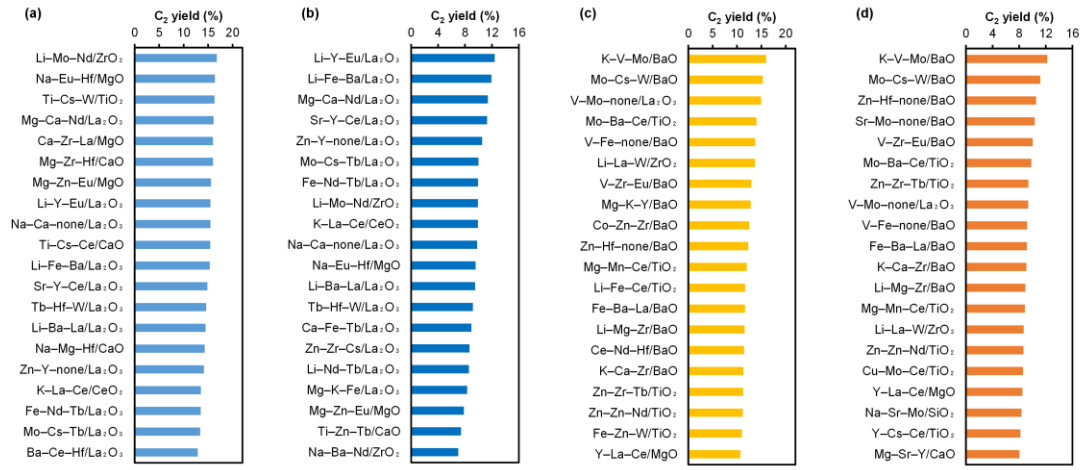


Figure 2.1. Best 20 catalysts at (Temp., CH_4/O_2 ratio) = (a) (700 °C, 2 mol/mol), (b) (700 °C, 6 mol/mol), (c) (900 °C, 2 mol/mol), and (d) (900 °C, 6 mol/mol). The data are extracted from a random catalyst dataset published in our previous paper[19].

This observation suggests the importance of oxometalate anions (e.g. WO_4^{2-}) as highly selective species at high temperatures[28,29]. On the other hand, the commonly selected supports were BaO and TiO_2. TiO_2-supported catalysts mostly accompanied basic oxide-forming elements as active elements, which suggests the formation of TiO_3^{2-} as an oxometalate anion. BaO has not been known to be a good support for OCM, especially at high temperatures. Here, it is important to note that MgO and CaO performed much better at 700 °C than at 900 °C. One potential way to explain the different behavior of these alkaline earth metal oxides is the difference in the thermal stability of their carbonates: CO_2, a major by-product of OCM, can react with these oxides to in-situ form carbonates. $BaCO_3$ has a much higher decomposition temperature than $MgCO_3$ and $CaCO_3$ (1350 °C vs. 350 °C and 825 °C). Note that the decomposition temperature of a carbonate significantly decreases in the presence of CH_4, e.g. 800–1100 °C for $BaCO_3$[30]. Hence, only BaO likely exists (partly) as $BaCO_3$.

The presence of $BaCO_3$ in the used catalyst was confirmed by FT-IR in Figure 2.2. Meanwhile, MgO and CaO exist as basic oxides, being suitable for low-temperature activation of CH_4. An oxometalate anion works as an active site that selectively mediates the coupling of CH_4. The same role can be played by oxoanions of typical elements such as PO_4^{3-} and SO_4^{2-}[31], i.e. CO_3^{2-} of $BaCO_3$ is considered to be similarly selective. Besides, in-situ formed $BaCO_3$ might be decomposed by the exotherm of the reactions to suppress the formation of hot spots[32–34].

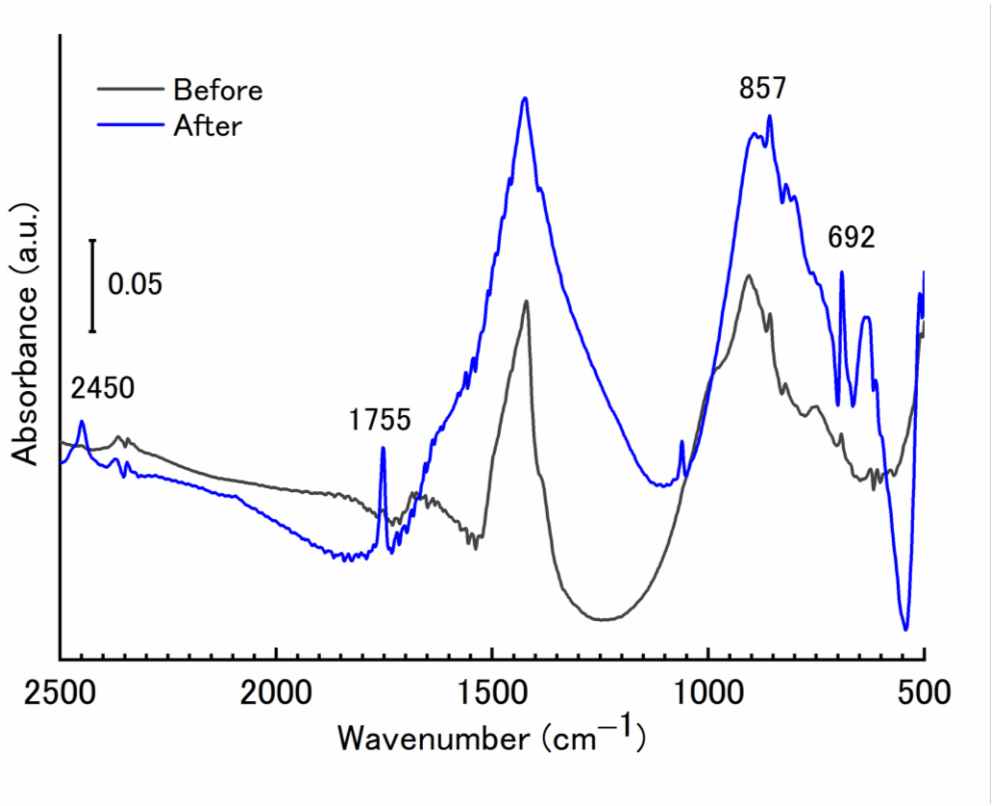


Figure 2.2. IR spectra of BaO before and after the use in OCM. After the use, new characteristic peaks appeared at around 2450 cm^{-1} ($BaCO_3$), 1755 cm^{-1} (C=O of CO_3^{2-}), 857 cm^{-1}, and 692 cm^{-1} (in-plane and out-of-plane bending of CO_3^{2-}).

In our previous paper, it was found that the OCM performance of a catalyst is hardly described only by the presence or absence of a specific component, but the combination is very critical[19]. For example, in Figure 2.1a, basic oxide-forming elements appear frequently in combination with La_2O_3. This means that these elements are synergistically combined with La_2O_3 rather than they are superior in low-temperature activation by their own. Thus, I analyzed binary combinations for the four condition sets, where the performance of a binary combination (A*B) is defined by averaging the C_2 yield over catalysts which contain both A and B. A combination was not considered unless more than one catalyst contains the combination. The best 20 binary combinations are summarized at each condition set in Figure 2.3. At 700 °C, the majority of the high performing combinations are pairs of active elements and supports (Figure 2.3a,b). They are subclassified into combinations of La_2O_3 with basic oxide-forming elements and those of CaO with redox oxide-forming elements (Ti, Ce, Hf). In the case of combination between active elements, the majority of the combinations were of between basic oxide-forming elements. At 900 °C, one half of the high performing combinations were of between active elements, and the other half were of between active elements and supports (Figure 2.3c,d). In the former case, most of the combinations arose from pairs of basic oxide-forming elements and early transition metals, validating the importance of oxometalate anions for high-temperature C_2 selectivity. In the latter case, the combinations mostly came from BaO with basic oxide-forming elements or early transition metals. These results suggest BaO, a basic oxide-forming element, and an early transition metal as a promising ternary combination.

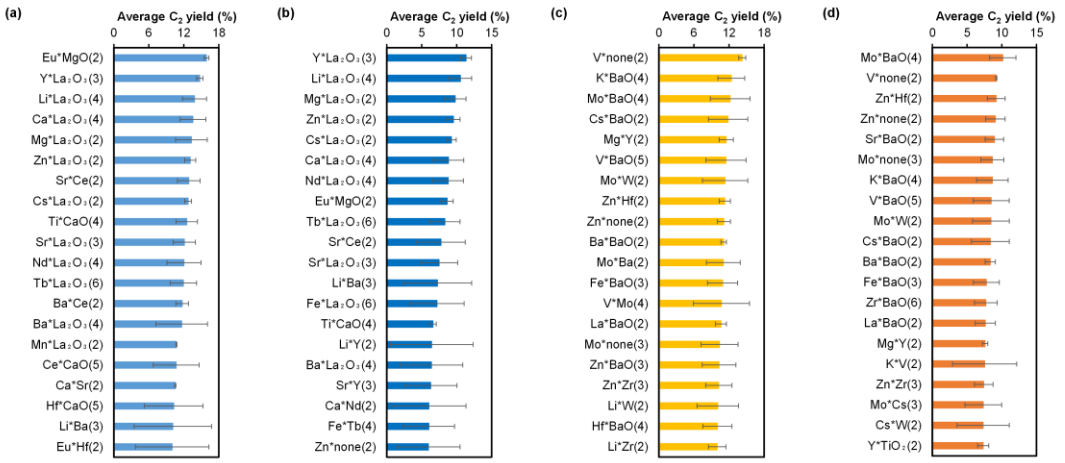


Figure 2.3. Best 20 binary combinations at (Temp., CH_4/O_2 ratio) = (a) (700 °C, 2 mol/mol), (b) (700 °C, 6 mol/mol), (c) (900 °C, 2 mol/mol), and (d) (900 °C, 6 mol/mol). The C_2 yields are averaged among catalysts containing a specific combination (A*B). The numbers in the parentheses correspond to the frequency of appearance of specific combinations within the dataset.

2.3.2. Catalyst development based on extracted heuristics

The analysis of the random catalyst dataset revealed that low-temperature activity and high-temperature selectivity could be associated with specific components and their combinations in the catalysts. Here, I aimed to develop catalysts by combining the heuristics related to the low-temperature activity and high-temperature selectivity. In detail, 20 catalysts were prepared based on the following four strategies.

1. Starting from Tb–Hf–W/La_2O_3, one or two active elements are replaced with other elements. Out of the La_2O_3-supported catalysts, Tb–Hf–W/La_2O_3 was ranked in the top 20 at 700 °C irrespective of the CH_4/O_2 ratio, and showed the highest C_2 yield at 900 °C. A quaternary catalyst, M1–M2–M3/Support, consists of 6 binary

combinations. The replacement of the active elements is carried out in a way to maximize the sum of the calculated average yield over the 6 binary combinations.

2. Starting from Li–Mg–Zr/BaO, one or two active elements are replaced with other elements. Li–Mg–Zr/BaO is selected and modified in a similar fashion to that described in the strategy 1.

3. La_2O_3-supported catalysts with the highest C_2 yields at 700 °C and the CH_4/O_2 ratio of 2 mol/mol are selected. Catalysts are prepared using a mixed support of BaO (0.5 g) and La_2O_3 (0.5 g) instead of La_2O_3 (1.0 g), while keeping the active elements.

4. BaO-supported catalysts with the highest C_2 yields at 900 °C and the CH_4/O_2 ratio of 2 are selected. Catalysts are prepared using a mixed support of BaO (0.5 g) and La_2O_3 (0.5 g) instead of BaO (1.0 g), while keeping the active elements.

These strategies are more or less based on the heuristics that La_2O_3 is a major cause of low-temperature activity and BaO is a major cause of high-temperature selectivity. The strategies 1 and 2 aim to impart high-temperature selectivity to La_2O_3 and low-temperature activity to BaO by judiciously combining active elements. The strategies 3 and 4 aim to combine the low-temperature activity of La_2O_3 and the high-temperature selectivity of BaO by means of a mixed support.

Table 2.1 summarizes the composition and the performance of the 20 catalysts that were prepared based on the above-mentioned strategies. Note that the performance is shown for only two representative conditions, but the catalysis was actually evaluated under the 135 conditions as described in Experimental (the full data is available in a web platform, Catalyst Acquisition by Data Science[35]). Also note that the performance at each specified temperature and CH_4/O_2 ratio corresponds to the data point showing the highest C_2 yield in terms of the variations of Q and P_{Ar}. Table 2.1

also includes catalysts that serve as references for each catalyst design strategy. The sample code is as described in the footnote.

Table 2.1. Performance of catalysts prepared based on four different strategies.[a,b,c]

Code	Name	700 ˚C, CH_4/O_2 = 2 mol/mol			900 ˚C, CH_4/O_2 = 2 mol/mol		
		C_2 yield (%)	CH_4 conversion (%)	C_2 selectivity (%)	C_2 yield (%)	CH_4 conversion (%)	C_2 selectivity (%)
1-1	Zn–Mo–Hf/La_2O_3	8.6	30.1	28.4	3.8	35.1	10.8
1-2	Zn–Cs–Hf/La_2O_3	13.1	33.9	38.7	4.2	36.2	11.5
1-3	Zn–Eu–Hf/La_2O_3	16.3	36.0	45.2	6.0	35.1	17.2
1-4	Zn–Hf–W/La_2O_3	9.6	32.5	29.4	1.3	33.8	4.0
1-5	Zn–Hf–none/La_2O_3	8.4	50.1	16.8	2.8	38.3	7.3
R1	Tb–Hf–W/La_2O_3	14.5	41.4	35.1	9.7	35.3	27.5
2-1	Li–Y–Eu/BaO	1.6	3.7	42.7	14.9	41.0	36.4
2-2	Li–Mo–Ba/BaO	5.3	10.2	52.4	13.4	39.9	33.6
2-3	Li–Mo–Nd/BaO	2.5	7.1	34.7	14.0	39.3	35.6
2-4	Li–Ba–Nd/BaO	5.0	15.2	32.5	10.6	38.7	27.4
2-5	Li–Nd–Tb/BaO	2.8	10.0	27.6	13.2	37.8	34.9
R2	Li–Mg–Zr/BaO	4.0	15.1	26.3	11.5	36.8	31.1
3-1	Mg–Ca–Nd/BaO–La_2O_3	19.5	39.9	48.7	10.1	34.0	29.6
3-2	Li–Y–Eu/BaO–La_2O_3	19.0	39.3	48.3	8.9	37.4	23.9
3-3	Na–Ca–none/BaO–La_2O_3	18.1	37.7	47.9	10.1	34.0	29.6
3-4	Li–Fe–Ba/BaO–La_2O_3	16.9	36.1	46.7	7.6	31.9	24.0
3-5	Sr–Y–Ce/BaO–La_2O_3	18.2	37.9	48.0	12.6	39.1	32.1
R3-1	Mg–Ca–Nd/La_2O_3	16.0	43.5	36.9	3.0	33.0	9.2
R3-2	Li–Y–Eu/La_2O_3	15.4	35.9	42.8	6.4	35.5	18.1
R3-3	Na–Ca–none/La_2O_3	15.4	33.5	45.9	5.0	32.2	15.4
R3-4	Li–Fe–Ba/La_2O_3	15.2	35.5	43.0	5.1	34.1	14.9
R3-5	Sr–Y–Ce/La_2O_3	14.8	37.1	39.9	6.2	33.4	18.4
4-1	K–V–Mo/BaO–La_2O_3	17.7	35.9	49.4	8.7	35.9	24.1
4-2	Mo–Cs–W/BaO–La_2O_3	19.1	41.3	46.3	5.7	32.4	17.7
4-3	V–Fe–none/BaO–La_2O_3	17.4	38.0	45.8	3.3	34.5	9.5
4-4	V–Zr–Eu/BaO–La_2O_3	17.4	39.1	44.5	4.3	32.6	13.3
4-5	Mg–K–Y/BaO–La_2O_3	20.3	40.8	49.7	10.6	37.4	28.2
R4-1	K–V–Mo/BaO	0	0	n.d.	15.9	38.2	41.6
R4-2	Mo–Cs–W/BaO	0	0	n.d.	15.2	39.3	38.7
R4-3	V–Fe–none/BaO	0.9	3.7	24.9	13.7	37.7	36.3
R4-4	V–Zr–Eu/BaO	0.9	1.9	45.2	12.9	36.2	35.6
R4-5	Mg–K–Y/BaO	1.8	5.4	34.0	12.7	35.8	35.6

[a]$LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, KNO_3, $Ca(NO_3)_2{\cdot}4H_2O$, $VOSO_4{\cdot}xH_2O$ (x = 3–5), $Fe(NO_3)_3{\cdot}9H_2O$, $Zn(NO_3)_2{\cdot}6H_2O$, $Sr(NO_3)_2$, $Y(NO_3)_3{\cdot}6H_2O$, $ZrO(NO_3)_2{\cdot}xH_2O$ (x = 2), $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, $CsNO_3$, $Ba(NO_3)_2$, $Ce(NO_3)_3{\cdot}6H_2O$, $Nd(NO_3)_3{\cdot}6H_2O$, $Eu(CH_3COO)_3{\cdot}xH_2O$ (x = 4), $Tb(NO_3)_3{\cdot}5H_2O$, $Hf(OEt)_4$, and $(NH_4)_{10}H_2(W_2O_7)_6$ were used as metal precursors. These reagents were purchased from either Sigma-Aldrich, Kanto Chemical, Wako Pure Chemical Industries, or Alfa-Aesar. Barium hydroxide ($Ba(OH)_2{\cdot}8H_2O$, 1.1 m^2/g, Wako Pure Chemical Industries), and lanthanum(III) oxide (La_2O_3, 8.3 m^2/g, Wako Pure Chemical Industries) were used as a support precursor or a support.

[b]The sample code, *x*-*y*, means the *y*-th catalyst in the *x*-th strategy. R*x*-(*y*) corresponds to the reference catalyst for the catalyst *x*-*y*: e.g. The catalysts from 1-1 to 1-5 were derived from R1 by replacing one or two active elements.

[c]The performance at each specified temperature and CH_4/O_2 ratio corresponds to the data point showing the highest C_2 yield in terms of the variations of Q and P_{Ar}.

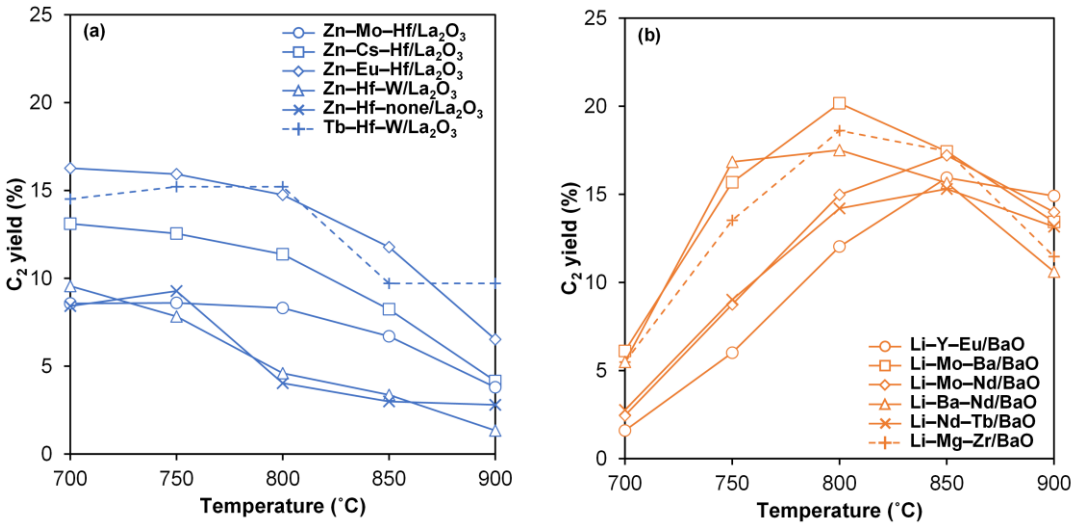


Figure 2.4. Performance of catalysts that are derived based on (a) the strategy 1 and (b) the strategy 2. The C_2 yield is plotted against the reaction temperature. The performance of reference catalysts is also shown for comparison (dashed lines).

Figure 2.4 compares the performance of the catalysts that were derived based on the strategies 1 and 2 for the replacement of active elements. Here, the C_2 yield corresponds to the highest value at each temperature in terms of the variations of Q, the CH_4/O_2 ratio, and P_{Ar} (i.e. among 27 conditions). The results for the CH_4 conversion and the C_2 selectivity and the distribution of the C_2 yield irrespective of reaction conditions are also shown in Figures 2.5 and 2.6.

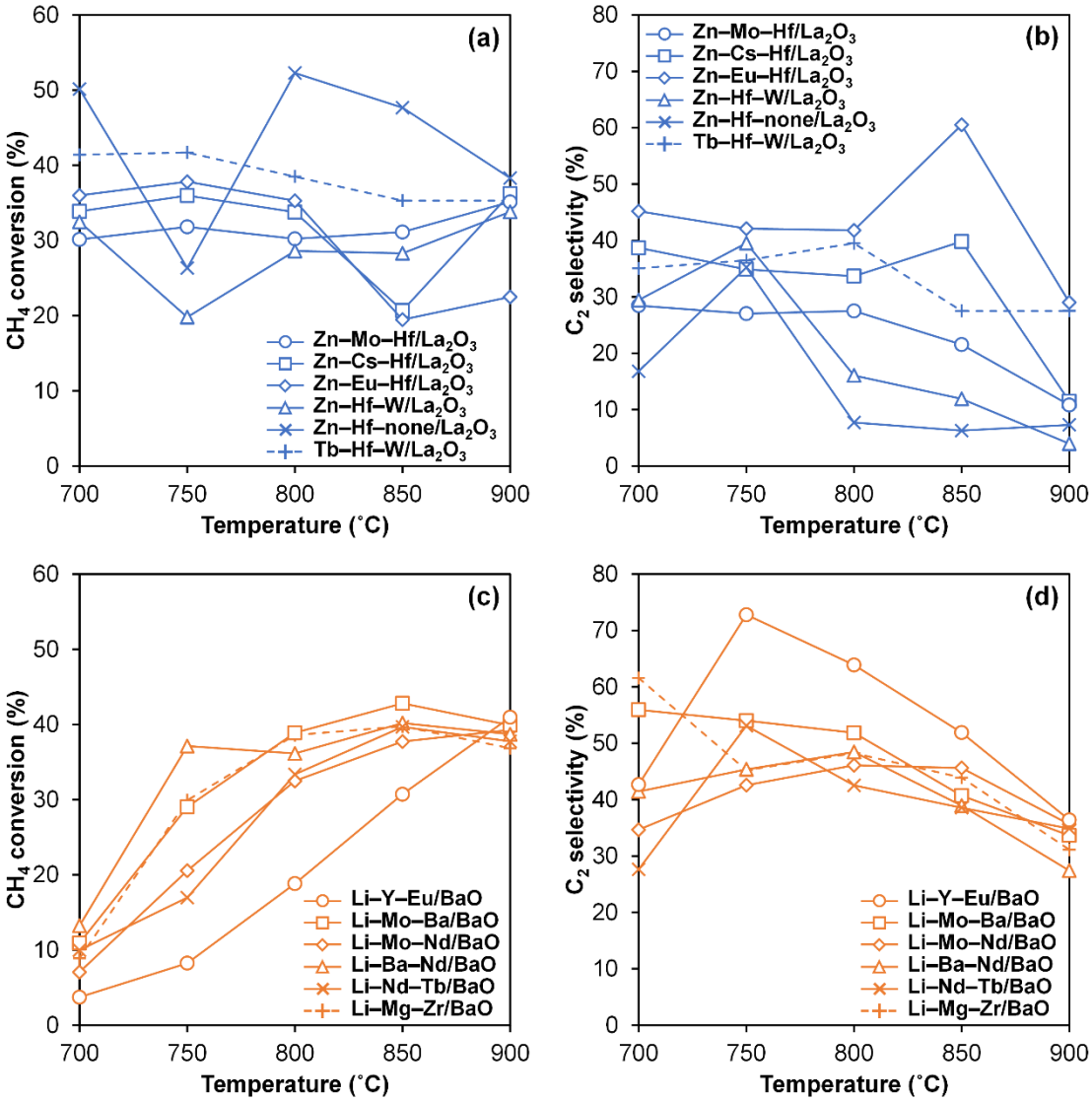


Figure 2.5. Performance of catalysts that are derived based on (a,b) the strategy 1 and (c,d) the strategy 2. The performance of reference catalysts is also shown for comparison (dashed lines).

