

Electrolytomics: A unified big data approach for electrolyte design and discovery

Ritesh Kumar¹, Minh Canh Vu¹, Peiyuan Ma¹, and Chibueze V. Amanchukwu^{1,*}

¹Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA

*Corresponding author: Chibueze V. Amanchukwu (chibueze@uchicago.edu)

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, composition) is practically infinite; hence discovery is primarily guided through trial-and-error which slows the deployment of new battery chemistries. Inspired by artificial intelligence (AI)-enabled drug discovery, we usher in a new paradigm for electrolyte discovery. We assemble the largest small molecule experimental liquid electrolyte ionic conductivity dataset and build highly accurate machine learning (ML) and deep learning models to predict ionic conductivity across a wide range of electrolyte classes. The developed models outperform molecular dynamic (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 new metric called the electrolyte score (*eScore*) to unify the predicted disparate electrolyte properties. Deploying these models on large unlabeled datasets, we discover new electrolyte solvents, experimentally validate that the electrolyte is conductive ($> 1 \text{ mS cm}^{-1}$), stable up to 6V, supports efficient anode-free LMB, and even LIB cycling at extreme temperatures. Our work heralds a new age in electrolyte design and battery materials discovery.

Introduction

Next generation energy-dense batteries are required to accelerate decarbonization and the transition to grid electrification¹. While the 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 severe battery capacity fade. Since conventional Li-ion electrolytes are poor for LMBs³, the design of novel liquid electrolytes have led to significant increases in Coulombic efficiency (CE)⁴, but progress has plateaued⁴. Even current Li-ion batteries (LIBs) are limited by the lack of electrolytes that support high voltage or low cobalt cathodes, work at low ($<-20^\circ\text{C}$) or high temperatures ($>60^\circ\text{C}$)⁶. Therefore, novel approaches to electrolyte design and discovery are vital.

An excellent electrolyte is required to be ionically conductive, oxidatively stable, reductively stable, support high CE, and should have additional properties such as low volatility, low viscosity, nonflammability, etc⁵. 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 ($\sim 10^{60}$)⁶. Hence, for the past century, electrolyte design has primarily been driven by trial-and-error. Computational approaches such as classical equilibrium molecular dynamics (MD) are unable to predict ionic conductivity accurately, are computationally prohibitive to explore this large search space, and generated force fields are also not universally applicable.⁷ Statistical methods such as the licensed Advanced Electrolyte Model by Gering *et. al.* has primarily been used for carbonate-based liquid electrolytes and struggle with other electrolyte classes⁸. Hence, these

approaches cannot be deployed on unexplored electrolyte classes. Density functional theory (DFT) can be used to predict oxidative stability that correlates with experimental observations^{9,10}; however large scale DFT calculations to screen unlabeled molecules are also computationally intractable. For measuring CE, no computational methods exist. Inspired by advances in drug¹¹ and materials^{12–14} discovery and other scientific domains¹⁵, artificial intelligence (AI)-based approaches¹⁴ provide a facile pathway to accelerate electrolyte property prediction and new electrolyte discovery. Data science approaches have gained traction to predict properties relevant for mostly solid^{16–19} and liquid^{20,21} electrolytes with prior studies focused primarily on one or two figures of merit with small experimental datasets or little to no experimental verification. Therefore, long standing challenges remain such as limited dataset size, the lack of multi-property predictive models, limited model interpretability, limited experimental validation, and no discoveries of new electrolyte classes.

Here, we develop a holistic AI-guided approach incorporating a confluence of data science, experiments, and computation for the simultaneous prediction of electrolyte properties and the discovery of new electrolyte classes. Akin to the omics type strategy utilized in genomics²² and proteomics²³; we term this approach “electrolyomics”, the large-scale systematic investigation of electrolytes. First, we build the largest liquid electrolyte ionic conductivity database (EDB-1) encompassing the complexity of typical electrolytes with different solvent types and mixtures, salts, concentration, and temperature. Coupled with an experimental CE dataset and a DFT-obtained oxidative stability dataset (Materials Project²⁴), we predict the three desired electrolyte properties using a combination of forward design ML and deep learning (DL) techniques with high confidence. The obtained conductivity predictions outperform MD simulations. Despite no encoding of 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 (*eScore*) that provides a unified view of the electrolytes’ performance in terms of the three figures of merit. Finally, we discover a novel class of sulfonyl fluoride-based electrolytes yielding CE as high as 99.5% that competes favorably against all known liquid electrolytes, support low lithium metal overpotentials, and can support LMBs with the environmentally benign LiFePO₄ (LFP) cathodes for long term cycling. To broaden the scope, we also evaluate these electrolytes for next generation LIBs that demonstrate compatibility with high voltage NMC811 (LiNi_{0.8}Mn_{0.1}Co_{0.1}O₂) 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 datasets for critical properties of next generation battery electrolytes

The data-driven electrolyte discovery workflow from data collection to experimental observations is summarized in Figures 1(a) and S3. 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 Coulombic efficiency (CE) among the most important (Figure 1(b)). We collected an extensive database of ionic conductivity for lithium-containing aprotic liquid electrolytes (EDB-1) reported over the past fifty 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 dataset is heavily sampling biased with carbonates constituting around 72% of the dataset and has deviations within the reported values itself from different sources (Table S2).

