ChemComm

ROYAL SOCIETY OF CHEMISTRY

FEATURE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2026, **62**, 9536

Discovering new materials knowledge from "old data"

Yuhang Wang, [ab] Qian Wang, [a] Seong-Hoon Jang, [a] Eric Jianfeng Cheng[a] and Hao Li *[a]

From "old data" to new knowledge discovery, this paradigm is fundamentally reshaping research in chemistry and materials. Unlike traditional trial-and-error approaches, knowledge mining driven by large-scale databases offers unprecedented potential in exploring complex compositional spaces and accelerating rational materials design. In this review, we highlight three significant progresses in discovering new materials knowledge from "old literature data": (1) In the field of catalysis, data-driven approaches reveal new phenomena and limitations of existing theoretical models, greatly accelerating materials design and screening. (2) In the field of solid-state electrolytes, data empowerment accelerates the understanding of underlying physical mechanisms. (3) In the field of hydrogen storage, we demonstrate a pathway from "old data" to structured knowledge and finally to autonomous design. Finally, we highlight the critical role of database construction in data intelligence and the development of AI agents for materials design. Looking ahead, such data-driven models will continue to deepen our knowledge generation and accelerate the discovery of target materials in the relevant field. By integrating knowledge generation from "old data", theoretical simulations, and experimental validation, this approach promises to establish a digital materials ecosystem for cross-disciplinary innovation, where materials discovery will be continuously accelerated.

Received 22nd March 2026,
Accepted 20th April 2026

DOI: 10.1039/d6cc01716a

rsc.li/chemcomm

1. Introduction

Over the past decade, extensive experimental and theoretical data have been accumulated around key reactions relevant to energy conversion and storage. In electrocatalysis, systems such as the CO_2 reduction reaction (CO_2RR),[1–5] oxygen reduction reaction (ORR),[6–11] and nitrate reduction reaction (NO_3RR)[12–16] have generated a wealth of reported metrics, including activity, selectivity, and adsorption energies of reaction intermediates. Similarly, in the fields of hydrogen storage and the development of solid-state electrolytes (SSEs),[17,18] extensive experimental and computational datasets have also been accumulated, encompassing structural stability, ion diffusion barriers, hydrogen storage capacity, and other key physicochemical parameters. Most of these data were originally

[a] *Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai, 980-8577, Japan. E-mail: li.hao.b8@tohoku.ac.jp*
[b] *College of Chemistry and Pharmaceutical Engineering, Nanyang Normal University, Nanyang 473061, PR China*

Yuhang Wang

Yuhang Wang is a Visiting Scientist at the Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Japan. He received his PhD in Materials Science and Engineering at City University of Hong Kong in 2025. His research interests include first-principles calculations, electrocatalysis, machine learning, and AI for catalysis.

Qian Wang

Qian Wang is a Specially-appointed Assistant Professor at the Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Japan. She received her PhD in Advanced Energy Materials and Devices at Sichuan University in 2024. Her research interests include solid electrolyte, data science, machine learning, and AI for battery.

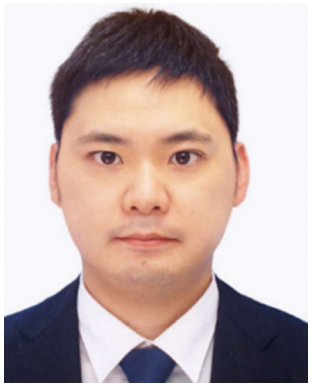
collected to address specific questions, such as testing mechanistic hypotheses, optimizing material performance, or rationalizing particular experimental observations. However, when viewed beyond the scope of individual studies and reanalyzed from a broader, cross-system perspective, their full potential has yet to be realized.[19,20] In this sense, existing datasets should not be viewed as outdated information, but rather as valuable knowledge resources that have yet to be fully explored.

Why is it valuable to revisit existing datasets? Materials research is inherently cumulative. Empirical trends, adsorption-energy scaling relationships,[21] and interfacial effects identified in one system often have relevance beyond a single reaction or material class. Yet, under conventional research paradigms, data are typically analyzed within the boundaries of a specific material and set of conditions. Limited cross-comparison and integration mean that potentially generalizable patterns across systems can remain unnoticed. As datasets continue to expand, the conclusions drawn from individual studies are increasingly insufficient to represent broader trends. Reorganizing and reanalyzing "old literature data" therefore serves not only to test the limits of established theories, but also to uncover previously unrecognized structure–property relationships and mechanistic insights.

Why has it become possible to extract new insights from "old data" at this stage? ("old data" primarily refers to published experimental/computational data.) A key reason lies in the availability of digital materials databases, which enable systematic integration of data from multiple publications and across different material systems.[22] The establishment of structured databases provides a unified foundation for large-scale statistical analysis and pattern recognition.[23–26] At the same time, tools such as machine learning, multiscale modelling, and interfacial physics-based models offer new interpretative frameworks through which data can be reexamined.[27–31] Equally important is the ongoing shift in the research paradigm from hypothesis testing towards data reinterpretation and knowledge discovery, especially when we reexamine vast amounts of legacy data from a broader perspective—things start to look different. Observations that might have once been dismissed as isolated incidents or specific phenomena can now be seen in a broader context. When re-evaluated, they often reveal deeper, cross-system knowledge that was previously hidden.

Against this backdrop, focusing on catalysis, hydrogen storage, and SSEs provides a representative and balanced perspective for discovering new knowledge from reported data. These areas occupy central positions in the energy transition and have accumulated extensive experimental and computational datasets over the past decade, covering a wide range of material systems. Moreover, each field has developed relatively systematic database resources,[23,32,33] creating favorable conditions for cross-system comparison and abstraction of general principles. By taking these three areas as our entry points, we can systematically illustrate how to uncover new knowledge from existing data and how to leverage extant databases to accelerate materials design.

Based on these considerations, this review discusses representative studies in the field of catalysis, SSEs, and hydrogen storage, with a particular focus on the new insights that emerge



Seong-Hoon Jang

Seong-Hoon Jang is an Associate Professor at the Unprecedented-Scale Data Analytics Center, Tohoku University. He received his PhD in Applied Physics from the University of Tokyo in 2020 and completed his postdoctoral training at the National Institute for Materials Science before joining Tohoku University in 2023. His research focuses on applied mathematics: higher-order perturbation theory in strongly correlated quantum spin systems, molecular dynamics simulations for energy materials, general white-box symbolic regression modeling, and the development of massively parallel programming techniques for applications in physics and chemistry.

Eric Jianfeng Cheng

Eric Jianfeng Cheng is an Associate Professor at the Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Japan. He earned his PhDs in Materials Engineering from Tohoku University and Energy Chemistry from Kyoto University, respectively. He pioneered the direct observation of Li dendrite growth in garnet solid electrolytes. His current research interests include solid-state batteries, ceramic science and processing, energy materials design, and AI for batteries.

Hao Li

Hao Li is a Distinguished Professor at the Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Japan. His research focuses on developing AI, materials theory, and autonomous experimentation for closed-loop materials design. He proposed the concept of the digital materials ecosystem and serves as the founding Editor-in-Chief of the journal AI Agent.

from the reorganization and reanalysis of "old data". We focus on the new knowledge that was not fully recognized in the initial research but becomes clear when the data is examined from a broader cross-system perspective. In the field of catalysis, we have conducted a systematic review of the latest research progress in new phenomenon discovery, new mechanistic understanding, and new material design based on old-data-mining-driven methods; in the field of SSEs and hydrogen storage materials, we have emphasized how "old data" lay the foundation for the generation of new chemical insights and the acceleration of material discovery. Finally, we have outlined the potential role of database construction in establishing a sustainable and self-optimizing digital materials ecosystem. In the future, the digital ecosystem combined with artificial intelligence (AI) agents will significantly accelerate the transformation of material design research in fields such as catalytic conversion and energy storage and beyond.

2. Discovery of new chemical knowledge in catalysis

Catalytic transformations are central to a variety of critical energy-conversion processes. Over the past decade, this field has developed rapidly, generating a substantial body of experimental data and materials information. Nevertheless, most studies have focused on specific material systems, emphasizing local structural regulation or individual mechanistic interpretations. As a result, the overall picture remains largely case-specific, and broadly applicable principles are still limited. In recent years, the emergence of data-driven approaches has opened new possibilities for systematically integrating and reanalyzing existing datasets. Such efforts have begun to uncover phenomena and mechanistic insights that were not fully recognized in limited publications. Simultaneously, they provide a more forward-looking route for the efficient screening and rational design of new catalytic materials. In the following section, wefocus on the discovery of new knowledge from existing data in various aspects.

2.1. Discovering new phenomena from "old data"

2.1.1. pH-dependent effect. With the establishment of digital catalysis databases, such systematic reanalysis has become feasible. For example, the Digital Catalysis Platform (DigCat: https://www.digcat.org/)[34,35] has integrated more than 900 000 experimental and computational data entries, providing a quantitative foundation for cross-system statistical comparisons. In the field of carbon reduction, electrocatalytic CO_2 reduction to formic acid has attracted significant research interest due to its high economic feasibility. Wang *et al.*[36] systematically summarized the experimental CO_2RR performance of 2348 reported catalysts *via* a large-scale data mining approach focused on tin-based catalysts (Fig. 1a and b). Interestingly, when plotting faradaic efficiency (FE) against pH, a clear trend emerges: Sn-based catalysts exhibit overall higher FE in higher pH ranges. In the ORR system (Fig. 1c), Zhang *et al.*[37] further demonstrated that M–N–C catalysts exhibit distinct pH-dependent behaviors. Certain Fe single-atom catalysts show limited sensitivity to pH variation, whereas Fe-pyrrole-N_4 outperforms Fe-pyridine-N_4 under acidic conditions. In contrast, Co-centered catalysts tend to favor H_2O_2 formation in acidic media.

The widely observed pH dependence cannot be reconciled within the conventional computational hydrogen electrode (CHE) framework based on static free-energy analysis. Incorporating interfacial electric field effects into free-energy and microkinetic models resolves this discrepancy: even at constant U_{RHE} (RHE: reversible hydrogen electrode) the field varies with pH, modulating intermediate stability and shifting the volcano peak.[38] Under the research framework of pH-electric field coupling, Wang *et al.* revealed the structural sensitivity of Sn-based catalysts and accordingly proposed differentiated performance optimization strategies for different types of catalysts.[39] Similar electric field-driven pH effects were reported for ORR, including the "acid trap" and catalyst-dependent responses.[37]

Taken together, pH-dependent behavior should be viewed not as an isolated feature, but as a recurring trend revealed by cross-system data integration. This underscores the untapped value of existing data and enables a deeper understanding of electrocatalysis under practical conditions.

