MSDE

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5me00188a

Using electrolyte solvent embeddings to guide battery electrolyte discovery

Ritesh Kumar, † Ke-Hsin Wang† and Chibueze V. Amanchukwu *

Electrolyte development for next-generation batteries has largely relied on knowledge-driven trial-and-error. A significant reason for this approach is due to the vast, high-dimensional molecular design space for solvents and molecular mixtures. Popular dimensionality reduction techniques such as principal component analysis (PCA), t-distributed stochastic neighbor embeddings (t-SNE), and uniform manifold approximation and projection (UMAP) can facilitate chemical exploration by projecting this space into low dimensions, yet their applicability in electrolyte discovery remains underexplored. Here, we present a framework to assess the faithfulness of various dimensionality reduction methods in preserving chemically meaningful relationships. Then, we apply this framework to guide electrolyte solvent discovery by searching for electrolyte solvents with high and low cosine similarities to state-of-the-art systems, followed by formulation refinement using domain knowledge. We then performed experimental validation which revealed different failure modes for different electrolyte selections. We specifically focus on next generation lithium metal batteries (LMBs), and identified a non-fluorinated ether electrolyte solvent that exhibited high coulombic efficiency (CE). Our findings demonstrate that rational navigation of the electrolyte chemical space, enabled by dimensionality reduction and guided by data, can complement traditional approaches and accelerate the discovery of advanced battery materials.

Received 12th October 2025,
Accepted 2nd February 2026

DOI: 10.1039/d5me00188a

rsc.li/molecular-engineering

Design, System, Application

The chemical space is practically infinite, high-dimensional, discrete and discontinuous making manual exploration for molecular design and discovery slow and inefficient. This study establishes a quantitative framework to assess the faithfulness of widely used dimensionality reduction methods in preserving molecular neighborhood relationships between the original and reduced representations. Our framework introduces metrics that evaluate neighborhood preservation at three hierarchical levels. We further justify our approach by comparing chemical similarity and distance metrics in the reduced embedding space for some representative molecules. We leveraged the most faithful reduced embedding (t-stochastic distributed neighbor embeddings (t-SNE)) to guide new electrolyte solvent discovery for next-generation battery applications from a large virtual search space. This approach generalizes beyond electrolytes and can be applied to any molecular dataset to identify promising chemical motifs or explore specific subregions of chemical space with confidence. Therefore, the immediate impact will be widespread adoption of our approach by the cheminformatics community to find faithful reduced representation of the desired region of their chemical space. The only practical constraint lies in the use of general-purpose molecular descriptors that capture chemical similarity but not always functional similarity, meaning molecules with different functionalities may still appear closer in the reduced embedding.

Introduction

Innovation in rechargeable batteries is critical to accelerate grid electrification, increase electric vehicle adoption, and power electronic devices. Unfortunately, while anode and cathode materials are often known, progress in next-generation battery (NGB) development has stagnated due to the lack of electrolytes compatible with both electrodes. Electrolytes in most lithium battery chemistries face common challenges, such as low coulombic efficiency (CE), rapid capacity fading, and anode (*e.g.*, lithium metal) and cathode (*e.g.*, higher voltages) incompatibility.[1,2] Optimizing one property often degrades others, highlighting the inherent trade-offs among electrolyte properties.

Electrolyte design has relied primarily on trial-and-error methods or chemical intuition, constituting the knowledge-driven paradigm. Unfortunately, this approach has explored only about 10^3 solvents,[3,4] focusing on improvements from a limited pool of promising electrolytes.[5] Hence, molecular design efforts by computational and experimental experts typically focus on specific solvent classes[5,6] or electrolyte components like diluents and additives.[7] For instance, Ma

Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA. E-mail: chibueze@uchicago.edu
† Equal contribution.

et al.[8] found that increasing the weight fraction of fluorine in symmetric terminally fluorinated ethers improves CE. Similarly, Yu *et al.*[9] demonstrated that locally polar and asymmetric fluorinated groups at terminal carbon atoms in ethers enhance ion conduction and cyclability. Despite these advances, state-of-the-art (SOTA) electrolyte solvents like bis(2-(2,2,2-trifluoroethoxy)ethyl)ether (E3F1),[8] fluorinated 1,4-dimethoxylbutane (FDMB),[10] and 2-[2-(2,2-difluoroethoxy) ethoxy]-1,1,1-trifluoroethane (F5DEE)[9] still face challenges, such as limited operating temperatures and environmental issues linked to fluorination (*i.e.*, per- and poly-fluoroalkyl substances or PFAS). The vast, discrete chemical space, coupled with continuously variable compositions, creates a heterogeneous, high-dimensional input space for electrolyte optimization, even without considering electrode effects. Current design approaches address only selected failure modes, sometimes causing undesirable side effects—for example, increasing fluorination is one of the popular strategies to improve CE but reduces salt solubility and/or increases viscosity.[11] Similarly, tailored electrolyte additives and engineered solid electrolyte interphase (SEI) layers that promote the formation of mechanically robust, LiF-rich interphases have been shown to improve uniform lithium deposition and suppress dendrite growth, enhancing cycle life and coulombic efficiency in Li metal batteries (LMBs).[12,13] However, design principles for one solvent class, such as ethers, may not generalize to others like carbonates (domain shift). Could we then develop a new approach that allows us to efficiently and quickly explore the combinatorial space of electrolytes and overcome our existing selection biases?

Recent efforts have focused on developing computational and experimental descriptors like electrostatic potential (ESP) maps,[7] reductive potentials,[6] lowest unoccupied molecular orbitals (LUMO),[5,14] radial distribution functions (RDFs),[14] and solvatochromic[15] descriptors. However, the universality of these descriptors across diverse electrolytes remains unproven. The electrolyte solvent design space alone is estimated to be around 10^{60},[16] with even greater complexity when accounting for other components such as salts and mixtures. Navigating this vast chemical space solely through computational approaches like molecular dynamics (MD) or density functional theory (DFT) is computationally prohibitive. It is thus essential to develop rational, data-driven methods to explore electrolyte spaces irrespective of solvent class. Leveraging existing labeled data, even if limited (10^2 to 10^4)[3,17,18] compared to other scientific domains such as proteomics (~200k protein structure data[19]), can uncover patterns for identifying better electrolytes. Given the high time and labor costs of obtaining new labels—such as long-term battery cycling tests—data-driven techniques are crucial. There are three potential approaches through which data-driven techniques can enhance the discovery of electrolytes, or materials, in general: forward, adaptive, and inverse design. Forward design focuses on predicting properties from structures without accounting for underlying model uncertainties.[3,20] In contrast, adaptive design incorporates these uncertainties, thereby refining predictions based on iterative feedback.[21] Lastly, inverse design takes a reverse approach, starting from desired properties to determine the corresponding structures. Other alternative promising approaches have also emerged, *e.g.*, leveraging large language models to build relevant labeled database,[22] combining machine learning (ML) with computational descriptors to narrow down the search space,[23] and developing unified data-driven frameworks for diverse electrolyte requirements such as retrosynthetic pathway and SEI formation mechanism.[24] Despite these advanced methodologies, the electrolyte representation within the models remains high-dimensional, typically ranging from 10^2 to 10^5 parameters.[3] This complexity poses challenges for researchers attempting to navigate the high-dimensional search space to select a few promising candidates for real-world testing. For instance, while numerous similar compounds may be predicted to exhibit favorable electrolyte properties, researchers often aim to choose representatives from diverse classes to maximize the variability within the selection pool. Consequently, dimensionality reduction techniques become essential, enabling the visualization of extensive chemical spaces in more manageable, lower dimensions. This simplification facilitates informed human decision-making based on specific experimental requirements.

Here, we develop an unsupervised data-efficient framework for identifying promising electrolyte candidates from vast chemical spaces and outline rational design principles by using dimensionality reduction techniques. We first establish and provide a recipe to gauge the faithfulness of popular dimensionality reduction techniques to obtain reliable electrolyte solvent embeddings. Here, embedding refers to the projection of high-dimensional molecular descriptors into a two-dimensional space using explicit dimension-reduction techniques. This reduced representation facilitates quantitative comparisons among solvents based on physicochemical similarity and should not be confused with neural network-learned embeddings. We then selected 15 promising solvents, covering a diverse class of functional groups, for real-world LMB testing using lithium–copper (Li||Cu) half-cell cycling. This led to the identification of multiple electrolyte solvents with greater than 95% CE and a non-fluorinated ether (1-ethoxy-2-methylpropane; EMP) with nearly 99% CE. Combined with the domain knowledge of increasing performance through formulation design, we improved its half-cell and full-cell performances further that rivals known state-of-the-art electrolytes for "anode-free" or "zero-excess" LMB cycling. Our work expands the electrolyte design space by demonstrating that non-fluorinated solvents can also deliver comparable or superior performance, consistent with recent reports highlighting high-performance non-fluorinated electrolyte solvents.[21,25] Through our work, we aspire to complement the slow and tedious conventional knowledge-driven electrolyte discovery paradigm with the alternative

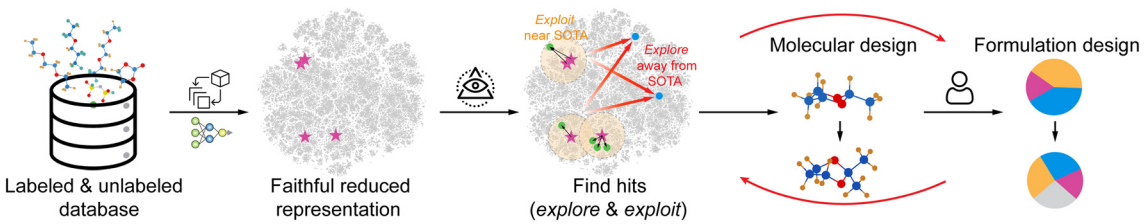
data-driven paradigm that can help navigate the vast, combinatorial uncharted electrolyte territory within a reasonable time frame. The data-driven component of this study lies in the systematic integration of data, descriptors, and dimensionality reduction to guide exploration of solvent space, rather than in black-box prediction of electrochemical performance. Our proposed approach will further accelerate the discovery of efficient electrolytes and deployment of a wide range of next generation battery chemistries.