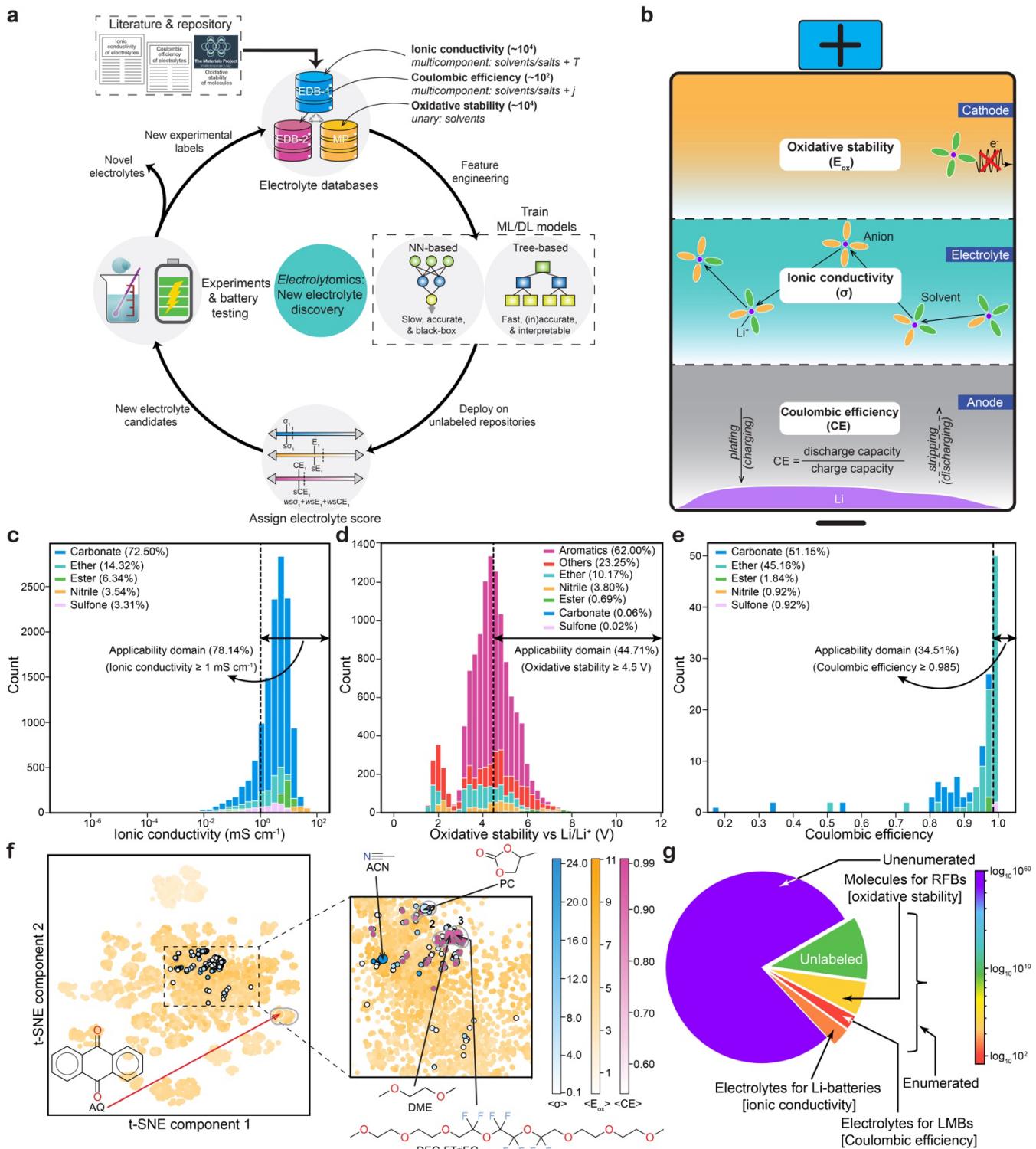


Figure 1. Electrolyomics workflow and visualizing chemical space of the explored and unexplored solvent molecules. **a** Schematic illustrating overall ML-based workflow for the discovery of new electrolytes. **b** Illustration depicting three key figures of merits of a typical electrolyte: ionic conductivity, oxidative stability, and Coulombic efficiency. MP = Materials Project, EDB-1 = ionic conductivity database, EDB-2 = Coulombic efficiency database. Histogram of **c** ionic conductivity, **d** oxidative stability, and **e** CE as a function of different solvent classes. **f** 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, orange for oxidative stability, and purple for CE.

Representative solvent molecules from the three datasets are also highlighted in larger circles. **g** Pie chart showing unexplored, and the approximate 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).

For oxidative stability, we utilized density functional theory (DFT)-obtained labels for solvents from the Materials Project (MP) database²⁴ since they correlate well with experiments^{9,10}, but should only be employed for screening purposes (Supplementary Note 1). 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) and obtained from the literature²⁵⁻²⁷. CE is the most important metric for an electrolyte as it encapsulates morphological, reductive stability, passivation effects, current density among others²⁸.

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 (Supplementary Note 3) as shown in Figure 1(f), where some of the typical solvent molecules present in the three databases are marked. A clear overlap between the EDB-1 and EDB-2 datasets 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 dataset lie in a different chemical space compared to the other two datasets (no overlap in t-SNE projection) because those molecules were primarily developed for redox flow batteries (RFBs) as active redox species and not ionically conductive electrolytes^{9,30}. Additionally, the three combined databases still constitute a small fraction compared to the chemical space of molecules that have been enumerated ($\sim 10^8$) or unenumerated ($\sim 10^{60}$) (Figure 1(g)).

Developing forward ML models for the critical electrolyte properties

The complete ML workflow is described in Supplementary Note 1. Separate ML regression models were developed to predict the three target properties as a function of different data splits (Supplementary Note 1). A common strategy for splitting all three datasets were employed through random and stratified splits. For examining data leakage³², two new splits were created for EDB-1 and EDB-2 datasets called cluster-I and cluster-II, as shown schematically in Figure 2(h), while oxidative stability prediction was also examined for out-of-domain performance through scaffold split. Comparisons of the shallow learning (SL) and deep learning (DL; Chemprop³³) models' performance on different splits of the EDB-1, MP, and EDB-2 datasets are summarized in Figures 2(a), 2(b), and 2(c), respectively. For all three target properties, ML models perform best on either random or stratified perform (that adhere to the independently and identically distributed (IID) principle) and worse on cluster-I or scaffold (that do not adhere to IID principle). For ionic conductivity property prediction, the Chemprop models perform only slightly better than the LightGBM models but are significantly better for oxidative stability, attributable to differences in dataset size (MP: 18133, EDB-1: 10196). For CE, the performance of Chemprop on some splits are slightly worse compared to the SL algorithm *viz* partial least squares regression (PLSR) using the graph embedding features obtained from pre-trained Chemprop model on a large LUMO dataset⁶⁰ because of low CE labels (142). The parity plots for the best performing models: Chemprop trained on EDB-1 (stratified), Chemprop trained on MP (random), and PLSR (stratified) trained on the EDB-2 datasets (stratified) are also shown in Figures 2(d), 2(e), and 2(f), respectively. Additionally, the test MAE 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 only electrolytes containing carbonates⁸ 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 is provided in Supplementary Notes 4, 5, and 6.

