CHEMISTRY OF MATERIALS

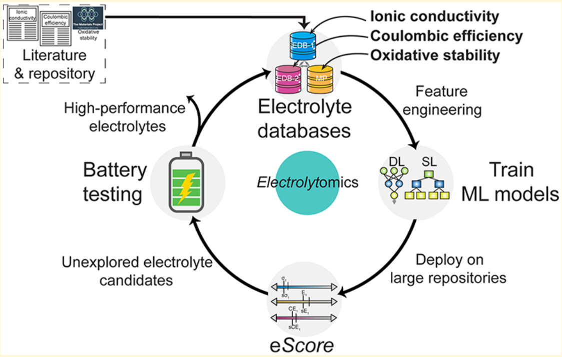
pubs.acs.org/cm Article

Electrolytomics: A Unified Big Data Approach for Electrolyte Design and Discovery

Ritesh Kumar, Minh Canh Vu, Peiyuan Ma, and Chibueze V. Amanchukwu*

Cite This: Chem. Mater. 2025, 37, 2720−2734 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



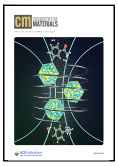
ABSTRACT: Electrolyte discovery is the bottleneck for developing next-generation batteries. For example, lithium metal batteries (LMBs) promise to double the energy density of current Li-ion batteries (LIBs), while next-generation LIBs are desired for operations at extreme temperature conditions and with high voltage cathodes. However, there are no suitable electrolytes to support these battery chemistries. Electrolyte requirements are complex (conductivity, stability, safety), and the chemical design space (salts, solvents, additives, concentration) is practically infinite; hence, discovery is primarily guided through trial and error, which slows the deployment of such next-generation battery chemistries. Inspired by artificial intelligence (AI)-enabled drug discovery, we adapt these machine learning (ML) approaches to electrolyte discovery. We assemble the largest small molecule experimental liquid electrolyte ionic conductivity data set and build highly accurate ML and deep learning models to predict ionic conductivity across a wide range of electrolyte classes. The developed models yield results similar to those of molecular dynamics (MD) simulations and are interpretable without explicit encoding of ionic solvation. While most ML-based approaches target a single property, we build additional models of oxidative stability and Coulombic efficiency and develop a metric called the electrolyte score (*eScore*) to unify the predicted disparate electrolyte properties. Deploying these models on large unlabeled data sets, we discover distinct electrolyte solvents, experimentally validate that the electrolyte is conductive (>1 mS cm^{-1}), stable up to 6 V, supports efficient anode-free LMB, and even LIB cycling at extreme temperatures. Our work marks a significant step toward efficient electrolyte design, accelerating the development and deployment of next-generation battery technologies.

■ INTRODUCTION

Next-generation energy-dense batteries are required to accelerate decarbonization and the transition to grid electrification.[1] While anode and cathode materials are often known, the lack of suitable electrolytes to enable long-term reversibility has stymied battery development. Lithium metal batteries (LMBs) illustrate these challenges. LMBs are considered the holy grail of batteries because they combine the high capacity and low reduction potential of lithium with either lithium-containing (transition metal oxide) or non-lithium-containing cathodes (sulfur, oxygen).[2] Unfortunately, current electrolytes support nonuniform high surface area deposits such as dendrites that exacerbate reaction with the electrolyte solvent and salt and lead to a severe battery capacity fade. Since conventional Li-ion electrolytes are poor for LMBs,[3] the design of previously unreported liquid electrolytes has led to significant increases in Coulombic efficiency (CE), but progress has plateaued.[4] Even current Li-ion batteries (LIBs) are limited by the lack of electrolytes that support high voltage or low cobalt cathodes and work at low (<−20 °C) or high temperatures (>60 °C).[5] Therefore, alternative approaches to electrolyte design and discovery are vital.

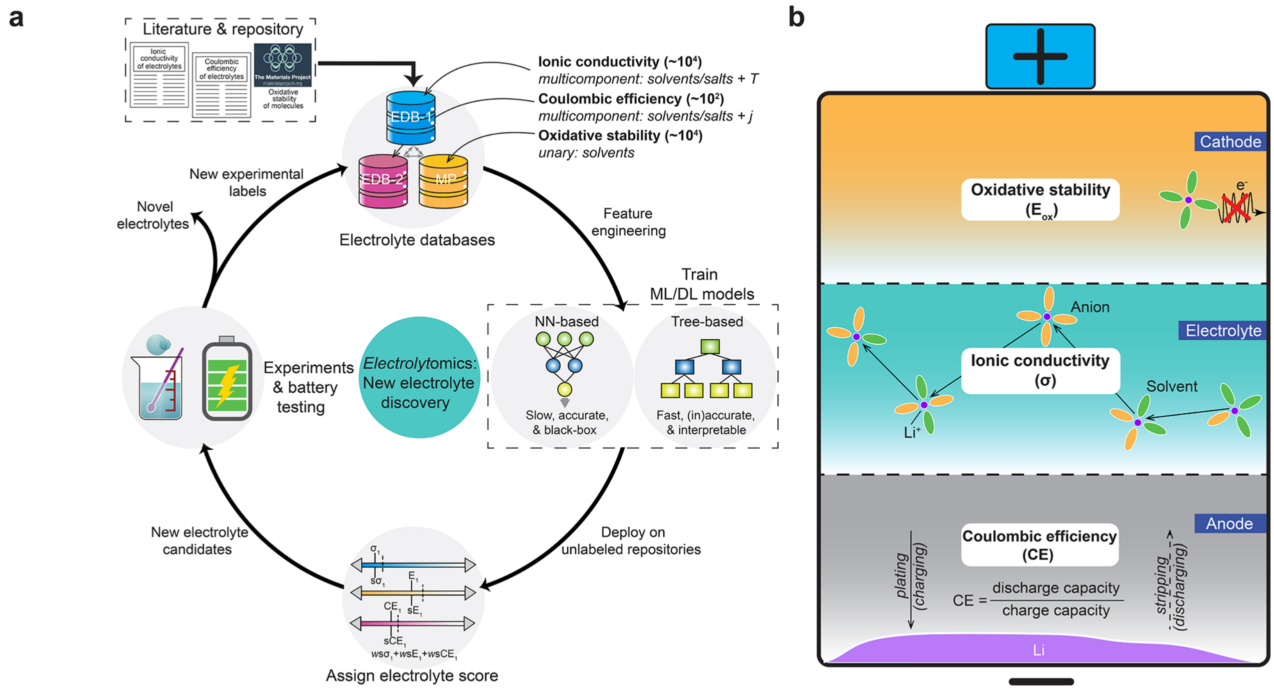
An excellent electrolyte for next-generation batteries such as LMBs is required to be ionically conductive, oxidatively stable, reductively stable, and support high CE.[6] Unfortunately, these properties are often in conflict (e.g., simultaneous oxidative and reductive stability), and the electrolyte solvent design space alone scales with the reported possible molecular space (~10^{60}).[7] Hence, for the past century, electrolyte design has primarily been driven by trial and error. For instance, in efforts to alleviate the poor reductive stability of carbonates with LMBs, ethers were studied but demonstrated limited oxidative stability.[8] To address this limitation, fluorinated ethers were developed, which showed promising compatibility with lithium anodes, albeit at the cost of poorer reductive stability compared to their nonfluorinated analogs.[9] Due to the lack of robust universal design principles for electrolytes, consid-

Received: November 20, 2024
Revised: March 15, 2025
Accepted: March 18, 2025
Published: April 1, 2025



https://doi.org/10.1021/acs.chemmater.4c03196
Chem. Mater. 2025, 37, 2720−2734

Scheme 1. Electrolyte Discovery Workflow (Electrolytomics) and Critical Electrolyte Properties[a]



[a](a) Schematic illustrating the overall ML-based workflow utilized in the present study for the discovery of unexplored LMB electrolytes. (b) Illustration depicting three key figures of merits of a typical electrolyte: ionic conductivity, oxidative stability, and Coulombic efficiency. MP = Materials project, EDB-1 = ionic conductivity database, EDB-2 = Coulombic efficiency database (explained in the text).

erable experimental and computational efforts are required to derive governing principles that are typically nontransferable to other electrolyte classes. Consequently, the conventional Edisonian approach is insufficient for the discovery of superior electrolytes within a reasonable time frame, requiring alternative data-driven approaches to accelerate progress.[10] Statistical methods such as the licensed Advanced Electrolyte Model by Gering et al. have primarily been used for carbonate-based liquid electrolytes and struggle with other electrolyte classes.[11] Inspired by advances in drug[12] and materials[13−15] discovery and other scientific domains,[16] artificial intelligence (AI)-based approaches[15] provide a facile pathway to accelerate electrolyte property prediction and electrolyte discovery. Approaches leveraging high-throughput[17−21] and data science techniques have gained traction to predict properties relevant for mostly solid[22−25] and liquid[26] electrolytes with prior studies focused primarily on LIBs and one or two figures of merit with small experimental data sets or little to no experimental verification. Therefore, long-standing challenges remain such as limited data set size, the lack of multiproperty predictive models, limited model interpretability, limited experimental validation, and limited discoveries of unexplored electrolyte classes.

While the focus of this study is on three key electrolyte properties—conductivity, oxidative stability, and CE—other attributes such as low melting point, high boiling point, low viscosity, and nonflammability have traditionally been emphasized as ideal.[27,28] However, strict adherence to these criteria may limit the scope of discovery and exclude promising candidates. For example, in the field of LIBs, ethylene carbonate (EC), with a melting point of approximately 38 °C, was initially disregarded in favor of propylene carbonate (PC) with a lower melting point (−55 °C)[29] but caused graphite exfoliation. Similarly, dimethyl carbonate (DMC) and methyl formate despite having low flash points (~16 and −19 °C[30]), are widely employed as cosolvents or additives in LIBs. While high boiling points are generally considered advantageous for electrolyte solvents, the emerging field of liquefied gas electrolytes challenges this assumption, as these solvents are gases at room temperature (e.g., fluoromethane, with a boiling point of −78 °C[31]). Despite this, liquefied gas electrolytes have shown excellent performance, particularly in LMBs,[31] and are suitable for specialized applications, including low-temperature batteries for space missions. Similarly, molten salt electrolytes, despite having very high viscosity compared to traditional liquid electrolytes, have demonstrated significant improvements in cycling performance.[32] These examples highlight that while conventional electrolyte design principles are valuable, deviations from standard criteria have frequently led to major advances. A flexible approach to electrolyte selection allows for the discovery of unconventional but highly effective solvents for specific energy storage applications.

Here, we develop a holistic AI-guided approach incorporating a confluence of data science, experiments, and computation for the simultaneous prediction of three electrolyte properties (conductivity, oxidative stability, and CE) and the discovery of distinct electrolyte classes. Akin to the omics-type strategy utilized in genomics[33] and proteomics;[34] we term this approach "electrolytomics", the large-scale systematic investigation of electrolytes. First, we build a large liquid electrolyte ionic conductivity database (EDB-1; 10^4 entries) encompassing the complexity of typical electrolytes with different solvent types and mixtures, salts, concentrations, and temperatures. Coupled with an experimental CE data set and a DFT-obtained oxidative stability data set (Materials Project[35]), we accurately predict the three desired electrolyte properties using

