Supplementary Information
Using electrolyte solvent embeddings to guide battery electrolyte discovery

Ritesh Kumar[1, †], **Ke-Hsin Wang**[1, †], **and Chibueze V. Amanchukwu**[1, *]

[1]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA
[†]Equal contribution
[*]Corresponding author: Chibueze V. Amanchukwu (chibueze@uchicago.edu)

Supplementary Notes

Supplementary Note 1: Enumeration of all possible acyclic ethers

To achieve enumeration of all possible ether, we start with dimethyl ether (simplest known ether) and then new compounds are sequentially generated in three different ways: add methyl group to increase alkyl chain length, add oxygen to increase number of ether fragments (ensure there is no peroxy (unstable) or hydroxyl (unsuitable for NGBs) bonds), and replace hydrogens on alkyl chain with fluorine atoms. This will lead to a growth of tree of ether compounds, for example, performing this sequential growth from dimethyl ether leads to nine unique compounds as shown in Figure S1. Continuing this for a total of three iterations only leads to a total of 2500 compounds with the constraint that the total number of carbon and oxygen atoms in the molecule should not exceed six (even while excluding the cyclic compounds). The code for enumerating the solvents molecules is available on the GitHub repository.

Algorithm 1 Sequential Molecule Generation Algorithm

Input: Initial molecule SMILES_0 (e.g., 'COC'), maximum total atoms $N_{C,O}$ (sum of carbon and oxygen atoms)
Initialize the parent molecule set with SMILES_0
Initialize an empty set for storing all generated molecules
while maximum number of carbon and oxygen atoms in generated molecule $\leq N_{C,O}$ **do**
    For each molecule in the parent set:
        i. **Ether Growth:** Add one oxygen atom (O)
        ii. **Alkyl Growth:** Add one carbon atom (C)
        iii. **Fluorination:**
            a. Identify all hydrogen atoms attached to carbon atoms
            b. Generate all unique molecules by substituting hydrogens with fluorines in all possible combinations
    Filter generated molecules for:
        i. Uniqueness (avoid duplicates)
        ii. Validity (respect valence and SMILES rules)
        iii. No O-O or O-H bonds
    Add valid molecules to the set of generated molecules
    Update parent set for the next iteration with the new molecules
end while
Output: Set of all unique generated molecules.

Supplementary Note 2: Virtual search space used for exploring promising electrolytes

Creating search space: The unlabeled chemical space to identify potential candidates for next-generation rechargeable lithium battery electrolytes were *partially* taken from Kumar *et al.*[1] The original chemical space consisted of molecules from two popular chemical repositories built for drug discovery applications — eMolecules and PubChem. The eMolecules database contains entries of molecules (total ~24 million, accessed on 27th May 2022) that have been experimentally synthesized and most of them are also available commercially. The PubChem repository contains information regarding physicochemical properties of molecules that have been reported in the literature (total ~110 million, accessed on 6th February 2023). However, since these datasets have primarily been curated for drug-discovery purposes, many of them may not be directly relevant for the domain of Li battery electrolytes. To refine the databases for our specific application, we systematically excluded molecules containing functional groups known to be incompatible with lithium battery chemistry, such as transition metals, alkenes, alkynes, and hydroxyl groups (all such undesired moieties listed in Table S1), resulting in a total of ~175k compounds. Ultimately, ML models for ionic conductivity (Chemprop), oxidative stability (Chemprop), and Coulombic efficiency (CE) (partial least squares regression; PLSR) from Kumar *et al.*[1] were used for making prediction of three properties on the remaining dataset. For the task of oxidative stability prediction, only the filtered SMILES from these repositories were input into the ML model. For tasks predicting ionic conductivity and CE, we provided additional input parameters including the choice of salt, concentration, temperature, current density, and cycling protocol, set to be LiFSA (canonical SMILES: O=S(=O)(F)[N-]S(=O)(=O)F.[Li+]), 1 M for ionic conductivity and 2 M for CE, 25°C, 0.5 mA cm^{-2}, and Modified Aurbach cycling protocol, respectively. Finally all three models were deployed on the pool of ~175k molecules and then filtered further based on a metric called e*Score* as defined in Kumar *et al*.[1] (compounds with non-zero e*Scores* were removed). These refined selections (total 59851 out of ~134 million initial compounds) are regarded as potential solvent search space for our study.

Supplementary Note 3: Reduced embedding techniques used in the present work

1. Principal component analysis (PCA): PCA is a linear embedding method. It projects the high-dimensional data onto lower-dimensional linear subspaces (called principal components) that capture the maximum variance. The first few principal components capture the directions of maximum variance and the *i*-th principal component is taken as a direction orthogonal to the first (*i*-1) principal components. Therefore, it aims to preserve the total variance in the data as much as possible. Global structures typically contribute more to the total variance than local structures. The linear approach utilized in the PCA makes it effective at identifying broad trends and overarching patterns (global structures) that contribute most to the data variance. Hence, PCA is more inclined to capture these large-scale patterns.

2. t-distributed Stochastic Neighbor Embedding (t-SNE): Compared to PCA, t-SNE is a non-linear embedding technique that has gained traction recently for reduced visualization of high-dimensional data, such as the molecular structures in the chemical space [38]. As a first step, the high-dimensional distances between data points are converted into conditional probabilities that represent similarities. The probability of similarity is higher for points closer to each other, thus forming local neighborhoods. Then the data points are mapped to a lower-dimensional space and computes a similar set of probabilities in this space. The final objective is to make the probabilities in the lower-dimensional space as close as possible to those in the high-dimensional space. For this purpose, Kullback-Leibler (KL) divergence between the two probability distributions (a measure of how one probability distribution is different from a second one) is minimized using a gradient descent optimization. While a Gaussian distribution is used in the original space, a Student's t-distribution (similar to a Gaussian distribution but has heavier tails) is applied in the reduced space to calculate probabilities. This heavy-tailed distribution allows a natural separation of clusters. Since t-SNE is particularly adept at preserving local neighborhoods in the reduced embeddings, it provides a much better picture of local structures such as clusters or groups in the data that might not be apparent in the original high-dimensional space, as compared to the PCA.

3. Uniform Manifold Approximation and Projection (UMAP): Similar to t-SNE, UMAP is also a non-linear embedding method. However, unlike t-SNE, the UMAP works directly with unnormalized similarities rather than normalized probabilities. The other major differences from t-SNE being the use of cross-entropy as a loss function rather than KL divergence and using stochastic gradient descent to minimize this loss function instead of regular gradient descent. The cross-entropy loss function has both attractive and repulsive forces, while KL divergence has only attractive forces and the repulsive forces only appear during calculation of the similarity matrix. The optimization process in UMAP starts by assuming a uniform distribution over the dataset, which helps in maintaining the global structure. It then refines the layout to represent finer local structures, whereas t-SNE focuses on local interactions from the start.

In summary, PCA, unlike non-linear methods such as t-SNE and UMAP, does not prioritize preserving the local neighborhood structures within the dataset. Its primary objective is to capture the overall variance and distribution, which can lead to a loss of nuanced, local geometries. While PCA's linear nature allows it to handle large datasets efficiently and with less computational demand than its non-linear counterparts, this efficiency may result in an oversimplification of intricate local patterns. Consequently, for a more detailed exploration of local and mesoscopic structures, t-SNE and UMAP are often preferred methods, as evident by their widespread adoption in the cheminformatics, cell transcriptomics, and materials informatics domains [37,38,39]. To determine the most faithful representation of the chemical space in reduced dimensions for our analysis, we compared embeddings derived from PCA, t-SNE, and UMAP as shown in Figure 1 in the main text.

Supplementary Note 4: Metrics for quantifying faithfulness of reduced embeddings, distance metric used for selecting compounds, and Tanimoto coefficient

Our assessment of the resulting embeddings' accuracy was based on three metrics: KNN, KNC, and CPD, which were introduced and detailed by Kobak *et al.*[2,3] , as described below:

1. KNN: This metric evaluates how many of the *k*-nearest neighbors (KNN) in the original multi-dimensional space remain closest in the reduced-dimensional representation. It serves as a measure of how well the immediate, intricate structures are maintained. Often also termed as the neighborhood preservation ratio (NPR), this metric is extensively adopted in the cell transcriptomics field [41]. However, optimizing solely on this metric can yield detailed local depictions while possibly distorting broader or intermediary structural representations due to inherent trade-offs. In alignment with the methodology of Kobak *et al.*, we set *k* to be 10 (ten nearest neighbors) for our analysis.

