HIERARCHICAL MOIETY-AWARE GRAPH TRANSFORMER FOR LI-METAL ELECTROLYTE FORMULATION DESIGN

Luca Mondonico[*]
Department of Materials Science and Engineering
Stanford University
Stanford, CA 94305
lmondo@stanford.edu

Hance Su[*]
Department of Materials Science and Engineering
Stanford University
Stanford, CA 94305
hancesu@stanford.edu

Jacob Florian
Department of Chemical Engineering
Stanford University
Stanford, CA 94305
florianj@stanford.edu

Hao Lyu
Department of Chemical Engineering
Stanford University
Stanford, CA 94305
haolyu@stanford.edu

Zhenan Bao
Department of Chemical Engineering
Stanford University
Stanford, CA 94305
zbao@stanford.edu

Yi Cui
Department of Materials Science and Engineering
Stanford University
Stanford, CA 94305
yicui@stanford.edu

ABSTRACT

Maximizing Coulombic efficiency (CE) is critical in improving battery life and performance, although existing predictive models overlook the very chemical design strategies that drive electrolyte performance. Here, we introduce HELENA (Hierarchical Electrolyte Learning with Embedded Neural Attention), the first electrolyte deep-learning framework to encode moiety-level chemistry and cooperative interactions across full formulations. HELENA constructs a three-tiered representation and employs attention mechanisms to learn both established heuristics and potentially unseen design principles. Trained on a substantially expanded CE dataset of lithium metal batteries, HELENA outperforms state-of-the-art graph- and tree-based models, while its attention mechanisms both recover established electrolyte trends and reveal novel moiety co-dependencies. Applying HELENA as an in-silico screening engine, we experimentally validate several model-predicted formulations achieving CE > 99.5%, demonstrating its practical utility. By fusing moiety-aware tokenizer with scalable attention-based learning, HELENA establishes a new paradigm for rational electrolyte discovery and paves the way for generative, closed-loop design workflows of exploring multicomponent chemical systems beyond energy.

---

[*]These authors contributed equally to this work.

1 Main

The accelerating electrification of transportation, grid storage, and data-intensive hardware—including next-generation artificial-intelligence hardware systems—is driving demand for batteries that surpass the energy-density ceiling of conventional Li-ion technology [1, 2]. This imperative has sparked renewed interest in lithium metal batteries (LMBs), prized for their exceptionally high theoretical specific capacity (3860 mAh/g) and low reduction potential, promising unprecedented energy density [3]. Yet, widespread adoption of LMBs remains severely constrained by critical challenges associated with the lithium metal anode, notably uncontrollable dendritic growth and detrimental side reactions with electrolytes [4, 5]. These issues compromise cycle life and safety. Coulombic efficiency (CE), one of the most important performance parameters, describes the ratio of discharge to charge capacity in each cycle and captures all faradaic side reactions in a single metric. Maximizing it is therefore essential for realizing long-lived, high-energy LMBs for practical use [6].

Electrolyte engineering has recently delivered striking gains in lithium-metal stability—ranging from high-concentration [7] and localized-high-concentration chemistries [8, 9, 10, 11, 12], to fluorination-tuned [13, 14] and high-entropy electrolytes [15, 16]—all of which register as improvements in CE through better SEI integrity, solvent resilience, and uniform Li plating/stripping. Yet collecting CE data is costly: weeks-long cycling tests yield only a handful of measurements, leaving the combinatorial search space of $\geq 10^{10}$ small molecules (and combinatorially far more multi-component blends) essentially unexplored. Efficient data-driven prediction is therefore the bottleneck. Previous work from our group by Kim et al. [17] have demonstrated the potential of coarse molecular descriptors like solvent oxygen ratio (sO) for CE predictions. Subsequently, the adaption of graph-based machine learning models on our group's dataset have progressively pushed accuracy further, mirroring the success of GNN in other small molecules tasks such as antibiotics design [18] and olfactory perception representation [19]. Nevertheless, two significant gaps persist. First, the existing datasets must be updated to reflect the rapidly expanding corpus of lithium metal electrolyte molecules and formulations. Second, current architectures (GCNs [20], MPNNs [21], even naïvely repurposed Transformers [22, 23]) treat each component in isolation, neglecting the cooperative interactions that battery chemists deliberately exploit. Bridging these gaps demands a modeling framework that scales with new data and encodes electrolyte-design priors. Addressing these gaps is crucial for the next-generation modeling frameworks necessary for accelerating electrolyte discovery and innovation.

In this work, we present HELENA (Hierarchical Electrolyte Learning with Embedded Neural Attention), the first deep-learning framework to apply attention explicitly at the moiety level for lithium-metal electrolyte design. HELENA constructs a hierarchical representation spanning moieties, whole molecules, and full-formulation context, enabling the model to learn both the chemistry of individual functional groups and the cooperative phenomena that arise in multi-component blends. Trained on our newly curated dataset that triples the previously available CE records, the model surpasses state-of-the-art graph model baselines, produces attention score trends that recover established electrolyte-design heuristics, and pinpoints candidate formulations predicted to achieve CE values above 99.5% and above. Together, these advances establish a scalable, chemistry-aware framework for rational electrolyte formulation discovery and optimization.

2 Overall Workflow of HELENA

2.1 Model Architecture

HELENA employs an innovative hierarchical graph transformer architecture explicitly tailored for battery electrolyte systems, capitalizing on moiety-level chemical information to address limitations inherent in conventional GCNs or MPNNs. Traditional GCN models often suffer from over smoothing [24], particularly when applied to smaller, structurally simpler electrolyte molecules, leading to diminished predictive performance. In contrast, HELENA leverages detailed chemical insights by decomposing each molecule into fundamental building blocks (moieties), each uniquely contributing to cycling performance. For example, chemists have explored LiTFSI analogues in which one trifluoromethanesulfonyl group is replaced by a glyme-like moiety [25], a sulfonyl moiety is converted to a sulfinyl moiety [26], or the central imide nitrogen is methylated to yield a TFSI-inspired solvent PreTFSI [27]. Such targeted, moiety-level edits exemplify the design space HELENA is built to capture. Our tokenizer inspired by these design strategies enables HELENA to effectively capture complex chemical interactions across multiple scales—specifically at the intramolecular moiety, intermolecular solvent–solute interaction, and formulation levels. The graph transformer architecture employed here is particularly advantageous as it utilizes attention mechanisms [28], effectively capturing long-range dependencies and intricate interactions within molecular graphs, which significantly enhances predictive accuracy in chemical modeling tasks.
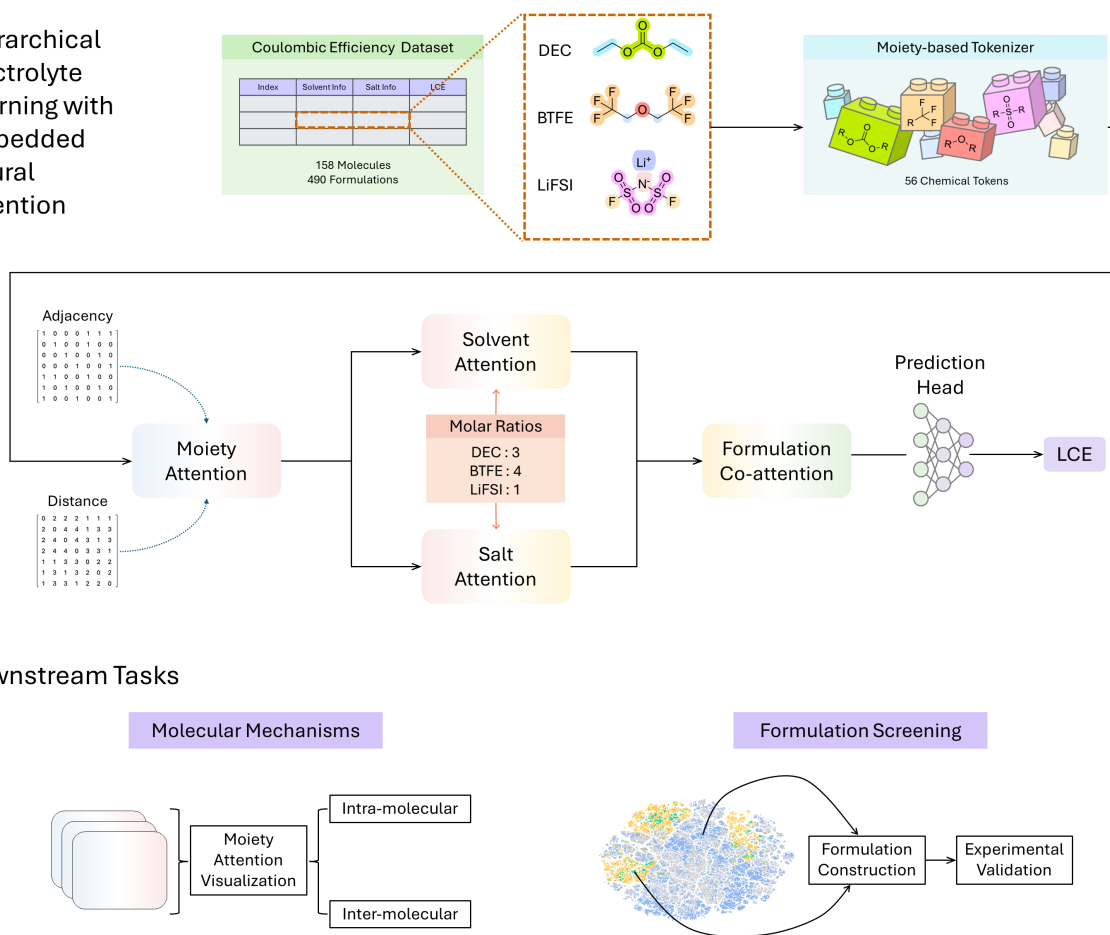
2

Figure 1: Schematic overview of HELENA workflow.