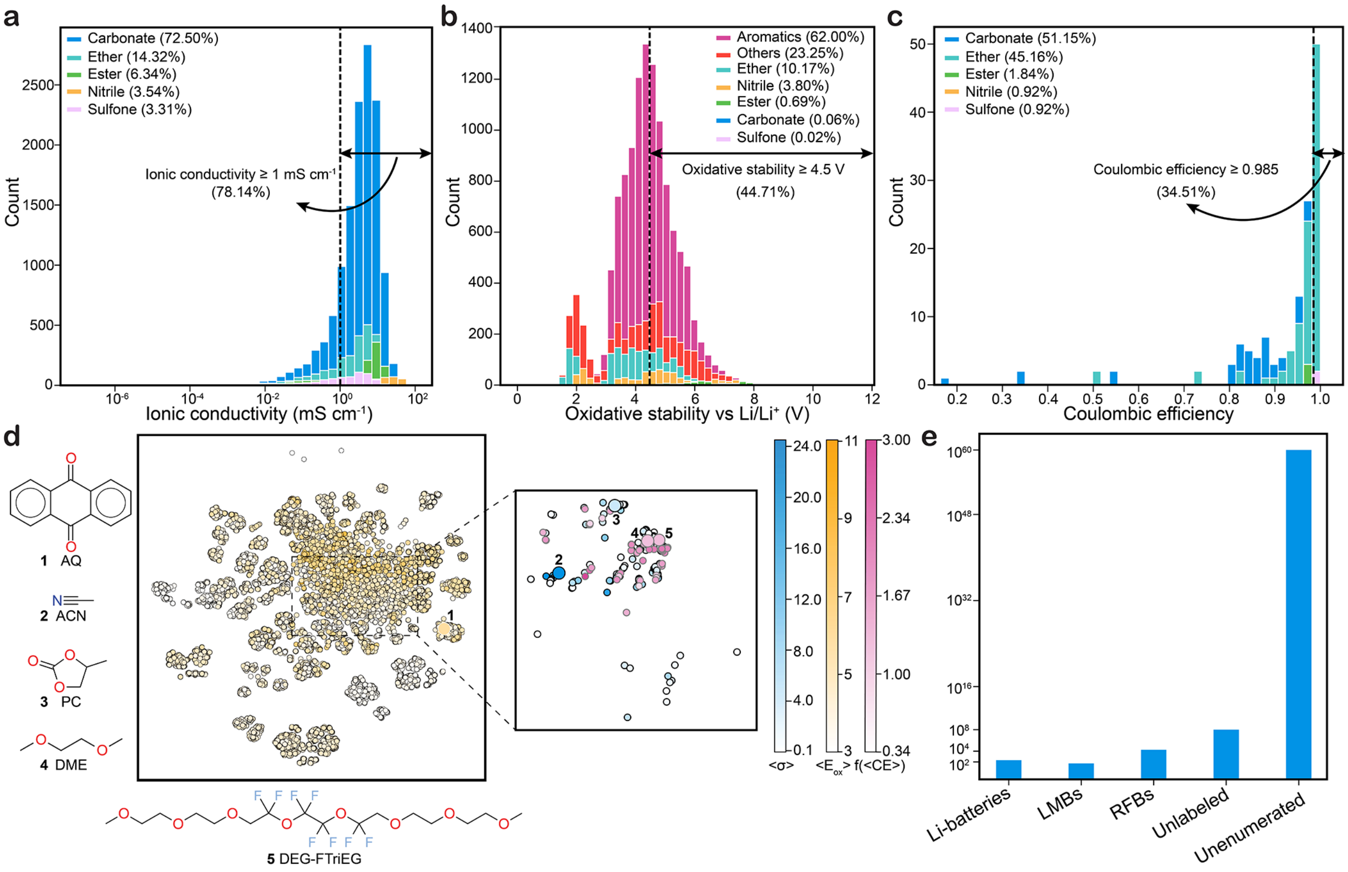


Figure 1. Visualizing and estimating the explored and unexplored electrolyte solvent molecules chemical space. Histogram of (a) ionic conductivity, (b) oxidative stability, and (c) CE as a function of different solvent classes. The histograms in (a−c) also show the fraction of data points surpassing respective performance thresholds (Supporting Information, Note 1). (d) A 2D t-SNE plot depicting chemical space of the three employed databases colored according to the average property values for each solvent molecule, i.e., blue for ionic conductivity ($\langle\sigma\rangle$), orange for oxidative stability ($<E_{ox}>$; only shown from 3.0 V onward for clarity), and purple for CE ($f(<CE> = -\log(1 - CE))$. The dashed box inside the left t-SNE denote the region spanned by ionic conductivity and CE data points (not shown for clarity), which has been zoomed out on the right t-SNE (oxidative stability data points not shown for clarity). Representative solvent molecules from the three data sets are also highlighted in larger circles and marked by numbers whose corresponding structures are shown on the left or bottom side. A detailed t-SNE plot is shown in Figure S6. (e) Bar plot showing the population of unique solvent molecules present in the three databases—EDB-1 (investigated for all Li-based batteries), MP (evaluated for redox flow battery (RFB) applications), and EDB-2 (investigated for LMBs). The EDB-1 and EDB-2 data sets are experimentally labeled, while the MP data set is derived from high-fidelity DFT calculations. Approximate populations of unexplored but enumerated compounds and all possible compounds are also shown for comparison.

a combination of forward design ML and deep learning (DL) techniques. From a plethora of ML algorithms and featurization techniques developed in the past decade for highly accurate molecular property prediction, we selected the most popular and highly accurate techniques such as LightGBM using hand-crafted features, e.g., fragments of common functional groups and message-passing neural networks using molecular graphs. These tools and techniques have been commonly employed in the drug[36] and materials[15] discovery domains. The objective of the present work is to examine state-of-the-art ML techniques for energy storage applications for wide community adoption and reproducibility. The obtained conductivity predictions are comparable to those calculated using MD simulations. Despite no encoding of the solvation structure, the interpretable model can predict the Arrhenius relationship for electrolytes as well as the volcano-like correlation between conductivity and salt concentration, among other phenomenological observations. We also establish a metric called the electrolyte score (e*Score*) that provides a unified view of the electrolytes' performance in terms of the three figures of merit. Finally, we discover a distinct class of sulfonyl fluoride-based electrolytes yielding CE as high as 99.5% that competes favorably against all known liquid electrolytes, supports low lithium metal overpotentials, and can support LMBs with $LiFePO_4$ (LFP) cathodes for long-term cycling. To broaden the scope, we also evaluated these electrolytes for next-generation LIBs that demonstrate compatibility with high voltage NMC811 ($LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$) cathodes and can also be cycled at low and high temperatures. Our proposed approach is transformative and will accelerate the discovery of electrolytes and deployment of a wide range of next-generation battery chemistries required to stem a changing climate.

■ RESULTS AND DISCUSSION

Curating Data Sets for Critical Properties of Next-Generation Battery Electrolytes. The data-driven electrolyte discovery workflow from data collection to experimental observations is summarized in Scheme 1a and Figure S1. An efficient electrolyte suitable for a general next-generation battery (NGB) is characterized by several conflicting properties, with ionic conductivity (σ), oxidative stability (E_{ox}), and

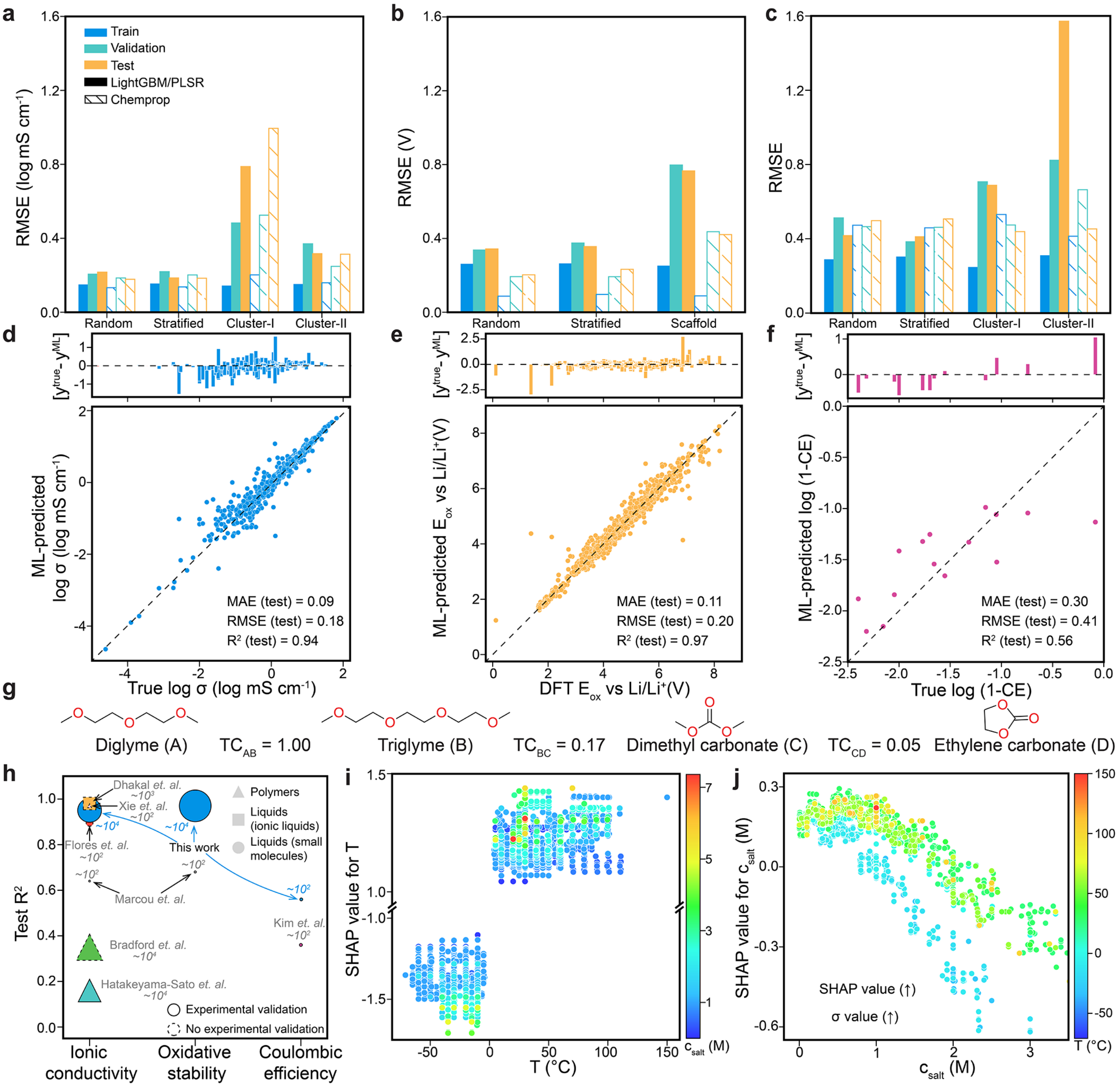


Figure 2. ML performance on the target properties using different data splits, comparison with literature, and interpretability. Bar plot comparing the performance of shallow learning (SL) and Chemprop models on different data splits for (a) EDB-1 (ionic conductivity), (b) MP (oxidative stability), and (c) EDB-2 (CE) data sets. Parity plots and residual histograms for best-performing models on test sets for predicting (d) ionic conductivity in the common log scale (log σ) (stratified split), (e) oxidative stability (random split), and (f) Coulombic inefficiency in the common log scale (log (1-CE)) (stratified split) labels. The parity plots for corresponding train and validation sets are shown in Figure S8 for more clarity. (g) Tanimoto coefficient (TC) values shown for a few typical solvent molecules present in the EDB-1 data set. (h) Comparison of ML-based research on Li battery electrolytes published in recent years, assessing the quality of ML performance and the approximate data set size (indicated by italicized numbers) used for model development. Different symbols are shown according to the electrolyte nature under consideration, e.g., polymer, liquid, and ionic liquid (note: ionic liquid data points did not contain added salts). SHAP dependence plots for (i) temperature (T) and (j) salt concentration (c_{salt}) (within moderate concentration range) features for the ionic conductivity model. The scatter points in (i,j) are colored according to the c_{salt} and T values, respectively, to highlight the interaction between the two features. RMSE = root mean squared error. MAE = mean absolute error.

Coulombic efficiency (CE) among the most important (Scheme 1b). We collected an extensive database of ionic conductivity for lithium-containing aprotic liquid electrolytes (EDB-1) reported over the past 50 years with greater diversity in solvent and salt classes, salt concentrations, and temperatures, compared to previous studies (Table S1). Because of the prominence of Li-ion battery research, the EDB-1 data set is heavily sampling biased with carbonate solvents constituting around 72% of the data set (Figure 1a) and $LiPF_6$ salts constituting around 26% of the data set. It also has deviations within the reported values from different sources (Table S2).

