Active learning accelerates electrolyte solvent screening for anode-free lithium metal batteries

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

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

[#]These authors contributed equally

*Corresponding author

Email: chibueze@uchicago.edu

Supplementary Note 1: Search space and classification based on functional groups

Chemical search space: The chemical search space for optimization of electrolytes for anode-free lithium metal batteries was constructed using unlabeled molecules contained in eMolecules and PubChem databases (shown schematically in Fig. 1c).

1. Molecule filtration: All molecules with undesirable chemical moieties such as nitriles, aromatics, radicals, etc. that are known to be incompatible with lithium metal negative electrodes due to their high reactivity were removed. A list of all such moieties or class of compounds are listed in Table S6. All molecules with molecular weight greater than 500 were also removed, as such large molecules could have high viscosities. Molecules with active hydrogens such as alcohols, primary and secondary amines, carboxylic acids were also augmented such that the active hydrogens are replaced by methyl groups (Table S7). After performing filtration and augmentation, we obtained a total of approximately 4.8 million unique molecules starting from initial ~135 million molecules contained in the eMolecules and PubChem databases.

2. RAscore: Retrosynthetic accessibility score (RAscore)[1] has been postulated to be one of the metrics for determining chemical synthesizability of compounds. It ranges from 0.00 to 1.00, with lower values indicating molecules that are very difficult to synthesize while higher values denote molecules that can be synthesized easily. Therefore, molecules with RAscore lower than 0.01 were removed, resulting in a total of approximately 3.8 million unique solvent molecules.

3. Ionic conductivity prediction: Electrolyte requirements are complex and multiple properties need to be satisfied simultaneously for them to work in batteries. For example, only electrolytes with sufficient ionic conductivity can cycle effectively. Using a deep learning model from Kumar *et al.*,[2] electrolytes with predicted ionic conductivities below 0.8 mS/cm were filtered out. Conductivity predictions were made for all remaining 3.8 million molecules (obtained from step 2) paired with 1 M LiFSA, 1 M $LiPF_6$, and 1 M LiDFOB, and the top one-third of a million electrolytes from each salt combination were selected. This process created a virtual search space of 1 million electrolytes, consisting of 388,013 unique solvent molecules, for optimization via the active learning approach.

However, we did not pre-screen the search space based on melting or boiling points due to the following reasons:

i) **Historical precedent in electrolyte selection** – certain solvents that are solid at room temperature (e.g., ethylene carbonate, EC, melting point ~38°C) have been widely used in lithium-ion batteries,[3] while some low-boiling-point solvents (e.g., dimethyl carbonate, DMC) remain key components of commercial formulations.[4] Additionally, localized high-concentration electrolytes (LHCEs) and liquefied gas electrolytes[5] have demonstrated that conventional temperature-based filtering does not always correlate with electrolyte viability. Thus, overly restrictive pre-screening could lead to the exclusion of potentially promising candidates.

ii) **Computational screening based on ionic conductivity predictions** – to ensure that only viable electrolytes were considered, we employed a deep-learning-based predictive model

(Chemprop) for ionic conductivity from ref [2]. This allowed us to filter out candidates that would be unlikely to support effective ion transport, thereby refining the search space in a more functionally relevant manner.

Classification in terms of functionality: To map the chemical functionality distribution in the pre-defined search space and high-ranking molecules suggested in each batch, RDKit python library was used to assign all common organic functional groups present in the individual molecules (Fig. 2a). Since a large fraction (~57%) of compounds in search space contains more than one functional group, an inclusive classification principle was used to avoid artificial bias, where any compounds containing a specific functional group is counted for the corresponding class. As a result, the fraction of functionality distribution shown in Fig. S6 adds up to over 100%. However, such inclusive classification may overestimate the fraction of ether solvents. Therefore, exclusive for the criterion was used when accounting fraction of ether solvent molecules to gauge the actual fraction of ethers being populated in the chemical space spanned by suggestions and subsequent acquisitions (Fig. 2(c,d)). The code snippets for performing both inclusive and exclusive functional group classification is provided in the GitHub repository.

Supplementary Note 2: Active learning (AL) workflow

The overall AL workflow we used for optimizing anode-free LMB performance is illustrated in Figure 1. It consists of the following steps:

1. Creation of initial labeled data and virtual search space: The process for creation of initial labeled data is described in the main text in the subsection "Design of experiments (DoE) for electrolyte optimization; datasets, and virtual search space" and that for the virtual search space in the Supplementary Note 1. The list of unique solvent molecules comprising the in-house dataset is tabulated in Table S2.

2. Featurization and building black-box surrogate models: The input data consists of solvent and salt identities, composition such as the salt concentration and amount of electrolyte and cycling conditions such as the positive electrode used. The heterogeneous input data was converted into a suitable numerical data as described in Supplementary Note 3. Gaussian process regression (GPR) was chosen as the surrogate model due to its probabilistic nature and ability to accurately yield both epistemic and aleatoric uncertainties, while the frequentist machine learning models such as graph neural networks suffer from inaccurate aleatoric uncertainty prediction[6] and overfitting on typical low data settings such as ours. We also evaluated the performance of common frequentist ML models, including K-nearest neighbor regressor (5-fold CV RMSE: 0.271 ± 0.077), decision trees (0.234 ± 0.059), random forests (0.208 ± 0.041), and LightGBM (0.269 ± 0.063). However, these models underperformed compared to the GPR surrogate models, where the ensemble of four surrogate models achieved a lower 5-fold CV RMSE of 0.159 ± 0.040 on the in-house dataset (Fig. S2a). This discrepancy may be attributed to the data-intensive nature of these frequentist models. Consequently, they were not incorporated as surrogate models in our AL workflow. The surrogate models are trained in the following steps (discussed in detail in the subsequent sections):

i. Selection of kernel functions: We selected four popular and standard prior kernels (or covariance functions; Supplementary Note 5) — i) combination of radial basis function and exp-sine squared kernels, ii) matern-3/2, iii) pairwise, and iv) rational quadratic. Each kernel functions have their own hyperparameters (details in Supplementary Note 5). We also added white noise to all the kernels.

ii. Choosing best hyperparameters for each kernel functions: The hyperparameters such as length scale and noise level were optimized using the negative log likelihood method. The log likelihood (L) is mathematically defined according to the following equation:

$$-log\ L(\theta|D) = -\sum_i^N logP(y_i|x_i, \theta) \quad \text{(S1)}$$

where D is the dataset, θ is the hyperparameter for the surrogate model, N is the number of data points, x_i represents the input features of the ith data point, and y_i is the observed outcome for that data point. The 'likelihood' is a mathematical way to measure how well the model's predictions match the actual data. When the model's predictions are close to the actual data, the likelihood is high, indicating a good match. However, when the predictions are far from the data, the likelihood decreases, showing a weaker match. In this task, we aim to find the parameter values that minimize this negative log likelihood (or maximize likelihood). Essentially, we search for the parameter

values that make the model's predictions most consistent with the observed data. Therefore, according to equation S1, negative log likelihood is calculated by summing up the negative logarithms of the probabilities of the observed data points given the model's predictions. It is carried out by summing up the log of probability distribution function (PDF; normal distribution) at each x_i, shifting by its mean (μ), and scaling by its standard deviation (s.d.; σ).

iii. Train each surrogate model: Each of four surrogate models were trained independently on the labeled data using individually optimized hyperparameters.

3. Obtain predictions on unlabeled data and model weights: The trained models were used to obtain individual predictions (mean) and uncertainties (s.d.) for the unlabeled dataset. Model weights for each data points were obtained using the log likelihood method. The process is described in detail in Supplementary Note 7.

4. Calculate expected improvement (EI) values using aggregated predictions and uncertainties (Bayesian model averaging): The individual predictions and uncertainties were used to obtain aggregated predictions (μ^{aggr}; equation S10), aggregated uncertainties (σ^{aggr}; equation S11), and aggregated EI (EI^{aggr}; equation S9) values using the Bayesian model averaging (BMA) technique (discussed in detail in Supplementary Note 7). The EI^{aggr} values were used to rank electrolytes and the higher EI^{aggr} values were prioritized for experimental battery testing.