2. KNC: This measure captures the proportion of *k*-nearest class means in the original dataset that stay closer in the reduced space and is evaluated exclusively for class means and then averaged across classes. This metric is instrumental in assessing the retention of intermediate-scale structures within the data. This requires assigning classes to each of the data points. For our dataset, we chose functional groups present in the molecules (assigned according to protocol discussed in Supplementary Note 7) to assign classes. Since a molecule may contain multiple functional groups, therefore, we assigned a unique class ID to each of the unique combination of functional groups present in the molecules, e.g., a molecule with only ether present will be assigned a different class compared to a molecule with both ether and nitrile present, or a molecule with ether, amine, and ester present. In total, we found 155 unique classes in the virtual search space. We adopted *k*=4, in line with Kobak *et al.*

3. CPD: The Spearman correlation of pairwise distances (CPD) can be used to compare the original space with the reduced space, over all possible pairs among a subset of 2500 points selected at random (higher than that used by Kobak *et al.*). It effectively measures the extent to which the global, macroscopic structures are preserved in the reduced-dimensional space.

The code for calculating the KNN, KNC, and CPD metrics was adopted from Kobak *et al.*[2,3] and is available on the GitHub repository.

Normalized cosine distance (d_{cos}^{norm}) in embedding spaces: It is defined according to the following expression:

$$d_{cos}^{norm} = {}^{d_{cos}}\big/_{d^{max}} \tag{S1}$$

where d_{cos} and d^{max} are cosine distance between given pair of molecules on the embedding space and maximum of Euclidean distance between any pair of molecules (found out by using 'pdist' function available in the scipy python library). The d_{cos} for a given pair of vectors is defined according to the following expression:

$$d_{cos} = 1 - \frac{u \cdot v}{||u||_2 ||v||_2} \tag{S2}$$

where $||\ ||_2$ refers to $\mathcal{L}_2$ norm of the given vector. In this case, u and v refers to the reduced embedding coordinates on the 2D t-SNE space. The lower the values of d_{cos} are, the similar the molecules are on the embedding space.

Tanimoto coefficient (TC): It is given by the following expression:

$$TC = \frac{N_{comm}}{N_{comm} + N_{diff}} \tag{S3}$$

where N_{comm} and N_{diff} denote the number of same fingerprint bits on both molecules and the number of different fingerprint bits on both molecules, where the bits 0 and 1 denote the absence and presence of a molecular substructure. The Morgan fingerprint with 2048 bits and radius = 2 were employed in evaluating TC (also called ECFP4 fingerprint).

Supplementary Note 5: Choosing promising electrolytes for experimental validation

Out of the existing list of 59851 unlabeled molecules in the search space, first commercially purchasable compounds were selected from emolecules.com, resulting in 3880 molecules. However, since even the space of ~3000 compounds is practically infeasible to be tested experimentally (in terms of both cost and time) for compatibility with LMBs, we further down selected candidates based on their proximity with respect to the SOTA electrolytes on the reduced embedding space (exploration and exploitation strategies). The primary goal in selecting candidate solvent molecules was to capture the broadest possible diversity within our choices. To achieve this, we focused on purchasable molecules from the eMolecules database, prioritizing those with a reasonable cost (approximately $300 as the upper limit) and available within a reasonable time (<6 weeks).

Supplementary Note 6: DFT analysis of electrolyte solvent molecules

Solvent-Li$^+$ binding energies ($\Delta E_{solv-Li^+}$): It was calculated according to the following expression:

$$\Delta E_{solv-Li^+} = E_{solv-Li^+} - E_{solv} - E_{Li}^{+} \quad \text{(S4)}$$

where $E_{solv-Li^+}$, E_{solv}, E_{Li}^{+} and are DFT-optimized energies of solvent-lithium-ion complexes, isolated solvent molecule, and the lithium ion, respectively, at the same level of theory. It cannot explain 1 M/2 M salt solubility (Figure S5a) or the trends in CE for LiFSA-based single solvent electrolytes (Figure S12a).

ESP: The ESP isosurfaces are obtained from the electronic density isosurfaces in DFT. The minimum and maximum points on the ESP surface denote the maximum electronic donation and acceptance capability of the molecule. However, even ESP_{solv}^{min} and ESP_{solv}^{max} values could not explain the trends in 1 M/2 M salt solubilities (Figure S5b).

Solvent-cation affinity (α_s): Li *et al.*[4] recently proposed a descriptor called solvent-cation affinity to predict whether a given organic molecule can dissolve salts or not. It is calculated according to the following expression:

$$\alpha_s = \int\int MEP_{min} dS_{min} \times S_{min}^{\frac{1}{2}} \times V_{tot}^{-\frac{1}{3}}$$

where MEP is the molecular electrostatic potential (obtained from ESP calculations) in the van der Waals surfaces, S_{min} is atomic-projected effective areas and V_{tot} is the volume. According to them, if the α_s value is below the anion-cation affinity of the salt (α_{salt}, e.g., LiFSA), then the solvent will have lower affinity to bind Li+ cation compared to the salt anion and hence should not dissolve the salt. We also calculated α_s (Figure S5c), however, we did not find it to explain the trends in 1 M/2 M salt solubility for all 15 solvents. We have provided the code for calculating solvent-cation affinity in the GitHub repository.

LUMO: Molecules with lower LUMO energy levels tend to accept molecules easily during reaction and can be reduced easily and hence is considered a proxy for reductive stability of molecules. LUMO can also not explain the trends in CE for LiFSA-based single solvent electrolytes (Figure S12b).

Supplementary Note 7: Effect of EMP-based electrolyte formulation

In this work, the champion single-solvent, single-salt electrolyte (2 M LiFSA in EMP) was further formulated into a high-concentration electrolyte (HCE) by increasing the salt concentration to 4 M, and into a localized high-concentration electrolyte (LHCE) by fixing a 1:3 molar ratio of salt to diluent and adding EMP until no salt precipitation was observed. After electrolyte formulation, half cell CE only shows slight improvement after formulation. One possible explanation is that 2 M LiFSA in EMP already operates near its optimal solvation regime, leaving limited room for further enhancement. Raman spectroscopy (Figure S16) reveals a slight red shift of the S–N–S stretching mode of LiFSA as the salt concentration increases from 2 M to 4 M, with a larger shift observed in the LHCE. However, the total shift is at most 1.4 cm^{-1}. Moreover, the center of mass (Cm) of the Raman peaks for all electrolytes falls within a range of 744 cm^{-1} to 760 cm^{-1}, indicating that aggregate-dominated solvation structures persist across all formulations and reflect their weakly solvating nature[5]. Together, these results suggest that significantly altering the solvation structure of 2 M LiFSA in EMP through formulation alone is inherently challenging.

Supplementary Figures

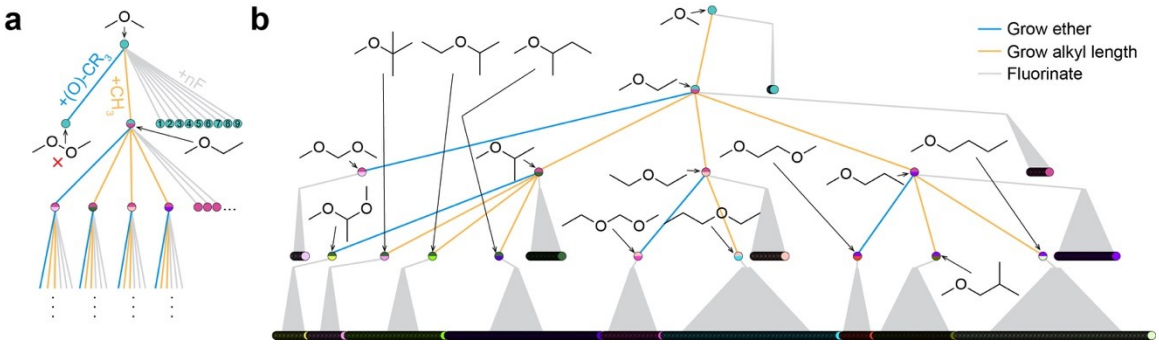

Figure S1: a Illustration of systematic ether molecule modification to all acyclic ethers through three pathways — adding oxygen atoms to extend ether chains (while avoiding unstable peroxides, indicated with a cross mark), elongating alkyl chains by adding methyl groups, or introducing fluorine atoms to create unique fluorinated ethers. **b** Tree visualization of sequentially designed non-fluorinated (structures shown) and fluorinated (structures not shown for brevity) ethers, starting from dimethyl ether. Molecules are generated under the constraint that the total number of carbon and oxygen atoms does not exceed six. Each node represents a unique molecule, and edges depict the relationships between parent and offspring molecules via different growth pathways: ether growth (blue), alkyl chain extension (orange), or fluorination (gray). Nodes share the same color as their parent, and bi-colored semicircles indicate molecules that act as both parents and offspring.