For oxidative stability, we utilized density functional theory (DFT)-derived labels for molecules from the Materials Project (MP) database,[35] as these values show good correlation with experimental results, demonstrating a mean absolute error (MAE) of 0.16 eV for isolated molecules,[17,37] even when employing coarse B3LYP/6-31*g(d) calculations. However, when comparing DFT-derived values for solvents to experimentally obtained values for electrolytes, a higher MAE of 0.68 eV has recently been reported.[38] Thus, it is crucial to exercise caution when using the MP data set, and it should be employed strictly for screening purposes (Supplementary Note 1). Additionally, it should be noted that the MP data set primarily consists of aromatic compounds, as it was initially developed for redox flow batteries (RFBs) (Figure 1b). CE, which is the ratio of the amount of charge delivered during discharge to the amount of charge stored during the charging cycle, was selected as the final metric (Li||Cu half-cell CE)[39−41] and obtained from the literature. CE is an important metric for an electrolyte as it encapsulates morphological, reductive stability, passivation effects, current density among others.[42] The EDB-2 data set for CE primarily consists of carbonates and ethers, with each class contributing roughly half of the entire data set. This composition reflects the recent rise in the use of ether-based electrolytes for LMBs due to their better compatibility.[4] Our approach is inherently flexible and can be readily extended to include additional properties such as nonflammability for desired applications, provided that suitable data sets are available.

The chemical space of the unique solvent molecules present in the three databases can be discerned through a two-dimensional (2D) t-distributed stochastic neighbor embedding (t-SNE) projection (using Morgan fingerprint representation of molecules; details in Supplementary Note 3) as shown in Figure 1d, where some of the typical solvent molecules present in the three databases are marked. A clear overlap between the EDB-1 and EDB-2 data sets is visible due to some of the common electrolytes that have been investigated for both properties. In contrast, most of the molecules in the MP data set lie in a different chemical space compared to the other two data sets (minimal overlap in the t-SNE projection) because such molecules are active redox species and not ionically conductive electrolytes.[37,43] Moreover, the number of unique solvents present in the experimental data sets of ionic conductivity and CE curated from the literature are quite low—202 and 53, respectively. All solvent and salt molecular structures present in the EDB-1 and EDB-2 data sets are provided in Figures S2 and S3, while all molecules in the MP database are publicly available. Additionally, the three combined databases still constitute a small fraction compared to the chemical space of molecules that have been enumerated (with chemical structures; $\sim 10^8$) but still unexplored or unenumerated (with unknown chemical structures; $\sim 10^{60}$) (Figure 1e). Moreover, in comparison to other domains, the data set sizes in this study (10^3−10^4 data points) are consistent with typical scales in materials science and drug discovery, where data sets often range from hundreds to tens of thousands of entries due to experimental and computational constraints.[44−46] This scale reflects the inherent challenges and costs of generating high-quality experimental data in materials and chemical sciences.

Developing Forward ML Models for the Critical Electrolyte Properties. For this work, we focus on the deployment of state-of-the-art ML models and not on the development of unique methods. Separate ML regression models were developed to predict the three target properties as a function of different data splits (Supporting Information, Note 1). A common strategy for splitting all three data sets was employed through random and stratified splits. For examining data leakage,[47] two other splits were created for EDB-1 and EDB-2 data sets called cluster-I and cluster-II, as shown schematically in Figure S7, while oxidative stability prediction was also examined for out-of-domain performance through scaffold split. We evaluated the performance of both shallow learning (SL) and deep learning (DL) models across the three target properties. In our approach, molecules were represented using two methods: either as tabular hand-crafted descriptors or as geometric molecular graphs suitable for SL and DL algorithms, respectively. For the MP data set, where the labels depend solely on individual molecules (as discussed in Supplementary Note 5), we utilized molecular hand-crafted descriptors (detailed in Supplementary Note 2 and Table S13) with the LightGBM algorithm, as well as directed molecular graph features with Chemprop, to develop predictive models for oxidative stability. In contrast, the EDB-1 and EDB-2 data sets include both solvent and salt components, as well as experimental conditions such as temperature and current density. Thus, molecular representations of both the solvent and salt (list of hand-crafted features tabulated in Table S13), along with numerical values for experimental conditions, e.g., temperature for ionic conductivity and current density or cycling protocol for CE, were collectively provided to the SL and DL models to enable comprehensive prediction (details in Supporting Information, Notes 4 and 6). A schematic representation of different featurization techniques used in the training of ML models is provided in Figure S5, the exact number of features used in Tables S5, S8, and S11, and discussed in detail in Supplementary Note 2. Comparisons of the SL and DL (Chemprop[48]) models' performance on different splits of the EDB-1, MP, and EDB-2 data sets are summarized in Figure 2a−c, respectively. For all three target properties, the ML models perform best on either random or stratified splits (that adhere to the independently and identically distributed (IID) principle) and worse on cluster-I/II or scaffold splits (that do not adhere to the IID principle). For ionic conductivity property prediction, the Chemprop models perform only slightly better than the LightGBM models (Figure 2a and Table S6) but are significantly better for oxidative stability (Figure 2b and Table S9), attributable to differences in data set size (MP: 18133, EDB-1:10196;). The EDB-2 data set for CE is the smallest data set compared to the other two target properties (142 data points); hence, we leveraged a pretraining strategy of molecular representation to overcome the data-hungry nature of the ML models. For this purpose, a Chemprop model was pretrained on a large LUMO data set containing >100,000 data points,[49] since LUMO is reported to be one of the main descriptors for CE.[39,50] Graph embeddings from this pretrained model for solvent and salt components were fed to the SL algorithm viz partial least-squares regression (PLSR; Supplementary Note 2). The actual number of data points varies across different splits due to the specific criteria used for data splitting (Tables S4, S7, and S10) but was maintained approximately at a ratio of 0.8:0.1:0.1 for train:validation:test sets, ensuring a fair basis for comparison. For CE, the performance of Chemprop on some splits is slightly worse compared to the PLSR model (Figure 2c and Table S12). The parity plots for the best performing models on respective test sets: Chemprop trained on EDB-1 (stratified; Table S6), Chemprop trained on MP (random; Table S9), and PLSR trained on EDB-2 (stratified; Table S12) are also shown in Figure 2d−f, respectively, and in Figure S8. These models were finally utilized for making predictions of unseen electrolytes. Additionally, the test MAE for conductivity prediction obtained over all electrolyte classes is 0.09 log mS cm^{-1} (1.23 mS cm^{-1}) is comparable to the Advanced Electrolyte Models' MAE of 0.66 mS cm^{-1} that was tested on electrolytes containing only carbonates[11] and to the error within the experimental measurements itself (0.80 mS cm^{-1} at room temperature; Table S2). Further details on performance for each split are provided in Supporting Information, Notes 4−6.

To rationalize the relative ML performance based on chemical similarity, Tanimoto coefficients (TC)[51] are used. The Tanimoto coefficients lie between 0 (chemically dissimilar) and 1 (chemically similar), where a TC value of 1 does not necessarily mean that the two molecules are identical but signifies maximum chemical similarity (Figure 2g). The average of maximum TC values ($\langle TC^{max} \rangle$; Supplementary Note 7) between train and validation and that between train and test splits for cluster/scaffold splits, in general, are very low compared to random/stratified splits (Figure S9), denoting that the cluster/scaffold split models struggle to predict on chemically dissimilar electrolytes in the validation and test sets, leading to highest prediction errors compared to the other splits (Figure 2). Finally, our best-performing models for the three target properties are compared against other recently reported ML-based studies in Figure 2h and Table S1. Compared to all these recent works, our work develops predictive ML models for all three figures of merit, with trained models that achieve accuracies either surpassing or matching previous efforts while enabling predictions across a wider chemical space. This justifies our choice of using Chemprop, LightGBM, and PLSR models for building different electrolyte property prediction tasks. Since our ML models can accept solvents, salts, and their compositions as inputs, they are capable not only of identifying unexplored solvent-based electrolytes but also of optimizing electrolyte formulations for specific properties. The focus of this study, however, is on discovering distinct solvent-based electrolytes, reflecting the significant challenge of identifying efficient solvents suitable for next-generation lithium batteries—a challenge emphasized by the fact that, to date, only approximately 10^2 solvents have been identified for general lithium battery applications (Figure 1e).

Obtaining Insights into the Trained Models. To establish whether the trained ML models are physically/chemically consistent with the existing scientific knowledge of electrolytes, we employed an interpretable ML approach called SHAP (Supplementary Note 8).[15,52] This approach was applied to traditional SL algorithms such as the LightGBM and PLSR since these models provide similar performance to Chemprop (Figure 2a,c, Tables S6 and S12 for ionic conductivity and CE prediction on stratified splits).

The SHAP summary plot of the LightGBM model for ionic conductivity (stratified) is shown in Figure S10a. The top five most relevant features for the conductivity property according to the SHAP method are the temperature (T), molecular weight of solvents ($MW_{solvent\text{-}1/2}$), the molecular weight of salt (MW_{salt}), and salt concentration (c_{salt}). From Figure 2i, we find that higher temperature values tend to have higher SHAP values, and vice versa, where higher SHAP values denote high ionic conductivity values. The SHAP dependence plot (Figure 2i) for temperature confirms the general trend found in Figure S10a which is also consistent with the widely known temperature dependence of electrolyte ionic conductivities.[53] Interestingly, it is to be noted that at low temperatures ($T < 0$ °C) and high salt concentrations, most data points have very low predicted conductivity values while at high temperatures, moderate to high salt concentrations lead to very high predicted values.[54] This trend is quite revealing and is seldom discussed in the literature. The SHAP values for the salt concentration feature—and hence ionic conductivity—increase initially as a function of the salt concentration and then start decreasing (Figure 2j). This behavior can be attributed to the classical conductivity law relating to ion mobility and conductivity.[55] The maximum SHAP values (or conductivity) are found near $c_{salt} \approx 1{-}1.5$ M, which is consistent with the findings for other conventional electrolytes.[11,55] This behavior can be attributed to the classical conductivity law:

$$\sigma = \mathrm{ne}\mu \tag{1}$$

where n, e, and μ are the amount of charge carriers, elementary charge, and mobility, respectively. Increasing salt concentration increases the amount of charge carriers and ionic conductivity until a concentration is reached where contact-ion pairs and aggregates decrease the number of available charge carriers, and high viscosity decreases mobility, leading to lower conductivities. The LightGBM model extracted this knowledge, even though it was not explicitly provided during the training process. Moreover, two distinct but nontrivial decreasing trends are observed at low and high temperatures, signifying different transport mechanisms at low and high temperatures, respectively. We emphasize that these trends emerge only upon post hoc analysis using the SHAP technique and are not readily apparent from a direct examination of the raw data. As illustrated in Figure S10, no clear trends are observed when ionic conductivity is plotted against salt concentration and temperature for all electrolytes in the raw data set, underscoring the necessity of the SHAP method to uncover subtle, physically consistent trends and correlations. Thus, SHAP analysis enables us to identify the complex, nontrivial interactions between the two most influential factors—temperature and salt concentration—and their coupled effects on ionic conductivity. Further SHAP analyses for oxidative stability and CE are discussed in Supplementary Note 8, where increased salt concentration is found to correlate with higher CE values (Figure S11); a design strategy used for high salt concentration electrolytes.[40,56] We also employed model sensitivity analysis (called Sobol analysis) to gauge the effect of the sensitivity of models toward the different features utilized (Supplementary Note 8), which are in good agreement with the SHAP analysis for each target property (Figure S12).