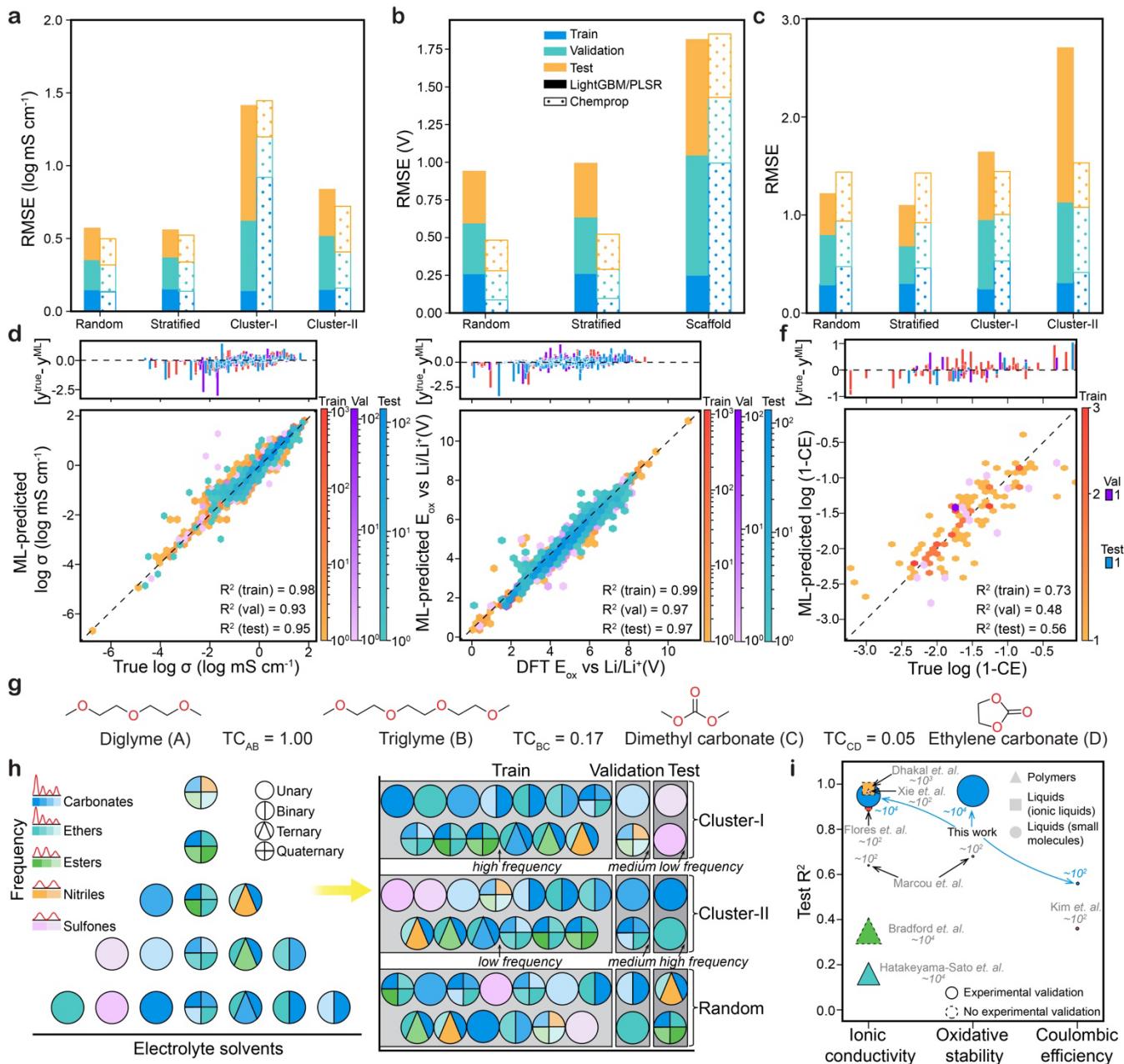


Figure 2. ML performance on the target properties using different data splits and comparison with literature. Bar plot comparing the performance of shallow learning (SL) and Chempop models on different data splits for **a** EDB-1 (ionic conductivity), **b** MP (oxidative stability), and **c** EDB-2 (CE) datasets. Hexagonal binned parity plots for best-performing models on random split datasets for predicting **d** ionic conductivity in common log scale ($\log \sigma$), **e** oxidative stability, and **f** Coulombic inefficiency in common log scale ($\log (1-CE)$) labels. **g** Tanimoto coefficient (TC) values shown for a few typical solvent molecules present in the EDB-1 dataset. **h** Schematic illustrating the three data splits utilized for inference on EDB-1 and EDB-2 datasets. Each circle represents a data point where the number of subdivisions within the circle denotes the number of solvents mixed in the electrolyte (details of schematic and other splits in Supplementary Note 4). **i** A comparison of ML-based research on Li battery electrolytes published in recent years, assessing the quality of ML performance and the approximate dataset size (indicated by italicized numbers) used for model development. RMSE = root mean squared error.

To rationalize the relative ML performance based on chemical similarity, Tanimoto coefficients (TC)³⁴ 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 2(g)). 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-I/ scaffold splits in general are very low compared to random/stratified splits (Figure S5), denoting that the cluster-I/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 the ML-based studies reported recently in Figure 2(i) and Table S1. Our work is the first to develop ML models for all three figures of merit and the models have accuracies that rival or outperform all prior works and allow for predictions across a wider chemical space¹⁴ (Table S1, Figure 2(i)).

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)^{14,38}. This approach was applied to traditional SL algorithms such as the LightGBM and PLSR since these models provide similar performance to Chemprop.

The SHAP summary plot of the LightGBM model for ionic conductivity (stratified) is shown in Figure 3(a). The top five most relevant features for the conductivity property according to the SHAP method are temperature (T), molecular weight of solvents ($MW_{solvent-1/2}$), molecular weight of salt (MW_{salt}), and salt concentration (c_{salt}). From Figure 3(a), 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 (Figures 3(b)) for temperature confirms the general trend found in Figure 3(a) that is also consistent with the widely known temperature-dependence of electrolyte ionic conductivities³⁹. Interestingly, it is to be noted that at low temperatures ($T < 0^{\circ}\text{C}$) and high salt concentrations, most data points have very low predicted conductivity values while at high temperatures and moderate to high salt concentrations lead to very high predicted values⁴⁰. The SHAP values for the salt concentration feature – and hence ionic conductivity – increase initially as a function of the salt concentration (and increased charge carrier concentration) and then starts decreasing (due to high viscosity and ion aggregates) (Figure 3(c)). This behavior can be attributed to the classical conductivity law relating ion mobility and conductivity³¹. The maximum SHAP values (or conductivity) are found near $c_{salt} \approx 1$ to 1.5 M, which is consistent with the findings for other conventional electrolytes^{8,31}. The LightGBM model extracted this knowledge even though it was not explicitly provided during the training process. Moreover, two distinct but non-trivial decreasing trends are observed at low and high temperatures, signifying different transport mechanism at low and high temperatures, respectively. No clear emerging trends are observed when ionic conductivity is plotted as a function of salt concentration and temperature for the raw data (Figure S6), demonstrating the need for the SHAP analysis to extract subtle and physically consistent trends and correlations. Corresponding 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 (consistent with Figure S7); a design strategy used for high salt concentration electrolytes^{26,41}. We also employed model sensitivity analysis (called Sobol analysis) to gauge the effect of sensitivity of models towards the different features utilized (Supplementary Note 8), which are in good agreement with the SHAP analysis for each target property (Figure S8).