Results

Faithful representation of electrolyte solvent chemical space

To utilize data-driven approaches for next-generation battery electrolyte discovery, it is imperative to increase the pool of existing electrolyte solvents, since the number of known solvents explored for lithium metal batteries remains limited. For instance, datasets used to train machine learning (ML) models for Li||Cu CE prediction in studies by Kumar *et al.*[3] and Kim *et al.*[17] included only 53 and 63 unique electrolyte solvents, respectively. A straightforward method to address this limitation involves enumerating the structures of all potential electrolyte solvents and utilizing trained ML models to identify promising candidates for real-world testing. To highlight the challenges of this approach, we manually enumerated all possible acyclic ethers, with the methodology illustrated in Fig. S1 and detailed in Note S1 in the SI. This process resulted in 2500 compounds constrained by a maximum of six carbon and oxygen atoms per molecule. Notably, only 23 of these compounds has been explored in the literature (Fig. S2), underscoring selection biases in current research. However, due to issues such as synthesizability/commercial availability, or salt solubility, not all enumerated compounds can be tested electrochemically in the real world. Therefore, there is a pressing need to integrate rational heuristics in the data-driven methodology to intelligently and efficiently guide the exploration of promising electrolyte solvents, as depicted in Scheme 1.

The first step in our unsupervised data-driven approach (without utilizing any end-to-end supervised ML models) for electrolyte discovery involved creating the virtual search space, consisting of approximately 60 000 unlabeled potential electrolyte solvent molecules, filtered from a collection of around 135 million drug-like molecules from eMolecules and PubChem repositories (refer to Note S2 for detailed discussion). The two repositories were chosen since many of the compounds listed are either commercially available or explored for different applications before. For molecular depiction of this search space, we used 2048-dimensional Morgan fingerprints.[26] Such a high-dimensional molecular representation necessitated dimensionality reduction while retaining valuable information simultaneously, for visualization and interpretation of molecular datasets. Moreover, the inherently discontinuous nature of chemical space makes the embedding techniques particularly useful for discerning discrete clusters of molecules with shared properties. An effective lower-dimensional representation of chemical space should, therefore, accurately place chemically similar molecules in proximity to one another and separate dissimilar ones, reflecting the local structure. It should also categorize related and unrelated molecular classes appropriately, indicative of the mesoscopic structure. Additionally, the overall relative positioning of different classes in this reduced space also holds significance, representing the global structure. While reduced embeddings should ideally be able to cluster into different regions based on functional groups, it is to be noted that the actual clustering of molecules is hierarchical, *e.g.*, linear or acyclic and cyclic compounds containing same functional groups have different physical and chemical properties and so should be placed separate from each other in the reduced embeddings space.[6] Therefore, we compared widely employed dimensionality reduction techniques—principal component analysis (PCA),[27] t-distributed stochastic neighbor



Scheme 1 Proposed electrolyte design paradigm. Proposed unsupervised data-driven paradigm combining current knowledge-based methods with ML approaches. Trained ML models from a previous study[3] were leveraged to screen large chemical space to electrolyte-relevant search space. This step represents an implemented preprocessing workflow for ionic conductivity, oxidative stability, and CE prediction to filter a large unlabeled chemical space into electrolyte-relevant solvent candidates. Full model details, input/output specifications, and filtering criteria are provided in Note S2. Faithful reduced embedding techniques are then used to map the electrolyte-relevant chemical space (including labeled and unlabeled electrolyte components) into a low-dimensional representation. Domain experts identify promising leads (hits) from the reduced representation using two strategies (oracle): selecting candidates closer to state-of-the-art (SOTA) electrolytes (exploitation) and selecting candidates away from the SOTA electrolytes (exploration). The best electrolyte candidates identified using this approach are further refined using traditional formulation-based and molecular design strategies. This integrated workflow aims to accelerate the discovery of high-performance electrolytes.

embedding (t-SNE),[28] and uniform manifold approximation and projection (UMAP)[29]—in terms of how they fare in preserving the high-dimensional representation at the three levels. The three techniques are discussed in detail in Note

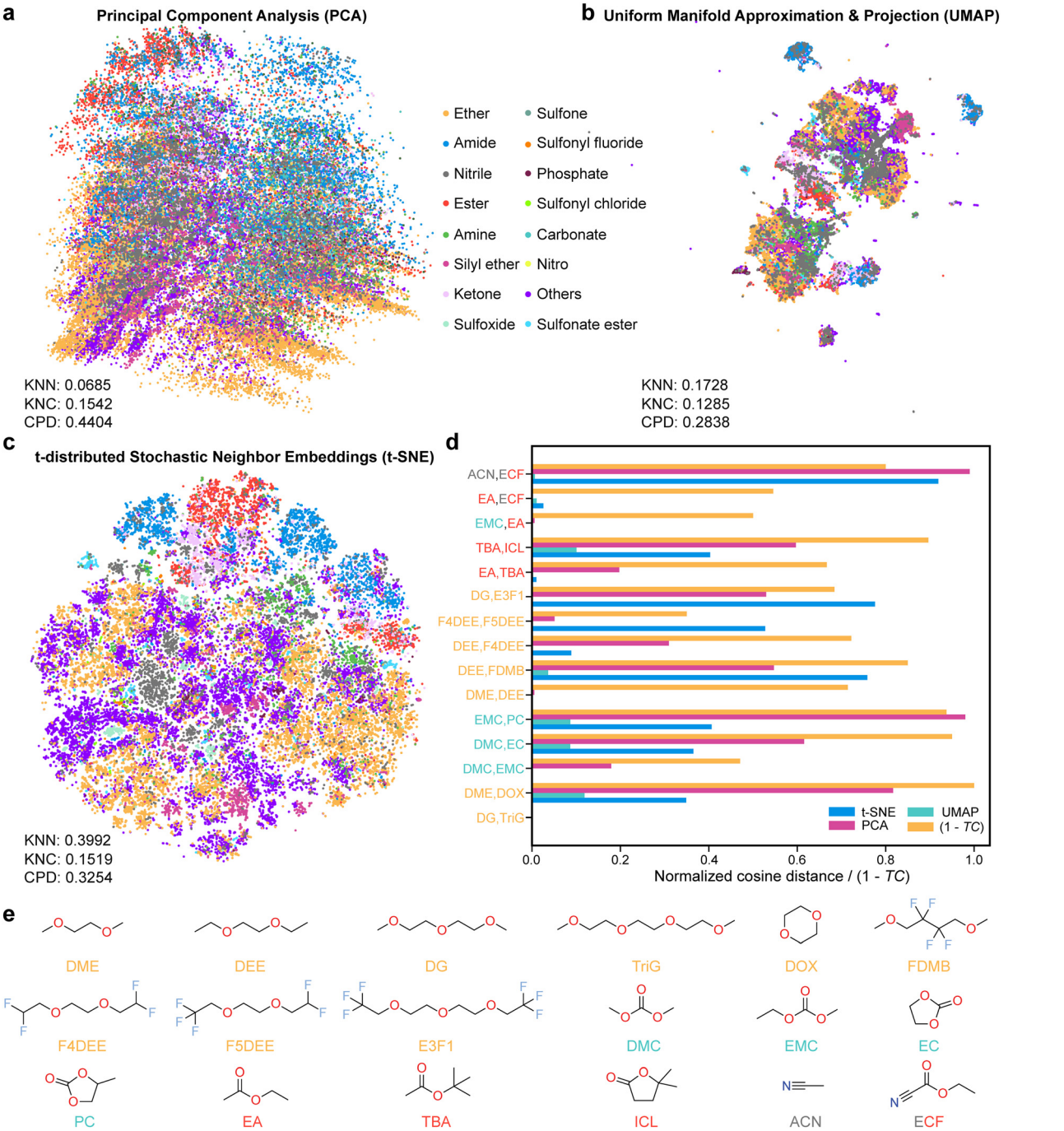


Fig. 1 Finding faithful representation of virtual chemical space of electrolyte solvents. Reduced embeddings of the virtual search space in two dimensions (2D) obtained from Morgan fingerprints using a PCA, b UMAP, and c t-SNE techniques, colored according to the functional groups present in the molecules (shown in legends). KNN = *k*-nearest neighbors, KNC = *k*-nearest classes, and CPD = Spearman correlation of pairwise distances. The higher the values, the better the embedding. d Comparison of normalized cosine distances between pairs of molecules (from a selected list of fifteen molecules) on the three embedding spaces and corresponding Tanimoto coefficient (T_C) values subtracted from unity. The closer the number is to 1, the more dissimilar the two compounds are. e Molecular structures of the solvent compounds chosen for comparison in d. DME = 1,2-dimethoxy ethane; DEE = 1,2-diethoxyethane; DG = diglyme; TriG = triglyme; DOX = 1,4-dioxane; F4DEE = 1,2-bis(2,2-difluoroethoxy)ethane; F5DEE = 2-[2-(2,2-difluoroethoxy)ethoxy]-1,1,1-trifluoroethane; FDMB = fluorinated 1,4-dimethoxylbutane; E3F1 = bis(2-(2,2,2-trifluoroethoxy)ethyl)ether; DMC = dimethyl carbonate; EMC = ethyl methyl carbonate; EC = ethylene carbonate; PC = propylene carbonate; EA = ethyl acetate; TBA = *tert*-butyl acetate; ICL = iso-caprolactone; ACN = acetonitrile; ECF = ethyl cyanoformate.