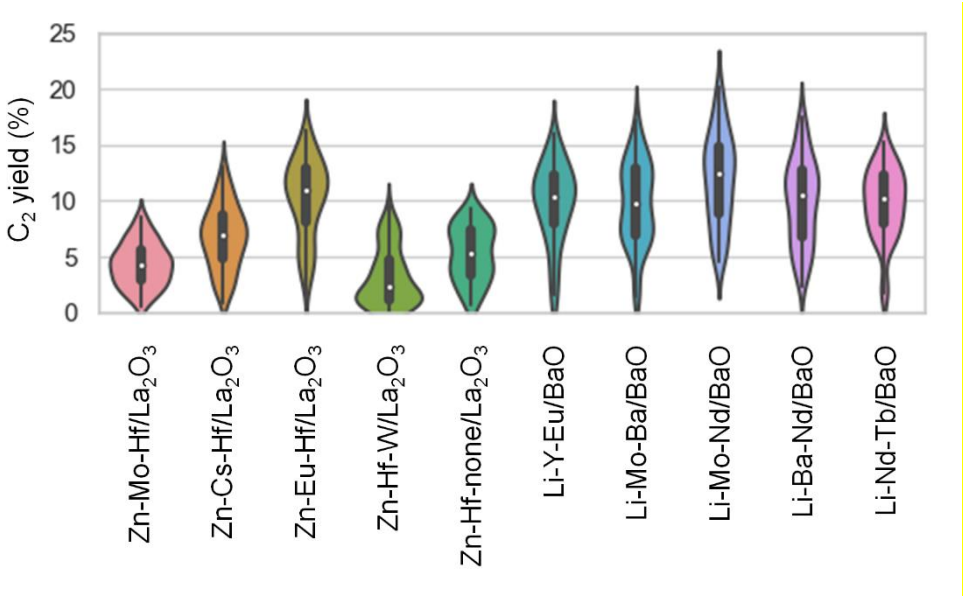


Figure 2.6. Violin plot of the C_2 yield for the catalysts derived based on the strategies 1 and 2.

The As shown in Figure 2.4, the response of the catalysts to the temperature was largely determined by the choice of the supports. When La_2O_3 was used as a support, the maximum yield was achieved at 700–750 °C, and an increase in the temperature caused a decrease in the C_2 selectivity and the C_2 yield. On the other hand, the maximum yield was achieved at 800–850 °C when BaO was used as a support, and the decrease in the temperature led to a significant decrease in the conversion and the C_2 yield. The choice of active elements did not significantly alter these support-specific responses, but greatly affected the yield level. For example, the lack of alkali metal or lanthanide group elements as M1–M3 significantly lowered the C_2 yield for the La_2O_3-supported catalysts. The presence of an alkaline earth metal element was essential for the BaO-supported catalysts, especially at lower temperatures. Here, one can see that most of the derived catalysts exhibited the performance inferior to that of the corresponding reference catalysts. Thus, the strategies 1 and 2 were judged to be

ineffective. The failure of these strategies is plausibly related to the lack of interaction among more than two components, e.g. interaction between two elements depends on the choice of the other element and the support.

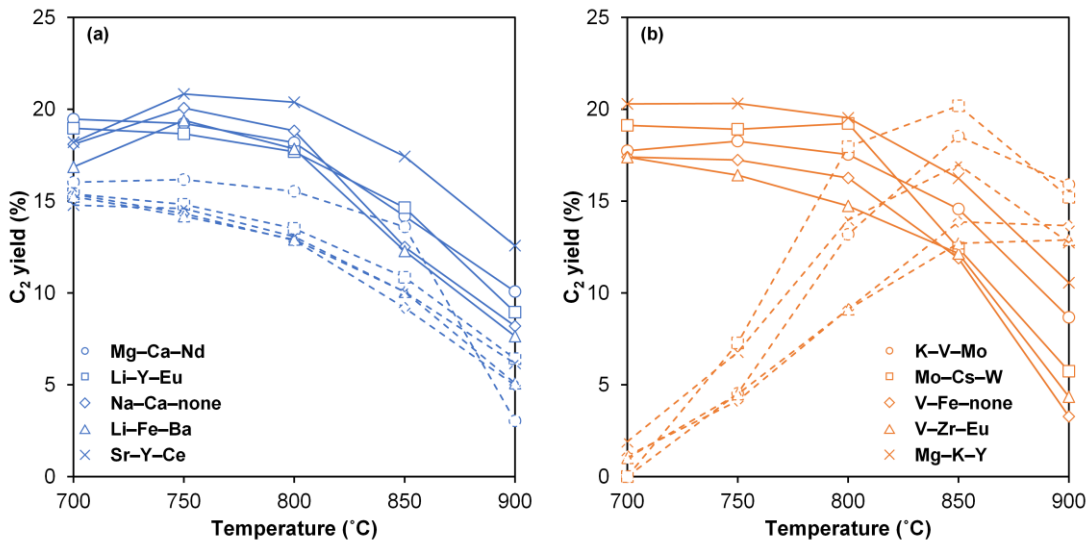


Figure 2.7. Performance of catalysts that are derived based on (a) the strategy 3 and (b) the strategy 4. The C_2 yield is plotted against the reaction temperature. Catalyst names are expressed based on active elements. Solid lines correspond to the use of BaO–La_2O_3 as a support while dashed lines correspond to the use of either (a) La_2O_3 or (b) BaO.

Figure 2.7 compares the performance of the catalysts that were derived based on the strategies 3 and 4 for the mixed support. The results for the CH_4 conversion and the C_2 selectivity and the distribution of the C_2 yield irrespective of reaction conditions are also shown in Figures 2.8 and 2.9.

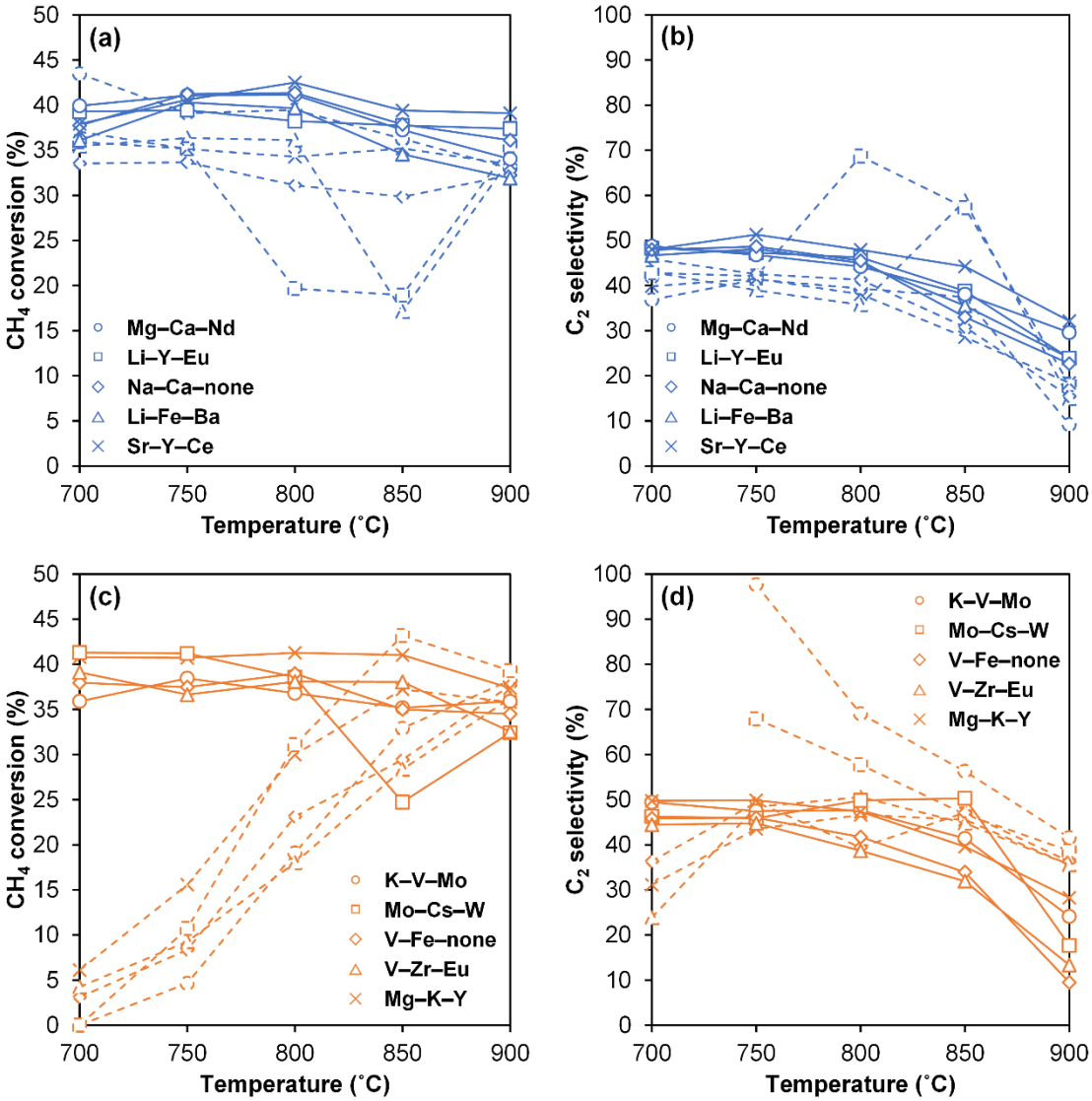


Figure 2.8. Performance of catalysts that are derived based on (a,b) the strategy 3 and (c,d) the strategy 4. Catalyst names are expressed based on active elements. Solid lines correspond to the use of BaO–La_2O_3 as a support while dashed lines correspond to the use of either (a,b) La_2O_3 or (c,d) BaO.

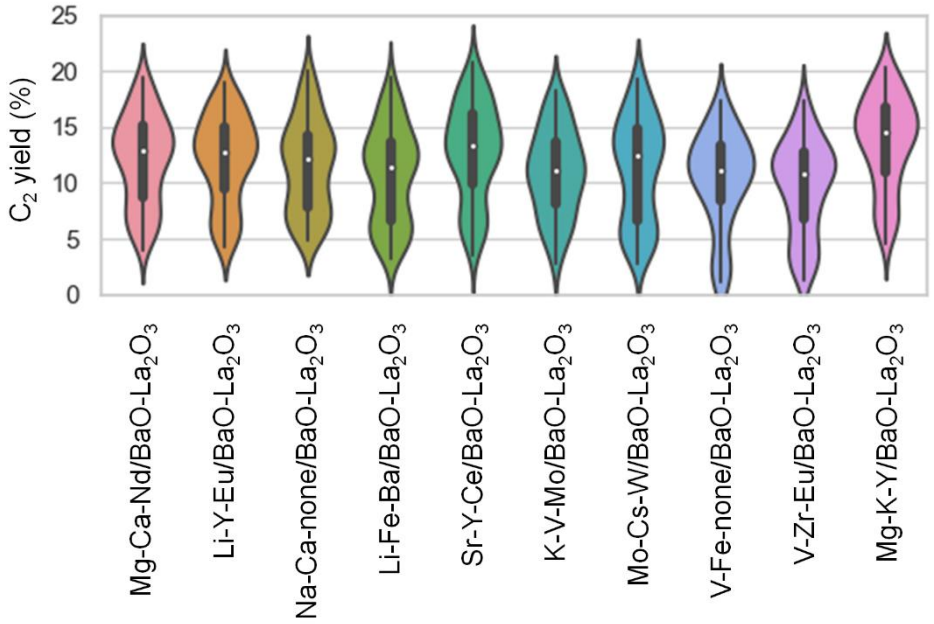


Figure 2.9. Violin plot of the C_2 yield for the catalysts derived based on the strategies 3 and 4.

The performance of the reference catalysts is also shown for comparison. The choice of supports is a major factor of the catalytic behavior. Hence, the effect of the mixed support was found to be dramatic. The addition of BaO to La_2O_3-supported catalysts significantly improved the C_2 selectivity at 900 °C. It also caused slight improvements in the conversion and the selectivity at lower temperatures. Consequently, the derived catalysts exhibited higher C_2 yields as compared to the corresponding reference catalysts at individual temperatures (Figure 2.7a). The addition of La_2O_3 to BaO-supported catalysts conferred high activity at 700 °C, as expected. Note that BaO alone exhibited almost no ability to activate CH_4 at 700 °C. On the other hand, the poor selectivity of La_2O_3 was also introduced, and this limited the advantage of the mixed support at higher temperatures (Figure 2.7b). In summary, it was found that the mixed support strategies are effective in improving the C_2 yield when applied to La_2O_3-

supported catalysts, and in lowering the operation temperature of the catalysts when applied to BaO-supported catalysts.

Based on the success of the mixed support strategy, additional experiments were carried out on the way of combining BaO and La_2O_3. In literature, it has been reported that the use of a two-layer bed with two different catalysts is effective for improving the C_2 yield[36–39]. Especially, Liang et al. adopted the same strategy in view of product concentration profiles along a flow direction[39]. Their aim was to enable the catalysis at a lower temperature while suppressing deep oxidation in the effluent side by using a more active catalyst as the first layer in the influencer side and a more selective catalyst as the second layer in the effluent side. In this light, 5 kinds of catalyst beds were prepared by using pristine supports without active elements: Single-layer beds of BaO and La_2O_3, a single-layer bed of a mixed support (denoted as BaO–La_2O_3), a two-layer bed of La_2O_3 as the first layer and BaO as the second layer (denoted as La_2O_3→BaO), and a two-layer bed with the opposite layer order (BaO→La_2O_3). The bed height was fixed at 10 mm, and in case of a two-layer bed, a 10 mm bed was composed of two layers of 5 mm.

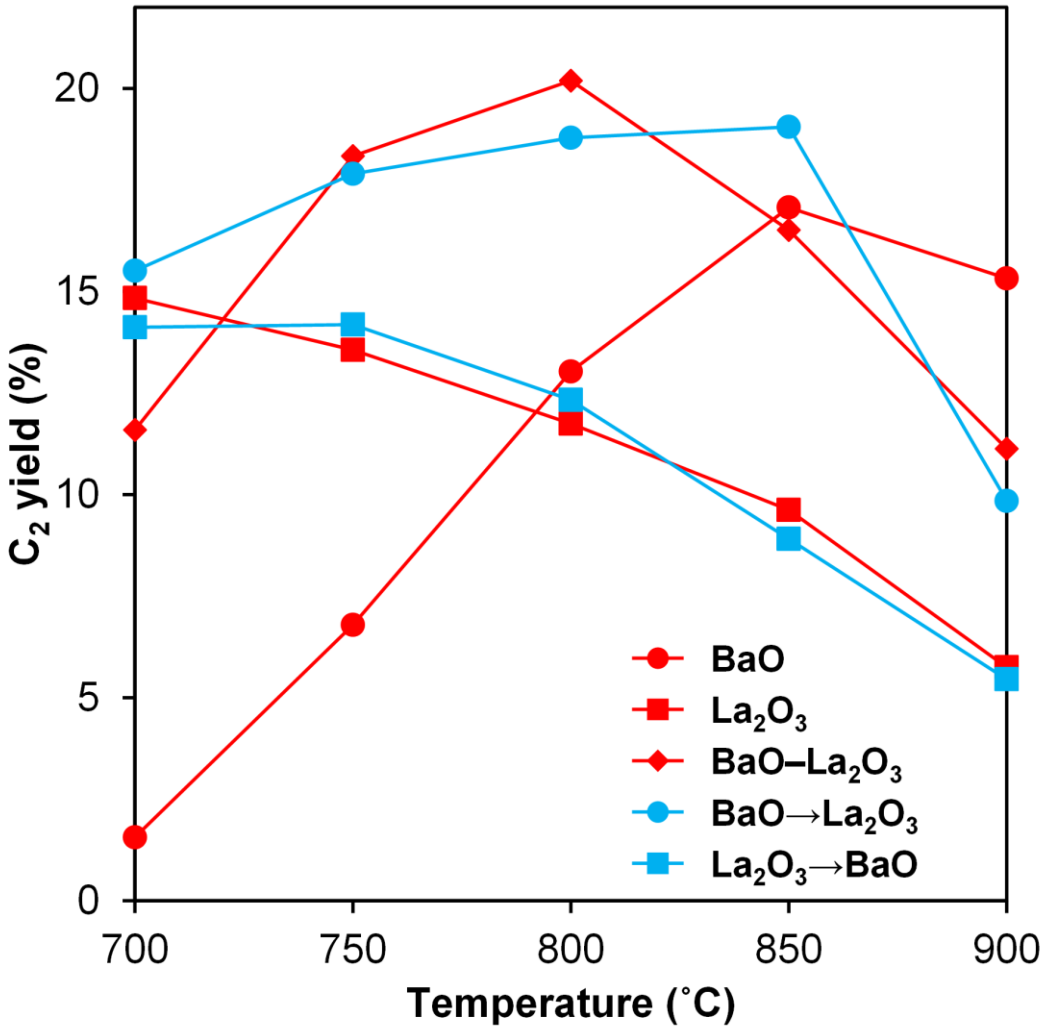


Figure 2.10. Effect of combining two supports on the OCM performance. BaO and La_2O_3 were combined in three different ways. BaO–La_2O_3: Single-layer bed of slurry-mixed powder; BaO→La_2O_3: Two-layer bed with BaO powder placed in the influent side; La_2O_3→BaO: Two-layer bed with La_2O_3 powder placed in the influent side.

The performance of the 5 catalyst beds is compared at different temperatures in Figure 2.10. Figure 2.11 is showing their CH_4 conversion and the C_2 selectivity.

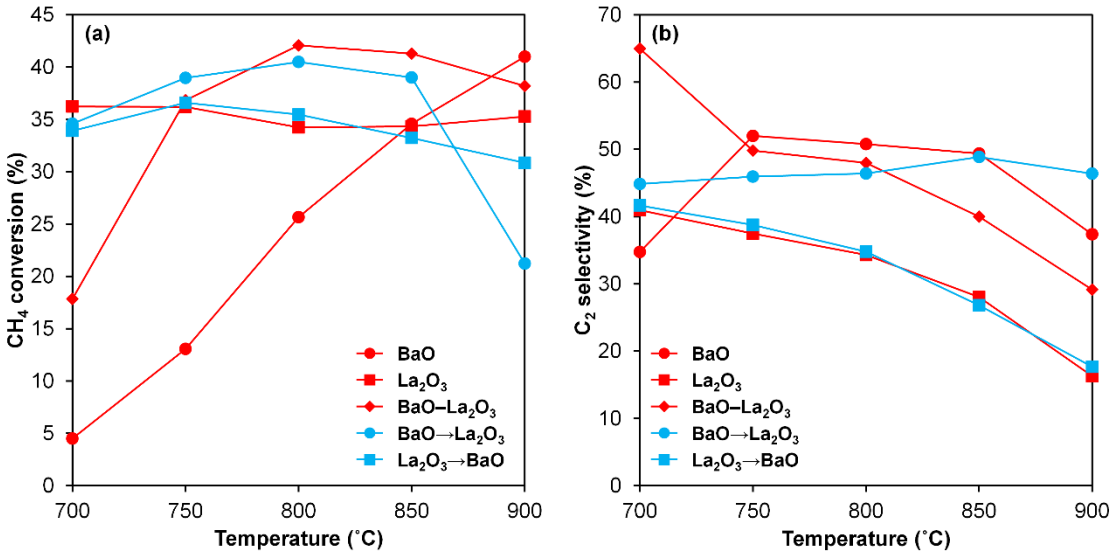


Figure 2.11. Effect of combining two supports on (a) the CH_4 conversion and (b) the C_2 selectivity. BaO and La_2O_3 were combined in three different ways. BaO–La_2O_3: Single-layer bed of slurry-mixed powder; BaO→La_2O_3: Two-layer bed with BaO powder placed in the influent side; La_2O_3→BaO: Two-layer bed with La_2O_3 powder placed in the influent side.

To start, the single-layer beds of La_2O_3 and BaO exhibited reasonably high C_2 yields even without active elements. This means that these supports are active for OCM by themselves while M1–M3 play a promoting role, which is in line with the observed consistent behaviors of various BaO- and La_2O_3-supported catalysts in response to the reaction temperature (Figure 2.4). The two-layer bed of La_2O_3→BaO showed almost the identical performance to the single-layer bed of La_2O_3 at each temperature. This is because La_2O_3, which is sufficiently active at 700 °C, fully consumed O_2 regardless of the temperature. In fact, a similar result was reported by Liang et al. when a large amount of a La–Ce mixed oxide was used in the first layer[39]. Contrary to La_2O_3→BaO, BaO→La_2O_3 revealed an advantage of using a two-layer bed. Its

performance at 700 °C was almost the same as that of the La_2O_3 single-layer bed, where BaO was inactive and hardly affected the operation of the La_2O_3 layer. However, the C_2 yields at 750 and 800 °C significantly exceeded those of La_2O_3 and BaO. It seemed that BaO partly consumed CH_4 and O_2 in a selective fashion, and this resulted in more selective catalysis in the following La_2O_3 layer. Note that lower concentrations of CH_4 and O_2 generally enhance the C_2 selectivity[18,40]. The performance at 900 °C was similar to that of BaO due to the complete consumption of O_2 in the first layer. The performance of the BaO–La_2O_3 mixed support was in between those of BaO and La_2O_3 at 700 °C, since the dilution of La_2O_3 by BaO partly suppressed the ignition. On the other hand, far greater C_2 yields were observed at 750 and 800 °C as compared to BaO and La_2O_3. It was considered that La_2O_3 acted as a heater, lowering the operating temperature of highly selective BaO[39]. The loss of the selectivity at 850 °C supports this consideration, while BaO alone still kept high selectivity at the same temperature. In summary, I demonstrated that the activity of La_2O_3 and the selectivity of BaO can be complementarily used when BaO and La_2O_3 are layered in this order, and when BaO and La_2O_3 are mixed within a single bed. Both of the combination methods led to improved C_2 yields, while the mechanisms of the improvements were seemingly different.

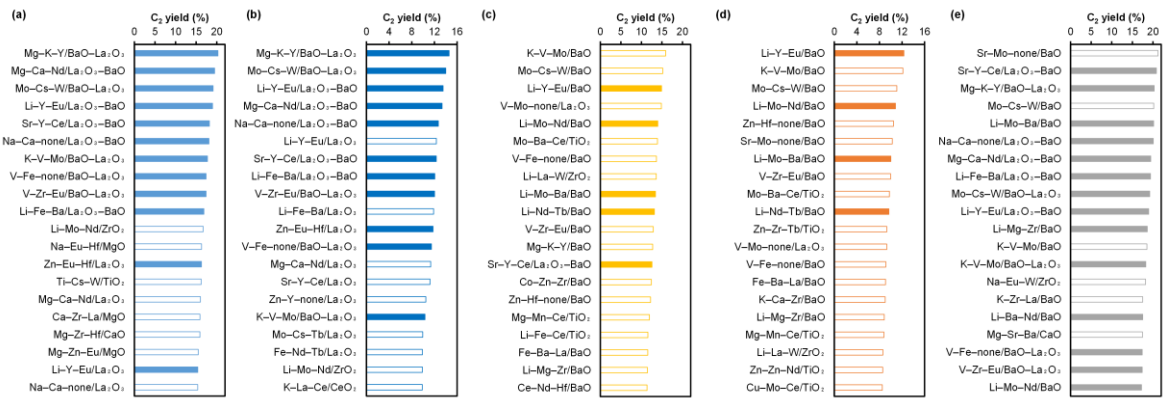


Figure 2.12. Best 20 catalysts at (Temp., CH_4/O_2 ratio) = (a) (700 ˚C, 2 mol/mol), (b) (700 ˚C, 6 mol/mol), (c) (900 ˚C, 2 mol/mol), and (d) (900 ˚C, 6 mol/mol). (e) The same in the absence of conditional restrictions. Filled and unfilled bars correspond to catalysts derived in this chapter and catalysts that were present in the original dataset, respectively.

