Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5dd00293a

Sparse identification of chemical reaction mechanisms from limited concentration profiles†

Shun Hayashi *

Automating the discovery of chemical reaction mechanisms can increase the efficiency of using experimental data to obtain chemical knowledge. In this study, a sparse identification approach was employed to determine reaction mechanisms, providing accurate and interpretable kinetic models while preventing overfitting. The main advantage of the proposed approach over conventional sparse identification algorithms is that it can be applied to cases with limited concentration profiles, which often occur in chemical reactions involving untraceable intermediates. To demonstrate its applicability to complex reaction mechanisms beyond the reach of classical kinetic analysis, the autocatalytic reduction of manganese oxide ions was selected as the target reaction. Although the concentrations of only two manganese species could be monitored *via* UV-vis absorption spectroscopy, the experimental data were sufficiently represented by 11 elementary steps involving 8 chemical species. This strategy enables the automated discovery of reaction mechanisms without relying on heuristic kinetic models, as the only assumption required is the composition of the intermediates.

Received 3rd July 2025
Accepted 19th July 2025

DOI: 10.1039/d5dd00293a

rsc.li/digitaldiscovery

Introduction

Understanding the mechanisms of chemical reactions is essential for the design of new synthetic processes, the exploration of novel catalysts, and the optimization of industrial processes.[1] Reaction mechanisms are determined by proposing kinetic models, which consist of a series of elementary steps and their associated rate constants, that reproduce temporal concentration profiles. Classic approaches involve the use of apparent kinetic models, in which a single rate equation that serves as a function of the concentration of the detectable species is used. This includes traditional approaches in which logarithmic plots of the initial reaction rates are used and modern approaches in which graphical overlays of entire concentration profiles are employed.[2–5] Methods involving nonlinear fitting,[6] symbolic regression,[7] and multiobjective optimization[8] have also been developed. These approaches provide the simplest forms of kinetic models, which are generally sufficient to explain and predict kinetic behavior. However, understanding the reaction mechanisms based on apparent kinetic models requires expert knowledge and experience.

Recent advances in both experimental and computational technologies have provided chemists with enhanced tools for analyzing reaction mechanisms. The development of analytical techniques has enabled *in situ*/*operando* identification of intermediates, providing crucial insights for manually constructing reaction mechanisms.[9] The use of both automated experiments and reaction monitoring techniques allows for the discovery and optimization of reactions and the collection of kinetic data.[10–13] The integration of reaction monitoring techniques with conventional kinetic analysis methods enables automated mechanistic studies.[14] Computational techniques, such as kinetic Monte Carlo simulations[15,16] and quantum chemical calculations,[17,18] can also be used to identify plausible reaction paths. Furthermore, machine learning technologies have shown potential for predicting reactivity on the basis of reaction and mechanism databases[19–21] and for automatically classifying reaction mechanisms.[22] However, with respect to discovering reaction mechanisms on the basis of experimental data, the existing theoretical frameworks are insufficient, particularly considering advances in data collection techniques. Automating the exploration of reaction mechanisms requires selecting the minimal number of elementary steps from complexly entangled reaction networks and estimating the rate constants from experimental data.

Recently, equation discovery, a computational approach for extracting governing ordinary differential equations (ODEs) from experimental data, has emerged as a powerful tool for understanding nonlinear dynamical systems.[23,24] Notably, an approach involving sparse regression, particularly sparse identification of nonlinear dynamics (SINDy), has attracted considerable attention because it allows for the identification of compact and interpretable models that capture the core dynamics of complex systems.[25] Since chemical reaction mechanisms can be uniquely converted into ODEs or rate

Division of Physical Sciences, Department of Science, National Museum of Nature and Science, Ibaraki 305-0005, Japan. E-mail: s-hayashi@kahaku.go.jp

† Electronic supplementary information (ESI) available: Detailed experimental procedures, computational methodologies, and optimized kinetic models. See DOI: https://doi.org/10.1039/d5dd00293a

equations, according to the law of mass action, sparse regression offers a promising approach for determining the reaction mechanisms. When concentration profiles for all chemical species are available, this approach can be applied to identify chemical reaction mechanisms.[26] However, only limited concentration profiles are typically available, as monitoring the concentrations of all chemical species, including dilute and reactive intermediates, is difficult.