S3. At the outset, we clarify that the reduced solvent embeddings refer to the projection of Morgan fingerprints into two-dimensional space using the afore-mentioned dimensionality reduction techniques. Prior applications of such techniques in the drug discovery/cheminformatics,[30,31] battery (electrolyte solvents),[3,6] or materials science[32] domains often relied on arbitrarily choosing one of the three techniques without justification based on any quantitative metric. Therefore, we devised a reliable approach to obtain reduced dimensional representation of the chemical datasets presented in our work that can, in principle, be applied to any molecular datasets. For this, we used an existing methodology developed in the cell transcriptomics domain—Kobak *et al.*[33,34] that proposed three robust metrics to measure faithfulness of low-dimensional representations of the high-dimensional data—*k*-nearest neighbors (KNN), *k*-nearest classes (KNC), and Spearman correlation of pairwise distances (CPD) for measuring local, mesoscopic, and global levels of pictures (detailed discussion in Note S4).

Along similar lines, we evaluated the PCA, t-SNE, and UMAP reduced embeddings of original high-dimensional molecular search space as displayed in Fig. 1a–c, along with the corresponding KNN, KNC, and CPD values. Each of the molecules in the virtual search space have been assigned all possible functional groups (necessary for quantifying KNC) as described in Note S4. PCA performs best in terms of capturing global picture (highest CPD value), however suffers from incorrect local picture (lowest KNN value of 0.0685). Major functional groups such as ethers, amides, and esters are present in different regions of the PCA embeddings (indicative of preservation at mesoscopic and global levels) but there is also significant overlap, due to molecules containing multiple functional groups or due to multiple molecules sharing similar hydrocarbon skeleton apart from the functional groups present. However, the points within each functional groups are arranged in a similar fashion, indicating that the PCA has not captured patterns at the local level due to its linear nature. The UMAP captures local picture better than PCA (unique pattern of points within each color), but inferior to the t-SNE, that performs the best in terms of KNN values (0.3992). The UMAP also has lower KNC and CPD values compared to the PCA. In trying to separate out dissimilar points, it overcompensates and separates them out much farther apart than is necessary, *e.g.*, the two amide clusters in the top left and right corners in Fig. 1b are much farther apart compared to PCA (Fig. 1a). On the other hand, both the mesoscopic and global pictures are well preserved in t-SNE embeddings compared to the PCA (Fig. 1c), as the functional groups are positioned correctly with respect to each other. Hence, the t-SNE technique performs the best with regards to balancing all the local, mesoscopic, and global structures, respectively, of our molecular datasets. Kobak *et al.*[33,34] also found t-SNE embeddings to be more reliable in preserving the reduced embeddings of single cell transcriptomics datasets at the three levels compared to UMAP and PCA, but it may vary depending on the nature of the available datasets. Therefore, it is critical to verify the faithfulness of the reduced embedding techniques before they can be used to navigate a significantly high-dimensional chemical space for enabling new molecular discovery.

To visualize if distances in embeddings spaces are correlated with chemical similarity, an analysis was performed as shown in Fig. 1d. We compared two chemical similarity metrics, *i.e.*, normalized cosine distances ($d_{\cos}^{\text{norm}}$; a popular metric for measuring vector similarity;[35] Note S4) on the reduced two-dimensional embedding coordinates obtained from t-SNE and Tanimoto coefficients (T_C; a widely used metric for measuring chemical similarity;[36] Note S4) subtracted from unity. Both metrics were calculated for 15 pairs of known representative electrolyte solvent molecules present in the virtual search space (structures shown in Fig. 1). We do not infer $d_{\cos}^{\text{norm}}$ from visual Euclidean distances in the 2D t-SNE plots or from the original high dimensional descriptor space. The relative trends in chemical similarities are in general agreement with all the reduced embedding techniques, with t-SNE matching to the highest degree. However, the maximum $d_{\cos}^{\text{norm}}$ values for the UMAP embeddings do not surpass value of 0.2, even though chemical dissimilarity obtained through $(1 - T_C)$ are quite high for some of the pairs of molecules in Fig. 1d. The $d_{\cos}^{\text{norm}}$ values obtained from PCA are unusually high for acyclic and cyclic pairs of compounds, *e.g.*, (EMC, PC) and (EA, TBA) even compared to pairs from different functional groups such as (EA, ECF). However, we also establish shortcomings of the T_C metric itself here. The Tanimoto dissimilarity values for (DME, DEE) and (DMC, EMC) pairs are unexpectedly high, while $d_{\cos}^{\text{norm}}$ from at least two out of the three reduced embeddings consistently come out to be close to zero. Only the $d_{\cos}^{\text{norm}}$ values from t-SNE embedding seem to trend along expected lines for molecules belonging to both same and different classes as well as for acyclic and cyclic compounds with the same functional group.

Selecting promising electrolytes and half-cell performance comparison

After establishing the faithfulness of chemical space of electrolyte solvent molecules, promising molecules based on their similarity with respect to the state-of-the-art (SOTA) electrolytes were selected for battery testing, subject to following commercial availability (determined using https://search.emolecules.com[37]) criteria: available in the United States with price <$300 per g and lead time <6 weeks (further details in Note S5). We chose fluorinated 1,4-dimethoxylbutane (FDMB),[10] 1,2-bis(2,2-difluoroethoxy)ethane (F4DEE),[9] 2-[2-(2,2-difluoroethoxy)ethoxy]-1,1,1-trifluoroethane (F5DEE),[9] bis(2-(2,2,2-trifluoroethoxy)ethyl)ether (E3F1),[8] bis(2-fluoroethyl) ether (BFE),[38] and ethanesulfonyl fluoride (ESF)[3] as representative SOTA electrolytes that exhibit high Li||Cu CEs (>99.0%; Fig. 2(a)). Clearly, most of the SOTA electrolytes (except for ESF) consist of fluorinated ether solvents, hence indicating a sampling bias in terms of the investigation of

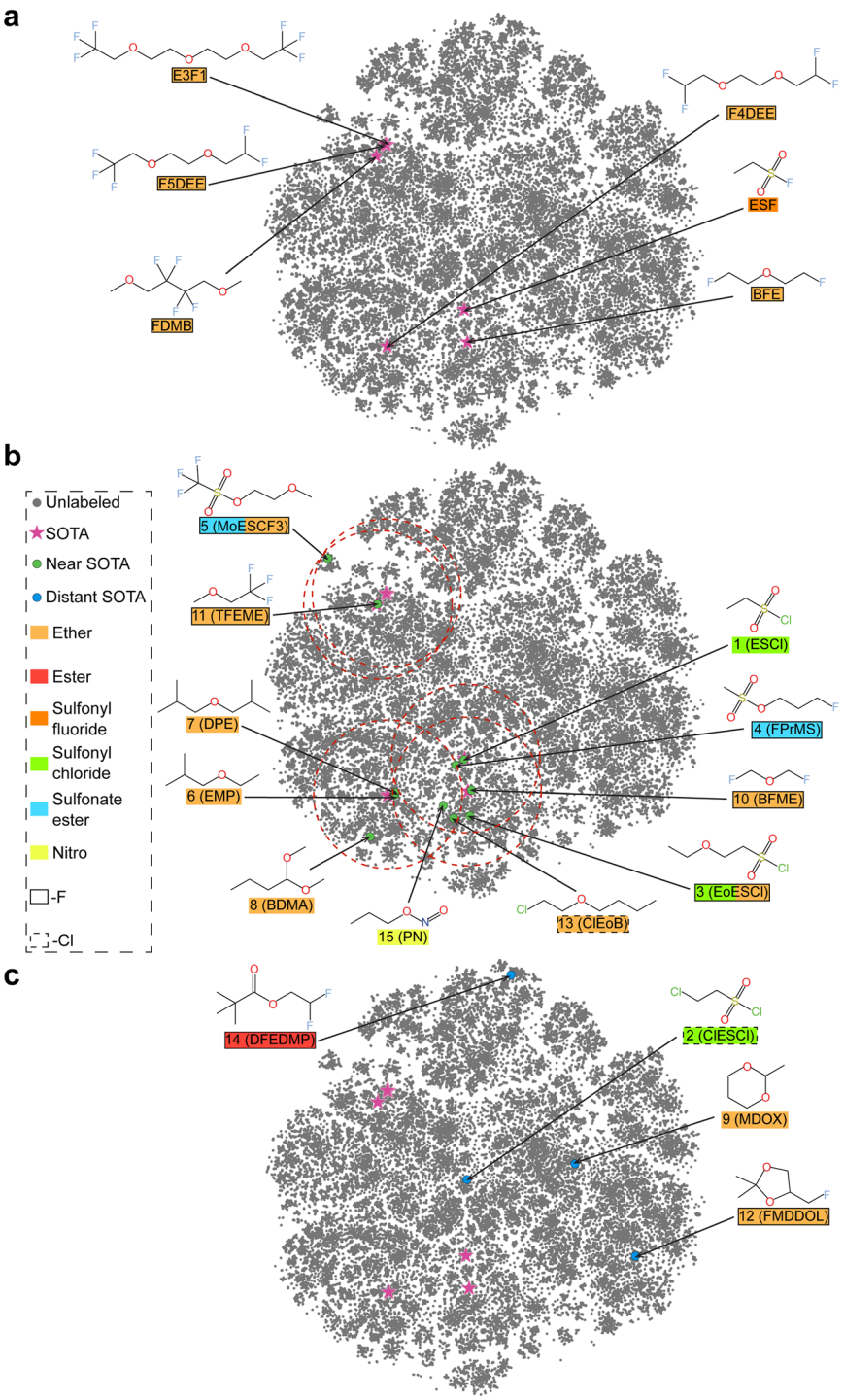


Fig. 2 Screening promising electrolyte solvents and experimental testing. 2D t-SNE plots depicting all molecules in the virtual search space (gray scatter points) along with a known state-of-the-art electrolytes (purple star-shaped scatter points; with acronyms in the colored boxes below the molecular structures), solvent candidate molecules selected for experimental validation (cyan scatter points; with numbers in colored boxes below the molecular structures) b in proximity (exploitation strategy) and c far away (exploration) from the SOTA electrolytes. The color of the boxes below the molecular structures denote the functional groups to which the molecules belong to (according to classification shown in Fig. 1). The circles in b are drawn as a guide to the eye to depict the SOTA electrolyte solvent molecule around which the new solvent molecules were chosen. Since the selection criteria was based on the normalized cosine distances, which do not correspond to the actual physical distance in the t-SNE plot, these circles are not drawn to scale.

chemical space till now. To circumvent this issue, we devised the following strategy to choose new electrolyte solvent molecules from the molecular search space for experimental validation: i) molecules chemically similar to the SOTA

molecules by selecting molecules with d_{cos}^{norm} values lower than a threshold (0.05), *i.e.*, closer to the SOTA molecules and hence termed as exploitation (Fig. 2(b)), ii) molecules chemically dissimilar to the SOTA molecules by selecting molecules far away from the SOTA, specifically with d_{cos}^{norm} values greater than the threshold and termed as exploration (Fig. 2(c)). We clarify here that the exploitation and exploration strategies defined here do not refer to the exploitation and exploration used in Bayesian optimization,[21] since we use a static reduced embedding similarity as oracle in a single-shot manner rather than explicit mathematical acquisition functions that update suggestions iteratively as new data emerge. The molecules selected under the exploitation strategy are—TFEME (#11; near FDMB, F5DEE, and E3F1), EMP (#6), DPE (#7) (both near F4DEE), ESCl (#1), FPrMS (#4) (both near ESF), BFME (#10), EoESCl (#3), ClEoB (#13), and PN (#15) (all near BFE). The remaining molecules selected under the exploration strategy are—ClESCl (#2), BDMA (#9), FMDDOL (#12), and DFEDMP (#14), leading to a total of 15 molecules selected for this study to investigate data-driven molecular design approach (full names along with SMILES listed in Table S2).

