GDCh Research Article Angewandte International Edition Chemie www.angewandte.org

How to cite: *Angew. Chem. Int. Ed.* **2025**, *64*, e202503105
doi.org/10.1002/anie.202503105

Artificial Intelligence for Batteries **Very Important Paper**

Uni-Electrolyte: An Artificial Intelligence Platform for Designing Electrolyte Molecules for Rechargeable Batteries

Xiang Chen[+],* *Mingkang Liu*[+], *Shiqiu Yin, Yu-Chen Gao, Nan Yao, and Qiang Zhang**

Abstract: Electrolytes are an essential part of rechargeable batteries, such as lithium batteries. However, electrolyte innovation is facing grand challenges due to the complicated solution chemistry and infinite molecular space (>10^{60} for small molecules). This work reported an artificial intelligence (AI) platform, namely Uni-Electrolyte, for designing advanced electrolyte molecules, which mainly includes three parts, i.e., EMolCurator, EMolForger, and EMolNetKnittor. New molecules can be designed by combining high-throughput screening and generative AI models from more than 100 million alternative molecules in the EMolCurator module. The molecular properties, including frontier molecular orbital information, formation energy, binding energy with a Li ion, viscosity, and dielectric constant, can be adopted as the screening parameters. The EMolForger and EMolNetKnittor modules can predict the retrosynthesis pathway and solid electrolyte interphase (SEI) formation mechanism for a given molecule, respectively. With the assistance of advanced AI methods, the Uni-Electrolyte is strongly supposed to discover new electrolyte molecules and chemical principles, promoting the practical application of next-generation rechargeable batteries.

Introduction

Rechargeable batteries, especially lithium (Li) ion batteries (LIBs), have been widely applied in modern society, from electronic devices, electric vehicles, and smart grids to the low-altitude industry. These wide applications increasingly put forward demanding requirements for next-generation batteries with high safety, high energy density, high power density, long lifespan, or wide temperature window (3H1L1 W). Among various battery technology innovations, designing advanced electrolyte molecules has been strongly considered as one of the most promising approaches due to the significant role of electrolytes in stabilizing battery interfaces and regulating battery performance.[1–5] Besides, new electrolytes can be directly adopted in current battery manufacturing without a huge change in the equipment.

Looking back in the history of LIB developments, the use of ethylene carbonate (EC)-based electrolyte is definitely a milestone due to its irreplaceable role in forming a stable solid electrolyte interphase (SEI) on the graphite anode.[6–8] EC-based ester electrolytes still dominate the current LIB battery market after more than 30 years of development. However, ester molecules are very active towards Li metal anodes, which possess the largest specific capacity as well as a very low electrode potential and are the most promising choice for constructing ultrahigh-energy-density batteries.[9] Beyond ester molecules, ether solvents such as 1,2-dimethoxyethane (DME) have been widely applied to Li metal batteries due to their relatively high stability against Li metal anodes and good Li salt solubility. High-concentration and localized high-concentration electrolytes have been further constructed mainly based on DME solvents and have demonstrated promising electrochemical performances.[5] Besides, many electrolyte additives have been proposed to improve battery 3H1L1 W performance. For example, vinyl carbonate (VC) and fluoroethylene carbonate (FEC) can be adapted to produce a polymeric layer on the electrode surface. Trimethyl phosphate (TMP) and biphenyl can be adopted as flame retardants and overcharge protective additives, respectively. All the above examples ensure the important role of new electrolyte innovation in battery developments.[10,11]

Tremendous electrolyte molecules have been explored for rechargeable batteries mainly through a conventional trial-and-error approach, while only dozens of them have been applied to current commercialized batteries. The time-consuming and low-efficiency approach to searching for advanced electrolytes is a major challenge for promoting next-generation batteries in the following years. Recently, the rise of artificial intelligence (AI) technology has afforded new and promising chances for the frontier research of chemistry and materials science.[12–18] Especially, the Nobel Prizes in both Physics and Chemistry in 2024 are rewarded to AI-related fields due to the great success of AI. In the electrolyte field, previous studies particularly focused on developing machine learning potentials to accelerate the simulation and expand length scales while maintaining first-principles accuracy.[19–21]

[*] X. Chen[+], Y.-C. Gao, N. Yao, Q. Zhang
Tsinghua Center for Green Chemical Engineering Electrification & Beijing Key Laboratory of Complex Solid State Batteries, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
E-mail: xiangchen@mail.tsinghua.edu.cn
zhang-qiang@mails.tsinghua.edu.cn

X. Chen[+], M. Liu[+], S. Yin
AI for Science Institute, Beijing 100080, China

[+] Both authors contributed equally to this work.

Additional supporting information can be found online in the Supporting Information section

© 2025 Wiley-VCH GmbH

Chen and authors developed data–knowledge-dual-driven and explainable machine learning models to directly predict electrolyte properties and establish molecular structure–function relationships. Besides, several electrolyte molecules have been predicted by combining machine learning models and high-throughput screening.[22,23] Although great successes have been demonstrated in applying AI to probing electrolyte chemistry and designing promising electrolyte molecules, a platform that integrates AI techniques and domain knowledge for designing advanced battery electrolytes has not been reported.

In this contribution, we reported the first AI platform, namely Uni-Electrolyte, for designing electrolyte molecules for rechargeable batteries. The Uni-Electrolyte integrates advanced AI algorithms for electrolyte molecule design, mainly including three modules. The EMolCurator module can assist in designing new molecules through high-throughput screening from an embedded electrolyte database or an uploaded molecular database and even beyond previous databases using AI-based generative models. When the target molecule is defined, the EMolForger module can predict the retrosynthesis pathways and reaction conditions, which are helpful in synthesizing the predicted molecules. Last but not least, the EMolNetKnittor module can unveil the reaction pathway of SEI products, which is supposed to be a critical factor in stabilizing the electrolyte–electrode interphase. Collectively, the Uni-Electrolyte is supposed to discover new electrolyte molecules, including solvents and additives, and further promote the practical application of next-generation batteries.