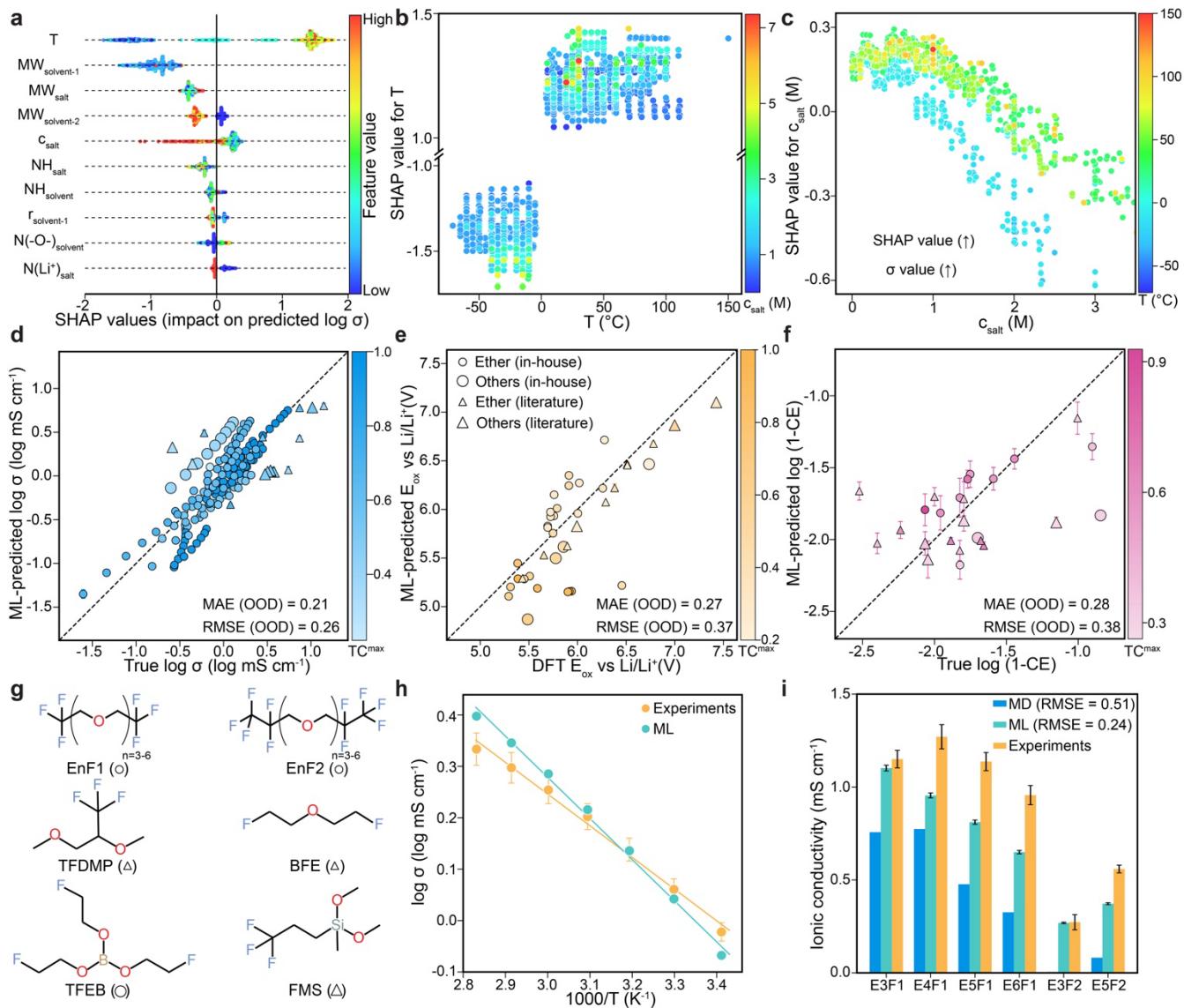


Figure 3. Interpretability and generalizability of the trained models. **a** SHAP summary plot for LightGBM model trained on ionic conductivity (random) dataset arranged in the decreasing order of importance towards ML prediction. The scatter points for each feature are colored according to the feature magnitudes. SHAP dependence plots for **b** temperature (T) and **c** salt concentration (c_{salt}) (within moderate concentration range) features for the ionic conductivity model. The scatter points in **b** and **c** are colored according to the c_{salt} and T values, respectively, to highlight the interaction between the two features. Performance of the trained best-models on in-house OOD datasets of **d** ionic conductivity, **e** oxidative stability, and **f** CE labels. The scatter points are colored according to the TC^{max} values with respect to the EDB-1, MP and EDB-2 datasets and the corresponding color bars (blue, orange, and purple) are shown alongside in **d**, **e** and **f**, respectively. **g** Representative molecular structures from OOD datasets, where symbols in the parenthesis correspond to solvent class and data source according to the legends shown in **e**. All OOD solvent molecules from literature are shown in Figure S9. **h** Plot showing Arrhenius fitting of the ML-predicted and experimental ionic conductivity data for 1 M lithium bis (fluorosulfonyl)amide (LiFSA)/E3F1 electrolyte. **i** Comparison of ML-predicted, MD-calculated, and experimental ionic conductivities for EnF1- and EnF2-based electrolytes at 30°C. The MD-derived values in **i** and the experimentally obtained values in **h**, **i** were extracted from Chen *et. al*⁴³ and Ma *et. al*³, respectively. TFDMP = 1,1,1-trifluoro-2,3-dimethoxypropane⁶⁴, BFE = bis(2-fluoroethyl) ether⁶⁵, TFE = tris(2-fluoroethyl) borate⁶⁶, FMS = dimethoxy(methyl)(3,3,3-trifluoropropyl)silane⁶⁷.

Examining generalizability on OOD data

To verify if the trained models are generalizable, i.e., make confident predictions on unseen electrolytes, they are deployed on out-of-distribution (OOD) datasets for each property, compiled in-house and from literature. Overall, the Chemprop models for both conductivity and oxidative stability performs reasonably well on all electrolyte solvent classes with lower error for solvents with high TC^{max} values with respect to the EDB-1 dataset (color bar in Figure 3(d)). 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 dataset (Figure 3(f)). Furthermore, the developed ML models successfully capture the relative trends in the ground truth data, affirming their reliability in identifying promising candidates within uncharted chemical territories.