5. Feedback and other considerations in acquisition of suggested candidates: It was found that most of the suggested molecules are not commercially purchasable within our budget constraints despite the search space being derived from well-known eMolecules and PubChem molecular repositories and removing difficult to synthesize molecules containing moieties such as radicals. Therefore, purchasable candidates were selected from a pool of molecules with top 5,000 (or even top 10,000 for batches 3 and 4) rank according to EI^{aggr}. It was performed by providing this list of solvent SMILES into the eMolecules search engine (https://search.emolecules.com/), in the 'List Search' tab and selecting 'SMILES' under 'Data Type/Search By' filter criteria. This provides a list of only purchasable molecules with all possible vendors it can find. For such molecules, we manually went through each of the vendors listed and selected molecules with price < \$1000/g and lead time < 2 weeks for battery testing (procedure discussed in the Materials and Electrochemical characterizations subsections under the Methods section in the main text). From the first batch of suggestion, some of the suggested molecules were presumed to not dissolve salt and hence were not purchased. In particular, 1,1-dimethylcyclohexane (1-11) fit this criterion due to the lack of any solvating centers in the molecule.[7] Other structurally similar compounds 1,1-dichlorocyclopentane (1-9), 1-chloro-1-methylcyclohexane (1-10), 1,1-difluorocycloheptane (1-12), 2-chloro-1,1-difluorocyclohexane (1-13), 1,1-difluorocyclohexane (1-14) were also presumed to not dissolve salt based on domain knowledge. One compound was also found to be solid under ambient conditions from the list of suggestions — spiro[4.5]decane-7,9-dione (1-15; https://www.jwpharmlab.com/usr/MSDS/MSDS_10R0469.pdf). The discharge capacities for such cases were assigned as zero and fed to the surrogate models in the second batch, so that such compounds are not suggested in subsequent rounds of suggestions. Moreover, as discussed in the main text, we only tested suggested solvent molecules from the virtual search space in combination

with the LiFSA salt due to the advantages offered by LiFSA in comparison with $LiClO_4$ and LiDFOB salts. This was done to quickly search through vast chemical space for inherently efficient solvent molecules. Additionally, the solvent candidate molecules in combination with the LiFSA salt always ranked higher compared to other salts in the suggestions. However, for cases where electrolytes failed to cycle in combination with the LiFSA salt, the discharge capacities for other virtual electrolytes containing same compounds but in combination with $LiClO_4$ or LiDFOB salts were also assigned as zero without actual experimental testing to increase size of labeled data, as significant improvement is not expected with these salts. The tested solvent molecules were removed from the virtual search space in subsequent batches as per standard practice followed in an AL-based optimization framework.

Supplementary Note 3: Featurization and data transformation

The heterogeneous input data, consisting of solvent and salt molecule identities, salt concentration, amount of electrolyte, and theoretical capacity of positive electrodes was featurized into suitable numerical form and the target property was normalized before feeding to the surrogate models. The molecules are represented using the popular Morgan fingerprints, specifically a 1024-dimensional binary vector with radius cutoff of 2 (also called ECFP4 variant), generated by feeding SMILES text representation of molecules into the RDKit python library. However, since Gaussian processes (GPs) tend to suffer from high computational costs when dealing with such high-dimensional data, the 1024-dimensional Morgan fingerprints for solvents and salts were reduced to 10 dimensions each using the principal component analysis (PCA) analysis technique. The other inputs — salt concentrations, amount of electrolyte, and full capacity of positive electrodes (See note *b* of Table S2), were directly used and concatenated with reduced molecular representation resulting in a tabular input data. Therefore, the input data consists of a combination of continuous (composition and cycling condition) and discrete variables (molecular fingerprints). These input features corresponding to both labeled and unlabeled datasets were then standardized using the Standard Scaler function (transformed features have mean = 0 and unit s.d.) available in the scikit-learn python library by first fitting with respect to the labeled data and then transforming with respect to it. Since, the objective of this work is to optimize (maximize) based on discharge capacity that differs for different positive electrodes, the output was normalized with respect to the theoretical capacity.

Supplementary Note 4: Gaussian process (GP) as surrogate model

GP is an ML technique that uses probabilistic modeling to make predictions. It uses a mean function, a covariance function (kernel), and observed data to establish the relationship between input points and target values in a probabilistic manner. The model starts with prior beliefs about functions, incorporates the observed data's likelihood, and updates its beliefs to create a posterior distribution over functions. This posterior distribution is used to make predictions with uncertainty estimates. Specifically, GPs consist of following components:

1. **Mean function and covariance function (kernel)**: GP can be considered as an infinite collection of random variables, where each point in the input space is associated with a random variable. The mean function represents the average or expected value of the target variable at each point in the input space. The covariance function (also called the kernel) defines the similarity between data points. It quantifies how correlated the target values are at different input points. For example, if two input points are close in the input space, a high covariance indicates that their corresponding target values are likely to be similar.
2. **Prior distribution**: Before trained on any data, a GP defines a prior distribution over functions. This distribution captures initial beliefs about how functions in the problem space behave. At this stage, the GP assumes no knowledge about the specific relationship between input points and target values. The mean function and covariance function determine the shape and properties of this prior distribution. Successful mapping of input to labels requires good estimates of the prior distribution.
3. **Likelihood**: The likelihood function describes how likely the observed target values are, given the function values predicted by the GP and a noise model. It incorporates the uncertainty associated with the observed target values, accounting for potential noise in the data. For each observed data point, the likelihood measures how well the GP's predictions match the actual observations, considering the expected noise.
4. **Posterior Distribution**: Upon observing data, the GP is updated using Bayes' theorem to obtain the posterior distribution over functions. It leads to the "probabilistic" nature of the GPs. The posterior distribution combines the prior distribution (from the mean and covariance functions) with the likelihood (how well the GP's predictions match the data) to produce an updated belief about the functions that could explain the observed data.
5. **Predictive Distribution**: Given a new input point, the predictive distribution provides a probabilistic estimate of the target value. It considers both the posterior distribution, and the uncertainty introduced by the data. The predictive distribution is a normal distribution with a mean and variance, where the mean is the predicted target value, and the variance quantifies the uncertainty in the prediction.

Supplementary Note 5: Different kernels used in GPR and performance comparison

1. Radial Basis Function (RBF) Kernel: It is also known as the squared exponential kernel. It assumes that points close in input space have similar outputs, leading to a smooth and continuous model. This kernel is commonly used when one does not have specific assumptions about the nature of the data and need a generic, flexible model. Therefore, it is ideal for capturing smooth and continuous functions where the underlying function is assumed to be infinitely differentiable. It has a length scale hyperparameter.

$$k(x, x') = exp\left(-\frac{\left||x-x'|\right|^2}{2 \cdot length\_scale^2}\right) \tag{S2}$$

2. Exponential-Sine Squared (ESS) Kernel: This kernel captures periodicity by measuring the distance between points in terms of sine functions, making it effective for periodic data, i.e., when the function has known repeating pattern. It has length scale and periodicity as hyperparameters.

$$k(x, x') = exp\left(-2 \cdot \frac{sin^2\left(\frac{\pi \cdot ||x - x'||}{periodicity}\right)}{length\_scale^2}\right) \tag{S3}$$

3. Matern Kernel: The Matern kernel is more flexible than the RBF kernel, allowing for varying levels of smoothness. It is commonly used when the function may not be perfectly smooth and might have rougher transitions. It has different variants according to degrees of smoothness — Matern-1/2, Matern-3/2, and Matern-5/2. The expression for Matern-3/2 (suitable for functions that are moderately smooth but can still have some roughness) is as follows:

$$k(x, x') = \left(1 + \sqrt{3} \cdot \frac{||x-x'||}{length\_scale}\right) \cdot exp\left(-\sqrt{3} \cdot \frac{||x-x'||}{length\_scale}\right) \tag{S4}$$

4. Rational Quadratic (RQ) Kernel: The RQ kernel is a mixture of RBF kernels with different length scales. It is useful when a combination of long-range and short-range dependencies is expected in the data. This kernel can handle functions that have varying levels of smoothness and allows for modeling a wider range of scales in the data. The hyperparameters involved are length scale and alpha (controls the relative weighting of different length scales).

$$k(x, x') = \left(1 + \frac{||x - x\prime||^2}{2 \cdot \alpha \cdot length\_scale^2}\right)^{-\alpha} \tag{S5}$$

5. Pairwise Kernels: This kernel computes the similarity between pairs of data points in a polynomial form. It is useful for modeling data with interactions that are more linear or polynomial in nature, rather than smooth and continuous. It allows combining other kernels elementwise, e.g., multiplying or adding two kernels together.

$$k(x, x') = (x^T \cdot x' + c)^d \quad \text{(S6)}$$

where c is a constant term, and d is the polynomial degree.

6. ConstantKernel (also known as WhiteKernel): It is a simple kernel that only affects the diagonal of the covariance matrix. It is particularly useful for handling noise in data, especially in experimental or observational settings where measurement errors or inherent noise are present. Adding white noise can improve the numerical stability of the GP model, especially when fitting models to noisy data. Without this term, the model might try to fit every minor fluctuation, resulting in overfitting. It has noise level as the only hyperparameter.

$$k(x, x') = \sigma^2 \delta(x, x') \quad \text{(S7)}$$

where σ^2 is the noise variance, and $\delta(x, x')$ is the Kronecker delta (1 if $x = x'$, 0 otherwise).

The four surrogate models used in the present work are RBF+ESS (combined), matern-3/2 (Mat), RQ, and pairwise (Pair) kernels. A white noise kernel was also added to all the kernels used.

Performance comparison: The details for comparing performance of GP surrogate models with different covariance kernels shown in Fig. S2 is described below:

Performance comparison on in-batch data (Fig. S2a): The input features in each batch were transformed using the Standard Scaler function fitted on the same data. The trained surrogate models from each batch were then deployed on 5-fold cross validation (CV) data splits (random state = 42) created out of labeled data in the same batch. The mean and s.d. of the 5-fold CV root mean squared error (RMSE) values for each individual surrogate model is shown in the figure.

Performance comparison on out-of-batch data (Fig. S2b): For this performance comparison, 'out-of-batch' refers to labeled data from subsequent batches, e.g., out-of-batch for batch 3 would refer to labeled data from batches 4 to 7. The input features in the out-of-batch data were transformed using the Standard Scaler function fitted on the in-batch data. The trained models from each batch were then deployed on 5-fold CV data splits of the out-of-batch data. The mean and s.d. of the 5-fold CV RMSE values for each individual surrogate model is shown in the figure.

