Infrared Spectroscopy Data-driven Machine learning for Unveiling Thermolysis Pathways of Metal-Organic Frameworks

By Yanzhang Zhao

School of Chemical Engineering and Advanced Materials

Faculty of Engineering, Computer and Mathematical Science

A Thesis submitted for the degree of Master of Philosophy

The University of Adelaide

Jan 2024

Table of Contents

Abstract

This project aims at developing machine-learning approach to unravel the thermolysis pathways of metal-organic frameworks (MOFs) into atomically doped metal oxide catalysts. The research methodology encompasses two main components, namely density functional theory (DFT) and a machine learning method based on the least absolute shrinkage and selection operator (LASSO). Starting with small amounts of experimental infrared spectroscopic data, the proposed machine-learning approach can extrapolate a more comprehensive infrared spectroscopy dataset. This augmented dataset is then employed to predict the pyrolysis pathways of MOF materials, providing valuable insight into the synthesis of single-atom catalysts (SACs). The innovation is that using infrared spectroscopy data-driven machine-learning for unveiling thermolysis pathways of MOFs is a powerful tool for understanding the behaviour of these materials and developing new applications for them.

Chapter 1 is the introduction and Chapter 2 presents a literature review. The advances and challenges of machine learning assisted reaction pathway finding in SACs. Then the application of machine learning tools, IR spectra, and MOFs thermolysis is studied and presented. Chapter 3 the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis. The synthesis of single-atom catalysts (SAC) through the pyrolysis of zeolitic imidazolate frameworks (ZIFs) represents a crucial pathway, and the mechanism can be examined using infrared (IR) spectroscopy. The results showed that the Pearson correlation exceeding 0.7 when compared to experimental data, the algorithm furnishes correlation coefficients for the chosen structures. This substantiates essential structural changes over time and temperature. Extends the study to other SACs formation from MOFs and the conclusions are drawn in Chapter 4, following the discussions of challenges and perspectives of machine learning on experimental graph recognition in reaction pathway exploration. The novel MOFs producing single atom catalyst provides a new platform for electrocatalyst development. This

approach possesses substantial potential for robustness and has the capability to be applied across a wide spectrum of applications for intelligent analysis of in situ experimental characterization data in the future.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

The author acknowledges that copyright of published works contained within the thesis resides with the copyright holder(s) of those works.

I acknowledge that the copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Name of Candidate: Yanzhang Zhao

Signature:

Date: 14/Jan/2023

Acknowledgments

Chapter 1: Introduction

1.1 Significance of the Project

Recently, there has been a remarkable advancement in the design of structures and foundational electrocatalytic studies concerning Single-Atom Catalysts (SACs)[1-4], since SACs provide an extremely high surface area, allowing for more active sites and efficient catalytic reactions, which sparked researchers interest and boosted extensive studies[2]. Nonetheless, single atoms have a propensity to aggregate during both synthetic and catalytic processes owing to their elevated surface energy. In this regard, the distinctive attributes of metal–organic frameworks (MOFs)[5], such as exceptional tailor ability, well-defined porous structures, effective design flexibility, and ultrahigh surface areas, make them ideally suited to meet the substrate requirements for stabilizing Single-Atom Catalysts (SACs) when compared to other conventional porous materials[6]. However, the exploration of their formation mechanisms is still in its infancy stage. This knowledge gap poses challenges when it comes to achieving precise and controlled syntheses of SACs from MOF thermolysis. As a result, Understanding the proper thermolysis reaction pathway which can forecast metal–metal interactions among densely populated metal single atoms are still urgently needed[2].

Over the past few years, machine learning has played a crucial role in extracting information from experimental graphs, particularly in-situ DRIFTS[7], showcasing considerable potential for data mining driven by ML techniques. Datasets in this procedure facilitate extraction of dynamic processes linked to the structural evolution of materials. A substantial database is essential for training the machine learning (ML) model, and density functional theory (DFT) calculations offer a rich source of data for this purpose[8]. Despite observed disparities between DFT-simulated IR spectra and experimental results, prior studies have uncovered a well-fitted linear scaling relationship between the calculated C–O bond stretching peak using DFT and

the corresponding experimental data[9]. As a result, the application of ML algorithms, particularly those based on linear regression, shows promise in bridging the gap between theoretical and experimental results, ultimately leading to superior fitting outcomes[10].

Discovering the vibrational mode in each atom bonding will facilitate the reaction pathway explanation and give positive feedback on practical instruction[11]. Due to substantial advancements in computing power, DFT calculations can now offer a precise depiction of the electronic structure of catalysts and vibration mode in each atom[12]. This can enrich the datasets for our machine-learning tool, it thus can help predict more accurate results to explain more complex reaction. This comparative analysis enables the deduction of the inherent formation mechanisms of SACs[13, 14].

The goal of this thesis is to study the reaction mechanisms of the Co_3O_4 from pristine ZIF-67 and Pt-doped ZIF-67 metal single atom supported MOF. Co_3O_4 was SAC product from the thermolysis during the study because of its essential role in the electrocatalysis field. Additionally, the amalgamation of machine learning algorithms with theoretical calculations for the analysis of in situ experimental spectra holds the potential to complement human expert interpretation. Consequently, this approach is poised to evolve into a more widely adopted and versatile methodology[15].

1.2 Research Objectives

The aim given to the study can be clearly obtained from the above statement is used machine learning method to explore the analysis thermolysis pathways of MOF. For this aim can be divided into three small objectives, they are:

1. To achieve a high correlation between the LASSO fitted IR spectroscopy and experimental data.

The objective of this statement is to state that the IR intensity of different bonds from DFT calculation, as identified by the LASSO model, has a high correlation with the experimental data. This suggests that the LASSO model can accurately predict the vibrational modes that contribute most strongly to the IR spectrum, and that these modes are important for understanding the molecular structure and chemical properties of the molecule.

To achieve this objective, a dataset of IR spectra and corresponding experimental data would need to be collected. The dataset would then be used to train a LASSO model, which would identify the vibrational modes that are most strongly correlated with the experimental data. The correlation between the predicted IR intensities and the experimental data would then be quantified using a statistical measure such as the Pearson correlation coefficient. If the correlation is high, it would suggest that the LASSO model is accurately predicting the important vibrational modes, and that these modes are crucial for understanding the molecule's structure and properties.

2. Analysis of the thermolysis mechanism of ZIF-67.

IR spectra provide information about the vibrational modes of molecules and are commonly used to identify the types of bonds present in a molecule. LASSO can be used to analyse IR spectra by identifying the most important vibrational modes that are correlated with experimental data.

By identifying the most important vibrational modes using LASSO, it is possible to gain insights into the chemical and physical processes that occur during the thermolysis of ZIF-67. For example, certain vibrational modes may be associated with specific chemical bonds that are broken or formed during the thermolysis process. By comparing the

vibrational modes to theoretical models or other experimental data, it may be possible to identify the intermediate species and reaction pathways that occur during thermolysis.

3. Analysis of the formation mechanism of single-atom catalyst.

The objective of this statement is to describe how the results of the thermolysis mechanism study of single-atom doped ZIF-67 can be compared with the previous thermolysis mechanism of ZIF-67.

The thermolysis mechanism of single-atom doped ZIF-67 was studied for Pt-ZIF-67. To investigate the thermolysis mechanism of Pt-doped ZIF-67, various techniques such as LASSO-fitted infrared spectroscopy can be employed. The results obtained from these techniques can provide insights into the chemical and physical processes that occur during thermolysis.

The results of the thermolysis mechanism study of Pt-doped ZIF-67 can be compared with the previous thermolysis mechanism of ZIF-67 to identify any differences that may arise from the presence of Pt. Comparing the vibrational modes and identifying any changes in the mechanism can provide insights into how the presence of Pt affects the thermolysis process.

Furthermore, the comparison of the results of the thermolysis mechanism study of Pt-doped ZIF-67 to theoretical models and computational simulations can provide a deeper understanding of the thermolysis mechanism and the factors that influence the formation of metal monomers and metal oxide nanoparticles.

1.3 Thesis Outline

This thesis is presented in the form of journal publications. The results of the reaction mechanisms and machine learning studied are included. Specifically, the chapters in the Thesis are presented in the following sequence:

➢ *Chapter 1* this thesis was introduced. It emphasizes the importance of this project and outlines the research objectives.

➢ *Chapter 2* reviews the recent progress and challenges of thermolysis pathways of MOFs and used for machine learning in reaction mechanism studies.

➢ *Chapter 3* presents machine learning confirms the formation mechanism of a Single-Atom Catalyst based on ZIF-67 and Pt doped ZIF-67.

➢ *Chapter 4* presents the conclusion and perspectives for further work on revealing the thermolysis mechanism in more kinds of MOFs.

1.4 References

1. Huang, H.; Shen, K.; Chen, F.; Li, Y. Metal–Organic Frameworks as a good platform for the fabrication of single-atom catalysts. ACS Catalysis 2020, 10 (12), 6579-6586. DOI: 10.1021/acscatal.0c01459.

2. Mitchell, S.; Perez-Ramirez, J. Single atom catalysis: a decade of stunning progress and the promise for a bright future. Nat Commun 2020, 11 (1), 4302. DOI: 10.1038/s41467-020-18182-5 From NLM PubMed-not-MEDLINE.

3. Li, X.; Yang, X.; Zhang, J.; Huang, Y.; Liu, B. In situ/operando techniques for characterization of single-atom catalysts. Acs Catalysis 2019, 9 (3), 2521-2531.

4. Kim, J.; Kim, H. E.; Lee, H. Single-atom catalysts of precious metals for electrochemical reactions. Chem Sus Chem 2018, 11 (1), 104-113.

5. Kitagawa, S. Metal–organic frameworks (MOFs). Chemical society reviews 2014, 43 (16), 5415-5418.

6. Shan, J.; Liao, J.; Ye, C.; Dong, J.; Zheng, Y.; Qiao, S. Z. The dynamic formation from metal-organic frameworks of high-density platinum single-atom catalysts with metal-metal interactions. Angew Chem Int Ed Engl 2022, 61 (48), e202213412. DOI: 10.1002/anie.202213412 From NLM PubMed-not-MEDLINE.

7. Fanning, P. E.; Vannice, M. A. A DRIFTS study of the formation of surface groups on carbon by oxidation. Carbon 1993, 31 (5), 721-730.

8. Bartolotti, L. J.; Flurchick, K. An introduction to density functional theory. Reviews in computational chemistry 1996, 187-216.

9. Cotton, F. A.; Kraihanzel, C. Vibrational spectra and bonding in metal carbonyls. I. Infrared spectra of phosphine-substituted group VI carbonyls in the CO stretching region. Journal of the American Chemical Society 1962, 84 (23), 4432-4438.

10. Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. Chemical science 2017, 8 (10), 6924-6935.

11. Kaiser, S. K.; Chen, Z.; Faust Akl, D.; Mitchell, S.; Perez-Ramirez, J. Single-atom catalysts across the periodic table. Chemical reviews 2020, 120 (21), 11703-11809.

12. Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. The journal of physical chemistry 1996, 100 (31), 12974-12980.

13. Porezag, D.; Pederson, M. R. Infrared intensities and Raman-scattering activities within density-functional theory. Physical Review B 1996, 54 (11), 7830.

14. Chen, H.; Zhang, Z.; Hu, D.; Chen, C.; Zhang, Y.; He, S.; Wang, J. Catalytic ozonation of norfloxacin using Co_3O_4/C composite derived from ZIF-67 as catalyst. Chemosphere 2021, 265, 129047.

15. Yang, X.-F.; Wang, A.; Qiao, B.; Li, J.; Liu, J.; Zhang, T. Single-atom catalysts: a new frontier in heterogeneous catalysis. Accounts of chemical research 2013, 46 (8), 1740-1748.

Chapter 2: Literature Review

2.1 Introductory Background

Due to the rising global concerns over environmental issues and the limited availability of fossil fuels, the search for sustainable and renewable energy sources has become a crucial challenge in today's world. One important aspect of achieving sustainable energy systems is through the use of electrochemical storage and conversion devices, such as fuel cells, water decomposition, and electrochemical reduction of carbon dioxide and nitrogen. In these high-performance energy devices, the performance of electrocatalysts plays a pivotal role in electrochemical reactions.[1-5]

Zeolitic imidazolate frameworks (ZIFs) are a class of metal-organic frameworks (MOFs) that have gained significant attention in recent years due to their unique structural properties and potential applications in various fields such as gas storage, separation, and catalysis.[6-8] One of the most promising applications of ZIFs is their use as precursors for the formation of SACs. ZIFs can serve as an excellent precursor for the formation of SACs due to their ability to encapsulate metal ions within their structure. By thermal treatment or chemical activation, the metal ions can be converted into single atoms, which are dispersed on the support material with a high degree of control and uniformity.[9] This allows to produce highly efficient SACs that can be tailored to specific applications.

The use of Infrared (IR) spectroscopy in the study of reaction mechanisms is based on the principle that different functional groups have characteristic vibrational frequencies.[10] Therefore, changes in the intensity or position of specific IR bands during a reaction can provide insights into the reaction mechanism and the formation of intermediates. By analysing

the changes in IR spectra at different reaction times, it is possible to obtain a detailed picture of the reaction pathway and the intermediates involved.[11]

In recent years, data-driven machine learning techniques, such as the LASSO algorithm, have been employed to analyse IR spectra and extract meaningful information about reaction mechanisms.[12-13] These methods can help to identify key features of the IR spectra that are important for understanding the reaction mechanism and can provide a more comprehensive analysis of complex reaction pathways.

2.2 Thermolysis mechanism of pristine ZIF-67

FTIR has been a great tool for analysing the bond information in material due to its manoeuvrability among all practical equipment. Wu et al.[14] conducted research on ZIF-67 to reveal those thermolysis routes in high temperature range.
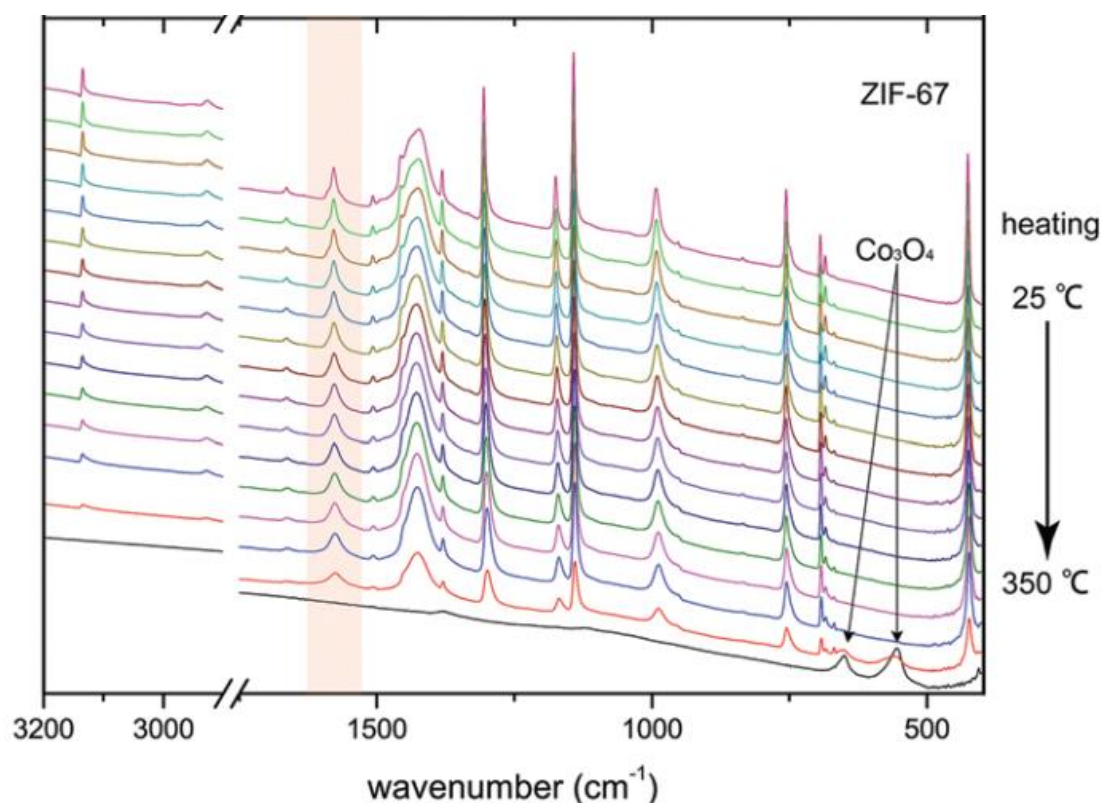

Figure 1 Temperature dependent in situ FTIR spectra of ZIF-8 and ZIF-67[14]

A noticeable difference appears in the peak related to "C=C" stretching (1580 cm^{-1}, as highlighted in Fig. 1). The peak was gradually broadened with the temperature rising to 350°C. Based on this information, the "C=C" was preliminary judged to be broken at this temperature, suggesting a direct collapse of its lattice structure without lattice deformation. However, this IR spectra contained different peaks and it may involve multiple bonds formation and break within one peak. Moreover, this research only provided the thermolysis routes in air from 25°C to 300°C, but it did not reveal how the metal-oxide bond was formed and the quantity that presents in the corresponded temperature.

2.3 Thermolysis mechanism of single-atom metal-doped ZIF-67

Understanding the mechanism by which the metal in ZIF-67 is oxidized to useful metal oxide catalysts is important in industrial catalyst production, as it enables the optimization of the synthesis process and facilitates control over the properties of the resultant catalysts. Shan et al. elucidated the formation of high-density Pt single atoms with inter-site interactions in derived Co_3O_4 host.[15] The resulting hybrid material is referred to as Pt-ZIF-67. In order to fully understand the driving force behind the construction of Pt-Pt interaction and the cleavage of organic ligands, the local evolution of metal-ligand coordination at Pt sites was analysed. The contour maps in Figures 2c, 2d and spectra revealed that a few characteristic peaks of ZIF-67 framework were observed in both Pt-ZIF-67 and ZIF-67.