Examining Generalizability on OOD Data. To verify if the trained models are generalizable, i.e., make accurate predictions on unseen electrolytes, they are deployed on out-of-distribution (OOD) data sets for each property, compiled in-house and from literature. To demonstrate that these data sets belong to different distributions compared to the training data sets, we conducted maximum TC with respect to training data points (TC^{max}), density estimation, and Jensen-Shannon divergence analyses (details in Supporting Information, Figure

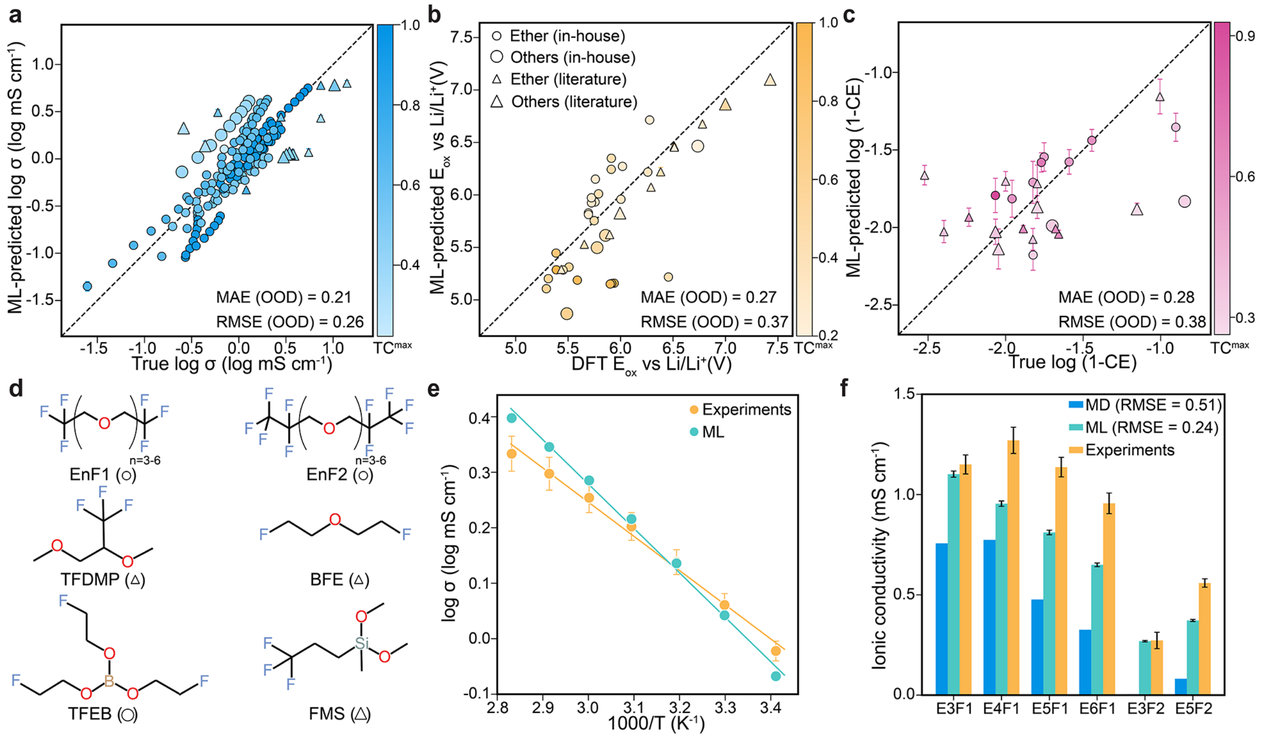


Figure 3. Generalizability of the trained models. Performance of the trained best-models on OOD data sets of (a) ionic conductivity, (b) oxidative stability, and (c) CE labels. The scatter points are colored according to the TC^{max} values with respect to the EDB-1, MP, and EDB-2 data sets and the corresponding color bars (blue, orange, and purple) are shown alongside in (a−c), respectively. (d) Representative molecular structures from OOD data sets, where symbols in the parentheses correspond to solvent class and data source according to the legends shown in (b). All OOD solvent molecules from the literature are shown in Figure S13. (e) Plot showing Arrhenius fitting of the ML-predicted and experimental ionic conductivity data for 1 M lithium bis (fluorosulfonyl)amide (LiFSA)/E3F1 electrolyte. (f) Comparison of ML-predicted, MD-calculated, and experimental ionic conductivities for EnF1- and EnF2-based electrolytes at 30 °C. The MD-derived values in (f) and the experimentally obtained values in (e,f) were extracted from Chen et al.[57] and Ma et al.,[3] respectively. TFDMP = 1,1,1-trifluoro-2,3-dimethoxypropane,[58] BFE = bis(2-fluoroethyl) ether,[59] TFEB = tris(2-fluoroethyl) borate,[60] FMS = dimethoxy(methyl)(3,3,3-trifluoropropyl)silane.[61] The error bars in the ML predictions correspond to the standard deviation in the predictions from ensembles of trained models (15 for Chemprop and 5 for PLSR; details in Supplementary Note 2).

S9). For oxidative stability calculations, we employed the exact same settings as those used in the MP data set to enable a direct one-to-one comparison. This included utilizing THF (tetrahydrofuran) solvent parameters in the implicit solvation calculations. The use of THF ensures that all compounds are evaluated consistently, providing a uniform basis for screening purposes. For each target property, the performance of the corresponding best models on the three OOD data sets is shown in Figure 3a−c, as they performed better compared to other models (refer to Tables S6, S9, and S12 for performance comparison). Overall, the Chemprop models for both conductivity and oxidative stability perform reasonably well on all electrolyte solvent classes with lower error for solvents with high TC^{max} values with respect to the EDB-1 and MP data sets (Figure 3a,b). The low CE performance on some of the OOD data points is not unexpected and the relative prediction errors also correlate with the TC^{max} values with respect to the EDB-2 data set (Figure 3c). Consequently, TC^{max} is proposed as a useful metric to gauge trustworthiness in ML predictions for uncharted chemical territories, with caution advised for molecules with $TC^{max} < 0.5$. The representative electrolyte solvent molecules present in the OOD data sets are illustrated in Figure 3d, while all the OOD molecules extracted from the literature are shown in Figure S13, to showcase the reasonable ML performance on diverse solvent classes such as fluorinated ethers, borate esters, and silyl ethers.

Interestingly, the Chemprop model also learned a phenomenological relation, namely, the Arrhenius equation, without having to be built into the architecture, e.g., Bradford et al.'s ChemArr architecture[25] (Figure 3e). The experimental and ML-predicted activation energies are also comparable for the unseen E3F1-based solvent electrolyte (11.80 and 15.33 kJ mol^{-1}, the difference is close to the thermal energy at room temperature). Moreover, the ML-predicted ionic conductivities for the E3F1-related compounds were also compared with those evaluated using classical molecular dynamics (MD) calculations using fixed-charge force fields by Chen et al.[57] as shown in Figure 3f. Despite the MD-derived ionic conductivities being evaluated based on ionic solvation structures, information inaccessible to the ML models, ML-predicted ionic conductivities are on par with the MD-derived ones. Using more accurate polarizable force fields could yield more precise predictions compared to the current results; however, these force fields are more computationally demanding (up to 3-fold higher) and are only available for a limited range of solvent classes.[63] Therefore, leveraging ML approaches such as ours or machine learning interatomic potentials[64] can mitigate computational costs associated with MD for obtaining key electrolyte properties.

Screening Promising Electrolytes from Molecular Repositories. The developed ML models for the three target properties were finally deployed on large unlabeled molecular

https://doi.org/10.1021/acs.chemmater.4c03196
Chem. Mater. 2025, 37, 2720−2734

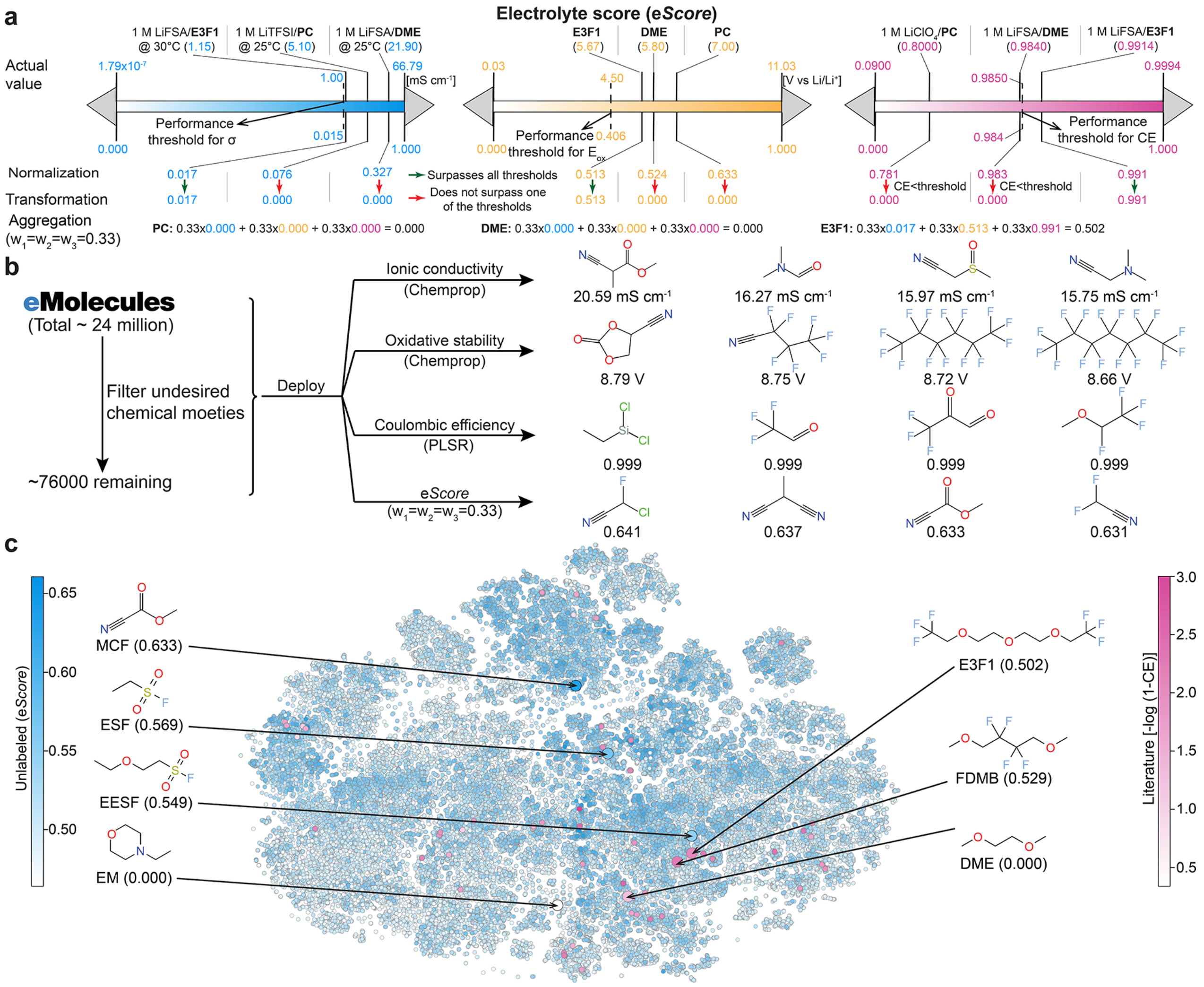


Figure 4. Assigning the electrolyte score and screening promising electrolytes. (a) Electrolyte score (e*Score*) concept demonstrated through the example of three common electrolytes: PC, DME, and E3F1. Equal weights of 0.33, 0.33, and 0.33 are considered for ionic conductivity, oxidative stability, and CE, respectively, in evaluating the aggregation step. The oxidative stability values for DME and PC have been taken from ref 62. (b) Top solvent molecules from the eMolecules repository, as predicted by the three ML models that fall within the respective performance thresholds. The fourth row shows the top solvent molecules according to the predicted e*Scores* (using equal weights of 0.33). (c) A 2D t-SNE plot comparing chemical space of the solvent molecules, comprising those reported in the literature for CE (from both EDB-2 and OOD data sets; in -log(1-CE) scale for clarity), and the candidate solvent molecules from the unlabeled repositories (eMolecules, PubChem, and GDB-17 data sets) having nonzero predicted e*Scores*, except for EM. Representative solvent molecules from the literature CE data sets and solvent molecules from the eMolecules repository selected for experimental verification are highlighted in larger circles with corresponding e*Scores* provided within parentheses below. All ML predictions were carried out by fixing the electrolyte formulation containing given solvent molecules with the LiFSA salt. Uncertainty estimates for the ML predictions are not shown for the sake of brevity.

repositories such as eMolecules,[65] PubChem,[66] and GDB-17.[67] The three combined databases make up a total of ~185 million. Since some chemical moieties are well-known for their incompatibility with LMBs, they were removed (Figure S14a), and the remaining molecules acted as a search space for screening unexplored promising electrolyte solvents. For predicting electrolyte properties, particularly ionic conductivity and CE, the identity of the salt, in addition to the solvent, is crucial. LiFSA salt was chosen for this study due to its reported advantages, including higher solubility, lower viscosity, greater ionic conductivity, and enhanced electrochemical stability, making it particularly suitable for our research objectives. The focus of this work is on optimizing electrolytes through the discovery of unexplored solvents rather than relying on formulations derived from existing solvents. Subsequently, the three most generalizable ML models were deployed on the compatible data sets created from filtered molecules (further details can be found in Supporting Information, Note 10). For this work, we focus on the eMolecules database, since it consists of commercially available compounds. We find that a total of 97 fluorinated molecules are found among the top 100 candidates predicted for E_{ox} (compared to 4 and 74 for conductivity and CE, respectively). This demonstrates that the ML models have grasped important chemical trends such as fluorination leading to high oxidative stability and a heuristic that many fluorinated solvents support LMBs with high CE. However, if only one figure of merit is used, Figure 4b shows that a top-predicted molecule for one property does not appear in the top predictions for another. This necessitates the need for simultaneous consideration of multiple figures of merit. This also demonstrates the need for a single metric to rank and acquire unexplored compounds. Therefore, to find molecules