HELENA is trained on a dataset of 490 electrolyte formulations with corresponding log-transformed Coulombic efficiency (LCE) values. For a representative formulation (1 LiFSI: 3 DEC: 4 BTFE), each molecule is tokenized into fundamental moieties (colored blocks) using SMARTS-based pattern recognition, and positional relationships are encoded via adjacency and shortest-path distance matrices. Moiety embeddings are generated through a multi-head self-attention block and aggregated into molecule-level representations—incorporating molar ratio information to distinguish solvent and salt components. These representations are fused by a formulation-level co-attention module, and a feedforward multilayer perceptron yields the final LCE prediction. Downstream analyses visualize moiety-level attention scores to reveal intra- and inter-molecular design trends, and an in-silico screening pipeline identifies candidate formulations with predicted CE > 99.5%, which are then experimentally validated.

The model pipeline initiates (**Figure 1**) with a custom tokenizer employing intelligent SMARTS pattern recognition on SMILES input strings, identifying 56 fundamental moieties prevalent in electrolyte formulations documented since the 1980s. An exemplary electrolyte formulation (LiFSI in DEC:BTFE) is illustrated within the workflow to clearly demonstrate the tokenization process. These identified moieties primarily include fluorinated groups, ethers, esters, and newer, emerging classes such as nitriles [29] or silyl [30, 31]-based functionalities. Notably, by dissecting molecules into a concise set of moieties rather than treating every atom individually, HELENA dramatically reduces the number of components per formulation and thereby slashes the computational cost of its attention modules.

Following tokenization, moiety-level features are processed through attention blocks employing both local and global graph representations, enabling the comprehensive capture of intra-molecular substructure interactions. Subsequently, these moiety-level outputs are aggregated at the molecule level to individually represent solvent and salt interactions. This step is augmented by integrating explicit compositional information, specifically molar ratios of solvents and salts, enhancing the model's ability to understand multi-component electrolyte interactions.

At the formulation stage, HELENA utilizes a sophisticated co-attention mechanism enriched by deep supervision to integrate solvent- and salt-level embeddings. This yields a comprehensive, chemically informative formulation-level embedding. Finally, the embedding feeds into a multi-layer perceptron (MLP) to output predictions of the log-transformed coulombic efficiency (LCE). The use of log scaling on CE, consistent with our prior methodologies [17], effectively distinguishes high-performing formulations, especially those achieving CE near or exceeding 99%, guiding targeted experimental validation and electrolyte optimization.

2.2 Downstream Tasks

After model training, we illustrate how molecular mechanisms can be uncovered through analysis of attention weights. By aggregating and visualizing these attention scores on a carefully selected set of molecules, we demonstrate that the model effectively learns established solvating power trends among various ether-based solvents. Furthermore, by revisiting experimental data from our group's previous publication, we uncover potential molecular interactions beyond the previously identified solvent oxygen ratio. To extend our findings, we applied a filtering strategy on PubChem compounds [32] to select molecules composed exclusively of existing electrolyte moieties. This enabled us to construct a comprehensive inference dataset for identifying promising electrolyte formulations, both single salt single solvent and beyond capable of achieving CE values greater than 99.5%. A few of related studies [21, 33] predominantly focused on single-salt and single-solvent electrolytes, thus leaving much of the broader electrolyte design space unexplored.

2.3 Coulombic Efficiency Dataset Curation

A key factor contributing to the project's success is the substantial expansion of our electrolyte formulation dataset, increasing from an initial 152 data points from our previous dataset [17] to 490. Correspondingly, the number of unique molecules represented rose from 63 to 158, and the identified moieties expanded from 37 to 56. Deep learning methodologies have found considerable success in drug discovery applications, making it logical to transfer these established strategies into electrolyte molecular modeling. However, as illustrated in **Figure 2a**, electrolyte molecules differ significantly from typical drug-like molecules [34]: they have substantially fewer heavy atoms (average significantly below the drug-like benchmark of 30.6) and generally contain fewer ring systems (fewer than one on average). Furthermore, drug discovery models rarely consider interactions beyond binary systems (e.g., drug-drug [35] or protein-protein [36] interactions), whereas electrolyte systems inherently involve at least binary interactions (solvent and salt), with more than 60% of formulations in our dataset involving more than two components. These distinct characteristics underscore the necessity of a specialized modeling approach tailored specifically for electrolyte molecules, such as HELENA. Additionally, **Figure 2b**, illustrating moiety frequency across our expanded dataset, highlights an enriched diversity in moiety–moiety interactions, notably including previously overlooked groups such as benzene [37, 38], silyl [39], and chloro [40, 41] functionalities. Detailed information on our dataset curation process and complete references can be found in the **Supplementary Note 1** and dataset hosting page.
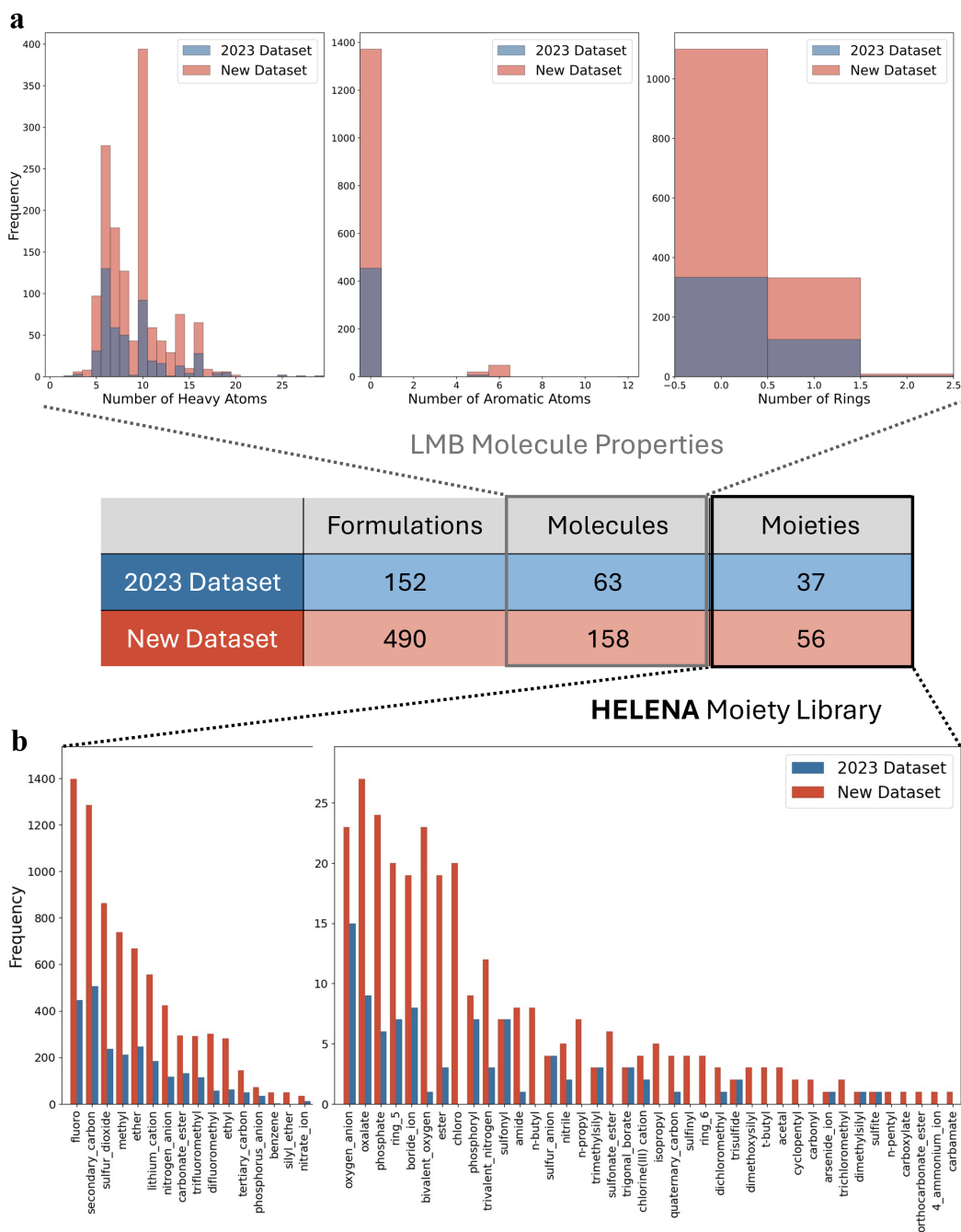
4

Figure 2: Dataset expansion and molecular property distributions.

The central table compares the 2025 and 2023 CE datasets, showing a threefold increase in formulations, more than a twofold rise in unique molecules, and a 50% expansion in identified moieties. **a.** Distributions of heavy-atom count and ring/aromatic content for the 158 molecules in the current dataset, highlighting an average heavy-atom count of ≈10 and a predominance of non-aromatic, acyclic structures. **b.** Frequency of each moiety across all formulations in 2025 versus 2023 (counts include multiple occurrences per formulation); every moiety is more abundant in the updated set, and new functional groups—benzene, silyl-ether, chloro, n-butyl, and n-propyl—appear for the first time.