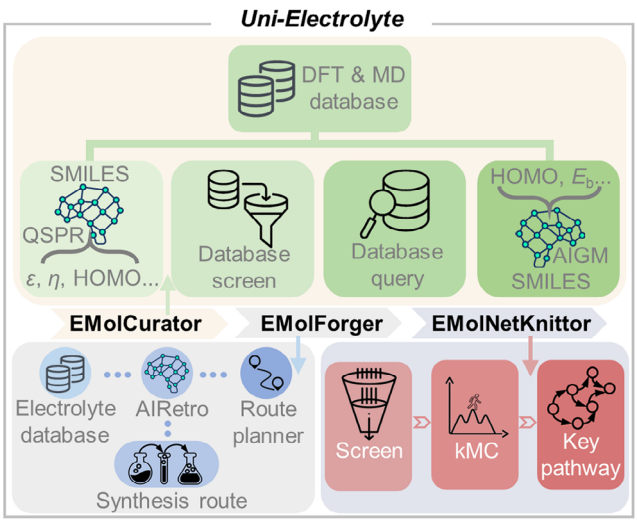

Figure 1. Schematic representation of the Uni-Electrolyte platform with three modules. The EMolCurator module aims to design new electrolyte molecules. Based on the embedded electrolyte database, QSPR and AI-based generative models were trained. The EMolForger module can predict the synthesis pathways and corresponding reaction conditions of potential electrolyte molecules. It was built with a synthetic route planner and AI-based single-step retrosynthesis predictor. The EMolNetKnittor module assesses the filtered electrolyte species and reaction database to propose chemical reaction networks and perform SEI-product analysis.

Results and Discussion

Framework of Uni-Electrolyte

The Uni-Electrolyte aims to expedite the discovery, synthesis, and analysis of novel electrolyte molecules using advanced AI technology (Figure 1). The platform mainly includes three interconnected modules: EMolCurator, EMolForger, and EMolNetKnittor. By integrating the three modules, Uni-Electrolyte offers a powerful and versatile platform for accelerating the discovery and development of advanced electrolyte materials. The detailed functions for each module will be introduced as follows.

The EMolCurator enables efficient and precise molecular design by combining advanced AI techniques and multi-scale simulations. Promising molecular candidates can be identified from vast chemical spaces by the four functions of EMolCurator, including predicting molecular properties, screening molecules according to their properties, searching similar molecules, and generating molecules by AI models.

1. Molecular Property Prediction. The cornerstone of EMolCurator is a robust quantitative structure–property relationship (QSPR) model, trained on a comprehensive electrolyte dataset constructed from density functional theory (DFT) and molecular dynamics (MD) calculations. The model accurately predicts various molecular properties, including binding energy, the highest occupied molecular orbital (HOMO) energy, the lowest unoccupied molecular orbital (LUMO) energy, viscosity (η), dielectric constant (ε), and other relevant properties, from both two-dimensional (2D) and three-dimensional (3D) molecular representations (Figure 2a).

2. Multi-Criterion Screening. A multi-criterion screening process is employed on a pre-designed database of electrolyte molecules (Figure 2b). The process involves filtering molecules based on multiple property criteria, such as desired HOMO–LUMO gap, binding energy, or specific structural features. By combining these criteria, a subset of molecules could be identified that are most likely to meet the desired specifications, accelerating the design process.

3. Molecular Similarity Search. A similarity search algorithm that leverages QSPR-predicted molecular properties was built upon a vector database (Figure 2c). A pgvector-enhanced PostgreSQL backend was constructed to store molecules and their corresponding property vectors. These vectors, representing a combination of physical and electrical properties, enable efficient similarity searches. When a query molecule is input, either it is retrieved from the database or its properties are predicted on the fly using a state-of-the-art (SOTA) model. The query vector is then compared to database vectors using a similarity metric, such as cosine similarity, to identify molecules with similar property profiles. The approach expands the search space while maintaining focus on relevant chemical regions.

4. Molecular Generation. To overcome the limitations of relying solely on existing datasets, an AI-driven molecular generation module (AIGM) was built. The module

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

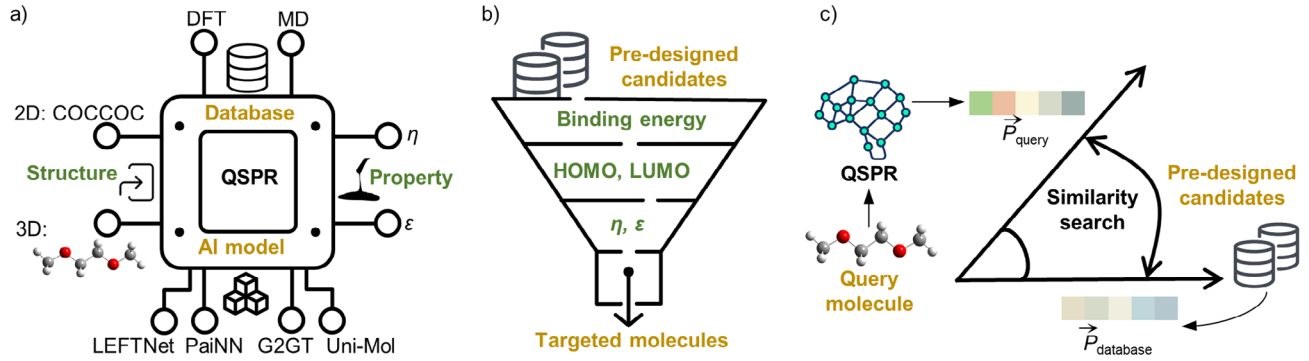

Figure 2. Three functions of the EMolCurator module. a) The QSPR model is benchmarked and trained on the DFT and MD databases. It intakes 2D or 3D molecular graphs and outputs their properties. b) Pre-designed candidate electrolyte molecules are screened with respect to user-defined intervals. c) Query similar-properties molecules with the vector database, the queried vector itself is composed of the predicted properties from the QSPR model.

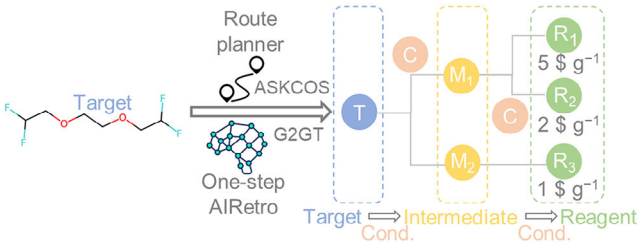

Figure 3. The illustration of the retrosynthesis module. The module includes two AI components, i.e., the G2GT One-Step AIRetro predictor and the Askcos Synthetic Route Planner. During inference, the module proposes purchasable starting reagents and potential intermediates, along with detailed reaction conditions and prices.

was trained on the same DFT and MD dataset and can generate novel molecules with targeted properties, including HOMO–LUMO gap optimization, binding energy optimization, and direct structural design. To ensure the practicality of the generated molecules, a rigorous filtering pipeline was implemented. For example, the stability of molecules (e.g., formation energy), the similarity of new molecules compared with existing molecules in the dataset, and the synthesizability of new molecules were considered.