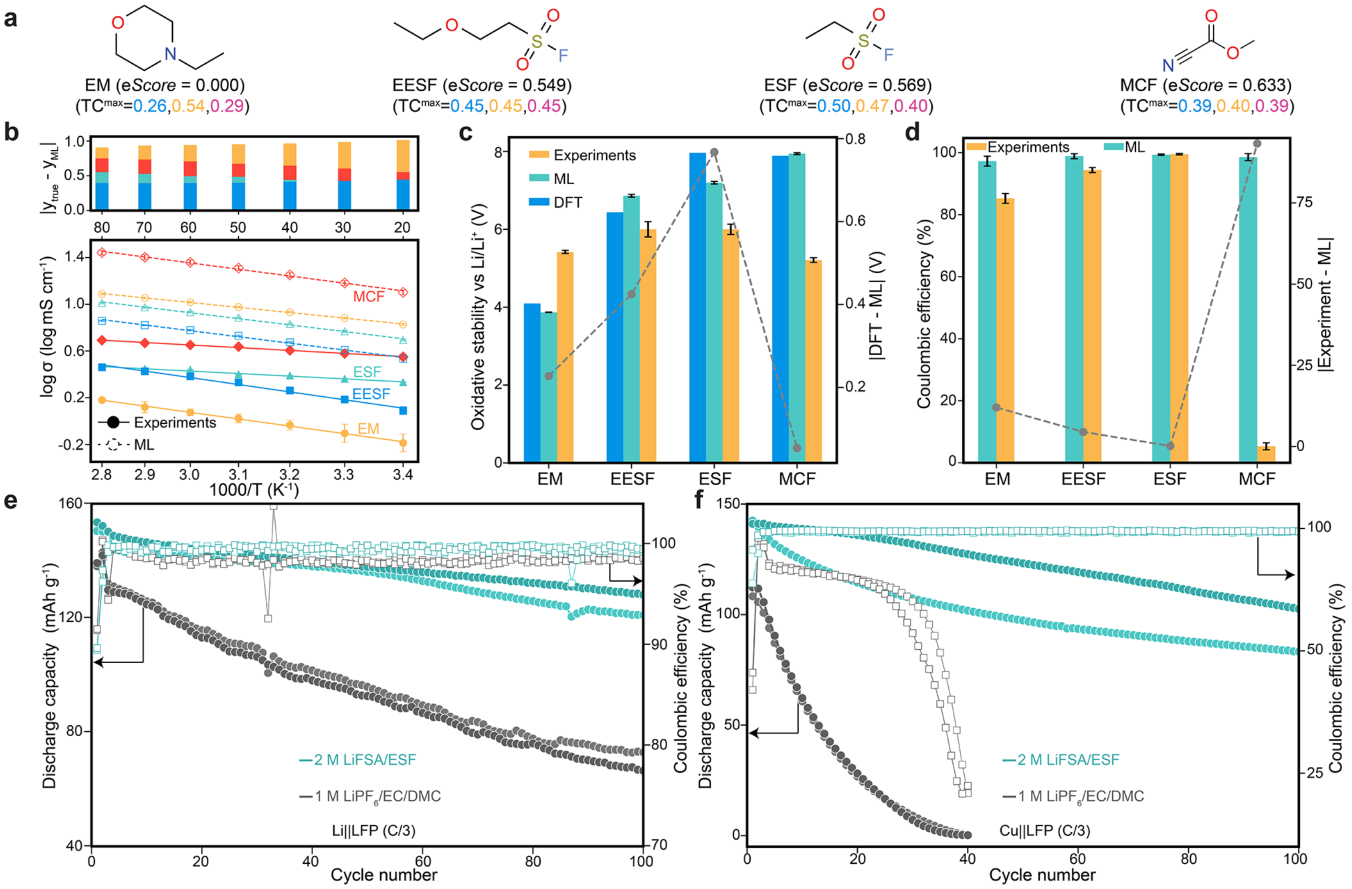


Figure 5. Experimental validation of the chosen solvent electrolyte candidates. (a) Molecular structure of EM, EESF, ESF, and MCF along with the predicted e*Scores* using equal weights ($w_1 = w_2 = w_3 = 0.33$). The TC^{max} values with respect to the EDB-1 (blue text), MP (orange text), and EDB-2 (purple text) data sets are also provided below the e*Scores*. (b) Comparison of ML-predicted and experimental ionic conductivities on a common log scale as a function of temperature for the four 1 M LiFSA solvent electrolyte candidates. (c) Comparison of DFT-calculated and ML-predicted solvent oxidative stabilities (absolute error shown by gray scatter points corresponding to the *y*-axis on the right side) and experimental 1 M LiFSA electrolyte oxidative stabilities measured using the LSV method with Li||Al cells. (d) Comparison of ML-predicted and experimental Li||Cu CE values for 2 M LiFSA solvent in electrolyte candidates (1 M LiFSA for MCF; absolute error shown by gray scatter points corresponding to the *y*-axis on the right side). Cycling performance of 2 M LiFSA/ESF and commercial carbonate electrolytes using (e) Li||LFP and (f) anode-free Cu||LFP coin cells. Cycling of two duplicate cells (in dark and light shades) for each electrolyte has been shown in (e,f). 1 C = 1.66 mA/cm^2. The square and circle scatter points in (e,f) represent CE and discharge capacity, respectively. All ML predictions and experimental tests were carried out by fixing the electrolyte formulation containing given solvent molecules with the LiFSA salt. The error bars in the ML predictions and experimental data in (b−d) correspond to standard deviation in the predictions from ensembles of trained models and two duplicate cells, respectively. The statistics of all cells employed for (e,f) are summarized in Table S18.

that satisfy all three performance thresholds simultaneously and to rank them based on the predicted values, we devised a unified metric called the electrolyte score (e*Score*). The e*Score* is computed in three steps: (1) normalization of property values based on data set ranges, (2) threshold transformation, where all properties are set to zero if any electrolyte fails to meet predefined performance criteria, and (3) weighted aggregation, allowing for flexible prioritization of different properties. This approach ensures that selected candidates perform well across multiple electrolyte design criteria. It reflects an electrolyte's ability to simultaneously meet these requirements, with higher values indicating higher performance across the three metrics. Therefore, an electrolyte with a higher e*Score* should perform better compared with electrolytes with lower values. The e*Score* is also quite flexible as the weights can be modified to give more weight to the desired target property and can even be extended to include more figures of merit such as viscosity, flammability, etc.; hence, it is data set- and property-agnostic. In the present study, we considered a total of four weights, and the top predicted molecules according to equal weights (of 0.33 for each property) are shown in Figure 4b. Additional details on the e*Score* formulation, rationale for different weighting strategies, and the representative structures with top predicted e*Scores* according to the four different weights for the eMolecules repository are provided in Figure S14b, Supplementary Note 10, and Figures S15−S18, respectively. It is indeed observed that different molecules are assigned different e*Scores* depending on the chosen weights. Therefore, as a heuristic scalarized ranking metric[68] for selecting promising candidates, e*Score* requires prior knowledge of the relative importance of the target properties. Future work will focus on developing a Bayesian multiobjective optimization[69] framework to select optimal electrolytes from the Pareto front.

We subsequently selected four solvent molecules from the filtered, unlabeled eMolecules corpus: three with a high predicted e*Score* and one with a predicted e*Score* of zero for experimental verification (discussed in detail in Supplementary

Note 10), none of which had previously been explored as electrolyte solvents: ethanesulfonyl fluoride (ESF), methyl cyanoformate (MCF), 4-ethylmorpholine (EM), and 2-ethoxyethane-1-sulfonyl fluoride (EESF) that are shown in Figure 4c along with their predicted e*Scores*. These molecules were chosen to cover different and unique classes of organic molecules based on availability and cost (detailed rationale in Supporting Information, Note 10). Due to high cost or commercial unavailability, we could not investigate every top 100 ranked compound (also explained in Supporting Information, Note 10). A DME (1,2-dimethoxyethane)-based and a commercial electrolyte (1 M $LiPF_6$ in ethylene carbonate (EC)/dimethyl carbonate (DMC)) were chosen as baselines for comparison. We find that the selected molecules span different regions of the chemical space (Figure 4c), and a vast majority of the remaining space remains unexplored. Most of the state-of-the-art LMB electrolytes are clustered into the ether-containing region, e.g., E3F1, DME, and fluorinated 1,4-dimethoxylbutane (FDMB). In contrast, all four selected compounds belong to distinct solvent classes ($TC^{max} \leq$ or ∼0.5), with two of them sulfonyl-based (ESF and EESF).

Unexplored Electrolyte Discovery. Next, electrolytes formulated with the four selected candidate solvent molecules (with LiFSA salt for both prediction and testing at the same conditions; Supplementary Note 10) underwent experimental validation. Comparison of experimentally validated and ML-predicted properties for EM, EESF, ESF, and MCF are shown in Figure 5. The ML-predicted ionic conductivities for electrolytes other than EM (lowest TC^{max} value compared to other solvent candidates; Figure 5a) are in reasonable agreement with the experimental labels at ambient temperatures with some deviation at higher temperatures (Figure 5b). The increased errors at higher temperatures can be attributed to the limited data points at these temperatures compared to the ambient ones in the EDB-1 data set (Figure S4a). For oxidative stability, good agreements are found between DFT-calculated and ML-predicted values, and they exhibit high experimental oxidative stabilities (Figure 5c). Using 2 M LiFSA electrolytes for CE measurements, the disagreements between the ML predictions and the ground truth for 2 M LiFSA solvent electrolytes decrease in the following order: EM < EESF < ESF. The error for the EM electrolyte can again be attributed to its lowest TC^{max} with respect to the EDB-2 data set. The MCF-based electrolyte was predicted to have an e*Score* of 0.633, attributed to the overestimated CE for MCF that likely arises from the lack of nitrile representation in the EDB-2 data set. Our domain knowledge suggested that it would likely perform poorly under low salt concentrations and without additives or cosolvents. Figure 5d confirms this, demonstrating that while models can effectively narrow down the chemical space, domain expertise remains critical to further filter out false positives in data-deficient scenarios. Nevertheless, ESF (0.569) having higher e*Scores* than EESF (0.549) and EM (0.000) performs better than the other two electrolytes in terms of all three properties—ionic conductivity (Figure 5b), oxidative stability (Figure 5c), and CE (Figure 5d). In fact, at least three out of the four selected electrolytes follow the order of their predicted ranking for the three properties by the respective ML models. As a result, the relative trends in all three experimentally obtained properties for EM, EESF, and ESF follow the order of their predicted e*Scores*, and hence ML predictions remain useful for identifying promising candidates. The discussion of structure−property relationships for the selected electrolytes using SHAP analysis is also provided in Supplementary Note 11 and shown in Figure S21. Given the wide variability in CE performance observed among the four selected electrolytes, despite using the same LiFSA salt, it is evident that the differences in performance are attributable to the solvents rather than the salt. Future work will focus on developing custom ML architectures tailored for electrolyte systems, aimed at improving predictions for unexplored compounds, even for properties with limited data availability, such as CE.