5

3 Model Performances and Comparisons

To evaluate the performance of HELENA, we constructed and trained a comprehensive suite of baseline models, including traditional methods from our previous work—such as linear regression (LinReg), support vector machines (SVM), random forests (RF), and the advanced tree-based model XGBoost [42]—followed by some well-studied graph-based models including GCN, GAT [43] and MPNN [44]. All models were trained and evaluated on the latest CE dataset using 5-fold cross-validation. HELENA achieved the lowest average RMSE 0.212 and highest average R^2 of 0.784 across all experiments (**Figure 3a,b**). To ensure robustness, each experiment was repeated five times with different random seeds; notably, HELENA exhibited minimal variance (standard deviation of 0.012 and 0.025), indicating reduced sensitivity to data segmentation and initialization. Furthermore, acknowledging that certain low CE values—particularly those reported in the early days of electrolyte engineering—may be influenced by outdated measurement techniques, we incorporated a penalty score for LCE values below a predefined threshold. Based on our parity plot (**Figure 3c**), HELENA demonstrates superior performance on what is, to the best of our knowledge, the largest LMB CE dataset ever assembled.

In our analysis, baseline models utilizing traditional atomic-ratio–based molecular descriptors failed to capture the full structural complexity necessary for accurate CE prediction, resulting in limited generalization to unseen test data. We further implemented previously reported graph-based models, including a weighted GCN and MPNN, following methodologies from our prior studies. Although tree-based models achieved high performance on the training set, their extrapolative ability on the test set was limited—likely due to overfitting and the inadequate expressiveness of simple descriptors. Even when using sophisticated architectures like the MPNN, which conveys detailed bond-level information, the performance remained suboptimal compared to HELENA. Similarly, while GAT incorporate attention mechanisms, our evaluations indicate that their design does not capture mechanistic insights as effectively as our moiety-based attention modules, which will be detailed in the following section.

4 Hierarchical Attention Enhances Molecular Interpretability

4.1 Ether Moiety-Attention Score Reflects Solvating Power

While basic linear models are simple to train and offer high interpretability for guiding future designs, they often fall short in capturing complex molecular interactions. In contrast, black-box models such as GCNs and their derivatives may achieve high predictive accuracy but provide limited interpretability. Our moiety-based transformer architecture overcomes these challenges by leveraging carefully engineered attention modules that uncover critical design metrics. This approach, successfully applied in early-stage NLP, computer vision, and molecular design tasks (e.g., drug discovery [45]), enables us to probe the internal decision-making process of our model. Here, we adopt a similar approach to interpret the electrolyte formulation predictions generated by HELENA, focusing specifically on moiety-level attention scores. As a representative example, the attention scores for each attention head at the moiety level of Dimethoxyethane (DME), one of the most studied electrolyte solvents, are presented in **Supplementary Fig. 9**.

Building from DME, we systematically curated and analyzed a set of structurally related non-fluorinated ether solvent molecules. By increasing the side chain lengths on both sides of DME, we derive Diethoxyethane (DEE). Conversely, shortening the backbone yielded Dimethoxymethane (DMM) and Diethoxymethane (DEM). To further probe the effect of side chain length, we sequentially examined ethers with increasingly larger side chains, progressing from DMM and DEM to Dipropoxymethane (DPM) and Dibutoxymethane (DBM). Additionally, methylation of molecules DMM and DEM resulted in three additional derivatives, namely Dimethoxypropane (DMP), Dipropoxyethane (DPE), Diethoxypropane (DEP) (**Figure 4a**). It is well established in electrolyte chemistry that shortening backbone length [46], extending side chains [47], and introducing methyl groups [48, 49] typically reduce the strength of oxygen-lithium cation interactions, thereby lowering solvating power and subsequently impacting CE [50]. Remarkably, our model's attention score on the ether moieties independently captured these well-known chemical trends. While attention scores should not be interpreted as direct proxies for chemical properties—nor are they linearly correlated with CE—they provide a window into the model's internal prioritization of chemically relevant substructures. In this case, ether moieties exhibiting structural features known to modulate Li$^+$-O interactions, such as backbone length or side chain bulk, received differentiated attention across heads and layers. This suggests that HELENA is able to selectively focus on functional groups whose variations are historically known to impact electrolyte performance, even in the absence of handcrafted descriptors. Rather than serving as a simple indicator of contribution magnitude, the attention mechanism appears to enhance model performance by facilitating more effective feature disentanglement at the moiety level, improving representation quality and generalization across diverse formulations.

In another case study focusing on dual-solvent systems, we revisited and reanalyzed experimental formulations from our group's prior publication [17]. We started with Toluene:MBE (methyl-n-butyl ether) and then (i) replaced the linear
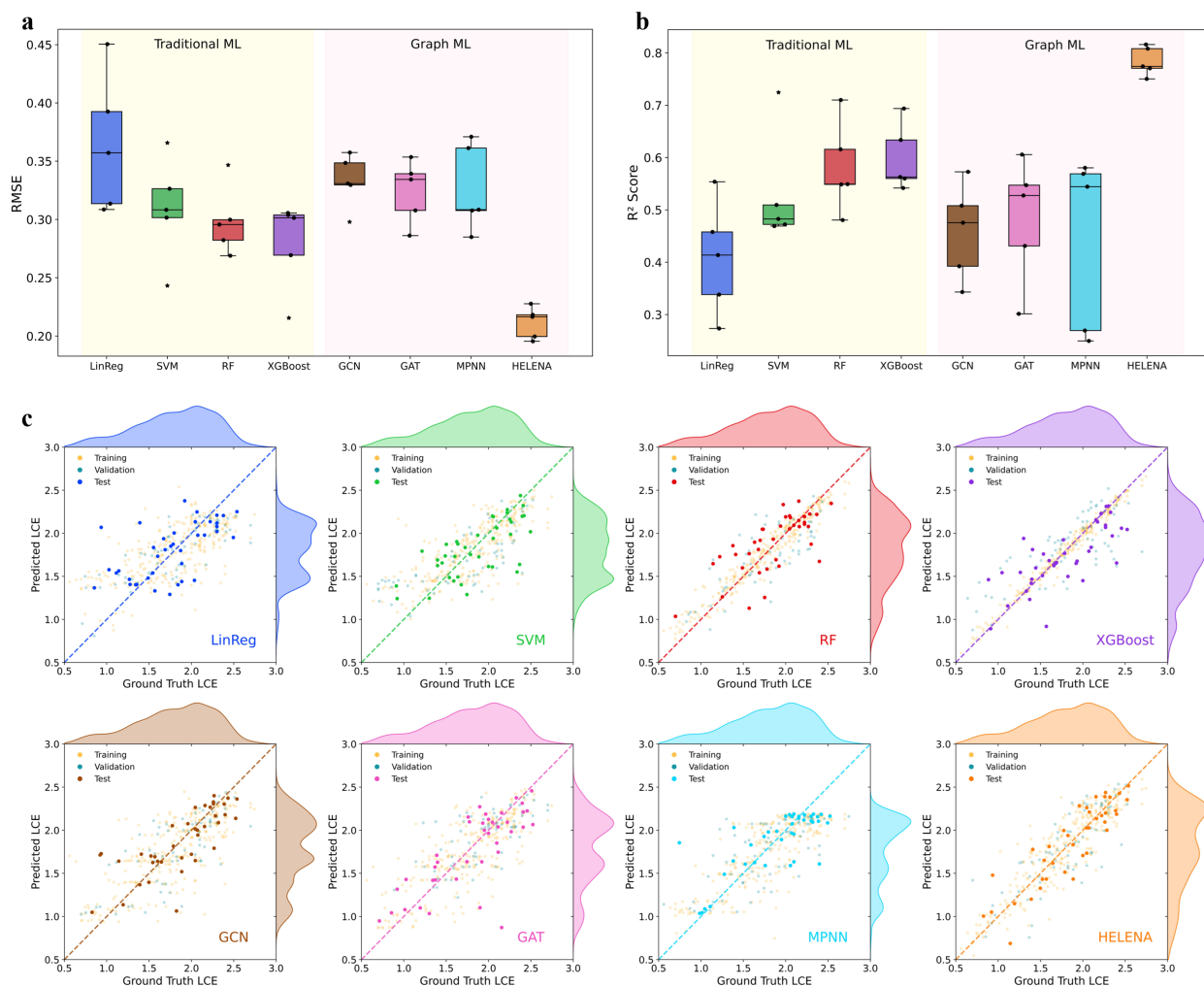
6

Figure 3: Performance evaluation of HELENA versus baseline models.

a. Distribution of root-mean-square error (RMSE) over five independent splits for traditional ML baselines (linear regression, SVM, random forest, XGBoost), graph-based models (GCN, GAT, MPNN), and HELENA; HELENA achieves the lowest mean RMSE and smallest variance. **b.** Distribution of coefficient of determination (R^2) across the same splits and models, with HELENA attaining the highest mean R^2 and minimal spread. **c.** Parity plot of predicted versus ground-truth log-transformed Coulombic efficiency (LCE) on held-out test data for all eight models in a representative run (test points highlighted), illustrating HELENA's closest agreement with experimental values.