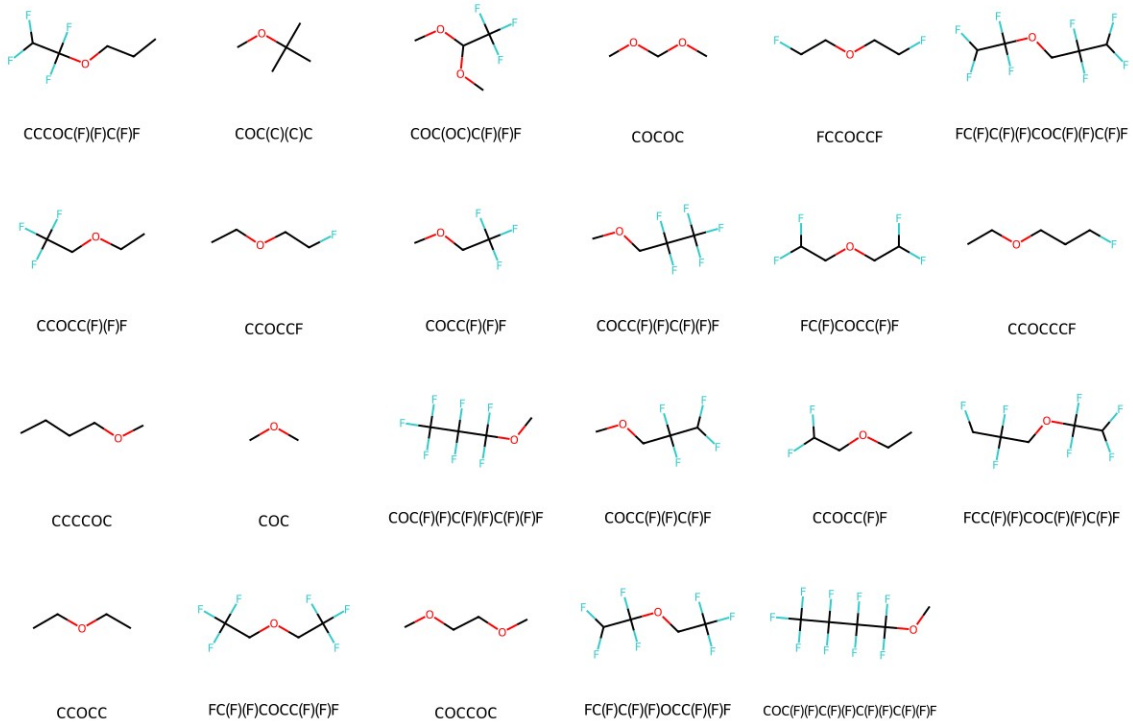

Figure S2: Molecular visualization of solvent molecules investigated for LMB applications so far, along with the corresponding SMILES representation below.

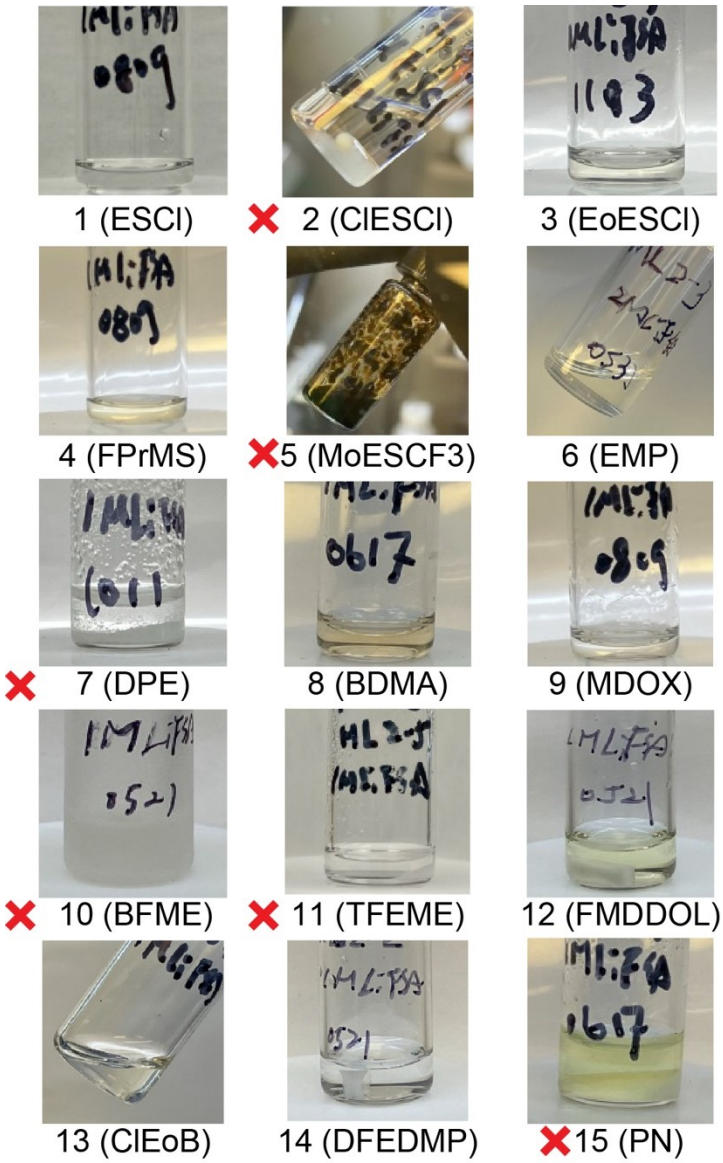

Figure S3: Optical images of LiFSA salt solubility tests with all the 15 solvent molecules selected in the present study.

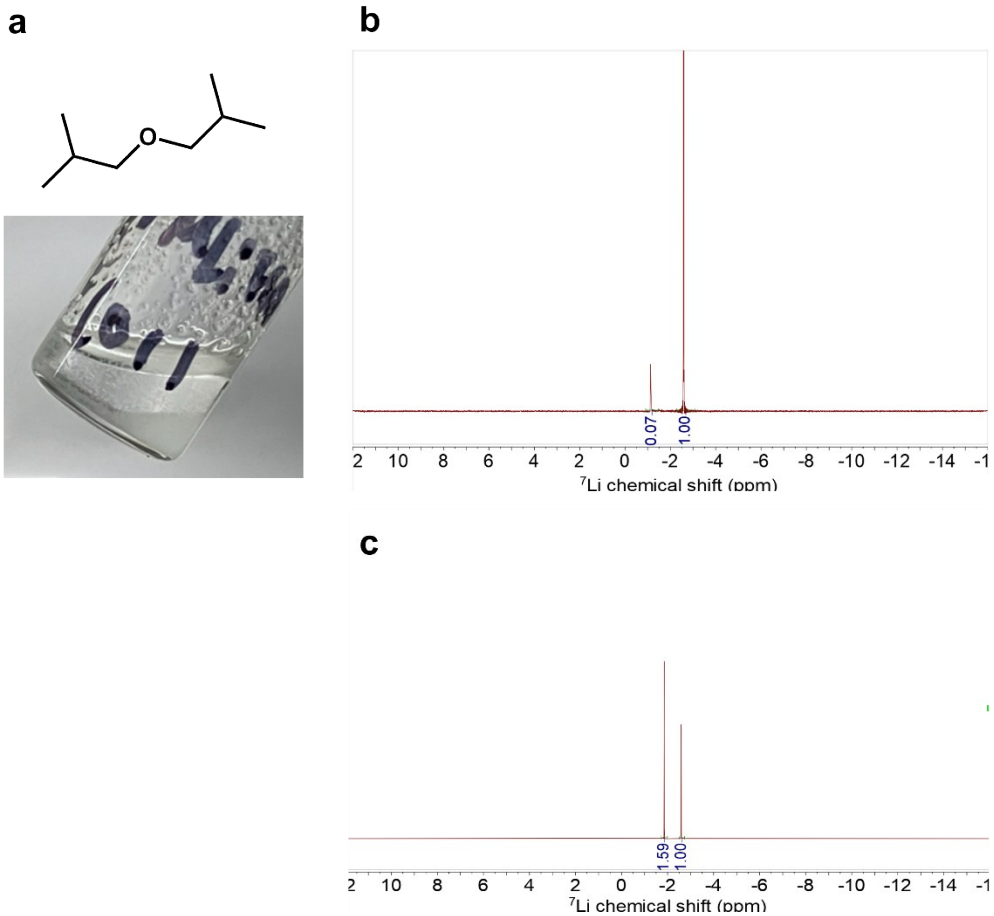

Figure S4: Solubility of DPE. a Image of 1 M LiFSA in DPE. Salt precipitates are visible at the bottom of the vial. **b** ^{7}Li NMR of the supernatant in Figure a using capillary NMR setup. 0.1 M $LiClO_4$ in d_3-ACN was used as the reference at –2.80 ppm. **c** ^{7}Li NMR of 1 M LiFSA in DME (fully dissolved). Solubility of DPE is determined by calculation shown below:

$$Solubility\ (M) = 1\ M\ \times \frac{integral\ of\ 7Li\ peak\ of\ the\ sample}{integral\ of\ 7Li\ peak\ of\ 1\ M\ LiFSA\ in\ DME} = \frac{0.07}{1.59} = 0.04\ M$$