Inspired by the sparse identification approach, this study presents a simple yet effective method for the data-driven discovery of chemical reaction mechanisms from limited concentration profiles. Since concentration profiles can be simulated by numerically solving initial value problems of rate equations,[27–29] the rate constants are estimated by fitting the simulated profiles to experimental data. The important elementary steps are selected by eliminating those with negligibly small rate constants. In summary, at the cost of numerically solving rate equations, the discovery of chemical reaction mechanisms from limited concentration profiles can be viewed as a sparse regression problem for determining rate constants. To demonstrate the potential of this strategy, the mechanism of the autocatalytic reduction of manganese oxide ions was studied. The experimental data can be sufficiently represented by 11 elementary steps involving 8 chemical species. The proposed approach holds promise as a framework for model discovery in various fields of physics where comprehensive measurements are impractical.

Results and discussion

Application to autocatalytic reactions

To demonstrate the applicability of the proposed method to distinct and complex reactions that cannot be adequately represented by apparent kinetic models, an autocatalytic reaction was selected as the target reaction. Autocatalysis, a chemical process in which the products of a reaction amplify their own formation rate, has attracted significant interest from researchers across a wide range of fields. It is widely accepted that autocatalytic processes are utilized in asymmetric synthesis[30] and systems chemistry[31] and play key roles in metabolism, self-replicating systems, crystal growth, and the origin of life.[32–34] The reduction of permanganate ions ($[MnO_4]^-$) with oxalic acid ($H_2C_2O_4$) was used as the target autocatalytic reaction because the mechanism was not fully understood, although the reactions could be easily monitored *via* UV-vis absorption spectroscopy.[35,36] The reduction of Mn^{7+} to Mn^{2+} proceeds *via* autocatalytic processes (*e.g.*, $Mn^{7+} + Mn^{2+} \rightarrow 2\ Mn^{2+}$); therefore, the addition of Mn^{2+} greatly accelerates this reaction. The presence of $[Mn(C_2O_4)_3]^{3-}$ as a traceable partially reduced intermediate has been reported.[37] In this study, the changes in the concentrations of Mn^{7+} ($[MnO_4]^-$, 525 nm) and Mn^{3+} ($[Mn(C_2O_4)_3]^{3-}$, 273 nm) over time were monitored (Fig. 1a and S1†). Twenty initial conditions, each with varying initial concentrations of Mn^{7+} and Mn^{2+} ($MnSO_4$), were used for data collection. The dataset was then split into eight training sets