7

n-butyl chain with a tert-butyl group to obtain Toluene:MTBE (methyl-tert-butyl ether) and (ii) swapped the methyl substituent for a second n-butyl chain to yield Toluene:DBE (di-n-butyl ether). These steric modifications corresponded to stepwise CE increases—from 99.51% (MBE) to 99.64% (MTBE) and 99.70% (DBE)—reflecting weaker Li^+-O interactions with bulkier ethers. We define the benzene–ether attention score as the mean cross-molecule attention weight between benzene rings and ether moieties, averaged across all heads and layers; this score decreased in lockstep with increasing steric hindrance, echoing the trend in the first case study, where attention scores for DME derivatives similarly captured backbone- and side-chain effects. Thus, HELENA's moiety-level attention recapitulates established steric effects and provides an interpretable metric for guiding electrolyte design beyond atomic ratios.

To further probe HELENA's mechanistic insights, we analyzed pairwise moiety attention weights for formulations predicted to achieve CE > 99.5% (**Supplementary Fig. 4a**). The nitrate moiety exhibited the highest average attention weight (0.51), in line with $LiNO_3$'s established role in forming Li_3N-rich SEI layers and enhancing cycle stability. Among non-solvating fragments, the benzene core—previously underrepresented in electrolyte datasets—showed a self-attention weight of 0.31. Notably, its functionalized derivatives received strong emphasis as well: benzene–fluoro interactions scored 0.34, and benzene–trifluoromethyl pairs 0.31. These elevated attention weights suggest that appending electron-withdrawing substituents to benzene confers partial solvating or interfacial benefits, thereby improving CE.

5   Screening Formulations from Inference Datasets

5.1   Small Molecule Filtering and Dataset Construction

Using HELENA as an in-silico screening engine, we first queried the entire PubChem collection ($\approx$111 million) and applied automated filters to remove invalid SMILES, non-organic structures, and multi-molecular entries. This reduced the pool to 1.63 million candidate molecules. Next, we tokenized each molecule and retained only those whose moieties were already represented in HELENA's vocabulary, yielding 243 thousand promising structures. From this refined library, we enumerated a large inference dataset spanning single-salt single-solvent to dual-salt ternary-solvent and yielding a total of 46 million formulations—and then applied expert-informed filters for commercial availability and reasonable toxicity (see **Supplementary Note 2**).
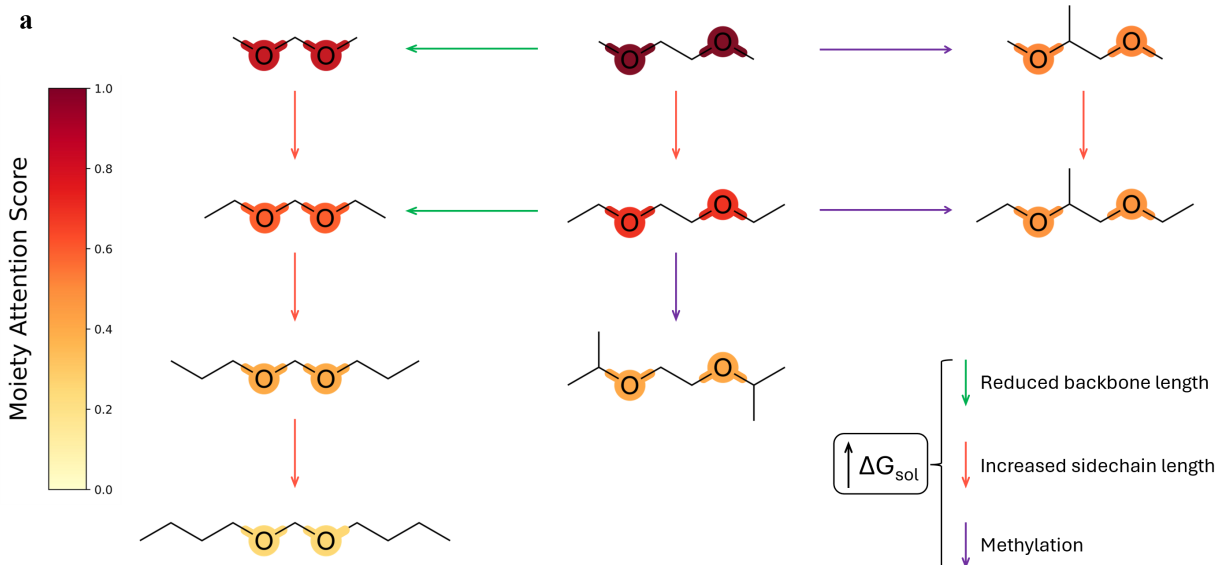
5.2   Experimental Validations

The inference and experimental results will be publicly disclosed upon publication.

6   Conclusion

In this work, we have introduced HELENA, the first deep-learning framework to apply moiety-level attention in predicting Coulombic efficiency (CE) across multi-component lithium-metal electrolytes. By constructing a three-tiered representation and training on a dataset that triples the available CE records, HELENA not only surpasses state-of-the-art graph-based baselines but also yields interpretable attention maps that recover and extend classic electrolyte-design heuristics. Through systematic interpretability studies, we demonstrated how HELENA's attention distinguishes key functional motifs—ethers, nitrates, phosphates, and fluorinated groups—and how co-attention reveals cooperative interactions that traditional descriptors overlook. Leveraging HELENA as an in-silico screening engine, we navigated a vast chemical landscape of over 46 million candidate formulations to pinpoint several novel solvent systems that experimentally delivered CE > 99.5%.

Looking ahead, HELENA lays the groundwork for generative electrolyte design: by coupling its chemically informed attention maps with optimization algorithms, entirely new moiety combinations tailored for various target performance metrics can be proposed. Integrating HELENA into closed-loop experimental workflows will accelerate the discovery of next-generation electrolytes—not only for lithium metal batteries but also for emerging multicomponent battery chemistries—ultimately enabling data-driven development of high-efficiency, high-energy-density storage solutions.
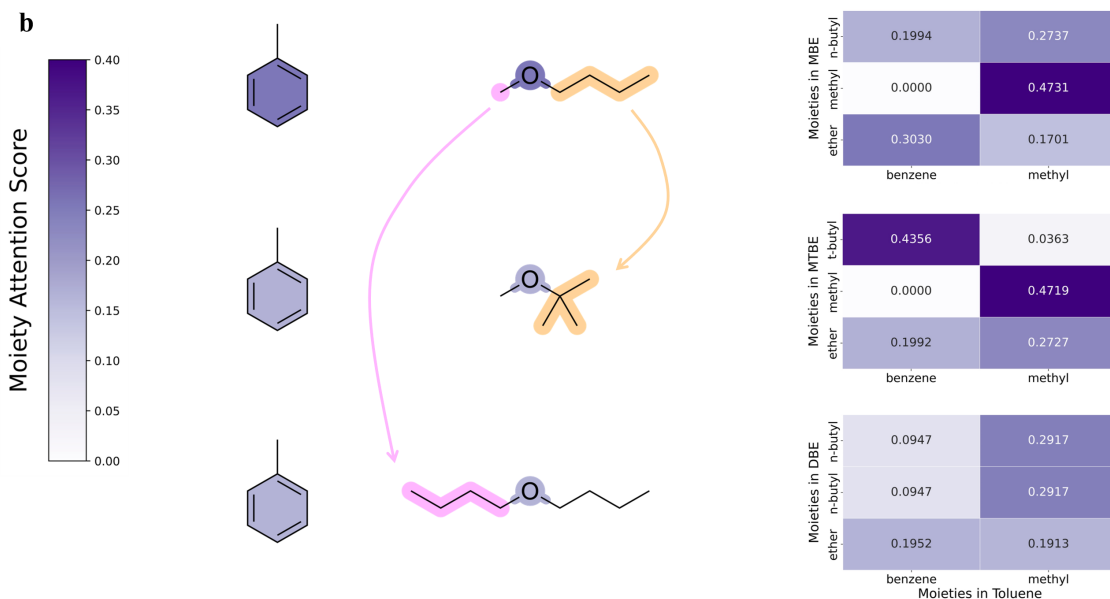
Figure 4: Case study of HELENA interpretability using moiety-level attention scores.

a. Non-fluorinated ether series organized by backbone and side-chain length: left column (top to bottom)—dimethoxymethane (DMM), diethoxymethane (DEM), dipropoxymethane (DPM), dibutoxymethane (DBM); center column—dimethoxyethane (DME), diethoxyethane (DEE), dipropoxyethane (DPE); right column—dimethoxypropane (DMP), diethoxypropane (DEP). The aggregated attention score on ether oxygen atoms—which coordinate Li^+—closely follows the experimentally determined solvation free-energy ranking. **b.** Dual-solvent formulations (Toluene:MBE, Toluene:MTBE, Toluene:DBE): benzene–ether attention scores are mapped onto each molecule's moieties, with numerical values shown at right, capturing the progressive decrease in attention as steric hindrance increases (same as in case study **a**).

7 Methods

7.1 Model Overview

The HELENA architecture consists of four primary components: (1) a moiety-level tokenizer that converts SMILES representations into functional group tokens, (2) a moiety encoder that processes these tokens using specialized local-global attention mechanisms, (3) a co-attention mechanism that captures interactions within solvent and salt components, and (4) a cross-attention formulation module that models interactions between solvent and salt components before making the final prediction (**Supplementary Fig. 1** and **Supplementary Fig. 2**). This design enables HELENA to effectively model complex formulations containing multiple solvent and salt molecules at varying concentrations.