The entire molecular design process is iterative, with user-defined targets guiding the generation and filtering steps (Figure S1). The workflow can be repeated until convergence is achieved, leading to the identification of optimal molecular candidates. By combining the above functions, the EMolCurator module can provide a powerful tool for accelerating the discovery of novel electrolyte molecules with tailored properties. The AI-driven nature of the framework enables rapid exploration of vast chemical spaces, leading to the identification of promising candidates that may be difficult to bed discovered through traditional methods.

After molecular design, obtaining the molecules is the next step, which is supported by the EMolForger module. The EMolForger module is an AI-powered retrosynthetic analysis platform, which includes two parts, i.e., the G2GT One-Step AIRetro Predictor and the Askcos Synthetic Route Planner (Figure 3). The G2GT One-Step AIRetro Predictor, built upon graph neural networks,[24] demonstrates exceptional capability in identifying and evaluating feasible one-step retrosynthetic transformations. Through extensive training on comprehensive chemical reaction databases, the predictor has developed robust algorithms for assessing chemical plausibility, establishing a reliable foundation for synthesis planning. The Askcos SOne-Step AIRetro Predictorynthetic Route Planner builds upon these initial predictions to optimize multi-step synthetic pathways.[25] The planner can conduct detailed analyses of crucial reaction parameters, including solvent selection, catalyst optimization, and temperature conditions, while prioritizing both synthetic efficiency and economic viability.

The practical implementation of EMolForger extends beyond theoretical route planning through its sophisticated reagent analysis capability. The system provides comprehensive specifications for chemical reagents while conducting detailed cost analyses, enabling researchers to make well-informed decisions regarding reagent selection and alternatives. The effectiveness of EMolForger is substantially enhanced by its domain-specific optimization, with models fine-tuned using reaction datasets specifically relevant to electrolyte molecule synthesis. The focused training approach ensures exceptional accuracy within the specialized domain of electrolyte chemistry, addressing the unique challenges and requirements of this field.

Following the molecular design strategies and synthetic planning capabilities discussed in previous sections, understanding the behavior and decomposition pathways of successfully synthesized electrolyte molecules becomes crucial. The SEI formation, a critical process affecting battery performance, requires sophisticated analytical tools for a comprehensive investigation. The EMolNetKnittor module was developed to probe the SEI formation mechanisms of electrolytes (Figure 4).

The EMolNetKnittor employs a two-pronged approach to analyze SEI formation. Initially, the input electrolyte molecule is queried against a comprehensive built-in database. If relevant species and reactions are identified, the platform utilizes stochastic kinetic Monte Carlo (kMC) simulations to construct a detailed reaction network. The

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

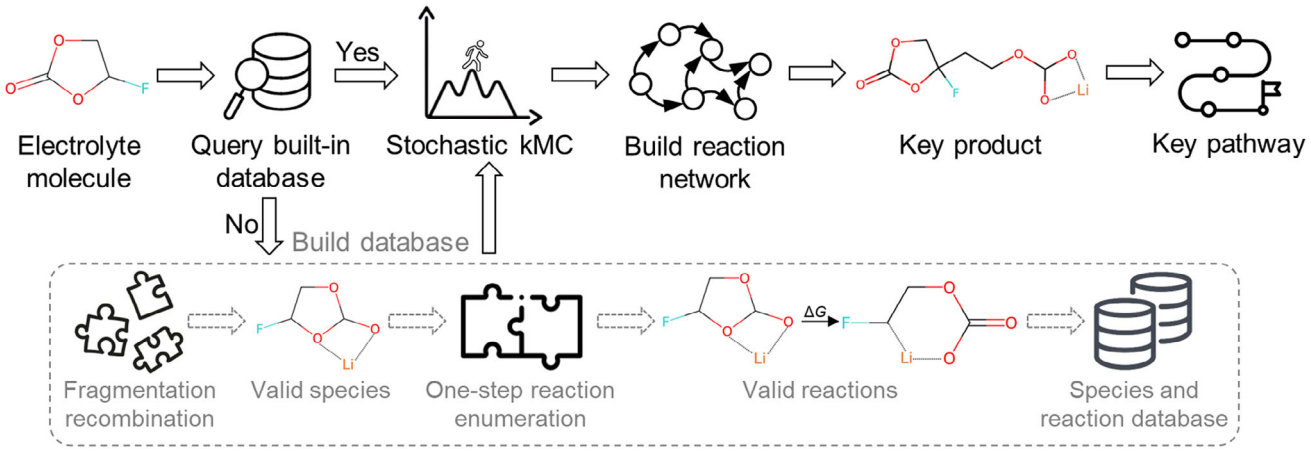

Figure 4. Illustration of the SEI-analysis module. The input electrolyte molecule is queried against a built-in database at first. If related species and reactions are found, the module utilizes stochastic kMC simulations to build a reaction network. The network then allows the identification of key products and their formation pathways. However, if no matches are found in the database, the module offers an on-the-fly database building option, which takes the queried molecule as input and generates species through fragmentation and recombination, followed by reaction enumeration. Note that filters are applied to ensure the validity of both generated species and reactions.

network facilitates the identification of key products and their formation pathways.

However, if the database lacks information about the input molecule, the EMolNetKnittor provides an on-the-fly database-building capability, which involves fragmenting and recombining the input molecule to generate potential species. Subsequently, one-step reactions are enumerated among these species, and rigorous filters are applied to ensure the validity of both species and reactions. The process results in a tailored database specific to the electrolyte molecule under investigation.

To address the limitations of existing tools like HiPRGen,[26] which supports only a subset of the LiBE[27] electrolyte database, the database in the EMolNetKnittor module was expanded to encompass the entire LiBE dataset. The expanding dataset enables the analysis of electrolyte molecules containing a wider range of elements, such as F, N, P, and S. Additionally, the database construction process has been automated, empowering users to add custom molecules and further broaden the application scope of EMolNetKnittor.

QSPR Model Performance

The QSPR models were assessed regarding their capacity in predicting critical physical and electronic properties, namely ε, η, HOMO, LUMO, and the binding energy with a Li ion. The initial evaluation sets the stage for a deeper exploration of the interplay between model architecture, dataset characteristics, and property-specific performance.