2.1.2. Magnetic field effects. Data mining of existing catalyst performance in ammonia synthesis reveals an underlying magnetic effect that governs promotion behavior. In recent years, a variety of emerging catalytic systems and promoters have been reported, including Ba/Ca-promoted Co, Fe, and Ru catalysts, electrides and hydride/amide systems.[40–42] Most studies attribute promotion to local structural or electrostatic effects within individual systems. Yet, when activity data are reanalyzed under unified criteria, a consistent trend emerges: diverse promoters induce comparable activity enhancements across different catalyst families (Fig. 2a). This cross-system similarity suggests a more general physicochemical origin underlying seemingly distinct promotional effects. Cao *et al.*[42] further compared magnetic Co with Ru catalysts and reached a more specific conclusion. On Co, promoters such as Li, Ba, Ca, and La display activities well beyond conventional electrostatic predictions, giving rise to a clear "excess promotion". In contrast, the behavior of Ru can be largely described within an electrostatic framework. Electronic structure analysis links the anomaly on Co to spin polarization: strong spin polarization induces d-band splitting and weakens coupling with the N–N transition state, whereas reduced polarization lowers the barrier and enhances kinetics. This case shows that systematic reanalysis of existing data can reveal spin-mediated promotion effects in magnetic catalysts, adding a new dimension beyond conventional promoter theories.

Data-driven analyses further indicate that magnetic effects are not isolated observations but are broadly present in electrocatalytic systems. Fig. 2b schematically illustrates the role of magnetic fields in electrochemical reactions. A systematic survey of the literature over the past few years shows that applying an external magnetic field can markedly influence catalytic performance by modulating the spin state of active sites. Based on this data-driven insight, You *et al.*[43] proposed tuning spin states with an external magnetic field to enhance electrochemical NO_3RR (Fig. 2c). Spin-state modulation promotes *NO hydrogenation to

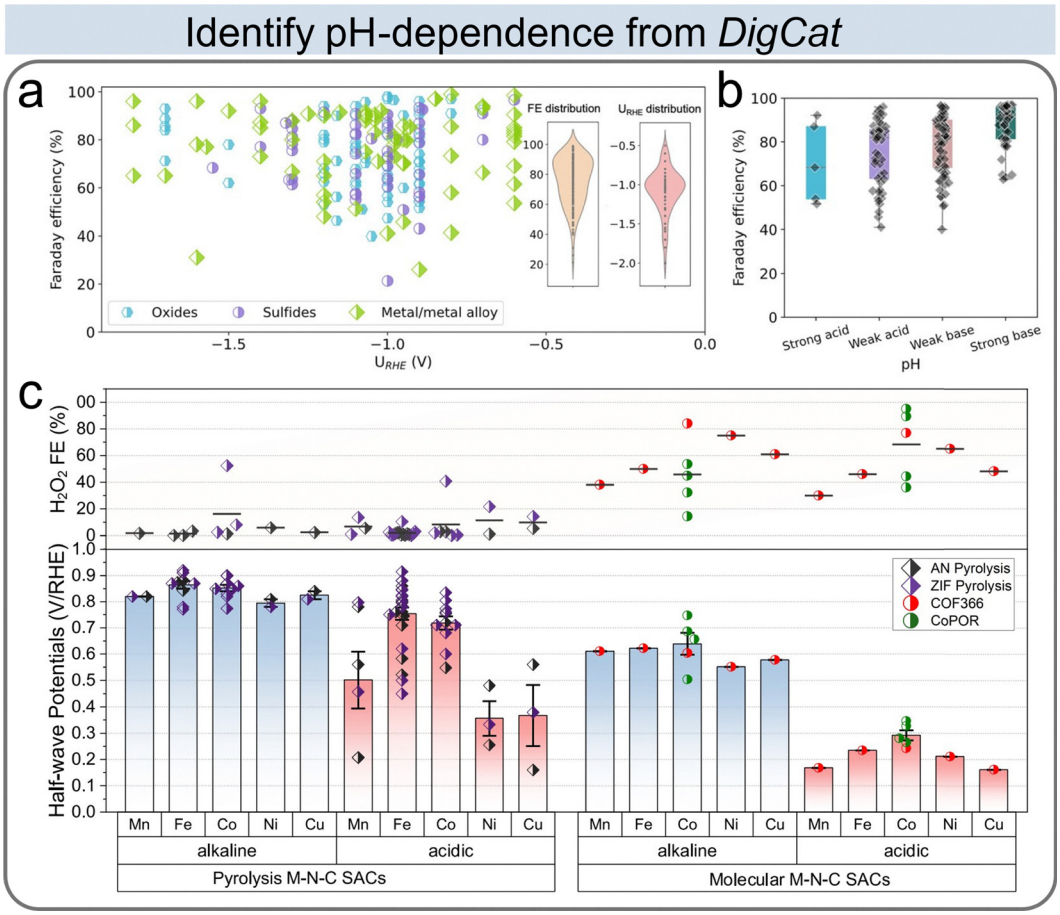


Fig. 1 (a) Summary of CO_2RR performance of Sn-based catalysts, including Sn oxides, sulfides, and Sn-based metals/alloys. For these Sn-based catalysts, the insets present violin plots of the FEs and potentials associated with HCOOH formation. (b) Box plot of maximum FEs at different pH values. Reproduced with permission from ref. 36 licensed under a Creative Commons License CC BY-NC 4.0. (c) ORR activity analysis of over 100 M–N–C catalysts sourced from the DigCat database, with an overview of representative M–N–C catalyst systems. Reproduced with permission from ref. 37 licensed under a Creative Commons License CC BY-NC 4.0.

*NHO and weakens excessive $*NH_2$ adsorption, accelerating key reaction steps. Under a magnetic field, the NH_3 yield reaches $\sim$38 mg L^{-1} h^{-1} with $\sim$95% FE and remains stable for over 200 h, markedly outperforming the field-free case.

2.1.3. Trade-off between ethylene oxide selectivity and partial oxidation activity. As a bulk chemical produced on the scale of over ten million tons per year, ethylene oxide (EO) is widely used in the pharmaceutical, personal care, and materials industries.[44] Its large-scale manufacture relies predominantly on the ethylene epoxidation route, rendering improvements in EO selectivity crucial for enhancing both resource efficiency and economic viability. Through large-scale data mining, Li *et al.*[45] identified a broadly observed experimental trend: EO selectivity decreases with increasing conversion (Fig. 3a). Although Ag-based catalyzed ethylene epoxidation has been investigated for decades, its atomic-scale origin remained unclear. As shown in Fig. 3b, theoretical analysis attributes Ag's superior performance to its unique surface electronic properties: moderate oxygen binding coupled with relatively weak carbon binding. This balance allows efficient O_2 activation while preventing strong adsorption of carbonaceous intermediates, thereby suppressing acetaldehyde formation and combustion. Such a balance between activity and selectivity explains why Ag outperforms other group IB metals, including Cu and Au. This data-driven analysis thus uncovers a fundamental trade-off between ethylene oxide selectivity and partial oxidation activity, revealing a constraint implicit in existing experimental data.

2.1.4. Limitations of transition metal oxides for ORR. At present, the most active and stable electrocatalysts for the ORR remain based on platinum-group metals. Although transition metal oxides (TMOs) exhibit favorable operational stability under oxidative conditions and are therefore considered promising alternatives, their intrinsic activity generally falls short of that of Pt.[49] Li *et al.*[46] conducted a large-scale high-throughput experimental screening of 7798 Mn-, Ni-, and Fe-based TMO compositions for alkaline ORR. Their results reveal a clear performance limitation for TMOs (Fig. 3c): even after optimization, the geometric current density remains about three times lower than that of Pt, with the gap becoming more pronounced when intrinsic activity is considered. Literature comparisons further indicate that this limitation persists in both acidic and alkaline media, preventing TMOs from matching Pt at moderate overpotentials.

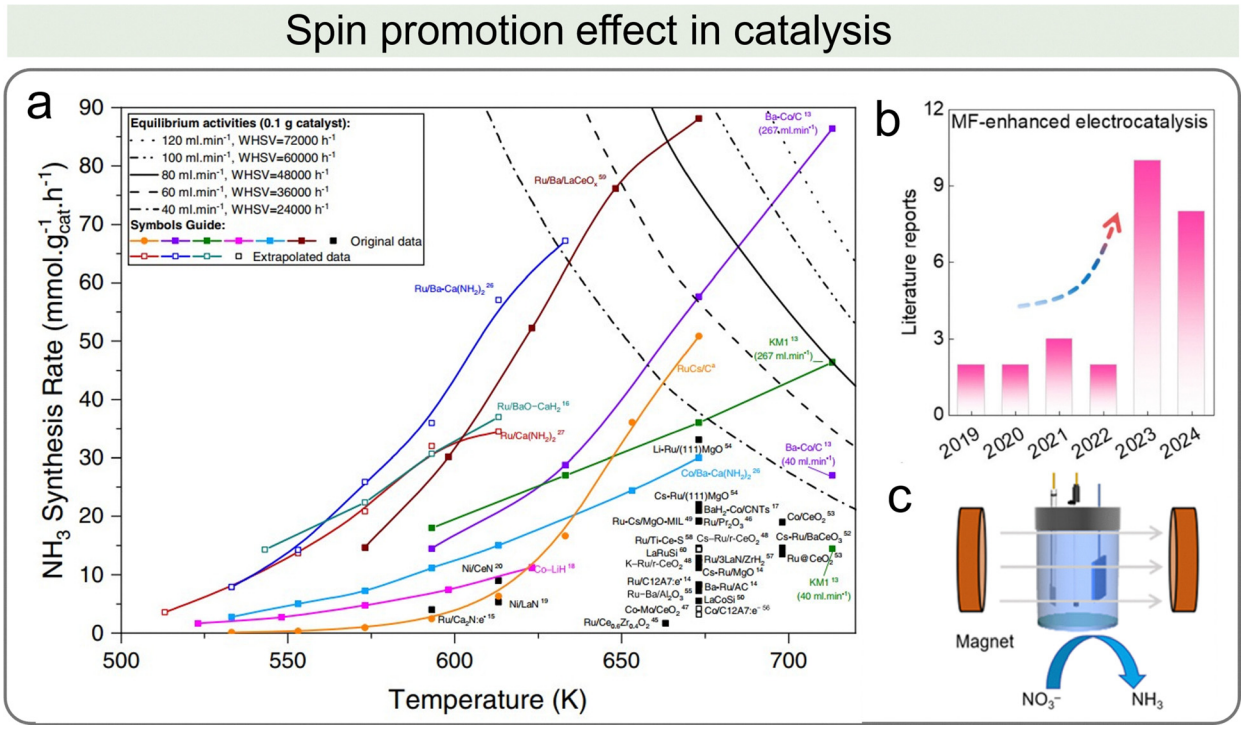


Fig. 2 (a) Data-mined summary of previously reported experimental activities for ammonia synthesis. Reproduced with permission from ref. 42 licensed under a Creative Commons License CC BY 4.0. (b) Overview of data mining results on magnetic-field-enhanced electrocatalysis. (c) Experimental illustration setup for NO_3RR under a constant external magnetic field. Reproduced with permission from ref. 43 licensed under a Creative Commons License CC BY-NC-ND 4.0.