7.1.1 Comprehensive Moiety Detection

Rather than processing molecules at the atomic level, HELENA employs a tokenization approach that decomposes molecules into chemically meaningful moieties.

The tokenization begins with a predefined dictionary of functional groups encoded as SMARTS patterns, covering approximately 130 distinct chemical moieties including carbon-based groups (e.g., methyl, ethyl, isopropyl), oxygen-containing groups (e.g., hydroxyl, carbonyl, ester), nitrogen-containing groups (e.g., amine, amide, nitro), and various other functional groups relevant to electrolyte chemistry. The tokenizer also identifies ring systems as distinct moiety types (**Supplementary Fig. 3**).

Given a molecule represented as a SMILES string, the tokenizer first converts it to an RDKit molecular object and then applies a systematic procedure to identify all matching moieties (see **Supplementary Information, Algorithm 1**).

A critical aspect of the moiety detection algorithm is its selective inclusion of non-carbon atoms in the identified functional groups. This is achieved by examining the SMARTS pattern for each atom in a potential match and only including atoms whose pattern does not correspond to a simple carbon atom (represented as "[#6]" in SMARTS notation). This selective inclusion ensures that the detected moieties focus on reactive functional groups rather than carbon backbones, which are generally less relevant for electrochemical properties.

7.1.2 Ring System Detection

Ring system detection complements functional group identification by capturing important molecular structures. The algorithm:

1. Groups shared-atom rings into unified systems using a disjoint set approach

2. Names systems by ring sizes (e.g., "ring_6" for benzene, "ring_6_&_6" for naphthalene)

3. Resolves overlapping functional groups by prioritizing standard moieties over rings and larger structures over smaller ones

This approach creates a non-overlapping partition of each molecule into distinct functional units, enabling the model to recognize patterns in common electrolyte components.

7.1.3 Moiety Graph Construction

After identifying non-overlapping moieties, the tokenizer constructs a moiety-level graph representation where:

- Each node represents a detected moiety

- Edges represent connectivity between moieties

The construction of this graph involves mapping atoms to their containing moieties and then determining inter-moiety connections (see **Supplementary Information, Algorithm 2**).

The adjacency matrix for the moiety graph is computed as:

$$A_{ij} = \begin{cases} 1 & \text{if moieties } i \text{ and } j \text{ are connected} \\ 1 & \text{if } i = j \text{ (self-connections)} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

10

The distance matrix is computed using NetworkX $all\_pairs\_shortest\_path\_length()$ function to determine the shortest path length between each pair of moieties [51]:

$$D_{ij} = \begin{cases} d(i,j) & \text{if a path exists between moieties } i \text{ and } j \\ 0 & \text{if } i = j \\ 10^9 & \text{if no path exists (effectively infinite distance)} \end{cases} \tag{2}$$

where $d(i,j)$ is the length of the shortest path between moieties i and j.

7.1.4 Vocabulary Management

The tokenizer maintains a dynamic vocabulary that includes both predefined functional groups and discovered ring systems. Each moiety type is assigned a unique integer identifier, with 0 reserved for padding:

- IDs 1 to N correspond to the predefined functional groups from the MOIETIES dictionary
- IDs $N + 1$ onward are assigned to ring systems as they are encountered

This approach allows the model to handle novel ring structures not defined in the predefined vocabulary, making it adaptable to diverse molecular structures. The mapping between moiety names and numeric IDs is bidirectional to facilitate both tokenization and interpretation of model outputs.

The end result of the tokenization process is a structured representation of each molecule as:

- A sequence of moiety identifiers $\mathbf{m} = [m_1, m_2, \ldots, m_n]$
- An adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ encoding moiety connectivity
- A distance matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ encoding shortest-path distances between moieties

This rich representation preserves critical chemical information while abstracting away atomic details, enabling the model to learn generalizable relationships between functional group combinations and electrochemical properties.

7.2 Moiety Encoder with Local-Global Self-Attention

7.2.1 Overview and Inspiration

The moiety encoder transforms the sequence of moiety identifiers into a rich, context-aware representation using a modified transformer architecture with a specialized dual-attention mechanism. Unlike standard transformers that employ a single self-attention mechanism [28], we take inspiration from the local-global self-attention mechanism introduced by Zhong et al. [52] and propose a moiety encoder where parallel local and global attention operations simultaneously capture connectivity patterns and long-range interactions.

7.2.2 Local-Global Self-Attention Mechanism

The key feature in the HELENA moiety encoder is the separation of each layer's representation into two equal halves that are processed by parallel multi-head attention mechanisms:

$$\mathbf{x}_l, \mathbf{x}_g = \text{Split}(\mathbf{x}, \text{dim} = -1) \tag{3}$$

where $\mathbf{x}_l \in \mathbb{R}^{(N+1) \times \frac{d_{\text{model}}}{2}}$ and $\mathbf{x}_g \in \mathbb{R}^{(N+1) \times \frac{d_{\text{model}}}{2}}$ correspond to the portions processed by local and global attention, respectively. Here, N represents the number of moieties and the "+1" accounts for a special [GLOBAL] node that serves as an aggregation point.

The local attention branch focuses on capturing interactions between topologically adjacent moieties:

$$\text{Attention}_{\text{local}}^{(i)}(\mathbf{Q}_l^{(i)}, \mathbf{K}_l^{(i)}, \mathbf{V}_l^{(i)}, \mathbf{A}) = \text{softmax}\left(\frac{\mathbf{Q}_l^{(i)} \mathbf{K}_l^{(i)T} \odot \mathbf{A}}{\sqrt{d_k}}\right) \mathbf{V}_l^{(i)} \tag{4}$$

where $\odot$ represents the Hadamard product, $\mathbf{A} \in \mathbb{R}^{(N+1) \times (N+1)}$ is the adjacency matrix, d_k is the dimensionality of the key vectors, and i is the i-th attention head.

The global attention branch captures long-range dependencies using a distance-weighted mechanism:

$$\hat{\mathbf{D}} = \frac{1 + \exp(w)}{1 + \exp(w - \mathbf{D})} \tag{5}$$

11

where $\mathbf{D} \in \mathbb{R}^{(N+1)\times(N+1)}$ is the shortest-path distance matrix computed during tokenization, and w is a hyperparameter (set to 1.0) that controls the influence of distance on attention weights. This rescaling function transforms raw distances to modulate attention scores without eliminating attention between distant moieties.

The global attention computation uses this distance weighting:

$$\text{Attention}_{\text{global}}^{(i)}(\mathbf{Q}_g^{(i)}, \mathbf{K}_g^{(i)}, \mathbf{V}_g^{(i)}, \hat{\mathbf{D}}) = \text{softmax}\left(\frac{\text{ReLU}(\mathbf{Q}_g^{(i)}\mathbf{K}_g^{(i)T}) \odot \hat{\mathbf{D}}}{\sqrt{d_k}}\right)\mathbf{V}_g^{(i)} \tag{6}$$

Both branches employ 8-head attention, and their outputs are concatenated:

$$\mathbf{z} = \text{Concat}[\text{MultiHead}_{\text{local}}, \text{MultiHead}_{\text{global}}] \tag{7}$$

This representation undergoes standard transformer processing with residual connections, layer normalization, and a feed-forward network:

$$\mathbf{x}_{\text{out}} = \text{LayerNorm}(\mathbf{z}' + \text{Dropout}(\text{FFN}(\mathbf{z}'))) \tag{8}$$

where

$$\mathbf{z}' = \text{LayerNorm}(\mathbf{x} + \text{Dropout}(\mathbf{z}))$$
$$\text{FFN}(\mathbf{z}') = \text{GELU}(\mathbf{z}'\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2$$

7.2.3 Attention Heatmap Analysis

For model interpretability, we visualize attention weights to identify strongly interacting moiety pairs (**Supplementary Fig. 8**). We use the minimum attention weight across all heads rather than mean values:

$$\mathbf{A}_{\text{local/global}}[i,j] = \min_{k \in \{1,\dots,h\}} \mathbf{A}_{\text{local/global}}^{(k)}[i,j] \tag{9}$$

This approach highlights consistent and significant interactions while filtering out weaker connections. **Supplementary Fig. 4**, **Supplementary Fig. 5**, **Supplementary Fig. 6** and **Supplementary Fig. 7** show these heatmaps for representative molecules with varying performance metrics.

We also visualize per-head attention maps, revealing how different heads specialize in capturing distinct chemical features—some focus on specific functional groups while others capture broader structural patterns (**Supplementary Fig. 9**, **Supplementary Fig. 10**, **Supplementary Fig. 11** and **Supplementary Fig. 12**). This specialization demonstrates the model's ability to simultaneously capture multiple aspects of molecular structure, enhancing both prediction reliability and interpretability.

7.3 Group Attention and Formulation Co-Attention

To effectively model complex electrolyte formulations with multiple solvent and salt molecules at varying concentrations, HELENA processes solvents and salts separately before modeling their interactions. This approach acknowledges their distinct but complementary roles: solvents provide the medium for ion transport and affect viscosity and stability, while salts supply lithium ions and influence conductivity [53, 54].

7.3.1 Concentration-Weighted Group Attention

The group attention processes molecular embeddings $\mathbf{M} \in \mathbb{R}^{B\times N\times D}$, concentration values $\mathbf{c} \in \mathbb{R}^{B\times N}$, and binary masks $\mathbf{m} \in \{0,1\}^{B\times N}$, where B is batch size, N is maximum molecules per formulation, and D is embedding dimension. We apply binary masking to differentiate solvent molecules ($\mathbf{m}_{\text{solvent}}$) and salt molecules ($\mathbf{m}_{\text{salt}}$), running the same attention mechanism twice with these different masks.