The battery cycling tests require dissolving salts in the selected solvent candidates to form electrolytes. We chose lithium bisfluorosulfonylamide (LiFSA) as salt at 1 or 2 M salt concentration since LiFSA has been widely reported to be compatible with NGBs[8,39,40] and all the SOTA electrolytes have also been investigated with the LiFSA salt at these salt concentrations. All solubility tests have been shown pictorially in Fig. S3 and listed in Table S2. A total of nine out of fifteen solvent molecules passed the 1 M/2 M solubility test (meaning they can dissolve up to at least 1 M). Of particular importance is the fact that BFME did not dissolve in 1 M LiFSA, despite lying close to the BFE molecule that forms a SOTA electrolyte with the LiFSA salt. It is also to be noted that 1 M LiFSA/EMP solution leads to phase separation possibly due to the asymmetry of the ether solvent molecule.[25] The issue was solved by increasing the salt concentration to 2 M. Interestingly, the solubility of LiFSA in DPE is as low as 0.04 M (Fig. S4), hence both examples illustrate that even small changes in molecular design can lead to a drastic change in properties. We also calculated solvent–Li^+ binding energies and recently proposed solvent–cation affinity (α_s; Note S6)[41] using density functional theory (DFT) seeking to explain the trends in 1 M/2 M LiFSA solubility as shown in Fig. S5—however, we did not find that they can universally explain the trends in contrast with previous reports.[39] The trends can only be qualitatively explained by electrostatic potential (ESP) maps as shown in Fig. 3(a) for all 15 molecules. The minimum (ESP_{solv}^{min}) and maximum (ESP_{solv}^{max}) point on the ESP isosurface values are also shown in Fig. S5(b). For example, electron donating density (red color) with lower intensity is observed on donor atoms in ClESCl (ESP_{solv}^{min} = −138.68 kJ mol^{-1}) and DPE (ESP_{solv}^{min} = −147.99 kJ mol^{-1}) that did not dissolve 1 M LiFSA salt compared to their close counterparts that did dissolve salt – ESCl (ESP_{solv}^{min} = −141.51 kJ mol^{-1}) and EMP (ESP_{solv}^{min} = −152.57 kJ mol^{-1}), respectively.

Next, electrolytes that formed homogeneous salt solutions with the selected promising solvents were validated by measuring CE in Li|Cu cells using the Aurbach method.[42] The electrolytes that successfully dissolved 1 M/2 M LiFSA salt could be classified into 4 groups based on the voltage profile, which represents their failure modes (Fig. 4a). First, chloride-containing electrolytes (ESCl, ClEoB, and EoESCl) exhibit poor reductive stability either chemically or electrochemically. None of these electrolytes yield a smooth plating/stripping voltage profile without either spiking or hitting the voltage cutoff immediately after stripping starts. Specifically, the voltage profile of EoEsCl indicates reductive degradation at potentials higher than 0 V *versus* Li/Li^+, as evidenced by additional peaks observed in the ^1H NMR spectrum (Fig. S6). Immersing lithium foil into this solution results in lithium loss and the formation of insoluble particles (Fig. S6). Moreover, compared to the S–F bond in

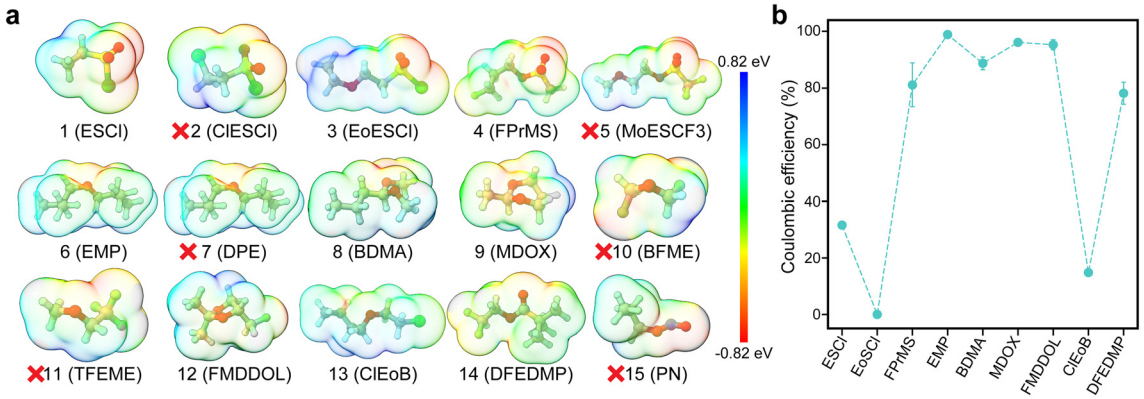


Fig. 3 Electronic structure and experimental CE performance of solvent molecules chosen. a Electrostatic potential maps (ESPs) for all the 15 solvent candidate molecules selected for experimental validation in the present work. The molecules that did not dissolve LiFSA salt up to 1 M are marked by cross marks to the left of molecule ID and acronyms (in parenthesis; refer Table S1 for full names). Isosurface value used for obtaining ESP maps = 1×10^{-3} e $Å^{-3}$. b Experimental Li||Cu CEs obtained using Aurbach test for the electrolytes corresponding to solvents that dissolved 1 M LiFSA (2 M LiFSA for EMP). The error bars correspond to standard deviation of at least three-coin cells.

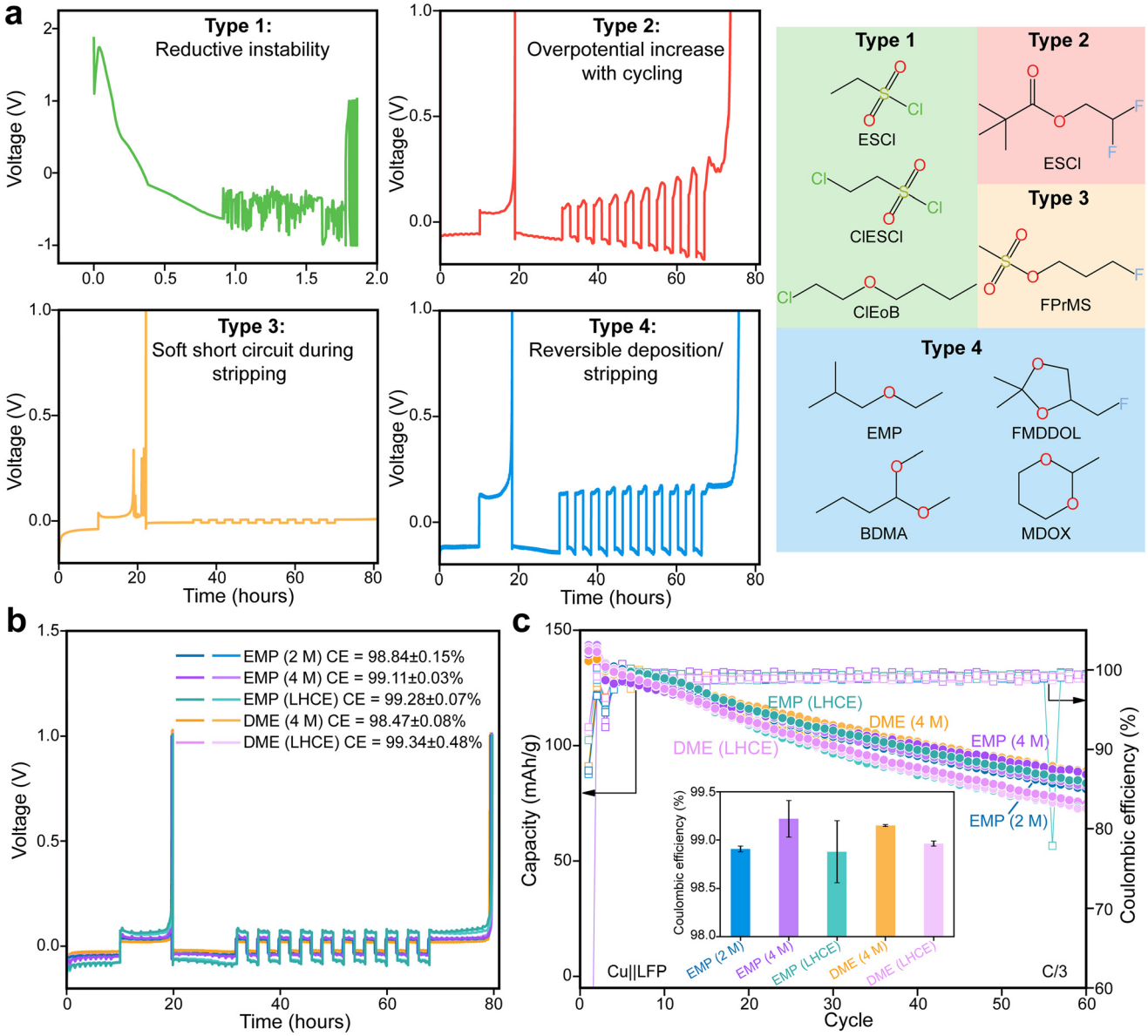


Fig. 4 Experimental battery cycling tests. a Classification of the nine single solvent-based electrolytes tested (that passed 1 M/2 M solubility test) into 4 distinct groups based on their (electro)chemical behavior. The representative Li||Cu voltage–time curves for electrolytes from each of the four groups are shown on the left side and the corresponding molecular structures of the solvents are shown on the right side. b Aurbach Li||Cu CE test for EMP and DME-based electrolytes. Multiple cells are shown. c Anode-free (Cu||LFP) cycling tests for EMP- and DME-based electrolytes. The CE averaged from cycles 11 to 60 are shown as bar plot in the inset for all electrolytes. Voltage is *versus* Li/Li^+. LFP: $LiFePO_4$, loading ~ 1.86 mA h cm^{-2}. 1*C* = 1.86 mA cm^{-2}.