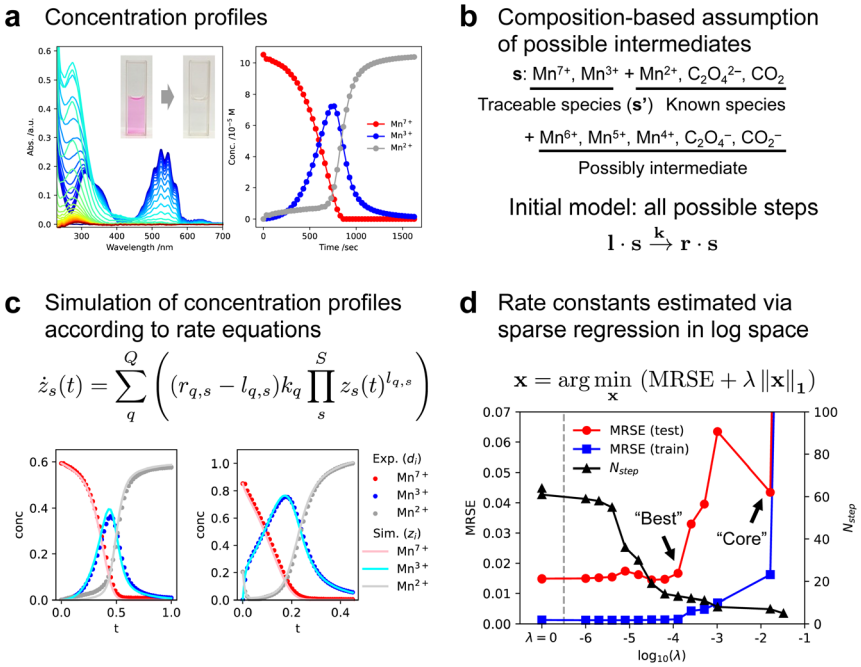

Fig. 1 Discovery of kinetic models representing chemical reaction mechanisms *via* a sparse identification approach. (a) Temporal concentration profiles of traceable species (Mn^{7+} and Mn^{3+}) were collected through spectroscopic reaction monitoring. (b) Based on the assumption regarding the intermediate compositions, the initial model is generated by including all possible steps involving 1 or 2 reactants that satisfy mass conservation. (c) The residual (mean relative squared error, MRSE) is evaluated by comparing the simulated concentration profiles derived from rate equations uniquely formulated based on the kinetic models with experimental data. (d) The rate constants were optimized *via* an L_1-regularized regression algorithm in log space ($\mathbf{x} = \log_{10}(\mathbf{k} + 1)$). By gradually increasing λ, the generalized model ("best") and a model with the minimal number of elementary steps ("core") are obtained.

and twelve test sets, with the initial conditions of the training data remaining within the range of the test data (Fig. S2†).

The only requirement for our approach is the assumption of the compositions of the intermediates, which are used to generate the initial kinetic models. On the basis of the compositions of five known compounds (Mn^{7+}, Mn^{3+}, Mn^{2+}, $C_2O_4^{2-}$, and CO_2), we assumed the presence of five possible intermediates, namely, partially reduced/oxidized species of Mn^{7+} and $C_2O_4^{2-}$ (Mn^{6+}, Mn^{5+}, Mn^{4+}, $C_2O_4^{-}$, and CO_2^{-}). The initial kinetic model was then generated by enumerating all possible elementary steps that satisfy mass conservation (Fig. 1b). The following two constraints were applied: (i) only steps involving 1 or 2 reactants were considered; (ii) certain steps were excluded on the basis of chemical knowledge. The first constraint is reasonable because reactions involving three or more reactants have lower collision probabilities than do reactions involving 1 or 2 reactants. According to the second constraint, the steps that involve CO_2 as a reactant are omitted. While knowledge of molecular structures and chemical properties is not essential for developing a kinetic model, this information is useful for reducing the number of elementary steps, which is crucial in the model optimization process. Among the 64 steps prepared with this approach, the key elementary steps were identified *via* sparse regression.

The kinetic models describing chemical reaction mechanisms can be represented *via* three arrays. A set of elementary steps can be denoted by two integer matrices: one for the left-hand side ($\mathbf{l}$) and one for the right-hand side ($\mathbf{r}$) of the chemical equations. The corresponding rate constants are stored in a vector ($\mathbf{k}$). The rate equations are uniquely generated from the three arrays according to the law of mass action (eqn (1)).

$$\dot{z}_s(t) = \sum_{q}^{Q} \left(\left(\mathbf{r}_{q,s} - \mathbf{l}_{q,s}\right)\mathbf{k}_q \prod_{s}^{S} z_s(t)^{\mathbf{l}_{q,s}} \right) \tag{1}$$

Here, z_s and Q represent the concentration of chemical species s and the number of elementary steps, respectively. An example of converting a reaction mechanism to rate equations is shown in Fig. S3.† Simulated concentration profiles are generated by numerically solving the initial value problems of the rate equations (Fig. 1c). The model's accuracy was estimated using the mean relative squared error (MRSE), which represents the average of the residuals between the experimental and simulated concentration profiles, as shown in eqn (2).

$$\mathrm{MRSE} = \frac{1}{N_i} \sum_{i}^{N_i} \frac{\left\| d_{i,s'} - z_{i,s'} \right\|_{\mathrm{F}}^{2}}{\left\| d_{i,s'} - \overline{d}_{i,s'} \right\|_{\mathrm{F}}^{2}} \tag{2}$$