Supplementary Note 6: Oracle or acquisition function (EI) and related terms

EI was used as the acquisition function in the present study, as it has been widely reported to perform well in Bayesian optimization[8–10] tasks compared to other acquisition functions. The expression for EI when used with a single surrogate model is given by:

$$EI(x;\xi) = (\mu - f(x^*) - \xi)\Phi\left(\frac{\mu - f(x^*) - \xi}{\sigma}\right) + \sigma\phi\left(\frac{\mu - f(x^*) - \xi}{\sigma}\right) \quad \text{(S8)}$$

where, $f(x^*)$ is maximum predicted value on the labeled datasets (different for each batch), ξ is parameter controlling trade-off between exploration and exploitation (=0.01). Φ and ϕ refer to the cumulative (CDF) and PDF of the normal distribution, respectively. The final ranking was based on aggregated EI (EI^{aggr}; shown in Fig. S5b as a distribution for each batch) values from the four GP models by obtained by aggregating based on model weights (w; Supplementary Note 7):

$$EI^{aggr} = EI^{RBF-ESS} \cdot w^{RBF-ESS} + EI^{Mat} \cdot w^{Mat} + EI^{RQ} \cdot w^{RQ} + EI^{Pair} \cdot w^{Pair} \quad \text{(S9)}$$

The aggregated uncertainty (σ^{aggr}; equation S11 and shown in Fig. S5a) was calculated from aggregated prediction (μ^{aggr}; equation S10) according to following expression:

$$\mu^{aggr} = \mu^{RBF-ESS} \cdot w^{RBF-ESS} + \mu^{Mat} \cdot w^{Mat} + \mu^{RQ} \cdot w^{RQ} + \mu^{Pair} \cdot w^{Pair} \quad \text{(S10)}$$

$$\sigma^{aggr} = \sqrt{\sum_{j=1}^{4} w_j \cdot \left(\sigma_j^2 + \mu_j^2\right) - (\mu^{aggr})^2} \quad \text{(S11)}$$

where j = 1 for RBF-ESS, 2 for Mat, 3 for RQ and 4 for Pair surrogate models.

To quantify extent of exploration or exploitation in the oracle-suggested molecules within each batch, a metric called exploitation ratio ($r_{exploit}$; shown in Fig. S6) was defined, which is obtained from the ratio of first term (r_1^j; tends to favor exploitation) over second term (r_2^j; tends to favor exploration) in equation S8 for each of the four surrogate models and then aggregating them based on model weights:

$$r_1^j = \left(\mu^j - f(x^*)^{\mathrm{j}} - \xi\right)\Phi\left(\frac{\mu^j - f(x^*)^{\mathrm{j}} - \xi}{\sigma^j}\right);\ r_2^j = \sigma^j\phi\left(\frac{\mu^j - f(x^*)^{\mathrm{j}} - \xi}{\sigma^j}\right) \quad \text{(S12)}$$

$$r_{exploit} = \sum_{j=1}^{4} {r_1^j}\Big/{r_2^j} \quad \text{(S13)}$$

Supplementary Note 7: Bayesian model averaging (BMA)

BMA is a statistical framework that addresses model uncertainty by combining the predictions or parameter estimates of multiple models to obtain a more robust and accurate result. It considers a set of competing models and uses probabilistic techniques to weigh their contributions based on their likelihood and prior credibility. This approach ensures that uncertainty in model selection is explicitly accounted for, reducing the risk of overfitting and improving the reliability of predictions. It involves the following steps:

1. **Model selection:** A set of competing models is chosen, each representing different hypotheses, parameter values, or assumptions about the data. These models may vary in complexity or functional forms, allowing for a diverse exploration of the problem space.

2. **Bayesian inference:** The likelihood of the data given each model is calculated. This involves comparing the model's predictions to the observed data, considering uncertainties in the predictions. The likelihood is obtained using the probability distribution function (PDF) of predicted values, scaled by their respective model uncertainties.

3. **Prior beliefs:** Each model is assigned a prior probability reflecting initial beliefs about its suitability before observing the data. These priors may be based on existing knowledge, domain expertise, or set to non-informative values (e.g., equal priors of 1) if no prior information is available. In the present work, we used non-informative priors due to the lack of existing knowledge.

4. **Model weights (Posterior probabilities):** Using Bayes' theorem, the prior probabilities and likelihoods are combined to compute the posterior probability of each model:

$$\text{Posterior probability} = \text{Likelihood} \times \text{Prior} \tag{S14}$$

   where likelihood is calculated from PDF using equation S1 (without taking logarithm) and prior is taken to be 1 (from previous step). These posterior probabilities after normalizing with respect to sum of posterior probabilities of individual models represent model weights (w) for each model.

5. **Model averaging and combined prediction:** Model weights are derived from their posterior probabilities, assigning higher weights to models better supported by the data and lower weights to those with less support. The final prediction or parameter estimate is computed as a weighted average of the individual model predictions, where the weights are the normalized posterior probabilities. This ensures that the combined prediction reflects both the individual model contributions and their associated uncertainties.

BMA offers several advantages. It explicitly accounts for uncertainty in model selection, avoiding over-reliance on a single model that might not be optimal. By integrating predictions across multiple models, it reduces the risk of overfitting to specific data. BMA is also particularly effective when prior information (e.g., GP covariance kernels) is unavailable or uncertain. By integrating multiple models with different priors, it ensures robust performance in data-scarce settings.

Supplementary Note 8: Structural diversity of acquired compounds

Wasserstein or earth mover's distance: It is a measure of the distance between two probability distributions. In simpler terms, it represents the minimum cost of transforming one distribution into another, where the cost is the amount of 'work' needed to move the 'mass' from one distribution to the other. The analytical expression for the Wasserstein distance between two distributions u and v is given by:

$$W(P,Q) = \int_{-\infty}^{\infty} |F_P(x) - F_Q(x)| dx \quad \text{(S15)}$$

where P and Q are distributions for which the Wasserstein distance is to be computed, $F_P(x)$ and $F_Q(x)$ are the CDFs of P and Q.

Diversity of each batch (Fig. S11a): To check structural diversity of each acquired batch, i.e., chemical dissimilarity of compounds within each batch were calculated using the Wasserstein distance according to following expression:

$$W(\boldsymbol{r_i}, \boldsymbol{r_j}) = \frac{1}{10}\sum_{k=1}^{10} |r_{i,k} - r_{j,k}| \quad \text{(S16)}$$

where r_i is the vector of features (PCA-reduced Morgan fingerprints for solvent) in row i, r_j is the vector of features in row j. The summation is carried out ten times, which corresponds to the length of solvent features. Each row is treated as a distribution, and the function computes the Wasserstein distance between each pair of rows. Therefore, it results in a distribution of pairwise distances for solvent molecules in each batch, where lower values indicate that the molecules are chemically similar while the higher values denote chemically dissimilar pair of molecules. We used 'wasserstein_distance' function available in stats module of scipy python library to evaluate Wasserstein distances. The full width at half maxima (FWHM) can be considered an estimate for the chemical diversity of each batch. Batch 3 is found to be the most diverse batch as it consists of electrolyte solvents belonging to different functional groups such as chlorinated, fluorinated, and linear ethers, ester, amine, and sulfonamide. After batch 3, FWHM decreases until batch 6 and batch 7 (same for both), signifying that the later batches are less diverse compared to the in-house and initial batches and that the AL framework has undergone from exploration of uncertain region to exploitation of promising region in the chemical search space (ethers).

Comparison of distribution of acquired batches with respect to the in-house dataset (Fig. S11b): In a similar way, to compute the distribution of batches of molecules acquired from our AL workflow with respect to the in-house dataset, we computed Wasserstein distances of solvent features of each batch with respect to that of the in-house dataset. The batches that are distributed towards the right are more chemically dissimilar with respect to the in-house dataset, compared to the ones that are distributed towards left. Expectedly, batch 1 is distributed farthest to the right, signifying that it is the most chemically dissimilar batch of acquired molecules and the surrogate models are most uncertain for molecules in this batch. Since, none of the electrolytes from this batch 1 worked, so the oracle switched towards more certain region (similar to the in-house dataset) and hence batch 2 is distributed towards left. However, the oracle still tried to find more molecules from uncertain regions in the later batches in an attempt to find global maxima. Therefore, batches

3, 4, and 6 are distributed towards right, while the last batch (batch 7) is distributed towards left, signifying that the sequential optimization framework has attained exploitation of most promising region.

We also performed Tanimoto coefficient (TC) analysis to further confirm the similarity analysis of the acquired molecules in the later batches. For this, we calculated average of maximum TC values of all electrolyte solvent molecules in each batch with respect to those in the in-house dataset ($< TC^{max} >$). A TC value near zero between two molecules indicates high structural dissimilarity, while a value approaching one signifies strong structural similarity. As shown in Fig. S11b, the obtained $< TC^{max} >$ values indicate that the framework initially explores more diverse molecules (batch 1: 0.12, batch 2: 0.38) but gradually converges toward structurally similar compounds (batches 3-7: ~0.5-0.6). This analysis supports the observation that the AL framework selects more structurally similar molecules as it refines its predictions.