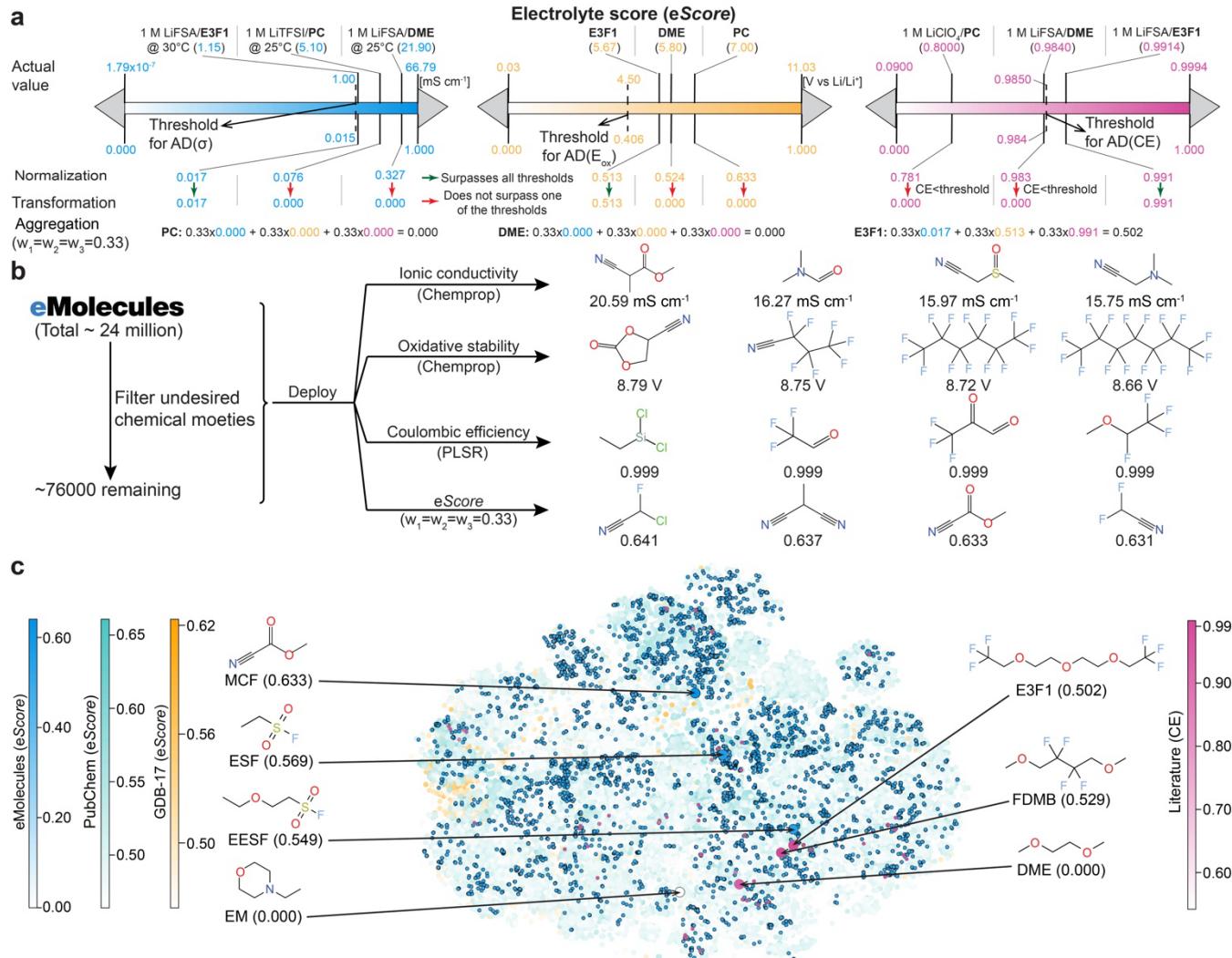


Figure 4. Assigning the electrolyte score and screening promising electrolytes. **a** The electrolyte score (eScore) concept demonstrated through the example of three common electrolytes: PC, DME, and E3F1. Equal weights of 0.33, 0.33, and 0.33 are considered for ionic conductivity, oxidative stability, and CE, respectively, in evaluating the aggregation step. The oxidative stability values for DME and PC have been taken from ref.⁴⁴ **b** Top solvent molecules from the eMolecules repository, as predicted by the three ML models that fall within the respective applicability domains. The fourth row shows the top solvent molecules according to the predicted eScores (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 (EDB-2 and reported in-house CE data points), and the candidate solvent molecules from the unlabeled repositories (eMolecules, PubChem, and GDB-17 datasets) having non-zero predicted eScores,

except for EM. Representative solvent molecules from the literature CE (both EDB-2 and OOD) datasets and solvent molecules from the eMolecules repository selected for experimental verification are highlighted in larger circles with corresponding *eScores* are provided within parentheses below. $[AD(\sigma)]$ = applicability domain for ionic conductivity, oxidative stability $[AD(E_{ox})]$, Coulombic efficiency $[AD(CE)]$, FDMB = fluorinated 1,4-dimethoxylbutane.

Interestingly, the Chemprop model also learnt phenomenological relations such as the Arrhenius equation as seen in Figure 3(h). The experimental and ML-predicted activation energies are also in good agreement for E3F1 (11.80 and 15.33 kJ mol⁻¹). Moreover, the ML-predicted ionic conductivities for the E3F1-related compounds were also compared with those evaluated using classical molecular dynamics (MD) calculations by Chen *et. al.*⁴³ as shown in Figure 3(i). Despite the MD-derived ionic conductivities being evaluated based on ionic solvation structures, information inaccessible to the ML models, ML-predicted ionic conductivities align more closely with the experimental values than the MD-derived ones. Therefore, leveraging ML approaches can mitigate computational costs associated with MD and address the challenge of generating force fields for unexplored electrolytes.

Screening promising electrolytes from molecular repositories

The developed ML models for the three target properties were finally deployed on large unlabeled molecular repositories such as eMolecules⁵², PubChem⁴⁵, and GDB-17⁴⁶ (Supplementary Note 10). The combined three databases make up a total of ~185 million, making it the largest search space for screening efficient electrolytes compared to previous studies (Table S1). For this work, we focus on the eMolecules database since it consists of commercially available compounds. If only one figure of merit is used, Figure 4(b) 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. Therefore, to find electrolyte solvent molecules that satisfy all the three applicability domain thresholds simultaneously and to rank them based on the predicted values, we devised a new unified metric called the electrolyte score (*eScore*) (Figure 4(a); Supplementary Note 10). The *eScore* is 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 volatility, flammability etc.; hence it is dataset- and property-agnostic. The representative structures with top predicted *eScores* according to four different weights for the eMolecules repository are shown Figures S11-S14. It is indeed observed that different molecules are assigned different *eScores* depending on the chosen weights. Furthermore, 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 knowledge such as fluorination leads to high oxidative stability and a heuristic that many fluorinated solvents support LMBs with high CE.

We subsequently selected four solvent molecules from the filtered, unlabeled eMolecules corpus, each with a high predicted *eScore*, 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 4(c) along with their predicted *eScores*. These molecules were chosen to cover different and novel classes of organic molecules and based on availability and cost (Supplementary Note 10). A DME (1,2-dimethoxyethane)-based and a commercial electrolyte (1 M LiPF₆ in ethylene carbonate (EC)/dimethyl carbonate (DMC)) were chosen as baselines for comparison. We find that the selected molecules span different regions of chemical space (Figure 4(c)), 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 the four selected compounds belong to new solvent classes, with two of them sulfonyl-based (ESF and EESF).