Here, $d_{s'}$ and $z_{s'}$ represent the experimental and simulated concentrations of the two traceable species, Mn^{7+} and Mn^{3+}. N_i represents the number of kinetic data with different initial conditions. $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Since the rate constants can have a wide range of values, the vector $\mathbf{k}$ was transformed into the logarithmic-like vector $\mathbf{x} = \log_{10}(\mathbf{k} + 1)$. To promote sparsity in the resulting models, the L_1-regularized regression algorithm (LASSO[38]) was also employed. The loss function was defined as the sum of MRSE and the L_1-norm of $\mathbf{x}$ ($\|\mathbf{x}\|_1$) with a regularization parameter λ (eqn (3)).

$$\mathbf{x} = \arg\ \min_{\mathbf{x}} (\mathrm{MRSE} + \lambda \|\mathbf{x}\|_1) \tag{3}$$

The sparse regression approach discovers the intrinsic elementary steps from a comprehensive set of candidates. The steps with minimal contribution can be eliminated, as the rate constant becomes negligibly small in the L_1-regularization algorithm. Among a variety of potentially analogous reaction paths, those involving traceable species are preferentially selected over those consisting solely of intermediates. This is because of the large rate constants associated with steps involving intermediates. Therefore, this approach eliminates redundant paths and generates models with reduced dependence on intermediates. The covariance matrix adaptation evolution strategy (CMA-ES),[39] a state-of-the-art optimizer for continuous black-box functions widely used for hyperparameter tuning in deep neural networks,[40] was used to solve the minimization problem. By gradually increasing λ, the elementary steps with rate constants below a certain threshold are progressively eliminated from the resulting models, thereby reducing the total number of elementary steps (N_{step}). Starting from the "full" model ($\lambda = 0$), the λ value is continuously increased to find the "best" model and the "core" model ($N_{\mathrm{step}} = 6$–7) (Fig. 1D). The "best" model is defined as the model with the minimum N_{step} that satisfies the condition $\mathrm{MRSE} < \mathrm{MRSE}_{\mathrm{full}} \times 1.2$.

As λ increases, the number of elementary steps decreases, whereas the residuals for the training data increase monotonically. The residuals for the test data initially showed little change but increased after λ reached 1.3×10^{-4} ($N_{\mathrm{step}} = 13$). This finding indicates that the full model ($N_{\mathrm{step}} = 64$) was overfitted, highlighting the importance of applying the sparse regression approach to identify generalized kinetic models. As λ is increased further, a core model with a minimal number of elementary steps is obtained, which is beneficial for interpreting the reaction mechanism. The advantage of this approach lies in its ability to provide both the generalized model and the minimal model from a set of all possible elementary steps.

Determination of key intermediates and elementary steps

The full models with 3, 4 or 5 intermediates were also studied to determine which intermediates are important for constructing the kinetic models. Among the 12 combinations, each consisting of 3, 4, or 5 intermediates, the accuracy of the model was evaluated based on the residuals for both the training and test datasets.

The best 4 combinations shared the same three intermediates: Mn^{6+}, Mn^{4+}, and $C_2O_4^{-}$ (Fig. 2A and S4†). Interestingly, among the best 4 models, the model with fewer intermediates better fits the test data, whereas the model with more intermediates better fits the training data. These findings suggest that three intermediates are sufficient to explain the experimental data. The 3 worst models among the 12 models do not include Mn^{6+}, implying that Mn^{6+} plays an important role in this reaction. Furthermore, for mechanisms involving 3 or 5 intermediates containing Mn^{6+}, Mn^{4+}, and $C_2O_4^{-}$, simplified models were generated by increasing λ, and the residuals were