As the thermolysis temperature increased, these peaks in Pt-ZIF-67 remained relatively stable until a sudden disappearance at around 300°C. This observation suggests that despite the host framework undergoing collapse from 250°C onwards, the organic ligands were retained until complete structured decomposition and transition to Co_3O_4 took place at 300°C. In contrast, ZIF-67 did not undergo such structural decomposition until being held at 300°C for more than 60 minutes. The X-ray diffraction analysis (XRD) and diffuse reflectance Fourier-transform

infrared spectroscopy (DRIFTs) observations confirmed that the fully phased transition of Pt-ZIF-67 occurred earlier than that of pristine ZIF-67, both of which were accompanied by the decomposition of 2-MeIm ligand. This result suggests that the incorporated Pt sites are possibly more vulnerable to be attacked by oxygen and therefore facilitate the phase transition to Co_3O_4.
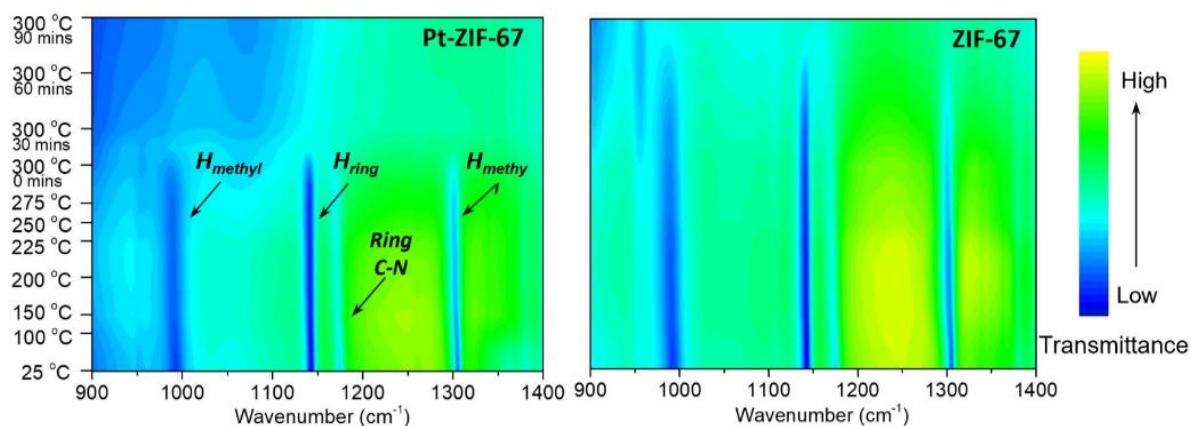

Figure 2 Pt-ZIF67 and pristine ZIF-67 with characteristic peaks indicated by arrows and labels.[15]

2.4 Machine-learning for Unveiling Thermolysis Pathways

The capability of machine learning (ML) to handle complex systems and make testable predictions has led to its application at the intersection of multiple disciplines. Although ML methods, including neural networks (NN), were proposed as early as the 1950s to 1970s, the application of ML in other fields was hindered by significant knowledge barriers. However, over the past decade, practical tools such as the Torch library, Scikit-Learn, and the Tensorflow library have reduced professional thresholds for users, [16] resulting in an increase in the application of ML in fields such as physics and chemistry[17-18], as shown in Figure 3.
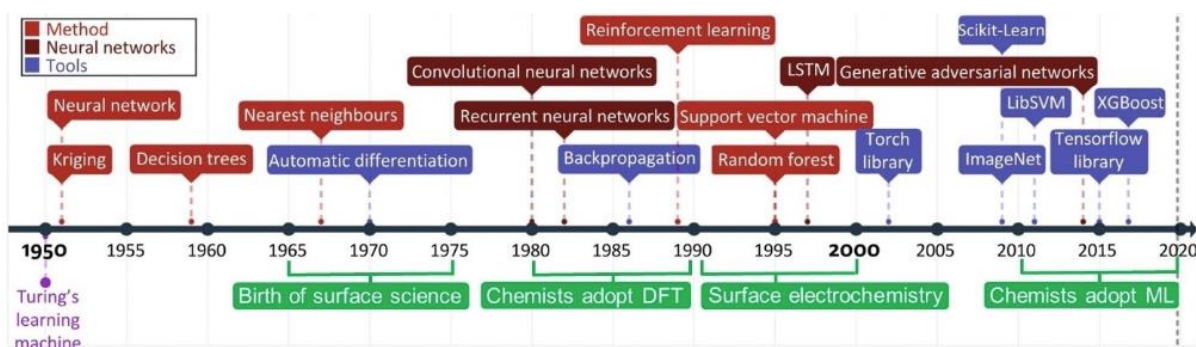
Figure 3. Pt-ZIF67 and pristine ZIF-67 with characteristic peaks indicated by arrows and labels.[19]

This increase in ML application has led to the emergence of machine learning interatomic potentials (MLIP)[20] and machine learning force fields (MLFF)[21], which have been used to accelerate DFT computation and improve precision in large systems. Furthermore, interpretable ML methods[22] are being used to understand various physical quantities on target properties and have resulted in suitable descriptors to make predictions. In addition, the inverse design approach[23] is being applied to derive theoretical structures from experimental characterizations with improved accuracy and efficiency.

 The optimal model is expected to be positioned in the upper left corner of the plot, even though it may lie in the infeasible region. Solutions that are located on the Pareto front near this ideal model are considered reasonable options. However, solutions with lower residuals may sacrifice the distribution of errors. On the other end of the Pareto front, some models may generate a nearly perfect normal error distribution, but they may not be effective in fitting the data. The best solution, therefore, must strike a balance between all objectives and should be chosen based on additional criteria and information. In the following section, we will discuss the distinctive characteristics of the Pareto front, which can serve as useful indicators for making valid decisions. A diagram summarizing all our approaches' relevant steps is included in the Figure 4.
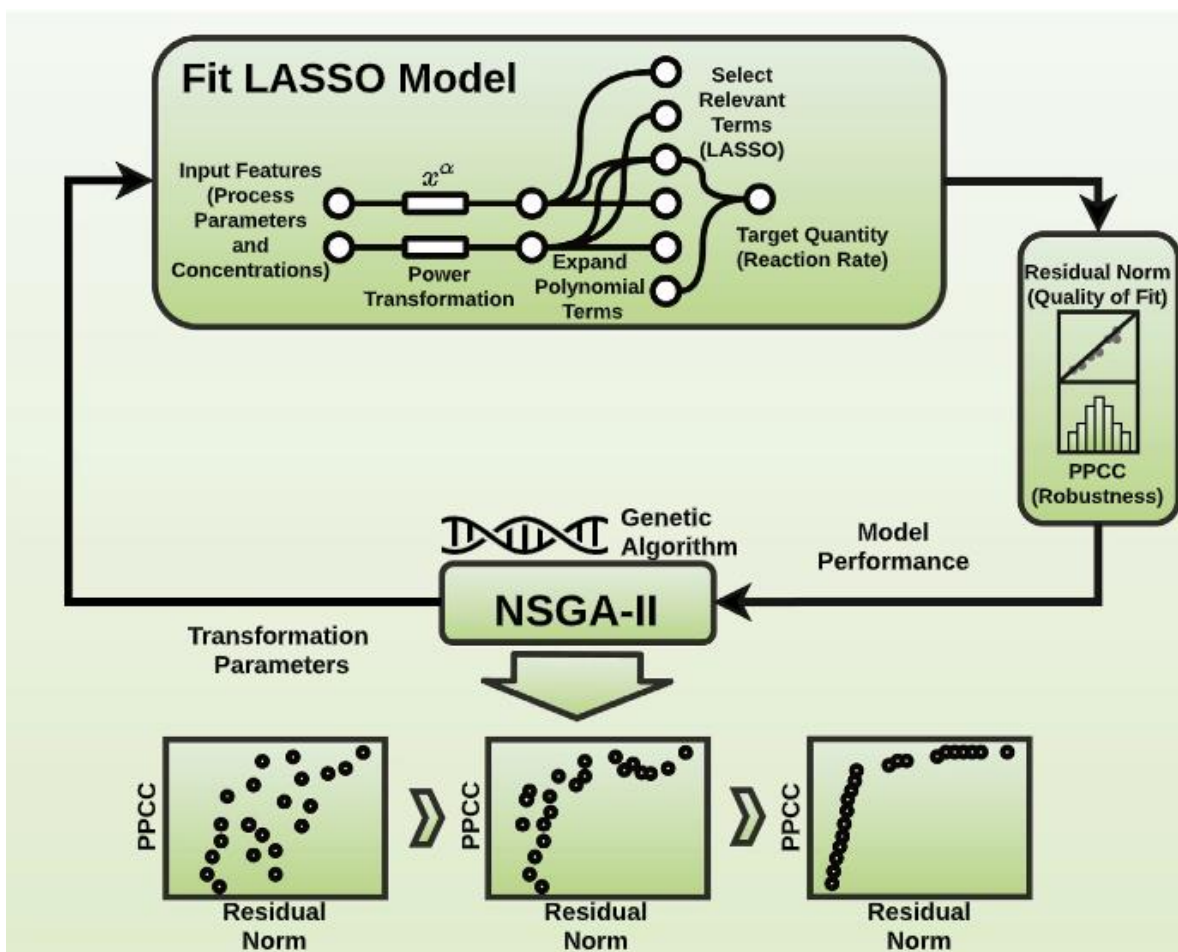
Figure 4. Pt-ZIF67 and pristine ZIF-67 with characteristic peaks indicated by arrows and labels.[24]

2.5 Methods

This section outlines the methods and procedures to be followed in order to achieve the aims of the study. This section may include a discussion of the sources to be consulted that will inform the selection and use of these methods.

DFT is used to simulate the Infrared Vibrational Spectrum

VASP (Vienna Ab initio Simulation Package) is a widely used DFT (Density Functional Theory) code for simulating the electronic structure and properties of materials. VASP uses a plane-wave basis set and the projector-augmented wave (PAW) method to describe the

electronic structure of a system. The PAW method is a variant of the pseudopotential method that includes a partial core correction to improve the accuracy of the calculation.

In practice, the process involves the following steps:

1.   Perform molecular dynamics simulations or quantum chemical calculations to obtain the positions and velocities of the atoms in the molecule.

2.   Calculate the dipole moment of the molecule as a function of time using the positions of the atoms.

3.   Calculate the dipole moment autocorrelation function from the dipole moment time series.

4.   Apply the DFT to the dipole moment autocorrelation function to obtain the IR vibrational spectrum.

The resulting IR spectrum can then be compared to experimental spectra or used to predict the vibrational properties of new molecules. The accuracy of the simulation depends on the quality of the input data and the level of theory used in the calculations.

Atomic Simulation Environment (ASE) is used as the interface to transform the DFT-calculated vibrational frequencies into spectra.

ASE is a Python library that provides a user-friendly interface for performing atomistic simulations and analysing the results. ASE includes various modules for performing DFT calculations, such as the popular DFT code VASP, as well as tools for analysing and visualizing the results.

To transform the DFT-calculated vibrational frequencies into spectra, ASE can be used in conjunction with other Python libraries such as NumPy and Matplotlib. The general process involves the following steps:

1.   Use ASE to perform a DFT calculation of the molecule of interest, including the calculation of the vibrational frequencies.

2.   Extract the vibrational frequencies from the DFT calculation using ASE's built-in tools.

3.   Use NumPy to convert the frequencies into wavelengths or wavenumbers that are suitable for plotting on an IR spectrum.

4.   Calculate the intensities of each vibrational mode based on the transition dipole moments, which can also be obtained from the DFT calculation.

5.   Plot the resulting IR spectrum using Matplotlib or another plotting library.

VESTA is used for visualization.

VESTA is a 3D visualization software designed for analysing electronic structures. Its full name, Visualisation for Electronic Structural Analysis, reflects its capabilities. The software is suitable for displaying structural models, volumetric data such as electron or nuclear densities, and crystal morphologies. The program's features allow users to work with many objects, such as atoms, bonds, polyhedral, and polygons on is surfaces. Multiple windows can be opened simultaneously, with each window capable of containing numerous tabs that correspond to different files. With VESTA, users have the flexibility to visualize and analyse complex structures with ease. So that I can use this software to view the structure and intermediate that been built in this project easily.

LASSO in machine learning:

LASSO regression tries to find the best fitting line while simultaneously shrinking the less important feature coefficients to zero. This makes LASSO a useful technique for feature selection in high-dimensional datasets, where the number of features is much larger than the number of observations. By setting the coefficients of unimportant features to zero, LASSO effectively eliminates those features from the model, reducing overfitting and improving generalization performance.

LASSO regression can be solved using optimization algorithms such as coordinate descent, which iteratively updates the regression coefficients. The strength of the penalty term is controlled by a hyperparameter called the regularization parameter, which can be tuned using techniques such as cross-validation.

The capability of machine learning (ML) to handle complex systems and make testable predictions has led to its application at the intersection of multiple disciplines. LASSO[24-25] is the machine learning algorithm used for fitting the experimental data and bond feature selection.

The simplest linear regression correlates with the infrared intensities. And the y-intercept β_0 and the forms were shown as:

$$y = \beta X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_p x_p$$

The LASSO algorithm aims to figure out the best-fit model while concurrently driving some of the IR's model coefficients to enclose to zero, which will lead to a better interpretable scattered model.

$$\sum_{i=1}^{n} (y_i - \beta X_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The X here represents the bonds IR intensities simulated by ASE Infrared intensities class. Y is the experimental data that can be regarded as the target set of LASSO machine learning. This process was repeated for 'X' IR intensities resulting in 'X' linear transformation. Then, a series of shrinkage in the coefficients is determined by the complexity parameter α. The strength of this 'balance' term, is hyperparameter λ. It aids in mitigating the overfitting of a model to the training data by striking a balance between fitting accuracy and model simplicity. These transformations facilitate the mapping of the source data's infrared intensities to the target data. Consequently, the coefficients become more resilient to collinearity. We can therefore predict the infrared intensities for the ZIF-67 by multiplying the β coefficient and then taking the summation of the contribution to the intensities.

2.6 References

1. Debe; M.K. Electrocatalyst approaches and challenges for automotive fuel cells. Nature 2012, 486(7401), pp.43-51.

2. Zhang, B.; Zheng, X.; Voznyy, O.; Comin, R.; Bajdich, M.; García-Melchor, M.; Han, L.; Xu, J.; Liu, M.; Zheng, L.; García de Arquer, F.P. Homogeneously dispersed multifetal oxygen-evolving catalysts. Science 2016, 352(6283), pp.333-337.

3. Roger, I.; Shipman, M. A.; Symes, M. D. Earth-abundant catalysts for electrochemical and photoelectrochemical water splitting. Nature Reviews Chemistry 2017, 1(1), p.0003.

4. Gao, S.; Jiao, X.; Sun, Z.; Zhang, W.; Sun, Y.; Wang, C.; Hu, Q.; Zu, X.; Yang, F.; Yang, S.; Liang, L. Ultrathin Co_3O_4 layers realizing optimized CO_2 electroreduction to formats. Angew Chem International Edition 2016, 55(2), pp.698-702.

5. Zhu, C.; Guo, S.; Dong, S. PdM (M= Pt, Au) bimetallic alloy nanowires with enhanced electrocatalytic activity for electro-oxidation of small molecules. Advanced materials 2012, 24(17), pp.2326-2331.

6. Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe; M.; Yaghi, O. M. Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. Proceedings of the National Academy of Sciences 2006, 103(27), pp.10186-10191.

7. Cheng, N.; Ren, L.; Xu, X.; Du, Y.; Dou, S. X. Recent development of zeolitic imidazolate frameworks (ZIFs) derived porous carbon based materials as electrocatalysts. Advanced Energy Materials 2018, 8(25), p.1801257.

8. Zhong, G.; Liu, D.; Zhang, J. The application of ZIF-67 and its derivatives: adsorption, separation, electrochemistry and catalysts. Journal of Materials Chemistry A 2018, 6(5), pp.1887-1899.

9. Liu, Y.; Zhao, H. Persulfate activation by single-atom catalysts for the removal of organic pollutants: A review. Resources Chemicals and Materials 2022.

10. Myneni, S. C.; Traina, S. J.; Waychunas, G. A.; Logan, T. J. Vibrational spectroscopy of functional group chemistry and arsenate coordination in ettringite. Geochimica et Cosmochimica Acta 1998, 62(21-22), pp.3499-3514.

11. Barth, A. Infrared spectroscopy of proteins. Biochimica et Biophysica Acta (BBA)-Bioenergetics 2007, 1767(9), pp.1073-1101.

12. Jain, R.K., Damoulas, T. and Kontokosta, C.E. Towards data-driven energy consumption forecasting of multi-family residential buildings: feature selection via the LASSO. In Computing in Civil and Building Engineering 2014, pp. 1675-1682.

13. Lee, C. Y.; Lin, J. Y.; Chang, R. I. Improve quality and efficiency of textile process using data-driven machine learning in industry 4.0. International Journal of Technology and Engineering Studies 2018, 4(2), pp.4-10004.

14. Wu, C., Xie, D.; Mei, Y.; Xiu, Z.; Poduska, K. M.; Li, D.; Xu, B.; Sun, D. Unveiling the thermolysis natures of ZIF-8 and ZIF-67 by employing in situ structural

characterization studies. 2019 Physical Chemistry Chemical Physics, 21(32), 17571-17577.

15. Shan, J.; Liao, J.; Ye, C.; Dong, J.; Zheng, Y.; Qiao, S. Z. The Dynamic Formation from Metal-Organic Frameworks of High-Density Platinum Single-Atom Catalysts with Metal-Metal Interactions. Angewandte Chemie 2022, 134(48), e202213412.

16. Yevick, A.; Hannel, M.; Grier, D. G. Machine-learning approach to holographic particle characterization. Optics express 2014, 22(22), 26884-26890.

17. Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K. R.; Tkatchenko, A. Combining machine learning and computational chemistry for predictive insights into chemical systems. Chemical reviews 2021, 121(16), 9816-9872.

18. Zubatiuk, T.; Isayev, O. Development of multimodal machine learning potentials: Toward a physics-aware artificial intelligence. Accounts of Chemical Research 2021, 54(7), 1575-1585.

19. Li, H.; Jiao, Y.; Davey, K.; Qiao, S. Z. Data-Driven Machine Learning for Understanding Surface Structures of Heterogeneous Catalysts. Angewandte Chemie, 135(9) 2023, e202216383.

20. Zubatiuk, T., & Isayev, O. Development of multimodal machine learning potentials: Toward a physics-aware artificial intelligence. Accounts of Chemical Research 2021, 54(7), 1575-1585.

21. Unke, O. T.; Chmiela, S., Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko A.; Müller, K. R. Machine learning force fields. Chemical Reviews 2021, 121(16), 10142-10186.

22. Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable machine learning for knowledge generation in heterogeneous catalysis. Nature Catalysis 2022, 5(3), 175-184.

23. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. 2018 Science, 361(6400), 360-365.

24. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software 2010, 33(1), p.1.

25. Kim, S.J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale ℓ_1-regularized least squares. IEEE journal of selected topics in signal processing 2007, 1(4), pp.606-617.

Chapter 3: Machine Learning Confirms the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis

3.1 Introduction and Significance

In this study, we present a pioneering endeavour wherein an artificial intelligence (AI)-driven analysis of in-situ temperature-dependent Fourier-transform infrared spectroscopy (DRIFTs) data is employed to confirm the pyrolysis mechanism of pristine ZIF and Pt-doped ZIF-67 to synthesize Co_3O_4 and $Pt-Co_3O_4$ SAC. The AI simulation exhibits a remarkable Pearson correlation value as high as 0.7~0.9 compared with experimental data. Furthermore, this algorithm provides correlation coefficients for the selected structural features, thereby facilitating the extraction of invaluable insights into the structural evolution along the temperature and time axis during the reaction process.