To ensure the rigor of these benchmarks, the DFT and MD database was partitioned into training, validation, and test datasets in a 3.79:0.37:1 ratio (34358:3435:9056). Further, the test dataset was split into Independent and Identically Distributed (IID) and Out-of-Distribution (OOD) subsets around a 1:1 ratio (4604:4452), enabling a nuanced evaluation of the models' generalizability. OOD molecules were identified using the Bemis–Murcko scaffold grouping method,[28] which classifies molecules based on the rarity of their molecular backbones. The methodology ensured a meaningful comparison of model performance across familiar and novel chemical spaces.

The benchmarking spanned both 2D and 3D QSPR models, revealing distinct patterns in their predictive capabilities. Table S1 summarizes the relative performance of these models, providing a detailed comparison of their mean absolute errors (MAE) on IID and OOD datasets. The relative error was calculated with Uni-Mol[29] serving as the baseline. Among the 2D models, which are input with simplified molecular input line entry system (SMILES) strings containing 2D topological information, those pre-trained on large public datasets exhibited obvious advantages. G2GT,[24] pre-trained on the USPTO[30] chemical reaction database, consistently outperformed Uni-Mol, pre-trained on QM9.[31] G2GT achieved a relative error of 0.97 compared to Uni-Mol's 1.0, a 3% improvement. Notably, G2GT demonstrates superior performance on all OOD metrics, indicating its enhanced extrapolation ability. The results suggest that the chemical reactivity data embedded in USPTO aligns better with the prediction of physical properties like dielectric constant and viscosity. Given its superior performance in capturing and leveraging 2D information, G2GT can serve as an excellent 2D information extractor.

For 3D models, the input 3D conformations were generated using RDKit. These 3D QSPR models demonstrated the pivotal role of geometric representation in capturing molecular features. Equivariant models such as LEFTNet[32] and PaiNN[33] significantly outperformed the invariant SchNet[34] model, highlighting the importance of symmetry-aware architectures. Notably, powerful invariant models such as SphereNet[35] were not considered here due to their extremely high computational cost, as mentioned in previous reports.[32] LEFTNet emerged as the

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

top-performing model, surpassing both SchNet and PaiNN across all metrics, demonstrating a relative improvement of approximately 6.0% and 6.6% over PaiNN and SchNet, respectively. The superiority is likely attributable to its architectural enhancements, such as the Local Substructure Encoding (LSE) and Frame Transition Encoding (FTE) modules, which enable comprehensive encoding of both local and global molecular information. Notably, LEFTNet, despite not being pre-trained, achieved comparable results to the pre-trained model G2GT. For instance, on the OOD dataset, LEFTNet achieved an MAE of 3.27 for dielectric constant and 12.97 mPa s for viscosity, while G2GT had an MAE of 3.31 and 13.28 mPa s, respectively. These results suggest that physical properties like dielectric constant and viscosity are also structure-sensitive properties, and LEFTNet can serve as an excellent 3D information extractor.

Furthermore, EMol-QSPR was developed as an ensemble model to leverage the complementary strengths of the best-performing 2D and 3D models. As demonstrated in the benchmark results, G2GT and LEFTNet emerged as the top performers in their respective categories. G2GT excelled in 2D topological information extraction through USPTO pre-training, and LEFTNet showcased superior 3D geometric representation through its advanced LSE and FTE modules. By averaging these two high-performing models, EMol-QSPR achieved a remarkable relative error of 0.94, demonstrating significant improvements over individual models, i.e., approximately 3% over G2GT and 4.4% over LEFTNet. The ensemble's effectiveness was particularly evident in challenging predictions, such as dielectric constant and viscosity in the OOD dataset. For the dielectric constant, EMol-QSPR achieved an MAE of 3.17, improving upon G2GT's 3.31 and LEFTNet's 3.27. Similarly, EMol-QSPR's MAE of 12.83 mPa s outperformed both G2GT (13.28 mPa s) and LEFTNet (12.97 mPa s). These results validate the strategic selection of G2GT and LEFTNet as ensemble base models, effectively combining their complementary strengths in 2D and 3D molecular representation to enhance predictive accuracy and robustness.

Similarity Query Results

Based on the above reliable QSPR models, a similarity query method has been developed to design new molecules with similar characteristics to current promising molecules, such as FEC used as electrolyte additives for Li metal anodes. The capability of the similarity query method is ensured by a comprehensive database containing 280000 molecules, carefully curated through multiple screening stages. The screening parameters include thermodynamic stability filtering (formation energy <0 eV atom^{-1}) and synthetic accessibility evaluation (RAScore > 0.9[36]).

Taking ethyl methyl carbonate (EMC), a widely used commercial electrolyte solvent, as an example, several new solvent molecules, including both linear carbonates and branched variants, can be screened out (Figure 5). The new molecules share similar characteristics to EMC ($E_b = -1.53$ eV, $\varepsilon = 1.34$, $\eta = 5.5$ mPa s, HOMO = −7.35 eV, and LUMO = 0.02 eV).

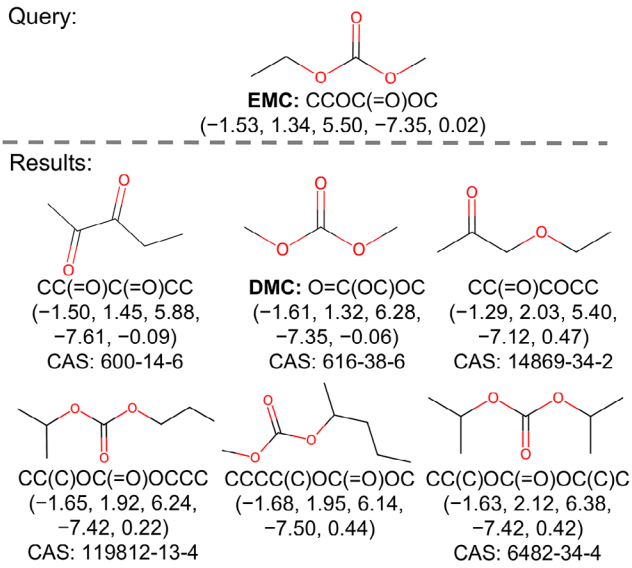

Figure 5. The query results of the EMC molecule. For each molecule, a 2D topology and a SMILES string are provided. The number of bottom the molecules is property vector ordered as binding energy (eV), dielectric constant, viscosity (mPa s), HOMO (eV), and LUMO (eV).