Mechanistically, weak oxygen binding renders O–O bond cleavage rate-determining, while stronger interfacial electric field effects further hinder dissociation. Large-scale data analysis supports this atomistic explanation for the intrinsic ORR activity bottleneck of TMOs.

2.1.5. Dual-atom electrocatalysts for CO_2RR to CO. Because dual atom catalysts (DACs) feature adjacent bimetal sites that can increase surface *CO coverage, C–C coupling would seem plausible.[50,51] However, after systematically analyzing the experimental CO_2RR data of 11 typical M–N–C DACs and combining it with DFT calculations as well as surface Pourbaix analysis, Yang *et al.*[43] reached an unexpected conclusion: despite the presence of adjacent bimetal sites in DACs, CO remains the dominant product in nearly all reported studies; even Cu-containing DACs fail to effectively generate C–C coupling products (Fig. 3d). Theoretical simulations further show that under reaction conditions, the stable surface of DACs is precovered by CO at the metal–metal bridge sites. Meanwhile, the hydrogenation of CO to *COH or *CHO is thermodynamically unfavorable, rendering C–C coupling a high-barrier process. In summary, both data mining and computational results indicate that the fundamental reasons why M–N–C DACs struggle to achieve C–C coupling are: strong *CO adsorption causing site poisoning, and the difficulty of *CO hydrogenation.

2.1.6. Dopant diversity *vs.* HER invariance in Cu-based SAAs for CO_2RR. Electrochemical CO_2RR toward multicarbon (C_{2+}) products represents a promising route to carbon neutrality. However, pure Cu catalysts are intrinsically limited by sluggish C–C coupling kinetics, resulting in unsatisfactory performance. Cu-based single-atom alloys (SAAs) have recently emerged as a potential strategy to overcome this bottleneck.

By statistically analyzing ~50 published studies and >80 data points collected on the DigCat platform, Wang *et al.*[48] found that incorporating 29 different elements into Cu-based SAAs consistently enhances C_{2+} selectivity, with an average FE exceeding 50%. This statistical analysis reveals a general trend: single-atom doping in Cu-based alloys broadly promotes C_{2+} selectivity, demonstrating the wide applicability of the SAA strategy for enhancing multicarbon formation in CO_2RR (Fig. 3e). Further data mining from the DigCat dataset shows that despite the pronounced differences in C_{2+} selectivity among various doped systems, the FE for the competing HER remains largely invariant across different dopant types (Fig. 3f). This observation is noteworthy. It suggests that the primary role of single-atom doping is not the suppression of HER, but rather the targeted promotion of C–C coupling kinetics. In other words, the improvement in C_{2+} selectivity originates predominantly from acceleration of the desired reaction pathway, rather than from inhibition of the competing side reaction. This insight provides a more precise target for catalyst optimization for Cu-based alloys in CO_2RR-to-C_{2+}.

Taken together, a comprehensive, systematic reanalysis of extensive existing datasets allows us to transcend the limitations inherent in individual studies, revealing pervasive trends across a broader chemical space. Critically, these data-driven observations provide direction for subsequent mechanistic investigations, such as pH-dependent microkinetic modeling

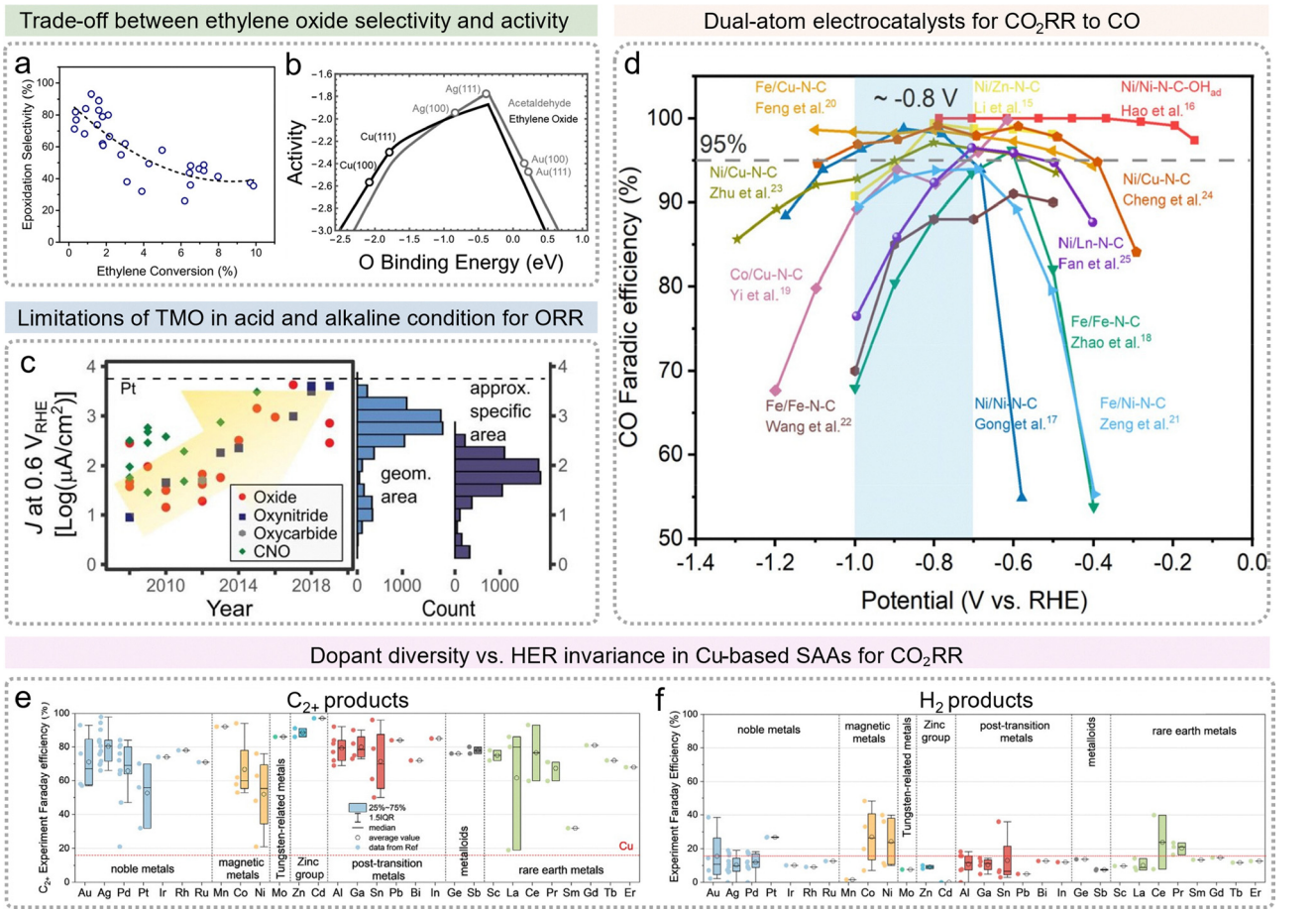


Fig. 3 (a) Epoxide selectivity as a function of ethylene conversion over unpromoted Ag-based catalysts, summarized from representative experimental studies. (b) Volcano activity model of ethylene partial oxidation as a function of O binding energy. Reproduced with permission from ref. 45 licensed under a Creative Commons License CC BY-NC-ND 4.0. (c) Current densities of TMOs measured at 0.6 V RHE^{-1}, categorized by host anion. The full high-throughput dataset is displayed, showing geometric current densities (blue) and estimated specific current densities (purple). Reproduced with permission from ref. 46 Copyright 2021, Springer Nature. (d) Comparison of CO FE for various DACs reported experimentally. Reproduced with permission from ref. 47 Copyright 2023, American Chemistry Society. Experimental faradaic efficiencies for C_2^+ products (e) and HER (f) from CO_2RR over Cu-based single-atom alloys (SAAs), summarized from the DigCat Platform. Reproduced with permission from ref. 48 licensed under a Creative Commons License CC BY 3.0.

or the proposal of mechanisms involving magnetic field effects on activity. The phenomena uncovered through data mining steer the development of electrocatalysts away from empirical trial-and-error and toward mechanism-guided, precise modulation. This approach establishes a scientifically grounded pathway for the targeted design of high-performance catalytic systems.

2.2. Identifying new mechanisms from "old data"

The value of large-scale data mining extends beyond summarizing empirical trends to probing the boundaries of existing theory. When diverse datasets are integrated, systematic deviations often emerge. Rather than treating these as noise, quantitative analysis can expose limitations in current descriptors and assumptions, prompting the revision of mechanistic hypotheses.

2.2.1. Applicability limitation of the single G_H descriptor in HER. In studies of the HER, the hydrogen adsorption free energy (G_H) has long been regarded as a key descriptor, successfully rationalizing the volcano relationship observed on noble metal surfaces. However, when the focus shifts to M–N–C SACs, the limitations of relying solely on G_H become apparent. A systematic integration of the representative experimental data has revealed a more nuanced reality. Materials with near-zero G_{*H} values often fail to deliver optimal activity, while some systems deviating from the ideal adsorption range show unexpectedly high performance (Fig. 4a).[52]

Further analysis reveals that traditional models overlook factors such as HO* poisoning, actual surface coverage, and the involvement of non-metal sites. For instance, V-pyrrole-N_4 at pH = 0 exhibits significant HO* poisoning (Fig. 4b). By introducing G_{HO} as a complementary descriptor and incorporating an electric-field–corrected microkinetic model, experimental observations can be more consistently rationalized. Under −0.5 V RHE^{-1} (Fig. 4c), considering HO* poisoning leads to a substantial decrease in the predicted activity of M-pyrrole-N_4 (M = Sc, Ti, V). This helps explain the limited experimental activity of systems like V–N–C, despite their "ideal" G_H values. This example illustrates that mining existing data can clarify the theoretical limitations of conventional descriptors and advance mechanistic understanding of HER.