Highlights of this work include:

1. Establish DFT-calculated IR spectrum database of all possible chemical bonds during the ZIF pyrolysis process and use machine learning algorithm to bridge the gap between theory and experimental results to directly simulate and analyse in-situ experimental IR spectrum.

2. Develop AI workflow to extract chemical bond information correlated with experimental data from the theoretical database. A successful fit of the experimental in-situ DRIFTs spectrum is achieved with high correlation of 0.7-0.9, indicating effective simulation of experimental results.

3. Through AI-analysed changes in chemical bond coefficients, the inferred reaction mechanism, encompassing ZIF decomposition and Pt-O bond formation, closely aligns with experimental findings, which demonstrates AI's potential in reasonably speculating the reaction mechanism.

4. The approach relies heavily on database construction and AI algorithms, with minimal manual intervention. As such, it holds significant potential for robustness and can be extended to intelligent analysis of experimental data to predict the formation mechanism of more SACs.

3.2 Machine Learning Confirms the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis

This Chapter is included as it appears as a journal paper published by Yanzhang Zhao, Huan Li, Jieqiong Shan, Zhen Zhang, Xinyu Li, Javen Qinfeng Shi, Yan Jiao, and Haobo Li*, "Machine Learning Confirms the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis." The Journal of Physical Chemistry Letters 14 (2023): 11058-11062.

Statement of Authorship

Title of Paper	Machine Learning Confirms the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Yanzhang Zhao, Huan Li, Jieqiong Shan, Zhen Zhang, Xinyu Li, Javen Qinfeng Shi, Yan Jiao, and Haobo Li*," Machine Learning Confirms the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis." The Journal of Physical Chemistry Letters 14 (2023): 11058-11062.

Principal Author

Name of Principal Author (Candidate)	Yanzhang Zhao		
Contribution to the Paper	Research plan, computation calculations, data analysis and manuscript draft.		
Overall percentage (%)	70		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	16/01/2024

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    i.   the candidate's stated contribution to the publication is accurate (as detailed above);

    ii.   permission is granted for the candidate in include the publication in the thesis; and

    iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Huan Li		
Contribution to the Paper	Provision of experimental support, and discussion of the manuscript evaluation.		
Signature		Date	16/01/2024

Name of Co-Author	Jieqiong Shan		
Contribution to the Paper	Support experimental data, and discussion of the manuscript evaluation.		
Signature		Date	16/01/2024

Name of Co-Author	Zhen Zhang			
Contribution to the Paper	Provide help with machine learning theory and techniques, and discussion of the manuscript evaluation.			
Signature			Date	16/01/2024

Name of Co-Author	Xinyu Li			
Contribution to the Paper	Provide help with machine learning theory and techniques, and discussion of the manuscript evaluation.			
Signature			Date	16/01/2024

Name of Co-Author	Javen Qinfeng Shi			
Contribution to the Paper	Provide help with machine learning theory and techniques, and discussion of the manuscript evaluation.			
Signature			Date	16/01/2024

Name of Co-Author	Yan Jiao			
Contribution to the Paper	Discussion of the manuscript evaluation.			
Signature			Date	16/01/2024

Name of Co-Author	Haobo Li			
Contribution to the Paper	Supervised the research project, and discussion of the manuscript evaluation.			
Signature			Date	16/01/2024

Please cut and paste additional co-author panels here as required.

Letter

Machine Learning Confirms the Formation Mechanism of a Single-Atom Catalyst via Infrared Spectroscopic Analysis

Yanzhang Zhao,[†] Huan Li,[†] Jieqiong Shan, Zhen Zhang, Xinyu Li, Javen Qinfeng Shi, Yan Jiao, and Haobo Li*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Single-atom catalysts (SACs) offer significant potential across various applications, yet our understanding of their formation mechanism remains limited. Notably, the pyrolysis of zeolitic imidazolate frameworks (ZIFs) stands as a pivotal avenue for SAC synthesis, of which the mechanism can be assessed through infrared (IR) spectroscopy. However, the prevailing analysis techniques still rely on manual interpretation. Here, we report a machine learning (ML)-driven analysis of the IR spectroscopy to unravel the pyrolysis process of Pt-doped ZIF-67 to synthesize Pt−Co_3O_4 SAC. Demonstrating a total Pearson correlation exceeding 0.7 with experimental data, the algorithm provides correlation coefficients for the selected structures, thereby confirming crucial structural changes with time and temperature, including the decomposition of ZIF and formation of Pt−O bonds. These findings reveal and confirm the formation mechanism of SACs. As demonstrated, the integration of ML algorithms, theoretical simulations, and experimental spectral analysis introduces an approach to deciphering experimental characterization data, implying its potential for broader adoption.
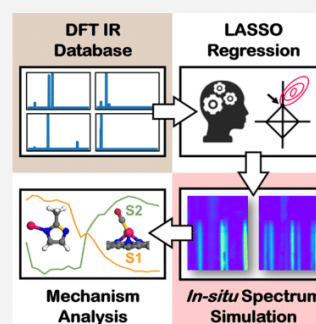
M achine-learning (ML) techniques are increasingly used as effective tools to analyze the experimental spectra, such as the X-ray absorption structure (XAS),[1−3] nuclear magnetic resonance (NMR),[4] infrared (IR),[5,6] ultraviolet (UV),[7] and Raman[8,9] spectroscopies, etc. A primary objective is to identify structures by comparing experimentally obtained spectra to theoretically predicted spectra. These experiments entail a series of multiple tests conducted under diverse conditions, especially *in situ* experiments. ML becomes a viable approach for reverse engineering atomic structures and, even more promisingly, for inferring underlying mechanisms.[10] Nevertheless, it is noteworthy that research in the latter domain, specifically concerning the inference of unclear mechanisms, remains relatively limited.

Single-atom catalysts (SACs) have attracted broad interests, particularly in energy conversion applications.[11,12] The primary focus has predominantly centered on elucidating the origin of their catalytic performance; however, the exploration of their formation mechanisms remains limited. This knowledge gap poses challenges when it comes to achieving precise and controlled syntheses of SACs. Metal−organic framework (MOF) materials are frequently employed as precursors for the synthesis of SACs, which are subsequently obtained through a pyrolysis process.[13,14] Shan et al.[15,16] conducted a comprehensive investigation into the evolution of cobalt-based zeolite imidazolate framework (ZIF-67) during pyrolysis, resulting in the formation of Pt-doped cobalt oxide (Pt−Co_3O_4) SAC. Their study employed *in situ* temperature-dependent diffuse reflectance Fourier transform infrared

spectroscopy (DRIFTS) in conjunction with complementary experiments, including X-ray diffraction (XRD) patterns, extended X-ray absorption fine structure (EXAFS) spectra, and X-ray absorption near-edge structure (XANES) spectra characterization. The collective findings demonstrated the key role of the formation of metal−oxygen−metal bonds in fostering metal−metal interactions among densely populated metal single atoms.

The aforementioned research includes a series of *in situ* DRIFTS, which hold significant potential for ML-driven data mining. Such data sets enable the extraction of dynamic processes associated with structural evolution in materials. A substantial database is required to train the ML model, and density functional theory (DFT) calculations provide a sufficient pool of data for this purpose. Despite disparities observed between DFT-simulated IR spectra and experimental results,[17] prior studies have revealed a well-fitted linear scaling relationship between the calculated C−O bond stretching peak using DFT and the corresponding experimental data.[18] Consequently, the application of ML algorithms, particularly those founded on linear regression, holds promise in bridging
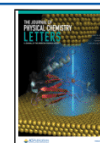
the gap between theoretical and experimental results, finally yielding superior fitting outcomes.

Herein, we have, for the first time, applied the least absolute shrinkage and selection operator (LASSO),[19−23] a regression algorithm, to analyze the *in situ* temperature-dependent DRIFTS data set. Our aim was to gain deeper insights into the underlying reaction mechanism. In comparison to other ML algorithms, LASSO offers distinct advantages in terms of interpretability and reliability, rendering it particularly well-suited for this application. Through the analysis of pristine ZIF-67 and Pt-doped ZIF-67, a comparative study is conducted to discern the distinctions in the pyrolysis mechanisms. Our investigation involved the examination of the pyrolysis of ZIF-67, resulting in the synthesis of Co_3O_4, and the pyrolysis of Pt-doped ZIF-67 with the objective of producing Pt−Co_3O_4 SACs. This comparative analysis allows for inference of the underlying formation mechanisms of SACs. Such integration of ML algorithms with theoretical calculations for analyzing *in situ* experimental spectra offers the potential to supplement human expert interpretation, thereby evolving into a more widely adopted and versatile methodology.

The workflow is shown in Figure 1. First, we constructed a data set encompassing all conceivable chemical bonds formed
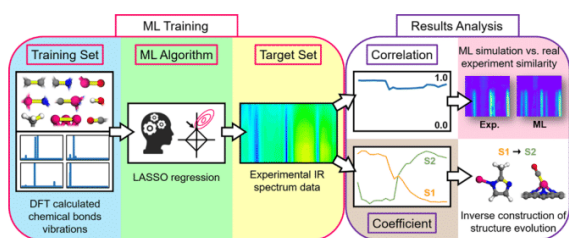
Figure 1. Workflow for data-driven IR spectroscopic analysis to extract the structural evolution mechanism using ML algorithms. The experimental IR spectrum is reproduced with permission from ref 15. Copyright 2022 Wiley-VCH.

during the pyrolysis of ZIF-67. This process involves elements, such as Co, C, N, O, and H, in pristine ZIF-67 and the addition of Pt in Pt-doped ZIF-67. Our data set covers a wide range of chemical bonds, including C−C, C−N, C−O, Co(Pt)−O, Co(Pt)−N, Co(Pt)−C, Co(Pt)−Co, C−H, and O−H bonds. These chemical bonds may exist in diverse structural environments, prompting us to build distinct models for them encompassing both three-dimensional (3D) ZIF-based and two-dimensional (2D) graphene-based models. Computational models for pristine ZIF-67 and additional models for Pt-doped ZIF-67 are shown in Figure 2. Our
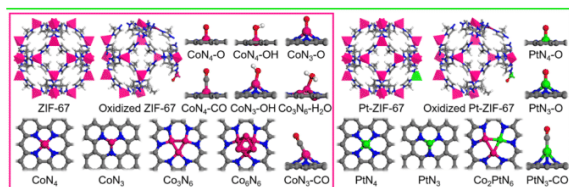

Figure 2. Computational models of the IR spectrum database for chemical bonds during ZIF-67 (in the pink frame) and Pt-doped ZIF-67 (in the green frame) pyrolysis. Atoms: Co, pink; Pt, green; C, gray; N, blue; O, red; and H, white.

database comprises all non-equivalent chemical bonds within these models, including 55 chemical bonds for ZIF-67 and 86 chemical bonds for Pt-doped ZIF-67, as detailed in Tables S1 and S2 of the Supporting Information, respectively. We conducted DFT calculations to simulate the IR spectra for each of chemical bond. The data for each chemical bond includes the 6638 wavenumber−absorbance pairs, which have been represented in 2D plots displayed in Figures S1−S32 of the Supporting Information. In total, the database encompasses 935 958 data points.

In the second step, we proceeded to train the ML model by fitting the LASSO model using DFT data as input and experimental data as the target output. The IR spectra for each chemical bond are regarded as ML features. The experimental target data are sourced directly from ref 15, encompassing the pyrolysis process of both ZIF-67 and Pt-doped ZIF-67. This process involves a gradual temperature increase from 20 to 300 °C, followed by a 90 min hold at 300 °C. The *in situ* DRIFTS spectra are thereby measured on the temperature−time axis, comprising 23 and 24 spectral lines for ZIF-67 and Pt-doped ZIF-67, respectively, totaling 47 lines. This experimental data set contains 41 830 data points, which are visualized in the form of 2D mapping.

In the third step, we further analyzed the results of ML training, where the LASSO algorithm provides two key parameters: correlation and coefficient. The detailed calculation methods for these parameters are specified in the Computational Methods of the Supporting Information. Correlation, which is referred to as the Pearson correlation in this study, ranging from 0.0 to 1.0, denotes the degree of relationship between the ML prediction set and the experimental data. Higher correlation values signify more accurate replication of experimental results by ML. Coefficient, particularly in our work, representing the significance of various chemical bonds, allows for the quantification of the influence of individual bonds along the temperature−time axis. Coefficients for all chemical bonds are summarized in Figures S1−S32 of the Supporting Information. Notably, some key bond coefficients decrease or increase across the temperature−time axis. This phenomenon suggests that some chemical structures associated with specific bonds diminish during the reaction, while others significantly increase. This implies the transformation of certain structural entities into alternative forms. This method facilitates the reverse construction[24] of the chemical structures based on spectral information.

Figure 3 shows the comparison between the ML-simulated *in situ* temperature-dependent DRIFTS spectra within the developed workflow and the corresponding experimental results. Remarkably, the ML-predicted spectrum closely resembles the experimental spectrum, a similarity quantitatively substantiated by correlation values in panels c and f of Figure 3, approximately 0.7 and 0.9 for ZIF-67 and Pt-doped ZIF-67, respectively. According to the literature reports,[25,26] Pearson coefficients of 0.60 and above indicate a strong correlation or association. *p* values significantly lower than 0.05 indicate the reliability of the linear correlation observed in ML (Figure S34 of the Supporting Information). The correlation values thus affirm the success of our ML predictions for the *in situ* DRIFTS spectra. Furthermore, as depicted in Figure 4, the majority of distinct features employed by our ML model demonstrate low correlations. This underscores the rationality of using these calculated IR spectra as features in our ML approach.
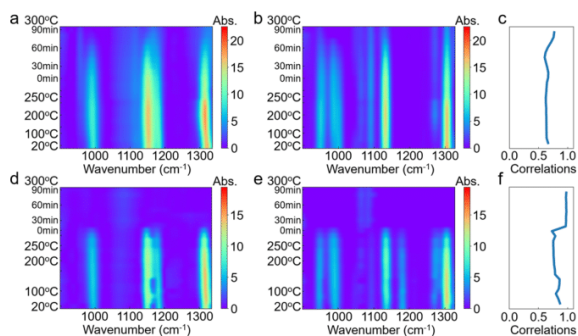
Figure 3. ML-simulated *in situ* temperature-dependent DRIFTS spectra in comparison to experimental characterization outcomes for the pyrolysis processes of (a−c) ZIF-67 and (d−f) Pt-doped ZIF-67, respectively. (a and d) Experimental results exhibited a striking resemblance to the (b and e) ML simulations, with (c and f) correlations reaching approximately 0.7 and 0.9, respectively.
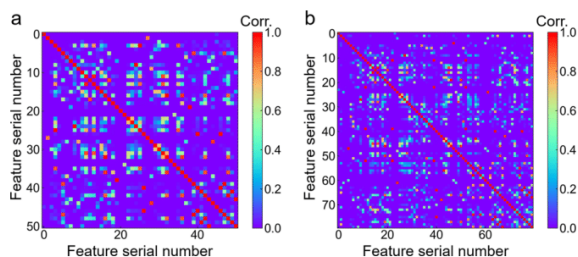

Figure 4. Pearson correlation between every pair of primary features, i.e., calculated IR spectrum for each chemical bond in the (a) ZIF-67 and (b) Pt-doped ZIF-67 pyrolysis process.

It was observed that both experimental and LASSO-predicted spectra presented similar IR peaks that appeared around 1000, 1150, and 1300 cm^{-1}. Additionally, the absorbance intensity is stronger at lower temperatures, and the sudden intensity decrease when the time surpassed 60 min held at 300 °C was associated with sample decomposition of ZIF-67 as a result of long-time thermolysis. Noticeably, Pt-doped ZIF-67 exhibited relatively less thermal stability in Figure 3d, and the sample was completely decomposed when the temperature rose to 300 °C. The positions of Fourier transform infrared spectroscopy (FTIR) peaks are affected by the energies of vibration modes within the solids, and the energy can be noticeably changed by alterations in atomic arrangements. The peak position and absorbance intensity provide a clear indication that the overall IR intensity absorbance of the three peaks at 1000, 1150, and 1300 cm^{-1} in Pt-doped ZIF-67 declined during the heating process. While the predicted IR spectra for both ZIF-67 and Pt-doped ZIF-67 exhibit significant similarity to their corresponding experimental spectra, it is worth noting that the LASSO regression analysis presents a higher overall correlation for Pt-doped ZIF-67 compared to pristine ZIF-67. Moreover, it exhibited a higher correlation value in the lower temperature zone.

Given the successful outcomes of our ML predictions, we then investigated the extraction of chemical structure information from the provided coefficient data. The coefficients associated with the 3D models based on ZIF exhibit a consistent decrease throughout the pyrolysis reaction process, as evident in Figures S1, S2, S14, and S15 of the

Supporting Information. This decline corresponds to the decomposition of the ZIF structure, wherein the Co−N bonds manifest high coefficients in comparison to other chemical bonds involved in the pyrolysis process. This observation underscores the prevalence of Co−N bonds as the predominant chemical bond. Concurrently, some bonds among intermediate products gradually emerge, including metal−C bonds, metal−OH bonds, and metal−O bonds. Interestingly, even though our training set encompasses intermetallic cluster structures, such as Co$_3$, Co$_2$Pt, and Co$_6$, their coefficients overwhelmingly remain at 0, as depicted in Figures S11−S13 and S29−S32 of the Supporting Information. This indicates that the ML rarely selects these specific structures, which aligns with experimental observations where metals tend not to agglomerate but form oxides or single-atom-doped oxides.[15]

To facilitate a more insightful analysis of the dynamic evolution process during pyrolysis, we selected chemical bonds that exhibit the most pronounced changing trends and compile them in Figure 5. For clarity of observation, all coefficient
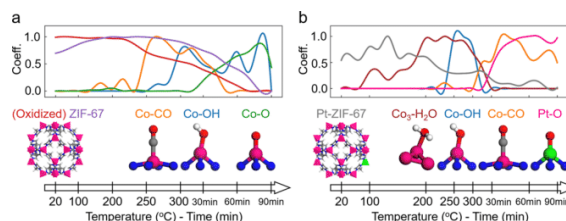

Figure 5. Coefficients of selected chemical bonds along the temperature−time axis for the pyrolysis processes of (a) ZIF-67 and (b) Pt-doped ZIF-67. The model structures corresponding to the different colors of the chemical bonds are labeled below, and they are positioned along the temperature−time axis based on the approximate peak positions of the coefficients. It illustrates the evolution process of the structures. Atoms: Co, pink; Pt, green; C, gray; N, blue; and O, red.

values have been normalized. In the case of pristine ZIF-67 (Figure 5a), the initial structures, ZIF-67 (purple line) and oxidized ZIF-67 (red line), prominently dominate and persist until the temperature reaches 300 °C. At this point, they begin to rapidly decline, implying that the structural integrity of ZIF-67 remains largely intact until the pyrolysis temperature reaches 300 °C. Around the temperature range of 250−300 °C, notable intermediates emerge, including the Co−CO (orange line) and Co−OH (blue line) structures. Subsequently, as the system is held at 300 °C over time, these intermediates undergo oxidation, transforming into a Co−O (green line) structure with a higher oxidation state.