Despite structural variations, all retrieved molecules maintain similar binding energies (from −1.29 to −1.68 eV) and HOMO energy levels (from −7.12 to −7.61 eV), suggesting these properties are primarily determined by the carbonate functional group. Especially, dimethyl carbonate, which is widely adopted in commercialized electrolytes and is experimentally recognized to be similar to EMC, was successfully identified by such a query process. Besides, the presence of CAS Registry Numbers for most retrieved molecules indicates that they have been officially registered and are likely commercially available, providing strong validation for the practical utility of our search algorithm.

Property-Targeted Generation Results

Beyond the above query method, an AI generation model was further developed to design property-targeted molecules out of the embedded electrolyte database. Three distinct molecular generation tasks were specially investigated to demonstrate their practicability. 1) The generation of electrolyte molecules targeting a specific HOMO–LUMO gap; 2) the generation of molecules characterized by low binding energy and specific chemical formulas; 3) the generation of molecules exhibiting structural similarity to user-provided templates, represented as fingerprints.

Given the inherent stochasticity and potential instability associated with AI-generated molecules, an automated molecular cleaning workflow was established. The workflow serves as a critical post-generation filter, ensuring the quality and reliability of the generated molecular candidates. The workflow sequentially removes molecules that fail topological checks, indicating structural inconsistencies; duplicate molecules, eliminating redundancy; molecules already present in the training and validation sets, ensuring novelty; and

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

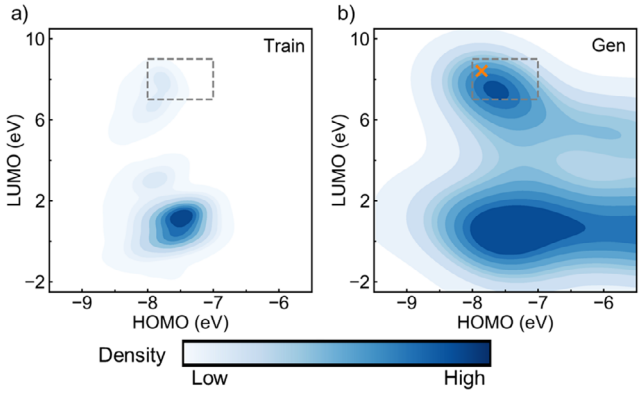

Figure 6. The out-of-domain performance of the generative model with respect to HOMO–LUMO properties. a) The training dataset is sparse in the targeted area and does not contain the DME molecule. b) The generated dataset is relatively dense in the targeted area and includes DME molecule.

molecules with low synthetic accessibility scores (RAScore <0.8), prioritizing molecules amenable to practical synthesis.

The evaluation of the generated molecules employs a suite of metrics tailored to the specific task. For property-targeted tasks (Tasks 1 and 2), SOTA pre-trained models were utilized to predict the targeted properties, and the distribution of these properties is compared to that of the training set. The comparison provides a quantitative measure of the model's ability to generate molecules with the desired characteristics. For the fingerprint-targeted task (Task 3), the Tanimoto coefficient[37] serves as the primary metric, quantifying the structural similarity between the fingerprints of the generated molecules and the target fingerprints. The coefficient provides a direct assessment of the model's success in replicating the desired structural features.

The evaluation of property-targeted generation commences with Task 1, focusing on the generation of molecules with a predefined HOMO and LUMO values. Two distinct generative model architectures were investigated: diffusion models, specifically the Equivariant Diffusion Model (EDM),[38] and autoregressive models represented by the conditional Generative-Schnet (cG-Schnet).[39] Initial comparative analysis reveals that the classifier-guided EDM demonstrates superior performance in generating molecules with targeted HOMO–LUMO gaps (Figure S2). Accordingly, DME, a molecule absenting from the original dataset yet residing in a relatively sparse region of the HOMO–LUMO gap distribution was considered (Figure 6a). When the model is tasked with generating molecules within this region, the resulting molecules exhibit a concentrated distribution around the targeted area and successfully include DME (Figure 6b), highlighting the model's sensitivity to conditional guidance and its ability to explore and populate sparse regions of the chemical space. Besides, examples of other generated molecules are provided in Table S8. The out-of-domain performance, the ability of the model to generate molecules with properties beyond the scope of the training data, is a crucial indicator of its generalizability and practical utility in molecular design.

Tasks 2 and 3 involve generating molecules with low binding energy and specific chemical formulas and generating molecules with structural similarity to target fingerprints, respectively. Autoregressive models are the focus of this investigation, given the current limitations in readily adaptable diffusion models for these specific tasks. Within the autoregressive framework, variants of cG-Schnet were explored, specifically focusing on the impact of encoder architecture. The original encoder was replaced with LEFTNet architecture. The results indicate that cG-LEFTNet exhibits comparable performance on Task 2 and superior performance on Task 3 (Figures S3 and S4, Table S3). The enhanced performance is likely attributable to the superior encoding capabilities of LEFTNet, enabling it to more effectively capture and represent the intricate relationships between molecular structure and properties.

Retrosynthesis Validation

Following the successful generation of molecules with targeted properties, the focus shifts to ensuring the practical feasibility of synthesizing these novel electrolyte candidates. A critical component of this feasibility assessment is the ability to predict viable synthetic pathways, a task typically addressed through retrosynthetic analysis. However, existing retrosynthesis methodologies are predominantly tailored for pharmaceutical molecules, relying heavily on template-based single-step predictors that often exhibit limited performance when applied to the structurally distinct domain of electrolytes. To overcome the limitation and enhance the applicability of retrosynthesis to electrolyte design, two key improvements were implemented.

First, the single-step predictor was replaced with the template-free model G2GT. The transition to a template-free approach is motivated by its potential for improved predictive capability outside the original training domain. To quantify the benefit of the substitution, a comparative analysis was performed against the established retrosynthesis method Askcos, utilizing an IID dataset comprising 100 electrolyte molecules. The performance of each method was evaluated based on its ability to correctly predict the immediate precursors (single-step prediction) for each target molecule, as measured by Top-k accuracy (where $k = 1, 2, 3, 4,$ and 5) and the number of molecules for which at least one correct precursor was identified within the Top-k predictions ("Number of Molecules Recalled").