In an advance for ML-enabled electrolyte discovery, the 2 M LiFSA/ESF electrolyte was found to have an exceptionally high CE of 99.5% (Figure S20) that performs on par with the state-of-the-art LMB liquid electrolytes such as FDMB[2], F5DEE,[70] etc., in terms of all three desirable figures of merit (Table S19). For LMBs, the 2 M LiFSA/ESF electrolyte enables much-improved cycling (Li||LFP in Figure 5e and anode-free Cu||LFP in Figure 5f) that outperforms carbonate electrolytes. The sulfonyl fluoride ($R\text{-}SO_2\text{-}F$) is chemically distinct from previously reported sulfamoyl fluorides ($(R)_2\text{-}N\text{-}SO_2\text{-}F$)[71] and sulfones ($R\text{-}SO_2\text{-}R$).[72] Contrary to Jiang et al.'s investigation of ESF as a 3% additive in conventional electrolytes,[73] which showed no improvements due to substrate dependency, our research focused on discovering unexplored electrolyte solvents. To broaden the scope of the work, we applied the proposed electrolytomics approach to address challenges with next-generation LIBs. The ESF-based electrolytes show highly reversible ambient (20 °C), high (60 °C), and low temperature (−20 °C) (Figure S23) cycling with high voltage NMC811 ($LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$) cathodes, outperforming conventional carbonate-based electrolytes (complete discussion in Supplementary Note 12). To explain the underlying mechanism of improved battery performance with ESF-based electrolytes, experimental characterizations in the form of scanning electron microscopy (SEM, Figure S24), nuclear magnetic resonance (NMR, Figure S25), and X-ray photoelectron spectroscopy (XPS, Figures S26 and S27) were further carried out (Supplementary Note 13). The SEM analysis confirms smoother lithium deposition morphology and larger particle size with minimum lithium dendrite formation, the NMR depicts no solvent degradation, and the XPS spectra signify a lithium or graphite interface enriched with inorganic decomposition products with the ESF-based electrolytes compared with commercial carbonate electrolytes. The 2 M LiFSA/ESF electrolyte also exhibits better flame-retardant properties compared to the carbonate electrolyte (Supplementary Note 13 and Figure S28). The impact of this work lies not solely in the discovery of ESF but in the quantitative platform to accelerate the discovery of electrolytes for all battery and electrochemical applications.

■ CONCLUSIONS

In summary, we devised a data-centric heuristic framework to guide the discovery of unexplored liquid electrolytes that are compatible with NGBs. The present study develops relevant data sets and ML models for crucial electrolyte properties—ionic conductivity, oxidative stability, and CE—for discovering efficient electrolytes compatible with next-generation batteries. Post hoc analyses using SHAP, and sensitivity analysis techniques reveal trivial and nontrivial trends in electrolytes. We define a quantitative metric (e*Score*) for electrolytes and showcase the discovery of a distinct class of electrolyte solvents based on sulfonyl fluorides for NGBs. Our work reveals the power of AI in conjunction with computation and experiments to accelerate battery materials discovery that solves a wide range of battery challenges and provides a foundational framework for the development of advanced forward ML design using transformers or chemistry-infused architectures, inverse ML design using generative algorithms, and closed-loop optimization using active learning.

■ METHODS

Data Collection and Curation of EDB Data Sets. A wide breadth of literature (over 250 research articles and review papers) pertaining to liquid electrolytes for lithium-based batteries was collected and processed for both the EDB-1 and the EDB-2 data sets. From the literature, data points were manually extracted from the text, tables, and/or figures present in either the main article or the Supporting Information. The existing natural language processing (NLP) tools face challenges when employed to extract data from such varied sources. Because the desired experimental parameters and measurements were often reported in different units between articles, all data points were standardized to a common unit when possible, e.g., Celsius for temperature, molarity for salt concentration, mS/cm for ionic conductivity, volume fraction (v:v:v:v) for solvent ratio, mA/cm^2 for current density. Particular attention was paid to the standardization of data points such as the solvent ratio due to the significant variety of reported units, e.g., mole fraction, weight fraction, molarity, and molality. To convert such units into the standardized unit-like volume fraction for the solvent ratio, the relevant information such as the density and molecular weight of solvent molecules was determined from credible sources, which were primarily PubChem and chemical vendors. The data points for which no density information could be extracted from anywhere and the solvent ratios were taken as reported in the literature. Several works also reported conductivities measured at 'room temperature' instead of the exact temperature: the temperature was assumed to be 25 °C in these cases. Data from plots (images) were extracted by utilizing WebPlotDigitizer.[74] The simplified molecular-input line-entry system (SMILES) for solvent, salt, and additive molecules were obtained from PubChem by entering the chemical names or structures reported in the literature. The molecules that were missing in the PubChem repository, were drawn manually on the Open Babel[75] utility to obtain the corresponding SMILES. Chemical drawing software such as Marvin was also used in determining SMILES or scientific names. The data sets then went through a rigorous editing process where the data sets were cross-checked by several group members to account for any human errors. Furthermore, as the literature included both primary research papers and secondary review articles, duplications of data were removed. Finally, the data sets were curated by removing entries that were missing key parameters such as temperature, solvent ratios, etc. The SMILES notation was standardized using the RDKit (version 2023.3.2)[76] Python library. The whole process is illustrated in detail in Supplementary Note 1 in SI.

ML Inference. All codes utilized in this study were built by using various libraries available in the Python programming language. The three types of popular molecular representations utilized in the manuscript, (i) common hand-crafted descriptors were generated using RDKit and fed to a shallow learning (SL) algorithm, e.g., LightGBM[77] (version 4.1.0) (ii) directed molecular graphs utilized in an end-to-end manner within the python library Chemprop[48] (and supplemented with additional RDKit features; version 1.5.2), which is an end-to-end deep learning framework used previously for antibiotic discovery,[12] as well as the graph embeddings obtained from trained Chemprop model. For an in-depth discussion on different settings employed in training and analyzing ML models, refer to Supporting Information, Notes 4–6, and 8. The RDKit library was also utilized for evaluating the Tanimoto coefficients. Scikit-learn (version 1.3.0)[78] library was employed for building the SL (LightGBM, RF, and PLSR) models. The hyperparameter optimization for the SL algorithms was carried out using the the Optuna (version 3.4.0) library. SHAP (version 0.44.1) library was utilized for interpreting SL models. Sobol analysis for conducting model sensitivity analysis was carried out using the SALib library (version 1.4.7). The reduced embeddings of chemical space for the t-SNE plots were generated using the OpenTSNE library (version 1.0.1). All plots in the manuscript were made using matplotlib, seaborn, or plotly libraries. All predicted values of ionic conductivity, oxidative stability, and CE for the eMolecules database, along with their e*Scores*, are uploaded to the GitHub repository. All codes have been tested to work on the Intel Xeon Gold 6346 CPU (32 cores) and Nvidia A100 (4 GPUs; for Chemprop).

DFT Calculations. The adiabatic ionization energy calculations were carried out using DFT as implemented in the Gaussian 16 (version 16.0.1) program.[79] To match the level of theory employed in the MP database for fair comparison, B3LYP and 6-31*g(d) were utilized as the functional and basis sets, respectively. Implicit solvation corrections considering tetrahydrofuran (THF) as the solvent were also included using the polarizable continuum model (PCM) method. The DFT oxidative stability values were obtained by referencing the adiabatic ionization energies with respect to the Li/Li$^+$ scale by subtracting a constant (1.4).[62] The oxidative stability values calculated at a higher fidelity level (B3LYP/6-311**g(d,p)/SMD(THF)) are also available at GitHub.

Materials. Lithium bis(fluorosulfonyl)amide (LiFSA, 99.5%) salt was provided by Arkema, and was dried under vacuum in the integrated glovebox heating chamber at 120 °C for 2 days and kept inside an argon-filled glovebox (VigorTech, O_2 and H_2O content <1 ppm). 4-ethylmorpholine (EM, 98%) was purchased from Thermo Fisher, and methyl cyanoformate (MCF, 99%) was purchased from Sigma-Aldrich. 2-ethoxyethane-1-sulfonyl fluoride (EESF, 99%) and ethanesulfonlyl fluoride (ESF, 99%) were purchased from Enamine US INC. All solvents were dried with 4 Å molecular sieves. A calculated amount of LiFSA salt was dissolved separately in 4 different solvents to prepare 1 M LiFSA liquid electrolytes. A commercial electrolyte (1 M $LiPF_6$ in EC-DMC, v/v, 1/1) was used as the reference electrolyte. LFP ($LiFePO_4$), NMC811 ($LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$), and Gr (graphite) electrodes were received from the Cell Analysis, Modeling, and Prototyping (CAMP) facility at Argonne National Laboratory. LFP electrode has a total mass loading of 12.1 mg cm^{-2} (coated on Al foil with 20 μm thickness) with 90 wt % Johnson Matthey LFP, 5 wt % Timcal C-45, and 5 wt % Solvay 5130 PVDF binder. NMC811 electrode has a total mass loading of 9.08 mg cm^{-2} (coated on Al foil with 20 μm thickness) with 90 wt % Targray NMC811, 5 wt % Timcal C-45, and 5 wt % Solvay 5130 PVDF binder. Gr electrode has a total mass loading of 6.35 mg cm^{-2} (coated on Cu foil with 10 μm thickness) with 91.83 wt % Hitachi MagE3, 2 wt % Timcal C-45, 6 wt % Kureha 9300 PVDF binder, and 0.17 wt % of oxalic acid. The theoretical capacity was 150 mAh cm^{-2}. The electrodes were cut into disks of 12 mm and dried under vacuum in the integrated heating chamber at 120 °C for 1 day before usage. All coin cell parts (CR2032) were ordered from Xiamen TOB New Energy Technology. Thick lithium foil (thickness: 500 μm, purity: 99.9%, MTI Corporation) and thin lithium foil (thickness: 20 μm, Argonne National Laboratory) were brushed to remove oxide layers and cut into disks of 12 mm in diameter. Celgard 3501 and Celgard 2325 separators were obtained from Celgard LLC, cut into 18 mm disks, and dried at 70 °C using a BUCHI B-585 glass oven before usage.

Ionic Conductivity Measurements Using Electrochemical Impedance Spectroscopy (EIS). Ionic conductivity of liquid electrolytes was measured using EIS in SS||SS coin cells, which were prepared in the argon-filled glovebox (O_2, H_2O < 1 ppm) with the following configuration: Stainless steel (SS)||25 μL of electrolyte|| 1 separator (Celgard 3501)||25 μL of electrolyte||SS. The EIS was performed using a Biologic VSP-300 Potentiostat in the frequency range of 7 MHz and 100 Hz at a temperature range of 80 to 20 °C with an interval of 10 °C. Coin cells were held at each temperature in an ESPEC environmental chamber (BTZ-133) for 1 h before EIS testing. A platinum-cell conductivity probe (Vernier) was used to measure the conductivity of electrolytes at 20 °C. A cell constant of ~10 was used to calculate the realistic ionic conductivity of electrolytes using the EIS method.

Oxidative Stability Measurements Using Linear Sweep Voltammetry (LSV). LSV of coin cells was performed using a Biologic MPG-2 potentiostat at 20 °C with two different working electrodes of SS and aluminum. Coin cells were prepared using the following configuration: SS||Li||25 μL of electrolyte||1 separator (Celgard 3501)||25 μL of electrolyte||SS and SS||SS||Li||25 μL of electrolyte||1 separator (Celgard 3501)||25 μL of electrolyte||Al. All cells were assembled in the glovebox and rested for 10 h before LSV testing. The cell voltage was scanned from the open-circuit voltage to 6 V at a rate of 1 mV/s.

Li||Cu Coulombic Efficiency (CE) Measurement Using the Modified Aurbach Method. Li||Cu coin cells were prepared with the following configuration: SS||Li||25 μL of electrolyte||1 separator (Celgard 2325)||25 μL of electrolyte||Cu||SS. A Biologic MPG-2 potentiostat was used to cycle the Li||Cu coin cells. In the CE measurement, a current density of 0.5 mA cm^{-2} was applied for all steps. First, the formation cycle was performed by depositing lithium on a copper electrode for 10 h and stripping to 1 V. Next, another 10 h of Li deposition on a copper electrode was performed. Next, the coin cells were cycled for 10 cycles of 2 h-depositing/2 h-stripping. Finally, Li was stripped from the Cu electrode until the voltage reached 1 V. CE was calculated as the ratio of total stripping capacity over total depositing capacity (excluding the formation cycle).

Full Cell Cycling for LMBs, AFBs, and LIBs. The Li||LFP and Cu||LFP coin cells were prepared with the following configurations: SS||Li||25 μL of electrolyte||1 separator (Celgard 3501)||25 μL of electrolyte||LFP||SS, and SS||SS||Cu||25 μL of electrolyte||1 separator (Celgard 3501)||25 μL of electrolyte||LFP||SS, respectively. Coin cells were rested for 10 h, continued to have 3 formation cycles at the C/10 rate, and cycled at the C/3 rate using a Neware BTS4000 battery tester at 20 °C. LFP has a full capacity of 150 mAh/g. The mass loading of LFP is 12.1 mg/cm^2 which is equivalent to 1 C = 1.81 mA/cm^2. Li||Gr coin cells were prepared with the following configurations: SS||Li||25 μL of electrolyte||1 separator||25 μL of electrolyte||Gr||SS. A graphite full capacity of 372 mAh/g is used for calculating the current rate of the Li||Gr cell (1 C $\approx$ 2.17 mA/cm^2). Gr||NMC811 coin cells were prepared with the following configurations: SS||SS||Gr||25 μL of electrolyte||1 separator||25 μL of electrolyte||NMC811||Al, and an aluminum-coated cathode case was used. An NMC811 full capacity of 203 mAh/g is used for calculating the current rate of Gr||NMC811 cell (1 C $\approx$ 1.66 mA/cm^2, n/p $\approx$ 1.3). Celgard 2325 was used for ESF electrolytes while Celgard 3501 was used for carbonate electrolytes. Two formation cycles at C/10 were performed at 20 °C before cycling at the desired current rate and temperature.