We compared this to the case of Pt-doped ZIF-67 (Figure 5b). Pt-ZIF-67 (gray line) predominantly characterizes the initial structure of the reaction. Subsequently, the Co$_3$ (dark red line) cluster becomes noticeable, but it diminishs above 300 °C. Furthermore, the Co−CO (orange line) and Co−OH (blue line) structures identified in pristine ZIF-67 appear as intermediates. More significantly, a prominent Pt−O (pink line) structure becomes evident above 300 °C. This aligns with experimental observations where the formation of Pt−O bonds play a key role as an intermediate step in generating SACs.[15] It suggests that Pt exhibits a tendency to avoid agglomeration and, instead, undergoes a process involving the formation of

30

Pt−O and Co−O bonds, ultimately incorporating into the crystal lattice of Co_3O_4.

On the basis of ML-assisted IR spectral analysis, we have gained comprehensive insights into the intricate processes involved in the formation of the SAC, Pt−Co_3O_4, from ZIF pyrolysis. The entire process was found to be primarily an oxidation process, wherein Co and Pt atoms initially form connections with the N atoms of the four methylimidazole ligands within the ZIF framework. As the ZIF skeleton disintegrates, a planar CoN_4 or PtN_4 structure is formed, albeit in a relatively small proportion. More prevalent are structures with oxygen-containing intermediates, such as CO or O, adsorbed onto them. When held at a high temperature of 300 °C over time, the four N atoms gradually detach, coinciding with the formation of Co−O bonds (green line in Figure 5a). In comparison, the coefficient of Pt−O bonds increases obviously faster (pink line in Figure 5b) than that of Co−O bonds, indicating that Pt−O bonds are formed earlier than Co−O bonds. This is consistent with experimental observations that the transformation from Co−N to Co−O starts at about 300 °C after Pt−N begins to gradually change into Pt−O. These Pt−O and Co−O bonds then connect with each other, culminating in the formation of Co_3O_4 or Pt-containing high-density Pt−Co_3O_4 structures.

We acknowledge certain limitations in the model, such as the fact that the DRIFTS data used for training only focus on the higher wavenumber organic ligand region and the absence of training data on oxide structures below 900 cm^{-1} causes the final step of oxide formation to be missed in the ML analysis. It is still noteworthy that the ML-driven analysis closely aligns with manual analysis results, especially concerning the mechanistic aspects up to Co−O and Pt−O bond formation.[15] This underscores the potential of ML, following training with substantial theoretical simulation data, to provide highly accurate insights into *in situ* experimental data analysis.

In summary, this work involves the utilization of a ML tool based on the LASSO algorithm for the comprehensive analysis of *in situ* temperature-dependent DRIFTS in experimental setting. The comprehensive modeling of chemical bond information and the establishment of a theoretical database have enabled the trained ML model to effectively emulate real experimental data. This emulation is reflected in correlation values as high as 0.7−0.9, confirming the effectiveness of ML to simulate experimental data. Simultaneously, the correlations provided by ML for different chemical bonds during the experimental data fitting process can be used to infer chemical structures. In combination with variations along the temperature−time axis, this information offers deep insights into the chemical reaction mechanisms. The observed degradation of the ZIF framework, the gradual oxidation of oxygen-containing species adsorbed on the metal, and the emergence of Co−O and Pt−O bonds align remarkably well with findings obtained from various experimental characterizations. It is worth highlighting that the entire workflow that we have developed relies heavily on database construction and ML algorithms, with minimal manual intervention. As such, we believe that this approach holds significant potential for robustness and can be extended to a broad range of applications for intelligent analysis of *in situ* experimental characterization data in the future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c02896.

Computational details, IR spectrum and corresponding LASSO coefficients in the database, tests for different regression algorithms, LASSO *p* values, summary of chemical bonds and temperature and time coordinates, and tests for different force convergence and Hubbard correction settings (PDF)

Transparent Peer Review report available (PDF)

AUTHOR INFORMATION

Corresponding Author

Haobo Li − *School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0002-9448-6771; Email: haobo.li@adelaide.edu.au

Authors

Yanzhang Zhao − *School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Huan Li − *School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Jieqiong Shan − *School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia 5005, Australia; Department of Chemistry, City University of Hong Kong, Kowloon 999077, Hong Kong Special Administrative Region of the People's Republic of China*

Zhen Zhang − *Australian Institute for Machine Learning, The University of Adelaide, Adelaide, South Australia 5000, Australia*

Xinyu Li − *Australian Institute for Machine Learning, The University of Adelaide, Adelaide, South Australia 5000, Australia;* orcid.org/0000-0003-1332-9203

Javen Qinfeng Shi − *Australian Institute for Machine Learning, The University of Adelaide, Adelaide, South Australia 5000, Australia*

Yan Jiao − *School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0003-1329-4290

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c02896

Author Contributions

†Yanzhang Zhao and Huan Li contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

■ REFERENCES

(1) Martini, A.; Hursán, D.; Timoshenko, J.; Rüscher, M.; Haase, F.; Rettenmaier, C.; Ortega, E.; Etxebarria, A.; Roldan Cuenya, B. Tracking the evolution of single-atom catalysts for the CO_2 electrocatalytic reduction using operando X-ray absorption spectroscopy and machine learning. *J. Am. Chem. Soc.* **2023**, *145*, 17351−17366.

(2) Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A. I. Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8*, 5091−5098.

(3) Kotobi, A.; Singh, K.; Höche, D.; Bari, S.; Meißner, R. H.; Bande, A. Integrating explainability into graph neural network models for the prediction of X-ray absorption spectra. *J. Am. Chem. Soc.* **2023**, *145*, 22584−22598.

(4) Guo, S.; Jiang, J.; Ren, H.; Wang, S. Fusion of multiple spectra for investigating chemical bonding properties *via* machine learning. *J. Phys. Chem. Lett.* **2023**, *14*, 7461−7468.

(5) Lansford, J. L.; Vlachos, D. G. Infrared spectroscopy data-and physics-driven machine learning for characterizing surface microstructure of complex materials. *Nat. Commun.* **2020**, *11*, 1513.

(6) Lansford, J. L.; Vlachos, D. G. Spectroscopic probe molecule selection using quantum theory, first-principles calculations, and machine learning. *ACS Nano* **2020**, *14*, 17295−17307.

(7) Ren, H.; Zhang, Q.; Wang, Z.; Zhang, G.; Liu, H.; Guo, W.; Mukamel, S.; Jiang, J. Machine learning recognition of protein secondary structuresbased on two-dimensional spectroscopic descriptors. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2202713119.

(8) Wang, X.; Jiang, S.; Hu, W.; Ye, S.; Wang, T.; Wu, F.; Yang, L.; Li, X.; Zhang, G.; Chen, X.; Jiang, J.; Luo, Y. Quantitatively determining surface−adsorbate properties from vibrational spectroscopy with interpretable machine learning. *J. Am. Chem. Soc.* **2022**, *144*, 16069−16076.

(9) Wang, S.; Jiang, J. Interpretable catalysis models using machine learning with spectroscopic descriptors. *ACS Catal.* **2023**, *13*, 7428−7436.

(10) Li, H.; Jiao, Y.; Davey, K.; Qiao, S.-Z. Data-driven machine learning for understanding surface structures of heterogeneous catalysts. *Angew. Chem. Int. Ed.* **2023**, *62*, No. e202216383.

(11) Wang, A.; Li, J.; Zhang, T. Heterogeneous single-atom catalysis. *Nat. Rev. Chem.* **2018**, *2*, 65−81.

(12) Li, H.; Liu, Y.; Chen, K.; Margraf, J. T.; Li, Y.; Reuter, K. Subgroup discovery points to the prominent role of charge transfer in breaking nitrogen scaling relations at single-atom catalysts on VS_2. *ACS Catal.* **2021**, *11*, 7906−7914.

(13) Shan, J.; Ye, C.; Jiang, Y.; Jaroniec, M.; Zheng, Y.; Qiao, S.-Z. Metal−metal interactions in correlated single-atom catalysts. *Sci. Adv.* **2022**, *8*, No. eabo0762.

(14) Cheng, N.; Ren, L.; Xu, X.; Du, Y.; Dou, S. X. Recent development of zeolitic imidazolate frameworks (ZIFs) derived porous carbon based materials as electrocatalysts. *Adv. Energy Mater.* **2018**, *8*, 1801257.

(15) Shan, J.; Liao, J.; Ye, C.; Dong, J.; Zheng, Y.; Qiao, S.-Z. The dynamic formation from metal−organic frameworks of high-density platinum single-atom catalysts with metal−metal interactions. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202213412.

(16) Shan, J.; Ye, C.; Zhu, C.; Dong, J.; Xu, W.; Chen, L.; Jiao, Y.; Jiang, Y.; Song, L.; Zhang, Y.; Jaroniec, M.; Zhu, Y.; Zheng, Y.; Qiao, S.-Z. Integrating interactive noble metal single-atom catalysts into transition metal oxide lattices. *J. Am. Chem. Soc.* **2022**, *144*, 23214−23222.

(17) Porezag, D.; Pederson, M. R. Infrared intensities and Raman-scattering activities within density-functional theory. *Phys. Rev. B* **1996**, *54*, 7830.

(18) Mairegger, T.; Li, H.; Grießer, C.; Winkler, D.; Filser, J.; Hörmann, N. G.; Reuter, K.; Kunze-Liebhäuser, J. Electroreduction of CO_2 in a non-aqueous electrolyte-the generic role of acetonitrile. *ACS Catal.* **2023**, *13*, 5780−5786.

(19) Santosa, F.; Symes, W. W. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Comput.* **1986**, *7*, 1307−1330.

(20) Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267−288.

(21) Candes, E.; Tao, T. The Dantzig selector: Statistical estimation when *p* is much larger than *n*. *Ann. Statist.* **2007**, *35*, 2313−2351.

(22) Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Software* **2010**, *33*, 1−22.

(23) Kim, S. J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale l_1-regularized least squares. *J. Mach. Learn. Res.* **2007**, *8*, 1519−1555.

(24) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360−365.

(25) Evans, J. D. *Straightforward Statistics for the Behavioral Sciences*; Brooks/Cole Publishing: Pacific Grove, CA, 1996.

(26) Akoglu, H. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **2018**, *18*, 91−93.

32

Supplementary Information for

Machine-Learning Confirms Formation Mechanism of Single-Atom Catalyst *via* Infrared Spectroscopic Analysis

Yanzhang Zhao[1,†]*, Huan Li*[1,†]*, Jieqiong Shan*[1,3]*, Zhen Zhang*[2]*, Xinyu Li*[2]*, Javen Qinfeng Shi*[2]*, Yan Jiao*[1]*, Haobo Li*[*,1]

[1] School of Chemical Engineering, The University of Adelaide, Adelaide, SA 5005, Australia

[2] Australian Institute for Machine Learning, The University of Adelaide, Adelaide, SA 5000, Australia

[3] Department of Chemistry, City University of Hong Kong, Kowloon, Hong Kong 999077, P. R. China

[†] These authors contributed equally: Yanzhang Zhao, Huan Li

[*] Correspondence to Email: haobo.li@adelaide.edu.au

Computational methods

DFT calculations

All Density functional theory (DFT) calculations were conducted by using the Vienna ab-initio simulation package (VASP)[1] with the Perdew-Burke-Ernzerhop (PBE)[2] exchange-correlation and all calculations used the projector-augmented wave (PAW)[3] potential to describe the ionic cores. Each 3D ZIF structure was calculated by setting a 600 eV plane-wave cutoff energy and a Monkhorst-Pack[4] k-point grid of 3×3×3. The lattice constants of the unit cell for those 3D structures including all of the possible intermates are $a = b = c =$ 16.91 Å, $\alpha = \beta = \gamma = 90°$. As for the 2D intermediate structures, a hexagonal unit cell was built with the lattice constants determined to $a = b = 12.3$ Å, vacuum layer 20 Å. A Monkhorst-Pack[4] k-point grid of 3×3×1 was employed. The DFT-D3[5] was utilized to address the Van der Waals interactions, the convergence of the energy was set to be $1×10^{-5}$ eV and that of geometry optimization was set to be maximum force <= 0.05 eV/A. Spin polarization was considered for calculations involved with Co, with initial magnetic moments of 2 μ_B. Different settings including force convergence and Hubbard correction (DFT+U) were tested for various chemical bonds. The tests show that these settings have only minor impact on the IR spectra calculations in this study (**Tables S4, S5**).

Vibrations for IR spectrum

The Infrared intensities class, as implemented in the Atomic Simulation Environment (ASE)[6], was applied in this study. The infrared modes[7] were utilized by a finite difference approximation of the Dynamical matrix by setting the magnitude of displacements with 0.01 Å and the width of 20 cm^{-1} with the Gaussian process regression to plot infrared spectra.

LASSO regression

LASSO[8] was used as a linear model algorithm to estimate the sparse coefficients and its objective function was shown:

$$\min_{w} \frac{1}{2n_{samples}} \| Xw - y \|_2^2 + \alpha \| w \|_1$$

The input array X was preprocessed extracting IR intensities features from the different chemical bonds, the input array y here is the experimental value from IR spectra. $n_{samples}$ represents the number of samples for training. A grid search was performed using a base-10 logarithm grid for α (ranging from 1e-9 to 10). It was observed that a low α value yielded the highest Pearson correlation values. Therefore, a universal α of 1e-9 is used in this study. The data is set to be centered; therefore, no intercept will be used in this calculation, and forced all coefficients to be positive. The Elastic Net was set to default where the norm of the coefficient vector is set to be 1.

As a comparison with LASSO, ridge regression and ordinary least squares (OLS) regression algorithms were tested with the same hyperparameters. As shown in **Figure S33**, LASSO is a more suitable choice compared to ridge regression due to the wide numerical ranges of the various features and the small, sparse coefficients present in the data of this study. Additionally, LASSO is more effective in preventing overfitting compared to OLS regression, making it a better fit for the requirements of this study.

Model validation

In statistics, the Pearson correlation can produce a value that measures linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations. thus, it is essentially a normalized measurement of the covariance. The Pearson correlation function was shown as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where n is the sample size, x_i and y_i are each sample point index, r_{xy} is the value correlation for the x, y value. In here, we will use this Pearson correlation value to reflect the feature that was presented in **Figure 4** in the main text. Additionally, this parameter was applied to identify the consistency of the experimental and LASSO predicted IR spectra.

S3

35
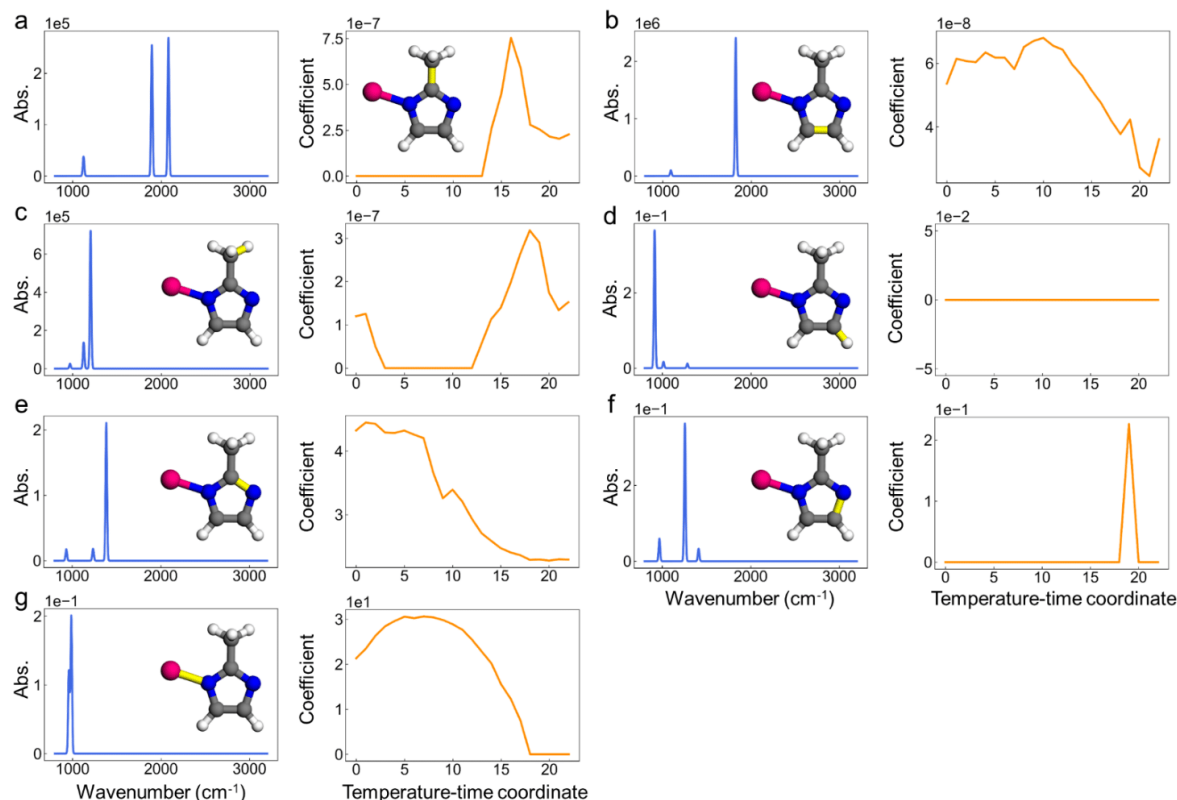
Figure S1. Calculated IR spectrum and corresponding LASSO coefficients for ZIF-67 in ZIF-67 pyrolysis. (a) C-C bond on the methyl; (b) C-C bond in the 5-membered ring; (c) C-H bond in the methyl; (d) C-H bond on the 5-membered ring; (e) C-N bond in the 5-membered ring; (f) C-N bond in the 5-membered ring; (g) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
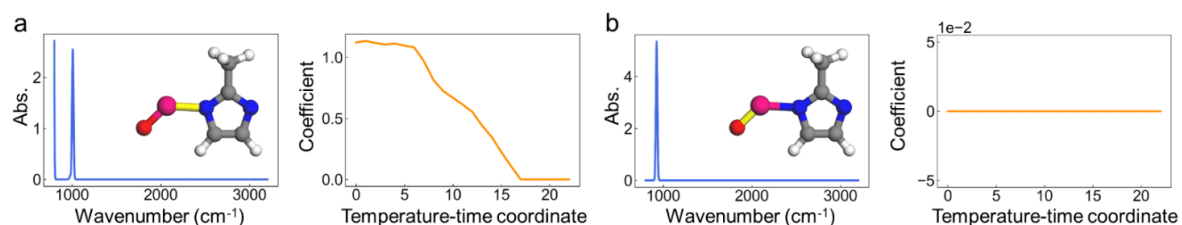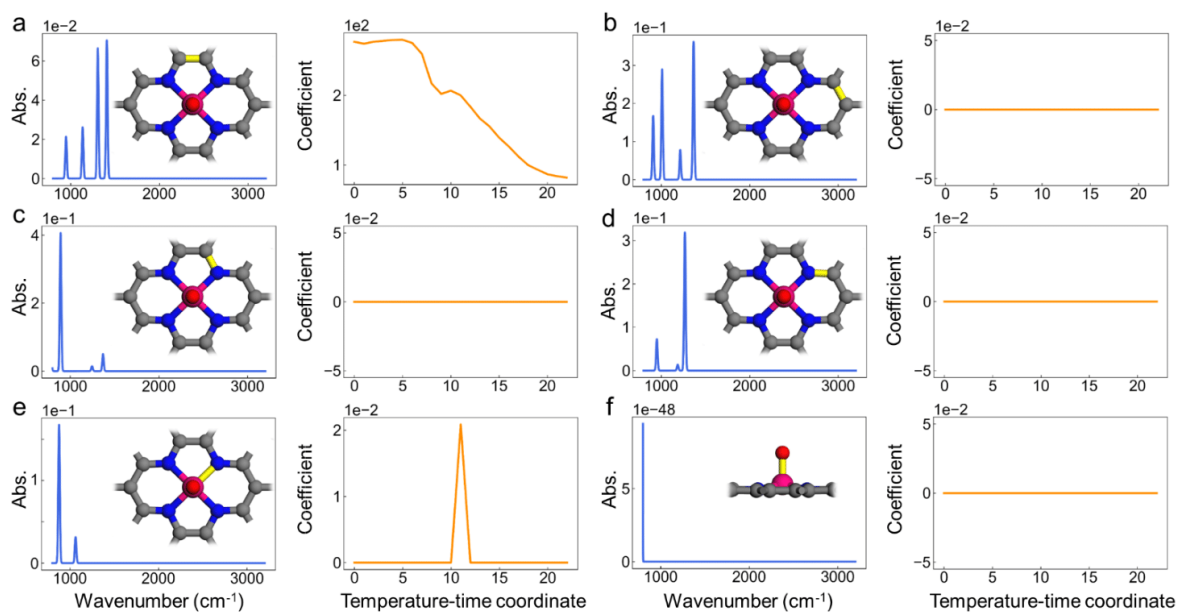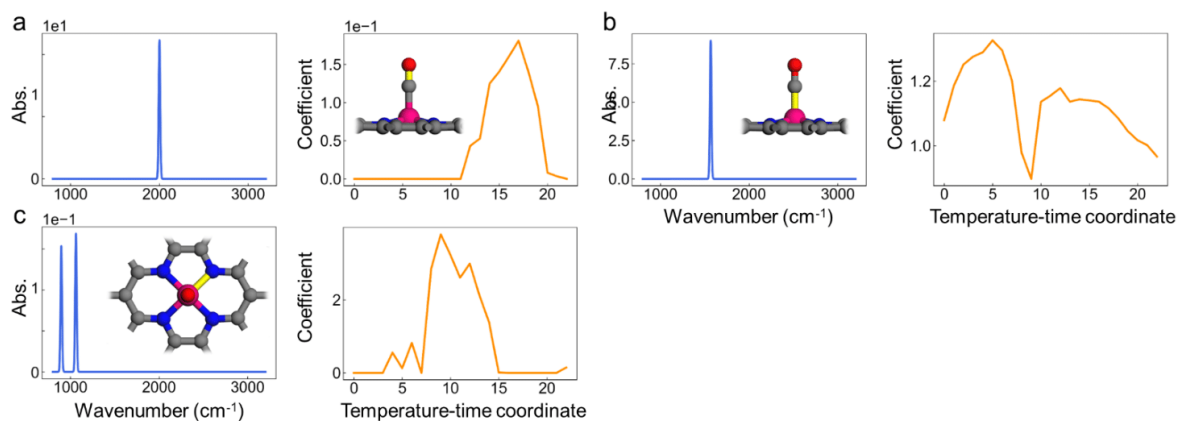