For each formulation, we compute query, key, and value projections (**Supplementary Fig. 21**):

$$\mathbf{Q} = \mathbf{M}\mathbf{W}_q, \quad \mathbf{K} = \mathbf{M}\mathbf{W}_k, \quad \mathbf{V} = \mathbf{M}\mathbf{W}_v \tag{10}$$

The attention scores are calculated for valid molecules:

$$\mathbf{S}[b] = \frac{\mathbf{Q}[b]_{\text{valid}} \cdot \mathbf{K}[b]_{\text{valid}}^T}{\sqrt{D'}} \tag{11}$$

We first normalize concentrations for valid molecules:

$$\mathbf{c}_{\text{norm}}[b,i] = \frac{\mathbf{c}[b,i]}{\sum_{j=1}^{N} \mathbf{c}[b,j] \cdot \mathbf{m}[b,j] + \epsilon} \tag{12}$$

12

The attention mechanism applies concentration weighting to the softmax attention scores:

$$\mathbf{A}[b] = \text{softmax}(\mathbf{S}[b]) \odot \mathbf{c}_{\text{norm}}[b]_{\text{valid}} \tag{13}$$

We compute the context vector $\mathbf{O}$ and apply feed-forward networks with residual connections:

$$\mathbf{O}[b] = \mathbf{A}[b] \cdot \mathbf{V}[b]_{\text{valid}} \tag{14}$$

$$\mathbf{O}'[b] = \text{LayerNorm}(\mathbf{O}[b] + \text{FFN}(\mathbf{O}[b])) \tag{15}$$

Finally, we create fixed-size group representations using residual connections:

$$\mathbf{h}_{\text{group}}[b] = \text{LayerNorm}\left(\mathbf{O}'[b] + \frac{\sum_{i=1}^{N} \mathbf{M}[b,i] \cdot \mathbf{m}[b,i]}{\sum_{i=1}^{N} \mathbf{m}[b,i] + \epsilon}\right) \tag{16}$$

This process yields separate representations for solvent and salt groups: $\mathbf{h}_{\text{solvent}} \in \mathbb{R}^{B \times D}$ and $\mathbf{h}_{\text{salt}} \in \mathbb{R}^{B \times D}$.

7.3.2 Formulation Co-Attention

The formulation co-attention module models interactions between solvent and salt components using directed attention from solvent to salt (**Supplementary Fig. 22**):

$$\mathbf{Q} = \mathbf{W}_q \mathbf{h}_{\text{solvent}}, \quad \mathbf{K} = \mathbf{W}_k \mathbf{h}_{\text{salt}}, \quad \mathbf{V} = \mathbf{W}_v \mathbf{h}_{\text{salt}} \tag{17}$$

$$\mathbf{A}_{\text{co}} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{\text{co}}}}\right) \tag{18}$$

$$\mathbf{h}_{\text{updated}} = \text{FFN}(\text{LayerNorm}(\mathbf{C} + \mathbf{h}_{\text{solvent}})) \quad \text{where} \quad \mathbf{C} = \mathbf{A}_{\text{co}}\mathbf{V} \tag{19}$$

We then combine information through residual connections:

$$\mathbf{h}_{\text{combined}} = \text{LayerNorm}(\mathbf{h}_{\text{updated}} + \mathbf{h}_{\text{solvent}} + \mathbf{h}_{\text{salt}}) \tag{20}$$

Finally, learnable attention pooling [55] with a trainable query vector focuses on the most relevant aspects of the combined representation:

$$\mathbf{h}_{\text{formulation}} = \text{MultiheadAttention}(\mathbf{q}_{\text{pool}}, \mathbf{h}_{\text{combined}}, \mathbf{h}_{\text{combined}}) \tag{21}$$

The output $\mathbf{h}_{\text{formulation}} \in \mathbb{R}^{B \times D}$ serves as the final representation of the entire electrolyte formulation, capturing component properties and their interactions.

7.4 Prediction Head

The final formulation representation is passed through a multi-layer prediction head to generate the target property prediction:

$$\mathbf{h}_1 = \text{Dropout}(\text{GELU}(\text{LayerNorm}(\mathbf{h}_{\text{formulation}}\mathbf{W}_1 + \mathbf{b}_1))) \tag{22}$$
$$\mathbf{h}_2 = \text{Dropout}(\text{GELU}(\text{LayerNorm}(\mathbf{h}_1\mathbf{W}_2 + \mathbf{b}_2))) \tag{23}$$
$$\hat{y} = \mathbf{h}_2\mathbf{W}_3 + \mathbf{b}_3 \tag{24}$$

To improve gradient flow, training stability and performance, HELENA incorporates a deep-supervision-inspired mechanism that generates an auxiliary prediction directly from the formulation representation [56]:

$$\hat{y}_{\text{deep}} = \mathbf{h}_{\text{formulation}}\mathbf{W}_{\text{deep}} + \mathbf{b}_{\text{deep}} \tag{25}$$

The final prediction is the average of the main and auxiliary predictions:

$$\hat{y}_{\text{final}} = \frac{\hat{y} + \hat{y}_{\text{deep}}}{2} \tag{26}$$

13

7.5 Loss Function

The model was trained to predict the logarithmic Coulombic Efficiency $LCE = -log_{10}(1 - CE)$ rather than raw Coulombic Efficiency (CE). This transformation expands the range of high-performance values and better reflects meaningful improvements in cell performance, making it more suitable as a regression target [17]. The model was trained using the custom MSE loss function $\mathcal{L}$:

$$\mathcal{L} = (1 - \beta) \cdot \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \cdot w_i \cdot m_i + \beta \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{|y_i|} \tag{27}$$

where:

- $w_i = 1 + \alpha \cdot \tanh(z_i)$ with the z-score $z_i = \dfrac{|y_i - \mu_y|}{\sigma_y + \epsilon}$

- μ_y is the mean of all target values in the batch

- σ_y is the standard deviation of all target values in the batch

- m_i represents method-specific weighting (1.0 for Aurbach method versus 0.75 for Cycle method)

- α is a hyperparameter controlling the emphasis on extreme LCE values (set to 2.0)

- β is a weighting factor for the relative error term (set to 0.4)

- ϵ is a small constant to prevent division by zero

This multi-component loss function provides overall error minimization and includes method-specific weighting accounting for differences in experimental reliability between different cell testing methods.

7.6 Training Protocol

We employed 5-fold cross-validation on 90% of the data, with performance assessed on a separate 10% test set. This process was independently repeated 5 times with different random seeds for test set selection. By repeating the cross-validation, we mitigated the risk that rare but potentially important moieties would be concentrated in the test set, which could bias evaluation metrics.

During training, we applied random noise of $\pm 3\%$ to all concentration values to account for variability in laboratory measurements, enhancing the model's robustness to experimental uncertainty.

After partitioning the dataset, we trained with a batch size of 32 for up to 100 epochs, implementing early stopping with a patience of 25 epochs based on validation loss. For optimization, we used AdamW [57] with a weight decay of 0.01 and a OneCycleLR scheduler [58], which employs cosine annealing with a maximum learning rate of 1e-5. This scheduler used a 20% warm-up phase followed by gradual decay, with momentum cycling between 0.85 and 0.95 (**Supplementary Fig. 23**). We applied gradient clipping with a maximum norm of 1.0 to prevent gradient explosions, particularly important for molecules with complex functional group arrangements.

The final evaluation reported average metrics for the test set for the best fold model across all repetitions.

7.7 Ablation Studies

To understand the contribution of each architectural component to the overall model performance, we conducted ablation studies by systematically removing or modifying key components of the HELENA architecture. Each ablation variant was evaluated using a 5-fold cross-validation protocol on 90% of the data, with the best model from each run evaluated on a separately held-out 10% test set. To ensure statistical robustness, the entire process was repeated 5 times with different random seeds for test set selection, with results averaged across these independent evaluations.

We investigated the following model variants:

1. **No Moiety Encoder**: Replaced the moiety-level attention mechanism with simple embedding and averaging, testing the importance of the specialized attention for capturing intra-molecular interactions.

2. **No Solvent Attention**: Substituted the concentration-weighted co-attention for solvent molecules with simple pooling to assess the importance of modeling solvent-solvent interactions.

3. **No Salt Attention**: Removed the salt-specific attention mechanism to evaluate its contribution to the model's performance.

14

Extended Table 1: Ablation results on the test sets via 5-fold cross-validation protocol (repeated 5 times).

Model Variant	RMSE	R^2
No Moiety Encoder	0.293 ± 0.034	0.624 ± 0.088
No Solvent Attention	0.298 ± 0.059	0.596 ± 0.171
No Salt Attention	0.319 ± 0.037	0.560 ± 0.056
No Formulation Co-Attention	0.278 ± 0.035	0.659 ± 0.094
No Deep Supervision	0.287 ± 0.046	0.637 ± 0.111
Single Layer Moiety Attention	0.290 ± 0.049	0.625 ± 0.133
Half Model Size	0.287 ± 0.042	0.634 ± 0.110
HELENA	$\mathbf{0.212 \pm 0.014}$	$\mathbf{0.784 \pm 0.027}$

4. **No Formulation Co-Attention**: Replaced the cross-attention between solvent and salt representations with simple concatenation or averaging, testing the importance of explicitly modeling solvent-salt interactions.