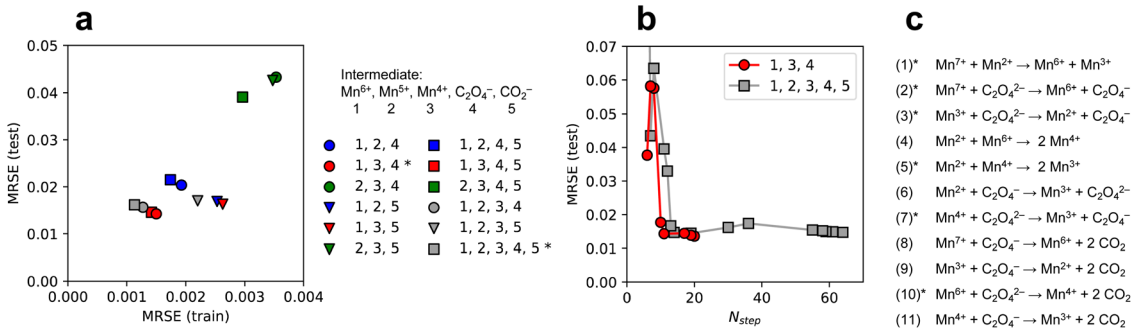

Fig. 2 Determination of key intermediates by comparison of the kinetic models. (a) The residuals (mean relative squared error, MRSE) are estimated for 3, 4, and 5 intermediate mechanisms for the test and training datasets with $\lambda = 0$. The points marked with asterisks in the legend represent the averages of 10 calculations (Fig. S4†). (b) Residuals plotted as a function of the number of elementary steps for 3 and 5 intermediate mechanisms containing three key intermediates, Mn^{6+}, Mn^{4+}, and $C_2O_4^{-}$. (c) A set of elementary steps in the best model with 3-intermediate mechanisms. The six steps marked with asterisks represent the core model.

plotted as a function of N_{step} (Fig. 2B and S6–11†). These plots displayed similar trends regarding the residuals' dependence on N_{step}: while the residuals remained relatively stable for N_{step} > 11–13, they increased considerably with each increase in N_{step} at smaller N_{step} values. This suggests that the accuracy of the models depends more on the number of elementary steps than on the number of intermediates considered. While redundancy in the reaction mechanism does not improve the model's accuracy, omitting even a single critical reaction can lead to a significant decrease in accuracy. The core model with three intermediates (Mn^{6+}, Mn^{4+}, and $C_2O_4^{-}$) can be interpreted as a combination of an autocatalytic reaction ($Mn^{7+} + 2\ Mn^{2+} \rightarrow 3\ Mn^{3+}$) and a sequential ($Mn^{7+} \rightarrow Mn^{3+}$) reduction of Mn^{7+}. This leads to the formation of Mn^{3+}, which is subsequently reduced to Mn^{2+} (Fig. 2C).

To highlight the advantages of the proposed method over empirical approaches, the kinetic models generated in this study were compared with previously developed models in the literature and numerical models (Fig. 3 and S12–15†). The reported mechanism, consisting of 7 elementary steps involving three intermediates (Mn^{6+}, Mn^{4+}, and partially oxidized $C_2O_4^{2-}$), was heuristically proposed based on existing chemical knowledge and kinetic studies of various manganese oxide species.[37] In this study, the initial literature model was developed by incorporating an essential step ($Mn^{7+} \rightarrow Mn^{6+}$) into the reported 7-step mechanism (Fig. S12†). The accuracy of the best model, which was developed based on the literature model, was better than that of the full model, indicating that our approach can be applied to heuristic models for estimating rate constants and reducing redundancy. The best model developed based on the 3-intermediate mechanism showed a better fit than the best model of the literature mechanism did, highlighting the importance of considering all possible elementary steps while avoiding overfitting. The numerical mechanism denotes the apparent kinetic model, which uses only the concentrations of known manganese oxide species (Mn^{7+}, Mn^{3+}, and Mn^{2+}). The numerical model allows for the inclusion of up to five reactant elementary steps and the displacement of the sum of the oxidation states in each elementary step (Fig. S14†). The advantage of numerical models is that the absence of untraceable species enables model optimization *via* the conventional SINDy method,[25,26] which does not require numerically solving ODEs. Nonetheless, the same algorithm was applied to our approach for comparison. The numerical model poorly fit the experimental data, suggesting the difficulty of constructing kinetic models from only the concentrations of traceable species. This suggests that our approach, which employs a composition-based assumption for undetected intermediates, is effective, as it accurately represents the underlying chemical reaction networks.