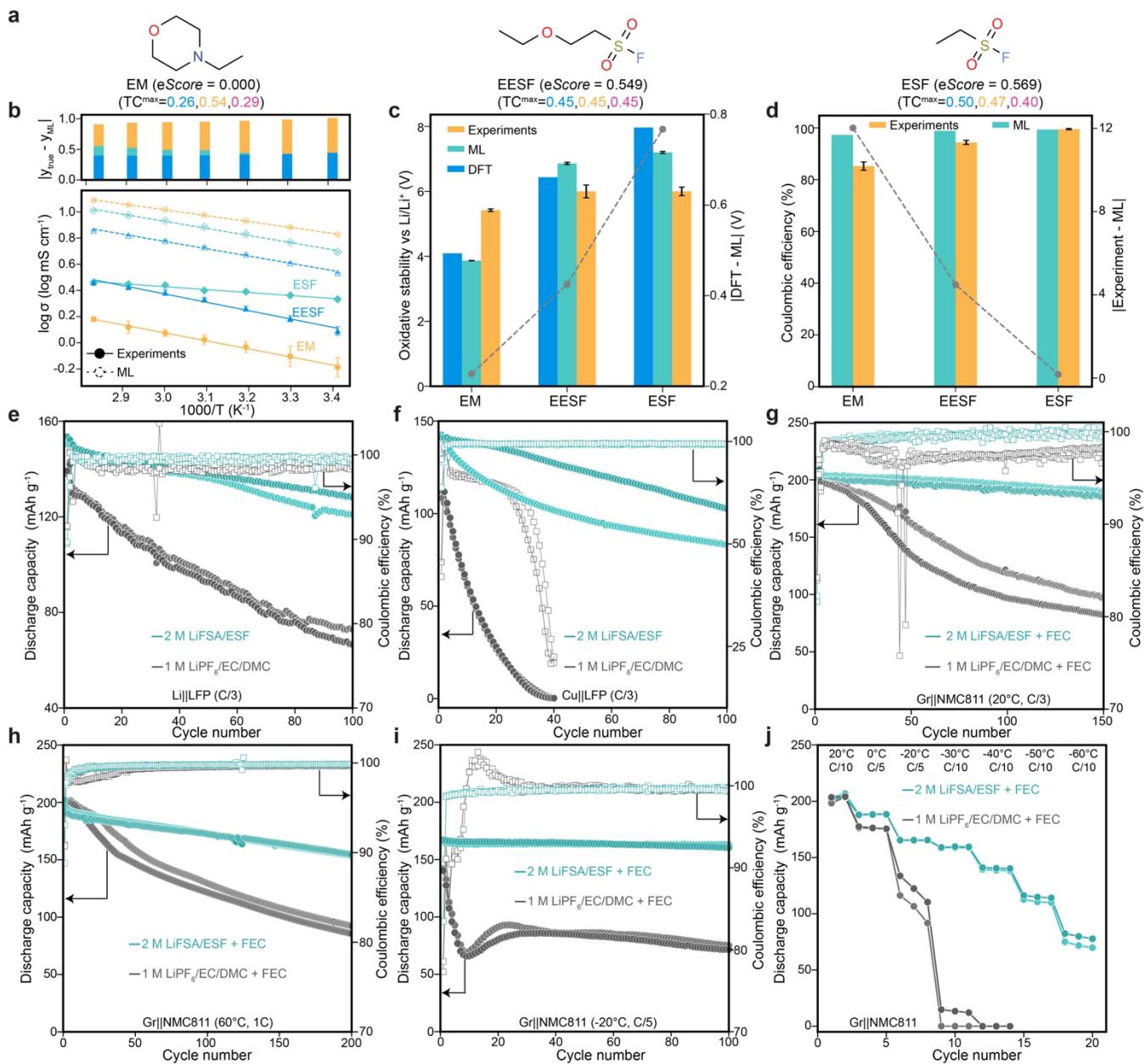


Figure 5. Experimental validation of the chosen solvent electrolyte candidates. a Molecular structure of EM, EESF and ESF along with the predicted eScores using equal weights ($w_1=w_2=w_3=0.33$). The TC^{max} values with respect to the EDB-1 (blue text), MP (orange text), and EDB-2 (purple text) datasets are also provided below the eScores. **b** Comparison of ML-predicted and experimental ionic conductivities in common log scale as a function of temperature for the three 1M LiFSA solvent electrolyte candidates. **c** Comparison of DFT-calculated and ML-predicted solvent oxidative stabilities (absolute error shown by gray scatter points) and experimental oxidative stabilities measured using LSV method with Li||Al cells. **d** Comparison of ML-predicted and experimental Li||Cu CE values for the three 2 M LiFSA solvent electrolyte candidates (absolute error shown by gray scatter points). Cycling performance of 2 M LiFSA/ESF and commercial electrolytes using **e** Li||LFP and **f** anode-free Cu||LFP coin cells. Cycling performance of FEC-containing ESF-based electrolyte in comparison with the commercial electrolyte using Gr||NMC811 cells at **g** 20°C, **h** 60°C, **i** -20°C, and **j** different temperatures. Cycling of duplicate cells (in dark and light shades) for each electrolyte have been shown in **e-j**. 1C = 1.66 mA/cm². 5wt% FEC (fluoroethylene carbonate).

New electrolyte discovery

Next, electrolytes formulated with the four selected candidate solvent molecules (with LiFSA salt) underwent experimental validation. The ML-predicted ionic conductivities for electrolytes other than EM (lowest TC^{max} value compared to other solvent candidates; Figure 5(a)) are in reasonable agreement with the experimental labels at ambient temperatures with some deviation at higher temperatures (Figure 5(b)). 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 dataset (Figure S4(a)). For oxidative stability, good agreements are found between DFT-calculated and ML-predicted values (Figure 5(c)) and they exhibit high experimental oxidative stabilities (Figure 5(c) and Figure S15). Using 2M 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 dataset. Although the models predicted MCF to have an eScore of 0.633, our domain knowledge and lack of nitrile representation in the EDB-2 dataset suggested that it would likely perform poorly under low salt concentration and without additives or co-solvents. Table S18 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. However, the relative trends in all the three experimentally obtained properties for EM, EESF, and ESF follow the order of their predicted eScores.

In a breakthrough for ML-enabled electrolyte discovery, the 2 M LiFSA/ESF electrolyte was found to have an exceptionally high CE of 99.5% (Figure S16) that performs on par with the state-of-the-art LMB liquid electrolytes such as FDMB², F5DEE⁴⁸, etc., in terms of all three desirable figures of merit (Table S20). For LMBs, the 2 M LiFSA/ESF electrolyte enables much improved cycling (Li||LFP and anode free Cu||LFP) outperforming carbonate electrolytes. The sulfonyl fluoride (R-SO₂-F) is chemically distinct from previously reported sulfamoyl fluorides ((R)₂N-SO₂-F)⁶² and sulfones (R-SO₂-R)⁶³. Contrary to Jiang et al.'s investigation of ESF as a 3% additive in conventional electrolytes⁵⁰, which showed no improvements due to substrate dependency, our research is dedicated to a much harder challenge of discovering novel electrolyte solvents. To broaden the scope of the work, we applied the proposed electrolyomics approach to address challenges with next generation LIBs. Incorporating additives such as 5 wt % FEC (fluoroethylene carbonate) within the ESF electrolyte shows highly reversible ambient (20°C), high (60°C), and low temperature (-20°C) (Figure 5(g-i)) cycling with high voltage NMC811 (LiNi_{0.8}Mn_{0.1}Co_{0.1}O₂) cathodes, outperforming conventional carbonate-based electrolytes (further discussion in Supplementary Note 11). Experimental characterizations in the form of scanning electron microscopy (SEM, Figure S19), nuclear magnetic resonance (NMR, Figure S20), and X-ray photoelectron spectroscopy (XPS, Figures S21-S22) provide insights into a smoother lithium deposition morphology, no solvent degradation, and a lithium or graphite interface enriched with inorganic decomposition products. Undoubtedly, our findings break new ground in electrolyte design. The impact of this work lies not solely in the discovery of ESF but 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 new liquid electrolytes compatible with NGBs. To our knowledge, this is the first application of ML that considers all crucial electrolyte properties – ionic conductivity, oxidative stability, and CE – using the largest assembled datasets to date. We define a quantitative metric (eScore) for electrolytes and showcase the discovery of a new 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 inverse design of novel electrolytes using generative algorithms.

Methods

Data collection and curation of EDB datasets

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 datasets. From the literature, data points were manually extracted from the text, tables, and/or figures present in either the main manuscript 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/cm² 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, 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, 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 WebPlotDigitizer⁵³. 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 Babel⁵⁶ utility to obtain the corresponding SMILES. Chemical drawing software such as Marvin were also used in determining SMILES or scientific names. The datasets then went through a rigorous editing process where the datasets 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 datasets 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 2021.09.5)⁵⁷ 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 using various libraries available in the Python (version 3.9.0) programming language. The three types of popular molecular representations utilized in the manuscript – i) quantitative structure-activity relationship (QSAR) descriptors were generated using RDKit and fed to a shallow learning (SL) algorithm, e.g., LightGBM⁵¹ ii) directed molecular graphs utilized in an end-to-end manner within the python library Chemprop³³ (and supplemented with additional RDKit features), which is an end-to-end deep learning framework used previously for antibiotic discovery¹¹, as well as the graph embeddings obtained from trained Chemprop model. For in-depth discussion on different settings employed in training and analyzing ML models, refer to Supplementary Notes 4, 5, 6, and 8. The RDKit library was also utilized for evaluating Tanimoto coefficients. Scikit-learn (version 0.23)⁵⁷ library was employed for building the SL (LightGBM, RF, and PLSR) models. The hyperparameter optimization for the SL algorithms were carried out using Optuna (version 2.3.0) library. SHAP (version 0.36.0) library was utilized for interpreting SL models. Sobol analysis for conducting model sensitivity analysis was carried out using SALib library. The reduced embeddings of chemical space for the t-SNE plots were generated using OpenTSNE library. All plots in the manuscript were made using matplotlib, seaborn, or plotly libraries.