sulfonyl fluoride (*i.e.*, the SOTA electrolyte ESF), the S–Cl bond of sulfonyl chloride in ESCl is more reductively unstable[43] even though it is structurally similar to ESF (very close in reduced embedding space, Fig. 2a). The excessive solvent degradation reaction caused from this reductive instability prevents the formation of a stable passivation layer on the lithium metal, leading to low CE (Fig. S7). Besides, the alkyl chloride electrolyte ClEoB, also suffers from limited reductive stability. After cycling, the lithium anode develops a dark gray color, and immersing lithium foil in this solution causes the solution to change color overnight (Fig. S8).

The second electrolyte group only contains fluorinated carbonate, DFEDMP. The voltage profile shows no spikes, however, there is a gradual increase in overpotential during ten cycles of plating/stripping with Li reservoir. This could indicate a constant rebuilding of poor SEI which accumulates into a higher transport barrier during cycling.

The third category only contains FPrMS, with the unique voltage profile including spikes at the stripping step of the precycle. This potentially represents soft short circuit[44] which could have come from dendritic lithium formed on the surface (Fig. S9).

The last group contains all the acetals and ethers (BDMA, MDOX, FMDDOL, EMP). All of them show smooth plating/stripping until the 1 V cutoff is reached. Electrolytes containing acetal groups (BDMA, MDOX, and FMDDOL) demonstrate moderate performance in Li||Cu cells, achieving CEs of 88.69%, 96.09%, and 95.24%, respectively. However, the primary limitation of this class is the instability of the acetal solvent in the presence of LiFSA salt. Over the course of a month, noticeable color changes occur in these solutions (Fig. S10a). Analysis of the degradation products *via* NMR spectra indicates that the hydrolysis of acetal is the dominant degradation pathway (Fig. S10b–d). We note that using other salts beyond LiFSA (*i.e.*, LiTFSI) can help mitigate this degradation. Among the selected electrolytes, EMP (2 M), MDOX, FMDDOL show the highest coulombic efficiencies (CE), with EMP (2 M) reaching 98.84%. While EMP is indeed found closer to a SOTA electrolyte solvent molecule (F4DEE, Fig. 2a), MDOX and MDDOL were distant compared to existing SOTA electrolyte solvents signifying their chemical uniqueness. Unfortunately, most of the molecules selected in unexplored regions (*e.g.*, DFEDMP) or electrochemically hitherto unexplored molecules, *e.g.*, nitro compound (PN) yielded unsatisfactory performance in Aurbach tests (Fig. S11). Interestingly, even some of the fluorinated solvent candidates cannot be used for LMBs such as FPrMS, BFME, and DFEDMP while the best performing electrolyte constituted a non-fluorinated solvent. We further checked if computational descriptors such as solvent–Li^+ binding energies (ΔE^{bind}; Fig. S12a) and LUMO energy levels (E_{solv}^{LUMO}; Fig. S12b) could act as descriptors to explain the trends in CEs observed; however, neither of them yields any clear trends, in contrast to recent reports[14] that have been investigated on much narrower sets of solvents.

We performed further improvements in molecular design based on chemical intuition, *e.g.*, introducing two methyl substitutions at the 2′ position of 1,3-dioxane could mitigate the degradation issues observed with MDOX. The synthesis route and the characterization of this new solvent are in Fig. S13. The modified molecule is expected to degrade into a ketone rather than an aldehyde, with the former being less susceptible to further oxidation. To test this hypothesis, we synthesized a modified version of MDOX (denoted as MDOX-M) and evaluated its performance under the same conditions. After decrimping the cycled cells, it is found that the concentration of degradation products remained close to the initial amount of 1 mol%, suggesting that the proposed structure offers greater resistance to high-voltage degradation compared to MDOX (Fig. S14).

Formulation design to enhance anode-free performance

Apart from Li||Cu half-cell performance (Fig. 4b), we also tested the physicochemical properties (Fig. S15) and the performance in anode-free LMB (Cu||LFP; no excess lithium) cells (Fig. 4c) of our best electrolytes to confirm their practical viability. The 2 M LiFSA/EMP electrolyte was explored in anode-free coin cells to test its full cell performance. Interestingly, the average CE (from the 11th to the 60th cycle) of anode-free cells (98.91 ± 0.03%) aligns well with CE (98.84 ± 0.15%) from the Aurbach test, indicating that the loss in capacity mainly comes from the plating/stripping inefficiency on the anode side for 2 M LiFSA/EMP. It is a common domain knowledge that plating/stripping inefficiency can be mitigated either by increasing the salt concentration to form a high-concentration electrolyte (HCE)[45] or by adding a diluent (1,1,2,2-tetrafluoroethyl 2,2,3,3-tetrafluoropropyl ether; TTE) to create a localized high-concentration electrolyte (LHCE).[46] To benchmark EMP as a solvent in electrolytes, its performance was compared with DME under identical conditions, specifically by evaluating HCE and LHCE prepared from both solvents. The improvement in CE achieved by formulating EMP into HCE and LHCE using TTE (details in the Methods section) is shown in Fig. 4b. 4 M LiFSA dissolved in EMP, denoted as EMP (4 M), shows CE of 99.14%, and the LHCE formed by EMP, denoted as EMP (LHCE), gives CE of 99.28%. The improvement in Li plating/stripping reversibility benefits the cycle life of anode-free cells using the same electrolytes. As shown in Fig. 4c, at the same salt concentration, EMP (4 M) shows an average CE of 99.22%, comparable to that of DME (4 M) (99.15%). EMP (LHCE) also gives comparably high CE to DME (LHCE). The details in the formulation effect on solvation structure are discussed in Note S7 and shown in Fig. S16. Both LHCE show higher capacity retention in anode-free configuration than other widely used carbonate or ether-based mixed-solvent systems (Fig. S17). Our result establishes EMP as a promising electrolyte solvent for next-generation LMBs as EMP-based electrolytes compare favorably with SOTA electrolytes.

Discussion

It is clear that fixed or static molecular embeddings such as t-SNE embeddings from Morgan fingerprints of solvent molecules can be nuanced when applied directly to finding new electrolytes even when the salt is fixed. Therefore, structural similarity of solvent molecules does not necessarily lead to functional similarity. This inequivalence of structural and functional similarity in reduced embeddings has also been reported in a prior study on molecular discovery for organic LED applications[47] (called activity cliffs). The above findings signify that data-driven molecular design approach to efficient electrolyte discovery is a significantly challenging task. This is further complicated by the bias in existing labeled data (lack of negative data), sparse exploration of chemical space, availability of low number of explored solvents, and lack of *a priori* knowledge about synthesizability and salt solubility of unexplored solvents. However, gaining physical insights and trends even from the limited labeled data (including failed solvents to overcome positive data bias) and leveraging it to propose design principles can be advantageous to accelerate the discovery process with higher success rate. Additionally, the mapping between solvent similarity in the reduced space and electrolyte performance

(*e.g.*, CE) is expected to be salt- and formulation-dependent, as the salt identity and concentration modulate solvation structure, ion association, transport, and interphase chemistry. While the primary focus of the present study is on finding high-performing LiFSA-based single solvent electrolytes, extension to alternative salts would require incorporating reduced representations of salts (including salt structure and concentration) and training supervised ML models with both solvent and salt descriptors as input features. However, such supervised ML frameworks typically require substantial experimental data,[21] which currently remain limited for LMB systems.[3,17] For multi-component electrolytes (co-solvents, diluents, LHCE-type formulations), mixture representations can be constructed from component embeddings (*e.g.*, composition-weighted[48] or interaction-aware), and mixture-to-performance relationships can be modeled. Nonetheless, mixture behavior may not be strictly additive,[49] and cross-interactions between solvents and ions can introduce additional complexity. Explicitly capturing these non-additive effects remains an important direction for future work.

Conclusions

In summary, we carried out an unsupervised data-driven approach for efficient electrolyte discovery for lithium-based next-generation batteries. The reduced embeddings from three different dimensionality reduction techniques underwent stringent faithfulness checks at preserving local, mesoscopic, and global pictures. The t-SNE embeddings emerged as the best among the three, which then allowed us to explore vast stretches in the chemical space of electrolyte solvents and test them in real-world batteries. Cheap computational descriptors were also calculated using DFT to explain trends in 1 M/2 M salt solubility and half-cell CE performance. The lithium metal battery performance of the best unary electrolyte solvent (1-ethoxy-2-methylpropane; EMP) found from our data-driven approach shows comparable capacity retention and high CE to DME-based electrolytes, with longer cycle life than widely used mixed-solvent electrolytes. Our results demonstrate that fluorination is not the only viable pathway to high-performance electrolytes, and that non-fluorinated solvents can be competitive when their solvation characteristics are appropriately engineered such as high ion-pairing. Our data-driven methodology combined with the knowledge paradigm embodies a transformative shift in battery materials research, moving from the slow, iterative trial-and-error methods to a rapid discovery process.