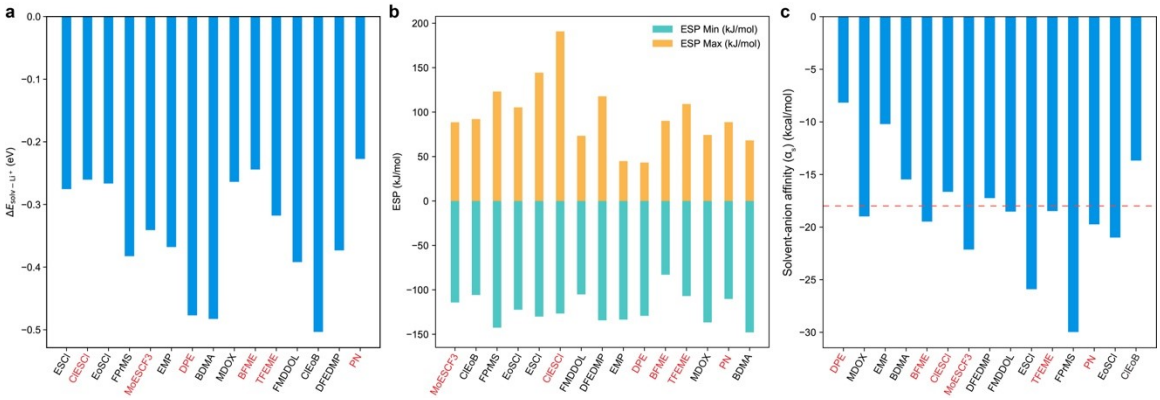

Figure S5: a Solvent-Li$^+$ binding energies ($^{\Delta E}_{solv-Li^+}$), **b** ESP^{min}_{solv} and ESP^{max}_{solv} values for all 15 solvent candidates investigated in the present study, and **c** solvent-cation affinity ($^{\alpha_s}$) values for all 15 solvent candidates investigated in the present study. The dashed line in **c** represents the anion-cation affinity for the LiFSA salt. The names of the solvents that did not pass the 1 M/2 M LiFSA salt solubility test are marked in red font.

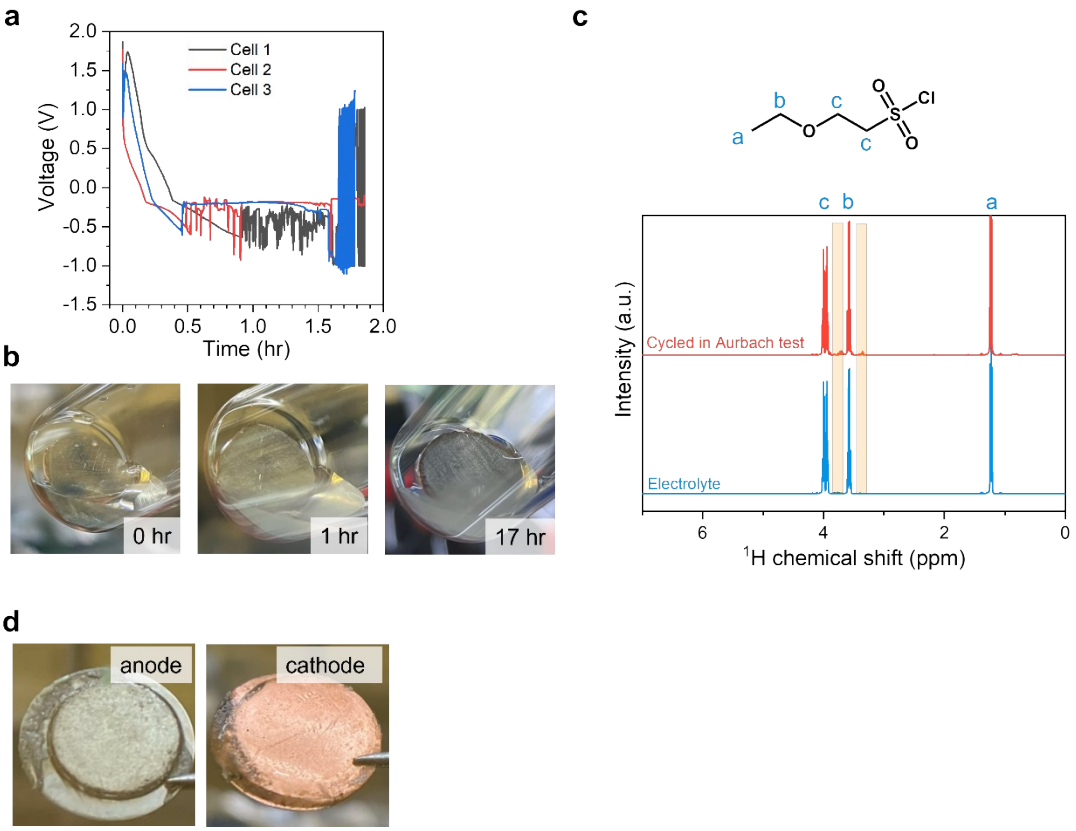

Figure S6: Cycling profile and degradation of EoESCl. a Voltage profiles of Li||Cu cells cycled with Aubach method using 1 M LiFSA in EoEsCl. Three cells using the same electrolyte are shown. **b** Digital images of 20 µm Li immersed in 1 M LiFSA in EoESCl at 0, 1, and 17 hours. Insoluble particles are visible after 1 hour immersion. **c** ^{1}H NMR of electrolyte extracted from Li||Cu coin cells after cycling with Aurbach method. Orange boxes indicate the positions of additional peaks after cycling. $CDCl_3$ was used as the deuterated solvent. **d** Images of the anode (Li, left) and cathode (Cu, right) after cycled in Aurbach test.

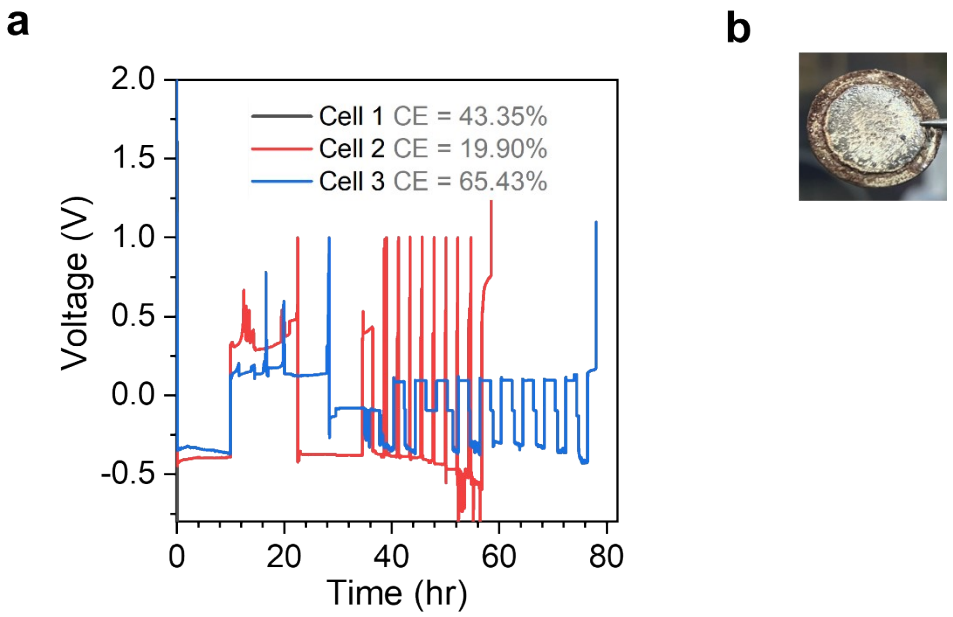

Figure S7: Unstable cycling and low CE of ESCl. a Voltage profile of Li||Cu cells using 1 M LiFSA in ESCl as the electrolyte. High overpotential (> 0.1 V) low CE (all 3 cells < 70%) potentially shows the lack of passivation layer on Li surface. **b** Image of Li anode after cycling.