To the end, the effectiveness of the heuristics-based catalyst design is illustrated in terms of the C_2 yield improvement from the original data. For this, the best 20 catalysts are extracted at specified conditions from a combined dataset of the 20 catalysts derived in this chapter and the 291 catalysts in the original dataset (Figure 2.12), where the new catalysts and the original catalysts are displayed as filled and unfilled bars, respectively. It can be seen that many of the new catalysts are ranked within the top 20 at each of the specified conditions (Figure 2.12a–d) and even without specifying conditions (Figure 2.12e), which clarifies the effectiveness of the heuristics-based catalyst design in improving the low-temperature activity, the high-temperature selectivity, as well as the maximum C_2 yield. The effectiveness is more evident at 700 °C, where more than half of the entries come from the new catalysts. This fact suggests that improving the selectivity of highly active catalysts is more feasible than improving the activity of selective catalysts without deteriorating the selectivity. The observations of Figure 2.12e are also confirmed in the scatter plot in Figure 2.13, where the horizontal axis is CH_4 conversion and the vertical axis is C_2 selectivity.

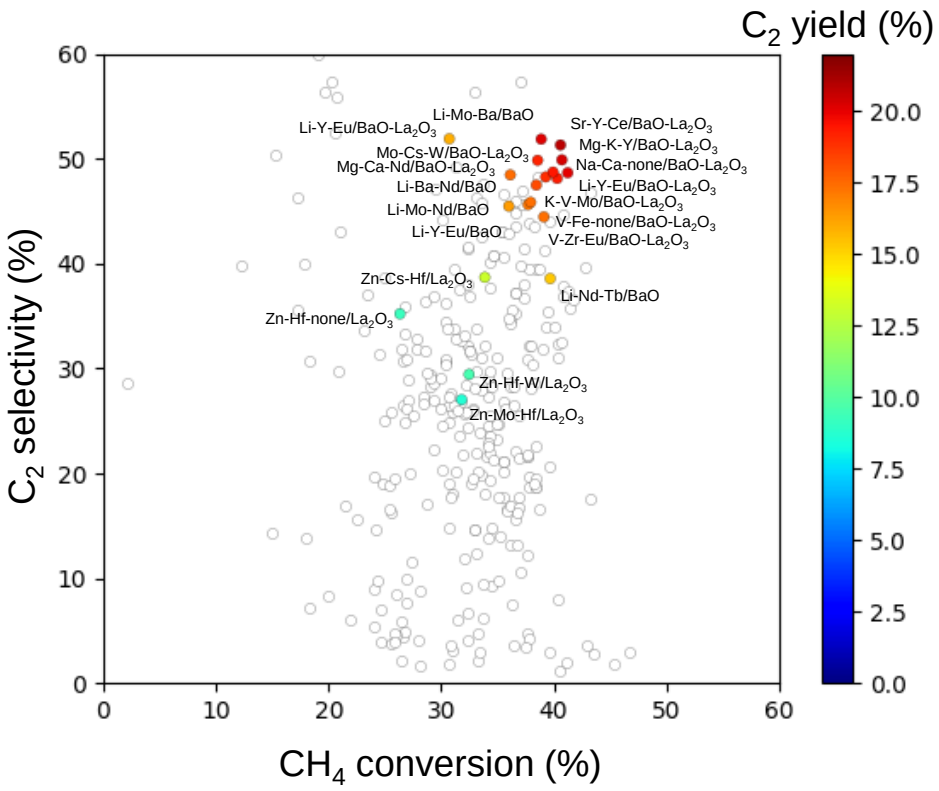


Figure 2.13. Scatter plot of the best C_2 yield data points for the CH_4 conversion and the C_2 selectivity. Filled makers represent catalysts derived in this chapter, while unfilled makers represent those in the original dataset.

2.4. Conclusions

In this chapter, catalyst design heuristics were extracted for methane activation at mild conditions and for suppression of the secondary oxidation at severe conditions by analyzing a dataset previously obtained in our group.[19] These heuristics were then combined to design better performing catalysts. The major conclusions are summarized below:

- Catalyst design guidelines vary greatly depending on the reaction temperature. At low temperatures, a combination of basic oxide-forming elements and La_2O_3 as a support tends to result in high performance. At high temperatures, BaO, which can in-situ form a stable carbonate, and/or a combination between basic oxide-forming elements and early transition metals, which can form an oxometalate anion, tends to result in high performance.

- Catalyst design guidelines are rather specific to the properties of supports, i.e. it is not possible to dramatically alter the catalytic behavior by the selection of active elements when the same support is used.

- A mixed support is an effective strategy to pursue both the activity and selectivity. In fact, by combining La_2O_3 and BaO, I succeeded in simultaneously improving the CH_4 conversion, C_2 selectivity, and C_2 yield.

- Various methods of combining La_2O_3 and BaO were evaluated, and it was found that the mixed support led to improved performance based on a different mechanism from that of a two-layer bed. It was considered that La_2O_3, which possesses an ability for low-temperature methane activation, works as a heater for highly selective BaO, and reduces its apparent operational temperature.

In the implementation of catalyst informatics, it is important to have data that is acquired uniformly across both materials and process conditions without placing anthropogenic bias. I have demonstrated here that once such data is obtained, various insights useful for catalyst development can be obtained by how the data is projected onto a materials space, i.e., how subsets are extracted.

In order to control complex chemical reactions, the development of solid catalysts often premises combinatorial design, which is effective to balance different aspects of the performance. Hence, the strategy of finding heuristics for improving individual aspects of the performance from data analysis and combining them in the framework of combinatorial design is believed to be widely applicable to any other catalysis as long as combinatorial design is effective.

References

[1] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. I. Shimizu. *ACS Catal.* **2020**, *10*, 2260–2297.

[2] A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, R. Fushimi. *ACS Catal.* **2018**, *8*, 7403–7429.

[3] T. Zhou, Z. Song, K. Sundmacher. *Engineering* **2019**, *5*, 1017–1026.

[4] G. Takasao, T. Wada, A. Thakur, P. Chammingkwan, M. Terano, T. Taniike. *ACS Catal.* **2019**, *9*, 2599–2609.

[5] G. Takasao, T. Wada, A. Thakur, P. Chammingkwan, M. Terano, T. Taniike. *J. Catal.* **2020**, *394*, 299–306.

[6] X. Jia, A. Lynch, Y. Huang, M. Danielson, I. Lang'at, A. Milder, A. E. Ruby, H. Wang, S. A. Friedler, A. J. Norquist, J. Schrier. *Nature* **2019**, *573*, 251–255.

[7] D. G. Brown, M. M. Gagnon, J. Boström. *J. Med. Chem.* **2015**, *58*, 2390–2405.

[8] U. Zavyalova, M. Holena, R. Schlögl, M. Baerns. *ChemCatChem* **2011**, *3*, 1935–1947.

[9] B. Zadrozny. *Proceedings, Twenty-First Int. Conf. Mach. Learn. ICML 2004* **2004**, 903–910.

[10] H. Jiang, O. Nachum. *arXiv* **2019**, *108*.

[11] G. E. Keller, M. M. Bhasin, "Synthesis of Ethylene via Oxidative Coupling of Methane. I. Determination of Active Catalysts."

[12] E. V. Kondratenko, T. Peppel, D. Seeburg, V. A. Kondratenko, N. Kalevaru, A. Martin, S. Wohlrab. *Catal. Sci. Technol.* **2017**, *7*, 366–381.

[13] Y. Gao, L. Neal, D. Ding, W. Wu, C. Baroi, A. M. Gaffney, F. Li. *ACS Catal.* **2019**, *9*, 8592–8621.

[14] S. Arndt, T. Otremba, U. Simon, M. Yildiz, H. Schubert, R. Schomäcker. *Appl. Catal. A Gen.* **2012**, *425–426*, 53–61.

[15] Y. Gambo, A. A. Jalil, S. Triwahyono, A. A. Abdulrasheed. *J. Ind. Eng. Chem.* **2018**, *59*, 218–229.

[16] J. Ohyama, T. Kinoshita, E. Funada, H. Yoshida, M. Machida, S. Nishimura, T. Uno, J. Fujima, I. Miyazato, L. Takahashi, K. Takahashi. *Catal. Sci. Technol.* **2020**, 17–21.

[17] L. Olivier, S. Haag, H. Pennemann, C. Hofmann, C. Mirodatos, A. C. van Veen. *Catal. Today* **2008**, *137*, 80–89.

[18] T. N. Nguyen, T. T. P. Nhat, K. Takimoto, A. Thakur, S. Nishimura, J. Ohyama, I. Miyazato, L. Takahashi, J. Fujima, K. Takahashi, T. Taniike. *ACS Catal.* **2020**, *10*, 921–932.

[19] T. N. Nguyen, S. Nakanowatari, T. P. N. Tran, A. Thakur, L. Takahashi, K. Takahashi, T. Taniike. *ACS Catal.* **2021**, 1797–1809.

[20] S. L. S. Kuś, M. Otremba, M. Taniewski. *Fuel* **2003**, *82*, 1331–1338.

[21] L. Pirro, A. Obradović, B. D. Vandegehuchte, G. B. Marin, J. W. Thybaut. *Ind. Eng. Chem. Res.* **2018**, *57*, 16295–16307.

[22] S. Wang, S. Li, D. A. Dixon. *Catal. Sci. Technol.* **2020**, *10*, 2602–2614.

[23] M. S. Palmer, M. Neurock, M. M. Olken. *J. Am. Chem. Soc.* **2002**, *124*, 8452–8461.

[24] V. R. Choudhary, S. A. R. Mulla, V. H. Rane. *J. Chem. Technol. Biotechnol.* **1998**, *72*, 125–130.

[25] R. C. Schucker, K. J. Derrickson, A. K. Ali, N. J. Caton. *Appl. Catal. A Gen.* **2020**, *607*, 117827.

[26] Z. Q. Wang, D. Wang, X. Q. Gong. *ACS Catal.* **2020**, *10*, 586–594.

[27] C. Karakaya, H. Zhu, B. Zohour, S. Senkan, R. J. Kee. *ChemCatChem* **2017**, *9*, 4538–4551.

[28] J. Wu, S. Li. *J. Phys. Chem.* **1995**, *99*, 4566–4568.

[29] S. Ji, T. Xiao, S. Li, L. Chou, B. Zhang, C. Xu, R. Hou, A. P. E. York, M. L. H. Green. *J. Catal.* **2003**, *220*, 47–56.

[30] H. W. Rahn, C. J. Sindlinger. *US Pat.* 2915366A, **1956** (to Columbia Southern Chemical Corp).

[31] S. Hou, Y. Cao, W. Xiong, H. Liu, Y. Kou. *Ind. Eng. Chem. Res.* **2006**, *45*, 7077–7083.

[32] D. Schweer, L. Meeczko, M. Baerns. *Catal. Today* **1994**, *21*, 357–369.

[33] A. Kooh, J. L. Dubois, H. Mimoun, C. J. Cameron. *Catal. Today* **1990**, *6*, 453–462.

[34] S. Pak, J. H. Lunsford. *Appl. Catal. A Gen.* **1998**, *168*, 131–137.

[35] J. Fujima, Y. Tanaka, I. Miyazato, L. Takahashi, K. Takahashi. *React. Chem. Eng.* **2020**, *5*, 903–911.

[36] Q. Sun, J. I. Di Cosimo, R. G. Herman, K. Klier, M. M. Bhasin. *Catal. Letters* **1992**, *15*, 371–376.

[37] C. Shi, Q. Sun, H. Hu, R. G. Herman, K. Klier, I. E. Wachs. *Chem. Commun.* **1996**, 663–664.

[38] W. P. Schammel, J. Wolfenbarger, M. Ajinkya, J. McCarty, J. M. Cizeron, S. Weinberger, J. D. Edwards, D. Sheridan, E. C. Scher, J. McCormick. *US Pat. 20140107385A1* **2014** (to Siluria Technologies, inc.)

[39] W. Liang, S. Sarsani, D. West, A. Mamedov, I. Lengyel, H. Perez, J. Lowrey. *Catal. Today* **2018**, *299*, 60–66.

[40] H. Wang, Y. Cong, W. Yang. *Catal. Today* **2005**, *104*, 160–167.

[41] J. Sun, J. W. Thybaut, G. B. Marin. *Catal. Today* **2008**, *137*, 90–102.

Chapter 3

Adaptive sampling for catalyst discovery

incorporating artificial serendipity

Abstract:

This chapter introduces a novel algorithmic approach to catalyst development, leveraging elemental substitution to guide the creation of a sophisticated catalyst recommender system. Integrating adaptive sampling, this system efficiently predicts catalyst performance, subsequently verified through high-throughput experimentation. A significant innovation of this work is a catalyst serendipiter system, which anticipates the occurrence of serendipitous catalysts within the development process—a phenomenon previously considered a product of chance. This results reveal that the serendipiter system increases the likelihood of serendipity to 43.8%, significantly outperforming the baseline adaptive sampling serendipity rate of 8.1%. Through this predictive model, I discovered catalysts with enhanced performances without any known performant counterparts, demonstrating the model's effectiveness in identifying novel catalyst combinations. This approach not only provides a structured framework for catalyst exploration and exploitation but also harnesses the unpredictable nature of serendipity, offering a new frontier in the systematic development of catalysts.

Keywords: Adaptive sampling, serendipity, evidence theory, oxidative coupling of methane, high-throughput experimentation.

3.1. Introduction

The challenge in developing solid catalysts often lies in diversifying their elemental combinations to achieve multifunctionality[1]. Catalytic reactions involve multiple elementary steps, and even a partial change in a catalyst's composition can affect all these stages[2,3]. Additionally, it's not only the individual elements that are important; the interactions between these elements can lead to unexpected effects on each elementary reaction, a phenomenon known as synergistic effects[4–6]. As the number of components in a catalyst increases, the potential combinations grow exponentially, making it impractical to explore all possible catalysts due to time and resource constraints. Researchers in solid catalysts frequently rely on trial-and-error methods to identify optimal catalysts from a vast pool of candidates[7,8]. Their approach is informed by a combination of past experience, insights from literature and analyses, and curiosity about unexplored combinations. Serendipity, or the unexpected discovery occurring during trial and error, has been a key driver in significant advancements for catalyst development[9–11]. However, these conventional manual methodologies, despite their high costs, do not guarantee success. Furthermore, as the complexity of the system increases, making groundbreaking developments becomes increasingly challenging. Moreover, knowledge-based approaches that aim for rational catalyst design find their applicability diminishing with the increasing complexity of the target system[1]. Developing a reproducible methodology that systematizes and streamlines the trial-and-error process and effectively harnesses serendipity could not only reduce costs but also expedite the achievement of objectives. Such a development could significantly speed up catalyst development and enhance the understanding of the principles governing catalyst design.

Catalyst informatics leverages statistical analysis and machine learning to accelerate catalyst development and understanding their mechanisms[12–15]. The advent of machine learning techniques like adaptive sampling and black-box optimization has transformed the approach to informatics in recent years[16,17]. These techniques are adept at efficiently labeling samples in previously undefined regions or iteratively converging on an optimal composition. This is achieved by incorporating unknown data points suggested by a surrogate model—trained on existing data—into the training set and then updating the model with this enhanced dataset. A crucial technique in reinforcement learning applied within this context is the exploration/exploitation trade-off[18]. Exploration is considered as a 'high-risk, high-return' strategy, where one learns about the potential reward of an untried option through direct experimentation. Conversely, exploitation is a 'low-risk, low-return' strategy that involves selecting the option with the highest predicted reward based on current knowledge. Effective decision-making in actual catalyst development often requires a balance between these strategies, especially when operating under constraints of limited time, resources, and budgets. When modeling the catalyst development process itself, incorporating the exploration/exploitation concept is essential. Several materials informatics (MI) studies exemplify the application of this concept, utilizing methods like Bayesian optimization[19,20], best-arm identification[21], and active learning[16,22,23]. However, despite these advancements, the occurrence of serendipity, which has historically been crucial in catalyst and material discovery, is still underrepresented[24,25]. Additionally, while some studies have acknowledged the importance of combinatorial effects in catalyst design, the causal modeling of these synergistic effects is still a challenge[26,27]. Developing an algorithm that can assess predictive uncertainties, elemental combination effects, and serendipitous interactions

would represent a significant breakthrough. Coupling with an adaptive sampling loop and such an algorithm could refine the precision of catalyst development methodologies.

In this chapter, I present a catalyst recommendation system that is adept at handling uncertainty and considers synergistic effects in its predictive model. This system is used in conjunction with adaptive sampling. Additionally, by gathering and incorporating unexpected findings from the system during the adaptive sampling process as training data, an another-layer machine learning has been introduced. This new layer is specifically designed to identify the conditions under which serendipitous discoveries are likely to occur, which this has been named the 'catalyst serendipiter system.' By integrating the recommendation system with the serendipiter system, I have established a reproducible method for catalyst development. This approach effectively systematizes and streamlines both the trial-and-error process and the harnessing of serendipitous discoveries, which are common in traditional catalyst development.

3.2. Methods

3.2.1. Catalyst preparation and evaluation

All catalysts were prepared using a parallelized co-impregnation method and evaluated with a high-throughput screening system[28]. The catalysts comprised quaternary compositions of M1-M2-M3/support, using elements selected from a predefined library of supported elements and oxides. These are briefly mentioned here; for detailed information, please refer to our previous publications.

The supported elements library includes 27 elements: 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, and W, with an additional 'none' option to indicate no selection, totaling 28 options. The oxide library consists of nine species: MgO, Al_2O_3, SiO_2, CaO, TiO_2, ZrO_2, BaO, La_2O_3, and CeO_2. As precursors to each of these, $LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, 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), $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, $Pd(OAc)_2$, $CsNO_3$, $Ba(NO_3)_2$, $La(NO_3)_3{\cdot}6H_2O$, $Ce(NO_3)_3{\cdot}6H_2O$, $Nd(NO_3)_3{\cdot}6H_2O$, $Eu(NO_3)_3{\cdot}5H_2O$, $Tb(NO_3)_3{\cdot}5H_2O$, $Hf(OEt)_4$, $(NH_4)_{10}H_2(W_2O_7)_6$, MgO, γ-Al_2O_3, SiO_2, $Ca(OH)_2$, TiO_2, ZrO_2 , $Ba(OH)_2$-$8H_2O$, La_2O_3 and CeO_2 were used. These reagents were purchased from Sigma-Aldrich, Kanto Chemical, Wako Pure Chemical Industries, Alfa-Aesar or Sumitomo Chemical Industry. Each element from M1 to M3 was loaded onto 1.0 g of support at 0.371 mmol. The carrier powder was impregnated with an aqueous solution containing the metal precursor at 50 °C for 6 h, followed by drying under reduced pressure at 90 °C for 4 h. The dried material was then placed in ceramic cups and calcined at 1000 °C for 3 h under air.

All prepared catalysts underwent testing in a high-throughput screening system under 135 distinct conditions, varying temperature (700 °C, 750 °C, 800 °C, 850 °C, and 900 °C), gas contact time (0.75 s, 0.50 s, and 0.38 s), CH_4/O_2 ratios (2, 4, and 6), and carrier gas partial pressures (0.15, 0.4, 0.7)[28]. The screening system consisted of a gas mixer, distributor, quartz reaction tubes, an electric furnace, an autosampler, and a quadrupole mass spectrometer (QMS)[29]. The mixer controlled the flow rates of methane, oxygen, and the carrier gas argon, distributing them evenly into 20 reaction tubes. The reaction tubes had a 4 mm internal diameter at the inlet and 2 mm at the outlet. Inside the electric furnace, 20 tubes each contained a 10 mm catalyst bed. The outlet gases were sampled in sequence by the autosampler and analyzed by the QMS. Argon was used as the internal standard, and the calibration curve was derived from signal ratios at specific argon partial pressures and those of the target gases. Using these ratios, the relative pressures of the gases compared to argon was back-calculated. This information, along with the theoretical argon partial pressure from the set reaction conditions, allowed us to calculate methane conversion and product yield. These values were then used to determine the selectivity.

$$CH_4\ conversion\ (\%) = \frac{P_{CH_4/Ar_{theoritical}} - P_{CH_4/Ar_{measured}}}{P_{CH_4/Ar_{theoritical}}} \times 100$$

$$C_2H_4\ or\ C_2H_6\ yield\ (\%) = \frac{2 \times P_{C_2H_4\ or\ C_2H_6/Ar_{measured}}}{P_{CH_4/Ar_{theoritical}}} \times 100$$

$$C_2\ yield\ (\%) = C_2H_4\ yield + C_2H_6\ yield$$

$$CO\ or\ CO_2\ yield\ (\%) = \frac{P_{CO\ or\ CO_2/Ar_{measured}}}{P_{CH_4/Ar_{theoritical}}} \times 100$$

$$X\ selectivity\ (\%) = \frac{X\ yield}{CH_4\ conversion} \times 100$$

3.2.2. Prediction accuracy metric calculation

When comparing predicted and measured values in building a machine learning model, there are four possible patterns of consistency: (1) Both predictions and measurements are positive; (2) predictions are positive but measurements are negative; (3) predictions are negative but measurements are positive; and (4) both are negative. Correct predictions occur in cases (1) and (4), while incorrect predictions are seen in (2) and (3). These outcomes are referred to as (1) True Positive (TP), (2) False Positive (FP), (3) False Negative (FN), and (4) True Negative (TN), respectively. These categories form the basis for calculating indicators of a machine learning model's accuracy. In this chapter, the performance metrics used - namely, accuracy, precision, recall, and the F1 score - are derived from these fundamental outcomes.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

The general meanings of these indicators are as follows: 'Accuracy' measures the model's overall predictive correctness across both classes; 'Precision' assesses the model's accuracy specifically for the positive (target) class predictions; 'Recall' gauges the model's ability to correctly identify all relevant instances of the positive (target) class (i.e., its capacity to avoid false negatives); and the 'F1 Score' is the harmonic mean of precision and recall, providing a balance between them.

3.2.3. Generating phylogenetic tree using the neighbor-joining method

The neighbor-joining method is a bottom-up clustering approach utilized for generating phylogenetic trees in molecular biology and bioinformatics[30]. This method begins by calculating the pairwise distances between all sequences (or catalysts), which are then arranged into a distance matrix. A phylogenetic tree is incrementally constructed by repeatedly connecting the pair of entities with the smallest distance. The ultimate goal is to produce a phylogenetic tree with the minimum possible sum of branch distances[31].

For the construction of the phylogenetic tree, I utilized the scikit-bio library within the Python programming environment to apply the neighbor-joining method. Subsequently, the ETE3 library was employed for the visualization of the phylogenetic tree.

3.2.4. Implementation of a catalyst serendipiter system

In implementing the catalyst serendipiter system, four distinct learner characterizations were trained in parallel, and their predictions were utilized as descriptors. Each learner was trained using catalyst composition as the explanatory variable and catalyst performance as the response variable. To designate catalyst performance as the objective for the classifiers, a threshold value was established and subsequently binarized. This threshold was set at a 13% C_2 yield; catalysts exceeding this value were labeled as 'positive,' while those falling short were labeled as 'non-positive.' For the explanatory variable of catalyst composition, one-hot encoding was applied. Below is a brief description of the training instruments used.

Logistic regression, despite its nomenclature, is a linear classifier for classification tasks, predicting outputs between 0 and 1 from continuous input data. It is commonly used in binary classification problems, with its outputs derived via a sigmoid function, allowing interpretation as probabilities. However, logistic regression's susceptibility to overfitting may increase with high-dimensional datasets.

Support vector machines (SVM) are algorithms designed to identify optimal data separation boundaries. They can handle both linear and non-linear datasets by utilizing the 'kernel trick.' SVMs are known for their effectiveness with small datasets.

Random forests, an ensemble method, combines multiple decision trees to enhance prediction accuracy and mitigate the risk of overfitting. This approach effectively captures feature interactions and non-linear relationships within the data.

Naive Bayes is a probabilistic classifier based on Bayes' theorem. Apt for tasks like spam mail detection due to its speed in training and prediction, it operates under the 'naive' assumption of feature independence. This assumption, while facilitating efficiency, may not yield high performance across all datasets.