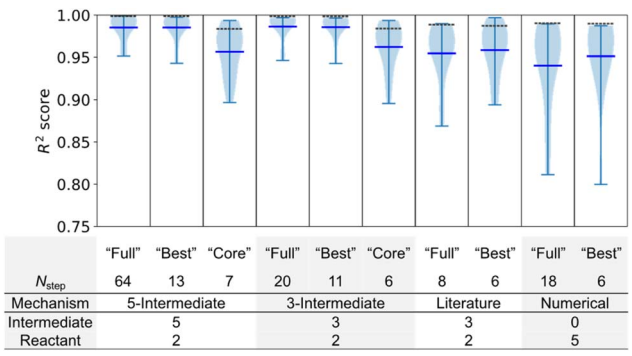

Fig. 3 Comparison of the kinetic models generated by considering various mechanisms. Three models with different numbers of elementary steps (N_{step}), denoted as "full", "best", and "core", were generated by gradually increasing λ. For the literature and numerical mechanisms, the "best" and "core" models are identical. The violin plot with a blue average line displays the R^2 score of the model with the test dataset. The width of each violin indicates the density of data points, and the whiskers represent the minimum and maximum R^2 scores. The grey dotted lines show the average R^2 scores for the training datasets.

The proposed approach transformed the determination of reaction mechanisms into a minimization problem involving rate constants. This strategy is theoretically applicable to kinetic studies of all types of reactions. However, a key challenge arises from the dimension of the minimization problem (N_{step}). Since the minimization problem was solved stochastically by CMA-ES, the reproducible convergence could not be guaranteed,

especially in problems with large dimensions. According to ten independent calculations, the full model with 3 intermediates ($N_{step} = 20$) converged to solutions with similar MRSE values, whereas that with 5 intermediates ($N_{step} = 64$) showed some variability (Fig. S4a†). However, the sparse models with 3 intermediates ($N_{step} = 11$) did not necessarily converge to mechanisms with the same set of elementary steps (Fig. S4b†). Starting from the full model ($N_{step} = 20$), seven steps were consistently eliminated, five were retained with similar rate constant values (coefficient of variation of $\log_{10}(\mathbf{k} + 1) < 0.02$), five were consistently retained but exhibited varying values, and one of the remaining three was included in each model, although the specific step varied across the ten calculations. These imply that the proposed approach is readily applicable to the discovery of simple mechanisms, although the resulting sparse model was not always consistent across calculations. For example, in a catalytic organic reaction involving two substrates (A and B), a coupled product (A–B), a catalyst (C), and three intermediates (A–C, B–C, and A–B–C), the total number of possible elementary steps is 29 (Fig. S5†). Therefore, the dimensionality of the optimization problem is not a critical constraint in this case. However, for the discovery of more complex reaction networks, such as those found in combustion chemistry and astrochemistry,[41,42] reducing the dimension through chemical knowledge and other computational approaches is needed.

Conclusions

In this study, we present a data-driven approach for determining kinetic models of chemical reaction mechanisms *via* a sparse identification strategy. The challenge of applying the sparse identification approach in mechanistic studies, which involves the limited availability of concentration profiles for traceable chemical species, was addressed by generating simulated concentration profiles from ODEs uniquely generated from reaction mechanisms according to the law of mass action. Starting with comprehensive models generated mechanistically on the basis of the conservation of mass, this approach provides generalized models that prevent overfitting and minimal models that are useful for interpreting reaction mechanisms. The application to the autocatalytic reduction of manganese oxide ions revealed that the experimental data can be sufficiently represented by 11 elementary steps involving 8 chemical species. This strategy enables the discovery of reaction mechanisms without relying on heuristic kinetic models, as the only assumption required is the composition of the intermediates. The approach to extract kinetic models from limited experimental data expands the potential for automated mechanistic studies, especially in situations where comprehensive reaction monitoring is impractical.

Data availability

The data supporting this article have been included as part of the ESI.† Data for this article, including experimental data and Python scripts, are available on Zenodo at **https://doi.org/10.5281/zenodo.15259062** and on GitHub at **https://github.com/shun-hayashi/sparse-chem-react-mec**.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was financially supported by a Grant-in-Aid for Scientific Research (24K17563) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (MEXT).