Figure S2. Calculated IR spectrum and corresponding LASSO coefficients for oxidized ZIF-67 in ZIF-67 pyrolysis. (a) Co-N bond; (b) Co-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
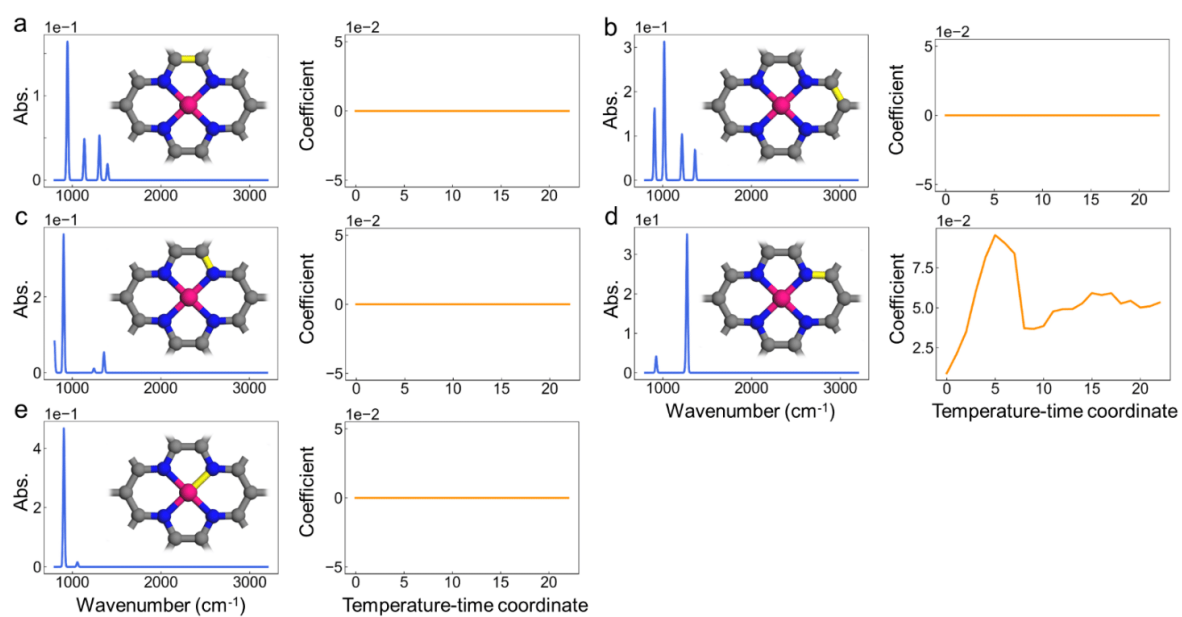
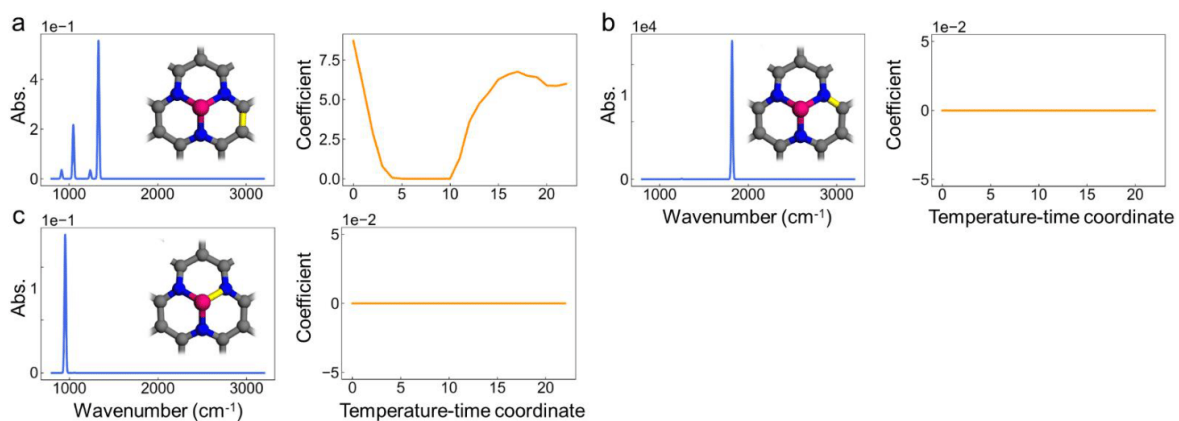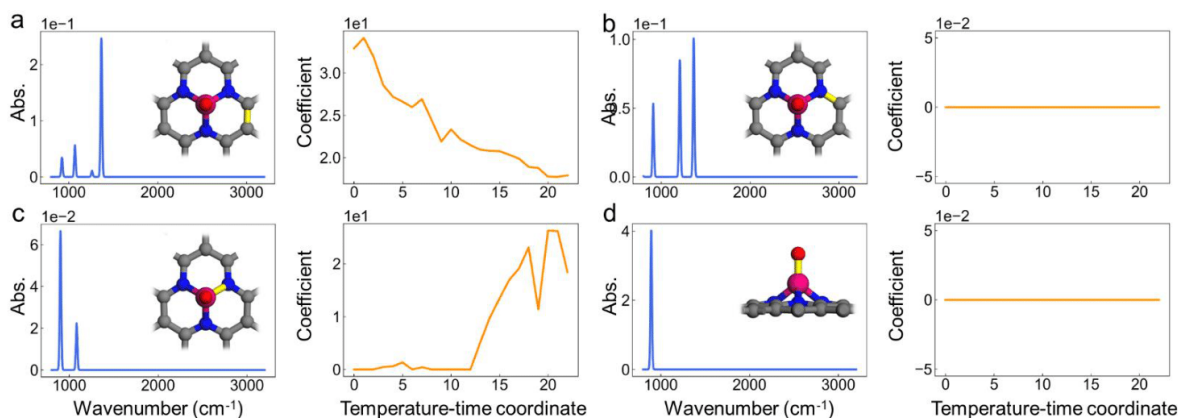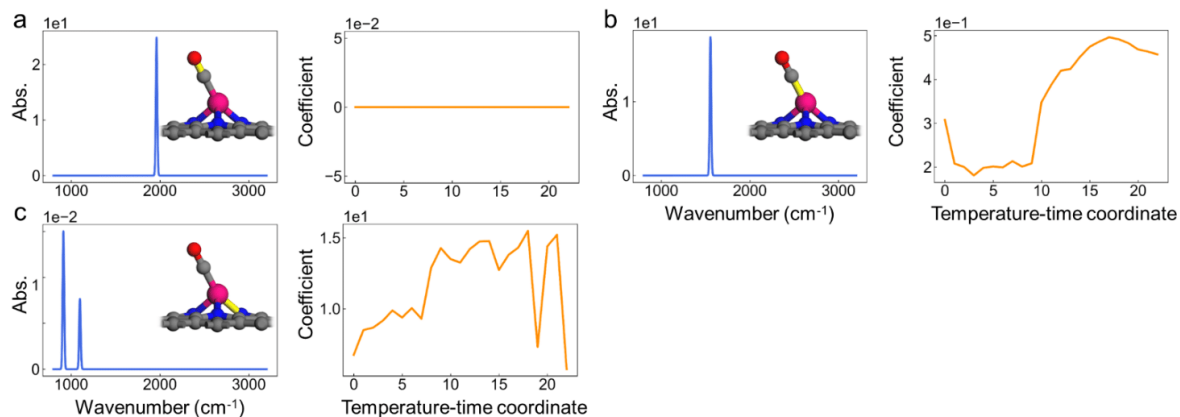**Figure S3.** Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$ in ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S4. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$-O in ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-N bond; (f) Co-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S5. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$-CO in ZIF-67 pyrolysis. (a) C-O bond; (b) Co-C bond; (c) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
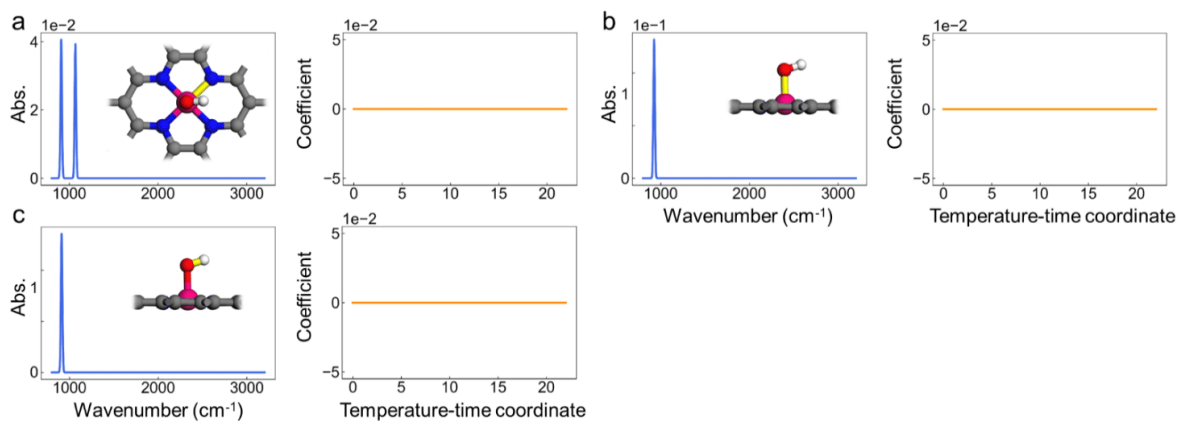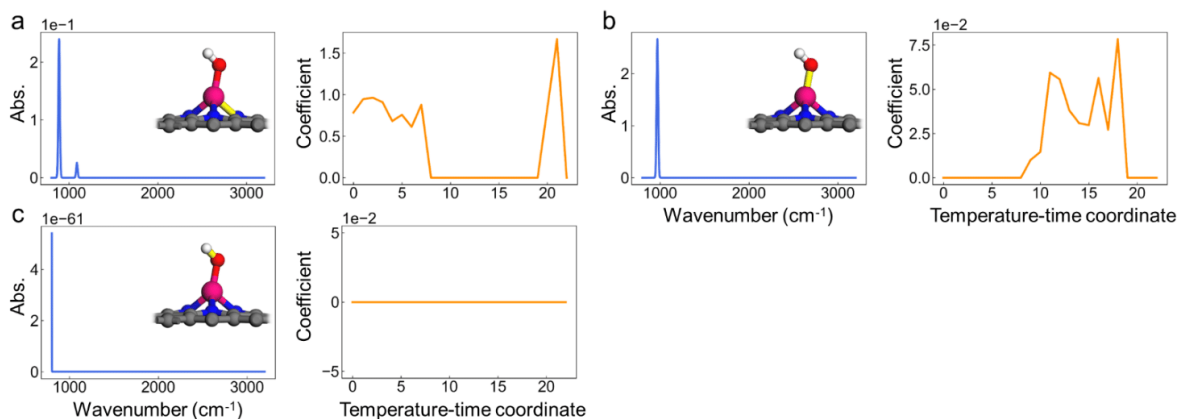
Figure S6. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$-OH in ZIF-67 pyrolysis. (a) Co-N bond; (b) Co-O bond; (c) O-H bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
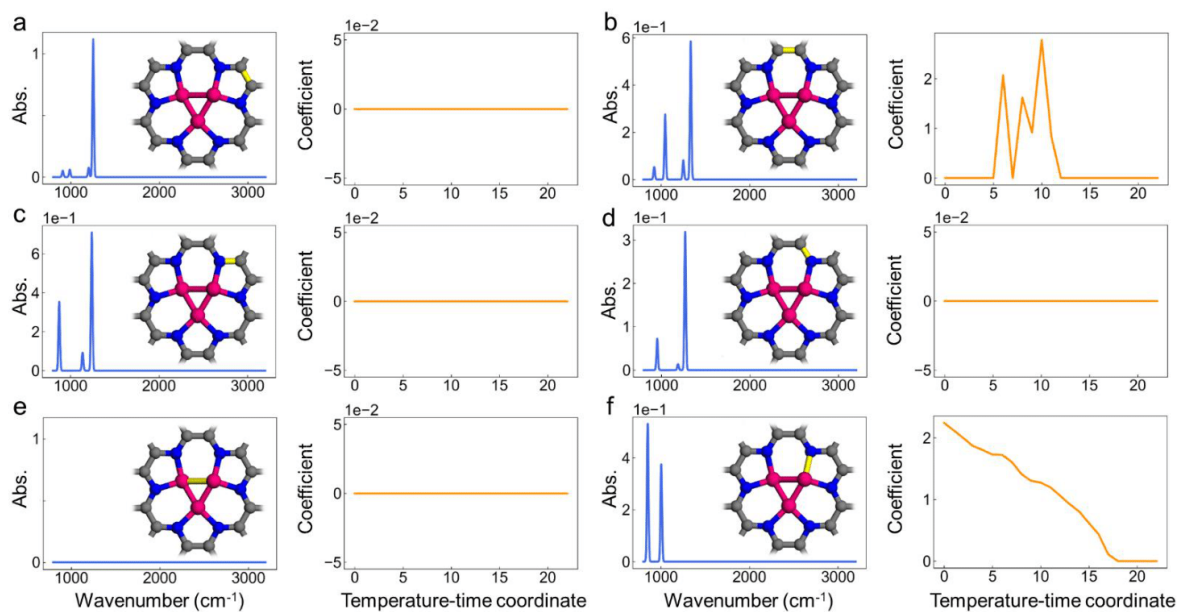