Methods

DFT calculations

All DFT calculations including binding energies, LUMO energy levels, ESP, and solvent–cation affinity were performed using Gaussian 16 (version 16.0.1) program. B3LYP and 6-311**g(d,p) was utilized as the functional and basis sets, respectively. van der Waals corrections were added using the GD3BJ method for initial optimization only. Implicit solvation corrections considering tetrahydrofuran (THF) as the solvent were also included using the polarizable continuum model (PCM) method. ESP map visualization was carried out using ChimeraX program.[50]

Data analysis

All codes utilized in this study were built using various libraries available in the Python (version 3.13.5) programming language. The RDKit library was utilized for obtaining Morgan fingerprints and for evaluating Tanimoto coefficients. The reduced embeddings of Morgan fingerprints corresponding to t-SNE, UMAP, and PCA techniques were generated using OpenTSNE, UMAP, and Scikit-learn (version 0.23) libraries. All plots in the manuscript were made using Matplotlib and Seaborn python libraries.

Materials

Methyl *p*-toluenesulfonate (98%, Ambeed), sodium sulfate (anhydrous, Fisher Scientific), sodium hydride (60%, in mineral oil, Sigma-Aldrich) 2,2-dimethyl-5-hydroxymethyl-1,3-dioxane (97%, Combi-blocks), lithium hexafluorophosphate solution (1.0 M $LiPF_6$ in EC/DMC = 1 : 1 in volume, battery grade, Sigma-Aldrich) were used as received. Lithium bis(fluorosulfonyl) amide (LiFSA, 99%, Arkema Inc.) was vacuum-dried at 30 °C in a glovebox antechamber before use and transferred to a glovebox without air exposure. Lithium bis(trifluoromethanesulfonyl)imide (LiTFSI, 99.99%, Sigma-Aldrich) was vacuum-dried at 120 °C in the same way. Lithium perchlorate ($LiClO_4$, 99%, Oakwood Chemical) was dried in the same way at 110 °C.

Dimethoxyethane (DME, 99.5%, anhydrous, inhibitor-free, Sigma-Aldrich), 1,1,2,2-tetrafluoroethyl 2,2,3,3-tetrafluoropropyl ether (TTE, Synquest), deuterated acetonitrile (d_3-ACN, 99.8 atom% D, Cambridge Isotope Laboratories), deuterated chloroform ($CDCl_3$, 99.5 atom% D, Cambridge Isotope Laboratories), deuterated dimethyl sulfoxide (DMSO-d_6, 99.9 atom% D, Cambridge Isotope Laboratories), 2-methoxyethyl trifluoromethanesulfonate (Asta Tech), 1-(2-chloroethoxy)butane (98%, Ambeed), 3-fluoroprop-1-yl methanesulfonate (98%, Ambeed), 2-ethoxyethanesulfonyl chloride (96%, Ambeed), ethanesulfonyl chloride (98%, Fisher Scientific), 2-chloroethanesulfonyl chloride (98%, Fisher Scientific), 2-fluoroethyl methanesulfonate (95%, Ambeed), 2-chloroethyl methanesulfonate (98%, Ambeed), 2,2-difluoroethyl methanesulfonate (95%, Ambeed), (*S*)-4-(fluoromethyl)-2,2-dimethyl-1,3-dioxolane (98%, Ambeed), 2,2-difluoroethyl 2,2-dimethylpropanoate (A2B), 1-ethoxy-2-methylpropane (95%, Asta Tech), diisobutyl ether (98%, TCI), bis(fluoromethyl)ether (Synquest), 2,2,2-trifluoroethyl methyl ether (98%, Fisher Scientific), 2-methyl-1,3-dioxane (98%, Ambeed), *N*-propyl nitrite (95%, Asta Tech), and

1,1-dimethoxybutane (95%, Asta Tech) were dried with 4 Å molecular sieves for at least 3 days before used. 1,3-Dioxolane (anhydrous, contains 75 ppm BHT, Sigma-Aldrich) was distilled and dried with 4 Å molecular sieves for at least 3 days before used.

Li (500 μm thickness, China Energy Lithium) and thin Li (20 μm on Cu, China Energy Lithium) was brushed to polish before cell assembly. $LiFePO_4$ (LFP, 92 wt%, active materials, 3.9 wt% Timcal C-45, 4 wt% Solvay 5130 poly(vinylidene difluoride) binder, 0.1 wt% Tuball SWCNT) and $LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$ (NMC811, 90 wt%, active materials, 5 wt% Timcal C-45, 5 wt% Solvay 5130 poly(vinylidene difluoride) binder) were kindly provided by Steve Trask and Andrew Jansen at Argonne's cell analysis, modeling, and prototyping facility. The cathodes were cut into disks (12 mm in diameter) and vacuum-dried at 120 °C in a glovebox antechamber before use.

The ether-based mixed-solvent control electrolyte was prepared by dissolving 0.5 M LiTFSI and 0.5 M LiFSA in 1,3-dioxolane and DME at a 2 : 1 ratio. All other electrolytes used in this work were prepared by dissolving LiFSA in the solvent of interest. Localized high-concentration electrolytes (LHCEs) were prepared by first mixing TTE and LiFSA at a 3 : 1 molar ratio (TTE : LiFSA = 3 : 1), followed by gradual addition of the solvent (either EMP or DME) until no visible salt precipitate remained, as per standard procedure.[51] The salt concentration of LHCE of EMP and DME electrolytes are 1.48 M and 1.96 M respectively.

Synthesis

MDOX-M was synthesized *via* Williamson ether synthesis. Dried tetrahydrofuran (70 mL) was added to a round-bottom flask, followed by sodium hydride (2.7 g, 1.2 equiv.), which was suspended in the solvent. The reaction mixture was cooled to 0 °C, and after 15 min of equilibration, 2,2-dimethyl-5-hydroxymethyl-1,3-dioxane (7.16 g, 1.0 equiv.) was added dropwise. The mixture was then warmed to room temperature, and methyl *p*-toluenesulfonate (12.2 g, 1.2 equiv.), pre-dissolved in 20 mL of dried THF, was added dropwise. The reaction was stirred overnight. The crude product was extracted once with water and diethyl ether, and the organic phase was subsequently washed twice with brine. The combined organic layers were dried over sodium sulfate, filtered through cotton, and concentrated by rotary evaporation. The residue was distilled twice at 50 mbar to yield the final product (b.p. ~90 °C, yield = 55.4%).

Characterization and battery tests

^1H, ^{13}C, ^7Li, and HSQC NMR are performed in Bruker Ascend 9.4 T/400 MHz instrument using either $CDCl_3$, d_3-ACN, or DMSO-d_6 as the deuterated solvent. Salt solubility in the work was determined by capillary NMR setup. The sample was put in a capillary and sealed with PTFE cap. A reference solution of 0.1 M $LiClO_4$ in d_3-ACN was added in a precision NMR tub outside the capillary. Sample preparation was done in the glovebox before the tube was sealed and taken out for measurement. Scanning electron microscopy (SEM) is performed on the Cu electrode of Li||Cu half cells using a Carl Zeiss Merlin field emission scanning electron microscope. The Cu foil surface was first electrochemically cleaned by applying 5 precycles from 0 to 1 V at 0.02 mA cm^{-2}. Then, Li was deposited to 1.5 mA h cm^{-2} at a rate of 1.0 mA cm^{-2}. Li deposit was washed with DME 3 times before dried under vacuum. Sample preparation was done in glovebox under Ar atmosphere.

All coin cells are in CR2032 cell type and crimped using the configuration described below: negative case||spring||spacer||anode||25 μL electrolyte||Celgard separator||25 μL electrolyte||cathode||positive case. Each part was added in the above order except for electrolyte containing EMP. While making EMP cells, 50 μL electrolyte was added only once after placing Celgard 2325. The EMP cells were crimped within 20 seconds after electrolyte addition to minimize electrolyte evaporation. Celgard 3501 was used in sulfonate-based and EC/DMC electrolytes to improve wetting. Celgard 2325 was used in all the other electrolytes.

Conductivity of the electrolyte was measured with electrochemical impedance spectroscopy using a Biologic VSP-300 potentiostat. The range of the spectra is between 7 MHz and 100 Hz. Stainless steel was used as both the anode and the cathode in the coin cell, and the cell is tested in an Espec oven (BTS-133) with controlled temperature sweeping from 20 °C to −60 °C. The measured conductivity was multiplied by a cell constant of 13 to give the conductivity presented in this work.

Linear sweep voltammetry was measured with Biologic MPG-2 potentiostat at 20 °C, using Li as the anode and aluminum foil as the cathode. The cell was first rested for 3 hours, then subjected to a voltage from the open-circuit potential to 6.0 V. The scan rate was 1 mV s^{-1}.

Li/Li^+ potential test was performed following the method proposed by Yamada *et al.*[14] Specifically, 1 mM ferrocene was dissolved in the electrolyte of interest, and the combined solution was injected into a sealed microbeaker in glovebox. Using Pt wire as the working electrode and Li as the counter electrode, a cyclic voltammetry was performed using a scan rate of 5 mV s^{-1}. The halfway potential was calculated by averaging over the oxidation and reduction peak positions.

Li||Cu, Cu||LFP, and thin Li||NMC811 cells were all tested with a Neware BTS4000 battery tester at 20 °C. Coulombic efficiency was determined using the Aurbach method of Li||Cu cells. After 10 h resting, a formation cycle was performed at 0.5 mA cm^{-2} to a capacity of 5 mA h cm^{-2}, followed by a 10 h deposition step and 10 subsequent cycles of 2 h Li plating/stripping at the same current density. Finally, Li was stripped from the Cu foil to a cutoff of 1.0 V. CE was determined as the ratio of the total stripping capacity to the plating capacity excluding the formation cycle. Cu||LFP cells were rested for 10 h before cycling. Two precycles at C/10 were performed before the cells were cycled at C/3. The LFP loading ≈1.86 mA h cm^{-2}, the NMC811 loading ≈1.66 mA h cm^{-2}, and the current density was set based on the actual mass of active

materials in the cathode. The average CE of Cu||LFP cells were calculated from CE between the 11th and the 60th cycle.