These algorithms were implemented using the Python library, scikit-learn. Hyperparameters were optimized through grid search. For the catalyst serendipiter system, descriptors included the catalyst recommender system predictions, the four traditional learners' predictions for each candidate catalyst, and the one-hot encoding of catalyst compositions. The correspondence between the catalyst recommender system predictions and the actual measurements served as the objective variables. It is important to note that the hyperparameters for each learner were fine-tuned for comparative analysis.

3.3. Results

In the context of adaptive sampling, it's crucial for the learning model to quantify uncertainty effectively to enable both exploration and exploitation strategies. To this end, I have developed a catalyst recommendation system, depicted in figure 3.1a, which employs evidence theory[32]. Evidence theory is a generalization of the Bayesian approach for dealing with situations of incomplete information and data[33,34]. Unlike Bayesian theory, which assigns weights to individual events, evidence theory allocates non-negative weights to subsets of possibilities within the framework of discernment. In the field of solid catalysts, the effectiveness of an active component can be significantly influenced by the presence of other elements[5,6]. The characteristics of evidence theory are particularly beneficial in depicting this phenomenon from a machine learning perspective. For example, consider a situation where there are two conflicting observations about a catalyst's performance, one positive and one negative. In Bayesian terms, this would be treated as two mutually exclusive events: 'the catalyst is positive' and 'the catalyst is negative.' Predictions would be made accordingly, but the contradicting observations would cancel each other out, leading the probability to revert to its prior value. Conversely, evidence theory considers three possibilities: the fundamental hypothesis that a catalyst is positive, that a catalyst is negative, and a combined hypothesis that acknowledges the catalyst could be either positive or negative. Here, 'the catalyst is positive or negative' represents uncertainty. The conflicting observations, in evidence theory, provide a measure of belief for each of the first two hypotheses while reducing confidence in the uncertainty. This approach allows for distinguishing between uncertainty due to contradictory information and uncertainty from a lack of information. Such a distinction is crucial

for maintaining specific catalytic information that arises from synergistic effects, hence preserving the nuances of catalyst performance prediction.

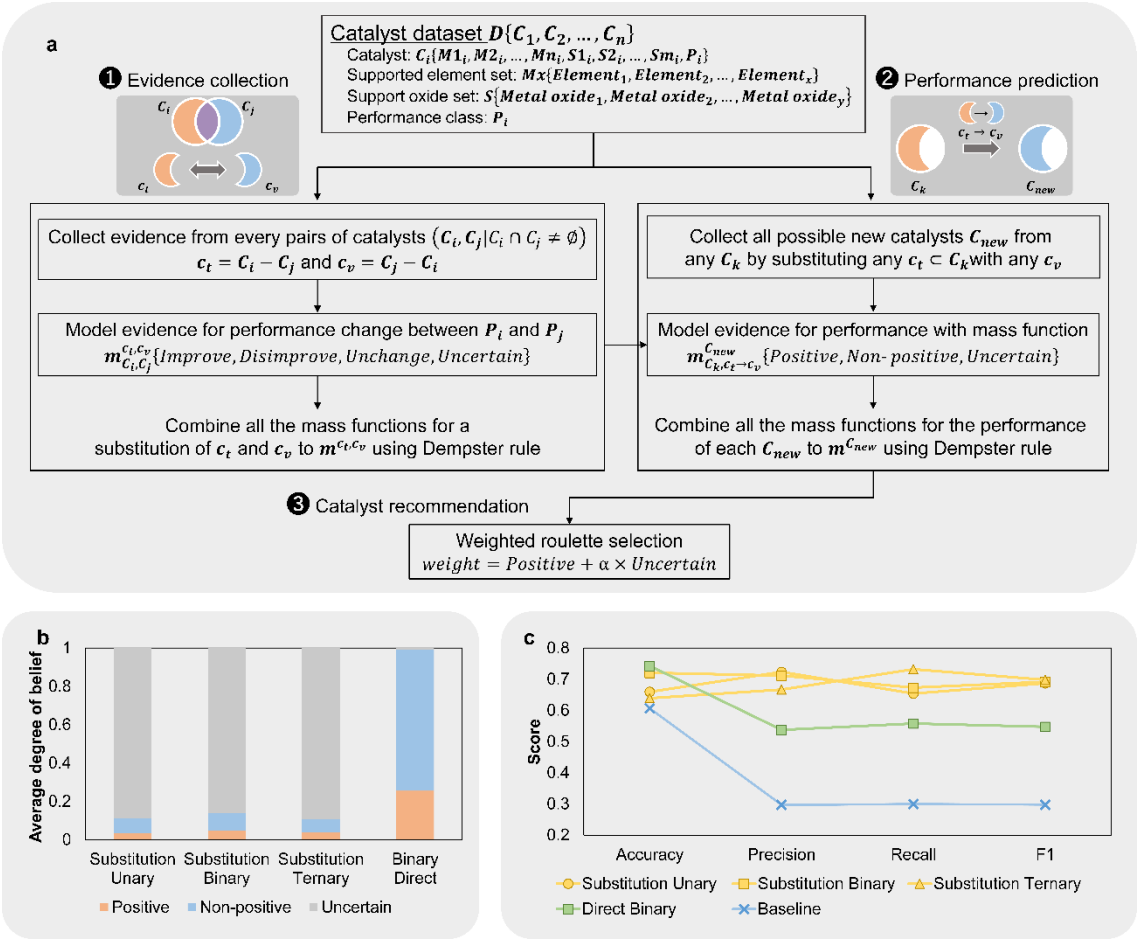


Figure 3.1. Implementation of catalyst recommender system. **a,** Developed algorithm for catalyst recommendation. Two catalysts, C_i and C_j, are compared. If they share common components, observations on performance changes due to the substitution of differing components are collected as evidence, denoted as $m^{ct, cv}$. This evidence and the third catalyst, C_k, are then used to make performance predictions. The performances of all catalysts within the configuration space are predicted, and weights calculated from these predictions are used to recommend the next catalyst for evaluation. **b.** Efficiency of performance prediction for each algorithm. The graph plots the average beliefs for both positive and non-positive across all 36,540 catalysts in the configuration space. These predictions are based on the 376 catalysts that have been evaluated at our previous study. **c,** Comparison of prediction abilities for all developed algorithms when varying the number of elements substituted or be common. "Accuracy" refers to the overall prediction rate, including both positive and non-positive results. "Precision" is defined as the proportion of true positives among the predicted positives, while "Recall" is the proportion of true positives among the actual positives. “F1 score” is the harmonic mean of precision and recall, serving as a measure of prediction accuracy specifically for positive cases.

In this approach to modeling catalyst development using evidence theory, I sought to replicate and algorithmize the human thought process used in developing new catalysts through elemental substitution[32]. This process involves two crucial steps, each crafted into an algorithm. The step 1 is 'compiling data on performance variations resulting from the substitution of non-common elements within two catalysts that share some elements in their composition.' This step aims to generate insights that can inform future catalyst designs. The step 2 is 'projecting the performance of a novel catalyst by applying these insights to existing data.' Both of these steps are meticulously structured into a comprehensive algorithm, which is described in detail in the subsequent paragraph.

The structure of the algorithm used in the catalyst recommender system is detailed in figure 3.1a. The process begins with the system categorizing the performance of evaluated catalysts into bins, based on a predefined threshold value. For clarity, I will describe the binarized implementation used in practice, where catalysts are classified as either 'positive' or 'non-positive' according to the threshold. The system then compares the compositions of two catalysts, C_i and C_j. If they share a common element, the system assesses the impact on performance resulting from the substitution of differing elements, c_t and c_v (e.g., improve, disimprove, or no change). This assessment is conducted using the belief mass function $m_{C_i,C_j}^{c_t,c_v}$ for the ct, cv substitution. This part corresponds to step 1 of the algorithm. Initially, all basic belief masses, m, are set to 1 for uncertainty and 0 otherwise. An observation increases the belief in the corresponding certainty hypothesis by a value x and decreases the uncertainty by the same amount. The optimal value for x is determined through cross-validation. Step 2

begins with the application of the belief mass function m^{c_t,c_v} for catalyst C_k, which contains element c_t. This application calculates the belief mass function $m^{C_{new}}_{C_k,c_t \to c_v}$ for the performance of a new catalyst, C_{new}, after the replacement using the original performance of C_k and m^{c_t,c_v}. It's important to note that when similar evidence is obtained from different observations for both the belief mass function in steps 1 and 2, these beliefs are combined using Dempster's rule of combination[33]. Beliefs related to both 'positive' and 'uncertain' outcomes are considered as weights for selection and exploration, respectively. These weights are calculated for all potential catalyst candidates. Catalyst recommendations are then made using a roulette selection method, where the calculated 'weight' forms the basis for selection. The balance between exploitation and exploration, as determined by these weights, is adjusted according to the specific goals and context of the research.

The recommender system developed in this chapter was trained using catalyst reaction data on the oxidative coupling of methane (OCM) reaction, comprising results from 376 catalysts[28,35–37]. OCM is a reaction that directly converts methane, the main component of natural gas, into ethylene, a key petrochemical material[38]. The difficulty of OCM lies in the fact that the reactants (methane) are chemically more stable than the main products (ethylene and ethane), leading to a kinetic trade-off [1,39–41]. The reaction must be conducted under conditions harsh enough to convert methane, but mild enough to avoid deep oxidation of ethylene. Our dataset for OCM includes results using a quaternary catalyst system consisting of M1-M2-M3/support[28]. M1 to M3 were chosen from a 'supported element set' of 27 elements (from Lithium to Tungsten) plus "none," allowing for overlap, and the support was selected from a 'support metal set' of nine different oxides. These oxides include three with acidic surfaces (TiO_2, ZrO_2, and CeO_2), four with basic surfaces (MgO, CaO, BaO, and La_2O_3),

and two with both acidic and basic properties (Al_2O_3 and SiO_2). The choice of components was informed by their frequent mention in previous OCM studies or their similar properties but lesser usage. Out of 36,540 potential catalyst combinations, 300 were randomly selected, and 76 had been proposed by machine learning using the 300 as training data in our previous study[28,35–37]. All catalysts were evaluated under 135 uniform reaction conditions using a high-throughput screening system[29]. The condition that yielded the highest C_2 yield for each catalyst was identified as a unique point for system training. A threshold value of 13% for C_2 yield was set for performance dichotomization, based on the highest yield achieved without a catalyst under these conditions. Catalysts exceeding this threshold were classified as positive, and those below as non-positive[28]. The belief mass function assigned a belief degree of 0.2 for a single observation. Considering that 51 out of the 300 randomly evaluated catalysts were positive, a similar proportion of positive catalysts was assumed within the entire candidate population. This proportion (approximately $0.2 \approx 51/300$) was used as the value of α in calculating the weights[28].

In the targeted quaternary catalysts, the dimensions that can be considered for elemental combinations to be substituted include unary substitution, binary substitution, and ternary substitution. While all these could be employed simultaneously, integrating and employing the prediction results of all substitution dimensionalities should be avoided when generalizing this system for all types of catalyst development. For instance, if the system were to be applied to develop a catalyst for a ten-element system, there would be approximately 0.8 billion nine-element combinations, assuming the supported element and support library remain the same as in this chapter. Furthermore, the number of feasible substitutions among the nine-element combinations would surge to about 23 quadrillion. If eight-element and seven-element combinations and so on

were to be considered additionally, the system would become impractical from a computational cost standpoint.

To ascertain the optimal dimensionality for each number of terms, the average acquisition certainty of the performance belief functions for each of the 36,540 candidate catalysts was compared. These results are displayed in figure 3.1b. Although the focus of this chapter has been primarily on performance estimation methods based on the substitution of non-common elements, consideration was also given to methods attributing performance directly to the common part, rather than the non-common part. Consequently, results for the former are presented as "substitution" and for the latter as "direct." Please note that in figure 3.1b, only binary results for the direct algorithm are shown. For results concerning unary and ternary approaches, as well as a detailed flowchart of the algorithm, reference is made to figure 3.2.

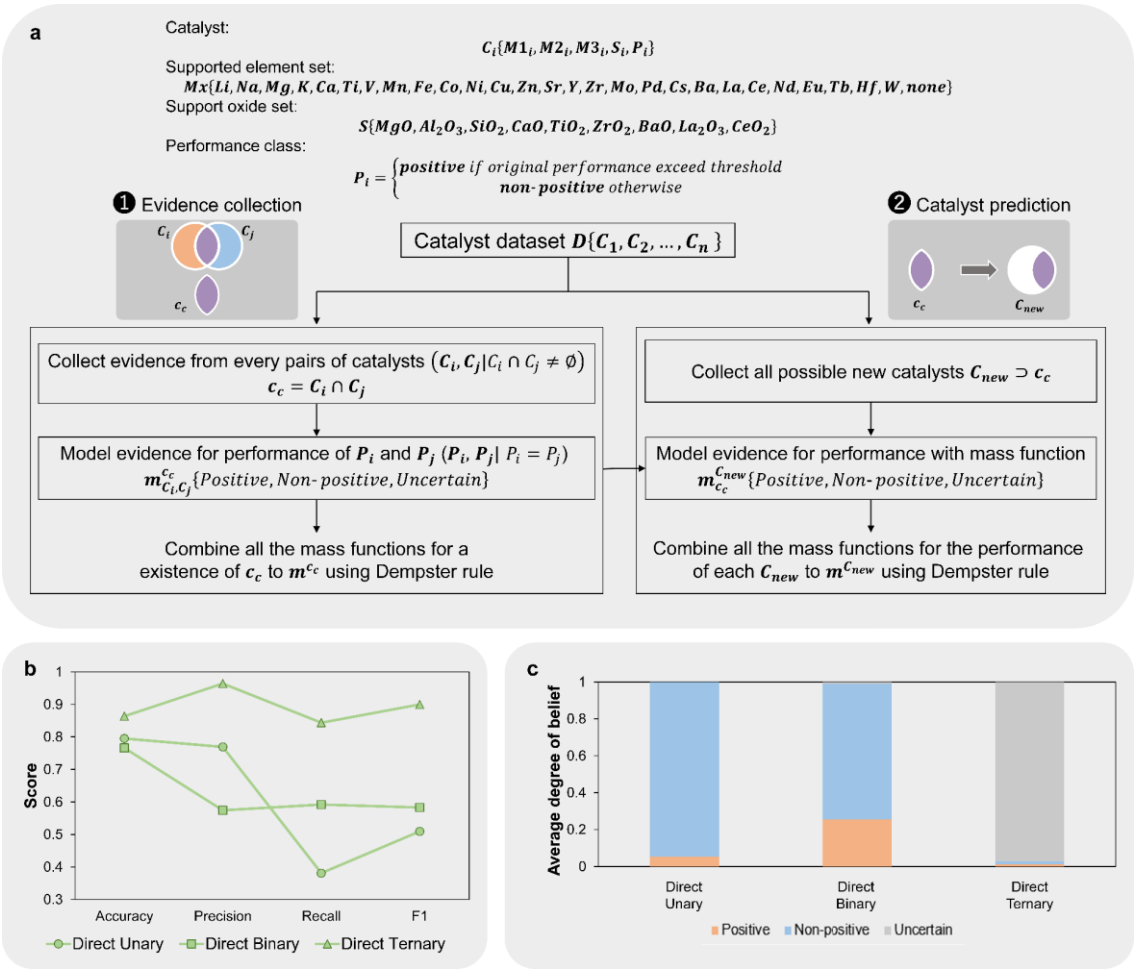


Figure 3.2. Implementation of catalyst recommender system based on direct algorithm for comparison with the substitution algorithm. **a.** Developed algorithm for catalyst recommendation based on direct estimation method. Two catalysts, C_i and C_j, are compared. If they share common components and their performances are the same class, evidence will be collected as the common part c_c is having the performance. This evidence and the third catalyst, C_k, which is having its common part c_c in the composition are then used to make performance predictions. **b.** Comparison of prediction abilities for all developed algorithms when varying the number of elements substituted or be common. **c.** Efficiency of performance prediction for each algorithm.

A comparison within the substitution algorithms reveals mostly same in the total amount of certainty acquired between unary and ternary substitution. This is attributed to the inherent connection between unary and ternary substitutions; unary substitution requires common elements of ternary to gather substitution evidence, resulting in fewer opportunities to acquire certainty and thus being less efficient in certainty acquisition. In contrast, for operating substitution evidence, only unary elements are needed as the source, offering more opportunities for evidence manipulation and hence greater efficiency in evidence operation. Ternary substitution, on the other hand, is characterized by high efficiency in certainty acquisition but low efficiency in evidence operation. Binary substitution has respective efficiencies between unary and ternary. The total number of combinations derivable from a single catalyst is four for both unary and ternary, but six for binary, leading to more opportunities for certainty acquisition and, therefore, a higher total amount of certainty acquired. When comparing the binary results of the substitution and direct methods, it becomes evident that the direct method yields more instances of performance inference formation and almost completely eliminates uncertainty. The ratio between positive certainty and non-positive certainty is almost consistent with the class balance in the training data; only 105 out of 376 catalysts in the training data are positive. The result is like conventional machine learning that does not account for uncertainty classes.

In the analysis of substitution algorithms—unary, binary, and ternary—it was observed that the binary dimension exhibited the highest efficiency in certainty acquisition. However, the value of this finding hinges on the accompanying prediction accuracy. Therefore, the prediction accuracy of each dimension was evaluated using 10-fold cross-validation, with the results displayed in figure 3.1c. For this validation, only catalysts obtained two or more evidence during performance estimation (where

uncertain confidence was less than 0.8) were included, and the higher of the two belief masses was selected as the prediction outcome. Additionally, a baseline method was considered, where the prediction outcome was randomly determined based on the class balance of the training data. A comparison between the Baseline and the substitution algorithms revealed that the algorithms surpassed the Baseline across all metrics—accuracy, precision, recall, and F1 score. This improvement is mainly attributed to the fact that most random predictions are non-positive according to the training data's low number of positive cases. When analyzing the substitution algorithms by the number of replacement dimensions, it was found that unary substitution exhibits high precision but low recall. In contrast, ternary substitution shows the opposite trend, while binary substitution lies between the two. This pattern emerges because unary substitution only predicts a catalyst as positive when there is a strong active component performing well independently. On the other hand, ternary substitution tends to generalize, identifying specific combinations that perform well and extrapolating those observations to predict other catalysts as positive. It issues a positive decision even with minimal evidence of a positive outcome, due to its specificity and a tendency toward generalization. The same logic applies to the binary dimension in the direct algorithm, where prediction accuracy follows a similar calculation. As shown in figure 3.1b, when extending performance prediction to candidate catalysts, the direct method mirrors the class balance of the training data and show the same trend with the baseline. Here, it is the positive class for which prediction accuracy is primarily sought. Precision, recall, and the F1 score, which reflect accuracy for this class, are lower in the direct method compared to substitution method. Based on these findings, the binary substitution algorithm was chosen as the practical recommender system.

Using the trained system, adaptive sampling was carried out. In each iteration, twenty catalysts were recommended through weighted roulette sampling and then evaluated under 135 uniform reaction conditions using a high-throughput catalyst screening system. This adaptive sampling loop, driven by the system, was repeated eight times, leading to the development of 160 catalysts. Figure 3.3a shows both the belief degrees and the actual performance data of these 160 recommended catalysts at the time of their recommendation. In the recommendation for roulette sampling, a catalyst was classified as recommended by the exploitation strategy if, at the time of recommendation, its positive certainty was greater than its uncertainty multiplied by 0.2. All other recommendations were considered as stemming from the exploration strategy. Following this criterion, 59 out of the 160 catalysts were categorized as exploitation catalysts, and the remaining 101 as exploration catalysts. Upon evaluation, 35 exploitation catalysts and 22 exploration catalysts demonstrated Positive performance. Of particular interest are catalysts with C_2 yields above 20% and exploration catalysts that exhibited positive performance despite a higher non-positive certainty than positive certainty. These catalysts are highlighted with green and red stars, respectively, in figure 3.3a, marking them as catalysts of interest. The red-starred catalysts, considered as 'serendipitous catalysts,' are those whose performance stems from synergies evident only in specific ternary or higher combinations because of the binary algorithm for estimation. These catalysts are difficult to predict but offer significant value due to their unexpected positive performance. They are also marked in the same way in figure 3.3b, which provides further details.

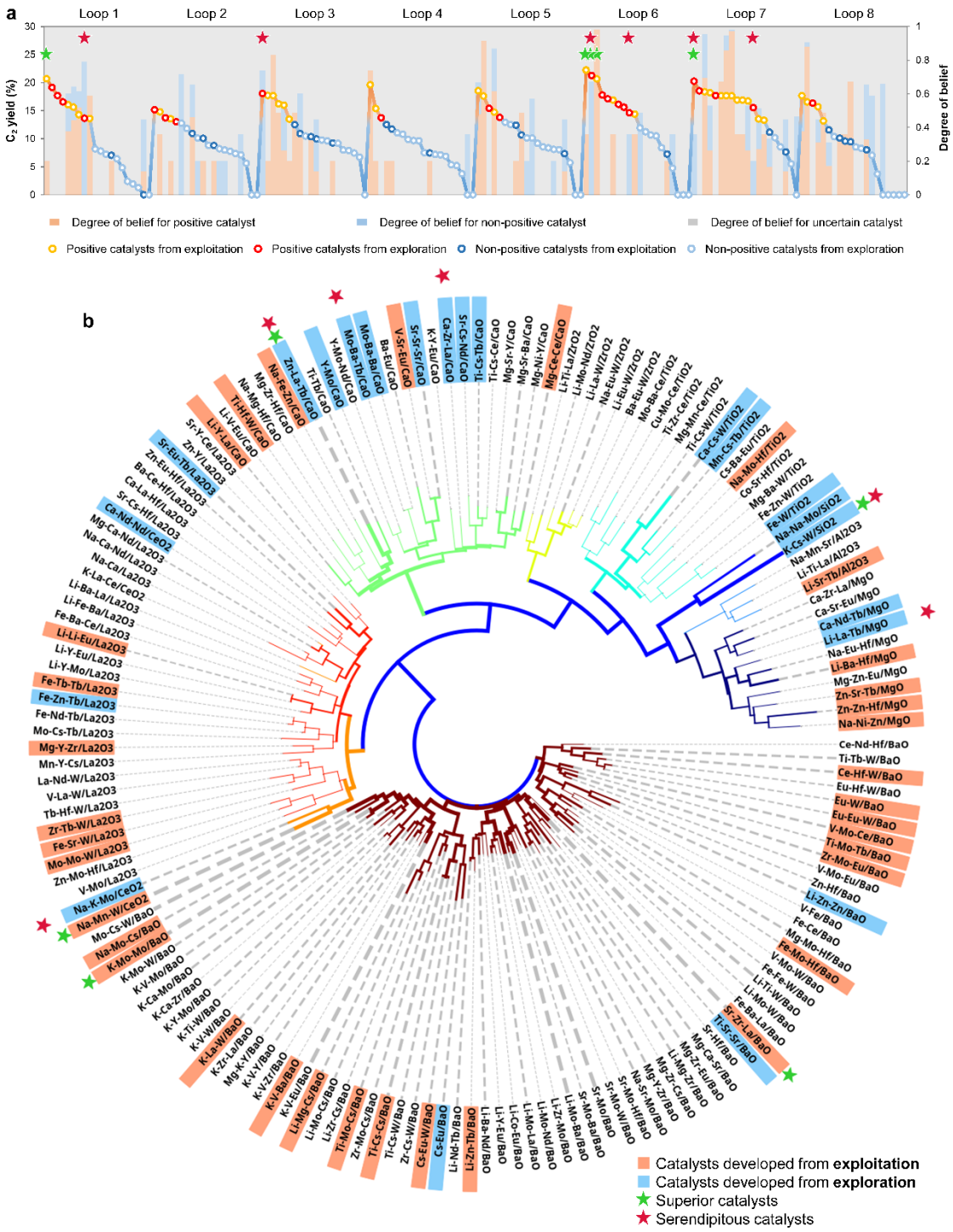


Figure 3.3. Developed catalyst by active learning. a, Prediction and testing results for each catalyst during active learning. The color of the background represents the degree of belief, and the color of the dots shows whether the catalyst was found from strategy of exploitation or exploration. The stars were marker for positive catalysts that is discovered from exploration strategy, especially degree of belief for non-positive catalyst is higher than degree of belief for positive. **b,** Catalyst phylogenetic tree for this chapter. Out of all 536 evaluated catalysts, only the positive ones were included here. Catalysts with similar physical properties are displayed adjacent to each other on the phylogenetic tree. Catalysts with more similar catalysts are placed at the center of the tree, while less similar ones are placed on the periphery. The branches are color-coded based on the type of catalyst support and controlled thickness by its C_2 yield.