DFT calculations

The adiabatic ionization energy calculations were carried out using DFT as implemented in Gaussian 16 (version 16.0.1) program⁵⁸. B3LYP and 6-31*g(d) was 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)⁴⁴. 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, O₂ and H₂O 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 ethanesulfonyl 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 1M LiFSA liquid electrolytes. A commercial electrolyte (1M LiPF₆ in EC-DMC, v/v, 1/1) was used as a reference electrolyte. The LFP (LiFePO₄) electrode has a total mass loading of 12.1 mg cm⁻² with 90 wt % Johnson Matthey LFP, 5 wt % Timcal C-45, and 5 wt % Solvay 5130 PVDF binder received from Cell Analysis, Modeling, and Prototyping (CAMP) facility at Argonne National Laboratory. The LFP electrode was 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 disks of 18 mm, 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 (O₂, H₂O < 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°C 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 realistic ionic conductivity of electrolytes using 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 using 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 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 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 copper electrode for 10 h and stripping to 1 V. Next, another 10 h of Li deposition on copper electrode was performed. Next, the coin cells were cycled for 10 cycles of two-hour-depositing/two-hour-stripping. Finally, Li was stripped from 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 μ L of electrolyte||LFP||SS, respectively. Coin cells were rested for 10 hours, 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/cm² which is equivalent to 1 C = 1.81 mA/cm². Li||Gr coin cells were prepared with the following configurations: SS||Li||25 μ L of electrolyte||1 separator||25 μ L of electrolyte||Gr||SS. Gr||NMC811 coin cells were prepared with the following configurations: SS||SS||Gr||25 μ L of electrolyte||1 separator||25 μ L of electrolyte||NMC811||Al and aluminum coated cathode case was used. Celgard 2325 was used for ESF electrolytes while Celgard 3501 was used for carbonate electrolyte. Two formation cycles at C/10 were performed at 20°C before cycling at desired current rate and temperature.

Symmetric Li||Li cycling

Li||Li coin cells were prepared with the following configurations: SS||Li||25 μ L of electrolyte||1 separator (Celgard 2325)||25 μ L 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 hours and had 5 formation cycles at 0.1 mA/cm² to 0.1 mAh/cm², and continued to cycle at 1 mA/cm² to 1 mAh/cm².

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 1M LiPF₆-EC-DMC (v/v, 1/1) and 2M LIFSA-ESF electrolytes. First, five pre-cycles with a current density of 0.02 mA cm⁻² within the voltage range of 0 V to 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 K α radiation beam ($h\nu = 1486.6$ eV) with 100 μ m 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.

Data Availability

The data will be made accessible through a public-facing website at amanchukwu.uchicago.edu.

Code Availability

The codes will be made publicly available on GitHub (<https://github.com/AmanchukwuLab/electrolytomics>) upon publication.

Acknowledgements

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 Futures program. The authors thank the Research Computing Center (RCC) at the University of Chicago and Argonne Leadership Computing Facility (ALCF) at Argonne National Laboratory for providing the required computational facilities. The authors thank Edward F. Barry at Argonne National Laboratory for providing the Li (20 μ m) 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 help with preliminary ML setup. The authors also acknowledge Emily Doyle and Michael Han for providing critical comments on the manuscript.

Author Contributions

C.V.A. conceived and supervised the project. R.K. curated the datasets, 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.

Competing Interests

The authors declare no competing interests.

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* **4**, 3243–3262 (2011).
2. Yu, Z. *et al.* Molecular design for electrolyte solvents enabling energy-dense and long-cycling lithium metal batteries. *Nat Energy* **5**, 526–533 (2020).
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 Cent. Sci.* **7**, 1232–1244 (2021).
4. Hobold, G. M. *et al.* Moving beyond 99.9% Coulombic efficiency for lithium anodes in liquid electrolytes. *Nat Energy* **6**, 951–960 (2021).
5. Amanchukwu, C. V. The electrolyte frontier: A manifesto. *Joule* **4**, 281–285 (2020).
6. Kirkpatrick, P. & Ellis, C. Chemical space. *Nature* **432**, 823–823 (2004).
7. Yao, N., Chen, X., Fu, Z.-H. & Zhang, Q. Applying Classical, Ab Initio, and Machine-Learning Molecular Dynamics Simulations to the Liquid Electrolyte for Rechargeable Batteries. *Chem. Rev.* **122**, 10970–11021 (2021).
8. Gering, K. L. Prediction of electrolyte conductivity: results from a generalized molecular model based on ion solvation and a chemical physics framework. *Electrochim. Acta* **225**, 175–189 (2017).
9. Cheng, L. *et al.* Accelerating electrolyte discovery for energy storage with high-throughput screening. *J. Phys. Chem. Lett.* **6**, 283–291 (2015).
10. Qu, X. *et al.* The Electrolyte Genome project: A big data approach in battery materials discovery. *Comp. Mater. Sci.* **103**, 56–67 (2015).
11. Stokes, J. M. *et al.* A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702 (2020).
12. Rajan, A. C. *et al.* Machine-learning-assisted accurate band gap predictions of functionalized MXene. *Chem. Mater.* **30**, 4031–4038 (2018).
13. 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.* **31**, 5145–5151 (2019).
14. Kumar, R. & Singh, A. K. Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts. *Npj Comput. Mater.* **7**, 197 (2021).
15. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
16. Sendek, A. D. *et al.* Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.* **10**, 306–320 (2016).
17. 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.* **142**, 3301–3305 (2020).