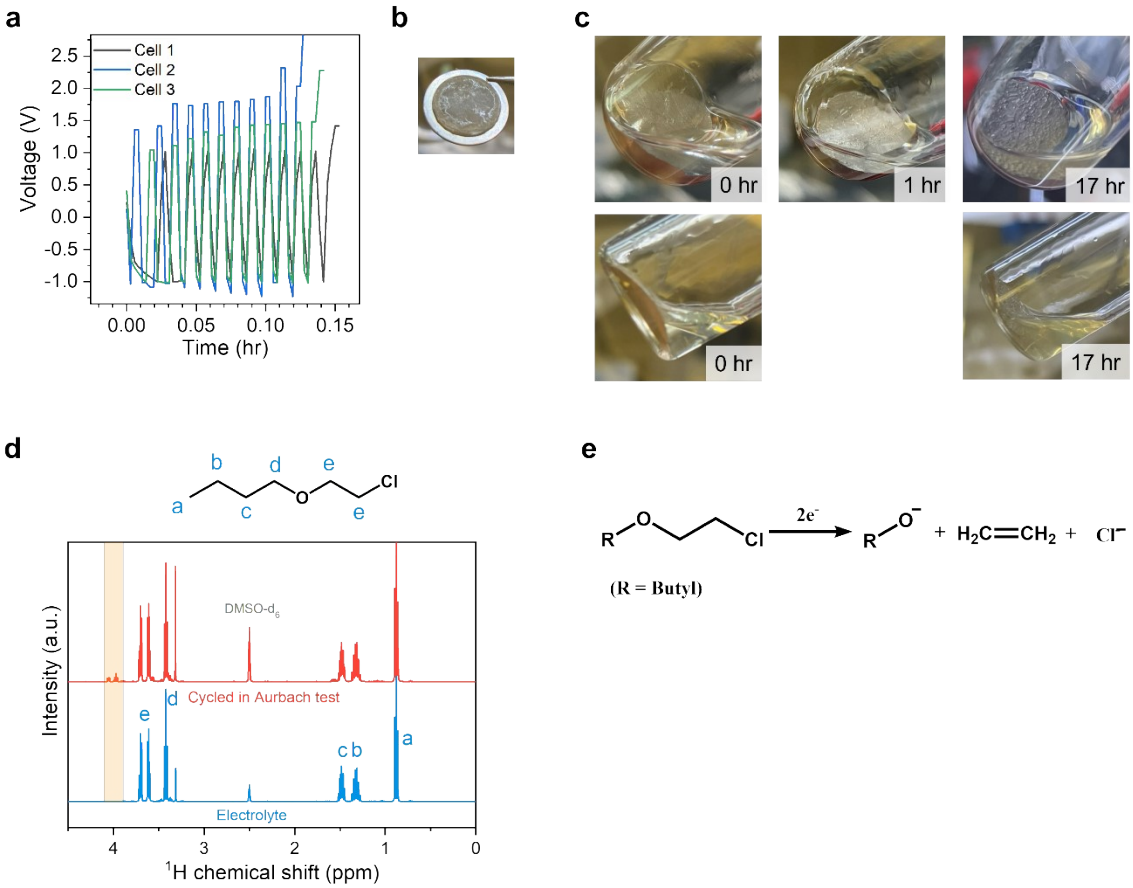

Figure S8: Degradation of ClEoB electrolyte. a Voltage profile of Li||Cu cells cycled with Aurbach method. The voltage cutoffs were immediately reached at each step of the method. Three Cells were shown in the figure. **b** Image of Li anode after cycling with Aurbach method. **c** Images of 20 μm Li immersed in 1 M LiFSA in ClEoB at 0, 1, and 17 hours. Bumps form on the lithium foil after submerging lithium into the electrolytes for 17 hours. The solution turns yellow after 17 hours of Li immersion. **d** ^{1}H NMR of electrolyte extracted from Li||Cu coin cells after Aurbach test. The range box marks the positions of additional peaks after cycling. **e** Possible degradation route of ClEoB. The alkoxide could be responsible for the addition peaks in **d**.

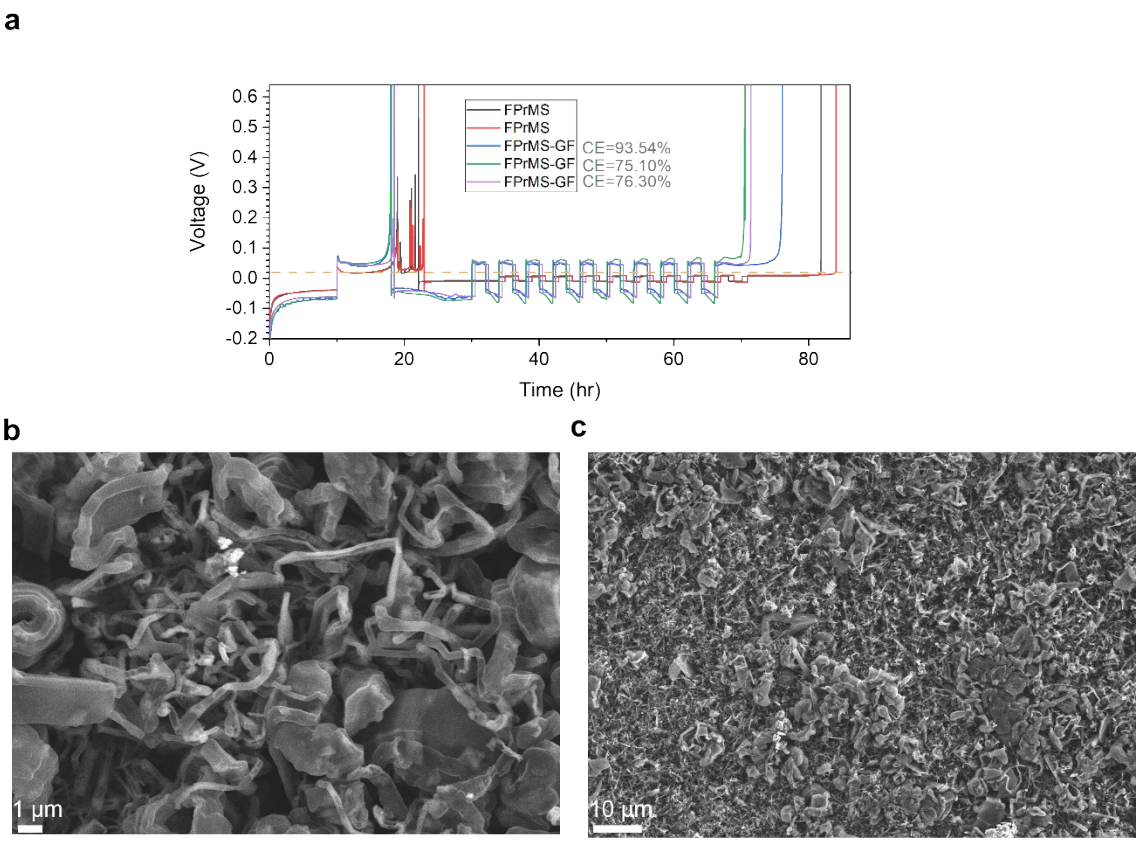

Figure S9: Cycling profile and dendritic Li deposition while using 1 M LiFSA in FPrMS as the electrolyte. a Voltage profile of Li||Cu cells using 1 M LiFSA in FPrMS as the electrolyte. Black and red traces refer to the cells using 1 separator (Celgard 3501). Spikes can be seen in the first stripping step. The cell overpotential drops after the spikes, indicating soft shorting inside the cells. Orange dashed horizontal line marks the overpotential before soft shorting. Blue, green, and purple traces in the same plot refer to the cells using glass fiber (GF) sandwiched by 2 Celgard 3501 as the separators. No soft shoring events were found in this cell configuration. **b** and **c** Scanning electron microscopy (SEM) image of Li deposition at 1.0 mA/cm^2 to 1.5 mA/cm^2.