References

1 J. I. Steinfeld, J. S. Francisco and W. L. Hase, *Chemical kinetics and dynamics*, Prentice Hall, 2nd edn, 1998.
2 D. G. Blackmond, Reaction progress kinetic analysis: A powerful methodology for mechanistic studies of complex catalytic reactions, *Angew. Chem., Int. Ed.*, 2005, **44**, 4302–4320.
3 D. G. Blackmond, Kinetic Profiling of Catalytic Organic Reactions as a Mechanistic Tool, *J. Am. Chem. Soc.*, 2015, **137**, 10852–10866.
4 A. Martínez-Carrión, M. G. Howlett, C. Alamillo-Ferrer, A. D. Clayton, R. A. Bourne, A. Codina, A. Vidal-Ferran, R. W. Adams and J. Burés, Kinetic Treatments for Catalyst Activation and Deactivation Processes based on Variable Time Normalization Analysis, *Angew. Chem., Int. Ed.*, 2019, **58**, 10189–10193.
5 C. D. T. Nielsen and J. Burés, Visual kinetic analysis, *Chem. Sci.*, 2019, **10**, 348–353.
6 G. E. Garrett and M. S. Taylor, A Nonlinear Ordinary Differential Equation for Generating Graphical Rate Equations from Concentration Versus Time Data, *Top. Catal.*, 2017, **60**, 554–563.
7 M. Á. de Carvalho Servia, I. O. Sandoval, K. K. Hii, K. Hellgardt, D. Zhang and E. Antonio del Rio Chanona, The automated discovery of kinetic rate models - methodological frameworks, *Digital Discovery*, 2024, **3**, 954–968.
8 F. Felsen, K. Reuter and C. Scheurer, A model-free sparse approximation approach to robust formal reaction kinetics, *Chem. Eng. J.*, 2022, **433**, 134121.
9 E. Groppo, S. Rojas-Buzo and S. Bordiga, The Role of In Situ/Operando IR Spectroscopy in Unraveling Adsorbate-Induced Structural Changes in Heterogeneous Catalysis, *Chem. Rev.*, 2023, **123**, 12135–12169.
10 A. Bédard, A. Adamo, K. C. Aroh, M. G. Russell, A. A. Bedermann, J. Torosian, B. Yue, K. F. Jensen and T. F. Jamison, Reconfigurable system for automated optimization of diverse chemical reactions, *Science*, 2018, **361**, 1220–1225.
11 B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick and A. I. Cooper, A mobile robotic chemist, *Nature*, 2020, **583**, 237–241.