5. **No Deep Supervision**: Eliminated the auxiliary prediction path to assess its impact on training stability and final performance.

6. **Single Layer Moiety Attention**: Reduced the number of attention layers in the moiety encoder from five to one to evaluate the importance of depth in the attention mechanism.

7. **Half Model Size**: Reduced all model dimensions by half to investigate the trade-off between model capacity and performance.

For each ablation variant, we tracked multiple performance metrics including RMSE and R² on both validation and test sets. **Extended Table 1** presents a comprehensive comparison of these variants across the 5 independent runs for the test sets, highlighting the relative contribution of each component to the overall model performance.

7.8 Large-Scale Inference

We deployed HELENA for large-scale inference on a comprehensive dataset of 46 million electrolyte formulations, leveraging an NVIDIA Grace Hopper GH200 equipped with a H100 GPU. During inference, we extracted and preserved the 104-dimensional penultimate-layer embeddings for the each formulation alongside its predicted Coulombic Efficiency (CE). This approach allowed us to analyze the latent space structure of high-performing electrolytes. For molecule discovery, we filtered the predictions to focus on formulations with predicted CE values above 99.5%, significantly narrowing the candidate space for experimental validation while maintaining diversity in chemical composition.

7.9 Model Hyperparameters

The HELENA hyperparameters determined through ablation studies and cross-validation are shown in **Extended Table 2**:

7.10 Implementation and Environment

HELENA was implemented in Python 3.12.9, PyTorch 2.4.1, RDKit 2024.09.2 [59] for molecular representation and cheminformatics tasks, NetworkX 3.4.2 for graph-based calculations, NumPy 2.1.3 and Pandas 2.2.3 for data manipulation, Scikit-learn 1.5.2 [60] and SciPy 1.14.1 for model evaluation and statistical analysis, Matplotlib 3.9.2 and Seaborn 0.13.2 for visualization, and tqdm 4.67.0 for progress tracking.

15

Extended Table 2: HELENA model hyperparameters.

Component	Parameters
Overall Architecture	Embedding dimension (d_{model}): 416 Maximum solvent molecules: 5 Maximum salt molecules: 3
Moiety Encoder	Number of layers: 5 Number of attention heads: 8 Dropout rate: 0.1
Solvent Attention	Attention dimension: 384 Dropout rate: 0.1
Salt Attention	Attention dimension: 384 Dropout rate: 0.1
Formulation Co-Attention	Attention dimension: 384 Dropout rate: 0.1
Final Layers	First dropout rate: 0.1 Second dropout rate: 0.1

8   Data Availability

The dataset will be made publicly available upon publication. All relevant information is included in the paper and its Supplementary Information. Supplementary Information will be expanded upon publication.

9   Code Availability

The codes will be made publicly available on GitHub upon publication.

10   Acknowledgments

11   Author Contributions

L.M., H.S., Y.C. and Z.B. conceived the idea. Y.C. and Z.B. directed the project. L.M. designed the model architecture, trained the model, and performed the inference study. H.S provided insights on the modeling, curated the training and inference dataset, performed model evaluation versus baseline models and selected the commercially available molecules for testing. J.F. performed the synthesis, material characterizations, electrochemical measurements and coin-cell tests. H.L. assisted with electrolyte synthesis, purification, and data analysis. All authors discussed and analyzed the data. L.M., H.S., Y.C. and Z.B. wrote and revised the manuscript.

References

[1] K. Bourzac. Fixing ai's energy crisis. *Nature*, 2024. doi:10.1038/d41586-024-03408-z.

[2] F. Degen, M. Winter, D. Bendig, and J. Tübke. Energy consumption of current and future production of lithium-ion and post lithium-ion battery cells. *Nature Energy*, 8:1284–1295, 2023.

[3] J. Liu, Z. Bao, Y. Cui, E.J. Dufek, J.B. Goodenough, P. Khalifah, Q. Li, B.Y. Liaw, P. Liu, A. Manthiram, Y.S. Meng, V.R. Subramanian, M.F. Toney, V.V. Viswanathan, M.S. Whittingham, J. Xiao, W. Xu, J. Yang, X.-Q. Yang, and J.-G. Zhang. Pathways for practical high-energy long-cycling lithium metal batteries. *Nature Energy*, 4:180–186, 2019.

[4] W. Zhang, J.R. Bowen, X. Lu, D. Qu, K. Dai, C. Blakley, T.P. Pollard, L. Chen, Y. Wang, O. Borodin, and C. Wang. Recovery of isolated lithium through discharged state calendar ageing. *Nature*, 626:306–312, 2024.

[5] M.D. Tikekar, S. Choudhury, Z. Tu, and L.A. Archer. Design principles for electrolytes and interfaces for stable lithium-metal batteries. *Nature Energy*, 1:1–7, 2016.

[6] D. Aurbach, Y. Gofer, and J. Langzam. The correlation between surface chemistry, surface morphology, and cycling efficiency of lithium electrodes in a few polar aprotic systems. *Journal of The Electrochemical Society*, 136:3198, 1989.

[7] X. Ren, S. Chen, H. Lee, D. Mei, M.H. Engelhard, S.D. Burton, W. Zhao, J. Zheng, Q. Li, M.S. Ding, M. Schroeder, J. Alvarado, K. Xu, Y.S. Meng, J. Liu, J.-G. Zhang, and W. Xu. High-concentration ether electrolytes for stable high-voltage lithium metal batteries. *ACS Energy Letters*, 4:896–902, 2019.

[8] C. Chang, X. Zhang, H. Zhao, L. Zhang, Z. Ma, Z. Wu, T. Li, Z. Liu, Y. Huang, F. Wu, P. He, and J. Chen. Stable lithium metal batteries enabled by localized high-concentration electrolytes with sevoflurane as a diluent. *Journal of Materials Chemistry: Materials for Energy and Sustainability*, 10:9001–9009, 2022.

[9] C.M. Efaw, D.M. Chapman, N. Aryal, C.P. Hensley, K.D. Myers, R.S. Andrus, J.N. Seitz, D.R. Baer, S.F. Son, D.-e Jiang, V. Murugesan, and W.A. Henderson. Localized high-concentration electrolytes get more localized through micelle-like structures. *Nature Materials*, pages 1–9, 2023. doi:10.1038/s41563-023-01700-3.

[10] X. Feng, X. Zhang, Y. Guo, F. Wang, B. Zhao, X. Peng, Q. Zhang, and Y. Xiao. New nonflammable tributyl phosphate based localized high concentration electrolytes for lithium metal batteries. *Sustainable Energy & Fuels*, 6:2198–2206, 2022.

[11] F. Hai, L. Qin, C. Zhu, Z. Li, Q. Ni, M. Xu, C. Yang, C. Ma, X.-B. Cheng, T. Liu, and Q. Zhang. A low-cost, fluorine-free localized highly concentrated electrolyte toward ultra-high loading lithium metal batteries. *Advanced Energy Materials*, 14:2304253, 2024.

[12] J. Wang, S. Jin, X. Feng, X. Han, J. Kang, S. Qiao, Z. Lin, C. Zhao, and S. Liu. 2h,3h-decafluoropentane endorsed localized high concentration electrolyte for low-temperature lithium-metal batteries. *ChemPhysChem*, page e202400920, 2024.

[13] Z. Yu, H. Wang, X. Kong, W. Huang, Y. Tsao, D.G. Mackanic, K. Wang, X. Wang, W. Huang, S. Choudhury, Y. Zheng, C.V. Amanchukwu, S.T. Hung, Y. Ma, E.G. Lomeli, J. Qin, Y. Cui, and Z. Bao. Rational solvent molecule tuning for high-performance lithium metal battery electrolytes. *Nature Energy*, 7:94–106, 2022.

[14] I.R. Choi, S. Jeong, J.S. Kim, S. Yuk, J. Chun, J. Song, B.C. Koo, J.S. Lee, J. Lee, S. Hwang, M. Jang, H.D. Lim, and K. Kang. Asymmetric ether solvents for high-rate lithium metal batteries. *Nature Energy*, 10:365–379, 2025.

[15] S.C. Kim, H. Ji, A.T. Ngo, M. Li, J. Zhang, K.S. Han, D. Su, V. Murugesan, and S.D. Han. High-entropy electrolytes for practical lithium metal batteries. *Nature Energy*, 8:814–826, 2023.

[16] Q. Wang, X. Sun, K. Xu, H. Zheng, L. Xue, Z. Zhang, Y. Fu, Y. Zhou, G. Chen, and W. Li. High entropy liquid electrolytes for lithium batteries. *Nature Communications*, 14:440, 2023.

[17] S.C. Kim, X. Kong, R.A. Vilá, W. Huang, Y. Chen, D.T. Boyle, H. He, Z. Bao, and Y. Cui. Data-driven electrolyte design for lithium metal anodes. *Proceedings of the National Academy of Sciences*, 120:e2214357120, 2023.

[18] J.M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N.M. Donghia, C.R. MacNair, S. French, L.A. Carfrae, Z. Bloom-Ackermann, V.M. Tran, A. Chiappino-Pepe, A.H. Badran, I.W. Andrews, E.J. Chory, G.M. Church, E.D. Brown, T.S. Jaakkola, R. Barzilay, and J.J. Collins. A deep learning approach to antibiotic discovery. *Cell*, 181:475–483, 2020.

[19] B.K. Lee, S.R. Datta, and D.D. Lee. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381:999–1006, 2023.