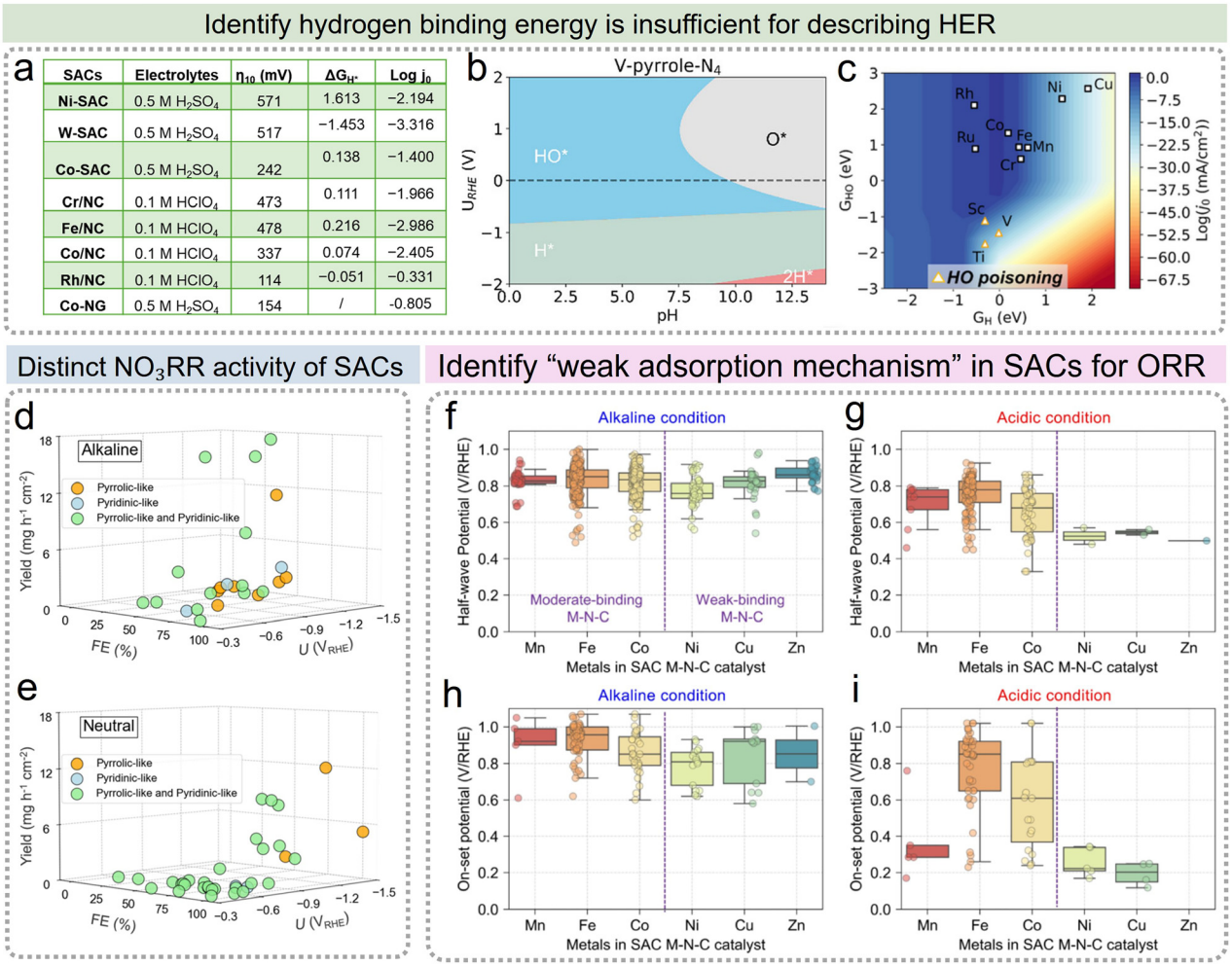


Fig. 4 (a) HER activity comparison of M–N–C catalysts documented in existing literature. Data on the experimental HER performance of typical M–N_4–C SACs, mined from available experimental literature published over the past decade are also included in DigCat. (b) Surface pourbaix diagram of V-pyrrole-N_4 as a function of pH and potential. (c) 2D HER microkinetic volcano plotted against (G_H, G_{HO}) at −0.5 V RHE^{-1} (pH = 0), with M-pyrrole-N_4 sites highlighted. Reproduced with permission from ref. 52 Copyright 2025, Wiley-VCH. Analysis of reported experimental performance of NO_3RR on >60 M–N–C catalysts: (d) alkaline conditions and (e) neutral conditions. Reproduced with permission[53] under the terms of the CC-BY-NC-ND 4.0. Summary of ORR performance for M–N–C catalysts, with half-wave potentials measured under alkaline (f) and acidic (g) conditions. Onset potentials under alkaline (h) and acidic (i) conditions, with the onset current density typically defined as 0.5 mA cm^{-2}. Reproduced with permission from ref. 54 licensed under a Creative Commons License CC-BY-NC-ND 4.0.

2.2.2. Coordination-dependent activity of single-atom catalysts in NO_3RR. In the context of sustainable ammonia production, NO_3RR has emerged as a promising route for both pollution mitigation and resource conversion. By systematically analyzing experimental data from over 60 M–N–C catalysts, Jiang *et al.*[53] identified a trend that had not been emphasized previously. Specifically, M–N–pyrrolic sites consistently exhibit higher ammonia production activity than M–N–pyridinic counterparts (Fig. 4d and e). While this distinction may not be apparent in individual studies, it becomes clear when comparing results across multiple reports, revealing a consistent structure–activity differentiation. However, it is worth noting that the classical thermodynamic-based limiting-potential model is insufficient to accurately capture the rate-determining step and the performance trends on M–N–pyrrolic and M–N–pyridinic catalysts. To address this limitation, Jiang *et al.*[53] developed a pH-dependent microkinetic model, which successfully revealed that the adsorption and protonation of nitrate constitute the rate-determining step in the NO_3RR. In fact, this step has often been overlooked or assumed to occur concomitantly in previous studies. These new insights, driven by mining existing experimental data, not only deepen the understanding of the NO_3RR reaction mechanism but also establish fundamental design principles for electrocatalytic ammonia synthesis.

2.2.3. Identifying "weak adsorption mechanism" in SACs for ORR. In studies of the ORR, M–N–C SACs typically follow the Sabatier principle and volcano plot predictions. Systems with moderate adsorption strength are considered optimal. As shown in Fig. 4f–i, Zhang *et al.*[54] analyzed 1018 experimental data points for M–N–C catalysts collected through extensive

data mining. The dataset shows that moderate-binding catalysts, such as Fe/Co–N–C, exhibit outstanding ORR activity in proton exchange membrane fuel cells, consistent with the Sabatier principle and volcano-type behavior. In contrast, weak-binding systems including Ni–N–C, Cu–N–C, and Zn–N–C display unexpectedly high activity in alkaline media and pronounced pH-dependent performance. These contrasting trends highlight a gap in the current understanding of the reaction pathway of weak-binding systems in ORR. Further investigation reveals that in weak-binding SACs such as Ni/Cu–N–C, atomic oxygen adsorption at the M–N bridge site induces significant changes in adsorption scaling relations, electric field responsiveness, and solvation effects. This distinguishes their reaction pathway from that of moderate-binding SACs like Fe/Co–N–C. The pH-electric field coupled model emphasizes that shifts in dipole moment orientation mitigate pH dependence and lower the kinetic barriers for the ORR. This effect is particularly pronounced for O* adsorption, thereby enhancing catalytic efficiency under both acidic and alkaline conditions.

Taken together, cross-system data integration helps clarify the limits of conventional descriptors and reveals potential missing mechanistic features. Whether it is the limitations of G_H in HER or the emergence of new pathways in NO_3RR and ORR, these examples demonstrate that revisiting existing data can drive theoretical refinement and deepen mechanistic understanding, rather than merely serving as a tool for empirical summary.

2.3. Designing new materials from "old data"

Beyond uncovering new mechanisms, data-driven approaches directly enable rational materials design. By integrating large-scale datasets, key structure–property relationships and stability limitations can be identified, allowing forward screening and targeted validation. Compared with intuition-based strategies, data-driven materials design narrows the search space and provides a more predictive route for materials development.

2.3.1. Data-driven screening of non-noble metal oxides catalysts. Non-precious metal oxides (MOs) catalysts have attracted considerable attention in electrocatalysis due to their low cost. However, the diversity of their potential structures presents a challenge for the rational design of high-activity catalysts. Data mining from materials databases has emerged as an effective strategy to identify unexplored catalysts. For example, Jia *et al.*[55] established a high-throughput screening workflow based on the Materials Project (Fig. 5a), identifying 1159 bulk-stable metal oxides from 19 737 candidates (Fig. 5b), including 38 unary, 470 binary, 621 ternary, and 30 quaternary compounds (Fig. 5c). Analysis of aqueous stability under different pH conditions (Fig. 5d) shows that acidic media stabilize only a few materials for HER and nitrogen reduction reaction (NRR), whereas ORR, oxygen evolution reaction (OER), and chlorine evolution reaction (CER) remain moderately stable. Alkaline conditions enhance stability for ORR, OER, and HER but reduce it for CER, while neutral media support the largest number of stable compounds. Encouragingly, data-driven screening identified Sb_2WO_6 as stable for acidic ORR, which was experimentally confirmed.

In addition, based on the data integration provided by the DigCat platform, Jia *et al.*[56] proposed a closed-loop design strategy for discovering non-noble metal electrocatalysts for water splitting. The workflow (Fig. 5e) links database screening, experimental validation, and advanced characterization through iterative feedback. Acid-stable oxides for OER and HER were first selected *via* aqueous stability analysis, followed by surface-state evaluation and microkinetic modeling to assess activity and possible transformations. This integrated framework of data mining, theoretical calculation, and experimental validation proved to be a powerful strategy, leading directly to the identification of $RbSbWO_6$ as an efficient bifunctional catalyst for acidic water splitting. It demonstrates how data-driven screening, when guided by mechanistic insight, can accelerate the discovery of high-performance materials.

2.3.2. High-accuracy computational reconstruction of aqueous phase stability. The advent of large materials databases, such as the materials project, has enabled high-throughput screening of materials based on their aqueous stability.[57,58] While Pourbaix diagrams based on PBE functionals are widely used, they can introduce systematic errors. Wang *et al.*[59] improved this framework by recalculating formation energies with SCAN(+*U*). As shown in Fig. 5f, SCAN(+*U*) reduces the mean absolute error (MAE) for 114 binary oxides to 0.072 eV per atom, outperforming PBE(+*U*) (0.182 eV per atom) and SCAN (0.094 eV per atom). For TMOs, SCAN+*U* further lowers the MAE to 0.025 eV per atom. Reassessment of MoS_2, MoP, CoP, and Ni_3Mo identifies open-circuit potential corrosion as a key instability source, underscoring the value of recalibrating existing data.

2.3.3. Universal machine-learning potential-driven structure exploration. Recently, universal machine-learning interatomic potentials (uMLIPs) trained on large materials databases have enabled a new paradigm for crystal structure prediction. Compared with conventional global search strategies based on density functional theory (DFT) calculations, which are computationally intensive for complex systems, An *et al.*[60] trained uMLIPs such as M3GNet using data from the Materials Project. These models were coupled with structure prediction algorithms to accelerate searches in quaternary systems including Sr–Li–Al–O and Ba–Y–Al–O. This approach successfully rediscovered known compounds absent from the training set and identified new candidates with favorable stability. As shown in Fig. 5g, Ba_4YAlO_7 was predicted to adopt a triclinic P1 structure with a formation energy of 28 meV per atom. Phonon calculations confirmed its dynamical stability (Fig. 5h). These results illustrate how database-trained uMLIPs can markedly reduce computational cost and promote predictive, data-driven materials discovery.