Symmetric Li||Li Cycling. Li||Li coin cells were prepared with the following configurations: SS||Li||25 μL of electrolyte||1 separator (Celgard 2325)||25 μL of electrolyte||Li||SS. Li||Li coin cells were cycled using a Neware BTS4000 battery tester at 20 °C. Coin cells were rested for 10 h and had 5 formation cycles at 0.1 mA/cm^2−0.1 mAh/cm^2, and continued to cycle at 1 mA/cm^2−1 mAh/cm^2.

Scanning Electron Microscopy (SEM) Analysis. SEM characterization utilized a Carl Zeiss Merlin field emission scanning electron microscope. The lithium-deposited samples were prepared using Li||Cu cells with three different electrolytes: commercial 1 M $LiPF_6$-EC-DMC (v/v, 1/1) and 2 M LIFSA-ESF electrolytes. First, five precycles with a current density of 0.02 mA cm^{-2} within the voltage range of 0−1 V were conducted before lithium deposition to cleanse the copper surface. Subsequently, lithium was deposited onto the copper electrode at 0.5 mA cm^{-2} for 1 mAh cm^{-2}. Li||Cu cells were opened in an argon-filled glovebox, and the copper foil-bearing deposited lithium underwent a rinsing process with 1,2-dimethoxyethane 5 times to eliminate lithium salt.

X-ray Photoelectron Spectroscopy (XPS) Analysis. XPS analysis was carried out using a PHI 5000 VersaProbe II System (Physical Electronics), employing an Al Kα radiation beam ($h\nu$ = 1486.6 eV) with 100 μm and 25 W, along with Ar$^+$ and electron beam sample neutralization in fixed analyzer transmission mode. Alignment of XPS spectra was conducted with respect to the C−C component in the C 1s spectra at 284.6 eV.

■ ASSOCIATED CONTENT

Data Availability Statement

The data are available at GitHub repository (https://github.com/AmanchukwuLab/electrolytomics).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.4c03196.

> Detailed discussion and related tables and figures on data extraction and curation, creation of data splits, different data featurization used, different ML models used, reduced representation of molecules used for building ML models, specific settings used for training ML models for the three properties (ionic conductivity, oxidative stability, and Coulombic efficiency), Tanimoto coefficients, SHAP and sensitivity analyses, OOD data sets, workflow for generating filtered set of molecules for search space, LIB cycling, SEM-, XPS-, and NMR-based comparative analysis for cells cycled with ESF-based and commercial electrolytes, and self-extinguishing tests for ESF-based and commercial electrolyte (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Chibueze V. Amanchukwu − *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-6573-1213; Email: chibueze@uchicago.edu

Authors

Ritesh Kumar − *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, United States*

Minh Canh Vu − *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, United States*

Peiyuan Ma − *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.4c03196

Author Contributions

C.V.A. conceived and supervised the project. R.K. curated the data sets, performed the ML and DFT calculations, and analyzed the results. M.C.V and P.M. performed the experiments. R.K., M.C.V., P.M., and C.V.A. contributed to the discussion of results and writing of the manuscript.

Notes

The authors declare no competing financial interest.

The codes (Jupyter notebooks) and model checkpoints are publicly available on the GitHub repository (https://github.com/AmanchukwuLab/electrolytomics).

■ ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation CAREER Award (CBET-2144454). C.V.A. acknowledges the Neubauer Family Assistant Professors Program. R. K. acknowledges support from the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences program. The authors thank the Research Computing Center (RCC) at the University of Chicago and the Argonne Leadership Computing Facility (ALCF) at

Argonne National Laboratory for providing the required computational facilities. The authors thank Steve Trask and Andrew Jansen at Argonne's Cell Analysis, Modeling, and Prototyping (CAMP) facility for providing the LFP electrode. The authors thank Edward F. Barry at Argonne National Laboratory for providing the Li (20 μm) electrode. The authors thank Michael Han, Christina Pirrotta, Walker Gillett, Chris Birch, Beamlak Lefebo, Benjamin Kash, Claire Fuschi, Cindy Xue, Uyen Le, Trinity Hixson-Wells, and Arnav Brahmasandra for help with extracting data from the literature for the electrolyte database. We also thank Zifeng Kang for helping with the preliminary ML setup. The authors also acknowledge Emily Doyle and Michael Han for providing critical comments on the manuscript.

REFERENCES

(1) Etacheri, V.; Marom, R.; Elazari, R.; Salitra, G.; Aurbach, D. Challenges in the Development of Advanced Li-Ion Batteries: A Review. *Energ Environ. Sci.* **2011**, *4* (9), 3243−3262.

(2) Yu, Z.; Wang, H.; Kong, X.; Huang, W.; Tsao, Y.; Mackanic, D. G.; Wang, K.; Wang, X.; Huang, W.; Choudhury, S.; Zheng, Y.; Amanchukwu, C. V.; Hung, S. T.; Ma, Y.; Lomeli, E. G.; Qin, J.; Cui, Y.; Bao, Z. Molecular Design for Electrolyte Solvents Enabling Energy-Dense and Long-Cycling Lithium Metal Batteries. *Nat. Energy* **2020**, *5* (7), 526−533.

(3) Ma, P.; Mirmira, P.; Amanchukwu, C. V. Effect of Building Block Connectivity and Ion Solvation on Electrochemical Stability and Ionic Conductivity in Novel Fluoroether Electrolytes. *Acs Central Sci.* **2021**, *7* (7), 1232−1244.

(4) Hobold, G. M.; Lopez, J.; Guo, R.; Minafra, N.; Banerjee, A.; Meng, Y. S.; Shao-Horn, Y.; Gallant, B. M. Moving beyond 99.9% Coulombic Efficiency for Lithium Anodes in Liquid Electrolytes. *Nat. Energy* **2021**, *6* (10), 951−960.

(5) Meng, Y. S.; Srinivasan, V.; Xu, K. Designing Better Electrolytes. *Science* **2022**, *378* (6624), No. eabq3750.

(6) Amanchukwu, C. V. The Electrolyte Frontier: A Manifesto. *Joule* **2020**, *4* (2), 281−285.

(7) Kirkpatrick, P.; Ellis, C. Chemical Space. *Nature* **2004**, *432* (7019), 823−823.

(8) Chen, J.; Zhang, D.; Zhu, L.; Liu, M.; Zheng, T.; Xu, J.; Li, J.; Wang, F.; Wang, Y.; Dong, X.; Xia, Y. Hybridizing Carbonate and Ether at Molecular Scales for High-Energy and High-Safety Lithium Metal Batteries. *Nat. Commun.* **2024**, *15* (1), 3217.

(9) Wu, M.; Wang, Z.; Zhang, W.; Jayawardana, C.; Li, Y.; Chen, F.; Nan, B.; Lucht, B. L.; Wang, C. High-Performance Lithium Metal Batteries Enabled by a Fluorinated Cyclic Ether with a Low Reduction Potential. *Angew. Chem., Int. Ed.* **2023**, *62* (8), No. e202216169.

(10) Benayad, A.; Diddens, D.; Heuer, A.; Krishnamoorthy, A. N.; Maiti, M.; Cras, F. L.; Legallais, M.; Rahmanian, F.; Shin, Y.; Stein, H.; Winter, M.; Wölke, C.; Yan, P.; Cekic-Laskovic, I. High-Throughput Experimentation and Computational Freeway Lanes for Accelerated Battery Electrolyte and Interface Development Research. *Adv. Energy Mater.* **2022**, *12*, No. 2102678.

(11) Gering, K. L. Prediction of Electrolyte Conductivity: Results from a Generalized Molecular Model Based on Ion Solvation and a Chemical Physics Framework. *Electrochim. Acta* **2017**, *225*, 175−189.

(12) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackerman, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180* (4), 688−702.e13.

(13) Rajan, A. C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K.-R.; Singh, A. K. Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized MXene. *Chem. Mater.* **2018**, *30* (12), 4031−4038.

(14) Juneja, R.; Yumnam, G.; Satsangi, S.; Singh, A. K. Coupling the High-Throughput Property Map to Machine Learning for Predicting Lattice Thermal Conductivity. *Chem. Mater.* **2019**, *31* (14), 5145−5151.

(15) Kumar, R.; Singh, A. K. Chemical Hardness-Driven Interpretable Machine Learning Approach for Rapid Search of Photocatalysts. *Npj Comput. Mater.* **2021**, *7* (1), 197.

(16) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.

(17) Qu, X.; Jain, A.; Rajput, N. N.; Cheng, L.; Zhang, Y.; Ong, S. P.; Brafman, M.; Maginn, E.; Curtiss, L. A.; Persson, K. A. The Electrolyte Genome Project: A Big Data Approach in Battery Materials Discovery. *Comput. Mater. Sci.* **2015**, *103*, 56−67.

(18) Korth, M. Large-Scale Virtual High-Throughput Screening for the Identification of New Battery Electrolyte Solvents: Evaluation of Electronic Structure Theory Methods. *Phys. Chem. Chem. Phys.* **2014**, *16* (17), 7919−7926.

(19) Husch, T.; Yilmazer, N. D.; Balducci, A.; Korth, M. Large-Scale Virtual High-Throughput Screening for the Identification of New Battery Electrolyte Solvents: Computing Infrastructure and Collective Properties. *Phys. Chem. Chem. Phys.* **2015**, *17* (5), 3394−3401.

(20) Schütter, C.; Husch, T.; Viswanathan, V.; Passerini, S.; Balducci, A.; Korth, M. Rational Design of New Electrolyte Materials for Electrochemical Double Layer Capacitors. *J. Power Sources* **2016**, *326*, 541−548.

(21) Schütter, C.; Husch, T.; Korth, M.; Balducci, A. Toward New Solvents for EDLCs: From Computational Screening to Electrochemical Validation. *J. Phys. Chem. C* **2015**, *119* (24), 13413−13424.

(22) Sendek, A. D.; Yang, Q.; Cubuk, E. D.; Duerloo, K.-A. N.; Cui, Y.; Reed, E. J. Holistic Computational Structure Screening of More than 12000 Candidates for Solid Lithium-Ion Conductor Materials. *Energ Environ. Sci.* **2017**, *10* (1), 306−320.

(23) Hatakeyama-Sato, K.; Tezuka, T.; Umeki, M.; Oyaizu, K. AI-Assisted Exploration of Superionic Glass-Type Li+ Conductors with Aromatic Structures. *J. Am. Chem. Soc.* **2020**, *142* (7), 3301−3305.

(24) Xie, T.; France-Lanord, A.; Wang, Y.; Lopez, J.; Stolberg, M. A.; Hill, M.; Leverick, G. M.; Gomez-Bombarelli, R.; Johnson, J. A.; Shao-Horn, Y.; Grossman, J. C. Accelerating Amorphous Polymer Electrolyte Screening by Learning to Reduce Errors in Molecular Dynamics Simulated Properties. *Nat. Commun.* **2022**, *13* (1), 3415.

(25) Bradford, G.; Lopez, J.; Ruza, J.; Stolberg, M. A.; Osterude, R.; Johnson, J. A.; Gomez-Bombarelli, R.; Shao-Horn, Y. Chemistry-Informed Machine Learning for Polymer Electrolyte Discovery. *Acs Central Sci.* **2023**, *9*, 206.

(26) Dhakal, P.; Shah, J. K. A Generalized Machine Learning Model for Predicting Ionic Conductivity of Ionic Liquids. *Mol. Syst. Des Eng.* **2022**, *7*, 1344.