Figure S7. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$ in ZIF-67 pyrolysis. (a) C-C bond; (b) C-N bond; (c) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S8. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$-O in ZIF-67 pyrolysis. (a) C-C bond; (b) C-N bond; (c) Co-N bond; (d) Co-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S9. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$-CO in ZIF-67 pyrolysis. (a) C-O bond; (b) Co-C bond; (c) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S10. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$-OH in ZIF-67 pyrolysis. (a) Co-N bond; (b) Co-O bond; (c) O-H bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S11. Calculated IR spectrum and corresponding LASSO coefficients for Co_3N_6 in ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-Co bond; (f) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
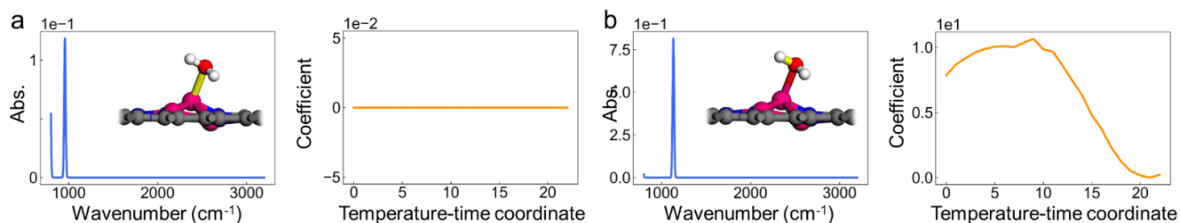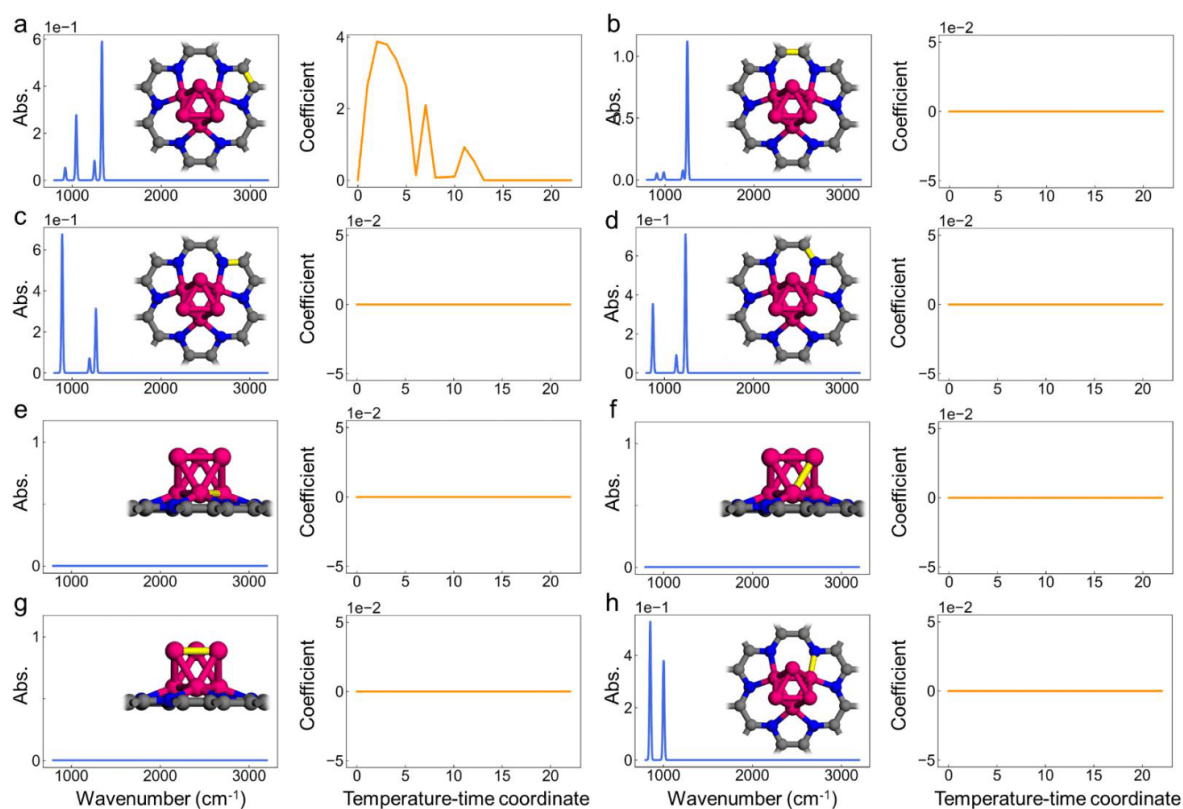

Figure S12. Calculated IR spectrum and corresponding LASSO coefficients for Co_3N_6-H_2O in ZIF-67 pyrolysis. (a) Co-O bond; (b) O-H bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S13. Calculated IR spectrum and corresponding LASSO coefficients for Co_6N_6 in ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-Co bond in the bottom layer; (f) Co-Co bond between the bottom and top layer; (g) Co-Co bond in the top layer; (h) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
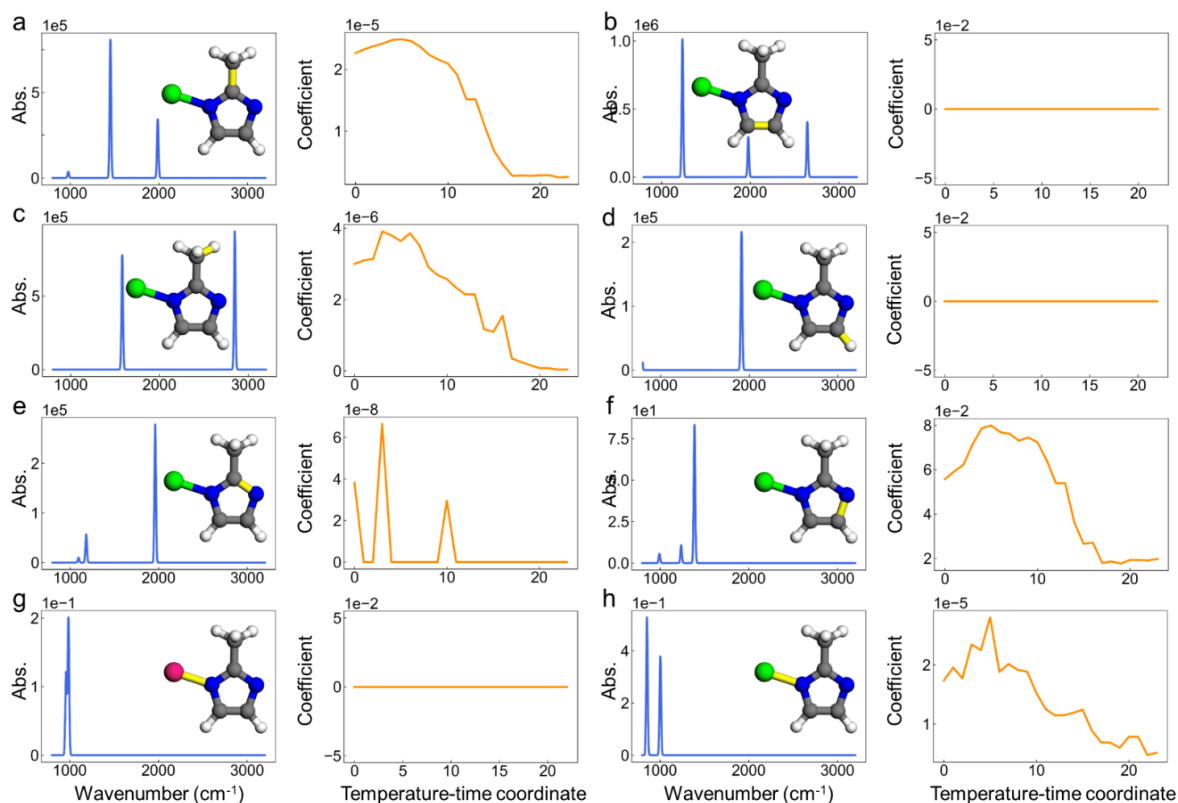
Figure S14. Calculated IR spectrum and corresponding LASSO coefficients for Pt-ZIF-67 in Pt-doped ZIF-67 pyrolysis. (a) C-C bond on the methyl; (b) C-C bond in the 5-membered ring; (c) C-H bond in the methyl; (d) C-H bond on the 5-membered ring; (e) C-N bond in the 5-membered ring; (f) C-N bond in the 5-membered ring; (g) Co-N bond; (f) Pt-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
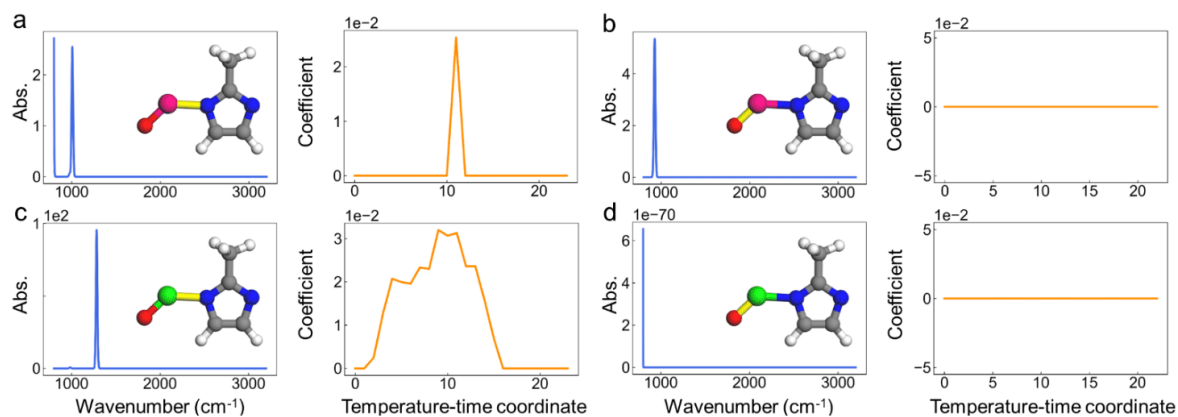

Figure S15. Calculated IR spectrum and corresponding LASSO coefficients for oxidized Pt-ZIF-67 in Pt-doped ZIF-67 pyrolysis. (a) Co-N bond; (b) Co-O bond; (c) Pt-N bond; (d) Pt-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
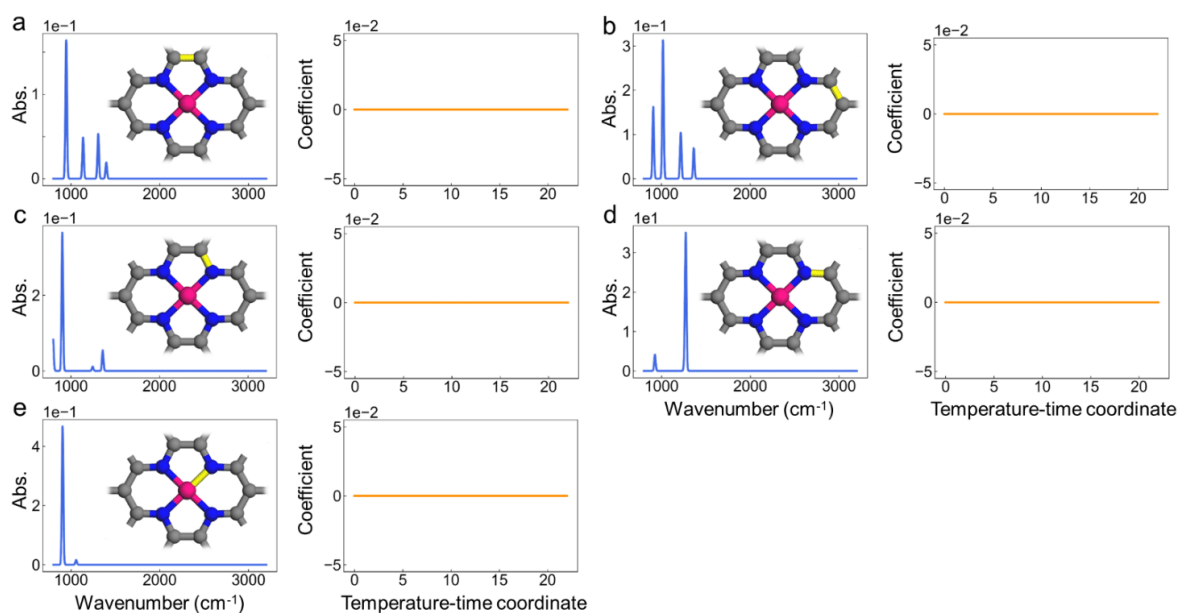
Figure S16. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$ in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
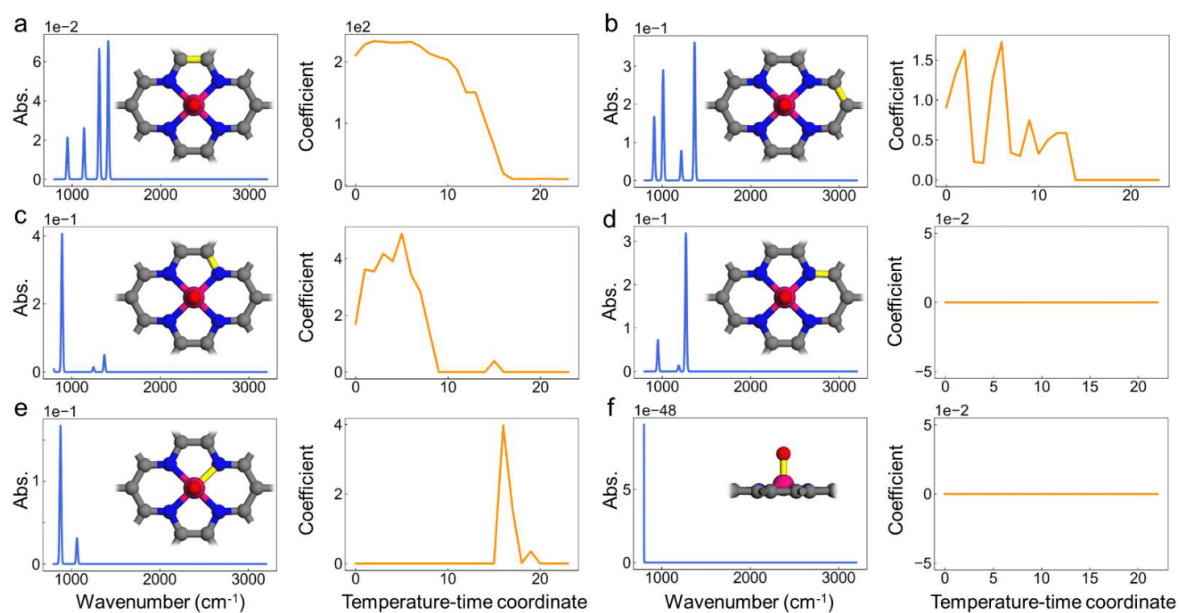

Figure S17. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$-O in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-N bond; (f) Co-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
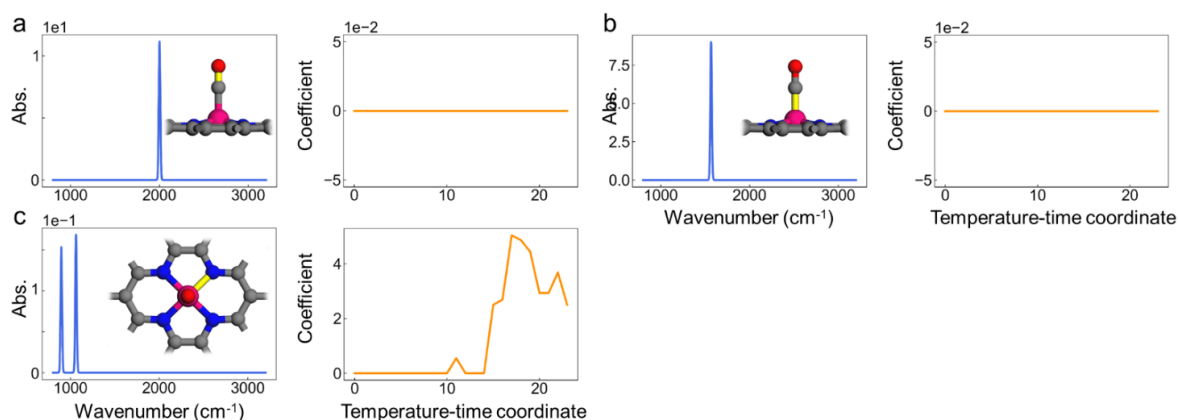
Figure S18. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$-CO in Pt-doped ZIF-67 pyrolysis. (a) C-O bond; (b) Co-C bond; (c) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
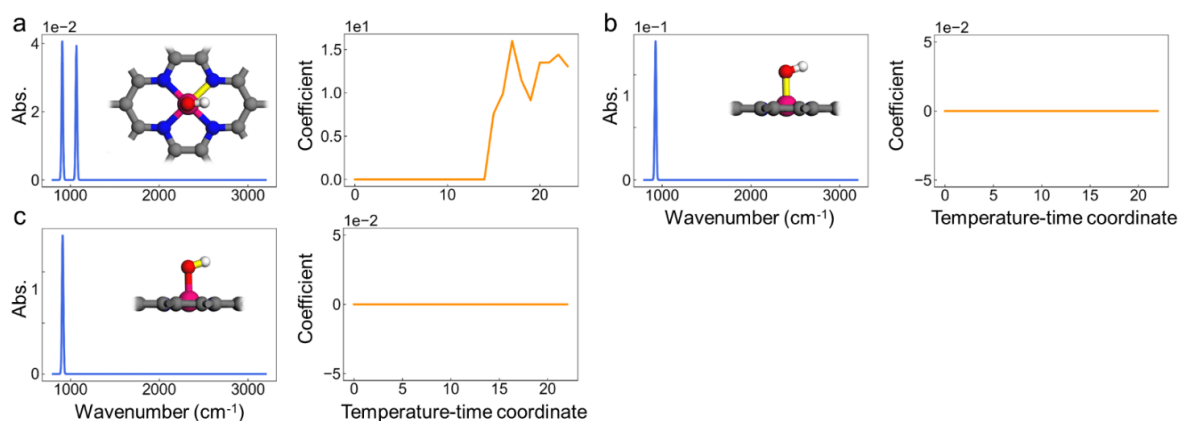