18. Xie, T. *et al.* Accelerating amorphous polymer electrolyte screening by learning to reduce errors in molecular dynamics simulated properties. *Nat. Commun.* **13**, 3415 (2022).
19. Bradford, G. *et al.* Chemistry-informed machine learning for polymer electrolyte discovery. *ACS Cent. Sci.* **9**, 206–216 (2023).
20. Dhakal, P. & Shah, J. K. A generalized machine learning model for predicting ionic conductivity of ionic liquids. *Mol. Syst. Des. Eng.* **7**, 1344–1353 (2022).
21. Marcou, G. *et al.* In silico Design, virtual screening and synthesis of novel electrolytic solvents. *Mol. Inform.* **38**, 1900014 (2019).
22. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).
23. Vaudel, M. *et al.* Exploring the potential of public proteomics data. *Proteomics* **16**, 214–225 (2016).
24. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
25. Yoo, D., Yang, S., Kim, K. J. & Choi, J. W. Fluorinated aromatic diluent for high-performance lithium metal batteries. *Angew. Chem. Int. Ed.* **59**, 14869–14876 (2020).
26. 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.* **32**, 5973–5984 (2020).
27. Leverick, G. & Shao-Horn, Y. Controlling electrolyte properties and redox reactions using solvation and implications in battery functions: A Mini-Review. *Adv. Energy Mater.* **2204094** (2023).
28. Xiao, J. *et al.* Understanding and applying coulombic efficiency in lithium metal batteries. *Nat. Energy* **5**, 561–568 (2020).
29. Selim, R. & Bro, P. Some observations on rechargeable lithium electrodes in a propylene carbonate electrolyte. *J Electrochem. Soc.* **121**, 1457–1459 (1974).
30. V., S. S. S. *et al.* Multi-objective goal-directed optimization of de novo stable organic radicals for aqueous redox flow batteries. *Nat. Mach. Intell.* **4**, 720–730 (2022).
31. Zhang, Y., Bier, I. & Viswanathan, V. Predicting electrolyte conductivity directly from molecular-level interactions. *ACS Ener. Lett.* **7**, 4061–4070 (2022).
32. Wang, A. Y.-T. *et al.* Machine learning for materials scientists: An introductory guide toward best practices. *Chem. Mater.* **32**, 4954–4965 (2020).
33. Yang, K. *et al.* Analyzing learned molecular representations for property prediction. *J. Chem. Inf. Model.* **59**, 3370–3388 (2019).
34. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminformatics* **7**, 20 (2015).
35. Strieth-Kalthoff, F. *et al.* Machine learning for chemical reactivity: The importance of failed experiments. *Angew. Chem. Int. Ed.* (2022).
36. Kerns, G. Jay. & Székely, G. J. Definetti’s theorem for abstract finite exchangeable sequences. *J. Theor. Probab.* **19**, 589–608 (2006).
37. Kim, S. C. *et al.* Data-driven electrolyte design for lithium metal anodes. *Proc. Natl. Acad. Sci. USA* **120**, e2214357120 (2023).

38. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
39. Landesfeind, J. & Gasteiger, H. A. Temperature and concentration dependence of the ionic transport properties of lithium-ion battery electrolytes. *J. Electrochem. Soc.* **166**, A3079–A3097 (2019).
40. Flores, E. *et al.* Learning the laws of lithium-ion transport in electrolytes using symbolic regression. *Digital Discovery* **1**, 440–447 (2022).
41. Cao, X., Jia, H., Xu, W. & Zhang, J.-G. Review—Localized high-concentration electrolytes for lithium batteries. *J. Electrochem. Soc.* **168**, 010522 (2021).
42. Ma, P. *et al.* Co-intercalation-free ether electrolytes for graphitic anodes in lithium-ion batteries. *Energy Environ. Sci.* **15**, 4823–4835 (2022).
43. Chen, Y. *et al.* Molecular engineering of fluoroether electrolytes for lithium metal batteries. *Mol Syst Des Eng* **8**, 195–206 (2022).
44. Feng, S. *et al.* Mapping a stable solvent structure landscape for aprotic Li–air battery organic electrolytes. *J. Mater. Chem. A* **5**, 23987–23998 (2017).
45. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395 (2020).
46. Ruddigkeit, L., Deursen, R. van, Blum, L. C. & Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
47. Adams, B. D., Zheng, J., Ren, X., Xu, W. & Zhang, J. Accurate determination of Coulombic efficiency for lithium metal anodes and lithium metal batteries. *Adv. Energy Mater.* **8**, 1702097 (2018).
48. Yu, Z. *et al.* Rational solvent molecule tuning for high-performance lithium metal battery electrolytes. *Nat. Energy* **7**, 94–106 (2022).
49. Xue, W. *et al.* FSI-inspired solvent and “full fluorosulfonyl” electrolyte for 4 V class lithium-metal batteries. *Energy Environ. Sci.* **13**, 212–220 (2020).
50. Jiang, K. S. *et al.* Probing the Functionality of LiFSI Structural Derivatives as Additives for Li Metal Anodes. *ACS Energy Lett.* **10**, 3378–3385 (2022).
51. Ke, G. *et al.* LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* **30**, 3146–3154 (2017).
52. eMolecules; eMolecules: La Jolla, CA, 2017. <https://www.emolecules.com>.
53. Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa: Large-scale self-supervised pretraining for molecular property prediction. *Arxiv* (2020) doi:10.48550/arxiv.2010.09885.
54. WebPlotDigitizer: Web based tool to extract data from plots, images, and maps. <https://automeris.io/WebPlotDigitizer>.
55. O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminf* **3**, 33 (2011).
56. RDKit: Open-source cheminformatics. <https://rdkit.org>.
57. scikit-learn: Machine learning in python. <https://scikit-learn.org/stable>.
58. Frisch, M. J. *et al.* Gaussian 16, Revision C.01. Gaussian, Inc., Wallingford CT (2016).

59. D. Wang, T. He, A. Wang, K. Guo, M. Avdeev, C. Ouyang, L. Chen, S. Shi. A Thermodynamic Cycle-Based Electrochemical Windows Database of 308 Electrolyte Solvents for Rechargeable Batteries. *Adv. Funct. Mater.* **33**, 2212342 (2023).
60. F. Pereira, K. Xiao, D.A.R.S. Latino, C. Wu, Q. Zhang, J. A.d.-Sousa. Machine learning methods to predict density functional theory B3LYP energies of HOMO and LUMO orbitals. *J. Chem. Inf. Model.* **57**, 11-21 (2017).
61. Y. S. Meng, V. Srinivasan, K. Xu. Designing better electrolytes. *Science* **378**, eabq3750 (2022).
62. Xue, W. *et al.* FSI-inspired solvent and “full fluorosulfonyl” electrolyte for 4 V class lithium-metal batteries. *Energ. Environ. Sc.i* **13**, 212–220 (2019).
63. F. Hai, X. Tian, Y. Yi, Z. Wu, S. Zheng, J. Guo, W. Tang, W. Hua, and M. Li. A sulfolane-based high-voltage electrolyte with dispersed aggregates for 5 V batteries. *Energy Stor. Mater.* **54**, 641-650 (2023).
64. Y. Zhao, T. Zhou, M. Mensi, J. W. Choi, A. Coskun. Electrolyte engineering via ether solvent fluorination for developing stable non-aqueous lithium metal batteries. *Nat. Commun.* **14**, 299 (2023).
65. G. Zhang, J. Chang, L. Wang, J. Li, C. Wang, R. Wang, G. Shi, K. Yu, W. Huang, H. Zheng, T. Wu, Y. Deng, J. Lu. A monofluoride ether-based electrolyte solution for fast-charging and low-temperature non-aqueous lithium metal batteries. *Nat. Commun.* **14**, 1081 (2023).
66. P. Ma, R. Kumar, M.C. Vu, K.-H. Wang, P. Mirmira, C.V. Amanchukwu. Fluorination promotes lithium salt dissolution in borate esters for lithium metal batteries. *J. Mater. Chem. A* **12**, 2479-2490 (2024).
67. Y. Li, M. Liu, K. Wang, C. Li, Y. Lu, A. Choudhary, T. Ottley, D. Bedrov, L. Xing, W. Li. Single-Solvent-Based Electrolyte Enabling a High-Voltage Lithium-Metal Battery with Long Cycle Life. *Adv. Ener. Mater.* **13**, 2300918 (2023).