The above comparison results on one-step retrosynthesis demonstrate that the G2GT predictor consistently outperforms Askcos in terms of Top-k accuracy across all values of k (Table S4). Specifically, G2GT achieved a Top-1 accuracy of 0.529 compared to Askcos's 0.452, indicating a higher probability of identifying the correct precursor in the first prediction. Furthermore, when evaluating the complete retrosynthetic planning, ASKCOS finds a smaller number of molecule retrosynthetic routes (8) compared to the G2GT-based retrosynthesis planner (17), suggesting G2GT can cover more electrolyte molecules. The combination of the G2GT single-step predictor and the Askcos path planner, termed

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

G2GT-Askcos (Figure 3), leverages the strengths of both approaches, combining accurate single-step predictions with efficient path planning capabilities.

Second, to further enhance the performance of the retrosynthetic predictor within the electrolyte domain, the original USPTO reaction dataset was augmented. A subset of 1500 reactions was extracted from a larger pool of 1.1 million reactions in Reaxys, based on the structural similarity between the reaction products and common electrolyte molecules. The similarity was quantified using ECFP[40] and the cosine similarity function. The selected reactions exhibit a minimum product similarity of 0.53 and an average similarity of 0.65, ensuring that the augmented dataset is relevant to the target chemical space.

Following fine-tuning with the augmented dataset, the G2GT single-step predictor demonstrates improved adaptation to the electrolyte domain. The adaptation was exemplified through a case study involving the retrosynthetic analysis of fluorinated diethoxyethane derivatives (F3DEE, F4DEE, F5DEE, and F6DEE), which were first reported in 2022. Importantly, the training dataset, i.e., USPTO, was published in 2016, effectively eliminating the possibility of data leakage and demonstrating the model's generalization capabilities. The reactions proposed by G2GT-Askcos for these molecules closely resemble those reported in the literature,[41] validating the efficacy of the proposed retrosynthesis framework (Figure 7). In each case, the predicted reagents are indicated above the reaction arrow, and the corresponding reaction conditions are noted below. Another retrosynthesis case is the only molecule missing a CAS ID in the EMC query result (SMILES: CCCC(C)OC(=O)OC) (Figure S5). The prices of the starting materials are indicated above the substances (Figure 7). Through cost analysis of different pathways, it was found that the synthesis costs of F3DEE, F4DEE, and F6DEE are relatively high. Besides, the key raw materials for these pathways are expensive, making it difficult to reduce the overall synthesis costs. The synthesis cost of F5DEE is lower, achieving an economically efficient synthesis.

The successful prediction of plausible synthetic pathways for these representative electrolyte molecules underscores the potential of the G2GT-Askcos framework to facilitate the experimental validation and subsequent synthesis.

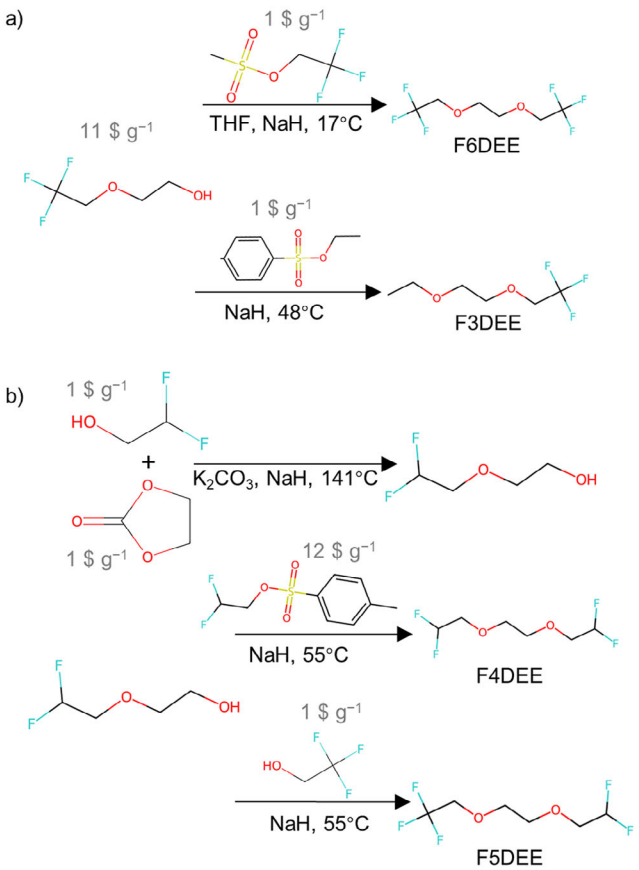

Figure 7. The retrosynthesis results of a) F6DEE and F3DEE, b) F4DEE and F5DEE. The proposed reagents are de-noted above the reaction arrow, while the reaction conditions are labeled underneath.

SEI Formation Analysis

Following the validation of synthetic accessibility, the investigation turns to predict the impact of the designed electrolyte molecules on battery performance, specifically focusing on the formation of SEI. Understanding SEI composition is crucial for designing electrolytes that enable stable battery cycling. To predict SEI formation pathways, the EMolNetKnittor module was developed.

Compared with previous approaches,[26] EMolNetKnittor offers significant advancements, including the integration of an expanded dataset encompassing the entire LiBE electrolyte database[27] and a fully automated workflow for custom database construction by users (Figure 4). These improvements enable the analysis of SEI formation across electrolyte chemistries with great diversity, including those containing F, N, P, and S. Additionally, EMolNetKnittor delivers high flexibility and scalability, allowing researchers to explore SEI formation pathways comprehensively and efficiently. By leveraging these capabilities, EMolNetKnittor provides deep insights into SEI formation mechanisms, paving the way for designing next-generation electrolytes for stable battery cycling.