Supplementary Note 9: Rationale for transition from EI-guided acquisition to greedy sampling

From Fig. S5b, we observe that EI^{aggr} stabilizes after batch 6, indicating diminishing returns from acquiring molecules suggested by the acquisition function beyond this point. Additionally, a significant overlap is observed in the recommendations for batches 6 and 7, with nearly 70% of the molecules being identical (as shown in Fig. 2e). Notably, only three of the top 5000 EI^{aggr}-ranked molecules in batch 7 are commercially purchasable. Given this, it became more practical to prioritize molecule selection based on predicted mean values rather than relying solely on the acquisition function, despite low uncertainties in batch 7 being incorporated into EI^{aggr}. This explains why some uncertain molecules, including non-ether compounds that had failed in previous batches, were still suggested in batch 7. To address this, we adopted a greedy sampling strategy, focusing exclusively on candidates with high predicted mean values while disregarding uncertainties. This approach proved effective: greedy sampling in batch 7 yielded four high-performing electrolytes, compared to just one identified using EI^{aggr}-based selection in batch 6.

Supplementary Note 10: SHAP discussion

The SHAP summary bar plot for C_{norm}^{20} prediction by the RQ model is shown in Fig. S14(a) (equivalent to Fig. 2g)

. It ranks the importance of features influencing target property predictions in descending order. The molecular fingerprints were derived by reducing 1024-dimensional extended connectivity fingerprints (ECFPs) to 10 dimensions using PCA (Supplementary Note 3). The specific substructures contributing the highest variance to f_{solv}^{6}, f_{solv}^{4}, and f_{solv}^{5} are shown in Fig. 2h: a tetrafunctional carbon center (a carbon bonded to four functional groups), a methoxy group, and a trifunctional carbon center (a carbon bonded to three functional groups and one hydrogen), respectively. The fractions of labeled molecules containing these substructures are shown in Fig. 2i, with f_{solv}^{4} being the most frequent, followed by f_{solv}^{6} and f_{solv}^{5}. However, the average C_{norm}^{20} (ground truth) in the labeled data follows a different trend: $f_{solv}^{6} > f_{solv}^{4} > f_{solv}^{5}$, consistent with the SHAP feature relevance order (Fig. 2g and S14a). This demonstrates that SHAP accurately captures the true impact of these features on the target property rather than merely reflecting their frequency in the dataset. Interestingly, while molecules containing these substructures often perform well, some fail due to incompatible fragments, such as chlorine, that counteract the desirable effects of these substructures. The presence of these substructures is also highlighted in the t-SNE plot (Fig. 2f), with tick marks indicating molecules containing tetrafunctional and trifunctional carbons. These substructures often correspond to bi- or trifluorinated groups, which are common in state-of-the-art electrolytes like F5DEE (2-[2-(2,2-difluoroethoxy)ethoxy]-1,1,1-trifluoroethane).

The SHAP dependence plots for three features among top eight most relevant features for normalized discharge capacity at 20th cycle (C_{norm}^{20}) prediction — molecular weight of solvents (MW_{solv}), salt concentration (c_{salt}), and solvent fingerprint-6 (f_{solv}^{6}) are shown in Fig. S14. The SHAP values denote the impact on target property prediction — higher SHAP values mean higher C_{norm}^{20} in this case. The color bars for each dependence plot depict the variation of corresponding most interacting feature picked by the SHAP. For MW_{solv}, SHAP values and hence C_{norm}^{20} decrease upon increase in the molecular weight of solvent molecules until ~200 g/mol, after which it increased slightly (Fig. S14b). On the other hand, C_{norm}^{20} increase significantly upon increase in salt concentration (Fig. S14b) — a widely known trend in the literature and this design strategy is used for investigating high-concentration electrolytes (HCEs)[11]. The dependence of f_{solv}^{6} with C_{norm}^{20} shows a linear correlation (Fig. S14c), which means higher fraction of tetrafunctional carbon centers (e.g., trifluorinated fragment) would lead to higher discharge capacities. The f_{solv}^{6} feature itself interacts with f_{solv}^{8} feature (substructure that contributes the highest variance to this feature also corresponds to tetrafunctional carbon centers but present on cyclic compounds, mostly attached to other functional groups, e.g., nitrogen or carbonyl as for molecules 1-3, 1-4, 1-7, 1-8 in batch 1 that did not cycle) — but has inverse correlation as higher f_{solv}^{8} values lie in negative SHAP value regions and vice versa. This denotes that the SHAP analysis also captured the incompatible molecular substructures and correctly predicted its trend on the desired target property.

Supplementary Note 11: Effect of exclusion of NMC positive electrode data on AL campaigns

To rigorously evaluate the influence of cycling data with NMC positive electrodes in the initial in-house data (Table S2), we conducted an additional AL campaign using the same settings as described in Supplementary Notes 2, 3, 4, 5, and 6 by excluding all NMC data from the initial dataset. We observed minimal degradation in the surrogate model performance for out-of-batch data upon removing the NMC data (Fig. S20), indicating limited influence of the NMC data inclusion, e.g., average 5-fold CV RMSE of 0.2558 and 0.2614 for batch 6 of NMC-included (Fig. S2b) and NMC-excluded AL campaigns, respectively. Finally, the total number of high-performing electrolytes discovered from this new AL campaigns remained unchanged at seven – compounds with IDs 5-7, 6-3, 7-1, 7-12, 7-14, 7-15, and 7-16. We have publicly shared the datasets, Jupyter notebooks and surrogate model checkpoints obtained from both the original and the NMC-excluded AL campaigns via our GitHub repository. Users can thereby select models and datasets based on their specific application needs.

Table S1. Literature reported anode-free LMB performance

Electrolyte	**Electrolyte amount**	**Cell condition**	**Cycling condition & capacity retention**	**Ref.**
4 M LiFSA in DME	~44 g/Ah	Cu||LFP coin cell, 1.71 mAh/cm^2	~C/8 charge/discharge, 60% after 50 cycles	[12]
1 M LiTFSA + 2 M LiFSA + 3 wt% $LiNO_3$ in DME/DOL	Not mentioned	Cu||LFP coin cell, 0.85 mAh/cm^2	39% after 100 cycles	[13]
1.7 M LiFSA in MeTHF/TTE	4.0 g/Ah	Cu||LFP pouch cell, 2.7 mAh/cm^2, 560 mAh	0.3 C charge 0.5 C discharge, 41.6% after 150 cycles	[14]
7 M LiFSA in FEC	Not mentioned	Cu||LNMO coin cell, 1.43 mAh/cm^2	54% after 50 cycles	[15]
2 M $LiPF_6$ in EC/DEC + 50% FEC	Not mentioned	Cu||NMC111 coin cell, ~1.6 mAh/cm^2	~C/8 charge/discharge, 40% after 50 cycles	[16]
1 M $LiPF_6$ in FEC/FEMC/HFE	~47 g/Ah	Cu||NMC811 coin cell, ~2.0 mAh/cm^2	~C/4 charge/discharge, 50% after 30 cycles	[17]
1LiFSA-1.2DME-3TTE	3 g/Ah	Cu||NMC811 coin cell, ~4.2 mAh/cm^2	C/10 charge C/3 discharge, 77% after 70 cycles	[18]
0.6 or 1 M LiDFOB + 0.6 or 0.2 M $LiBF_4$ in FEC/DEC	~2 g/Ah	Cu||NMC532 pouch cell, ~245 mAh	~80% DOD (depth of discharge), C/5 charge C/2 discharge, 40°C, 80% after 80 or 90 cycles (low pressure)	[19]

1.8 M LiDFOB + 0.4 M $LiBF_4$ in FEC/DEC	~2 g/Ah	Cu||NMC532 pouch cell, ~210 mAh	~70% DOD, C/5 charge C/2 discharge, 40°C hot formation, 80% after 90 cycles (low pressure) / 80% after 195 cycles (high pressure)	[20]
0.6 M LiDFOB + 0.6 M $LiBF_4$ in FEC/DEC	~2 g/Ah	Cu||NMC532 pouch cell, ~210 mAh	~70% DOD, C/5 charge C/2 discharge, 80% after ~16 cycles (low pressure) / 80% after 50-60 cycles (high pressure)	[20]
2.0 M LiDFOB and 1.4 M $LiBF_4$ in FEC/DEC	~2.6 g/Ah	Cu||NMC532 pouch cell, ~210 mAh	~70% DOD, C/5 charge C/2 discharge, 80% after 200 cycles (high pressure)	[21]
1 M LiFSA in DMTMSA	~3 g/Ah	Cu||NMC811 coin cell, 4.5 mAh/cm^2	100% DOD (3-4.5 V), C/10 charge C/3 discharge, 73% after 65 cycles	[22]
1 M LiFSA in FDMB	~2 g/Ah	Cu|| NMC532, Cu||NMC622, and Cu||NMC811 pouch cells, 200-250 mAh	100% DOD, C/5 charge C/3 discharge, 80% after 100 cycles (low pressure)	[23]
1.2 M LiFSA in F5DEE	~2.4 g/Ah	Cu||micro-LFP pouch cells, ~ 170 mAh	100% DOD, 0.2 C charge 2 C discharge, 80% after 140 cycles (high pressure)	[24]
1 M LiFSA in BFE	Not mentioned	Cu||NMC811 coin cell, 3.5 mAh/cm^2	100% DOD, C/10 charge/discharge, 88% after 40 cycles	[25]