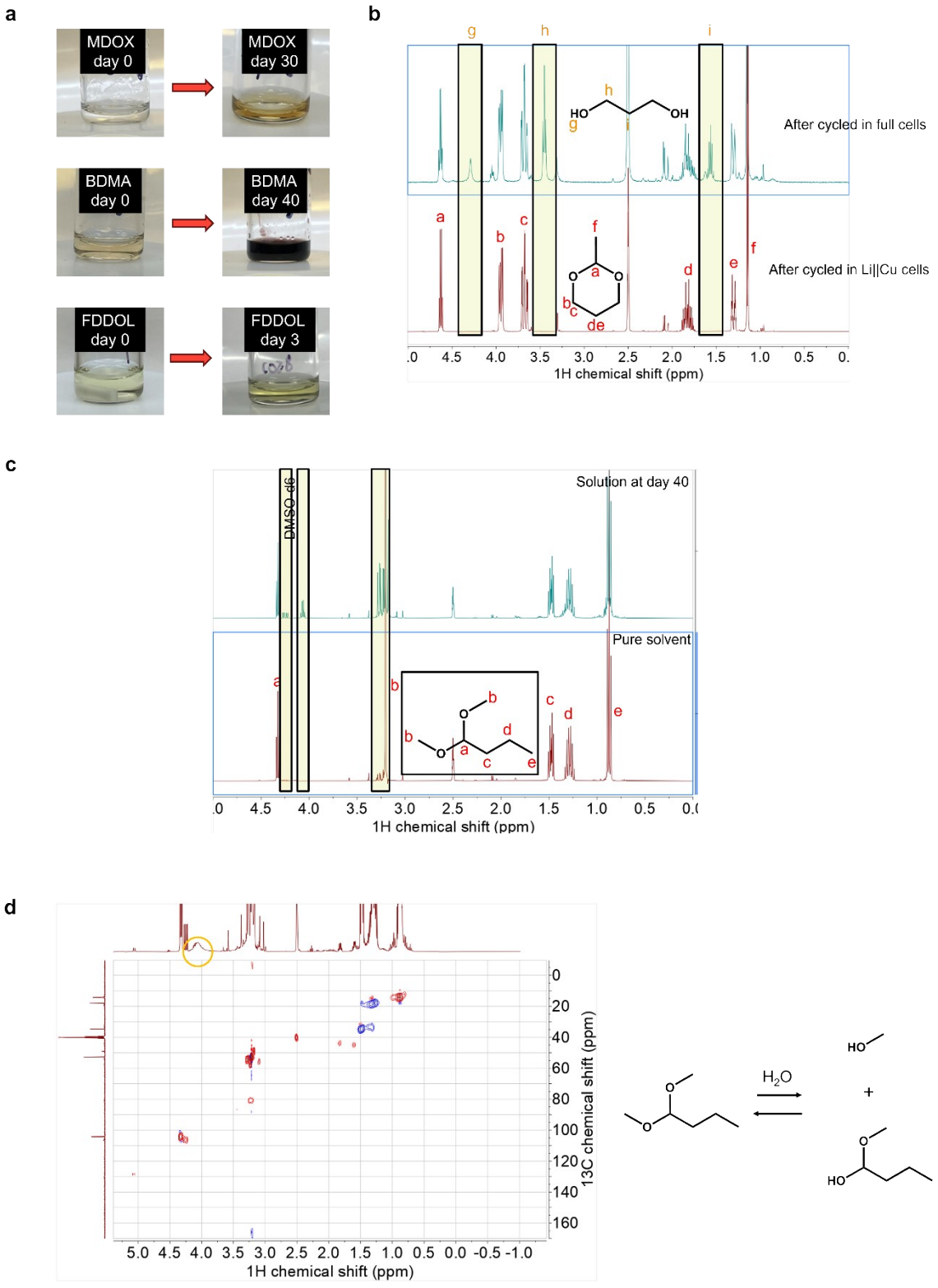

Figure S10: Chemical and electrochemical degradation of acetal-based electrolytes studied in this work. a Images of electrolytes containing 1 M LiFSA in MDOX, BDMA, and FDDOL showing color changes over time. **b** ^{1}H NMR showing electrochemical degradation of MDOX electrolyte which predominantly occurs when the cells were cycled in full cells (Cu||LFP). After

dissolving the cycled electrolyte in DMSO-d_6, 1, 3 propanediol was found to be the major decomposition product. **c** ^{1}H NMR showing chemical degradation of the BDMA electrolyte after the solution was kept in glovebox for 40 days. Orange squares indicate additional peaks after electrolyte aging. The aged electrolyte was dissolved in DMSO-d_6 for NMR measurement. **d** HSQC (Heteronuclear Single Quantum Coherence) spectra of aged BDMA electrolyte. A broad peak which does not correlate with any carbon is found (circled with orange marker). A possible hydrolysis pathway is shown on the right, which indicates the formation of methanol that could be responsible for the broad peak.

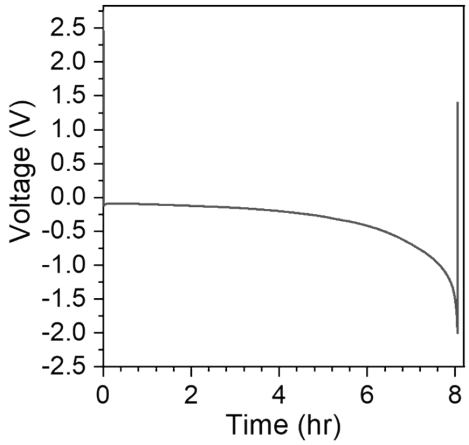

Figure S11: Voltage profile of Li||Cu cell using saturated LiFSA in PN as the electrolyte.

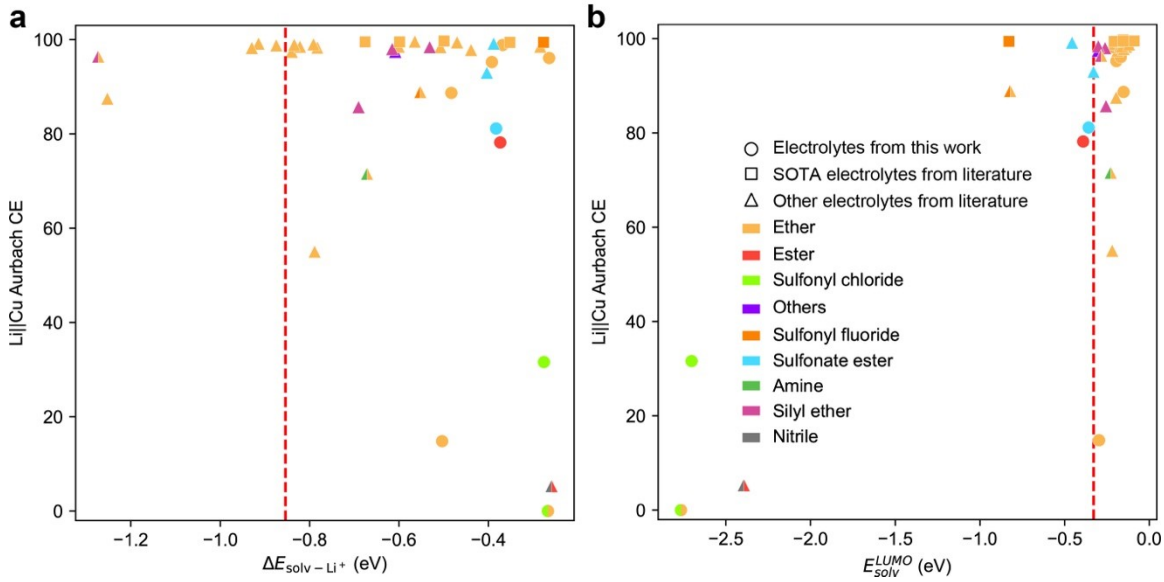

Figure S12: Variation of Li||Cu CE values as a function of **a** solvent-Li^+ binding energies ($\Delta E_{solv-Li^+}$) and **b** lowest unoccupied molecular orbital (LUMO) energy levels (E_{solv}^{LUMO}) (both calculated using DFT) for several LiFSA-based (single solvent) electrolytes reported in the literature and for 15 solvent-based electrolytes investigated in the present study. The dashed lines in a represent $\Delta E_{solv-Li^+}$ between FSA^- and Li^+ and in b represent E_{solv}^{LUMO} for FSA^-.