Figure S19. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_4$-OH in Pt-doped ZIF-67 pyrolysis. (a) Co-N bond; (b) Co-O bond; (c) O-H bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
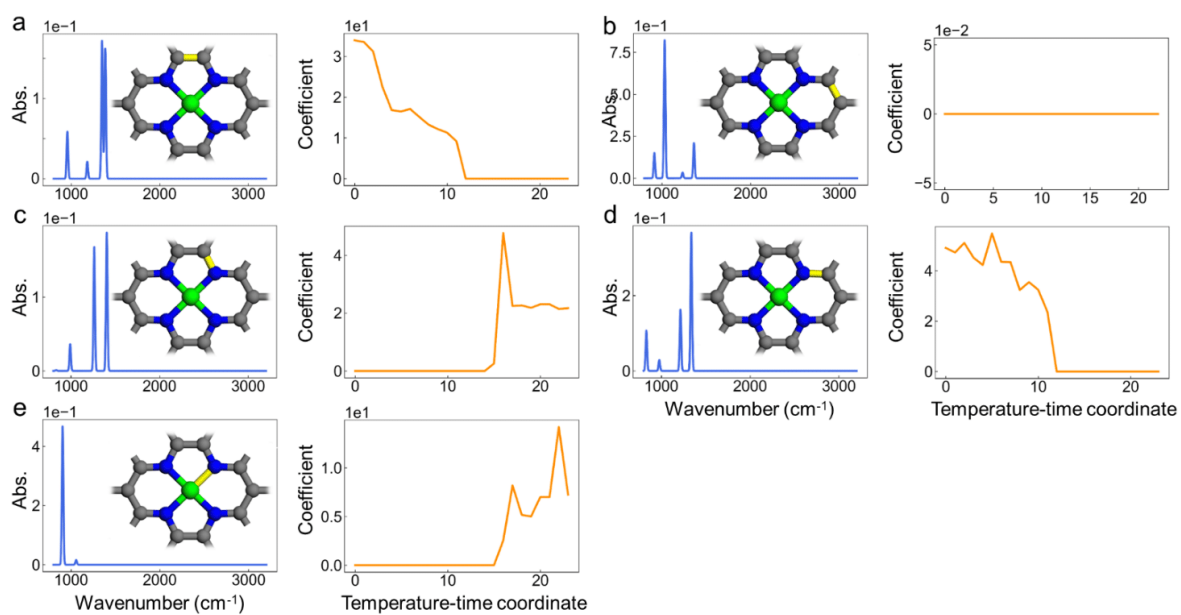
Figure S20. Calculated IR spectrum and corresponding LASSO coefficients for PtN$_4$ in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Pt-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
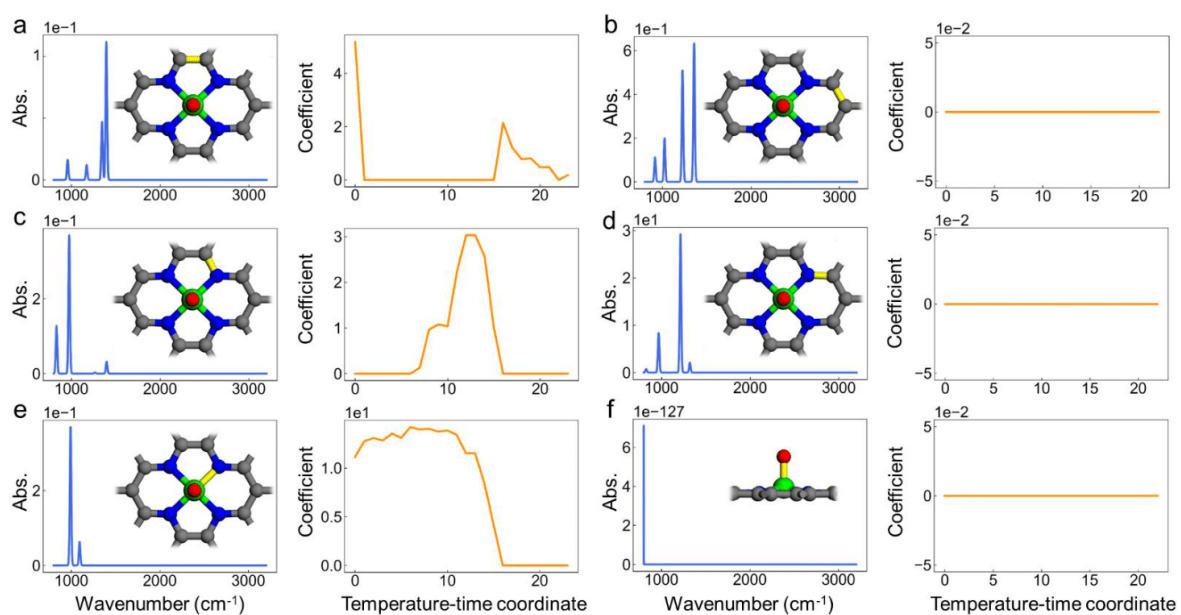

Figure S21. Calculated IR spectrum and corresponding LASSO coefficients for PtN$_4$-O in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Pt-N bond; (f) Pt-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
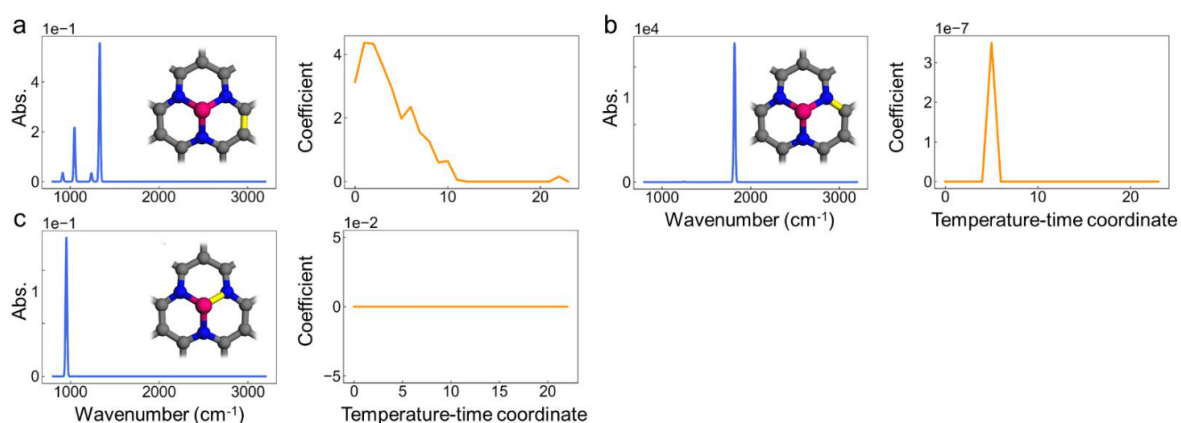
Figure S22. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$ in Pt-doped ZIF-67 pyrolysis. (a) C-C bond; (b) C-N bond; (c) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S23. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$-O in Pt-doped ZIF-67 pyrolysis. (a) C-C bond; (b) C-N bond; (c) Co-N bond; (d) Co-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S24. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$-CO in Pt-doped ZIF-67 pyrolysis. (a) C-O bond; (b) Co-C bond; (c) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
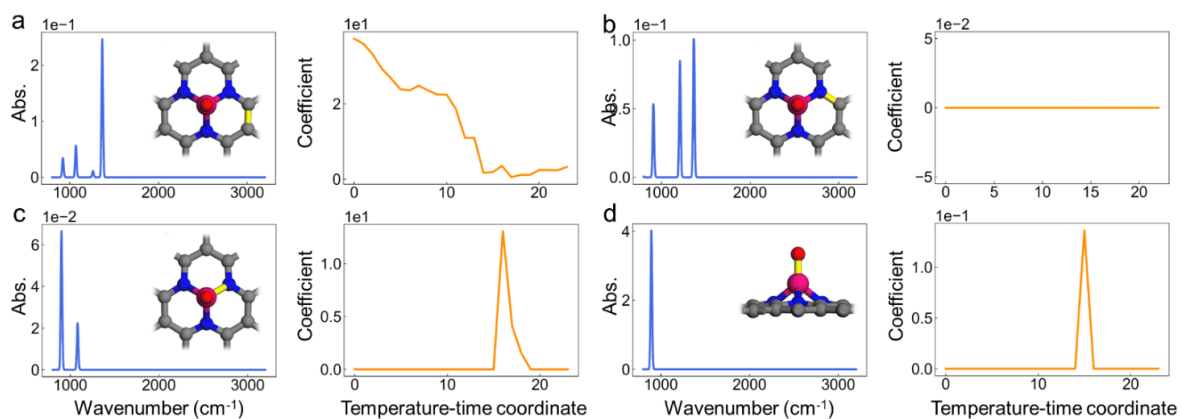

Figure S25. Calculated IR spectrum and corresponding LASSO coefficients for CoN$_3$-OH in Pt-doped ZIF-67 pyrolysis. (a) Co-N bond; (b) Co-O bond; (c) O-H bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

Figure S26. Calculated IR spectrum and corresponding LASSO coefficients for PtN$_3$ in Pt-doped ZIF-67 pyrolysis. (a) C-C bond; (b) C-N bond; (c) Pt-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

S15

Figure S27. Calculated IR spectrum and corresponding LASSO coefficients for PtN$_3$-O in Pt-doped ZIF-67 pyrolysis. (a) C-C bond; (b) C-N bond; (c) Pt-N bond; (d) Pt-O bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
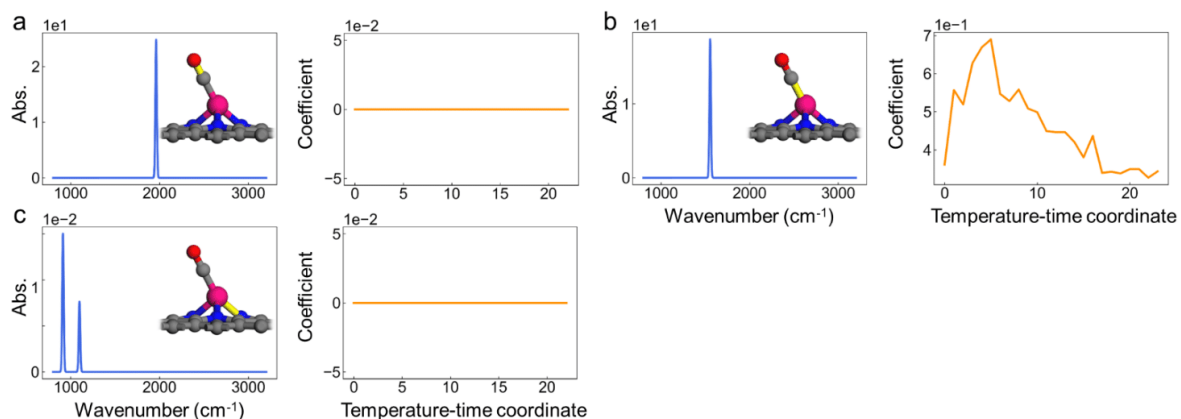

Figure S28. Calculated IR spectrum and corresponding LASSO coefficients for PtN$_3$-CO in Pt-doped ZIF-67 pyrolysis. (a) C-O bond; (b) Pt-C bond; (c) Pt-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
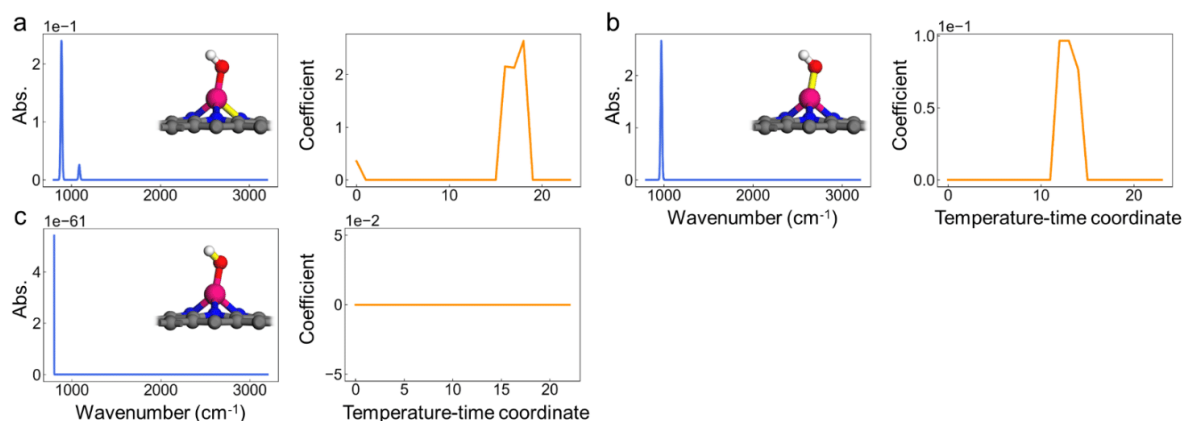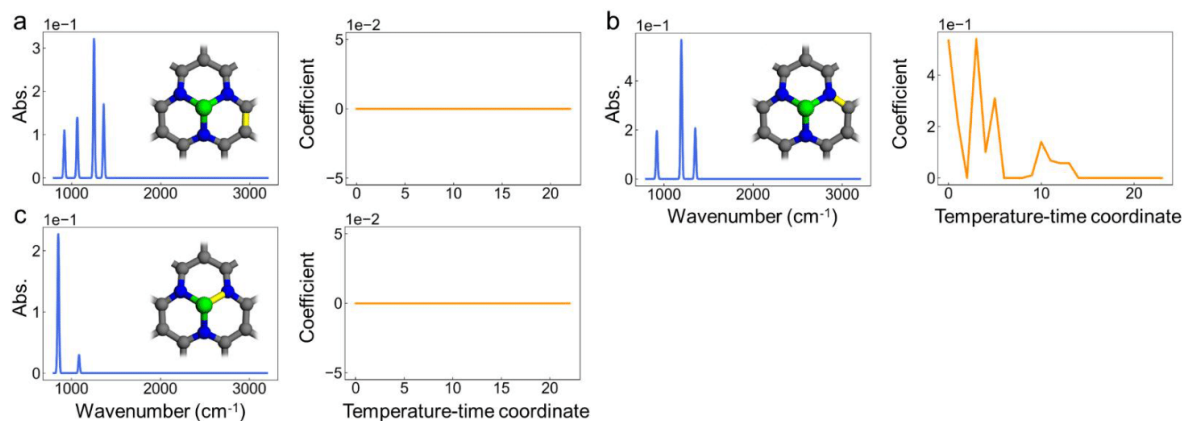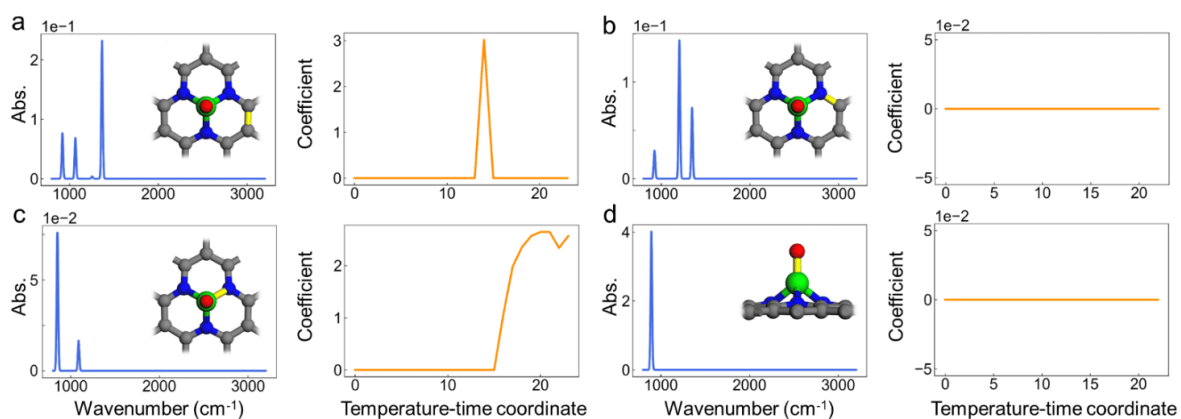
Figure S29. Calculated IR spectrum and corresponding LASSO coefficients for Co_3N_6 in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-Co bond; (f) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
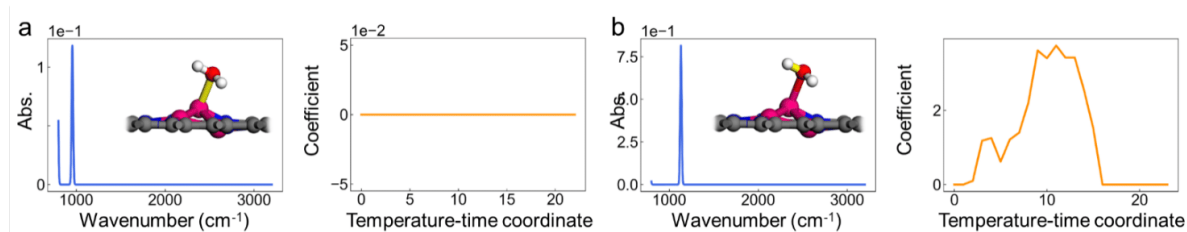
Figure S30. Calculated IR spectrum and corresponding LASSO coefficients for Co_3N_6-H_2O in Pt-doped ZIF-67 pyrolysis. (a) Co-O bond; (b) O-H bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
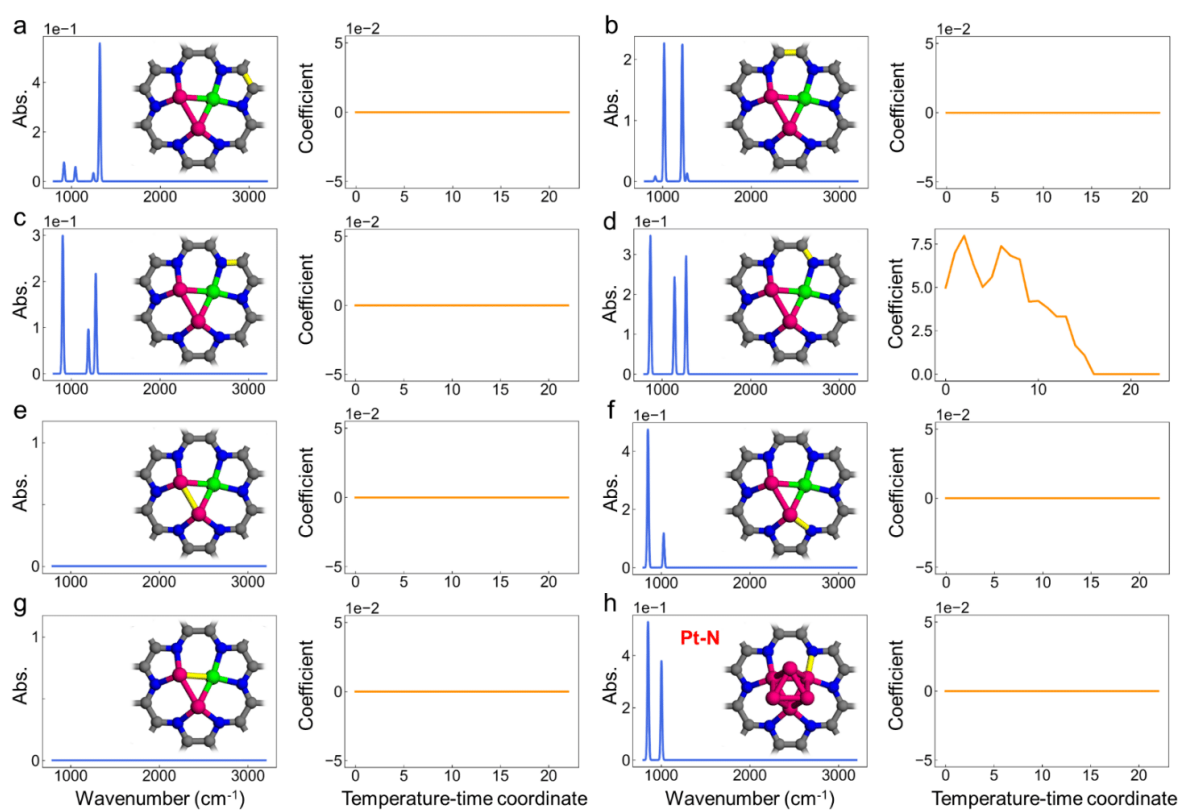
Figure S31. Calculated IR spectrum and corresponding LASSO coefficients for Co_2PtN_6 in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-Co bond; (f) Co-N bond; (g) Co-Pt bond; (h) Pt-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.
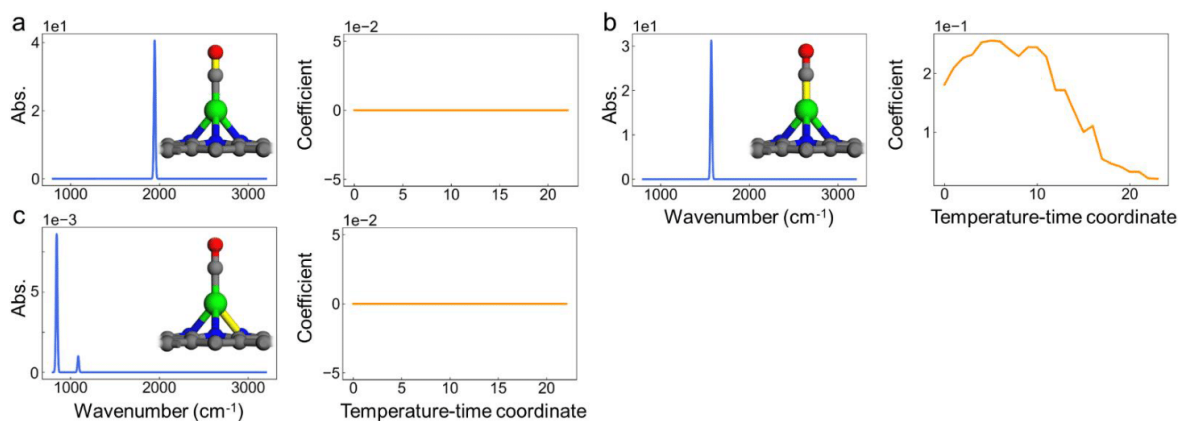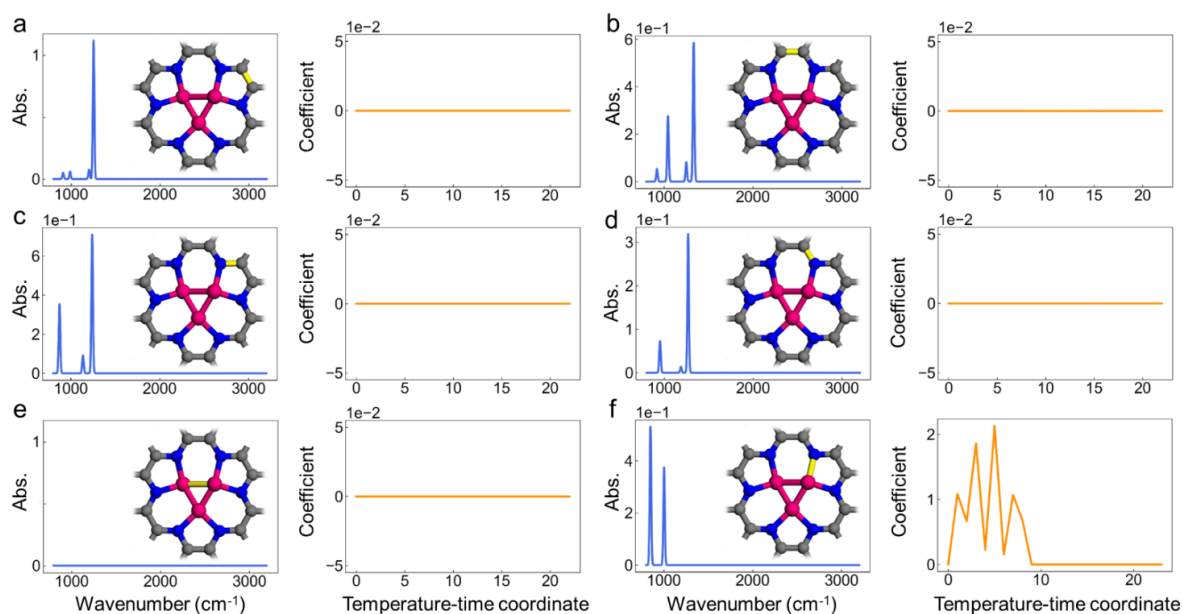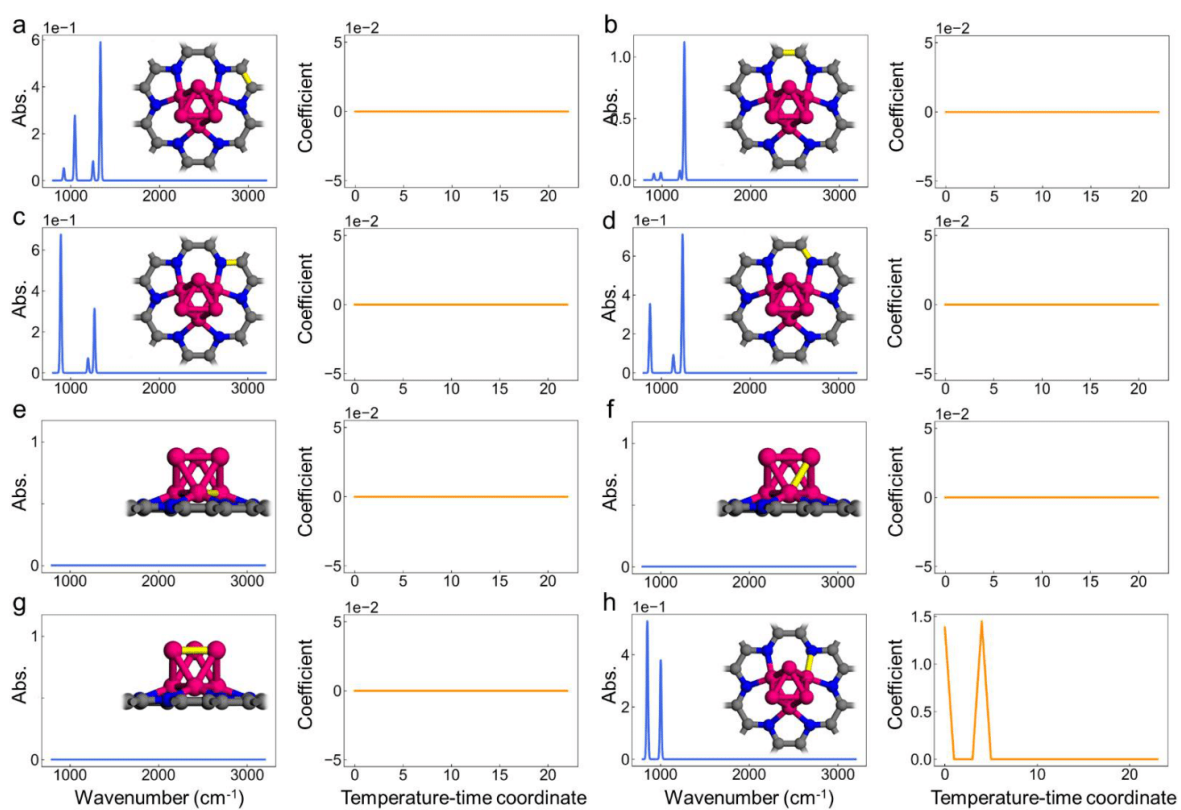
Figure S32. Calculated IR spectrum and corresponding LASSO coefficients for Co_6N_6 in Pt-doped ZIF-67 pyrolysis. (a) C-C bond in the 5-membered ring; (b) C-C bond in the 6-membered ring; (c) C-N bond in the 5-membered ring; (d) C-N bond in the 6-membered ring; (e) Co-Co bond in the bottom layer; (f) Co-Co bond between the bottom and top layer; (g) Co-Co bond in the top layer; (h) Co-N bond. The calculated chemical bonds are indicated in yellow, as shown in the insets.