2 M LiFSA in 3BTFM-1DME	Not mentioned	Cu||NMC811 coin cell, ~1.6 mAh/cm^2	100% DOD, 1 C charge/discharge, 64% after 80 cycles	[26]
2.1 M LiFSA in F2EMP	~ 2.3 g/Ah	Cu||NMC811 coin cell, 3.885 mAh/cm^2	~83% DOD, 0.1 C charge 0.3 C discharge, 80% after 129 cycles	[27]
2 M LiFSA in DEP	~ 5 g/Ah	Cu||NCA coin cell, 4 mAh/cm^2	100% DOD, 0.3 C charge/discharge, 61.7% after 250 cycles	[28]

Table S2. In-house anode-free cell cycling data set

Solvent	**Salt**	**Cell type** [a]	**Discharge capacity at 20th cycle (mAh/g)** [b]
1,2-Dimethoxyethane (SMILES: COCCOC)	1 M LiFSA	Cu||LFP	41.49
iPrME[29] (SMILES: CC(C)OCCOC)	1 M LiFSA	Cu||LFP	99.22
iBME[29] (SMILES: CC(C)COCCOC)	1 M LiFSA	Cu||LFP	26.76
iBME[29] (SMILES:CC(C)COCCOC)	1 M LiFSA	Cu||LFP	32.54
tBEE[29] (SMILES: CC(C)(C)OCCOCC)	2.2 M LiFSA	Cu||LFP	3.93
iPrEE[29] (SMILES: CC(C)OCCOCC)	2 M LiFSA	Cu||LFP	93.07
iPrEE[29] (SMILES: CC(C)OCCOCC)	2 M LiFSA	Cu||LFP	96.56
nBME[29] (SMILES: CCCCOCCOC)	1 M LiFSA	Cu||LFP	42.75
cPME[29] (SMILES: C1CCCC1OCCOC)	1 M LiFSA	Cu||LFP	69.89
iPrBE[29] (SMILES: CC(C)OCCOCCCC)	2 M LiFSA	Cu||LFP	84.9
Diglyme (SMILES: COCCOCCOC)	1 M LiFSA	Cu||LFP	16.80
E3F1[30] (SMILES: FC(F)(F)COCCOCCOCC(F)(F)F)	1 M LiFSA	Cu||LFP	55.04
E3CH2F[31] (SMILES: FCCOCCOCCOCCF)	1 M LiFSA	Cu||LFP	1.40
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA	Cu||LFP	97.92
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA	Cu||LFP	115.58
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA	Cu||LFP	105.13

E4F1[30] (SMILES: FC(F)(F)COCCOCCOCCOCC(F)(F)F)	2 M LiFSA	Cu||LFP	71.67
1,1,1,2,2,3,3-heptafluoro-4-(2-methoxyethoxy)butane (SMILES: C(COCCOC)(C(F)(F)C(F)(F)F)(F)F)	1 M LiFSA	Cu||LFP	112.76
1,1,1,2,2,3,3-heptafluoro-4-(2-methoxyethoxy)butane (SMILES: C(COCCOC)(C(F)(F)C(F)(F)F)(F)F)	1 M LiFSA	Cu||LFP	102.73
E2CH2F[31] (SMILES: FCCOCCOCCF)	1 M LiFSA	Cu||LFP	36.525
1,1,2,2-Tetrafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(F)(F)C(F)F)	1 M LiFSA	Cu||LFP	122.52
1,1,2,2-Tetrafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(F)(F)C(F)F)	1 M LiFSA	Cu||LFP	103.74
FDMB[32] (SMILES: COCC(F)(F)C(F)(F)COC)	1 M LiFSA	Cu||LFP	104.91
FDMB[32] (SMILES: COCC(F)(F)C(F)(F)COC)	1 M LiFSA	Cu||LFP	101.62
FDMB[32] (SMILES: COCC(F)(F)C(F)(F)COC)	1 M LiFSA	Cu||LFP	88.11
FDMB[32] (SMILES: COCC(F)(F)C(F)(F)COC)	1 M LiFSA	Cu||LFP	85.60
TMEB[33] (SMILES: B(OCCOC)(OCCOC)OCCOC)	1 M LiFSA	Cu||LFP	0.112
Methyl 2,2,2-trifluoroethyl carbonate (FEMC, SMILES: COC(=O)OCC(F)(F)F)	1 M LiFSA	Cu||LFP	7.617
Trifluoro propylene carbonate (TFPC, SMILES: C1OC(=O)OC1C(F)(F)F)	1 M LiFSA	Cu||LFP	0.744
TFEB[33] (SMILES: B(OCCF)(OCCF)OCCF)	1 M LiFSA	Cu||LFP	58.13
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA	Cu||LFP	116.46

2-(2-Methoxyethoxymethyl) furan (SMILES: COCCOCC1=CC=CO1)	1 M LiFSA	Cu||LFP	0.031
t-Butyl 2-methoxyethyl ether (SMILES: CC(C)(C)OCCOC)	1 M LiFSA	Cu||LFP [c]	41.42
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA	Cu||NMC811	149.42
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA, 0.2 M LiDFOB	Cu||NMC811	0.11
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA	Cu||NMC811	150.84
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M LiFSA, 0.2 M $LiBF_4$	Cu||NMC811	19.80
1,1,1,2,2-Pentafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(C(F)(F)F)(F)F)	1 M $LiBF_4$	Cu||NMC811	0.002
1,1,1,2,2,3,3-heptafluoro-4-(2-methoxyethoxy)butane (SMILES: C(COCCOC)(C(F)(F)C(F)(F)F)(F)F)	1 M LiFSA	Cu||NMC811	127.22
1,1,2,2-Tetrafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(F)(F)C(F)F)	1 M LiFSA	Cu||NMC811	124.24
1,1,2,2-Tetrafluoro-3-(2-methoxyethoxy)propane (SMILES: COCCOCC(F)(F)C(F)F)	1 M LiFSA	Cu||NMC811	130.43
1,1,1,2,2,3,3-heptafluoro-4-(2-methoxyethoxy)butane (SMILES: C(COCCOC)(C(F)(F)C(F)(F)F)(F)F)	1 M LiFSA	Cu||NMC111	90.206

1,1,1,2,2,3,3-heptafluoro-4-(2-methoxyethoxy)butane (SMILES: C(COCCOC)(C(F)(F)C(F)(F)F)(F)F)	1 M LiFSA	Cu||NMC111	90.21
1,1,1,2,2,3,3-heptafluoro-4-(2-methoxyethoxy)butane (SMILES: C(COCCOC)(C(F)(F)C(F)(F)F)(F)F)	2 M LiFSA	Cu||NMC111	99.47
1,2-Diethoxyethane (SMILES: CCOCCOCC)	1 M LiFSA	Cu||LFP	115.52
1,2-Dimethoxyethane (SMILES: COCCOC)	1 M LiFSA	Cu||LFP	55.63
iPrME[29] (SMILES: CC(C)OCCOC)	2 M LiFSA	Cu||LFP	109.53
iPrME[29] (SMILES: CC(C)OCCOC)	2 M LiFSA	Cu||LFP	111.86
1,2-Dimethoxyethane (SMILES: COCCOC)	2 M LiFSA	Cu||LFP	102.06
1,2-Dimethoxyethane (SMILES: COCCOC)	2 M LiFSA	Cu||LFP	106.45
1,2-Dimethoxyethane (SMILES: COCCOC)	2 M LiFSA	Cu||LFP	95.3
iPrEE[29] (SMILE: CC(C)OCCOCC)	2 M LiFSA	Cu||LFP	40.21
EESF[2] (SMILES: CCOCCS(=O)(=O)F)	1 M LiFSA	Cu||LFP	3.55
EESF[2] (SMILES: CCOCCS(=O)(=O)F)	2 M LiFSA	Cu||LFP	11.63
ESF[2] (SMILES: CCS(=O)(=O)F)	1 M LiFSA	Cu||LFP	74.23
ESF[2] (SMILES: CCS(=O)(=O)F)	2 M LiFSA	Cu||LFP	114.49
ESF[2] (SMILES: CCS(=O)(=O)F)	4 M LiFSA	Cu||NMC811	145.87
1,2-Dimethoxyethane (SMILES: COCCOC)	4 M LiFSA	Cu||NMC811	154.52

[a] Without additional note, all cycling was performed in CR2032 coin cells at 20 °C, where two formation cycles were first performed at C/10 and followed by cycling at C/3. 1 C = 150 mA/g, 161 mA/g and 203 mA/g for LFP, NMC111 and NMC811 positive electrodes, respectively.

[b] The discharge capacity was normalized to the full cell capacity (150 mAh/g, 161 mAh/g and 203 mAh/g for LFP, NMC111 and NMC811 positive electrode, respectively) when used to train the AL model to minimize the effect of positive electrode selection.

[c] Cycling was performed at 60 °C.