Author contributions

R. K. and K.-H. W. contributed equally to this work. R. K. curated the raw extracted data, performed ML training and inference, DFT calculations, and analyzed the results. K.-H. W. performed all experiments (synthesis, characterization, battery cycling) and analyzed the experimental results. R. K., K.-H. W., and C. V. A. contributed to the discussion of results and writing of the manuscript. C. V. A. supervised the work.

Conflicts of interest

The authors declare no competing interests.

Code availability

The Jupyter notebooks/codes necessary for reproducing main results of this study are also freely available on the GitHub repository (https://github.com/AmanchukwuLab/data-driven-electrolyte-solvents-discovery).

Data availability

The data supporting this article have been included as part of the supplementary information (SI). The data, including virtual search space and labeled experimental data along with computational descriptors, is also available at Zenodo at https://doi.org/10.5281/zenodo.17330301.[52]

Supplementary information: 1) description of pseudocode to manually enumerate ether solvent molecules; 2) description of virtual search space used; 3) description of the three reduced embedding techniques used; 4) metrics used for quantifying faithfulness of reduced embeddings; 5) DFT-based analyses carried out; 6) effect of formulating EMP-based electrolytes; 7) salt solubility tests; 8) several experimental characterizations, *e.g.*, NMR, Raman, SEM, *etc.* to support various hypotheses proposed; 8) ionic conductivity, Li||Cu half-cell, and Cu||LFP anode-free cell performance of different electrolytes investigated. See DOI: https://doi.org/10.1039/d5me00188a.

Acknowledgements

This work was partially supported by the National Science Foundation CAREER Award (CBET-2144454) and the University of Chicago Data Science Institute AI + Science Research Initiative. C. V. A. acknowledges the Neubauer Family Assistant Professors Program. R. K. acknowledges support from the Eric and Wendy Schmidt AI in Science Fellowship, a Schmidt Sciences 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.

References

1 G. M. Hobold, J. Lopez, R. Guo, N. Minafra, A. Banerjee, Y. Shirley Meng, Y. Shao-Horn and B. M. Gallant, Moving beyond 99.9% Coulombic Efficiency for Lithium Anodes in Liquid Electrolytes, *Nat. Energy*, 2021, **6**(10), 951–960, DOI: 10.1038/s41560-021-00910-w.

2 Y. S. Meng, V. Srinivasan and K. Xu, Designing Better Electrolytes, *Science*, 2022, **378**(6624), eabq3750, DOI: 10.1126/science.abq3750.

3 R. Kumar, M. C. Vu, P. Ma and C. V. Amanchukwu, Electrolytomics: A Unified Big Data Approach for Electrolyte Design and Discovery, *Chem. Mater.*, 2025, **37**(8), 2720–2734, DOI: 10.1021/acs.chemmater.4c03196.

4 D. Wang, T. He, A. Wang, K. Guo, M. Avdeev, C. Ouyang, L. Chen and S. Shi, A Thermodynamic Cycle-Based Electrochemical Windows Database of 308 Electrolyte Solvents for Rechargeable Batteries, *Adv. Funct. Mater.*, 2023, **33**(11), 2212342, DOI: 10.1002/adfm.202212342.

5 D.-J. Yoo, S. Yang, K. J. Kim and J. W. Choi, Fluorinated Aromatic Diluent for High-Performance Lithium Metal Batteries, *Angew. Chem., Int. Ed.*, 2020, **59**(35), 14869–14876, DOI: 10.1002/anie.202003663.

6 Y.-C. Gao, N. Yao, X. Chen, L. Yu, R. Zhang and Q. Zhang, Data-Driven Insight into the Reductive Stability of Ion–Solvent Complexes in Lithium Battery Electrolytes, *J. Am. Chem. Soc.*, 2023, **145**(43), 23764–23770, DOI: 10.1021/jacs.3c08346.

7 Y. Wu, Q. Hu, H. Liang, A. Wang, H. Xu, L. Wang and X. He, Electrostatic Potential as Solvent Descriptor to Enable Rational Electrolyte Design for Lithium Batteries, *Adv. Energy Mater.*, 2023, **13**(22), 2300259, DOI: 10.1002/aenm.202300259.

8 P. Ma, P. Mirmira and C. V. Amanchukwu, Effect of Building Block Connectivity and Ion Solvation on Electrochemical Stability and Ionic Conductivity in Novel Fluoroether Electrolytes, *ACS Cent. Sci.*, 2021, **7**(7), 1232–1244, DOI: 10.1021/acscentsci.1c00503.

9 Z. Yu, P. E. Rudnicki, Z. Zhang, Z. Huang, H. Celik, S. T. Oyakhire, Y. Chen, X. Kong, S. C. Kim, X. Xiao, H. Wang, Y. Zheng, G. A. Kamat, M. S. Kim, S. F. Bent, J. Qin, Y. Cui and Z. Bao, Rational Solvent Molecule Tuning for High-Performance Lithium Metal Battery Electrolytes, *Nat. Energy*, 2022, **7**(1), 94–106, DOI: 10.1038/s41560-021-00962-y.

10 Z. Yu, H. Wang, X. Kong, W. Huang, Y. Tsao, D. G. Mackanic, K. Wang, X. Wang, W. Huang, S. Choudhury, Y. Zheng, C. V. Amanchukwu, S. T. Hung, Y. Ma, E. G. Lomeli, J. Qin, Y. Cui and Z. Bao, Molecular Design for Electrolyte Solvents Enabling Energy-Dense and Long-Cycling Lithium Metal Batteries, *Nat. Energy*, 2020, **5**(7), 526–533, DOI: 10.1038/s41560-020-0634-5.

11 L. Kim, T. Jang and H. R. Byon, Fluorinated Ether-Based Co-Solvent Electrolytes for Lithium-Metal Batteries: High Ionic

Conductivity and Suppressed Dissolution of Fragmented Anions, *J. Power Sources*, 2023, **576**, 233237, DOI: 10.1016/j.jpowsour.2023.233237.

12 Y. Jeoun, K. Kim, S.-Y. Kim, S.-H. Lee, S.-H. Huh, S. H. Kim, X. Huang, Y.-E. Sung, H. D. Abruña and S.-H. Yu, Surface Roughness-Independent Homogeneous Lithium Plating in Synergetic Conditioned Electrolyte, *ACS Energy Lett.*, 2022, **7**(7), 2219–2227, DOI: 10.1021/acsenergylett.2c00974.

13 E. Park, Y.-H. Lee, S.-H. Huh, J. Huh, Y.-E. Sung and S.-H. Yu, Bifunctional Trimethylsilyl-Modified Fluorinated Ester Additive for LiF-Rich Solid Electrolyte Interphase in Lithium Metal Batteries, *Energy Storage Mater.*, 2025, **78**, 104271, DOI: 10.1016/j.ensm.2025.104271.

14 S. Ko, T. Obukata, T. Shimada, N. Takenaka, M. Nakayama, A. Yamada and Y. Yamada, Electrode Potential Influences the Reversibility of Lithium-Metal Anodes, *Nat. Energy*, 2022, **7**(12), 1217–1224, DOI: 10.1038/s41560-022-01144-0.

15 Z. Zheng, X. Liu, X.-Q. Zhang, S.-Y. Sun, J.-L. Li, Y.-N. Wang, N. Yao, D.-H. Zhan, W.-J. Feng, H.-J. Peng, J.-K. Hu, J.-Q. Huang and Q. Zhang, Deciphering Coulombic Efficiency of Lithium Metal Anodes by Screening Electrolyte Properties, *Angew. Chem.*, 2025, **64**(30), e202507387, DOI: 10.1002/ange.202507387.

16 P. Kirkpatrick and C. Ellis, Chemical Space, *Nature*, 2004, **432**(7019), 823–823, DOI: 10.1038/432823a.

17 S. C. Kim, S. T. Oyakhire, C. Athanitis, J. Wang, Z. Zhang, W. Zhang, D. T. Boyle, M. S. Kim, Z. Yu, X. Gao, T. Sogade, E. Wu, J. Qin, Z. Bao, S. F. Bent and Y. Cui, Data-Driven Electrolyte Design for Lithium Metal Anodes, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**(10), e2214357120, DOI: 10.1073/pnas.2214357120.

18 M. Zohair, V. Sharma, E. A. Soares, K. Nguyen, M. Giammona, L. Sundberg, A. Tek, E. A. V. Vital and Y.-H. La, Chemical Foundation Model Guided Design of High Ionic Conductivity Electrolyte Formulations, *arXiv*, 2025, preprint, arXiv:2503.14878, DOI: 10.48550/arXiv.2503.14878.

19 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**(1), 235–242, DOI: 10.1093/nar/28.1.235.

20 N. N. Rajput, R. Atwi, D. Gribble, D. T. Nguyen, M. Song, A. Babbar, M. Vijayakumar and V. Pol, Knowledge-Driven Design of Fluorinated Ether Electrolytes via a Multi-Model Approach, *Research Square*, 2024, DOI: 10.21203/rs.3.rs-4364688/v1.

21 P. Ma, R. Kumar, K.-H. Wang and C. V. Amanchukwu, Active Learning Accelerates Electrolyte Solvent Screening for Anode-Free Lithium Metal Batteries, *Nat. Commun.*, 2025, **16**(1), 8396, DOI: 10.1038/s41467-025-63303-7.

22 M. Zhu, Y. Lin, Z. Pang, Y. Zhao, J. Li, B. Zhang, B. Xu, F.-Z. Xuan, H. Liu and C. Lian, An Automated Workflow for Lithium Metal Oxide Screening Assisted by Large Language Models, *Chin. Chem. Lett.*, 2025, 112217, DOI: 10.1016/j.cclet.2025.112217.