12 Y. Shi, P. L. Prieto, T. Zepel, S. Grunert and J. E. Hein, Automated Experimentation Powers Data Science in Chemistry, *Acc. Chem. Res.*, 2021, **54**, 546–555.
13 C. J. Taylor, A. Pomberger, K. C. Felton, R. Grainger, M. Barecka, T. W. Chamberlain, R. A. Bourne, C. N. Johnson and A. A. Lapkin, A Brief Introduction to Chemical Reaction Optimization, *Chem. Rev.*, 2023, **123**, 3089–3126.
14 B. M. Matysiak, D. Thomas and L. Cronin, Reaction Kinetics using a Chemputable Framework for Data Collection and Analysis, *Angew. Chem., Int. Ed.*, 2024, **63**, 1–8.
15 M. Stamatakis and D. G. Vlachos, Unraveling the complexity of catalytic reactions via kinetic monte carlo simulation: Current status and frontiers, *ACS Catal.*, 2012, **2**, 2648–2663.
16 M. Pineda and M. Stamatakis, Kinetic Monte Carlo simulations for heterogeneous catalysis: Fundamentals, current status, and challenges, *J. Chem. Phys.*, 2022, **156**, 120902.
17 C. W. Gao, J. W. Allen, W. H. Green and R. H. West, Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms, *Comput. Phys. Commun.*, 2016, **203**, 212–225.
18 W. M. C. Sameera, S. Maeda and K. Morokuma, Computational Catalysis Using the Artificial Force Induced Reaction Method, *Acc. Chem. Res.*, 2016, **49**, 763–773.
19 C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green and K. F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning, *ACS Cent. Sci.*, 2017, **3**, 434–443.
20 M. Wen, E. W. C. Spotte-Smith, S. M. Blau, M. J. McDermott, A. S. Krishnapriyan and K. A. Persson, Chemical reaction networks and opportunities for machine learning, *Nat. Comput. Sci.*, 2023, **3**, 12–24.
21 J. F. Joung, M. H. Fong, J. Roh, Z. Tu, J. Bradshaw and C. W. Coley, Reproducing Reaction Mechanisms with Machine-Learning Models Trained on a Large-Scale Mechanistic Dataset, *Angew. Chem., Int. Ed.*, 2024, **63**, e202411296.
22 J. Burés and I. Larrosa, Organic reaction mechanism classification using machine learning, *Nature*, 2023, **613**, 689–695.
23 J. Bongard and H. Lipson, Automated reverse engineering of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9943–9948.
24 M. Schmidt and H. Lipson, Distilling Free-Form Natural Laws from Experimental Data, *Science*, 2009, **324**, 81–85.
25 S. L. Brunton, J. L. Proctor and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 3932–3937.
26 M. Hoffmann, C. Fröhner and F. Noé, Reactive SINDy: Discovering governing reactions from concentration data, *J. Chem. Phys.*, 2019, **150**, 025101.
27 E. Bezemer and S. C. Rutan, Multivariate curve resolution with non-linear fitting of kinetic profiles, *Chemom. Intell. Lab. Syst.*, 2001, **59**, 19–31.
28 N. Jiscoot, E. A. Uslamin and E. A. Pidko, Model-based evaluation and data requirements for parallel kinetic experimentation and data-driven reaction identification and optimization, *Digital Discovery*, 2023, **2**, 994–1005.
29 M. Á. d. C. Servia, K. K. Hii, K. Hellgardt, D. Zhang and E. A. d. R. Chanona, Simplest Mechanism Builder Algorithm (SiMBA): An Automated Microkinetic Model Discovery Tool, *arXiv*, 2024, preprint, arXiv:2410.21205, DOI: **10.48550/arXiv.2410.21205**.
30 C. Girard and H. B. Kagan, Nonlinear effects in asymmetric synthesis and stereoselective reactions: Ten years of investigation, *Angew. Chem., Int. Ed.*, 1998, **37**, 2922–2959.
31 R. F. Ludlow and S. Otto, Systems chemistry, *Chem. Soc. Rev.*, 2008, **37**, 101–108.
32 A. J. Bissette and S. P. Fletcher, Mechanisms of autocatalysis, *Angew. Chem., Int. Ed.*, 2013, **52**, 12800–12826.
33 K. Ruiz-Mirazo, C. Briones and A. De La Escosura, Prebiotic systems chemistry: New perspectives for the origins of life, *Chem. Rev.*, 2014, **114**, 285–366.
34 A. I. Hanopolskyi, V. A. Smaliak, A. I. Novichkov and S. N. Semenov, Autocatalysis: Kinetics, Mechanisms and Design, *ChemSystemsChem*, 2021, **3**, e2000026.
35 H. F. Launer, The kinetics of the reaction between potassium permanganate and oxalic acid. I, *J. Am. Chem. Soc.*, 1932, **54**, 2597–2610.
36 H. F. Launer and D. M. Yost, The Kinetics of the Reaction between Potassium Permanganate and Oxalic Acid. II, *J. Am. Chem. Soc.*, 1934, **56**, 2571–2577.
37 S. J. Adler and R. M. Noyes, The Mechanism of the Permanganate-Oxalate Reaction, *J. Am. Chem. Soc.*, 1955, **77**, 2036–2042.
38 R. Shrinkage, Regression Shrinkage and Selection via the Lasso, *J. Roy. Stat. Soc. B*, 1996, **58**, 267–288.
39 N. Hansen and A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.*, 2001, **9**, 159–195.
40 I. Loshchilov and F. Hutter, CMA-ES for Hyperparameter Optimization of Deep Neural Networks, *arXiv*, 2016, preprint, arXiv:1604.07269, DOI: **10.48550/arXiv.1604.07269**.
41 T. Lu and C. K. Law, Toward accommodating realistic fuel chemistry in large-scale computations, *Prog. Energy Combust. Sci.*, 2009, **35**, 192–215.
42 H. M. Cuppen, L. J. Karssemeijer and T. Lamberts, The kinetic Monte Carlo method as a way to solve the master equation for interstellar grain chemistry, *Chem. Rev.*, 2013, **113**, 8840–8871.