Table S3. Summary of experimentally tested electrolyte solvent molecules

ID	Name	CAS number	1 M LiFSA solubility	Discharge capacity of Cu||LFP coin cell at 20th cycle (mAh/g)		Source
				Cell1	Cell2	
Batch #1						
1-1	4,4-Difluoro-1-methylpiperidine	1186194-60-4	N	0	0	Combi-block (95%)
1-2	N-Methylpiperidine	626-67-5	N	0	0	Sigma-aldrich (99%)
1-3	3,3-Dimethylcyclohexanone	2979-19-3	Y	0.01	0.01	Ambeed (97%)
1-4	1,2,2,6,6-Pentamethylpiperidine	79-55-0	N	0	0	Ambeed (98%)
1-5	1-ethylpyrrolidine	7335-06-0	N	0	0	Ambeed (97%)
1-6	Methyl 1-Methylcyclopentanecarboxylate	4630-83-5	Y	0.14	0.18	Combi-block (97%)
1-7	1-Methylcyclohexanecarbonyl chloride	2890-61-1	N	0	0	Ambeed (95%)
1-8	1-chloro-2,2,6,6-tetramethylpiperidine	32579-76-3	N	0	0	Synthonix (97%)
Batch #2						
2-1	Trifluoracetaldehyd-dimethylacetal	42415-20-3	N	0	0	Synquest Lab (97%)
2-2	Methyl 2,2,3,3-tetrafluoropropyl ether	60598-17-6	N	0	0	Sigma-aldrich (97%)
2-3	3,3,3-Trifluoropropanal dimethylacetal	116586-94-8	Y	56.641	76.59	Synquest Lab (95%)
2-4	1,3,3-Trimethoxybutane	6607-66-5	Y	0.001	0	Oakwood Chemical (97%)

2-5	1,2-dimethoxypropane	7778-85-0	Y	0.001	0.001	Sigma-aldrich (⩾ 99%)
2-6[a]	1H,1H,2H,2H-Perfluoropentyltrimethoxysilane	109134-39-6	Y	2.06	8.233	Synquest Lab (97%)
2-7	2-((2-(methoxy)ethoxy)methoxy)-1,1,1-trifluoroethane	130156-55-7	Y	4.378	7.218	Synthonix (96%)
2-8[b]	chloroacetaldehyde dimethyl acetal	97-97-2	Y	0	0	Sigma-aldrich (98%)
2-9	2-chloro-1,1,2-trifluoroethyl methyl ether	425-87-6	N	0	0	1PlusChem (>99.0%)
2-10	3-Methoxybutyraldehyde dimethyl acetal	10138-89-3	Y	0.09	1.27	Oakwood Chemical (98%)
2-11	methyl propyl ether	557-17-5	Y	67.787	37.18	Sigma-aldrich (97%)
Batch #3						
3-1[b]	2-Chloroethyl Ethyl Ether	628-34-2	Y	0	0	TCI America (>98%)
3-2	N,N-dimethyl-2-methoxyethylamine	3030-44-2	Y[c]	0.001	0.001	Ambeed (97%)
3-3	Ethyl N-(2-methoxyethyl)-N-methylglycinate	616882-60-1	Y	0.001	0.0014	Sigma-aldrich (95%)
3-4	1-methoxyhexane	4747-07-3	Y	84.488	76.474	TCI America (>98%)
3-5	Triethylene Glycol Butyl Methyl Ether	7382-30-1	Y	32.01	42.77	TCI America (>98%)
3-6	Diethylene glycol methyl 2,2,3,3-tetrafluoropropyl ether (SMILES: COCCOCCOCC(F)(F)C(F)F)	-	Y	90.025	74.854	Synthesized

3-7[b]	N-methyl-N-(2-methoxyethyl)methanesulfonamide	691004-53-2	Y	0	0	Synthesized
Batch #4						
4-1	Hexyl Formate	629-33-4	Y	0.02	0.01	TCI America (>98%)
4-2	Hexyl Methanesulfonate	16156-50-6	Y	0.421	0.48	Ambeed (97%)
4-3[a]	Amyl Ether	693-65-2	Y	9.6	6.89	TCI America (>98%)
4-4[b]	Butyl 2-Chloroethyl Ether	10503-96-5	Y	0	0	TCI America (>98%)
4-5	Ethylene Glycol Dibutyl Ether	112-48-1	Y	21.20868	25.06872	TCI America (>98%)
4-6	Butyl Methyl Ether	628-28-4	Y	96.01	94.28	TCI America (>99%)
4-7	Hexyl Nitrite	638-51-7	N	0	0	TCI America (>95%)
4-8	Trimethoxy(1H,1H,2H,2H-tridecafluoro-n-octyl)silane	85857-16-5	N	0	0	TCI America (>97%)
4-9	Valeraldehyde-diethylacetal	3658-79-5	Y	0	0	Ambeed (95%)
Batch #5						
5-1	(2-Methoxyethoxy)acetaldehyde dimethyl acetal	94158-44-8	Y	0.01	0.01	Ambeed (95%)
5-2[b]	1-Chloro-4-methoxybutane	17913-18-7	Y	0	0	Ambeed (98%)
5-3	2,2,3,3,3-Pentafluoropropyl methyl ether	378-16-5	N	0	0	Sigma-aldrich (98%)
5-4	Ethylene glycol ethyl methyl ether	5137-45-1	Y	98.57	103.81	Sigma-aldrich (98%)
5-5	Methoxycyclohexane	931-56-6	Y	87.2	88.91	TCI America (>98%)
5-6	1,5-Dimethoxypentane	111-89-7	Y	52.59	61.31	TCI America (>97%)
5-7	1,3-Dimethoxypropane	17081-21-9	Y	110.48	113.59	Ambeed (95%)

5-8	1-Methoxy-3-(3-methoxypropoxy)propane	66226-74-2	Y	87.5	89.05	Ambeed (95%)
5-9[b]	Methanesulfonyl fluoride	558-25-8	Y	0	0	Oakwood Chemical (98%)
5-10	1,3-Dichloro-2,2-dimethoxypropane	6626-57-9	N	0	0	Ambeed (97%)
Batch #6						
6-1	Triethylene glycol dimethyl ether	112-49-2	Y	64.63695	75.83542	Sigma-aldrich (99%)
6-2	Tetraethylene glycol dimethyl ether	143-24-8	Y	75.31111	76.565	Sigma-aldrich (≥ 99%)
6-3	Cyclopentyl methyl ether	5614-37-9	Y	124.74	123.53	Sigma-aldrich (inhibitor-free, anhydrous, ≥99.9%)
6-4	1,6-Dimethoxyhexane	13179-98-1	Y	80.97	57.75	TCI America (>97%)
6-5	Methyl 2-Oxo-1-cycloheptanecarboxylate	52784-32-4	Y	0	0	Ambeed (97%)
6-6	Bis(2-methoxyethoxy)methane	4431-83-8	Y	107.364	106.795	Ambeed (97%)
6-7	Diethylene Glycol Butyl Methyl Ether	7382-32-3	Y	55.208	77.688	TCI America (>99%)
6-8	Diethylene Glycol Ethyl Methyl Ether	1002-67-1	Y	93.376	106.61	TCI America (>98%)
6-9	2-Methoxycyclohexanone	7429-44-9	Y	0	0	Ambeed (97%)
6-10	2-methoxyethane-1-sulfonyl fluoride	1087410-86-3	Y	36.09892	43.31997	Enamine (95%)
6-11[b]	1-methoxy-2-(methylsulfanyl)ethane	35332-09-3	Y	0	0	Enamine (95%)

Batch #7						
7-1	1,1-Difluoro-2-(2-methoxyethoxy)ethane	1270009-33-0	Y	115.21	115.97	Ambeed (95%)
7-2	2-Methoxytetrahydropyran	6581-66-4	Y	36.01	32.75	Ambeed (98%)
7-3	2-Methoxyethyl trifluoromethanesulfonate	112981-50-7	N	0	0	Sigma-aldrich (95%)
7-11	3-methoxypropyl neoPentyl ether (SMILES: COCCCOCC(C)(C)C)	-	Y	66.5	66.08	Synthesized
7-12	4-methoxybutyl 2,2,2-trifluoroethyl ether (SMILES: COCCCCOCC(F)(F)F)	-	Y	107.56	110.28	Synthesized
7-13	3-fluoropropyl methyl ether (SMILES: COCCCF)	-	Y	0.01	0.01	Synthesized
7-14	3-methoxypropyl 1-butyl ether	90951-92-1	Y[c]	116.35	116.64	Synthesized
7-15	1-Ethoxy-4-methoxy butane	36865-47-1	Y	116.64	114.92	Synthesized
7-16	3-methoxypropyl 2,2,3,3-tetrafluoropropyl ether (SMILES: COCCCOCC(F)(F)C(F)F)	-	Y	124.22	122.82	Synthesized

[a] Cu||LFP cells were tested at current rate of C/5 instead of C/3 due to high overpotential (1 C = 150 mA/g).

[b] Cu||LFP cells failed to complete the 1st cycle.

[c] Phase separation was observed at 1 M concentration of LiFSA, higher concentration solutions (~1.1 M for 3-2 and 2 M for 7-14) were used for testing.