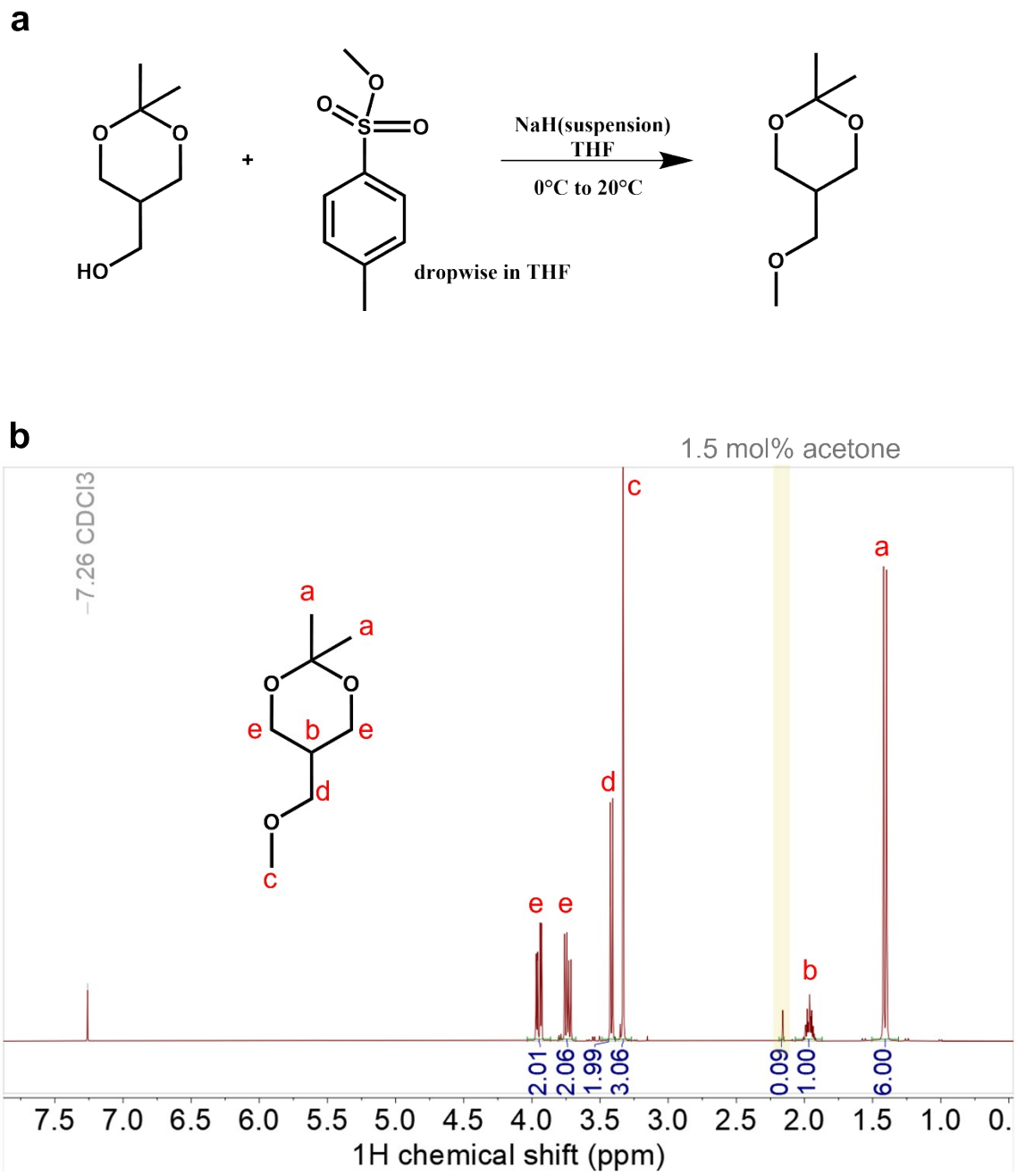

Figure S13: Synthesis of MDOX-M. a Synthetic route of MDOX-M **b** ^{1}H NMR of MDOX-M. $CDCl_3$ is the deuterated solvent. A peak at 2.2 ppm indicates the hydrolysis product, acetone (1.5 mol%), which cannot be removed by distillation.

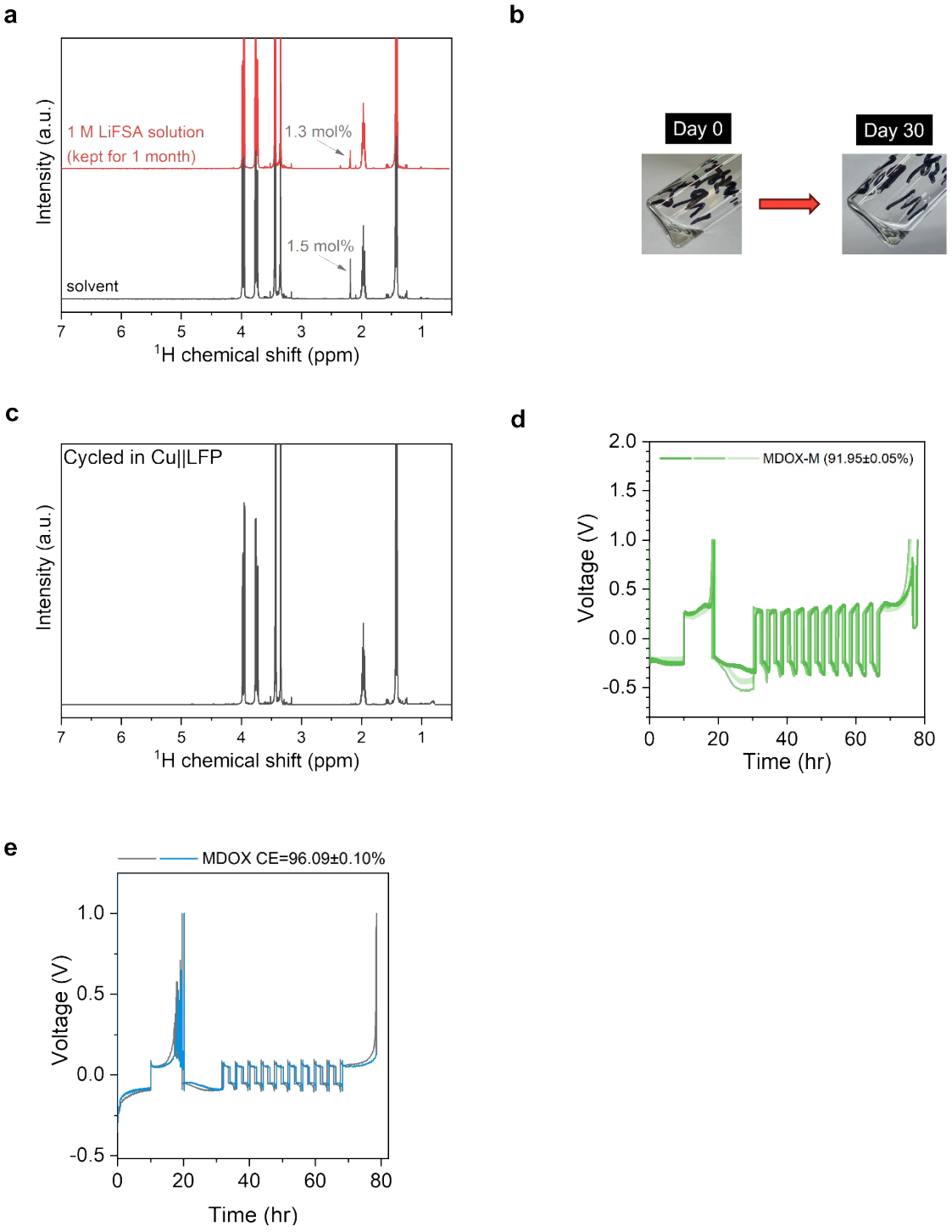

Figure S14: Electrolyte stability and half cell cycling of MDOX-M. a ^{1}H NMR of the electrolyte containing 1 M LiFSA in MDOX-M after being kept in glovebox for 1 month (red trace), which is compared to ^{1}H NMR of pure solvent (grey trace). The amount of impurity remains approximately constant, which are indicated by grey arrows. **b** Image of the electrolyte after being kept in glovebox for 1 month. No visible change in color was found. **c** ^{1}H NMR of the electrolyte extracted from the cell cycled in Cu||LFP configuration. No impurity is found after cycling. **d** and **e** Voltage profile of Li||Cu cells cycled with Aurbach method using 1 M LiFSA in MDOX-M (d) and MDOX (e). While

both chemical and electrochemical degradation are suppressed after substituted MDOX by MDOX-M, other reasons such as low conductivity of MDOX-M electrolyte prevent it from demonstrating reversible Li plating/stripping.