To demonstrate the utility of EMolNetKnittor for predicting SEI formation pathways, the decomposition of FEC was focused on as an example as it is a widely used electrolyte additive known to promote the formation of stable SEI layers (Figure 8). The analysis begins with the initial coordination of FEC with a Li ion (Li-FEC). EMolNetKnittor predicts that the transition from a single coordination state to a double coordination state (Figure 8a) is thermodynamically favorable. The prediction aligns with previous computational and experimental studies[42] that have identified the double coordination state of Li-FEC as a key intermediate during the formation process of SEI. Following the formation of the stable double coordination complex, EMolNetKnittor predicts a ring-opening reaction of Li-FEC (Figure 8b), leading to the formation of an open-chain intermediate. This intermediate then interacts with another double-coordinated Li-FEC molecule, resulting in the formation of a stable dimeric

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

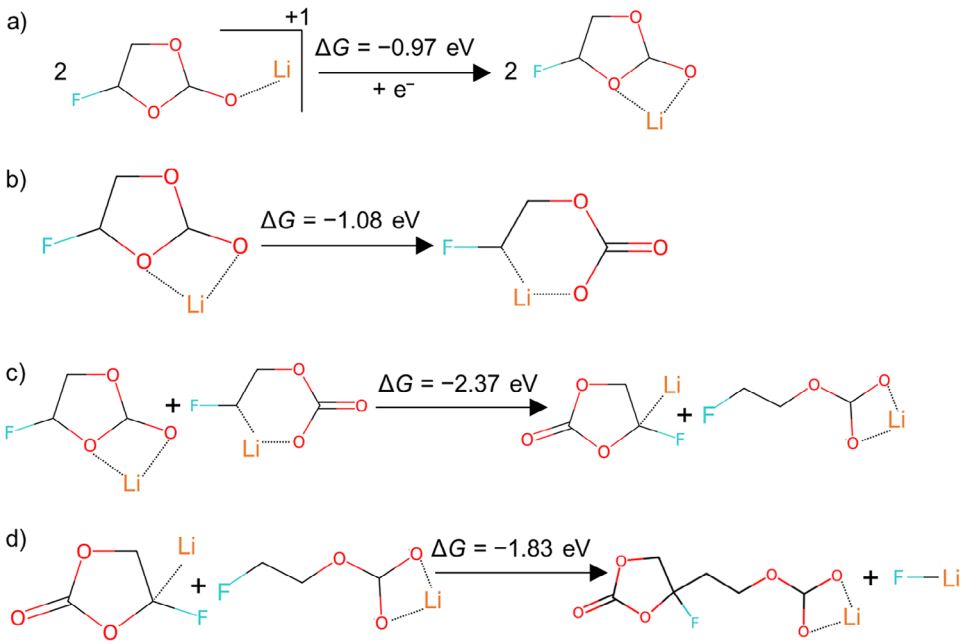

Figure 8. The retrosynthesis results of a) F6DEE and F3DEE, b) F4DEE and F5DEE. The proposed reagents are de-noted above the reaction arrow, while the reaction conditions are labeled underneath.

intermediate (Figure 8c). Finally, these dimeric intermediates undergo further reactions to form a polymeric SEI component and LiF (Figure 8d). The predicted polymeric structure is consistent with experimental observations of SEI layers formed from FEC-containing electrolytes, which often exhibit a significant fraction of polymeric and oligomeric species.

Another case study of a new molecule (SMILES: O=C1COC(=O)O1) was considered for integrating the three functions of the Uni-Electrolyte platform. The new molecule was generated by AI models taking EC as the target. After multi-criteria screening including topology check, duplication removement, and stability checking, the top ten molecules were confirmed and ranked according to their similarity to the EC molecule (Figure S6 and Table S6). The EMolForger module further predicted the retrosynthesis pathway and reaction conditions of the new molecule, and the EMolNetKnittor module predicted possible decomposition pathways in the presence of lithium oxide. A ring-opening reaction mechanism was predicted, and the Gibbs free energy of the reaction is around −4.98 eV, indicating a strong trend for such a reaction.

The case studies vividly illustrate the potential of EMolNetKnittor to provide valuable insights into the complex decomposition pathways of electrolyte molecules and the resulting SEI composition. The ability to predict the potential SEI products formed under operational conditions offers a powerful tool for guiding the design and optimization of electrolyte formulations, enabling the development of batteries with enhanced performance and longevity. Furthermore, the automated database construction and expansion capabilities of EMolNetKnittor make it readily adaptable to the exploration of novel electrolyte chemistries and the investigation of SEI formation in emerging battery technologies.

Conclusion

An AI platform, namely the Uni-Electrolyte, has been developed to design battery electrolyte molecules based on a data-driven approach. The Uni-Electrolyte integrates three synergistic modules, i.e., EMolCurator, EMolForger, and EMolNetKnittor. EMolCurator, the AI-assisted molecular design framework, leverages QSPR models, multi-criterion screening, similarity search, and AI-driven molecular generation to efficiently explore the vast chemical space and identify promising electrolyte candidates with tailored properties. EMolForger, the AI-powered retrosynthetic analysis module, bridges the gap between computational design and experimental synthesis by providing optimized synthetic routes and detailed reagent information. Last but not least, EMolNetKnittor, the comprehensive SEI formation analysis platform, enables detailed investigation of the complex interfacial processes that govern battery performance, providing crucial insights into the SEI composition and formation mechanisms. The integration of the three modules within the Uni-Electrolyte platform affords a powerful and versatile toolset, enabling an AI-driven design of advanced electrolytes for next-generation batteries.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (T2322015, 92472101, 22393903, 22393900, and 52394170), the National Key Research and Development Program (2021YFB2500300), and the Beijing Municipal Natural Science Foundation (L247015 and L233004).

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: Artificial intelligence · Electrolyte molecules · High-throughput design · Machine learning · Rechargeable batteries