Table S4. Specifications of electrodes

	$LiFePO_4$	**$LiNi_{1/3}Mn_{1/3}Co_{1/3}O_2$**	**$LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$**
Active material (wt%)	92%	90%	90%
Timcal C-45 (wt%)	3.9%	5%	5%
PVDF binder (wt%)	4% (Solvay 5130)	5% (Solvay 5130)	5% (Solvay 5130)
Other additives (wt%)	0.1% Tuball SWCNT	-	-
Total mass loading (mg/cm^2)	12.83	11.22	9.08
Areal capacity (mAh/cm^2)	1.77	1.62	1.66
Reversible specific capacity (mAh/cm^2)	150	161	203

Table S5. Physical properties of the best performing AL discovered electrolytes

	Density (g/cm^3)	**Boiling point (°C)**	**Ionic conductivity with 1 M LiFSA at 20 °C (mS/cm)**	**Viscosity with 1 M LiFSA at ~21 °C (mPa s)**
6-3	0.86	106 (1 atm)	0.67	1.331
7-1	1.09	62 (~ 110 mbar)	3.51	3.302
7-15	0.83	47 (~20 mbar)	1.05	2.540
7-16	1.15	37 (~10 mbar)	0.75	7.545

Table S6. List of undesirable chemical moieties that were removed

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
Unsaturated compounds	alkene (C=C), alkyne (C≡C), nitrile (C≡N)
Aromatic and anti-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 or radical compounds	cation (+), anion (–), radicals (•)
Enantiomers	chiral stereocenters (*, e.g., D-(+) or L-(–))

Table S7. List of chemical moieties that were augmented

Initial chemical moieties	**Augmented chemical moieties**
Alcohols (–OH)	Methoxy ethers (–OCH_3)
Carboxylic acids (-COOH)	Methyl esters (–$COOCH_3$)
Primary and secondary amines (–NH_2; –NHR)	Tertiary amines (–$N(CH_3)_2$; –$N(CH_3)$R)
Primary and secondary amides (–$CONH_2$; –CONHR)	Tertiary amides (–$CON(CH_3)_2$; –$CON(CH_3)$R)
Thiols (–SH)	Thioethers (–SCH_3)

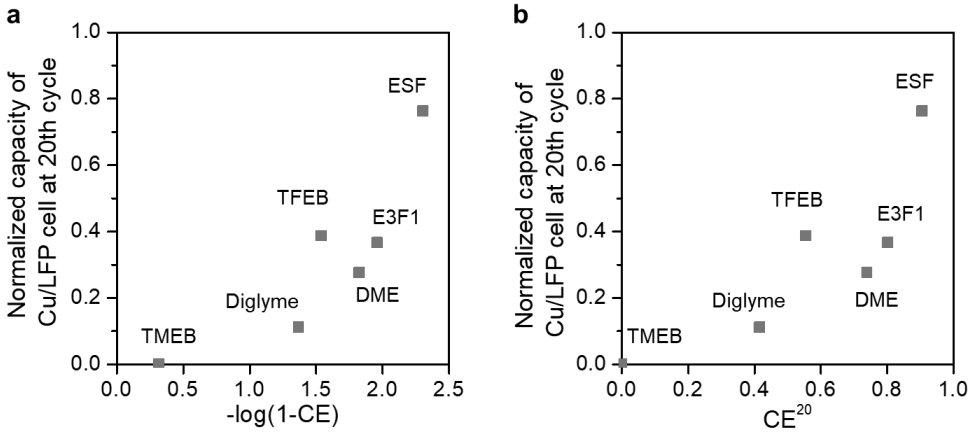


Figure S1. **Comparison of anode-free cell cycling performance and Coulombic efficiency (CE) measured in Li||Cu cell. a**) Using logarithmic CE as independent variable. **b**) Using the 20th power of CE as independent variable. In either case, CE has a positive correlation with anode-free cell performance but the order in CE does not match exactly to the anode-free cell performance order. Cu||LFP cell capacity was normalized to the full capacity of LFP (150 mAh/g). DME: 1 M Li FSA in 1,2-dimethoxy ethane; E3F1 and diglyme data are retrieved from ref. [30]; TFEB and TMEB data are retrieved from ref. [33]; ESF data are retrieved from ref. [2].

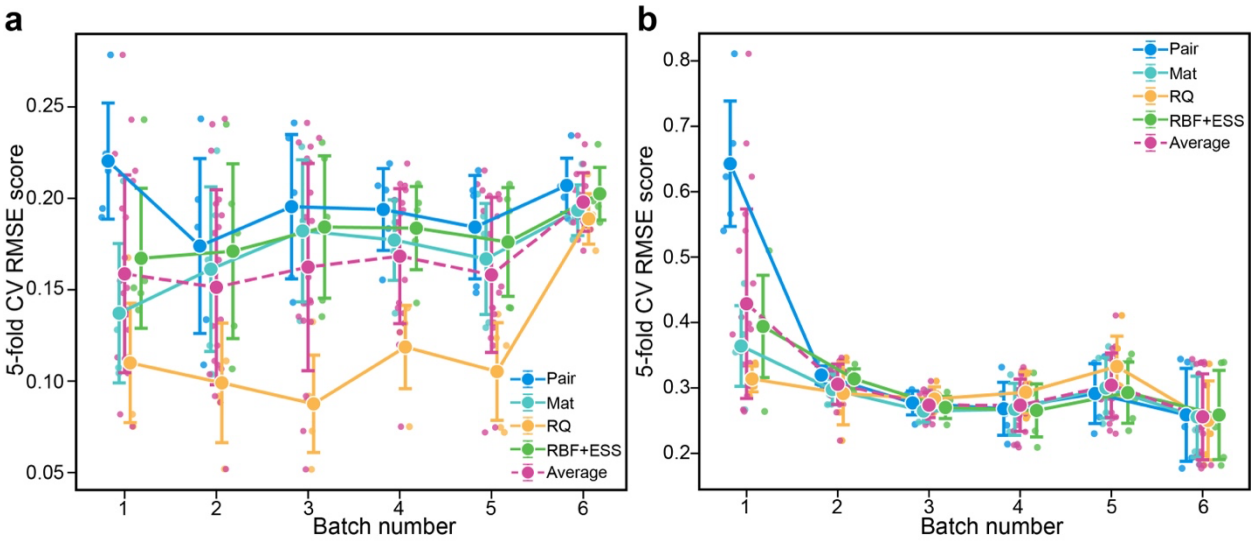


Figure S2. Surrogate model performance comparison. 5-fold cross-validation (CV) **a**) in-batch and **b**) out-of-batch performance of GP surrogate models as a function of batches. The error bars for each kernel (solid lines) denote standard deviation (s.d.) in root mean squared error (RMSE) of model predictions on 5-fold CV splits, and error bars for the average performance (dashed purple line) denote s.d. in RMSE of all 5-fold CV models. Since there are 5 models, one for each of the folds, the performance from each individual fold models are also shown as smaller scatter points for each batch. Pair = pairwise kernel, Mat = matern-3/2 kernel, RQ = rational quadratic kernel, RBF = radial basis function, ESS = exponential-sine squared kernel.