Catalyst names with no background color indicate those developed in our previous study, while those with red and blue backgrounds signify catalysts developed in this chapter. Specifically, catalysts with red backgrounds correspond to those developed using exploitation strategies, and those with blue backgrounds were developed using exploration strategies. Among the catalysts developed in this chapter, serendipitous catalysts were marked with stars.

To enhance catalyst chemists' intuitive understanding of the catalysts developed in this chapter from a chemical viewpoint, I constructed a catalysis phylogenetic tree[30,31]. This tree was developed by first calculating a distance matrix based on predefined distances between catalysts, followed by hierarchical clustering. The matrix for the phylogenetic tree was compiled by exhaustively calculating the distances between pairs of catalysts across 19 physicochemical properties that are crucial in catalytic chemistry[44]. These distances were then clustered using the neighbor-joining method, a technique commonly used in molecular biology for creating phylogenetic trees. The 19 attributes considered in this analysis included atomic number, atomic radius, atomic weight, boiling point, density, electron affinity, Pauling's electronegativity, first ionization energy, molar heat capacity, melting point, molar volume, total valence electrons, valence electrons in the d, f, p, and s shells, period, thermal conductivity at 25°C, and Van der Waals radius. These features were sourced from the XenonPy library in Python and were standardized for consistency[45]. The distance between the components of two catalysts, A (represented as A{a_A, b_A, c_A, d_A, ...}) and B (represented as B{a_B, b_B, c_B, d_B, ...}), was calculated using a following formula. This formula took into account the differences in the aforementioned physicochemical properties between each pair of catalysts, allowing for a systematic and quantitative comparison of their chemical characteristics.

$$Distance_{AB} = \sum_{i}^{|A|} \min \left(\sum_{j}^{|B|} \sqrt{\sum_{k}^{|Features|} \left(feature_{k,A_i} - feature_{k,B_j} \right)^2} \right)$$

The distances are calculated by identifying and summing the distances between each component of Catalyst A and its closest corresponding component in Catalyst B. The identification of the closest corresponding component is conducted separately for each of the supporting metal and the support. For example, when comparing Catalyst A (K-Cs-W/SiO_2) with Catalyst B (Na-Zr-Mo/SiO_2), distances are computed and summed between K and Na, Cs and Na, W and Mo, and SiO_2 and SiO_2. This approach ensures that catalysts with similar properties are grouped closely on adjacent branches of the phylogenetic tree, while those with differing properties are placed further apart. Catalysts sharing more properties are positioned nearer to the center of the tree, and those with more unique compositions are placed further outwards. For the analysis, only catalysts yielding over 13% C_2 were included, totaling 160 out of the 536 evaluated (376 original plus 160 additional). The tree is color-coded based on the type of catalyst support, with thicker branches indicating higher C_2 yields. Catalysts designed using exploitation strategies are indicated with an orange background, while those from exploration strategies have a blue background.

BaO-supported catalysts represent approximately 44% of the total, with La_2O_3-supported catalysts at about 19% and CaO-supported catalysts at 15%. In comparison, Al_2O_3-, CeO_2-, and ZrO_2-supported catalysts account for only 2%, 3%, and 6% of the entries, respectively. The solid surface acid-base properties of the aforementioned oxide supports and their distribution in the entries are strongly correlated. The trend in the number of entries suggests that solid bases are more prevalent than solid acids, which in turn are more common than amphoteric solids. Indeed, the surface basicity of the catalyst is a crucial factor in OCM, aligning with these findings[46–48]. On the supported element side, acidic and amphoteric solid-based catalysts mainly feature alkali metals, alkaline earth metals, or lanthanides, as well as early transition metals.

Basic solid-based catalysts also exhibit similar combinations but perform well even solely with strongly basic metals. Catalysts exclusively using early transition metals are primarily found in basic solid-based catalysts. This suggests that basic solids are intrinsically active and accommodate various combinations, whereas acidic and amphoteric solids are more inert and admit fewer applicable combination rules. Activation of the supported elements often requires the combination of basic metals and early transition metals. An example is the Mn-Na_2WO_4/SiO_2 catalyst, a high-performing standard in the OCM field, which combines a basic metal with early transition metal elements in the form of Na_2WO_4. Some researchers posit that the active site is the tetrahedral WO_4^{2-}[40,41]. This phylogenetic tree analysis offers a comprehensive view of the catalysts' properties, supporting elements, and their impact on OCM performance.

Next, I focus on catalysts denoted with stars. Two outstanding catalysts marked with green stars are found on BaO supports and one each on CeO_2, CaO, and SiO_2 supports. Among all five catalysts, three—Sr-Zr-La/BaO, K-Mo-Mo/BaO, and Na-Mn-W/CeO_2—were developed using the exploitation strategy. In contrast, the two exceptional catalysts created using the exploration strategy, K-Cs-W/SiO_2 and Zn-La-Tb/CaO, can be identified as both superior and serendipitous. These two catalysts are unique in that they exhibit extremely high performance unexpectedly, despite the non-positive performance of analogous catalysts. Examining the composition of each of the four catalysts, it is apparent that Sr-Zr-La/BaO, K-Mo-Mo/BaO, Na-Mn-W/CeO_2, and K-Cs-W/SiO_2 include the previously mentioned combinations of basic metals and pre-transition metals as support elements. The latter two catalysts' compositions closely resemble those of the standard catalysts mentioned earlier, while Zn-La-Tb/CaO does not contain any early transition metals.

Shifting focus to serendipitous catalysts as a whole, discoveries include one each from CeO_2-supported, SiO_2-supported, and MgO-supported catalysts, as well as three from CaO-supported catalysts. Among the serendipitous catalysts, Na-K-Mo/CeO_2 and Ca-Zr-La/CaO closely resemble the superior catalysts Na-Mn-W/CeO_2 and Sr-Zr-La/BaO, respectively, in terms of composition. Among the six serendipitous catalysts, Zn-La-Tb/CaO, Mo-Ba-Tb/CaO, Ca-Zr-La/CaO, and Li-La-Tb/MgO all share the commonality of being "lanthanide-supported alkaline earth metal oxides." Serendipitous catalysts exhibit a synergistic effect, distinguishing them from analogous catalysts, and may serve as a stepping stone to discovering superior catalysts. Predicting the occurrence of serendipity could therefore be invaluable for catalyst development. To closely examine the conditions that give rise to serendipity, Table 3.1 presents the prediction results for developed catalysts using catalyst recommendation systems and traditional classifiers, each with varying numbers of substitution dimensions during elemental substitution.

Table 3.1. Difference of character of various classifiers

	Binary recommender system (main)			Unary recommender system			Traditional classifiers				
Catalysts	Positive	Non-positive	Uncertain	Positive	Non-positive	Uncertain	LR	SVM	RF	NB	C2 yield (%)
Li-La-Tb/MgO	0	0.79	0.21	0	0.2	0.8	Non-positive	Non-positive	Non-positive	Non-positive	13.7
Na-K-Mo/CeO2	0	0.74	0.26	0	0.54	0.46	Non-positive	Non-positive	Non-positive	Non-positive	18.1
K-Cs-W/SiO2	0	0.67	0.33	0	0.36	0.64	Non-positive	Non-positive	Non-positive	Non-positive	21.3
Mo-Ba-Tb/CaO	0	0.36	0.64	0.2	0	0.8	Positive	Non-positive	Positive	Positive	14.6
Ca-Zr-La/CaO	0	0.2	0.8	0.17	0	0.83	Non-positive	Non-positive	Positive	Positive	15.5
Zn-La-Tb/CaO	0	0.36	0.64	0	0.2	0.8	Non-positive	Non-positive	Non-positive	Non-positive	20.2

It appears that two out of six catalysts can predict the true class with some classifiers, while the remaining four catalysts still fail to predict the true class accurately. However, it may be possible to identify common features of serendipitous catalysts, such as 'lanthanide-supported alkaline earth metal oxides,' within the composition of

the catalysts. Therefore, a machine learning system has been developed to predict the occurrence of serendipity, which utilizes the following four descriptors: (1) predictions from the main recommender, (2) predictions from a recommender with an algorithm slightly different from the main one, (3) predictions from traditional classifiers, and (4) the catalyst's composition. This system, termed the catalyst serendipiter system, is depicted in figure 3.4a.

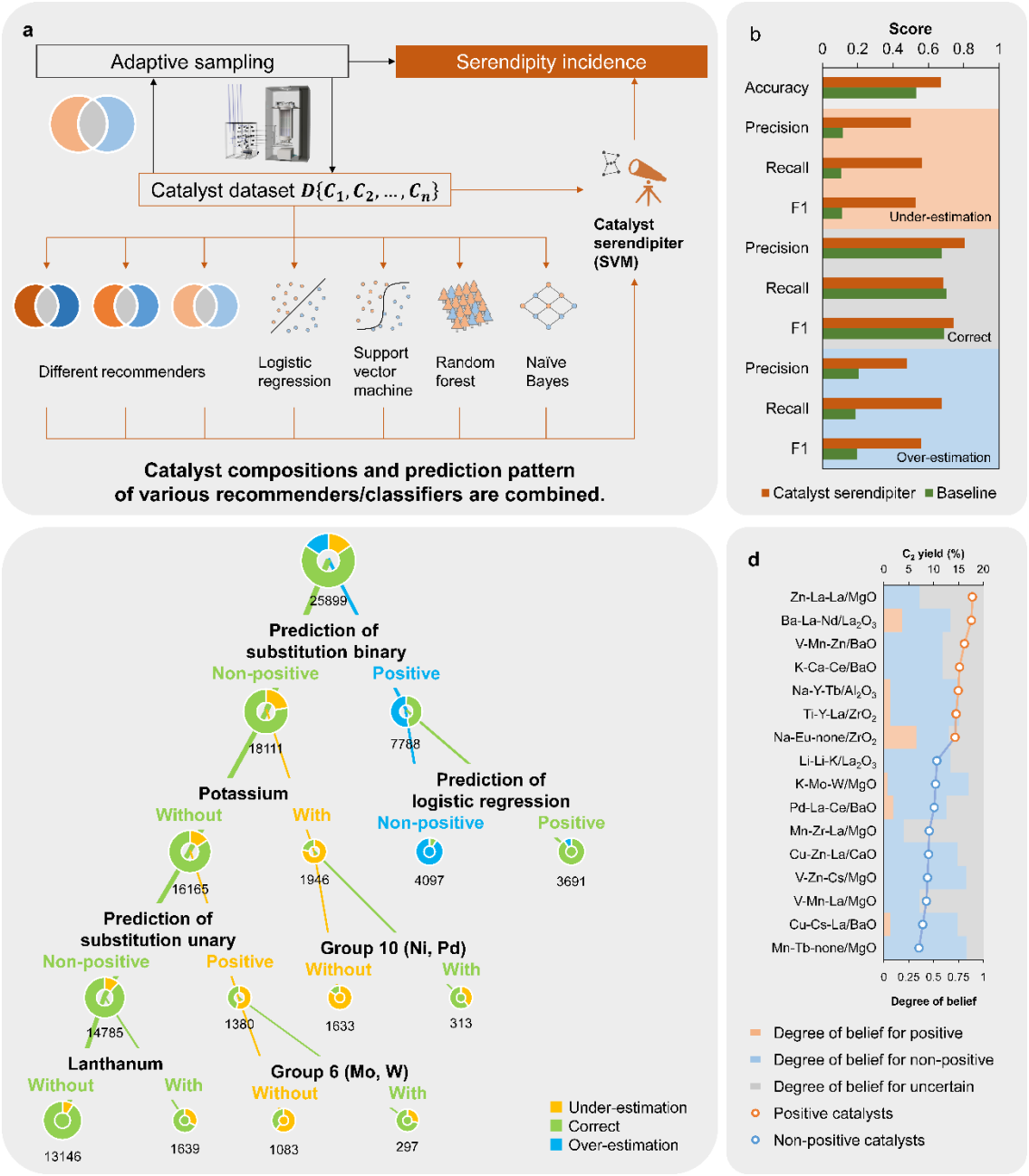


Figure 3.4. Catalyst serendipiter system and serendipity induction. **a,** Concept of serendipity detection. Prediction errors during active learning are considered serendipitous events and serve as the objective variable. I categorize these serendipities into three patterns: under-estimation, correct estimation, and over-estimation by the main recommender system, relative to actual measurements. To predict these occurrences, I use catalyst composition and prediction results from various learners as the explanatory variables. This approach enables us to predict serendipity based on both the common catalyst composition and the unique characteristics of each learner's algorithm. **b,** Each evaluation function of candidate classifiers for catalyst serendipiter. Four candidates were used: logistic regression, support vector machine, random forest classification, and naïve Bayes. Performance measures of precision, recall, and F1 score were computed for each of the three prediction classes of serendipiter. **c,** Model of model for catalyst serendipiter system. It is a decision tree constructed with descriptors which are same with setendipiter and target valuable which is prediction results of serendipiter. **d,** Prediction and testing results of recommended catalyst. 7 case of serendipities were actually induced.

The alignment of the recommender system's predictions with actual measurements is classified as "correct," "over-estimation," or "under-estimation." A "correct" classification means the recommender's prediction matched the actual result. "Over-estimation" occurs when the recommender predicts a positive outcome, but the actual result is non-positive. "Under-estimation," which is defined as serendipity, is the reverse scenario of over-estimation, where the recommender predicts a non-positive outcome, but the result is positive. The serendipiter system then categorizes each prediction from the main recommender as over-estimation, correct, or under-estimation. In cross-validating the serendipiter, samples with an uncertainty of 1 at the time of recommendation were excluded. Among the 160 catalysts evaluated during adaptive sampling, 74 were relevant for this analysis, including 6 under-estimations, 18 over-estimations, and 50 correct predictions. To increase the instances of under-estimation, an additional 80 catalysts were assessed using the serendipiter, thus augmenting the training data. This additional evaluation selected catalysts with an uncertainty not equal to 1, aiming for a balanced distribution among under-estimation, over-estimation, and correct categories. With this expansion, the training data included 16 under-estimations, 31 over-estimations, and 102 correct classifications. Leave-one-out cross-validation

results with this enlarged dataset are presented in figure 3.4b. A baseline, established based on class balance for comparison, is also shown in figure 3.4b. While the baseline scores are slightly higher than the serendipiter for correct recall, this is primarily due to the predominance of correct cases in the sample. The serendipiter exhibits superior performance in predicting both under-estimations and over-estimations.

To understand how the serendipiter system predicts serendipitous discoveries, a 'model of model' approach was introduced[49]. The model of model is a method used to interpret the prediction process of a highly opaque learning unit, such as a neural network. It involves using the inputs and outputs of this complex learning unit as training and target data, respectively, for decision tree analysis. This approach helps to shed light on how predictions are made by the original learning unit. Decision tree classification is an algorithm that classifies data by dividing it into increasingly specific subsets. It uses a tree-like structure where each internal node denotes a separation based on a particular feature, and each leaf node corresponds to a class label. The partitioning process is designed to either maximize or minimize specific criteria, such as Gini impurity or information gain, to optimize the classification. Figure 3.4c illustrates this constructed model of model, showcasing the decision-making process of the serendipiter system.

Initially, classification starts based on the prediction results from the main recommender. Given that over-estimations usually arise from positive predictions, and under-estimations from non-positive predictions, this initial classification is a logical step. Subsequently, within the branch of positive predictions, logistic regression outcomes are employed for further branching. This indicates that logistic regression is notably effective among various learners and catalyst compositions, particularly in predicting over-estimation. It also visualizes how other predictors refine the

recommender's initial predictions. Moving to the non-positive predictive branch of the main recommender, the first adjustment is made based on the presence of potassium. If potassium is included, the purity of under-estimation predictions is enhanced at the next node by the absence of group 10 elements. In contrast, in the branch where potassium is not included, the unary recommender steps in to correct the prediction. Each algorithm learns different rules for catalyst design, and the serendipiter leverages these to predict the occurrence of serendipity. Essentially, a decision based on elemental composition reveals that the main recommender is more accurate with catalysts containing a specific element, as evidenced by a higher number of correct predictions on the 'with' side. This implies that the main recommender has a better understanding of catalysts containing that element, resulting in fewer errors. From this observation, I infer that the main recommender is more susceptible to prediction errors when elements such as potassium or lanthanum are present. Catalysts with combinations of three or more elements that include them are more likely to exhibit altered catalytic performance due to synergistic effects.

An attempt was made to practically induce serendipity using the Serendipiter. Sixteen catalysts were chosen by roulette sampling based on their recommendation weight among those predicted as under-estimation. The degree of belief and actual performances of these 16 catalysts are depicted in figure 3.4d. Successfully, serendipity was induced in 7 out of the 16 cases, demonstrating that serendipity, typically a chance occurrence, could be artificially induced with a probability of 43.8%. This induction rate represents a fivefold increase in efficiency compared to the serendipity probability during adaptive sampling, which was 8.1% (approximately 6 out of 74). Moreover, focusing on the seven serendipitous catalysts identified, I note that three were discovered using Al_2O_3 and ZrO_2 catalysts, substances are sensitive to the

combinatorial law. Notably, the Serendipiter's effectiveness is confirmed for Na-Y-Tb/Al_2O_3 and Ti-Y-La/ZrO_2, as these combinations were not paralleled by any similar catalysts on the same supports in the phylogenetic tree presented in figure 3.3b.

3.4. Conclusion

In summary, this chapter has developed a framework for accelerated catalyst discovery by combining the evidence theory and ideas of elemental substitution to develop an algorithm for catalyst performance estimation, which is then used in conjunction with adaptive sampling. The algorithm, named the catalyst recommender system, is able to balance exploration and exploitation strategies by quantifying uncertainties. This feature has led to the development of several high-performance catalysts that have been difficult to find in our data so far.

Additionally, a supplementary layer of learning, the catalyst serendipiter system, was established. This system focuses on learning the emergence of high-performance catalysts that were challenging for the recommender system to predict. It enables the artificial reproduction of serendipity, a traditionally unpredictable but vital element in catalyst development. A practical outcome of this methodology is the successful identification of promising and unfamiliar OCM catalyst candidates such as Na-Y-Tb/Al_2O_3 and Ti-Y-La/ZrO_2.

References

[1] R. Schlögl, "Heterogeneous catalysis," *Angew. Chem. Int. Ed.*, **2015**, 54, 11, 3465–3520. doi: 10.1002/anie.201410738.

[2] J. W. Thybaut, J. Sun, L. Olivier, A. C. Van Veen, C. Mirodatos, G. B. Marin, "Catalyst design based on microkinetic models: Oxidative coupling of methane," *Catal. Today*, **2011**, 29–36. doi: 10.1016/j.cattod.2010.09.002.

[3] V. I. Alexiadis et al., "Oxidative coupling of methane: Catalytic behaviour assessment via comprehensive microkinetic modelling," *Appl. Catal. B*, **2014**, 150–151, 496–505. doi: 10.1016/j.apcatb.2013.12.043.

[4] C. Yu, J. He, "Synergic catalytic effects in confined spaces," *Chem. Commun.*, **2012**, 48, 41, 4933–4940. doi: 10.1039/c2cc31585h.

[5] Y. Ying, X. Luo, J. Qiao, H. Huang, "'More is Different:' Synergistic Effect and Structural Engineering in Double-Atom Catalysts," *Adv. Funct. Mater.*, **2021**, 31, 3. doi: 10.1002/adfm.202007423.

[6] A. G. Dedov, A. S. Loktev, I. I. Moiseev, A. Aboukais, J. F. Lamonier, I. N. Filimonov, "Oxidative coupling of methane catalyzed by rare earth oxides: Unexpected synergistic effect of the oxide mixtures," *Appl. Catal. A Gen.*, **2003**, 245, 2, 209–220. doi: 10.1016/S0926-860X(02)00641-5.

[7] E. J. Ras, G. Rothenberg, "Heterogeneous catalyst discovery using 21st century tools: A tutorial," *RSC Adv.*, **2014**, 4, 12, 5963–5974. doi: 10.1039/c3ra45852k.

[8] J. S. Kim, B. Kim, H. Kim, K. Kang, "Recent Progress on Multimetal Oxide Catalysts for the Oxygen Evolution Reaction," *Adv. Energy Mater.*, **2018**, 8, 11. doi: 10.1002/aenm.201702774.

[9] V. Busico, R. Pellecchia, F. Cutillo, R. Cipullo, "High-throughput screening in olefinpolymerization catalysis: from serendipitous discovery towards rational understanding," *Macromol. Rapid Commun.*, **2009**, 30, 20, 1697–1708. doi: 10.1002/marc.200900246.

[10] J. M. Venegas, W. P. McDermott, I. Hermans, “Serendipity in Catalysis Research: Boron-Based Materials for Alkane Oxidative Dehydrogenation,” *Acc. Chem. Res.*, **2018**, 51, 10, 2556–2564. doi: 10.1021/acs.accounts.8b00330.

[11] J. M. Venegas, W. P. McDermott, I. Hermans, “Serendipity in Catalysis Research: Boron-Based Materials for Alkane Oxidative Dehydrogenation,” *Acc. Chem. Res.*, **2018**, 51, 10, 2556–2564. doi: 10.1021/acs.accounts.8b00330.

[12] K. Takahashi et al., “The Rise of Catalyst Informatics: Towards Catalyst Genomics,” *ChemCatChem*, **2019**, 11, 4, 1146–1152. doi: 10.1002/cctc.201801956.

[13] K. Takahashi et al., “Catalysts informatics: paradigm shift towards data-driven catalyst design,” *Chem. Commun.*, **2023**, 59, 16, 2222–2238. doi: 10.1039/d2cc05938j.

[14] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. I. Shimizu, “Machine Learning for Catalysis Informatics: Recent Applications and Prospects,” *ACS Catal.*, **2020**, 10, 3, 2260–2297. doi: 10.1021/acscatal.9b04186.

[15] B. R. Goldsmith, J. Esterhuizen, J. X. Liu, C. J. Bartel, C. Sutton, “Machine learning for heterogeneous catalyst design and discovery,” *AIChE J.*, **2018**, 64, 7, 2311–2323. doi: 10.1002/aic.16198.

[16] T. Lookman, P. V. Balachandran, D. Xue, R. Yuan, “Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design,” *npj Comput. Mater.*, **2019**, 5, 1. doi: 10.1038/s41524-019-0153-8.

[17] K. Terayama, M. Sumita, R. Tamura, K. Tsuda, “Black-Box Optimization for Automated Discovery,” *Acc. Chem. Res.*, **2021**, 54, 6, 1334–1346. doi: 10.1021/acs.accounts.0c00713.

[18] C. B. Browne et al., “A survey of Monte Carlo tree search methods,” *IEEE Trans. Comput. Intell. AI Games*, **2012**, 4, 1, 1–43. doi: 10.1109/TCIAIG.2012.2186810.

[19] T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, K. Tsuda, “COMBO: An efficient Bayesian optimization library for materials science,” *Mater. Discov.*, **2016**, 4, 18–21. doi: 10.1016/j.md.2016.04.001.

[20] K. Homma et al., “Optimization of a Heterogeneous Ternary Li3PO4-Li3BO3-Li2SO4 Mixture for Li-Ion Conductivity by Machine Learning,” *J. Phys. Chem. C*, **2020**, 124, 24, 12865–12870. doi: 10.1021/acs.jpcc.9b11654.

[21] K. Terayama, H. Iwata, M. Araki, Y. Okuno, K. Tsuda, “Machine learning accelerates MD-based binding pose prediction between ligands and proteins,” *Bioinformatics*, **2018**, 34, 5, 770–778. doi: 10.1093/bioinformatics/btx638.

[22] K. Terayama et al., “Efficient construction method for phase diagrams using uncertainty sampling,” *Phys. Rev. Mater.*, **2019**, 3, 3. doi: 10.1103/PhysRevMaterials.3.033802.

[23] R. Batra, L. Song, R. Ramprasad, “Emerging materials intelligence ecosystems propelled by machine learning,” *Nat. Rev. Mater.*, **2021**, 6, 8, 655–678. doi: 10.1038/s41578-020-00255-y.