[1] N. Yao, X. Chen, Z.-H. Fu, Q. Zhang, *Chem. Rev.* **2022**, *122*, 10970–11021.
[2] X. Chen, Q. Zhang, *Acc. Chem. Res.* **2020**, *53*, 1992–2002.
[3] J.-G. Zhang, W. Xu, J. Xiao, X. Cao, J. Liu, *Chem. Rev.* **2020**, *120*, 13312–13348.
[4] X. Fan, C. Wang, *Chem. Soc. Rev.* **2021**, *50*, 10486–10566.
[5] Y. Yamada, J. Wang, S. Ko, E. Watanabe, A. Yamada, *Nat. Energy* **2019**, *4*, 269–280.
[6] M. Winter, B. Barnett, K. Xu, *Chem. Rev.* **2018**, *118*, 11433–11456.
[7] K. Xu, *Chem. Rev.* **2014**, *114*, 11503–11618.
[8] K. Xu, *Chem. Rev.* **2004**, *104*, 4303–4418.
[9] X.-B. Cheng, R. Zhang, C.-Z. Zhao, Q. Zhang, *Chem. Rev.* **2017**, *117*, 10403–10473.
[10] X. He, D. Bresser, S. Passerini, F. Baakes, U. Krewer, J. Lopez, C. T. Mallia, Y. Shao-Horn, I. Cekic-Laskovic, S. Wiemers-Meyer, F. A. Soto, V. Ponce, J. M. Seminario, P. B. Balbuena, H. Jia, W. Xu, Y. Xu, C. Wang, B. Horstmann, R. Amine, C.-C. Su, J. Shi, K. Amine, M. Winter, A. Latz, R. Kostecki, *Nat. Rev. Mater.* **2021**, *6*, 1036–1052.
[11] H. Wang, Z. Yu, X. Kong, S. C. Kim, D. T. Boyle, J. Qin, Z. Bao, Y. Cui, *Joule* **2022**, *6*, 588–616.
[12] X. Yang, Y. Wang, R. Byrne, G. Schneider, S. Yang, *Chem. Rev.* **2019**, *119*, 10520–10594.
[13] K. M. Jablonka, D. Ongari, S. M. Moosavi, B. Smit, *Chem. Rev.* **2020**, *120*, 8066–8129.
[14] T. Lombardo, M. Duquesnoy, H. El-Bouysidy, F. Årén, A. Gallo-Bueno, P. B. Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide, M. Reynaud, J. Carrasco, A. Grimaud, C. Zhang, T. Vegge, P. Johansson, A. A. Franco, *Chem. Rev.* **2022**, *122*, 10899–10969.
[15] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, W. C. Chueh, *Nature* **2020**, *578*, 397–402.
[16] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547–555.
[17] B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams, A. G. Doyle, *Nature* **2021**, *590*, 89–96.
[18] C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka, T. Xie, *Nature* **2025**, *639*, 624—632.
[19] I.-B. Magdău, D. J. Arismendi-Arrieta, H. E. Smith, C. P. Grey, K. Hermansson, G. Csányi, *npj Comp. Mater.* **2023**, *9*, 146.
[20] S. Dajnowicz, G. Agarwal, J. M. Stevenson, L. D. Jacobson, F. Ramezanghorbani, K. Leswing, R. A. Friesner, M. D. Halls, R. Abel, *J. Phys. Chem. B* **2022**, *126*, 6271–6280.
[21] S. Gong, Y. Zhang, Z.-H. Mu, Z. Pu, H. Wang, Z. Yu, M. Chen, T. Zheng, Z. Wang, L. Chen, X. Wu, S. Shi, W. Gao, W. Yan, L. Xiang, ArXiv **2024**, https://arxiv.org/abs/2404.07181, https://www.nature.com/articles/s42256-025-01009-7.
[22] Y.-C. Gao, Y.-H. Yuan, S. Huang, N. Yao, L. Yu, Y.-P. Chen, Q. Zhang, X. Chen, *Angew. Chem. Int. Ed.* **2025**, *64*, e202416506.
[23] Y.-C. Gao, N. Yao, X. Chen, L. Yu, R. Zhang, Q. Zhang, *J. Am. Chem. Soc.* **2023**, *145*, 23764–23770.
[24] Z. Lin, S. Yin, L. Shi, W. Zhou, Y. J. Zhang, *J. Chem. Inf. Model.* **2023**, *63*, 1894–1905.
[25] C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, K. F. Jensen, *Science* **2019**, *365*, eaax1566.
[26] D. Barter, E. W. Clark Spotte-Smith, N. S. Redkar, A. Khanwale, S. Dwaraknath, K. A. Persson, S. M. Blau, *Digit. Discov.* **2023**, *2*, 123–137.
[27] E. W. C. Spotte-Smith, S. M. Blau, X. Xie, H. D. Patel, M. Wen, B. Wood, S. Dwaraknath, K. A. Persson, *Sci. Data* **2021**, *8*, 203.
[28] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887–2893.
[29] G. Zhou, Z. Gao, Q. Ding, H. Zheng, H. Xu, Z. Wei, L. Zhang, G. Ke, ICLR ChemRxiv, **2023**, https://openreview.net/forum?id=6K2RM6wVqKu.
[30] N. Schneider, N. Stiefl, G. A. Landrum, *J. Chem. Inf. Model.* **2016**, *56*, 2336–2346.
[31] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *Sci. Data* **2014**, *1*, 140022.
[32] W. Du, Y. Du, L. Wang, D. Feng, G. Wang, S. Ji, C. P. Gomes, Z. Ma, *ArXiv* **2023**, https://arxiv.org/abs/2304.04757.
[33] K. T. Schütt, O. T. Unke, M. Gastegger, in Int. Conf. on Machine Learning PLMR, New York **2021**, 9377-9388, https://proceedings.mlr.press/v139/schutt21a.html.
[34] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, K.-R. Müller, *J. Chem. Phys.* **2018**, *148*, 241722.
[35] Y. Liu, L. Wang, M. Liu, X. Zhang, B. Oztekin, S. Ji, ICLR **2021**, https://openreview.net/forum?id=givsRXsOt9r.
[36] A. Thakkar, V. Chadimová, E. J. Bjerrum, O. Engkvist, J.-L. Reymond, *Chem. Sci.* **2021**, *12*, 3339–3349.
[37] D. Bajusz, A. Rácz, K. Héberger, *J. Cheminformatics* **2015**, *7*, 20.
[38] E. Hoogeboom, V. G. Satorras, C. Vignac, M. Welling, *ArXiv* **2022**, https://arxiv.org/abs/2203.17003.
[39] N. W. A. Gebauer, M. Gastegger, S. S. P. Hessmann, K.-R. Müller, K. T. Schütt, *Nat. Commun.* **2022**, *13*, 973.
[40] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
[41] Z. Yu, P. E. Rudnicki, Z. Zhang, Z. Huang, H. Celik, S. T. Oyakhire, Y. Chen, X. Kong, S. C. Kim, X. Xiao, H. Wang, Y. Zheng, G. A. Kamat, M. S. Kim, S. F. Bent, J. Qin, Y. Cui, Z. Bao, *Nat. Energy* **2022**, *7*, 94–106.
[42] S. M. Blau, H. D. Patel, E. W. C. Spotte-Smith, X. Xie, S. Dwaraknath, K. Persson, *Chem. Sci.* **2021**, *12*, 4931–4939.

Manuscript received: February 07, 2025
Revised manuscript received: April 05, 2025
Accepted manuscript online: April 15, 2025
Version of record online: May 05, 2025

© 2025 Wiley-VCH GmbH

15213773, 2025, 30, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202503105 by Nankai University, Wiley Online Library on [25/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License