S19

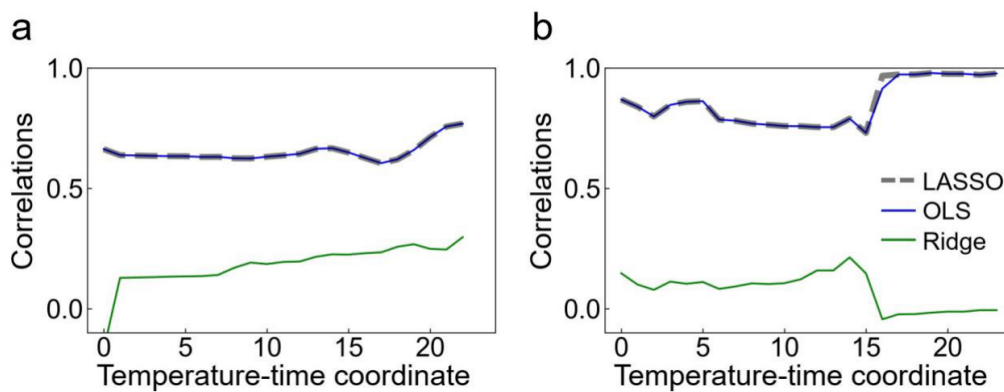**Figure S33.** Pearson correlations across the temperature-time coordinates using different regression algorithms in ZIF-67 (a) and Pt-doped ZIF-67 (b) pyrolysis.
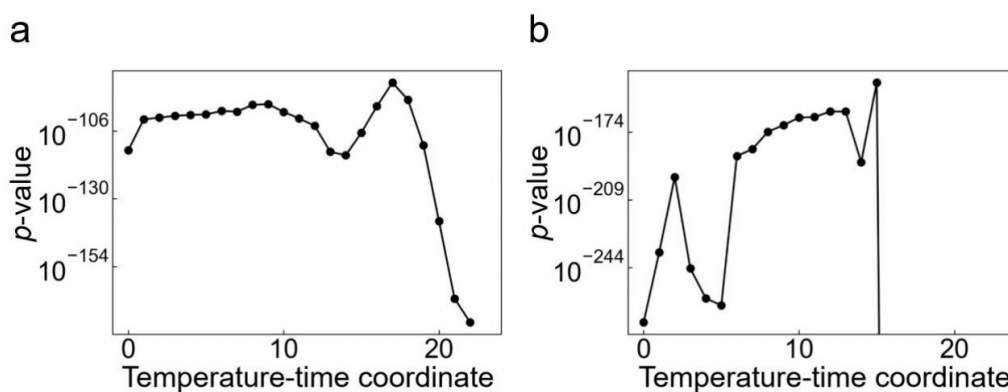

Figure S34. LASSO *p*-value across the temperature-time coordinates in ZIF-67 (a) and Pt-doped ZIF-67 (b) pyrolysis.

Table S1. Chemical bonds in the database for ZIF-67 pyrolysis.

Structure	Number of bonds	Type of bond	Corresponding Figure
ZIF-67	7	C-C (2); C-H (2); C-N (2); Co-N (1)	Figure S1
Oxidized ZIF-67	2	Co-N (1); Co-O (1)	Figure S2
CoN_4	5	C-C (2) ; C-N (2); Co-N (1)	Figure S3
CoN_4-O	6	C-C (2); C-N (2); Co-N (1); Co-O (1)	Figure S4
CoN_4-CO	3	C-O (1); Co-C (1); Co-N (1)	Figure S5
CoN_4-OH	3	Co-N (1); Co-O (1); O-H (1)	Figure S6
CoN_3	3	C-C (1); C-N (1); Co-N (1)	Figure S7
CoN_3-O	4	C-C (1); C-N (1); Co-N (1); Co-O (1)	Figure S8
CoN_3-CO	3	C-O (1); Co-C (1); Co-N (1)	Figure S9
CoN_3-OH	3	Co-N (1); Co-O (1); O-H (1)	Figure S10
Co_3N_6	6	C-C (2); C-N (2); Co-Co (1); Co-N (1)	Figure S11
Co_3N_6-H_2O	2	Co-O (1); O-H (1)	Figure S12
Co_6N_6	8	C-C (2); C-N (2); Co-Co (3); Co-N (1)	Figure S13
SUM	55		

Table S2. Chemical bonds in the database for Pt-doped ZIF-67 pyrolysis.

Structure	Number of bonds	Type of bond	Corresponding Figure
Pt-ZIF-67	8	C-C (2); C-H (2); C-N (2); Co-N (1); Pt-N (1)	Figure S14
Oxidized Pt-ZIF-67	4	Co-N (1); Co-O (1); Pt-N (1); Pt-O (1)	Figure S15
CoN_4	5	C-C (2); C-N (2); Co-N (1)	Figure S16
CoN_4-O	6	C-C (2); C-N (2); Co-N (1); Co-O (1)	Figure S17
CoN_4-CO	3	C-O (1); Co-C (1); Co-N (1)	Figure S18
CoN_4-OH	3	Co-N (1); Co-O (1); O-H (1)	Figure S19
PtN_4	5	C-C (2); C-N (2); Pt-N (1)	Figure S20
PtN_4-O	6	C-C (2); C-N (2); Pt-N (1); Pt-O (1)	Figure S21
CoN_3	3	C-C (1); C-N (1); Co-N (1)	Figure S22
CoN_3-O	4	C-C (1); C-N (1); Co-N (1); Co-O (1)	Figure S23
CoN_3-CO	3	C-O (1); Co-C (1); Co-N (1)	Figure S24
CoN_3-OH	3	Co-N (1); Co-O (1); O-H (1)	Figure S25
PtN_3	3	C-C (1); C-N (1); Pt-N (1)	Figure S26
PtN_3-O	4	C-C (1); C-N (1); Pt-N (1); Pt-O (1)	Figure S27
PtN_3-CO	3	C-O (1); Pt-C (1); Pt-N (1)	Figure S28

Co₃N₆	6	C-C (2); C-N (2); Co-Co (1); Co-N (1)	Figure S29
Co₃N₆-H₂O	2	Co-O (1); O-H (1)	Figure S30
Co₂PtN₆	7	C-C (2); C-N (2); Co-Co (1); Co-N (1); Co-Pt (1)	Figure S31
Co₆N₆	8	C-C (2); C-N (2); Co-Co (3); Co-N (1)	Figure S32
SUM	86		

Table S3. Temperature and time coordinate in **Figures S1-S13**, **S14-S32**.

	ZIF-67		Pt-ZIF-67	
Coordinate	**Temperature (°C)**	**Time (min)**	**Temperature (°C)**	**Time (min)**
0	20	-	20	-
1	50	-	50	-
2	100	-	75	-
3	150	-	100	0
4	175	-	100	10
5	200	0	100	20
6	200	10	120	-
7	225	-	150	-
8	250	-	175	-
9	275	0	200	0
10	275	10	200	15
11	300	0	225	-
12	300	20	250	0
13	300	30	250	10
14	300	40	270	-
15	300	50	300	0
16	300	60	300	15
17	300	70	300	30
18	300	80	300	40
19	300	90	300	50
20	300	100	300	60
21	300	110	300	70
22	300	120	300	80
23			300	90

S22

Table S4. Comparison of IR spectra frequency and intensity under different convergence of force criteria.

Convergence of force (eV/Å)	Co-O bond in **CoN₃-OH**		Pt-O bond in **PtN₃-OH**	
	Frequency (cm⁻¹)	Intensity (a.u.)	Frequency (cm⁻¹)	Intensity (a.u.)
0.05	969.9	2.66	1115.5	5.83
0.01	969.9	2.67	1115.5	5.83
0.005	969.9	2.68	1115.5	5.94

Table S5. Comparison of IR spectra frequency and intensity under different U values. $U = 3.5$ eV for Co[9] and 7.5 eV for Pt[10] were applied as the median values of the tests, respectively.

Co-O bond in **CoN₃-OH**		
U (eV)	Frequency (cm⁻¹)	Intensity (a.u.)
0	969.9	2.66
3.0	989.8	2.68
3.5[9]	978.6	2.75
4.0	990.0	2.67
Pt-O bond in **PtN₃-OH**		
U (eV)	Frequency (cm⁻¹)	Intensity (a.u.)
0	1115.0	5.83
7.0	1143.9	5.79
7.5[10]	1141.3	5.79
8.0	1136.4	5.82

S23

Supplementary References

(1) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane wave basis set. *J. Phys. Rev B*. **1996**, 54, 11169-11186.

(2) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, 77, 3865.

(3) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B*. **1994**, 50, 17953-17979.

(4) Monkhorst, H. J.; Pack J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B*. **1976**, 13, 5188–5192.

(5) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Chem. Phys.* **2010**, 132, 154104.

(6) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D. The atomic simulation environment-a Python library for working with atoms. *J. Phys.: Condens. Matter.* **2017**, 29, 273002.

(7) Frederiksen, T.; Paulsson, M.; Brandbyge, M.; Jauho, A. P. Inelastic transport theory from first principles: Methodology and application to nanoscale devices. *Phys. Rev. B*. **2007**, 75, 205413.

(8) Kim, S. J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale l1-regularized least squares. *J. Mach. Learn. Res*. **2007**, 1, 606-617.

(9) Capdevila-Cortada, M.; Łodziana, Z.; López, N. Performance of DFT+U approaches in the study of catalytic materials. *ACS Catal*. **2016**, 6, 8370-8379.

(10) Tian, Z. A DFT+U study of the segregation of Pt to the CeO_{2-x} Σ3[110]/(111) grain boundary. arXiv:1710.03929, **2017**.

Chapter 4: Conclusions and Perspectives

4.1 Conclusions

In summary, this thesis involves the utilization of a ML tool based on the LASSO algorithm for the comprehensive analysis of in situ temperature-dependent DRIFTS in experimental setting. In chapter 3, this thesis are achieved the comprehensive modelling of chemical bond information that relevant to the pristine ZIF-67 and Pt-doped ZIF-67, and the establishment of a theoretical database have enabled the trained ML model to effectively emulate real experimental data. The emulation is evident in correlation values can be highest to 0.9, affirming the proficiency of machine learning in replicating experimental data. Moreover, ML-generated correlations, obtained while fitting experimental data for various chemical bonds, can serve as valuable indicators for deducing chemical structures.

Integrating variations along the temperature–time axis provides profound insights into chemical reaction mechanisms. The degradation of the ZIF framework, the gradual oxidation of oxygen-containing species on the metal, and the formation of Co–O and Pt–O bonds observed align closely with results from diverse experimental characterizations, which has been shown in chapter 3. Notably, the comprehensive workflow we have established heavily depends on database construction and ML algorithms, requiring minimal manual intervention in chapter 3.

The understanding of the reaction mechanisms, vibrational mode of chemical bonds between ZIF-67 and Pt-doped ZIF-67, IR designed LASSO machine learning algorithm and reaction pathway for producing more single atom catalysts helps guide the design of other electrocatalyst materials in MOFs.

4.2 Perspectives

Therefore, we posit that this approach has substantial potential for practical and can be expanded to a diverse array of applications for intelligent analysis of in situ experimental characterization data in the future. These include:

1. Currently, we only process ZIF-67 and Pt-doped ZIF-67 in our machine learning model There lack of enough data for us to enable the model to reveal the single atom catalysts from other metal doped ZIF-67. Therefore, we plan to expand the training set which can enable us to predict more kinds of SACs from different MOFs.

2. Linear regression model such as LASSO has its restriction in predicting certain range of IR spectra, so further experiments can be focused on using different machine learning model to conduct the research, this also can benefit on verifying the common application on different types of MOFs.

3. IR spectra as the training set is still not enough to reveal the reaction pathway in more complex reaction, so we plan to feed more experimental graphs such as UV-Visible Spectra, X-ray Diffraction (XRD) Patterns, etc. This can enhance the accuracy of the model and reveal more reaction details.