[24] S. Makri, A. Blandford, M. Woods, S. Sharples, D. Maxwell, “‘Making my own luck’: Serendipity strategies and how to support them in digital information environments,” *J. Assoc. Inf. Sci. Technol.*, **2014**, 65, 11, 2179–2194. doi: 10.1002/asi.23200.

[25] Z. L. Wang, T. Funada, T. Onda, Z. C. Chen, “Knowledge extraction and performance improvement of Bi_2Te_3-based thermoelectric materials by machine learning,” *Mater. Today Phys.*, **2023**, 31. doi: 10.1016/j.mtphys.2023.100971.

[26] J. Yu et al., “Mining the synergistic effect in hydrothermal co-liquefaction of real feedstocks through machine learning approaches,” *Fuel*, **2023**, 334. doi: 10.1016/j.fuel.2022.126715.

[27] Z. Ni et al., “Investigation of the co-pyrolysis of coal slime and coffee industry residue based on machine learning methods and TG-FTIR: Synergistic effect, kinetics and thermodynamic,” *Fuel*, **2021**, 305. doi: 10.1016/j.fuel.2021.121527.

[28] T. N. Nguyen et al., “Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane,” *ACS Catal.*, **2021**, 11, 3, 1797–1809. doi: 10.1021/acscatal.0c04629.

[29] T. N. Nguyen et al., “High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane,” *ACS Catal.*, **2020**, 10, 2, 921–932. doi: 10.1021/acscatal.9b04293.

[30] P. Kapli, Z. Yang, M. J. Telford, “Phylogenetic tree building in the genomic age,” *Nat. Rev. Genet.*, **2020**, 21, 7, 428–444. doi: 10.1038/s41576-020-0233-0.

[31] N. Saitou, M. Nei, “The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees.” *Mol Biol Evol.* **1987**, 4, 406-25. doi: 10.1093/oxfordjournals.molbev.a040454.

[32] M. Q. Ha et al., “Evidence-based recommender system for high-entropy alloys,” *Nat. Comput. Sci.*, **2021**, 1, 7, 470–478. doi: 10.1038/s43588-021-00097-w.

[33] A. P. Dempster, “A Generalization of Bayesian Inference,” *J. R. Stat. Soc. Ser. B Methodol.* **1968**, 30, 205–232.

[34] A Guided Tour of Artificial Intelligence Research, *Springer International Publishing*, **2020**. doi: 10.1007/978-3-030-06164-7.

[35] S. Nakanowatari et al., “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. doi: 10.1002/cctc.202100460.

[36] K. Takahashi et al., “Catalysis Gene Expression Profiling: Sequencing and Designing Catalysts,” *J. Phys. Chem. Lett.*, **2021**, 12, 30, 7335–7341. doi: 10.1021/acs.jpclett.1c02111.

[37] L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike, K. Takahashi, "Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts," *Chem. Sci.*, **2021**, 12, 38, 12546–12555. doi: 10.1039/d1sc04390k.

[38] G. E. Keller, M. M. Bhasin, "Synthesis of Ethylene via Oxidative Coupling of Methane. I. Determination of Active Catalysts."

[39] V. I. Lomonosov, M. Y. Sinev, "Oxidative coupling of methane: Mechanism and kinetics," *Kinet. Catal.*, **2016**, 57, 5, 647–676. doi: 10.1134/S0023158416050128.

[40] D. Kiani, S. Sourav, J. Baltrusaitis, I. E. Wachs, "Oxidative Coupling of Methane (OCM) by SiO2-Supported Tungsten Oxide Catalysts Promoted with Mn and Na," *ACS Catal.*, **2019**, 9, 7, 5912–5928. doi: 10.1021/acscatal.9b01585.

[41] S. Arndt, T. Otremba, U. Simon, M. Yildiz, H. Schubert, R. Schomäcker, "Mn-Na $_2WO_4/SiO_2$ as catalyst for the oxidative coupling of methane. What is really known?," *Appl. Catal. A Gen.*, **2012**, 425–426, 53–61. doi: 10.1016/j.apcata.2012.02.046.

[42] U. Zavyalova, M. Holena, R. Schlögl, M. Baerns, "Statistical analysis of past catalytic data on oxidative methane coupling for new insights into the composition of high-performance catalysts," *ChemCatChem*, **2011**, 3, 12, 1935–1947. doi: 10.1002/cctc.201100186.

[43] S. Mine et al., "Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts," *ChemCatChem*, **2021**, 13, 16, 3636–3655. doi: 10.1002/cctc.202100495.

[44] S. Ishioka, A. Fujiwara, S. Nakanowatari, L. Takahashi, T. Taniike, K. Takahashi, "Designing Catalyst Descriptors for Machine Learning in Oxidative Coupling of Methane," *ACS Catal.*, **2022**, 12, 19, 11541–11546. doi: 10.1021/acscatal.2c03142.

[45] S. Wu, G. Lambard, C. Liu, H. Yamada, R. Yoshida, "iQSPR in XenonPy: A Bayesian Molecular Design Algorithm," *Mol. Inform.*, **2020**, 39, 1–2. doi: 10.1002/minf.201900107.

[46] S. L. S. Kuś, M. Otremba, M. Taniewski, “The catalytic performance in oxidative coupling of methane and the surface basicity of La_2O_3, Nd_2O_3, ZrO_2 and Nb_2O_5,” *Fuel*, **2003**, 82, 11, 1331–1338. doi: 10.1016/S0016-2361(03)00030-9.

[47] J. Xu et al., “Design of strontium stannate perovskites with different fine structures for the oxidative coupling of methane (OCM): Interpreting the functions of surface oxygen anions, basic sites and the structure–reactivity relationship,” *J. Catal.*, **2022**, 408, 465–477. doi: 10.1016/j.jcat.2021.04.007.

[48] V. R. Choudhary, V. H. Rane, M. Y. Pandit, “Comparison of alkali metal promoted MgO catalysts for their surface acidity/basicity and catalytic activity/selectivity in the oxidative coupling of methane,” *J. Chem. Technol. Biotechnol.*, **1997**, 68, 2, 177–186. doi: 10.1002/(SICI)1097-4660(199702)68:2<177::AID-JCTB645>3.0.CO;2-U.

[49] J. Li et al., “Deep Learning Accelerated Gold Nanocluster Synthesis,” *Adv. Intell. Syst.*, **2019**, 1, 3. doi: 10.1002/aisy.201900029.

Chapter 4

Visualization for Evolutionary History of Catalysis

Abstract: The design of solid catalysts is a complex challenge, heavily reliant on uncovering synergistic component combinations where individual elements contribute to the reaction cycle by their own, stabilize each other, and/or generate specific structures. The prediction of such combinatorial effects spans multiple spatiotemporal scales, rendering it a formidable task given the historically trial-and-error nature of catalyst development. Data science holds promise for discerning these combinatorial laws, yet the small, inconsistent, and often incomplete datasets present obstacles for deploying robust machine learning predictions. To address these challenges, this chapter introduces a method that employs phylogenetic trees—traditionally used in biology and linguistics—to map the evolutionary development of catalysts. This method calculates catalyst "distances" to visualize incremental advancements in the field, leveraging data from literature and high-throughput experiments. By creating a catalytic phylogenetic tree, I trace the lineage of catalyst development and introduce a horizontal propagation strategy for catalyst design. The application of this strategy to oxidative coupling of methane has yielded high-performing catalysts at low temperatures, underscoring the potential of phylogenetic trees as a powerful tool in navigating combinatorial catalyst designs.

Keywords: Data visualization, Phylogenetic tree, Catalyst informatics, Catalyst design, Oxidative coupling of methane.

4.1. Introduction

One of the keys to solid catalyst design lies in the discovery of synergistic combinations of components[1–3]. In these combinations, each component plays a distinct role in the overall reaction cycle[4–7], specific structures are generated upon combining components[8–12], and one component may stabilize another[13,14]. On the other hand, the combinatorial effect arises from a complex interplay across multiple spatiotemporal scales with other catalyst components and all elementary reactions present in the system, making the prediction of the combinatorial effect extremely challenging[15,16]. Historically, the development of solid catalysts has relied heavily on trial and error[17]. Currently, attempts to statistically model the combinatorial laws of catalysts using data science techniques are yielding promising results[18–20]. In this context, several datasets compiling catalytic data from the literature have been published[21–24]. However, the data science approach has unveiled critical issues: the datasets' small size, the low homogeneity in the data acquisition process, and the scarcity of negative results, which significantly limit the application of reliable and meaningful machine learning techniques[25,26]. Approaches to this problem include using homogeneous and naturally distributed data from laboratory notebooks and high-throughput experiments, and, in the long term, efforts to standardize the data[27–31]. Alternatively, literature data could also serve as input for supervised machine learning as it stands — not for quantitative prediction of unknown catalysts, but to provide useful hints for researchers contemplating their next experiments, such as combination rules or overlooked ideas.

The phylogenetic tree is technique developed in molecular biology and bioinformatics to represent the evolutionary pathways of organisms in the form of a tree[32]. They infer the phylogeny of organisms through cluster analysis, using

distances calculated between organisms based on gene and amino acid sequences. These trees operate under the assumption that by connecting closely related species among existing organisms, one can infer their differentiation or evolutionary processes. The concept of phylogenetic trees, which presupposes a common ancestor from which all diversification has occurred, is also applied in the field of linguistics[33]. In linguistics, phylogenetic trees are used to trace the evolution of languages by assuming a common origin and comparing features such as phonology, lexicon, and grammar. Furthermore, the principles of phylogenetic trees are adaptable for tracing processes in areas where developments are thought to have emerged through incremental changes. For instance, combination searches in catalyst development are conducted through stepwise modifications to existing combinations, making the research framework particularly amenable to visualization using phylogenetic trees. Additionally, unlike in biology and linguistics, it is relatively straightforward to branch out new lineages on the phylogenetic tree in catalytic chemistry. This characteristic may provide a valuable roadmap for strategizing in the design of new catalysts.

I therefore propose a method to visualize the development history of solid catalysts using phylogenetic trees, based on combinatorial laws. By establishing a method to calculate distances between catalysts and creating a catalytic phylogenetic tree from the publicly available catalytic reaction literature database, I can visualize the incremental improvements that have occurred throughout the history of catalyst development. Furthermore, by incorporating our large-scale catalytic data obtained from high-throughput experiments into the catalytic reaction database, I facilitate a phylogenetic tree-based comparison between historical data and data self-generated by individual laboratories. I aim to demonstrate the utility of phylogenetic trees in catalytic chemistry by designing and evaluating catalysts based on the insights gained.

4.2. Methods

4.2.1. Data pretreatment

The OCM literature database, initially constructed by Zavyalova et al.[21] and later re-edited by Mine et al.[22], was utilized in this chapter. This database compiles all published catalytic reaction data from articles spanning from 1982 to 2019, comprising a total of 4,759 data points. Each data point includes detailed information on the catalyst, such as the type and compositional ratio of the catalyst components (precursor cations, anions, promoters, and supports) and the active elements of the catalyst. It also encompasses process condition information like the catalyst preparation method and reaction temperature. This is linked to the output data on conversion, selectivity, and yield when the catalyst is assessed under these specified conditions. However, some data, such as certain process conditions or output information, may be missing depending on the source article. In this chapter, yield was considered the most critical output information. Therefore, data points lacking yield information were omitted from this analysis. Consequently, 44 data points were removed, resulting in a revised total of 4,715 data points. Moreover, discrepancies exist in the database regarding the presence of supports or active elements in some catalysts. To standardize the comparison, I temporarily disregarded the predefined framework of active elements and supports. The element with the highest compositional ratio was designated as the main component, i.e., the support, while others were considered modifiers. Elements with a compositional ratio exceeding half of that considered a support were also categorized as supports. Precursor anions were excluded from the catalyst composition, as they are typically replaced by oxygen during the sintering process.

Catalyst improvement can be categorized into two screening approaches: one in which the elements constituting the catalyst are varied to search for new combinations, and

another where the types of elements are fixed, focusing on optimizing their compositional ratios and process conditions. The latter approach is often undertaken after a promising elemental composition has been identified by the former. In this chapter, results from catalysts with identical elemental compositions are aggregated and treated as outcomes from a single catalyst to visualize the screening direction process. When integrating results from catalysts of the same composition, the median yield from a set of catalysts with a specific elemental combination was used as the performance representative for that composition. Additionally, the number of overlap on a single catalyst was recorded, as it can serve as an indicator of the level of attention that the catalyst has garnered in the particular catalytic reaction.

4.2.2. Distance definition

To generate the phylogenetic tree, I first defined a method for calculating the distance between catalysts. In molecular phylogenetics, phylogenetic trees utilize the "edit distance" to calculate the distance between two gene sequences or amino acid sequences. For instance, with genes, the simplest method involves counting the number of substitutions, insertions, or deletions of adenine (A), thymine (T), guanine (G), and cytosine (C) required to make two gene sequences identical. Inspired by this concept, I calculate the distance between two catalysts based on their non-common elements. Instead of merely tallying the number of substitutions and other operations, I incorporated 19 chemical parameters traditionally considered significant in catalysis into the distance calculation. These 19 parameters are listed in Table 1. It's important to note that these values were sourced from the Python library XenonPy and were standardized for each parameter[34]. Utilizing these parameters to calculate the distance allows for catalysts that are perceived as similar by catalysis researchers to be

placed in close proximity on the phylogenetic tree, even if they have undergone the same number of substitutions. The specific method to determine the distance involves calculating the sum of the Euclidean distances for the 19 parameters for the non-common elements of each modifier and main component, respectively, applying appropriate weightings, and then combining these sums. The weights are determined based on the visualization objective. If catalyst A is represented by A{a_A, b_A, c_A, d_A, ...} and catalyst B by B{a_B, b_B, c_B, d_B, ...}, the $distance_{AB}$ is calculated as follows:

$$Distance_{AB} = \sum_{i}^{|A|} \min \left(\sum_{j}^{|B|} \sqrt{\sum_{k}^{|Features|} \left(feature_{k,A_i} - feature_{k,B_j} \right)^2} \right)$$

Table 1. Employed features

k	$Feature_k$
1	Atomic number
2	Atomic radius
3	Atomic weight
4	Boiling point
5	Density
6	Electron affinity
7	Pauling's electron negativity
8	First ion energy
9	Molar heat capacity
10	Melting point
11	Molar volume
12	Total valance electron
13	Valance electron in d shell
14	Valance electron in f shell

15	Valance electron in p shell
16	Valance electron in s shell
17	period
18	Thermal conductivity at 25 ℃
19	Van der Waals radius

4.2.3. Construction of phylogenetic trees

After exhaustively calculating the distances between all catalyst combinations and compiling the distance matrix, a phylogenetic tree was constructed using the neighbor-joining (NJ) method[35]. The NJ method constructs the phylogenetic tree based on the premise that the most accurate phylogenetic tree has the shortest total branch length. It is a greedy algorithm that, unlike other methods, does not assume a constant rate of evolution. Instead, it identifies the pair of leaves (defined as neighbors) that adhere to the principle of minimum evolution at each step of the phylogenetic tree construction. It should be noted that, due to considerations of data reliability and the size of the phylogenetic tree, only catalysts with an overlap count of five or more are displayed.

Catalysts that are frequently investigated in the targeted catalytic reaction are considered mainstream within that context. I posit that the support, which constitutes the main component of the catalyst, dictates the broad behavior of the catalyst. A support represented by two or more catalysts with an overlap count of 30 or more was designated as the main catalytic system within that catalytic reaction. Branches linked to catalysts identified as part of the major catalytic system were color-coded accordingly. The thickness of the branches was modulated to reflect the number of overlaps, facilitating an intuitive discernment of the major catalysts in a given catalytic

reaction. Additionally, to depict the performance of catalysts on the phylogenetic tree, the background color of the catalyst names was varied on a gradient from white to red to represent increasing performance.

4.2.4. Catalyst preparation and evaluation

All catalysts were prepared using the co-impregnation method. Precisely 0.37 mmol of the modifier precursor was measured and added to 1 g of support and 5 mL of H_2O, then stirred at 50 °C for 6 hours. The mixture was subsequently dried under reduced pressure at 90 °C for 4 hours. If necessary, the process of impregnation and reduced pressure drying was repeated with a metal alkoxide-type precursor in ethanol. Following this, the catalyst was calcined at 1000°C for 4 hours. The produced catalysts were evaluated under 135 different conditions using a high-throughput catalyst evaluation system. For the preparation of Ca-La catalysts, equimolar amounts of $Ca(OH)_2$ and La_2O_3 were used and impregnated.

Metal precursors used in this chapter were $LiNO_3$, $NaNO_3$, $Mg(NO_3)_2$, KNO_3, $Ca(NO_3)_2{\cdot}4H_2O$, $Ti(O\mathit{i}Pr)_4$, $VOSO_4{\cdot}\mathit{x}H_2O$ (x = 3–5), $Mn(NO_3)_2{\cdot}6H_2O$, $Fe(NO_3)_3{\cdot}9H_2O$, $Co(NO_3)_2{\cdot}6H_2O$, $Ni(NO_3)_2{\cdot}6H_2O$, $Cu(NO_3)_2{\cdot}3H_2O$, $Zn(NO_3)_2{\cdot}6H_2O$, $Sr(NO_3)_2$, $Y(NO_3)_3{\cdot}6H_2O$, $ZrO(NO_3)_2{\cdot}\mathit{x}H_2O$ (x = 2), $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, $Pd(OAc)_2$, $CsNO_3$, $Ba(NO_3)_2$, $La(NO_3)_3{\cdot}6H_2O$, $Ce(NO_3)_3{\cdot}6H_2O$, $Nd(NO_3)_3{\cdot}6H_2O$, $Eu(CH_3COO)_3{\cdot}5H_2O$, $Tb(NO_3)_3{\cdot}5H_2O$, $Hf(OEt)_4$, and $(NH_4)_{10}H_2(W_2O_7)_6$. They were purchased from either of Sigma–Aldrich, Kanto Chemical, Wako Pure Chemical Industries, and Alfa-Aesar. The following solid powders were used as supports or as support precursors: Calcium hydroxide ($Ca(OH)_2$, Wako Pure Chemical Industries), titanium(IV) oxide (TiO_2, anatase, Kanto Chemical), barium hydroxide ($Ba(OH)_2{\cdot}8H_2O$, Wako Pure Chemical

Industries), lanthanum(III) oxide (La_2O_3, Wako Pure Chemical Industries), and Tin(IV) oxide (SnO_2, Wako Pure Chemical Industries).

4.2.5. Synthesis of Ti-Sn-Ba Support

$Ba(OH)_2 \cdot 8H_2O$, TiO_2, and SnO_2 were measured out in a 1:1:1 molar ratio, moistured, and ground in an agate mortar for 20 minutes. The resulting paste was dried at 110 °C overnight. The dried material was then reground in an agate mortar for an additional 20 minutes. Subsequently, the resulting powder was calcined at 1000 °C for 5 hours. After calcination, it was ground again using a mortar and used for catalyst preparation. The XRD pattern of the synthesized material, presented in figure 4.1, confirms the presence of $BaTiO_3$ and SnO_2, with minimal contributions from the starting materials $Ba(OH)_2$ and TiO_2. It is noted that no halide precursor was utilized in the synthesis to prevent potential damage to the reactor.

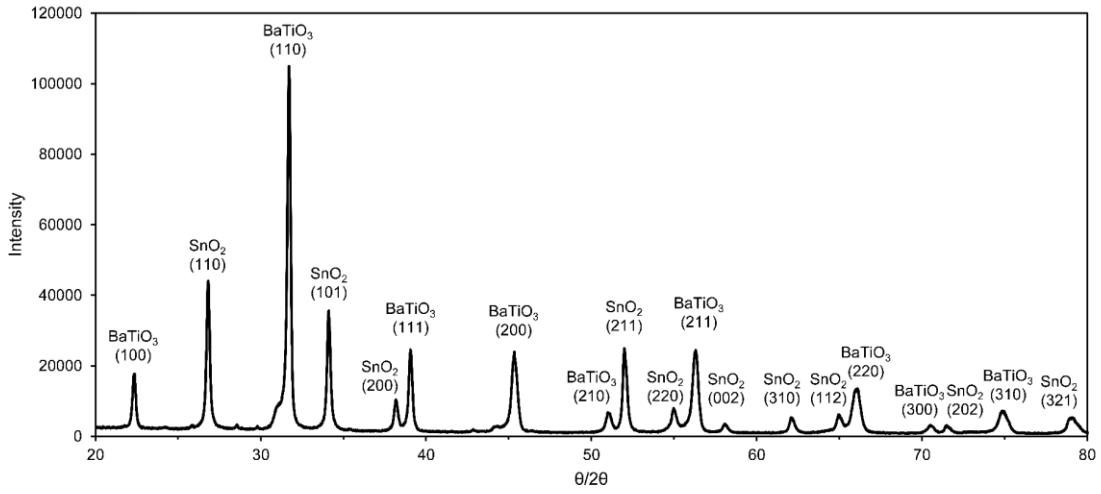


Figure S4.1. XRD pattern of prepared Ti-Sn-Ba support. measured from 3° to 90° by 0.01° at 5 degree/min; the small peak overlapping $BaTiO_3$(110) near 30° is $BaSnO_3$(110) (perovskite-type compound) is possible.

4.3. Results

4.3.1. Visualizing history using phylogenetic trees

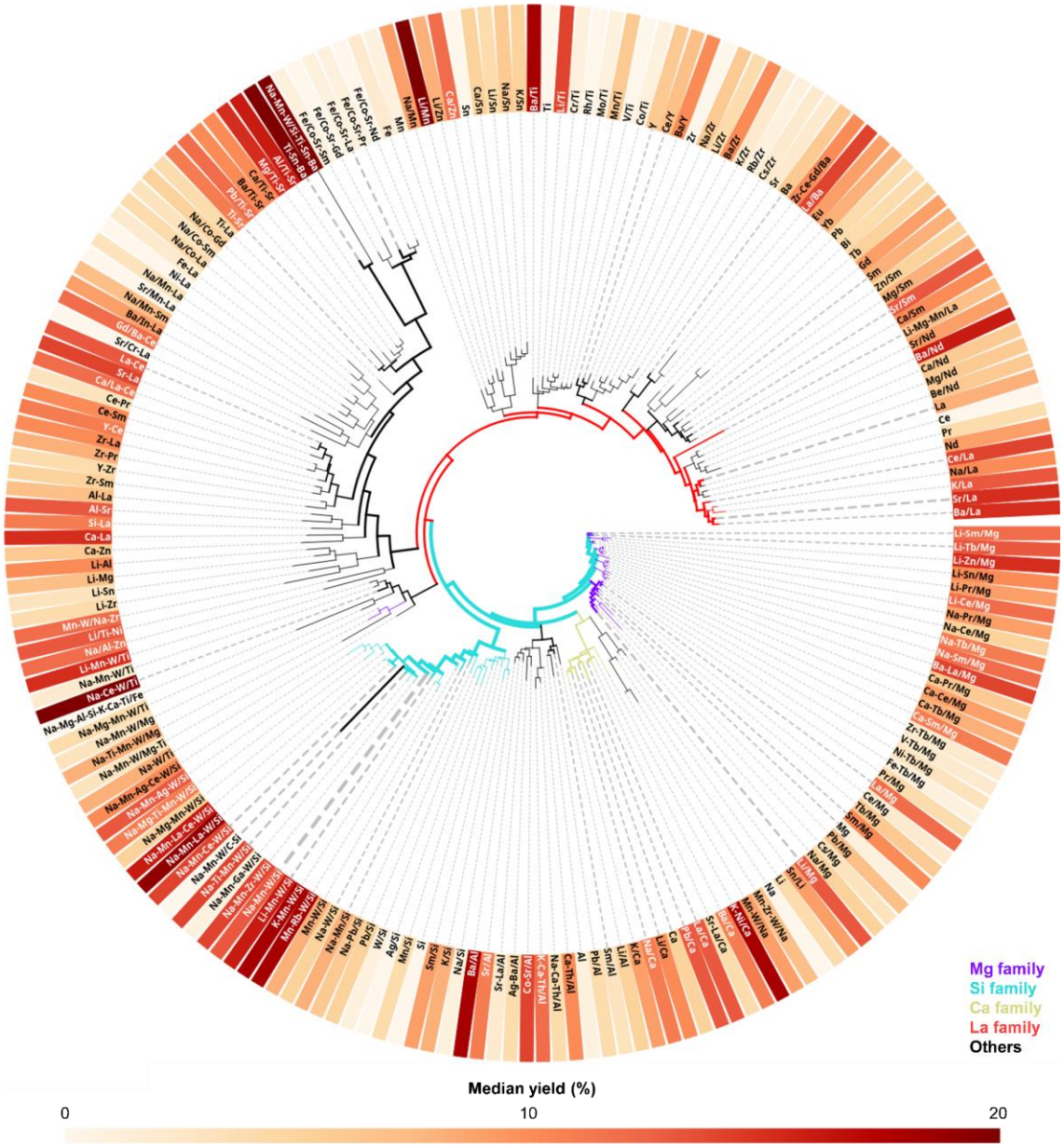


Figure 4.2. Catalyst phylogenetic tree for OCM literature data. In the phylogenetic tree, only catalysts with an overlap count of five or more are included. The thickness of a branch correlates with the number of overlaps—the thicker the branch, the higher the overlap count. A total of 198 catalysts met the criteria for inclusion. The median performance representation is limited to a maximum of 20% C_2 yield.