In summary, data-driven strategies leveraging existing databases are accelerating materials design. For example, high-throughput screening rapidly identifies promising candidates, while universal machine learning potentials trained on open-source data substantially reduce the cost of structure prediction for complex systems. These trends underscore that reanalysis of existing data is driving a paradigm shift in materials design, moving it from empirical trial-and-error toward a more precise and efficient approach.

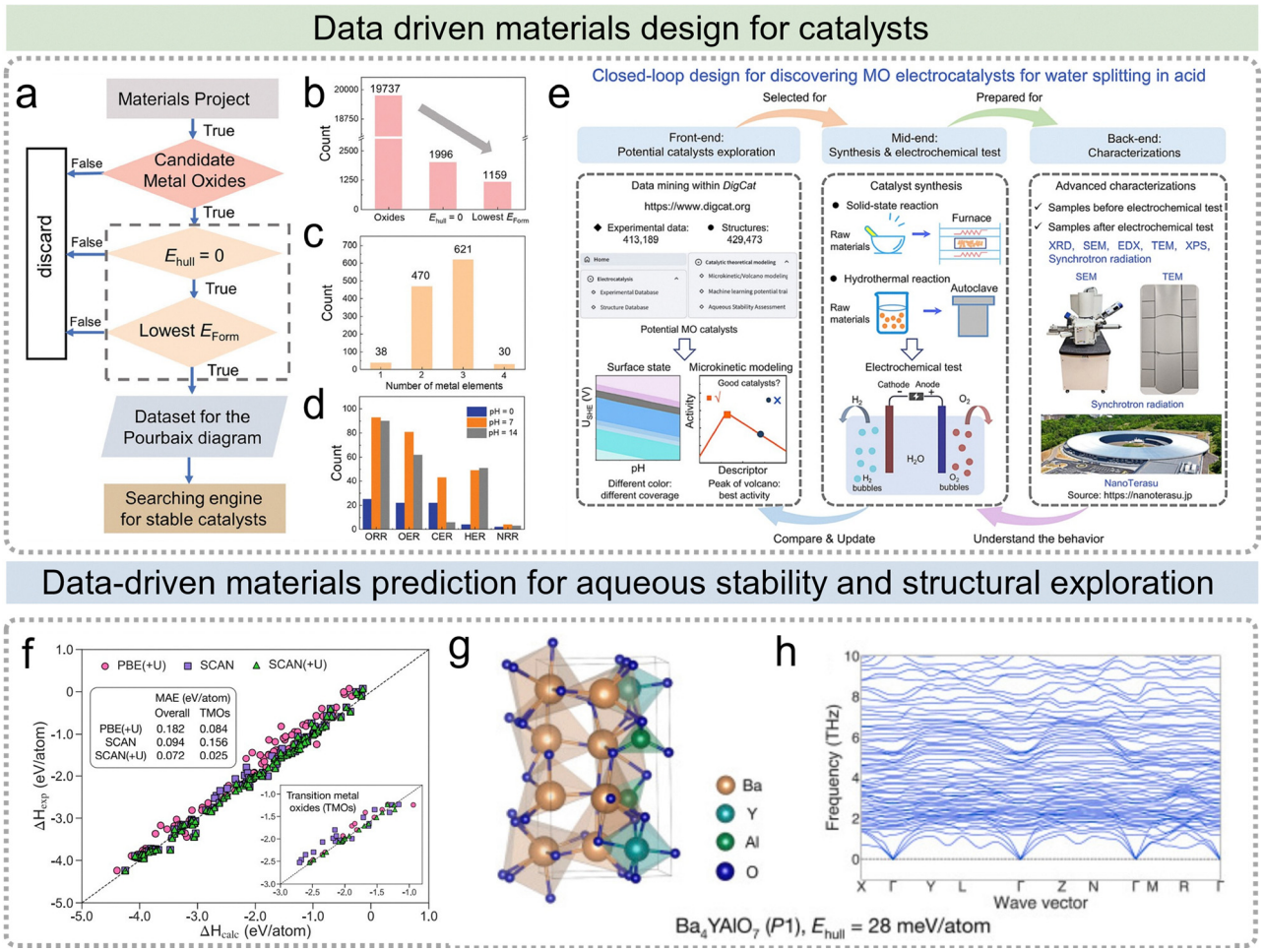


Fig. 5 (a) Schematic illustration of the data-mining workflow for screening aqueous-stable metal oxides (MOs). (b) Number of candidate MOs retained after each filtering stage and (c) distribution of bulk-stable MOs categorized by the number of constituent metal elements. (d) Count of aqueous-stable MOs under various pH and potential conditions. Reproduced with permission from ref. 55 licensed under a Creative Commons License CC BY 3.0. (e) Closed-loop discovery framework for non-noble MOs electrocatalysts toward overall water splitting. Reproduced with permission from ref. 56 licensed under a Creative Commons License CC BY-NC-ND 4.0. (f) Comparison between experimental and calculated formation enthalpies for 114 binary oxides. Inset highlights results with Hubbard U corrections for selected transition-metal oxides (TM = V, Cr, Mn, Fe, Co, Ni, Mo). Reproduced with permission from ref. 59 licensed under a Creative Commons License CC BY-NC-ND 4.0. (g and h) Crystal structure of Ba_4YAlO_7 and its corresponding computed phonon dispersion spectrum. Reproduced with permission from ref. 60 Copyright 2025, Elsevier.

3. Data-enabled physical understanding and AI-assisted discovery of SSEs

The discovery of SSEs has traditionally relied on incremental experiments and isolated simulations. As experimental measurements and first-principles calculations accumulate, a new paradigm is emerging: mechanistic understanding can be reconstructed from "old data". Rather than treating historical results as disconnected reports, systematic data integration, large language model (LLM)-assisted extraction, and physics-informed analysis can reveal hidden structure–property relationships and uncover methodological biases that were previously overlooked.[61]

A representative example is the dynamic database of solid-state electrolytes (DDSE, recently renamed DigBat: https://www.digbat.org), developed by Li and co-workers.[18] As of February 2026, it contains >3000 experimental materials, >26 000 ionic-conductivity measurements, and >850 computational entries. Through automated extraction, normalization, and descriptor standardization, DDSE converts scattered reports into a unified descriptor space that enables cross-family comparison (Fig. 6a).[62] Importantly, computational provenance is retained, preventing activation energies obtained using different methods from being treated as equivalent. In this way, "old data" are transformed into an interpretable and evolving knowledge base.

When hydride-based SSEs are analyzed within this structured framework, new insights emerge that are not evident in individual studies.[63] Reorganization of conductivity data across monovalent and divalent hydrides reveals statistically distinct transport regimes and a distribution shift toward higher conductivity in systems containing neutral molecules (Fig. 6b).

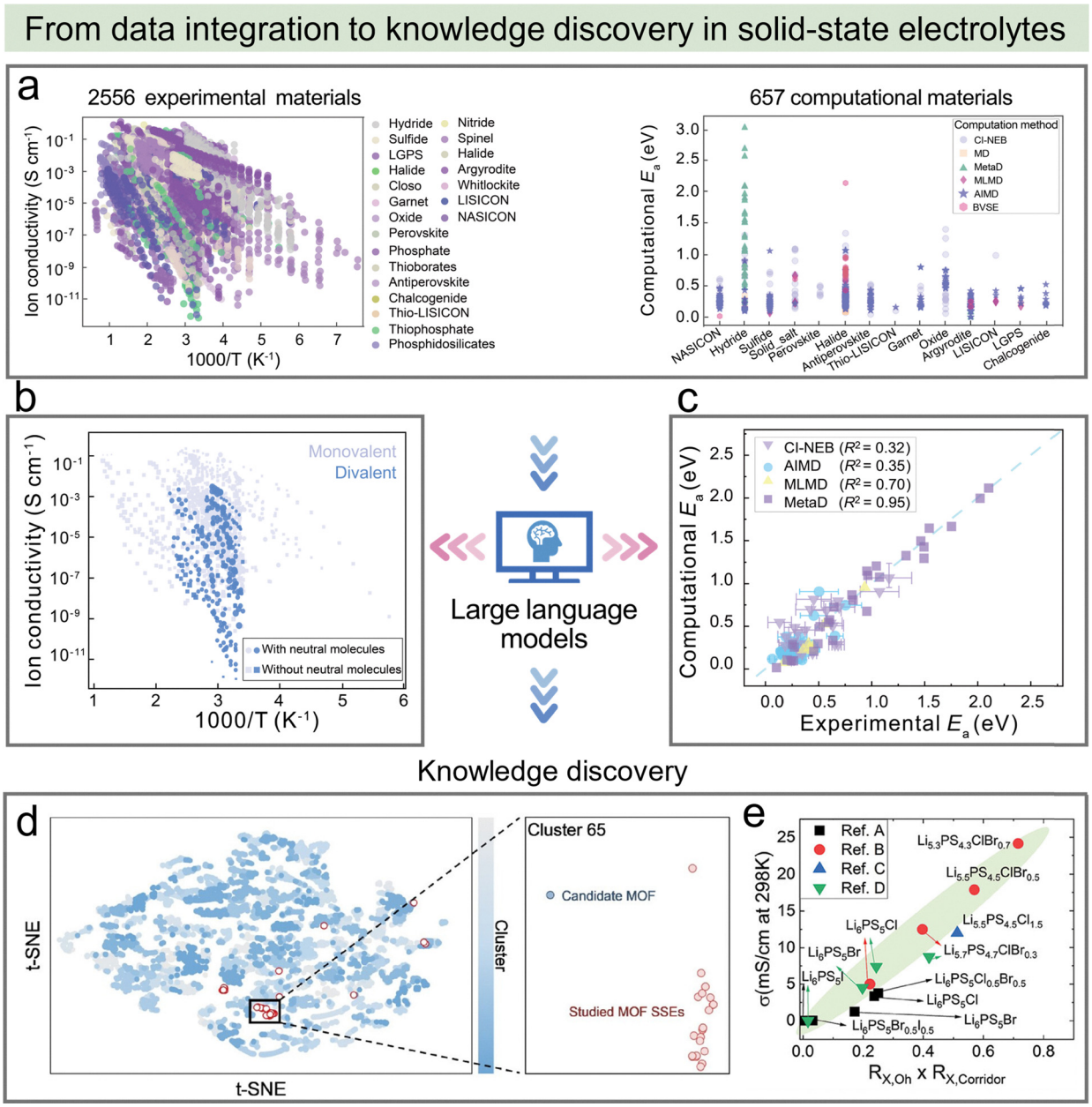


Fig. 6 Data integration-enabled knowledge reconstruction and AI-assisted discovery of SSEs. (a) Unified landscape of temperature-dependent ionic conductivity for ~3000 experimental materials and activation energies for ~700 computational materials. (b) Temperature-dependent ion conductivity for hydride SSEs, with/without neutral molecules. (c) Comparison of computational and experimental activation energies for different methods. Reproduced with permission from ref. 63 Copyright 2025, Wiley-VCH. (d) Representation clustering for mining MOF SSE candidates. Reproduced with permission from ref. 64 Copyright 2025, American Chemistry Society. (e) Ionic conductivity at 298 K *vs.* structural descriptors for sulfide-based SSEs. $R_{X,Oh} \times R_{X,Corridor}$ refers to the product of halogen substitution ratios at octahedral and corridor cage centers. Reproduced with permission from ref. 65 Copyright 2024, Wiley-VCH.