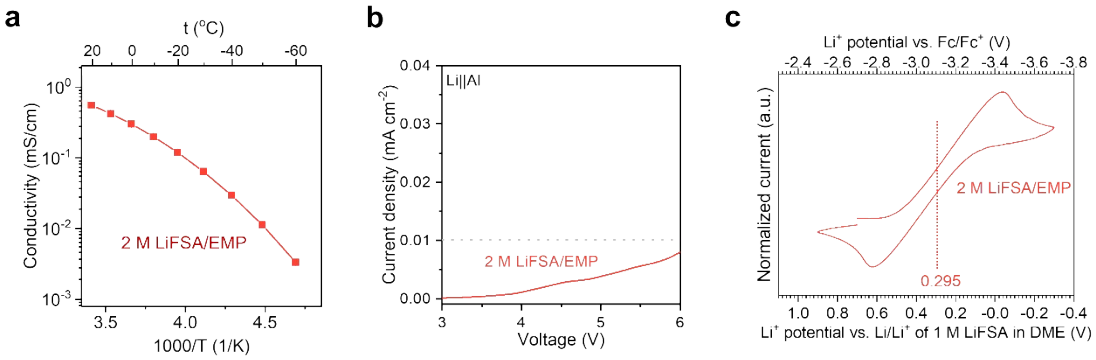

Figure S15: Physicochemical properties of EMP. a Conductivity of 2 M LiFSA/EMP measured in SS||SS cells. The values were obtained by averaging over 3 cells. Error bars are shown in the Figure. **b** Oxidative stability of 2 M LiFSA/EMP measured by linear sweep voltammetry (LSV) of Li||Al cells. A horizontal dashed line indicating 0.01 mA/cm^2 are drawn to guide eyes. **c** Li/Li$^+$ potential measured by Fc/Fc$^+$ test. The Li deposition potential of 2 M LiFSA/EMP is 0.295 V larger than that in 1 M LiFSA/DME, indicating more ion pairs in the solvation structure[6].

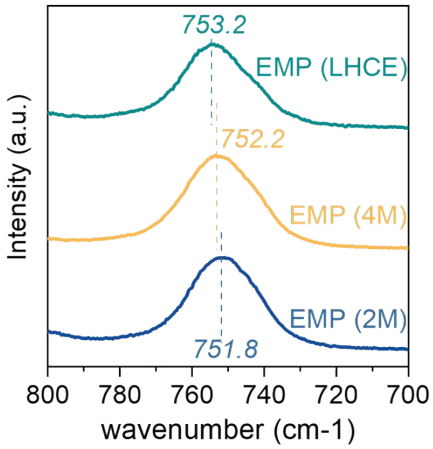

Figure S16 Raman Spectroscopy of EMP-based electrolytes. Only the peak representing the S–N–S stretching mode of LiFSA is shown. The center of mass of each peak is labeled in the same color as the corresponding data curve.

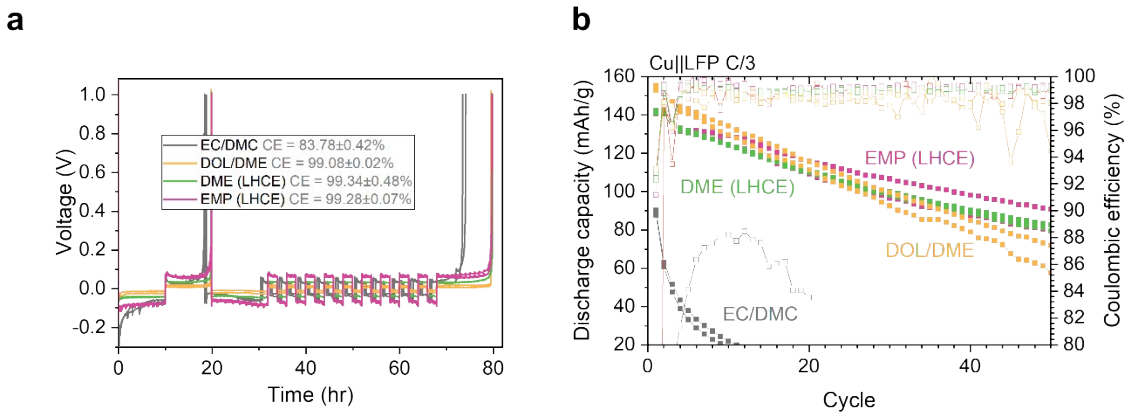

Figure S17: **a** Aurbach Li||Cu tests for EMP- and DME-based LHCEs compared with carbonate- and ether-based mixed-solvent electrolytes. **b** Anode-free cycling using EMP- and DME-based LHCE with comparison to carbonate- and ether-based mixed-solvent electrolytes. Carbonate-based mixed-solvent electrolyte (EC/DME): 1 M $LiPF_6$ in EC/DMC. Ether-based mixed-solvent electrolyte (DOL/DME): 0.5 M LiFSA and 0.5 M LiTFSA in DOL/DME 2v1[7].

Supplementary Tables

Table S1: List of undesirable chemical moieties that were filtered out

Type of undesirable moieties	**Example**
Noble gas elements	He, Ne, Ar, Kr, Xe, Rn
Alkali metals	Li, Na, K, Rb, Cs, Fr
d-block elements	Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd, Lu, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Lr, Rf
f-block elements	La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No
Other elements	Md, Ca, Sr, Ba, Ra, Ga, In, Tl, Ge, Sn, Pb, As, Sb, Bi, Se, Te, Po, Br, I, At
Hydrogen bond donor groups	hydroxyl (–OH), primary amine ($–NH_2$), secondary amine (–NHR), thiol (–SH)
Unsaturated compounds	alkene (C=C), alkyne (C≡C), imine (C≡N)
Aromatic compounds	All aromatic compounds, e.g., benzene derivatives, pyridine, pyrrole, thiophene, etc.
Isotopic compounds	carbon isotope ($^{13}CR_3$), hydrogen isotope ($^{2}H–$)
Charged compounds	cation (+), anion (–)
Enantiomers	chiral stereocenters (*, e.g., D-(+) or L-(–))

Table S2. Summary of experimentally tested electrolyte solvent molecules

ID	Name & abbreviation	SMILES	CAS number	Source	Solubil-ity test with 1 M/2 M LiFSA
1	Ethanesulfonyl chloride (ESCl)	CCS(=O)(=O)Cl	594-44-5	Ambeed	Y
2	2-Chloroethanesulfonyl chloride (ClESCl)	C(CCl)S(=O)(=O)Cl	1622-32-8	Ambeed	N
3	2-Ethoxyethanesulfonyl chloride (EoESCl)	CCOCCS(=O)(=O)Cl	69371-75-1	Ambeed	Y
4	3-Fluoropropyl methanesulfonate (FPrMS)	CS(=O)(=O)OCCCF	372-04-3	Ambeed	Y
5	2-Methoxyethyl trifluoromethanesulfonate (MoESCF3)	COCCOS(=O)(=O)C(F)(F)F	112981-50-7	Astatech Inc.	N
6	1-Ethoxy-2-methylpropane (EMP)[#]	CCOCC(C)C	111-181-1	Astatech Inc.	Y
7	Diisopropyl ether (DPE)	CC(C)OC(C)C	108-20-3	Astatech Inc.	N
8	Butyraldehyde dimethyl acetal (BDMA)	CCCC(OC)OC	4461-87-4	Ambeed	Y
9	2-Methyl-1,3-dioxane (MDOX)	CC1OCCCO1	626-68-6	Ambeed	Y
10	Bis(fluoromethyl)ether (BFME)	C(OCF)F	462-51-1	SynQuest Laboratories	N

11	2,2,2-Trifluoroethyl methyl ether (TFEME)	COCC(F)(F)F	460-43-5	Synquest	N
12	4-(Fluoromethyl)-2,2-dimethyl-1,3-dioxolane (FMDDOL)	CC1(OCC(O1)CF)C	33644-24-5	Combi-Blocks	Y
13	1-(2-Chloroethoxy)butane (ClEoB)	CCCCOCCCl	10503-96-5	Ambeed	Y
14	2,2-Difluoroethyl 2,2-dimethylpropanoate (DFEDMP)	CC(C)(C)C(=O)OCC(F)F	1874200-52-8	Astatech Inc.	Y
15	Propyl nitrite (PN)	CCCON=O	543-67-9	Astatech Inc.	N

Note: # (EMP) forms biphasic solution in 1 M LiFSA but dissolves completely in 2 M LiFSA solution.

References

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

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

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

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

(5) Han, S.-D.; Sommer, R. D.; Boyle, P. D.; Zhou, Z.-B.; Young, V. G.; Borodin, O.; Henderson, W. A. Electrolyte Solvation and Ionic Association: Part IX. Structures and Raman Spectroscopic Characterization of LiFSI Solvates. *J. Electrochem. Soc.* **2022**, *169* (11), 110544. https://doi.org/10.1149/1945-7111/ac9a07.

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

(7) Miao, R.; Yang, J.; Feng, X.; Jia, H.; Wang, J.; Nuli, Y. Novel Dual-Salts Electrolyte Solution for Dendrite-Free Lithium-Metal Based Rechargeable Batteries with High Cycle Reversibility. *J. Power Sources* **2014**, *271*, 291–297. https://doi.org/10.1016/j.jpowsour.2014.08.011.