In the OCM literature database[22], there were 198 catalysts with more than five overlaps. The phylogenetic tree generated using the 198 catalysts is shown in figure 4.2. Based on the defined criteria of a support having two or more catalysts with an overlap frequency of thirty or more as indicative of the main catalytic family for a given reaction, four main families of OCM have been identified.

The first family is characterized by Mg as the support, corresponding to the catalysts linked to the purple branch in the lower right region of the phylogenetic tree. The four most extensively studied catalysts in this category are Mg, Li/Mg, La/Mg, and Li-Sm/Mg, with support metals primarily comprising alkali metals, alkaline earth metals, and lanthanides. Performance trends suggest that high efficiency is often achieved with at least two supporting elements, forming a ternary system. However, the introduction of transition metal elements into a ternary system seems to adversely affect performance. Combinations of alkali or alkaline earth metals with lanthanides are common, with Ba-La/Mg displaying particularly high efficacy. The most notable Mg-based catalyst in the literature is Li/Mg, recognized as one of the top-performing catalysts in OCM[36]. Discovered by Ito and Lunsford in 1985[37], the active site of this catalyst is believed to be the locally generated O^- by the doping of Li^+ into $Mg^{2+}O^{2-}$ [36]. However, the Li+ that contributes to the active site is considered highly volatile at elevated temperatures, leading to deactivation through evaporation as LiOH or migration into the quartz reactor tube as Li_2SiO_3[36].

The second group of catalysts that have been extensively investigated in the context of OCM is the Ca-supported catalysts. These are associated with the yellow-green branch located to the left of the Mg family on the phylogenetic tree. The names of the catalysts suggests that there has been limited exploration in diversifying the combinatorial rule for Ca-based systems. However, several catalysts within this group,

such as Ca, La/Ca, Na/Ca, and Pb/Ca, are evidenced by thick branches on the tree, indicating substantial research focused on their compositional ratios or reaction conditions. Indeed, Ca-supported catalysts are recognized in the literature for their high methane conversion capacity and their exceptional performance at low temperatures[38].

The third phylogenetic lineage is the La-supported catalyst group. These catalysts are identified by a red branch in the top right-hand corner of the phylogenetic tree. In contrast to the other catalytic systems discussed, this group appears less clustered, suggesting a broader consideration of rare-earth metal oxides, like La, in OCM research. Within this category, Sm and Nd are frequently used as catalysts[39]. On the combinatorial development front, Li-Mg-Mn/La stands out as the sole higher-order combination designed among lanthanide-based catalysts, while the rest are either only the support or feature a single element on the support. Typically, alkali metals and alkaline earth metals are chosen for combination with lanthanide supports. Literature reports lanthanide oxide catalysts, particularly La_2O_3, as being highly active at low temperatures[36,39]. Contrary to most OCM catalysts that exhibit methane conversion capacity only between 700 °C and 900 °C, La_2O_3's activity is notable even below 600°C[40], [41]. This high methane conversion capacity is attributed to the ability of the catalyst to stabilize oxygen adsorbed from the gas phase onto the catalyst surface in various ionic forms, including O^-[36,39]. Surface basicity among lanthanide supports is regarded as a parameter closely related to catalytic activity[7,42].

The fourth phylogenetic group of catalysts employs Si as a support. This group is linked to the light blue branch situated in the bottom left area of the phylogenetic tree. Within this group, one particular catalyst, Na-Mn-W/Si, is distinguished by a thicker branch, signifying its status as the catalyst with the highest performance and stability

currently known in OCM[43,44]. It is hypothesized that the collective action of all components of this catalyst is essential for its performance, and the absence of any component leads to an immediate decline in efficacy[43,44]. This is corroborated by the background colors of Mn-W/Si, Na-W/Si, and Na-Mn/Si on the phylogenetic tree. The tree facilitates a gradual examination of the composition of the present catalyst. For instance, in Na-Mn-W/Si, replacing Na with another alkali metal is a common research focus. When substituting alkali metals, performance generally improves with the heavier substitutes. With the addition of another supporting element, it is noted that elements such as Ti, Zr, Ce, La, Ga, Mg, and Ag have been explored. From the phylogenetic tree, it is evident that the addition of Ga significantly hinders catalytic performance, while the incorporation of La leads to the most notable performance enhancement. Ti, Zr, and Ce are elements known for their multiple electronic states. In OCM, the transformation of methane into methyl radicals is a crucial initial reaction step that involves electron transfer, suggesting that these elements might be added to the catalyst to facilitate a redox reaction with methane on the catalyst's surface. The standard catalyst in OCM is recognized for its high selectivity at elevated temperatures, and introducing La to the current catalyst likely aims to combine this selectivity with lanthanum's low-temperature activity. For a catalyst with five elements loaded onto Si, the selection of combinations draws from those elements considered at the four-element stage. As the number of components reaches this level, the surface phenomena become exceedingly complex, rendering it challenging to deduce the processes at play. Thus, similar to the stage with four elements supported, catalyst design must rely on empirical combinations of the elements used to date.

A notable group of catalysts that has not yet been mentioned includes those with supports composed of two or more metals. These are typically referred to as mixed

oxide or perovskite-type catalysts. Mixed oxide catalysts essentially aim to combine two catalyst systems. In contrast, perovskite-type catalysts serve as model systems for understanding OCM due to their simple and definitive structure compared to other solid catalysts, such as Mn-Na_2WO_4/SiO_2[45,46]. Moreover, when used in conjunction with certain dopants or halogen-based precursors, these catalysts are also recognized for their considerable performance[47–49]. These catalysts are said to perform by producing electron holes in the perovskite due to the influence of dopants and precursors, causing the catalyst itself to behave like a p-type semiconductor and adsorbing gas-phase oxygen to the surface with form for high C_2 selectivity by electron holes moving across the surface[48].

In general, it is apparent that the standard OCM catalyst leads the way in terms of combinatory development compared to other catalyst families. The challenge in understanding combination development within OCM reactions is largely due to the extremely high reaction temperatures, typically around 800 °C, which complicates in situ or operando analysis. Consequently, development strategies have been predominantly empirical. To break through these challenges, there is a need to either refine analytical techniques for a deeper understanding or to establish statistical methods that can navigate the complexity of catalyst combinations as they currently exist. As discussed, the deployment of phylogenetic trees has allowed us to visualize the evolutionary history of OCM catalysts effectively.

4.3.2. Phylogenetic tree reconstructed with the addition of our catalysts

Our group, inclusive of my contributions, has integrated a total of 636 catalysts into the phylogenetic tree[26,50–52]. It is important to note that all catalysts developed

by our group were assessed under 135 uniform reaction conditions. However, for the sake of equity in the representation on the tree, the branch thickness has been normalized to the finest level. This adjustment is due to the fact that occurrences from a single research group should not be weighted equally with those more widely reported. To maintain clarity and prevent the tree from becoming overwhelmingly large, data representing less than 15% C_2 yield were omitted prior to determining the median catalytic performance. Subsequently, only those catalysts with an overlap occurrence of five or more were included in the tree. Out of the historically measured catalysts, and those evaluated in our study, 115 catalysts fulfilled these criteria. A phylogenetic tree was constructed using the selected catalysts, with those developed by our team denoted by a red star next to the catalyst name.

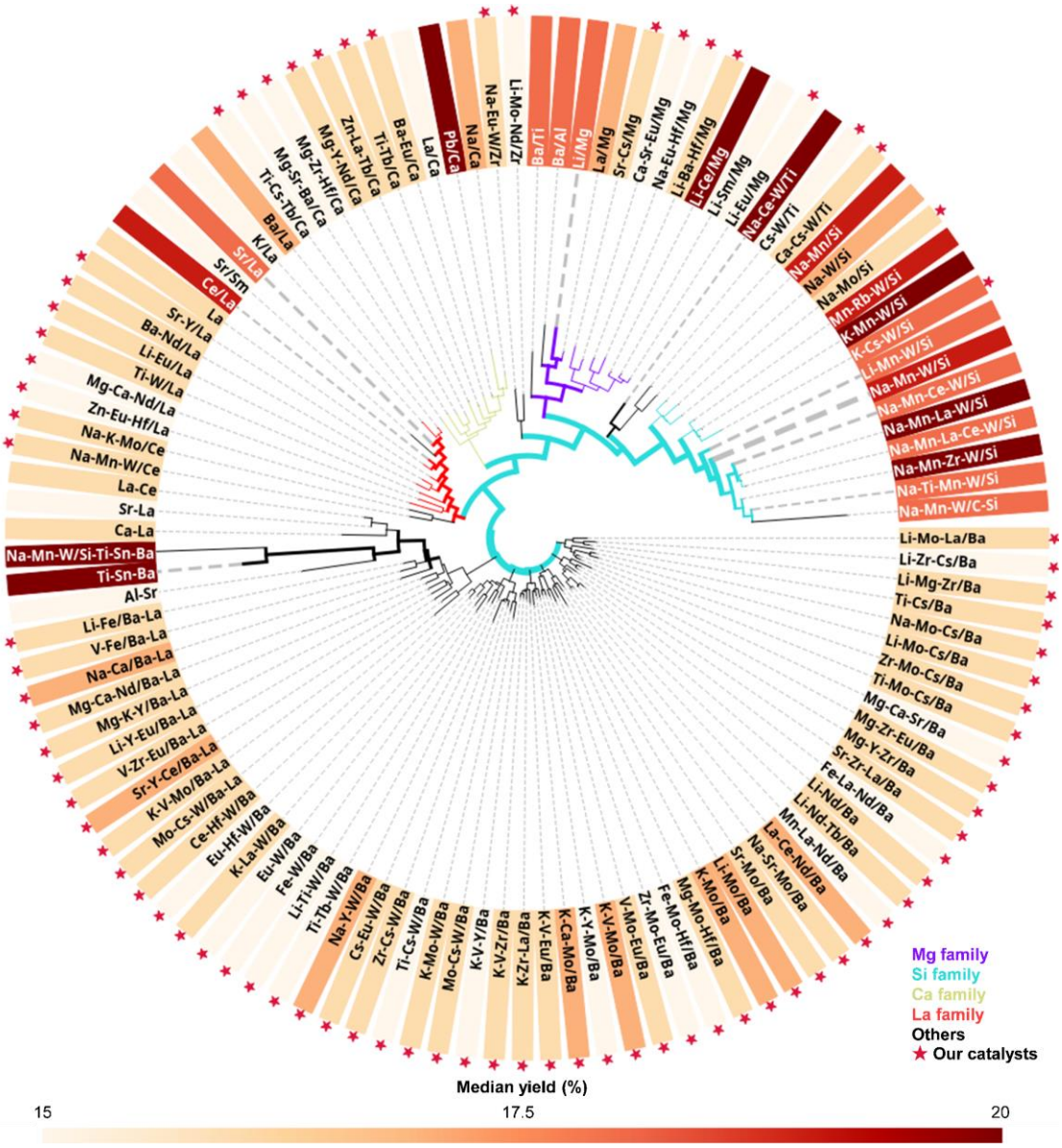


Figure 4.3. Catalyst phylogenetic tree featuring our catalysts.

Among the existing catalyst systems, I have successfully developed a number of catalysts based on Mg, Ca, and La. For Mg-based catalysts, I have reinforced the general rule that a binary support combining an alkali or alkaline earth metal with a lanthanide tends to perform well, as evidenced by our development of the Li-Eu/Mg catalyst. The Sr-Cs/Mg catalyst deviates slightly from this rule, as it does not incorporate lanthanides. Mg-based catalysts with three supported elements are scarce in published data, suggesting that our study has pioneered the discovery of numerous

such catalysts. Catalysts like Ca-Sr-Eu/Mg, Na-Eu-Hf/Mg, and Li-Ba-Hf/Mg all adhere to the general rule for binary support. The latter two, which include Hf, hint at the potential for higher-order combinations. Regarding Ca-based catalysts, prior to our study, the phylogenetic tree in figure 4.2 displayed only K-Ni/Ca as a catalyst with more than two supported elements. Our research has expanded this category significantly. Two noteworthy binary-supported catalysts are Ba-Eu/Ca and Ti-Tb/Ca, both lanthanide-supported. The addition of Cs to create Ti-Cs-Tb/Ca seemed promising, but a comparison of the background colors indicates superior performance without Cs. Other ternary-supported catalysts primarily incorporate alkali metals, alkaline earth metals, and lanthanides, following a combination rule similar to that for Mg. Unlike Mg, however, many Ca-based catalysts include group 4 elements like Ti and Zr, besides Hf. La-based catalysts, which were previously not represented in the phylogenetic tree with binary element combinations, have now been included thanks to our research. Generally, the performance follows rules akin to the historical design principles for Mg-based catalysts, with the exception of Ti-W/La, which appears to obey a different rule. For ternary combinations, Hf is a common inclusion, as seen in the previously mentioned examples. In the existing catalyst system based on Si, I have successfully developed Na-Mo/Si and K-Cs-W/Si. Both of these align with the Na-W component found in standard catalysts.

In beyond existing catalyst systems, a number of catalysts have been successfully developed using Ba as a support. In total, there are 45 Ba catalysts that meet the criteria displayed on the phylogenetic tree. Historically, Ba-based catalysts have not been extensively explored in the context of OCM, except for the Ti-Sn-Ba system. Considering that catalyst surface basicity is linked to performance, it would be expected that alkaline earth metals like Mg, Ca, Sr, and Ba would be ideal; however,

Sr and Ba tend to form stable carbonates under OCM conditions, which are thought to cover the active sites, hindering their effectiveness as catalysts[36]. Given that the OCM reaction is exothermic, the temperature at the reaction site is likely higher than the set reaction conditions[36]. Conversely, the endothermic decomposition of $BaCO_3$ could mitigate excessive temperatures by cooling the reaction zone, potentially preventing deep oxidation of C_2 products. Indeed, the catalytic system demonstrates remarkable selectivity at elevated temperatures. When examining the combination rule, a significant number of Ba-based catalysts are found to support group 6 elements, diverging from Mg and Ca catalysts. Conversely, there are scarce instances where the supported elements are solely alkali metals, alkaline earth metals, and lanthanides, with just three cases being Li-Nd/Ba, Li-Nd-Tb/Ba, and La-Ce-Nd/Ba. The incorporation of group 6 elements such as Mo and W, also present in standard OCM catalysts, indicates a shared design principle. Additionally, the performance of Si-based catalysts hinges on the inclusion of alkali metals and group 6 elements or Mn, suggesting that Ba-based catalysts are amenable to a broader array of combination possibilities. Beyond Ba-based catalysts, two catalysts were identified with Ti, Zr, and Ce as supports. The Ti-based system's Cs-W/Ti and Ca-Cs-W/Ti catalysts both feature a Cs-W/Ti ternary combination. Similarly, the historically high-performing Na-Ce-W/Ti catalyst also exhibits a combination of alkali metals and group 6 elements. It is notable that catalysts with Zr and Ce supports have been seldom investigated, with the only one meeting the criteria for inclusion on the phylogenetic tree being one I developed. The Zr-based system includes the Na-Eu-W/Zr and Li-Mo-Nd/Zr catalysts, both demonstrating lanthanide support alongside alkali metals and group 6 elements. In the Ce system, the Na-K-Mo/Ce and Na-Mn-W/Ce catalysts were developed. These three systems exhibit design rules highly akin to those of standard catalysts.

In summary, there are shared design principles across Mg-, Ca-, and La-based catalysts, where a foundational rule involves supporting alkali metals, alkaline earth metals, and lanthanides. The inclusion of group 4 elements is also effective once these basic criteria are satisfied. In the systems involving Si, Ti, Zr, and Ce, the combination rules differ from those for Mg, Ca, and La systems. The essential combination rule for high performance is the presence of an alkali metal coupled with a redox-active transition metal from a prior period. Ba-based catalysts conform to both aforementioned sets of combination rules; however, the first rule's applicability is more limited, while the second rule enjoys broader applicability.

4.3.3. Catalyst discoveries based on phylogenetic tree

The phylogenetic tree facilitates a detailed visual exploration of the development history of OCM catalysts. Reviewing the high-performance catalysts, examples such as Na-Mn-W/Si-Ti-Sn-Ba, and Na-Ce-W/Ti, Na-Mn-La-W/Si emerge, which were conceived by applying combination rules and empirical knowledge derived from other catalyst systems. Hence, the combination rules for supported metals identified through historical OCM catalyst development and this doctoral research had been employed for the design of catalysts, through lateral application across different supports. The catalyst design is concentrated on mixed oxide support catalysts, which are relatively under-explored in OCM history. Special attention was given to Ca-La and Ti-Sn-Ba, which show promising performance on the phylogenetic tree. I restricted the investigation to these supports, assessing whether the design rules established for each catalyst system could be extrapolated to these mixed oxides. The designed catalysts, their reference, and design strategies are summarized in Table 4.2.

Table 4.2. Designed catalysts based on the extracted knowledge from the phylogenetic tree.

	Designed catalysts	Reference catalysts	comment
1	Mg-Zr-Hf/Ca-La	Mg-Zr-Hf/Ca	our Ca catalyst supporting Hf
2	Zn-Eu-Hf/Ca-La	Zn-Eu-Hf/La	our La catalyst supporting Hf
3	Mg-Sr-Ba/Ca-La	Mg-Sr-Ba/Ca	our Ca catalyst not to support Hf
4	Mg-Ca-Nd/Ca-La	Mg-Ca-Nd/La	our La catalyst not to support Hf
5	Ti-W/Ca-La	Ti-W-W/La	our La catalyst supporting only transition metal
6	-	Ca-La	for comparison
7	La-Ce-Nd/Ti-Sn-Ba	La-Ce-Nd/Ba	our Ba catalyst supporting only lanthanides
8	Ca-Cs-W/Ti-Sn-Ba	Ca-Cs-W/Ti	our Ti catalyst
9	Zr-Cs-W/Ti-Sn-Ba	Zr-Cs-W/Ba	our Ba catalyst supporting 2 transition metals
10	Na-Y-W/Ti-Sn-Ba	Na-Y-W/Ba	our Ba catalyst supporting Na-W
11	Na-Eu-W/Ti-Sn-Ba	Na-Eu-W/Zr	our Zr catalyst
12	Na-Eu-Hf/Ti-Sn-Ba	Na-Eu-Hf/Mg	our Mg catalyst
13	Mg-Zr-Hf/Ti-Sn-Ba	Mg-Zr-Hf/Ca	our Ca catalyst
14	-	Ti-Sn-Ba	for comparison

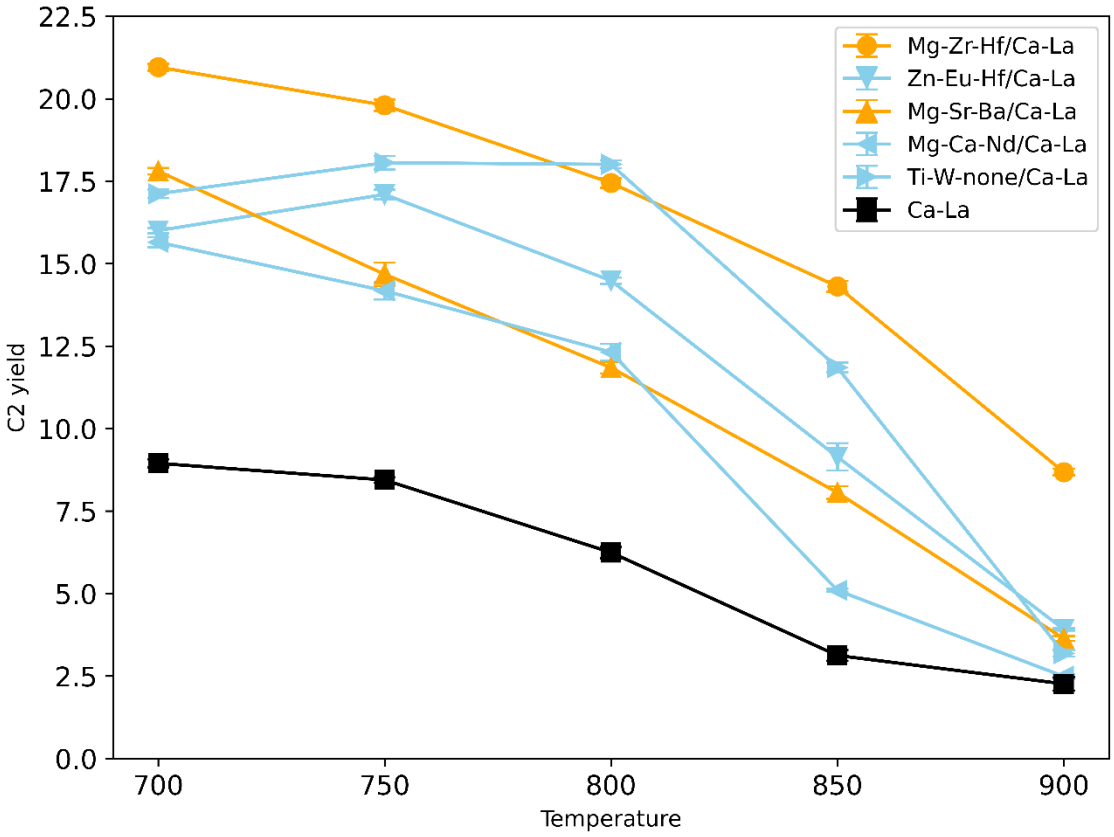


Figure 4.4. Relationship between temperature and performance of Ca-La-based catalysts

The results for catalysts with a Ca-La support are summarized in figure 4.4, and those for the Ti-Sn-Ba catalyst are outlined in figure 4.5. Figure 4.4 displays the points of highest yield at each temperature for each catalyst, with varying colors indicating the support material of the reference catalysts. A distinct divergence in the behavior of the catalysts is observed, contingent on whether the reference catalyst support was Ca or La. The Mg-Zr-Hf/Ca-La and Mg-Sr-Ba/Ca-La catalysts, referencing Ca-based supports, exhibited enhanced performance at lower temperatures, with their peak performance at 700 °C. Notably, Mg-Zr-Hf/Ca-La achieved a yield of 20.96% at this temperature. In the case of Zn-Eu-Hf/Ca-La, Mg-Ca-Nd/Ca-La, and Ti-W/Ca-La,

where the reference catalyst support was La-based, the temperature range for optimal performance shifts to higher temperatures with an increasing count of supported all-period transition metal elements. This shift could be ascribed to the selective formation of active sites akin to those in standard catalysts through the incorporation of early transition metal elements. Another explanation could be that the increased selectivity at higher temperatures results from the supported metal partially blending with the support to create a perovskite-like structure, thereby inducing electron holes. In either case, the low-temperature activity instead be reduced due to a decrease in the La contribution on the surface.