This separation becomes significant only after large-scale integration, linking molecular incorporation and coordination environments to migration barriers and suggesting that hydrides may represent a viable direction for multivalent transport when examined through a data-integrated perspective.

Beyond statistical trends, the same infrastructure enables benchmarking of computational methodologies. Direct comparison between computed and experimental activation energies reveals systematic discrepancies across methods (Fig. 6c). Metadynamics (MetaD) shows the strongest agreement with experiment, whereas climbing image nudged elastic band (CI-NEB) and *ab initio* molecular dynamics simulations (AIMD) exhibit weaker correlations, suggesting that static simulations may underestimate configurational entropy contributions in flexible lattices. Specifically, taking the CI-NEB method as an example, it can only locate a transition state when the initial pathway is artificially preset. This often leads to a significant discrepancy between the calculated activation energy (E_a) and the experimental measurement, and in severe cases, the coefficient of determination R^2 can be as low as 0.32. A typical example is using CI-NEB to

calculate the activation energy of Li^+ in γ-Li_3PS_4, yielding a value of 0.7 eV,[66] while the experimental values are 0.22 eV or 0.49 eV. The discrepancies between the calculated and experimental values reach 0.48 eV and 0.21 eV, respectively.[67,68] Through comparative analysis by data mining, we can intuitively observe this discrepancy, effectively demonstrating the great potential of extracting new knowledge from "old data".

Data-driven knowledge reconstruction can also extend exploration beyond established material classes. In the discovery of metal–organic framework (MOF) SSEs, LLM-based text mining and representation clustering were applied to a large candidate set, enabling targeted identification and validation of NOTT-400 with Li^+ conductivity of 2.23×10^{-4} S cm^{-1} and an electrochemical stability window of 0–4.79 V (Fig. 6d).[64] Here, literature-derived data function as a navigable knowledge space that guides hypothesis generation and experimental validation.

Finally, integrated datasets can yield transferable physical rules. In sulfide argyrodites, ionic conductivity scales with the product of halogen substitution ratios at octahedral and corridor cage centers ($R_{X,Oh} \times R_{X,Corridor}$), defining a compact descriptor that captures the balance between mobility enhancement and insulating phase formation (Fig. 6e).[65] This behavior illustrates how structured data can produce generalizable design principles across material families.

Collectively, Fig. 6 illustrates a data-to-knowledge pathway: integration and normalization enable statistical reinterpretation of material classes (Fig. 6a and b), benchmarking refines mechanistic reliability (Fig. 6c), and artificial intelligence (AI)-assisted analysis expands exploration while extracting transferable descriptors (Fig. 6d and e). The value of legacy data lies not in accumulation but in knowledge reconstruction through integration, provenance tracking, and critical reassessment, providing a foundation for future closed-loop discovery workflows.

4. From "old data" to structured knowledge and autonomous design in hydrogen storage

Hydrogen storage research has accumulated an extensive and heterogeneous body of experimental and computational results over several decades.[69–71] Mg-based hydrides alone have generated thousands of kinetic curves measured under diverse thermal conditions, particle sizes, catalytic modifications, and cycling histories.[72] In parallel, DFT calculations have mapped surface reactions, bulk diffusion pathways, and transition states for hydrogen release and uptake.[71] Despite this abundance of information, much of the knowledge has remained compartmentalized: computational barriers reported in isolated studies, kinetic curves interpreted within single experiments, and thermodynamic data embedded in graphical form across dispersed literature. A unifying theme in recent work is that such legacy data, when reorganized into structured and descriptor-linked formats, can yield mechanistic coherence and predictive capability that exceeds the scope of the original individual studies.

At the atomistic scale, descriptor-based modeling of MgH_2 dehydrogenation kinetics provides a clear example of this transformation.[73] In that study, a systematic set of activation barriers for sequential hydrogen removal was constructed using spin-polarized DFT and CI-NEB calculations. Rather than treating each transition state as a standalone result, the authors extracted physically meaningful descriptors from the calculated structures, including H–H distances, Mg–H bond strengths quantified *via* the integral value of the crystal orbital Hamiltonian population (ICOHP) analysis, and vacancy-related environmental factors. These descriptors were then regressed against the computed barriers to derive a closed-form surrogate model capable of reproducing DFT activation energies with high fidelity (Fig. 7a). This framework effectively compresses a set of expensive transition-state calculations into an interpretable kinetic expression, enabling rapid assessment of compositional and defect variations without repeated NEB searches. In doing so, reported computational results are converted from isolated barrier values into a transferable kinetic rule.

While first-principles studies operate at the atomic level, experimental kinetic data encode information across multiple length scales. The burst effect in MgH_2 desorption kinetics, analyzed in detail by Dong *et al.*[77] illustrates how transient regimes can be embedded within nominally smooth hydrogen-release curves. Through atomistic modeling and kinetic interpretation, they demonstrated that rapid depletion of surface-layer hydrogen can generate an apparent early-stage acceleration in hydrogen evolution before bulk diffusion becomes rate-limiting. Their analysis connects surface reaction energetics with subsequent diffusion-controlled stages, showing that a two-regime process naturally produces a burst-like signature in macroscopic kinetics. This work reframes the burst not as an experimental artifact, but as an emergent consequence of coupled surface and bulk processes during phase transformation. The implication is that kinetic curves, when examined with appropriate mechanistic awareness, encode layered information about surface reaction fronts, nucleation behavior, and diffusion barriers.

The burst effect was subsequently examined from a complementary statistical perspective by Cai *et al.*[74] Instead of focusing on a single mechanistic system, that study analyzed a broad collection of experimental desorption curves to evaluate whether transient acceleration constitutes a reproducible kinetic regime (Fig. 7b). By standardizing curve analysis and separating early-stage transients from later diffusion-dominated behavior, the authors demonstrated that burst-like features appear systematically across multiple datasets. Moreover, they showed that neglecting the early-stage acceleration obscures the multi-regime nature of the desorption process and complicates kinetic parameter extraction. In this sense, large archives of experimental curves become analyzable datasets rather than anecdotal observations. The combination of the two studies[74,77] establishes the burst effect as both physically grounded and reproducible across systems.

The ability to extract such insights at scale depends critically on converting literature-embedded information into structured

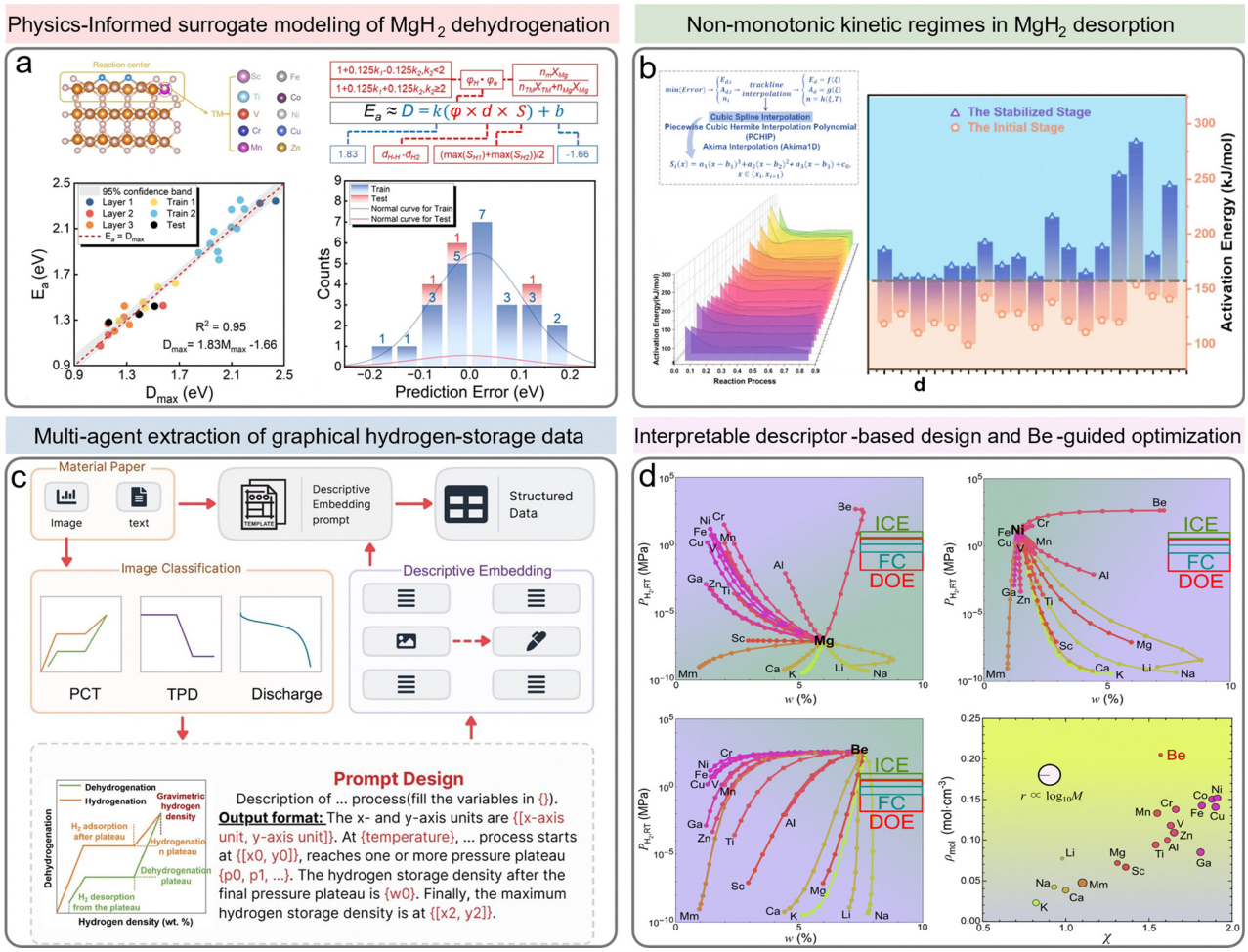


Fig. 7 From archived hydrogen storage data to structured design frameworks. (a) Descriptor-based surrogate modeling of MgH_2 dehydrogenation barriers derived from DFT/CI-NEB calculations, illustrating regression between computed activation energies and physically motivated descriptors. Reproduced with permission from ref. 73 Copyright 2024, Wiley-VCH. (b) Burst-effect behavior in MgH_2 desorption kinetics showing early-stage rate acceleration and multi-regime kinetic evolution. Reproduced with permission from ref. 74 Copyright 2025, Elsevier. (c) DIVE workflow for converting graphical hydrogen-storage data into structured machine-readable representations. Reproduced with permission from ref. 75 licensed under a Creative Commons License CC BY-NC 3.0. (d) Descriptor-based design maps for gravimetric capacity and equilibrium pressure highlighting Be-containing systems as potential outliers mitigating intrinsic thermodynamic trade-offs. Reproduced with permission from ref. 76 licensed under a Creative Commons License CC BY-NC 3.0.

form. The DIVE (descriptive interpretation of visual expression) framework addresses this challenge by introducing a multi-agent workflow capable of translating graphical hydrogen-storage data into machine-readable representations.[75] Through multimodal language-model pipelines, figure-caption parsing, and structured validation protocols, DIVE extracts quantitative data from pressure–composition–temperature (PCT), temperature programmed desorption (TPD), and discharge curves across thousands of publications (Fig. 7c). Applied to more than 4000 publications, the system generated a large-scale hydrogen storage database with improved extraction accuracy relative to direct multimodal parsing approaches. This infrastructure transforms graphical archives into searchable and analyzable datasets, enabling cross-study comparison and large-scale regression analysis. In effect, it bridges the gap between dispersed experimental figures and coherent data platforms.