23 X. Gao, A. Yang, W. Yu, J. Zhou, M. Pei, J. Chen, W. Yan, G. Luo, Y. Liu, J. Xiao and J. Zhang, Generative Artificial Intelligence Navigated Development of Solvents for Next Generation High-Performance Magnesium Batteries, *Adv. Mater.*, 2025, e10083, DOI: 10.1002/adma.202510083.

24 X. Chen, M. Liu, S. Yin, Y.-C. Gao, N. Yao and Q. Zhang, Uni-Electrolyte: An Artificial Intelligence Platform for Designing Electrolyte Molecules for Rechargeable Batteries, *Angew. Chem., Int. Ed.*, 2025, **64**(30), e202503105, DOI: 10.1002/anie.202503105.

25 P. Ma, U. Le, K.-H. Wang, M. C. Vu, P. Mirmira and C. V. Amanchukwu, Probing the Influence of Steric Hindrance in Nonfluorinated Ether Electrolytes for Lithium Metal Batteries, *J. Electrochem. Soc.*, 2024, **171**(12), 120536, DOI: 10.1149/1945-7111/ad9cca.

26 D. Rogers and M. Hahn, Extended-Connectivity Fingerprints, *J. Chem. Inf. Model.*, 2010, **50**(5), 742–754, DOI: 10.1021/ci100050t.

27 H. Hotelling, Analysis of a Complex of Statistical Variables into Principal Components, *J. Educ. Psychol.*, 1933, **24**(6), 417–441, DOI: 10.1037/h0071325.

28 L. van der Maaten and G. Hinton, Visualizing Data Using T-SNE, *J. Mach. Learn. Res.*, 2008, **9**(86), 2579–2605.

29 L. McInnes, J. Healy and J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv*, 2020, preprint, arXiv:1802.03426, DOI: 10.48550/arXiv.1802.03426.

30 F. Wong, E. J. Zheng, J. A. Valeri, N. M. Donghia, M. N. Anahtar, S. Omori, A. Li, A. Cubillos-Ruiz, A. Krishnan, W. Jin, A. L. Manson, J. Friedrichs, R. Helbig, B. Hajian, D. K. Fiejtek, F. F. Wagner, H. H. Soutter, A. M. Earl, J. M. Stokes, L. D. Renner and J. J. Collins, Discovery of a Structural Class of Antibiotics with Explainable Deep Learning, *Nature*, 2024, **626**(7997), 177–185, DOI: 10.1038/s41586-023-06887-8.

31 J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay and J. J. Collins, A Deep Learning Approach to Antibiotic Discovery, *Cell*, 2020, **180**(4), 688–702.e13, DOI: 10.1016/j.cell.2020.01.021.

32 K. Li, B. DeCost, K. Choudhary, M. Greenwood and J. Hattrick-Simpers, A Critical Examination of Robustness and Generalizability of Machine Learning Prediction of Materials Properties, *npj Comput. Mater.*, 2023, **9**(1), 55, DOI: 10.1038/s41524-023-01012-9.

33 D. Kobak and P. Berens, The Art of Using T-SNE for Single-Cell Transcriptomics, *Nat. Commun.*, 2019, **10**(1), 5416, DOI: 10.1038/s41467-019-13056-x.

34 D. Kobak and G. C. Linderman, Initialization Is Critical for Preserving Global Data Structure in Both T-SNE and UMAP, *Nat. Biotechnol.*, 2021, **39**(2), 156–157, DOI: 10.1038/s41587-020-00809-z.

35 G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley Longman Publishing Co., Inc., USA, 1989.

36 D. Bajusz, A. Rácz and K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity

calculations?, *J. Cheminf.*, 2015, **7**, 20, DOI: 10.1186/s13321-015-0069-3.

37 Molecules, Buy Research Compounds|eMolecules, https://www.emolecules.com, (accessed 2025-08-11).

38 G. Zhang, J. Chang, L. Wang, J. Li, C. Wang, R. Wang, G. Shi, K. Yu, W. Huang, H. Zheng, T. Wu, Y. Deng and J. Lu, A Monofluoride Ether-Based Electrolyte Solution for Fast-Charging and Low-Temperature Non-Aqueous Lithium Metal Batteries, *Nat. Commun.*, 2023, **14**(1), 1081, DOI: 10.1038/s41467-023-36793-6.

39 P. Ma, R. Kumar, M. C. Vu, K.-H. Wang, P. Mirmira and C. V. Amanchukwu, Fluorination Promotes Lithium Salt Dissolution in Borate Esters for Lithium Metal Batteries, *J. Mater. Chem. A*, 2024, **12**(4), 2479–2490, DOI: 10.1039/D3TA06228G.

40 C. V. Amanchukwu, Z. Yu, X. Kong, J. Qin, Y. Cui and Z. Bao, A New Class of Ionically Conducting Fluorinated Ether Electrolytes with High Electrochemical Stability, *J. Am. Chem. Soc.*, 2020, **142**(16), 7393–7403, DOI: 10.1021/jacs.9b11056.

41 R. Li, H. Zhang, S. Zhang, Y. Li, R. Guo, H. Pei, M. Yang, J. Zhang, L. Chen, X. Xiao, L. Chen, Y. Shen, T. Deng and X. Fan, Unified Affinity Paradigm for the Rational Design of High-Efficiency Lithium Metal Electrolytes, *Nat. Energy*, 2025, **10**(9), 1155–1165, DOI: 10.1038/s41560-025-01842-5.

42 B. D. Adams, J. Zheng, X. Ren, W. Xu and J.-G. Zhang, Accurate Determination of Coulombic Efficiency for Lithium Metal Anodes and Lithium Metal Batteries, *Adv. Energy Mater.*, 2018, **8**(7), 1702097, DOI: 10.1002/aenm.201702097.

43 S. Wilson Lucas, R. Zijian Qin, K. P. Rakesh, K. S. Sharath Kumar and H.-L. Qin, Chemical and Biology of Sulfur Fluoride Exchange (SuFEx) Click Chemistry for Drug Discovery, *Bioorg. Chem.*, 2023, **130**, 106227, DOI: 10.1016/j.bioorg.2022.106227.

44 S. Menkin, J. B. Fritzke, R. Larner, C. de Leeuw, Y. Choi, A. B. Gunnarsdóttir and C. P. Grey, Insights into Soft Short Circuit-Based Degradation of Lithium Metal Batteries, *Faraday Discuss.*, 2024, **248**, 277–297, DOI: 10.1039/D3FD00101F.

45 L. Suo, Y.-S. Hu, H. Li, M. Armand and L. Chen, A New Class of Solvent-in-Salt Electrolyte for High-Energy Rechargeable Metallic Lithium Batteries, *Nat. Commun.*, 2013, **4**(1), 1481, DOI: 10.1038/ncomms2513.

46 X. Ren, P. Gao, L. Zou, S. Jiao, X. Cao, X. Zhang, H. Jia, M. H. Engelhard, B. E. Matthews, H. Wu, H. Lee, C. Niu, C. Wang, B. W. Arey, J. Xiao, J. Liu, J.-G. Zhang and W. Xu, Role of Inner Solvation Sheath within Salt–Solvent Complexes in Tailoring Electrode/Electrolyte Interphases for Lithium Metal Batteries, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**(46), 28603–28613, DOI: 10.1073/pnas.2010852117.

47 F. Strieth-Kalthoff, H. Hao, V. Rathore, J. Derasp, T. Gaudin, N. H. Angello, M. Seifrid, E. Trushina, M. Guy, J. Liu, X. Tang, M. Mamada, W. Wang, T. Tsagaantsooj, C. Lavigne, R. Pollice, T. C. Wu, K. Hotta, L. Bodo, S. Li, M. Haddadnia, A. Wołos, R. Roszak, C. T. Ser, C. Bozal-Ginesta, R. J. Hickman, J. Vestfrid, A. Aguilar-Granda, E. L. Klimareva, R. C. Sigerson, W. Hou, D. Gahler, S. Lach, A. Warzybok, O. Borodin, S. Rohrbach, B. Sanchez-Lengeling, C. Adachi, B. A. Grzybowski, L. Cronin, J. E. Hein, M. D. Burke and A. Aspuru-Guzik, Delocalized, Asynchronous, Closed-Loop Discovery of Organic Laser Emitters, *Science*, 2024, **384**(6697), eadk9227, DOI: 10.1126/science.adk9227.

48 V. Sharma, M. Giammona, D. Zubarev, A. Tek, K. Nugyuen, L. Sundberg, D. Congiu and Y.-H. La, Formulation Graphs for Mapping Structure-Composition of Battery Electrolytes to Device Performance, *J. Chem. Inf. Model.*, 2023, **63**(22), 6998–7010, DOI: 10.1021/acs.jcim.3c01030.

49 H. Zhang, T. Lai, J. Chen, A. Manthiram, J. M. Rondinelli and W. Chen, Learning Molecular Mixture Property Using Chemistry-Aware Graph Neural Network, *PRX Energy*, 2024, **3**(2), 023006, DOI: 10.1103/PRXEnergy.3.023006.

50 E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris and T. E. Ferrin, UCSF ChimeraX: Tools for Structure Building and Analysis, *Protein Sci.*, 2023, **32**(11), e4792, DOI: 10.1002/pro.4792.

51 X. Ren, P. Gao, L. Zou, S. Jiao, X. Cao, X. Zhang, H. Jia, M. H. Engelhard, B. E. Matthews, H. Wu, H. Lee, C. Niu, C. Wang, B. W. Arey, J. Xiao, J. Liu, J.-G. Zhang and W. Xu, Role of Inner Solvation Sheath within Salt–Solvent Complexes in Tailoring Electrode/Electrolyte Interphases for Lithium Metal Batteries, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**(46), 28603–28613, DOI: 10.1073/pnas.2010852117.

52 R. Kumar, Using Electrolyte Solvent Embeddings to Guide Battery Electrolyte Discovery, *Zenodo*, 2025, preprint, DOI: 10.5281/zenodo.17330302.