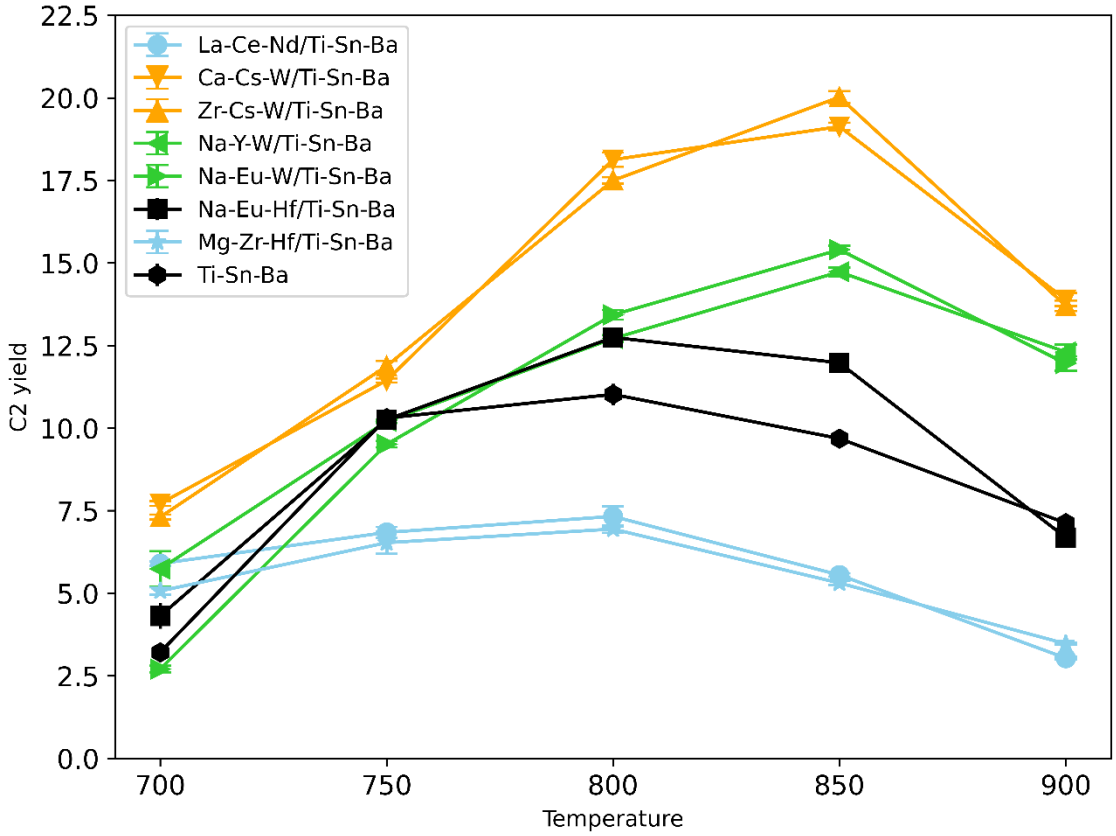


Figure 4.5. Relationship between temperature and performance of Ca-La-based catalysts

Figure 4.5 similarly illustrates the maximum performance of each catalyst across various temperature ranges. Due to the large number of reference catalyst supports in figure 4.5, the colors are categorized based on similarities in temperature response. This response suggests that the catalysts can be classified into four types, with discernible commonalities among them. Firstly, the two catalysts highlighted in orange, which represent the cluster with the highest performance, are both loaded with Cs-W. Our studies have revealed that the Cs-W/Ti combination exhibits high performance. The next two catalysts, displayed in green and offering medium performance, are Na-Y-W/Ti-Sn-Ba and Na-Eu-W/Ti-Sn-Ba, both incorporating Na-W. While Na-W is acknowledged as an active site in standard catalysts, its performance was inferior to that of Cs-W. This may be attributed to the ionic radius of Cs being more similar to that of Ba than Na is, enhancing the likelihood of substitution or doping between lattice Ba and Cs on the catalyst surface. The introduction of Cs^{+} ions, which differ in valence from Ba^{2+} sites, might induce an electronic distortion that enhances OCM selectivity and contributes to heightened activity at elevated temperatures[48]. Our attention then shifts to the two light blue catalysts, La-Ce-Nd/Ti-Sn-Ba and Mg-Zr-Hf/Ti-Sn-Ba. The reference catalyst for the former is Ba-based, with support comprised solely of lanthanides. This suggests that the design rule for basic metal-based catalysts may also be applicable to Ba. Conversely, the Mg-Zr-Hf/Ti-Sn-Ba catalyst, with its reference being Ca-based, implies that the design rule for basic metal supports is not transferable to the Ti-Sn-Ba support. The Na-Eu-Hf/Ti-Sn-Ba, depicted with a black line, was initially a Mg-based catalyst; hence, the performance improvement was not as pronounced as it might have been.

4.4. Conclusions

In this chapter, I demonstrated a new method for visualizing the development of elemental combinations in catalysts using a phylogenetic tree approach. This method was applied to the study of OCM, enabling a representation of the historical progression in the development of these combinations. Additionally, the integration of their distinctive catalysts into the existing framework of OCM development helped the researchers to modestly demonstrate the uniqueness of their contributions. Building on these insights, I implemented a horizontal propagation strategy for the design of new catalysts, which resulted in catalysts showing promising performance at lower temperatures.

References

[1] C. Yu, J. He, "Synergic catalytic effects in confined spaces," *Chem. Commun.*, **2012**, 48, 41, 4933–4940. doi: 10.1039/c2cc31585h.

[2] Y. Ying, X. Luo, J. Qiao, H. Huang, "'More is Different:' Synergistic Effect and Structural Engineering in Double-Atom Catalysts," *Adv. Funct. Mater.*, **2021**, 31, 3. doi: 10.1002/adfm.202007423.

[3] A. G. Dedov, A. S. Loktev, I. I. Moiseev, A. Aboukais, J. F. Lamonier, I. N. Filimonov, "Oxidative coupling of methane catalyzed by rare earth oxides: Unexpected synergistic effect of the oxide mixtures," *Appl. Catal. A Gen.*, **2003**, 245, 2, 209–220. doi: 10.1016/S0926-860X(02)00641-5.

[4] P. Chandra, A. M. Jonas, A. E. Fernandes, "Spatial Coordination of Cooperativity in Silica-Supported Cu/TEMPO/Imidazole Catalytic Triad," *ACS Catal.*, **2018**, 8, 7, 6006–6011. doi: 10.1021/acscatal.8b01399.

[5] J. M. Notestein, A. Katz, "Enhancing heterogeneous catalysis through cooperative hybrid organic-inorganic interfaces," *Chem. - A Eur. J.*, **2006**, 12, 15, 3954–3965. doi: 10.1002/chem.200501152.

[6] A. Thakur, R. Baba, T. Wada, P. Chammingkwan, T. Taniike, "Cooperative Catalysis by Multiple Active Centers of a Half-Titanocene Catalyst Integrated in Polymer Random Coils," *ACS Catal.*, **2019**, 9, 4, 3648–3656. doi: 10.1021/acscatal.9b00214.

[7] J. W. Thybaut, J. Sun, L. Olivier, A. C. Van Veen, C. Mirodatos, G. B. Marin, "Catalyst design based on microkinetic models: Oxidative coupling of methane," *Catal. Today*, **2011**, 29–36. doi: 10.1016/j.cattod.2010.09.002.

[8] G. Kumar, E. Nikolla, S. Linic, J. W. Medlin, M. J. Janik, "Multicomponent Catalysts: Limitations and Prospects," *ACS Catal.*, **2018**, 8, 4, 3202–3208. doi: 10.1021/acscatal.8b00145.

[9] T. Taniike, M. Terano, “Coadsorption model for first-principle description of roles of donors in heterogeneous Ziegler-Natta propylene polymerization,” *J. Catal.*, **2012**, 293, 39–50. doi: 10.1016/j.jcat.2012.06.001.

[10] J. M. Thomas, “Design, synthesis, and in situ characterization of new solid catalysts,” *Angew. Chem. Int. Ed.*, **1999**, 38, 24, 3588–3628. doi: 10.1002/(SICI)1521-3773(19991216)38:24<3588::AID-ANIE3588>3.0.CO;2-4.

[11] G. P. Szijjártó, Z. Pászti, I. Sajó, A. Erdohelyi, G. Radnóczi, A. Tompos, “Nature of the active sites in Ni/MgAl2O4-based catalysts designed for steam reforming of ethanol,” *J. Catal.*, **2013**, 305, 290–306. doi: 10.1016/j.jcat.2013.05.036.

[12] J. C. Védrine, “Revisiting active sites in heterogeneous catalysis: Their structure and their dynamic behaviour,” *Appl. Catal. A Gen.*, **2014**, 474, 40–50. doi: 10.1016/j.apcata.2013.05.029.

[13] T. Kobayashi, T. Yamada, K. Kayano, “Effect of basic metal additives on NOx reduction property of Pd-based three-way catalyst,” *Appl. Catal. B* **2001**, 30, 3-4, 287-292. doi: 10.1016/S0926-3373(00)00240-X.

[14] T. P. N. Tran, A. Thakur, D. X. Trinh, A. T. N. Dao, T. Taniike, “Design of Pd@Graphene oxide framework nanocatalyst with improved activity and recyclability in Suzuki-Miyaura cross-coupling reaction,” *Appl. Catal. A Gen.*, **2018**, 549, 60–67. doi: 10.1016/j.apcata.2017.09.026.

[15] R. Schlögl, “Heterogeneous catalysis,” *Angew. Chem. Int. Ed.*, **2015**, 54, 11, 3465–3520. doi: 10.1002/anie.201410738.

[16] B. R. Goldsmith, J. Esterhuizen, J. X. Liu, C. J. Bartel, C. Sutton, “Machine learning for heterogeneous catalyst design and discovery,” *AIChE J.*, **2018**, 64, 7, 2311–2323. doi: 10.1002/aic.16198.

[17] J. S. Kim, B. Kim, H. Kim, K. Kang, “Recent Progress on Multimetal Oxide Catalysts for the Oxygen Evolution Reaction,” *Adv. Energy Mater.*, **2018**, 8, 11. doi: 10.1002/aenm.201702774.

[18] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. I. Shimizu, “Machine Learning for Catalysis Informatics: Recent Applications and Prospects,” *ACS Catal.*, **2020**, 10, 3, 2260–2297. doi: 10.1021/acscatal.9b04186.

[19] M. Erdem Günay, R. Yıldırım, “Recent advances in knowledge discovery for heterogeneous catalysis using machine learning,” *Catal. Rev. Sci. Eng.*, **2021**, 63, 1, 120–164. doi: 10.1080/01614940.2020.1770402.

[20] Y. Y. Chen, M. Ross Kunz, X. He, R. Fushimi, “Recent progress toward catalyst properties, performance, and prediction with data-driven methods,” *Curr. Opin. Chem. Eng.*, **2022**, 37. doi: 10.1016/j.coche.2022.100843.

[21] U. Zavyalova, M. Holena, R. Schlögl, M. Baerns, “Statistical analysis of past catalytic data on oxidative methane coupling for new insights into the composition of high-performance catalysts,” *ChemCatChem*, **2011**, 3, 12, 1935–1947. doi: 10.1002/cctc.201100186.

[22] S. Mine et al., “Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts,” *ChemCatChem*, **2021**, 13, 16, 3636–3655. doi: 10.1002/cctc.202100495.

[23] Ç. Odabaşi, M. E. Günay, R. Yildirim, “Knowledge extraction for water gas shift reaction over noble metal catalysts from publications in the literature between 2002 and 2012,” *Int. J. Hydrogen Energy*, **2014**, 39, 11, 5733–5746. doi: 10.1016/j.ijhydene.2014.01.160.

[24] A. N. Şener, M. E. Günay, A. Leba, R. Yıldırım, “Statistical review of dry reforming of methane literature using decision tree and artificial neural network analysis,” *Catal. Today*, **2018**, 299, 289–302. doi: 10.1016/j.cattod.2017.05.012.

[25] T. N. Nguyen et al., “High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane,” *ACS Catal.*, **2020**, 10, 2, 921–932. doi: 10.1021/acscatal.9b04293.

[26] T. N. Nguyen et al., “Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane,” *ACS Catal.*, **2021**, 11, 3, 1797–1809. doi: 10.1021/acscatal.0c04629.

[27] P. Raccuglia et al., “Machine-learning-assisted materials discovery using failed experiments,” *Nature*, **2016**, 533, 7601, 73–76. doi: 10.1038/nature17439.

[28] C. D. Christ, M. Zentgraf, J. M. Kriegl, “Mining electronic laboratory notebooks: Analysis, retrosynthesis, and reaction based enumeration,” *J. Chem. Inf. Model.*, **2012**, 52, 7, 1745–1756. doi: 10.1021/ci300116p.

[29] S. Ma, X. Guo, L. Zhao, S. Scott, and X. Bao, “Recent progress in methane dehydroaromatization: From laboratory curiosities to promising technology,” *J. Energy Chem.* **2013**, 22, 1, 1-20. doi: 10.1016/S2095-4956(13)60001-7

[30] V. Busico, R. Pellecchia, F. Cutillo, R. Cipullo, “High-throughput screening in olefinpolymerization catalysis: from serendipitous discovery towards rational understanding,” *Macromol. Rapid Commun.*, **2009**, 30, 20, 1697–1708. doi: 10.1002/marc.200900246.

[31] R. Vicente-Saez, C. Martinez-Fuentes, “Open Science now: A systematic literature review for an integrated definition,” *J. Bus. Res.*, **2018**, 88, 428–436. doi: 10.1016/j.jbusres.2017.12.043.

[32] P. Kapli, Z. Yang, M. J. Telford, “Phylogenetic tree building in the genomic age,” *Nat. Rev. Genet.*, **2020**, 21, 7, 428–444. doi: 10.1038/s41576-020-0233-0.

[33] H. Fangerau, H. Geisler, T. Halling, W., Classification and evolution in biology, linguistics and the history of science;concepts, Franz Steiner Verlag, **2013**.

[34] S. Wu, G. Lambard, C. Liu, H. Yamada, R. Yoshida, "iQSPR in XenonPy: A Bayesian Molecular Design Algorithm," *Mol. Inform.*, **2020**, 39, 1–2. doi: 10.1002/minf.201900107.

[35] N. Saitou, M. Nei, "The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees." *Mol Biol Evol.* **1987**, 4, 406-25. doi: 10.1093/oxfordjournals.molbev.a040454.

[36] C. A. Ortiz-Bravo, C. A. Chagas, F. S. Toniolo, "Oxidative coupling of methane (OCM): An overview of the challenges and opportunities for developing new technologies," *J. Nat. Gas Sci. Eng.*, **2021**, 96. doi: 10.1016/j.jngse.2021.104254.

[37] T. Ito, J.-X. Wang, C.-H. Lin, J. H. Lunsford, "Oxidative Dimerization of Methane over a Lithium-Promoted Magnesium Oxide Catalyst," *J. Am. Chem. Soc.* **1985**, 107, 18, 5062–5068. doi: 10.1021/ja00304a008

[38] L. Thum et al., "Oxygen Activation in Oxidative Coupling of Methane on Calcium Oxide," *J. Phys. Chem. C*, **2019**, 123, 13, 8018–8026. doi: 10.1021/acs.jpcc.8b07391.

[39] J. Liu et al., "From fundamentals to chemical engineering on oxidative coupling of methane for ethylene production: A review," *Carbon Resour. Convers.*, **2022**, 5, 1, 1–14. doi: 10.1016/j.crcon.2021.11.001.

[40] J. Ohyama et al., "Direct design of active catalysts for low temperature oxidative coupling of methane via machine learning and data mining," *Catal. Sci. Technol.*, **2021**, 11, 2, 524–530. doi: 10.1039/d0cy01751e.

[41] L. Pirro, A. Obradović, B. D. Vandegehuchte, G. B. Marin, J. W. Thybaut, "Model-Based Catalyst Selection for the Oxidative Coupling of Methane in an Adiabatic Fixed-Bed Reactor," *Ind. Eng. Chem. Res.*, **2018**, 57, 48, 16295–16307. doi: 10.1021/acs.iecr.8b04242.

[42] V. I. Alexiadis et al., "Oxidative coupling of methane: Catalytic behaviour assessment via comprehensive microkinetic modelling," *Appl. Catal. B*, **2014**, 150–151, 496–505. doi: 10.1016/j.apcatb.2013.12.043.

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

[44] S. Arndt, T. Otremba, U. Simon, M. Yildiz, H. Schubert, R. Schomäcker, "Mn-Na_2WO_4/SiO_2 as catalyst for the oxidative coupling of methane. What is really known?," *Appl. Catal. A Gen.*, **2012**, 425–426, 53–61. doi: 10.1016/j.apcata.2012.02.046.

[45] Z. Taheri, N. Seyed-Matin, A. A. Safekordi, K. Nazari, S. Z. Pashne, "A comparative kinetic study on the oxidative coupling of methane over LSCF perovskite-type catalyst," *Appl. Catal. A Gen.*, **2009**, 354, 1–2, 143–152. doi: 10.1016/j.apcata.2008.11.017.

[46] N. Yaghobi, M. Hamid, R. Ghoreishy, "Oxidative coupling of methane in a fixed bed reactor over perovskite catalyst: A simulation study using experimental kinetic model," *J Nat Gas Sci Eng* **2008**, 17, 1, 8-16. doi: 10.1016/S1003-9953(08)60019-5.

[47] J. Xu et al., "Design of strontium stannate perovskites with different fine structures for the oxidative coupling of methane (OCM): Interpreting the functions of surface oxygen anions, basic sites and the structure–reactivity relationship," *J. Catal.*, **2022**, 408, 465–477. doi: 10.1016/j.jcat.2021.04.007.

[48] J. Xu et al., "Promoting the surface active sites of defect $BaSnO_3$ perovskite with $BaBr_2$ for the oxidative coupling of methane," *Catal. Today*, **2021**, 374, 29–37. doi: 10.1016/j.cattod.2020.12.021.

[49] Z. Wang et al., "Oxidative coupling of methane over $BaCl_2$-TiO_2-SnO_2 catalyst," *J. Nat. Gas Chem.*, **2012**, 21, 1, 49–55. doi: 10.1016/S1003-9953(11)60332-0.

[50] S. Nakanowatari et al., "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. doi: 10.1002/cctc.202100460.

[51] K. Takahashi et al., “Catalysis Gene Expression Profiling: Sequencing and Designing Catalysts,” *J. Phys. Chem. Lett.*, **2021**, 12, 30, 7335–7341. doi: 10.1021/acs.jpclett.1c02111.

[52] L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike, K. Takahashi, “Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts,” *Chem. Sci.*, **2021**, 12, 38, 12546–12555. doi: 10.1039/d1sc04390k.

Chapter 5

General conclusion

The historical development of solid catalysts has largely depended on trials and errors, along with occasional serendipitous discoveries, due to their inherent complexity. In this traditional approach, researchers tend to focus on superior catalysts that have been discovered in the course of the research Unfortunately, the valuable tacit knowledge that researchers accumulate during the discovering process often goes unshared. This lack of sharing the knowledge makes it challenging to establish guidelines for future catalyst development, trapping the field in a continuous cycle of trials and errors. In the current era, with increasing demands for catalysts, there is a push to develop catalysts with more intricate compositions. This necessitates exploring a vast search space, making the traditional approach not anymore feasible in the face of time and resource constraints. My thesis challenged this limitation by establishing a reproducible and efficient method for catalyst development, which is based on element substitution-centered catalyst informatics . The following is a summary of the work conducted in each chapter.

In **Chapter 2**, the development of catalysts focused on enhancing low-temperature activity and high-temperature selectivity, employing four strategies. The first strategy involved modifying the La_2O_3 catalyst by replacing active elements to optimize performance. The second strategy followed a similar modification approach with BaO. The third and fourth strategies involved selecting the best performing La_2O_3 and BaO-supported catalysts, respectively, and substituting their supports with a mix of BaO and La_2O_3, while retaining the active elements. These approaches are rooted in the heuristic that La_2O_3 boosts low-temperature activity and BaO enhances high-temperature selectivity. The aim was to improve selectivity in La_2O_3 and activity in BaO by combining active elements or using mixed supports. In addition, the OCM

activity of the supports alone was also assessed to determine the most effective way to combine them.

In **Chapter 3**, I presented a significant advancement in catalyst development with an algorithm focusing on elemental substitution. This algorithm is a key component of a newly designed catalyst recommender system, which is further enhanced by integrating an adaptive sampling strategy. The system's catalyst recommendations were tested through high-throughput experimentation. A notable point introduced in this chapter is the 'catalyst serendipiter system.' This advanced feature within the recommender system is adept at predicting serendipitous discoveries, which, while rare, can occur during adaptive sampling. This development marks a comprehensive and systematic approach to catalyst development, uniquely combining exploration, exploitation, and the serendipity. A practical outcome of this methodology was the identification of promising and novel OCM catalyst candidates, such as Na-Y-Tb/Al_2O_3 and Ti-Y-La/ZrO_2. This chapter showed the potential of this integrated approach in discovering innovative catalysts, showcasing how combining systematic algorithms with an element of unpredictability can lead to groundbreaking results in catalyst research.

Chapter 4 introduced a methodology that traces the evolutionary path of elemental combinations in catalysts using a phylogenetic tree approach. This novel method was applied to the study of OCM, providing a clear and detailed visual representation of the historical development of elemental combinations in OCM catalysts. Additionally, the set of catalysts developed in this thesis was incorporated into the analysis to visualize the contribution of this study to the field of OCM. Building upon the insights from phylogenetic tree, the catalyst design focused on mixed oxide support catalysts, which have historically not been examined very much. This approach

resulted in the creation of catalysts that exhibit superior performance at lower temperatures.

I believe that the research conducted in this thesis demonstrates the benefits of using catalyst informatics based on the elemental substitution strategy for catalyst development. This concept has broad applicability across various catalyst systems. The thesis is poised to offer a novel direction in catalyst research and development, potentially transforming how we approach and innovate in this field.

List of Publications and Other Achievements

Sunao Nakanowatari

A) PUBLICATION

1. "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.

2. "Direct Design of Catalysts in Oxidative Coupling of Methane via High‐Throughput Experiment and Deep Learning", Kanami Sugiyama, Thanh Nhat Nguyen, **Sunao Nakanowatari**, Itsuki Miyazato, Toshiaki Taniike, Keisuke Takahashi*, ChemCatChem, 2021, 13, 952-957.

3. "Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane", Thanh Nhat Nguyen, **Sunao Nakanowatari**, Thuy Phuong Nhat Tran, Ashutosh Thakur, Lauren Takahashi, Keisuke Takahashi, Toshiaki Taniike*, ACS Catalysis, 2021, 11, 1797-1809.

4. "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.

5. "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 (selected as Front Cover).

6. "Factors to influence low-temperature performance of supported Mn–Na_2WO_4 in oxidative coupling of methane", Thanh Nhat Nguyen, Kalaivani Seenivasan, **Sunao Nakanowatari**, Priyank Mohan, Thuy Phuong Nhat Tran, Shun Nishimura, Keisuke Takahashi, Toshiaki Taniike*, Molecular Catalysis 2021, 516, 111976 (pp. 1-9).

7. "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.

B) INTERNATIONAL CONFERENCE

1. "Catalyst design based on heuristics extracted from catalyst big data for oxidative coupling of methane", **Sunao Nakanowatari**, Thanh Nhat Nguyen, Ashutosh Thakur, Toshiaki Taniike, First international symposium on High-Throughput Catalyst Design (HTCD2021), online, Jun. 14-15, 2021, poster.
2. "Implementation of catalyst recommender system for oxidative coupling of methane", **Sunao Nakanowatari**, Keisuke Takahashi, Hieu Chi Dam, Toshiaki Taniike, ACS Spring 2023, Indianapolis, USA, March 26-30, 2023, poster.

C) DOMESTIC CONFERENCE

1. "証拠理論に基づく触媒推薦システムの開発とハイスループット実験を用いた実証", **中野渡 淳**, 藤原 綾, 髙橋 啓介, Dam Hieu Chi, 谷池 俊明, 日本化学会 第 102 春季年会, オンライン開催, 2022 年 3 月 23-26 日, ポスター.

D) BOOK

1. "ハイスループット実験と機械学習を用いた触媒設計", **中野渡 淳**, 髙橋 啓介, 谷池 俊明, ペトロテック, 2022, 45(12), 775-780.
2. "ハイスループット実験を基盤とするデータ駆動型触媒開発", 谷池 俊明, **中野渡 淳**, メタンと二酸化炭素 ～その触媒的化学変換技術の現状と展望～ (監修:上田 渉), シーエムシー出版, 2023, 322-329.