Once structured datasets are established, descriptor-based thermodynamic modeling can be performed at scale. Using the DigHyd database comprising over 5000 curated pressure–composition isotherm entries collected in the DIVE framework, physically interpretable regression models were constructed to predict gravimetric capacity and equilibrium pressure at room temperature.[76] Minimal descriptor sets (including atomic mass, electronegativity difference, molar density, and ionic filling factor) were shown to achieve predictive performance comparable to modern machine-learning approaches while preserving full chemical transparency. The resulting design maps reveal an intrinsic trade-off between capacity and equilibrium pressure, distinguishing saline-type and interstitial-type hydrides and identifying Be-containing systems as potential outliers capable of partially mitigating this constraint (Fig. 7d). Here, structured experimental archives are translated into compositional design

landscapes, demonstrating how interpretable descriptors can guide materials exploration within chemically vast spaces.

Taken together, these developments outline a hierarchical transformation in hydrogen storage research. DFT calculations are condensed into descriptor-based surrogate models; archived kinetic curves are reinterpreted to identify transient mesoscale regimes; graphical literature is converted into structured datasets *via* multimodal extraction; and curated thermodynamic data are transformed into interpretable design maps. The unifying advance is not merely the accumulation of additional measurements or simulations, but the systematic restructuring of existing knowledge into forms that enable regression, comparison, and increasingly autonomous exploration. Hydrogen storage thus exemplifies how legacy data, when rigorously organized and analyzed, can serve as a foundation for new chemical insight and accelerated materials discovery.

5. Summary and outlook

The integration of databases with artificial intelligence is reshaping materials research. By reanalyzing dispersed experimental and computational data, data-driven approaches enable the identification of hidden structure–property relationships and accelerate the development of catalysts, SSEs, and hydrogen storage materials. Beyond efficiency, they provide a more systematic route toward performance-oriented materials design. Overall, the current data-driven knowledge discovery is evolving along two interrelated directions: one is to strengthen the data foundation, and the other is to move toward AI agent development.

5.1. Databases remain the foundation of data intelligence

High-quality databases are a necessary condition for the development of reliable AI agents. Although data mining based on AI has shown certain potential so far, the accuracy of data extraction achieved by AI depends on the type of raw data it is based on. In AI-driven data mining, inconsistent annotation forms or incomplete experimental details entered into the database can easily undermine the later data-driven knowledge discovery. Therefore, the current construction of databases still relies on meticulous manual organization and expert judgment. In this sense, the continuous human investment in the database determines the long-term value of data-driven knowledge discovery. For example, the Digital Materials Ecosystem[78–80] developed through extensive standardization and cross-document integration demonstrates how carefully organized data can support mechanism analysis and rational design. Future competitiveness will be more dependent on whether a durable and continuously developing data infrastructure can be constructed.

5.2. From digital platforms to AI agents

With a solid data foundation, AI can function as an effective accelerator rather than a replacement for researchers.[81] Scientists define questions and hypotheses, while intelligent agents assist in mining correlations, testing scenarios, and refining predictions. Looking forward, the next stage is expected to be characterized by a tighter coupling of old-data-driven discovery, theoretical calculations, and experimental validations. This will establish a self-improving framework, thereby advancing materials discovery from data accumulation toward knowledge generation.

Conflicts of interest

The authors declare no competing financial interest.

Data availability

No new data is generated in this paper.

Acknowledgements

The authors thank the support by JSPS KAKENHI (No. JP25H01508 and JP25K01737).

References

1 S. Li, M. Sun, K. Zhang, X. Cai, Y. Chen, C. Yang, Z. Yang, X. Tang, B. Huang and S. Yang, *Angew. Chem., Int. Ed.*, 2025, **64**, e202508366.
2 S. Li, J. Yu, S. Zhang, W. Qiu, X. Tang, Z. Lin, R. Cai, Y. Fang, S. Yang and X. Cai, *Adv. Funct. Mater.*, 2023, **34**, 2311989.
3 S. Liang, L. Huang, Y. Gao, Q. Wang and B. Liu, *Adv. Sci.*, 2021, **8**, e2102886.
4 Y. Wang, N. Ma, Y. Zhang, B. Liang and J. Fan, *Appl. Surf. Sci.*, 2023, **626**, 157126.
5 Y. Wang, Y. Zhang, N. Ma, J. Zhao, Y. Xiong, S. Luo and J. Fan, *Surf. Interfaces*, 2024, **50**, 104498.
6 Y. C. Wang, Y. J. Lai, L. Song, Z. Y. Zhou, J. G. Liu, Q. Wang, X. D. Yang, C. Chen, W. Shi, Y. P. Zheng, M. Rauf and S. G. Sun, *Angew. Chem., Int. Ed.*, 2015, **54**, 9907–9910.
7 L. Cao, Z. Zhao, Z. Liu, W. Gao, S. Dai, J. Gha, W. Xue, H. Sun, X. Duan, X. Pan, T. Mueller and Y. Huang, *Matter*, 2019, **1**, 1567–1580.
8 C. Liang, J. Yao, X. Hou, L. Li, N. Gao, H. Lu, R. Zhao, X. Guo, N. Xue, L. Peng, X. Guo, Y. Zhu and W. Ding, *J. Am. Chem. Soc.*, 2026, **148**, 5115–5124.
9 S. Luo, N. Ma, J. Zhao, Y. Wang, Y. Zhang, Y. Xiong and J. Fan, *J. Mater. Sci. Technol.*, 2024, **199**, 145–155.
10 N. Ma, H. Liu, L. Yu, Q. Yu, J. Cheng, Y. Ren, J. Fan and Z. Wei, *Small*, 2026, e13102, DOI: 10.1002/smll.202513102.
11 S. Luo, N. Ma, J. Zhao, Y. Zhang, Y. Wang, Y. Xiong and J. Fan, *Electrochim. Acta*, 2024, **499**, 144712.
12 Y. Li, J. Ma, T. D. Waite, M. R. Hoffmann and Z. Wang, *Environ. Sci. Technol.*, 2021, **55**, 10695–10703.
13 Z. Chen, Y. Zhao, H. Huang, G. Liu, H. Zhang, Y. Yan, H. Li, L. Liu, M. Liu, D. Wang and J. Zeng, *J. Am. Chem. Soc.*, 2025, **147**, 18737–18746.
14 K. Liu, Z. Sun, X. Peng, X. Liu, X. Zhang, B. Zhou, K. Yu, Z. Chen, Q. Zhou, F. Zhang, Y. Wang, X. Gao, W. Chen and P. Chen, *Nat. Commun.*, 2025, **16**, 2167.
15 Y. Zhao, S. Gao, X. Chen, Y. Liu, M. Zhang, Y. Liu, J. Kang, T. Guo, L. M. Liu and L. Guo, *J. Am. Chem. Soc.*, 2025, **147**, 31722–31730.
16 Y. Zhu, Y. Li, Y. Tian, B. Johannessen, P. Ramkissoon, Q. Chang, A. Zhang and S. Zhang, *Angew. Chem., Int. Ed.*, 2026, **65**, e17715.
17 Z. Gao, X. Yang, Z. Zhuang, Y. Zhang, J. Cai, Y. Li, W. Fu, H. Li and W. Yang, *Chem. Catal.*, 2026, **6**, 101692.
18 F. Yang, E. Campos dos Santos, X. Jia, R. Sato, K. Kisu, Y. Hashimoto, S.-I. Orimo and H. Li, *Nano Mater. Sci.*, 2024, **6**, 256–262.
19 L. Takahashi, T. N. Nguyen, S. Nakanowatari, A. Fujiwara, T. Taniike and K. Takahashi, *Chem. Sci.*, 2021, **12**, 12546–12555.
20 K.-Q. Zhong, F.-Y. Yu, D. Zhang, Z.-H. Li, D.-H. Xie, T.-T. Li, Y. Zhang, L. Yuan, H. Li, Z.-Y. Wu and G.-P. Sheng, *Angew. Chem., Int. Ed.*, 2025, **64**, e202500004.
21 J. H. Montoya, C. Tsai, A. Vojvodic and J. K. Norskov, *ChemSusChem*, 2015, **8**, 2180–2186.