[20] V. Sharma, R.P. Cunha, C.M. Marques, J. Shah, S. Aggarwal, Y. Liu, E. Soares, E.V. Brazil, Y.-H. Na, and R. Cerqueira. Formulation graphs for mapping structure-composition of battery electrolytes to device performance. *Journal of Chemical Information and Modeling*, 63:6998–7010, 2023.

[21] R. Kumar, M.C. Vu, P. Ma, and C.V. Amanchukwu. Electrolytomics: A unified big data approach for electrolyte design and discovery. *Chemistry of Materials*, 2025. doi:10.1021/acs.chemmater.4c03196.

[22] E. Soares, V. Sharma, E.V. Brazil, R. Cerqueira, and Y.-H. Na. Capturing formulation design of battery electrolytes with chemical large language model. 2023.

[23] Anonymous. Improving performance prediction of electrolyte formulations with transformer-based molecular representation model. arXiv:2406.19792, 2024.

[24] T.K. Rusch, M.M. Bronstein, and S. Mishra. A survey on oversmoothing in graph neural networks. arXiv preprint, 2023.

[25] Y. Xia, A. Yin, Y. Cao, A. Lee, M. Shen, A. Walker, A. Dolocan, Z. Liu, J. Wang, H. Wang, Z. Liang, V. Arumugam, and J.B. Goodenough. Designing an asymmetric ether-like lithium salt to enable fast-cycling high-energy lithium metal batteries. *Nature Energy*, 8:934–945, 2023.

[26] Y. Lu, S. He, X. Feng, J. Zhou, L. Qin, T. Liu, X.-B. Cheng, Z. Hong, and Q. Zhang. Breaking the molecular symmetricity of sulfonimide anions for high-performance lithium metal batteries under extreme cycling conditions. *Nature Energy*, 10:191–204, 2025.

[27] A.-M. Li, P.Y. Zavalij, F. Omenya, X. Li, and C. Wang. Salt-in-presalt electrolyte solutions for high-potential non-aqueous sodium metal batteries. *Nature Nanotechnology*, 20:388–396, 2025.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. arXiv preprint, 2023.

[29] B. Kim, G. Nam, T.J. Lee, J. Kwon, I. Hwang, J. Kim, H.-T. Kim, H.-K. Song, and J.W. Choi. Succinonitrile-rich electrolyte solvation structure enables wide-temperature-range operation of lithium-metal batteries. *Small Methods*, page 2401957, 2024.

[30] Z. Piao, K.-H. Chen, M. Liu, Z. Wei, G.-L. Xu, C. Zhang, Y. Liu, Z. Zhang, Y. Zhu, C. Luo, H. Wang, Y. Ren, L.A. Curtiss, X. Zhang, and K. Amine. A semisolvated sole-solvent electrolyte for high-voltage lithium metal batteries. *Journal of the American Chemical Society*, 145:24260–24271, 2023.

[31] L. Chen, C. Zhai, D. Wang, S. Ning, C. Hu, C.-C. He, Z. Tao, P. Chen, A.L. Hong, W. Zhai, X. Li, D.G. Huang, S. Huang, W. Sun, H. Wang, and T.-Y. Zhang. Dynamic shielding of electrified interface enables high-voltage lithium batteries. *Chem*, 10:1196–1212, 2024.

[32] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S.H. Bryant. Pubchem substance and compound databases. *Nucleic Acids Research*, 2016.

[33] B. Zeng, Z. Jin, M. Li, X. Liu, H.-P. Cheng, and Q. Li. Uni-elf: A multi-level representation learning framework for electrolyte formulation design. arXiv preprint, 2024.

[34] Qmugs, quantum mechanical properties of drug-like molecules. *Scientific Data*, 2022.

[35] Y. Wang, Z. Yang, and Q. Yao. Accurate and interpretable drug-drug interaction prediction enabled by knowledge subgraph learning. *Communications Medicine*, 4:1–12, 2024.

[36] P. Bryant, G. Pozzati, and A. Elofsson. Improved prediction of protein-protein interactions using alphafold2. *Nature Communications*, 13:1265, 2022.

[37] X. Liu, Y. Deng, X. Zhang, J. Zhang, S. Lou, J. Ma, F. Li, Y. Lu, N. Zhang, Q. Li, H. Yuan, Y. Ma, J. Zheng, and G. Cui. Difluorobenzene-based locally concentrated ionic liquid electrolyte enabling stable cycling of lithium metal batteries with nickel-rich cathode. *Advanced Energy Materials*, 12:2200862, 2022.

[38] P. Liu, Y. Xu, S. Wang, W. Wang, Y. Zhao, W. Wang, L. Gao, H. An, X. Zeng, Y. Wu, Y. Gao, and Y. Yang. Difluorobenzene as an antisolvent for fluorinated electrolyte to achieve unparalleled cycle life of lithium metal battery. *ACS Applied Materials & Interfaces*, 16:63628–63637, 2024.

[39] Electrolyte with single salt and single solvent enables fes2-based lithium metal batteries with all-climate high performance. *Energy Storage Materials*, 2025. doi:10.1016/j.ensm.2025.104124.

[40] B. Ma, J. Wu, S. Ma, K. Chen, Y.-T. Li, Y. Zhang, T. Chen, and J. Xie. Molecular-docking electrolytes enable high-voltage lithium battery chemistries. *Nature Chemistry*, 16:1427–1435, 2024.

[41] D. Ruan, X. Meng, D. Liu, Z. Xu, X. Su, X. Guo, Z. Li, C. Chen, W. Chen, G. Zhu, W. Huang, C. Nan, and F. Lin. Solvent versus anion chemistry: Unveiling the structure-dependent reactivity in tailoring electrochemical interphases for lithium-metal batteries. *JACS Au*, 3:953–963, 2023.

[42] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016. doi:10.1145/2939672.2939785.

[43] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. arXiv preprint, 2018.

[44] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, and G.E. Dahl. Neural message passing for quantum chemistry. arXiv preprint, 2017.

[45] J. Cadow, J. Born, M. Manica, A. Oskooei, and M. Rodríguez Martínez. Paccmann: a web service for interpretable anticancer compound sensitivity prediction. *Nucleic Acids Research*, 48:W502–W508, 2020.

[46] Y. Chen, R. Sun, A. Dai, W. Zou, Y. Zheng, J. Zhang, C. Chen, M. Li, H. Kang, Y. Yang, J. Li, J. Mao, Y. Ding, L. Huang, Y. Jin, S. Zhang, J. Ma, and G. Cui. Hyperconjugation-controlled molecular conformation weakens lithium-ion solvation and stabilizes lithium metal anodes. *Chemical Science*, 15:19805–19819, 2024.

[47] Z. Li, X. Zhang, Z. Tu, X. Ren, Y. Sun, P. Okuniewski, M. Avdeev, K.S. Han, C. Liang, and W. Xu. Non-polar ether-based electrolyte solutions for stable high-voltage non-aqueous lithium metal batteries. *Nature Communications*, 14:868, 2023.

[48] A.-M. Li, L. Chen, Y. Wang, J.D. Lee, J. Vatamanu, C. Blakley, J. Cui, P.F. Wang, G.-L. Xu, K. Dai, J. Zhu, M. Wang, Y. Ju, X. Li, O. Borodin, and C. Wang. Methylation enables the use of fluorine-free ether electrolytes in high-voltage lithium metal batteries. *Nature Chemistry*, 16:922–929, 2024.

[49] E. Park, M.S. Kim, S.-Y. Lee, C.S. Yoon, and Y.-K. Sun. Exploiting the steric effect and low dielectric constant of 1,2-dimethoxypropane for 4.3 v lithium metal batteries. *ACS Energy Letters*, 8:179–188, 2023.

[50] S.C. Kim, X. Kong, R.A. Vilá, W. Huang, Y. Chen, D.T. Boyle, H. He, Z. Bao, and Y. Cui. Potentiometric measurement to probe solvation energy and its correlation to lithium battery cyclability. *Journal of the American Chemical Society*, 143:10301–10308, 2021.

[51] A. Hagberg, P.J. Swart, and D.A. Schult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab, 2008. URL `https://www.osti.gov/biblio/960616`.

[52] Y. Zhong, X. Zhou, J. Li, H. Dai, W. Gu, and J. Yang. Learning motif-based graphs for drug–drug interaction prediction via local–global self-attention. *Nature Machine Intelligence*, 6:1094–1105, 2024.

[53] A.J. Bard and L.R. Faulkner. *Electrochemical Methods: Fundamentals and Applications*. Wiley, 3 edition, 2023.

[54] L. Suo, Y.-S. Hu, H. Li, M. Armand, and L. Chen. A new class of solvent-in-salt electrolyte for high-energy rechargeable metallic lithium batteries. *Nature Communications*, 4:1481, 2013.

[55] H. Touvron, P. Bojanowski, M. Caron, M. Cord, A. El-Nouby, E. Grave, A. Joulin, G. Synnaeve, J. Verbeek, and H. Jégou. Augmenting convolutional networks with attention-based aggregation. arXiv preprint, 2021.

[56] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-supervised nets. arXiv preprint, 2014.

[57] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. arXiv preprint, 2019.

[58] L.N. Smith and N. Topin. Super-convergence: Very fast training of neural networks using large learning rates. arXiv preprint, 2018.

[59] Greg Landrum et al. rdkit/rdkit: 2025_03_1 (q1 2025) release. Zenodo, 2025.

[60] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in python. arXiv preprint, 2018.