(27) Qin, T.; Yang, H.; Wang, L.; Xue, W.; Yao, N.; Li, Q.; Chen, X.; Yang, X.; Yu, X.; Zhang, Q.; Li, H. Molecule Design for Non-Aqueous Wide-Temperature Electrolytes via the Intelligentized Screening Method. *Angew. Chem., Int. Ed.* **2024**, *63* (37), No. e202408902.

(28) Husch, T.; Korth, M. Charting the Known Chemical Space for Non-Aqueous Lithium−Air Battery Electrolyte Solvents. *Phys. Chem. Chem. Phys.* **2015**, *17* (35), 22596−22603.

(29) Winter, M.; Barnett, B.; Xu, K. Before Li Ion Batteries. *Chem. Rev.* **2018**, *118* (23), 11433−11456.

(30) Ein-Eli, Y.; Thomas, S. R.; Chadha, R.; Blakley, T. J.; Koch, V. R. Li-Ion Battery Electrolyte Formulated for Low-Temperature Applications. *J. Electrochem. Soc.* **1997**, *144* (3), 823−829.

(31) Rustomji, C. S.; Yang, Y.; Kim, T. K.; Mac, J.; Kim, Y. J.; Caldwell, E.; Chung, H.; Meng, Y. S. Liquefied Gas Electrolytes for Electrochemical Energy Storage Devices. *Science* **2017**, *356* (6345). .

(32) Vu, M. C.; Mirmira, P.; Gomes, R. J.; Ma, P.; Doyle, E. S.; Srinivasan, H. S.; Amanchukwu, C. V. Low Melting Alkali-Based Molten Salt Electrolytes for Solvent-Free Lithium-Metal Batteries. *Matter* **2023**, *6* (12), 4357−4375.

(33) Barrick, J. E.; Lenski, R. E. Genome Dynamics during Experimental Evolution. *Nat. Rev. Genet* **2013**, *14* (12), 827−839.

(34) Vaudel, M.; Verheggen, K.; Csordas, A.; Ræder, H.; Berven, F. S.; Martens, L.; Vizcaíno, J. A.; Barsnes, H. Exploring the Potential of Public Proteomics Data. *Proteomics* **2016**, *16* (2), 214−225.

(35) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *Apl Mater.* **2013**, *1* (1), No. 011002.

(36) Wong, F.; Zheng, E. J.; Valeri, J. A.; Donghia, N. M.; Anahtar, M. N.; Omori, S.; Li, A.; Cubillos-Ruiz, A.; Krishnan, A.; Jin, W.; Manson, A. L.; Friedrichs, J.; Helbig, R.; Hajian, B.; Fiejtek, D. K.; Wagner, F. F.; Soutter, H. H.; Earl, A. M.; Stokes, J. M.; Renner, L. D.; Collins, J. J. Discovery of a Structural Class of Antibiotics with Explainable Deep Learning. *Nature* **2024**, *626*, 177−185.

(37) Cheng, L.; Assary, R. S.; Qu, X.; Jain, A.; Ong, S. P.; Rajput, N. N.; Persson, K.; Curtiss, L. A. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *J. Phys. Chem. Lett.* **2015**, *6* (2), 283−291.

(38) Wang, D.; He, T.; Wang, A.; Guo, K.; Avdeev, M.; Ouyang, C.; Chen, L.; Shi, S. A Thermodynamic Cycle-Based Electrochemical Windows Database of 308 Electrolyte Solvents for Rechargeable Batteries. *Adv. Funct. Mater.* **2023**, *33* (11). .

(39) Yoo, D.; Yang, S.; Kim, K. J.; Choi, J. W. Fluorinated Aromatic Diluent for High-Performance Lithium Metal Batteries. *Angew. Chem., Int. Ed.* **2020**, *59* (35), 14869−14876.

(40) Beltran, S. P.; Cao, X.; Zhang, J.-G.; Balbuena, P. B. Localized High Concentration Electrolytes for High Voltage Lithium−Metal Batteries: Correlation between the Electrolyte Composition and Its Reductive/Oxidative Stability. *Chem. Mater.* **2020**, *32* (14), 5973−5984.

(41) Leverick, G.; Shao-Horn, Y. Controlling Electrolyte Properties and Redox Reactions Using Solvation and Implications in Battery Functions: A Mini-Review. *Adv. Energy Mater.* **2023**, *13*, No. 2204094.

(42) Xiao, J.; Li, Q.; Bi, Y.; Cai, M.; Dunn, B.; Glossmann, T.; Liu, J.; Osaka, T.; Sugiura, R.; Wu, B.; Yang, J.; Zhang, J.-G.; Whittingham, M. S. Understanding and Applying Coulombic Efficiency in Lithium Metal Batteries. *Nat. Energy* **2020**, *5* (8), 561−568.

(43) Shree Sowndarya, S. V.; Law, J.; Tripp, C. E.; Duplyakin, D.; Skordilis, E.; Biagioni, D.; Paton, R. S.; St John, P. Multi-Objective Goal-Directed Optimization of de Novo Stable Organic Radicals for Aqueous Redox Flow Batteries. *Nat. Mach. Intell.* **2022**, *4* (8), 720−730.

(44) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547−555.

(45) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9* (2), 513−530.

(46) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), No. 145301.

(47) Wang, A. Y.-T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brgoch, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chem. Mater.* **2020**, *32* (12), 4954−4965.

(48) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf Model* **2019**, *59* (8), 3370−3388.

(49) Pereira, F.; Xiao, K.; Latino, D. A. R. S.; Wu, C.; Zhang, Q.; Aires-de-Sousa, J. Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals. *J. Chem. Inf Model* **2017**, *57* (1), 11−21.

(50) Ko, S.; Obukata, T.; Shimada, T.; Takenaka, N.; Nakayama, M.; Yamada, A.; Yamada, Y. Electrode Potential Influences the Reversibility of Lithium-Metal Anodes. *Nat. Energy* **2022**, *7*, 1−8.

(51) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminformatics* **2015**, *7* (1), 20.

(52) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach Intell* **2020**, *2* (1), 56−67.

(53) Landesfeind, J.; Gasteiger, H. A. Temperature and Concentration Dependence of the Ionic Transport Properties of Lithium-Ion Battery Electrolytes. *J. Electrochem. Soc.* **2019**, *166* (14), A3079−A3097.

(54) Flores, E.; Wölke, C.; Yan, P.; Winter, M.; Vegge, T.; Cekic-Laskovic, I.; Bhowmik, A. Learning the Laws of Lithium-Ion Transport in Electrolytes Using Symbolic Regression. *Digital Discov* **2022**, *1* (4), 440−447.

(55) Zhang, Y.; Bier, I.; Viswanathan, V. Predicting Electrolyte Conductivity Directly from Molecular-Level Interactions. *ACS Energy Lett.* **2022**, *7*, 4061.

(56) Cao, X.; Jia, H.; Xu, W.; Zhang, J.-G. Review—Localized High-Concentration Electrolytes for Lithium Batteries. *J. Electrochem. Soc.* **2021**, *168* (1), No. 010522.

(57) Chen, Y.; Lee, E. M. Y.; Gil, P. S.; Ma, P.; Amanchukwu, C. V.; Pablo, J. J. de Molecular Engineering of Fluoroether Electrolytes for Lithium Metal Batteries. *Mol. Syst. Des. Eng.* **2022**, *8*, 2.

(58) Zhao, Y.; Zhou, T.; Mensi, M.; Choi, J. W.; Coskun, A. Electrolyte Engineering via Ether Solvent Fluorination for Developing Stable Non-Aqueous Lithium Metal Batteries. *Nat. Commun.* **2023**, *14* (1), 299.

(59) Zhang, G.; Chang, J.; Wang, L.; Li, J.; Wang, C.; Wang, R.; Shi, G.; Yu, K.; Huang, W.; Zheng, H.; Wu, T.; Deng, Y.; Lu, J. A Monofluoride Ether-Based Electrolyte Solution for Fast-Charging and Low-Temperature Non-Aqueous Lithium Metal Batteries. *Nat. Commun.* **2023**, *14* (1), 1081.

(60) Ma, P.; Kumar, R.; Vu, M. C.; Wang, K.-H.; Mirmira, P.; Amanchukwu, C. Fluorination Promotes Lithium Salt Dissolution in Borate Esters for Lithium Metal Batteries. *J. Mater. Chem. A* **2024**, *12*, 2479.

(61) Li, Y.; Liu, M.; Wang, K.; Li, C.; Lu, Y.; Choudhary, A.; Ottley, T.; Bedrov, D.; Xing, L.; Li, W. Single-Solvent-Based Electrolyte Enabling a High-Voltage Lithium-Metal Battery with Long Cycle Life. *Adv. Energy Mater.* **2023**, *13* (30). .

(62) Feng, S.; Chen, M.; Giordano, L.; Huang, M.; Zhang, W.; Amanchukwu, C. V.; Anandakathir, R.; Shao-Horn, Y.; Johnson, J. A. Mapping a Stable Solvent Structure Landscape for Aprotic Li−Air Battery Organic Electrolytes. *J. Mater. Chem. A* **2017**, *5* (45), 23987−23998.

(63) Atwi, R.; Chen, Y.; Han, K. S.; Mueller, K. T.; Murugesan, V.; Rajput, N. N. An Automated Framework for High-Throughput Predictions of NMR Chemical Shifts within Liquid Solutions. *Nat. Comput. Sci.* **2022**, *2* (2), 112−122.

(64) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat. Commun.* **2022**, *13* (1), 2453.

(65) *Buy purchasable compounds | eMolecules*. https://www.emolecules.com/ (accessed 2024-10-07).

(66) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem. in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2021**, *49* (D1), D1388.

(67) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52* (11), 2864−2875.

(68) Fromer, J. C.; Graff, D. E.; Coley, C. W. Pareto Optimization to Accelerate Multi-Objective Virtual Screening. *Digit. Discovery* **2024**, *3* (3), 467−481.

(69) Agarwal, G.; Doan, H. A.; Robertson, L. A.; Zhang, L.; Assary, R. S. Discovery of Energy Storage Molecular Materials Using Quantum Chemistry-Guided Multiobjective Bayesian Optimization. *Chem. Mater.* **2021**, *33*, 8133.

(70) Yu, Z.; Rudnicki, P. E.; Zhang, Z.; Huang, Z.; Celik, H.; Oyakhire, S. T.; Chen, Y.; Kong, X.; Kim, S. C.; Xiao, X.; Wang, H.; Zheng, Y.; Kamat, G. A.; Kim, M. S.; Bent, S. F.; Qin, J.; Cui, Y.; Bao, Z. Rational Solvent Molecule Tuning for High-Performance Lithium Metal Battery Electrolytes. *Nat. Energy* **2022**, *7* (1), 94−106.

(71) Xue, W.; Shi, Z.; Huang, M.; Feng, S.; Wang, C.; Wang, F.; Lopez, J.; Qiao, B.; Xu, G.; Zhang, W.; Dong, Y.; Gao, R.; Shao-Horn, Y.; Johnson, J. A.; Li, J. FSI-Inspired Solvent and "Full Fluorosulfonyl" Electrolyte for 4 V Class Lithium-Metal Batteries. *Energ Environ. Sci.* **2020**, *13* (1), 212−220.

(72) Hai, F.; Tian, X.; Yi, Y.; Wu, Z.; Zheng, S.; Guo, J.; Tang, W.; Hua, W.; Li, M. A Sulfolane-Based High-Voltage Electrolyte with Dispersed Aggregates for 5 V Batteries. *Energy Storage Mater.* **2023**, *54*, 641−650.

(73) Jiang, K. S.; Hobold, G. M.; Guo, R.; Kim, K.-H.; Melemed, A. M.; Wang, D.; Zuin, L.; Gallant, B. M. Probing the Functionality of LiFSI Structural Derivatives as Additives for Li Metal Anodes. *Acs Energy Lett.* **2022**, *7*, 3378−3385.

(74) *WebPlotDigitizer: Web based tool to extract data from plots, images, and maps.* https://automeris.io/ (accessed 2024-10-07).

(75) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, 3 (1), 33.

(76) *RDKit: Open-source cheminformatics.* https://rdkit.org/ (accessed 2024-10-07).

(77) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: 2017; Vol. *30*, pp 3146−3154.

(78) *Scikit-learn: Machine learning in python.* https://scikit-learn.org/stable/ (accessed 2024-10-07).

(79) *Gaussian 16, Revision C.01*; Gaussian, Inc.: Wallingford, CT,2016. https://gaussian.com/citation/ (accessed 2024-10-07).