22 Y. Zhuang, X. Yang, C. Zhang, X. Jia, D. Zhang, M. Li, T. Yao, J. Peng, Z. Gao, W. Yang and H. Li, *Precis. Chem.*, 2026, DOI: 10.1021/prechem.5c00449.
23 L. Ward, S. Babinec, E. J. Dufek, D. A. Howey, V. Viswanathan, M. Aykol, D. A. C. Beck, B. Blaiszik, B.-R. Chen, G. Crabtree, S. Clark, V. De Angelis, P. Dechent, M. Dubarry, E. E. Eggleton, D. P. Finegan, I. Foster, C. B. Gopal, P. K. Herring, V. W. Hu, N. H. Paulson, Y. Preger, D. Uwe-Sauer, K. Smith, S. W. Snyder, S. Sripad, T. R. Tanim and L. Teo, *Joule*, 2022, **6**, 2253–2271.
24 A. Belsky, M. Hellenbrandt, V. L. Karen and P. Luksch, *Acta Crystallogr. B*, 2002, **58**, 364–369.
25 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
26 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, *Sci. Data*, 2019, **6**, 75.
27 Y. Wang, Y. Zhang, N. Ma, J. Zhao, Y. Xiong, S. Luo and J. Fan, *J. Mater. Sci. Technol.*, 2025, **213**, 14–23.
28 Y. Zhang, Y. Wang, N. Ma and J. Fan, *J. Energy Chem.*, 2024, **97**, 139–148.
29 Y. Z. Yu Xiong, Y. Wang, N. Ma, Q. Wang, J. Z. Deshuai Yang, S. Luo and J. Fan, *Surf. Interfaces*, 2025, **59**, 105932.
30 N. Ma, Y. Zhang, Y. Wang, C. Huang, J. Zhao, B. Liang and J. Fan, *Appl. Surf. Sci.*, 2023, **628**, 157225.
31 M. Núñez and D. G. Vlachos, *Ind. Eng. Chem. Res.*, 2018, **58**, 6146–6154.
32 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, *APL Mater.*, 2013, **1**, 011002.
33 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, *npj Comput. Mater.*, 2015, **1**, 1–15.
34 D. Zhang, Z. Bao, Y. Chu, Z. Guo, X. Jia, Q. Jiang, H. Liu, T. Liu, T. Lu, Y. Lu, D. Devang Shah, Y. Wang, Y. Wang, Y. Wang, S. Ye, S. Ying, Z. Yu, L. Zhang, S. Zhao and H. Li, *ChemRxiv*, 2026, preprint, DOI: 10.26434/chemrxiv-2024-9lpb9.
35 D. Zhang, X. Jia, H. Liu, Y. Wang, S. Ye, Q. Jiang, Y. Wang, Z. Guo, L. Zhang, L. Wei, W. Yang, H. Liu, S. Zhao, H. Xu, D. Cheng, Y. Hashimoto, T. Tomai and H. Li, *AI Agent*, 2025, **1**, 2.
36 Y. Wang, Z. Wu, Y. Jiang, D. Zhang, Q. Wang, C. Wang, H. Li, X. Jia, J. Fan and H. Li, *Adv. Funct. Mater.*, 2025, e06314.
37 D. Zhang, Z. Wang, F. Liu, P. Yi, L. Peng, Y. Chen, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2024, **146**, 3210–3219.
38 S. Ye, Y. Wang, H. Liu, D. Zhang, X. Jia, L. Zhang, Y. Zhang, A. Kumatani, H. Shiku and H. Li, *J. Mater. Chem. A*, 2025, **13**, 37821–37832.
39 Y. Wang, D. Zhang, B. Sun, X. Jia, L. Zhang, H. Cheng, J. Fan and H. Li, *Angew. Chem., Int. Ed.*, 2024, e202418228, DOI: 10.1002/anie.202418228.
40 K.-i Aika, *Catal. Today*, 2017, **286**, 14–20.
41 B. A. Rohr, A. R. Singh and J. K. Nørskov, *J. Catal.*, 2019, **372**, 33–38.
42 A. Cao, V. J. Bukas, V. Shadravan, Z. Wang, H. Li, J. Kibsgaard, I. Chorkendorff and J. K. Norskov, *Nat. Commun.*, 2022, **13**, 2382.
43 X. You, Z. Guo, Q. Jiang, J. Xia, S. Wang, X. Yang, Z. Zhuang, Y. Li, H. Xiang, H. Li and B. Yu, *Nano Lett.*, 2025, **25**, 8704–8712.
44 T. Pu, H. Tian, M. E. Ford, S. Rangarajan and I. E. Wachs, *ACS Catal.*, 2019, **9**, 10727–10750.
45 H. Li, A. Cao and J. K. Nørskov, *ACS Catal.*, 2021, **11**, 12052–12057.
46 H. Li, S. Kelly, D. Guevarra, Z. Wang, Y. Wang, J. A. Haber, M. Anand, G. T. K. K. Gunasooriya, C. S. Abraham, S. Vijay, J. M. Gregoire and J. K. Nørskov, *Nat. Catal.*, 2021, **4**, 463–468.
47 W. Yang, Z. Jia, B. Zhou, L. Chen, X. Ding, L. Jiao, H. Zheng, Z. Gao, Q. Wang and H. Li, *ACS Catal.*, 2023, **13**, 9695–9705.
48 X. Wang, Z. Li, D. Zhang, H. Li, H. Xu and D. Cheng, *Angew. Chem., Int. Ed.*, 2026, e24612.
49 K.-i Ota, Y. Ohgi, K.-D. Nam, K. Matsuzawa, S. Mitsushima and A. Ishihara, *J. Power Sources*, 2011, **196**, 5256–5263.
50 D. Chen, Z. Chen, Z. Lu, J. Tang, X. Zhang and C. V. Singh, *J. Mater. Chem. A*, 2020, **8**, 21241–21254.
51 M. He, W. An, Y. Wang, Y. Men and S. Liu, *Small*, 2021, **17**, 2104445.
52 S. Ye, F. Liu, F. She, J. Chen, D. Zhang, A. Kumatani, H. Shiku, L. Wei and H. Li, *Angew. Chem., Int. Ed.*, 2025, **64**, e202425402.
53 Q. Jiang, M. Gu, S. Pei, T. Wang, F. Liu, X. Yang, D. Zhang, Z. Wu, Y. Wang, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2025, **147**, 26029–26039.
54 D. Zhang, F. She, J. Chen, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2025, **147**, 6076–6086.
55 X. Jia, Z. Yu, F. Liu, H. Liu, D. Zhang, E. Campos Dos Santos, H. Zheng, Y. Hashimoto, Y. Chen, L. Wei and H. Li, *Adv. Sci.*, 2023, e2305630.
56 X. Jia, Z. Zhou, F. Liu, T. Wang, Y. Wang, D. Zhang, H. Liu, Y. Wang, S. Ye, K. Amezawa, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2025, **147**, 22642–22654.
57 A. K. Singh, J. H. Montoya, J. M. Gregoire and K. A. Persson, *Nat. Commun.*, 2019, **10**, 443.
58 X. Guo, Z. Wang, Z. Deng, B. Wang, X. Chen and S. P. Ong, *Chem. Mater.*, 2020, **32**, 6875–6885.
59 Z. Wang, X. Guo, J. Montoya and J. K. Nørskov, *npj Comput. Mater.*, 2020, **6**, 160.
60 Y. An and Z. Wang, *Mater. Today Energy*, 2025, **54**, 102059.
61 Q. Wang, R. Sato, R. García-Méndez, W. Jang, P. Ou, A. Soon, J. Zhao, X. Wang, S.-I. Orimo and E. J. Cheng, *AI Agent*, 2025, **1**, 10.
62 Q. Wang, F. Yang, Y. Wang, D. Zhang, R. Sato, L. Zhang, E. J. Cheng, Y. Yan, Y. Chen, K. Kisu, S.-I. Orimo and H. Li, *Angew. Chem., Int. Ed.*, 2025, **64**, e202506573.
63 F.-L. Yang, R. Sato, E. J.-F. Cheng, K. Kisu, Q. Wang, X. Jia, S.-I. Orimo and H. Li, *J. Electrochem*, 2024, **30**, 3.
64 J. Zhang, J. Li, G. Zhao, Q. Wang, Y.-G. Guo and C. Yang, *J. Am. Chem. Soc.*, 2025, **147**, 40496–40506.
65 H. J. Lee, H. Kim, S. Ji, K. Choi, H. Choi, W. Lim and B. Lee, *Adv. Energy Mater.*, 2024, **14**, 2402396.
66 N. D. Lepley, N. A. W. Holzwarth and Y. A. Du, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **88**, 104103.
67 M. Tachez, J.-P. Malugani, R. Mercier and G. Robert, *Solid State Ionics*, 1984, **14**, 181–185.
68 K. Homma, M. Yonemura, T. Kobayashi, M. Nagao, M. Hirayama and R. Kanno, *Solid State Ionics*, 2011, **182**, 53–58.
69 A. Züttel, *Mater. Today*, 2003, **6**, 24–33.
70 M. Hirscher, V. A. Yartys, M. Baricco, J. Bellosta von Colbe, D. Blanchard, R. C. Bowman, D. P. Broom, C. E. Buckley, F. Chang, P. Chen, Y. W. Cho, J.-C. Crivello, F. Cuevas, W. I. F. David, P. E. de Jongh, R. V. Denys, M. Dornheim, M. Felderhoff, Y. Filinchuk, G. E. Froudakis, D. M. Grant, E. M. Gray, B. C. Hauback, T. He, T. D. Humphries, T. R. Jensen, S. Kim, Y. Kojima, M. Latroche, H.-W. Li, M. V. Lototskyy, J. W. Makepeace, K. T. Møller, L. Naheed, P. Ngene, D. Noréus, M. M. Nygård, S.-I. Orimo, M. Paskevicius, L. Pasquini, D. B. Ravnsbæk, M. Veronica Sofianos, T. J. Udovic, T. Vegge, G. S. Walker, C. J. Webb, C. Weidenthaler and C. Zlotea, *J. Alloys Compd.*, 2020, **827**, 153548.
71 N. O. Hassan, A. Ghais, M. H. M. Ahmed, M. Adam, R. Ahmed, A. Abdelatti, D. E. P. Klenam, M. O. Bodunrin and A. Koko, *Int. J. Hydrogen Energy*, 2026, **217**, 153928.
72 I. P. Jain, C. Lal and A. Jain, *Int. J. Hydrogen Energy*, 2010, **35**, 5133–5144.
73 C. Li, W. Yang, H. Liu, X. Liu, X. Xing, Z. Gao, S. Dong and H. Li, *Angew. Chem., Int. Ed.*, 2024, **63**, e202320151.
74 J. Cai, H. Wang, X. Tang, Z. Miao, T. Yao, Y. Liu, H. Wang, Z. Gao and W. Yang, *J. Energy Storage*, 2025, **136**, 118450.
75 D. Zhang, X. Jia, H. B. Tran, S. H. Jang, L. Zhang, R. Sato, Y. Hashimoto, T. Sato, K. Konno, S. I. Orimo and H. Li, *Chem. Sci.*, 2026, **17**, 3031–3042.
76 S. H. Jang, D. Zhang, H. B. Tran, X. Jia, K. Konno, R. Sato, S. I. Orimo and H. Li, *Chem. Sci.*, 2025, **16**, 23111–23120.
77 S. Dong, C. Li, J. Wang, H. Liu, Z. Ding, Z. Gao, W. Yang, W. Lv, L. Wei, Y. Wu and H. Li, *J. Mater. Chem. A*, 2022, **10**, 22363–22372.
78 D. Zhang, X. Jia, H. B. Tran, F. Yang, Q. Wang, H. Liu, Y. Qi, E. J. Cheng and H. Li, *ChemRxiv*, 2024, preprint, DOI: 10.26434/chemrxiv-2024-p9fvc.
79 D. Zhang, Y. Chen, C. Liu, Y. Liu, H. Xin, J. Peng, P. Ou and H. Li, *Angew. Chem., Int. Ed.*, 2026, e26150.
80 D. Zhang, X. Jia, Y. Wang, H. Liu, Q. Wang, S.-H. Jang, D. Shah, S. Ye, H. B. Tran and H. Li, *Chem. Sci.*, 2026, **17**, 5782–5804.
81 D. Zhang and H. Li, *Mol. Chem. Eng.*, 2025, **1**, 100003.