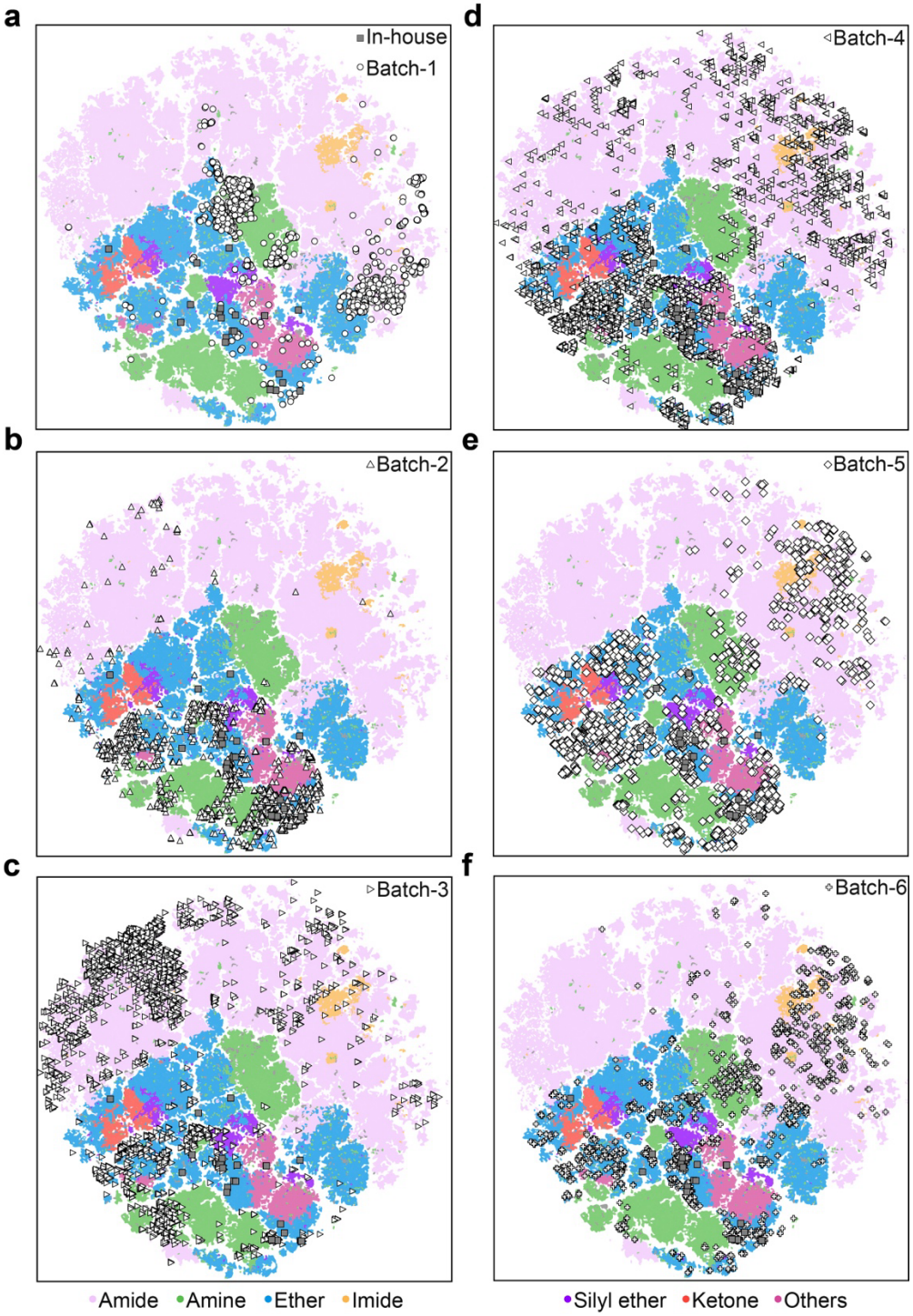


Figure S3. Visualization of progression of suggested molecules on the chemical space. The t-SNE plots depicting location of suggested molecules on the virtual chemical search space (~380000 unique molecules) for **a**) batch 1, **b**) batch 2, **c**) batch 3, **d**) batch 4, **e**) batch 5, and **f**) batch 6. The complete virtual chemical space in the background has been colored according to different functional group classes (only top 7 populated classes shown here for clarity).

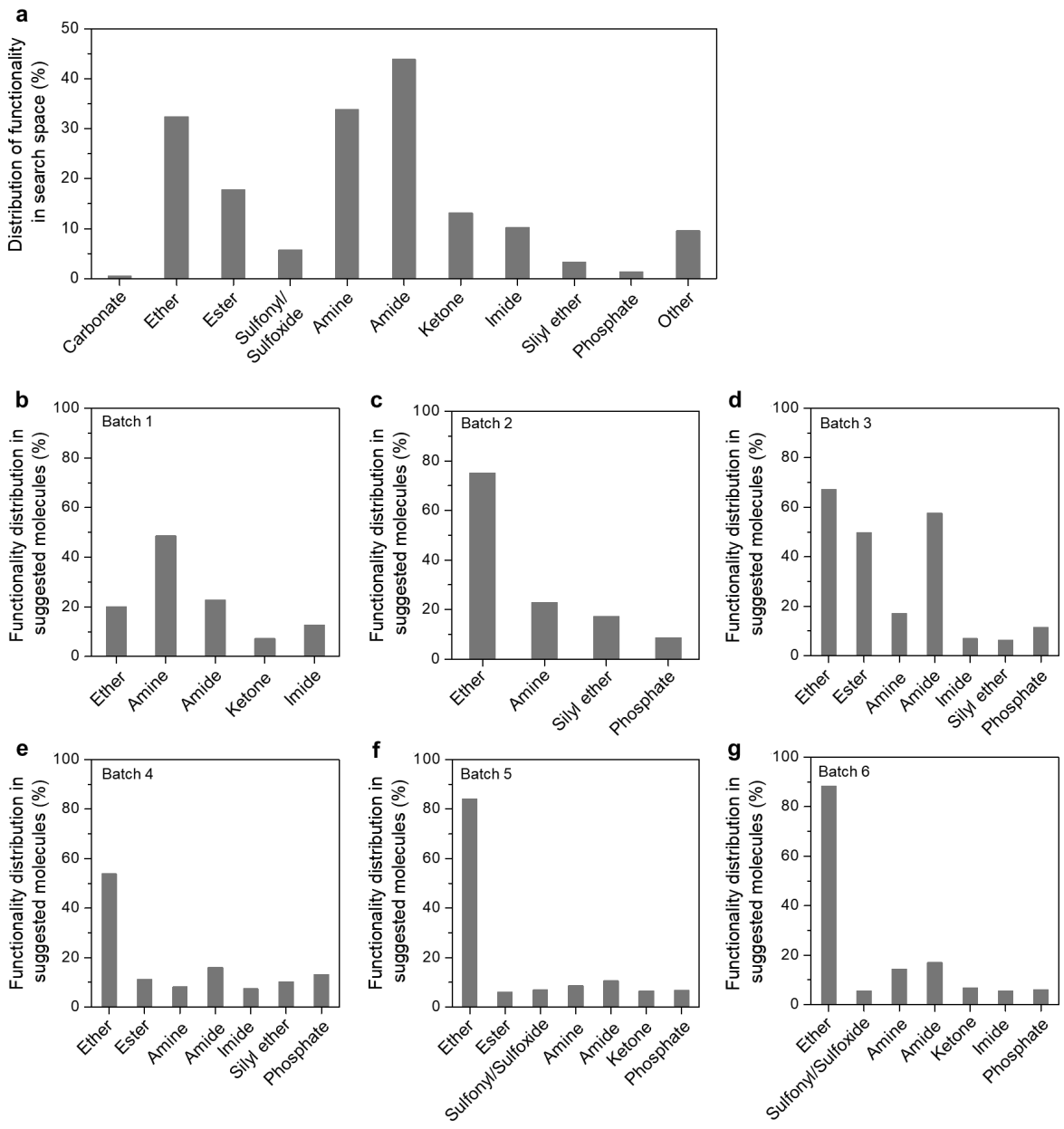


Figure S4. **Functional group distribution of suggested solvent molecules in different AL batches.** Bar plot depicting distribution of functional group classes (inclusive classification, Supplementary Note 1) in: **a**) virtual search space; **b**) batch 1; **c**) batch 2; **d**) batch 3; **e**) batch 4; **f**) batch 5; **g**) batch 6. For clarity, only categories with > 5% fraction are shown in (**b**-**g**). The functional group distribution in suggested molecules of the early batches is closer to the average of search space, where ether, amine and amide show the highest frequencies. From batches 4 to 6, ether functionality gradually dominates in the suggested molecules, in agreement with the trend of exclusive ether fraction shown in Fig. 2c.

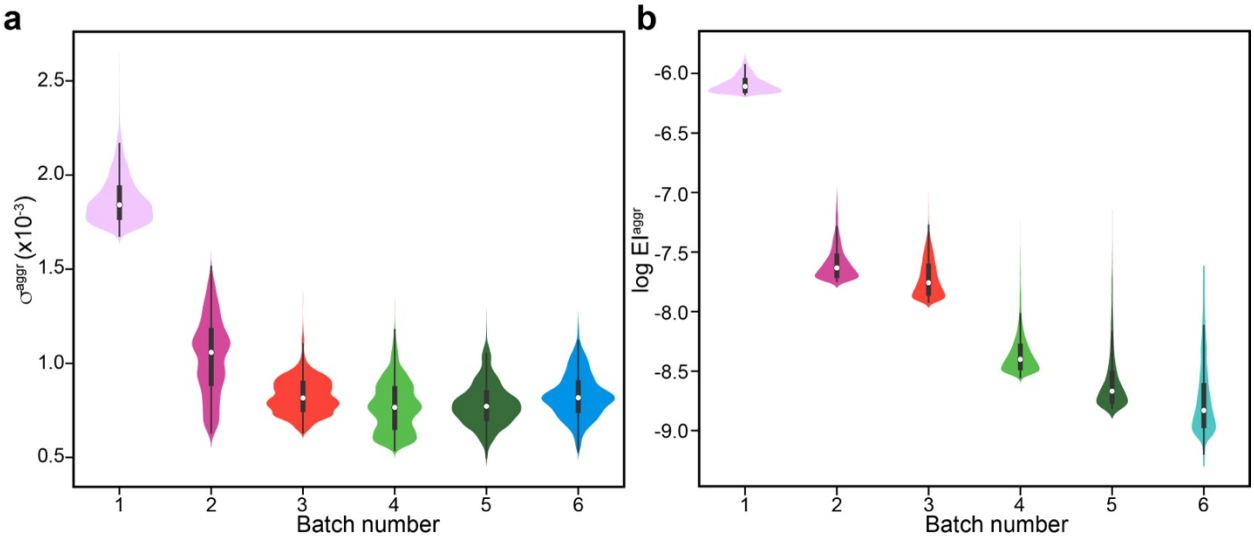


Figure S5. **Distribution of aggregated uncertainties and expected improvement in different AL batches**. Violin plot depicting distribution of **a**) aggregated uncertainty (σ^{aggr}) and **b**) log EI^{aggr} values for different batches (top 5000 for all batches except for 3 and 4, where top 10000 compounds were selected). The outer shells, thick lines, and thin lines span the complete data distribution, the interquartile range (i.e., between 25% and 75% of data), and 1.5 times the interquartile range, respectively. The white circles represent the median values.

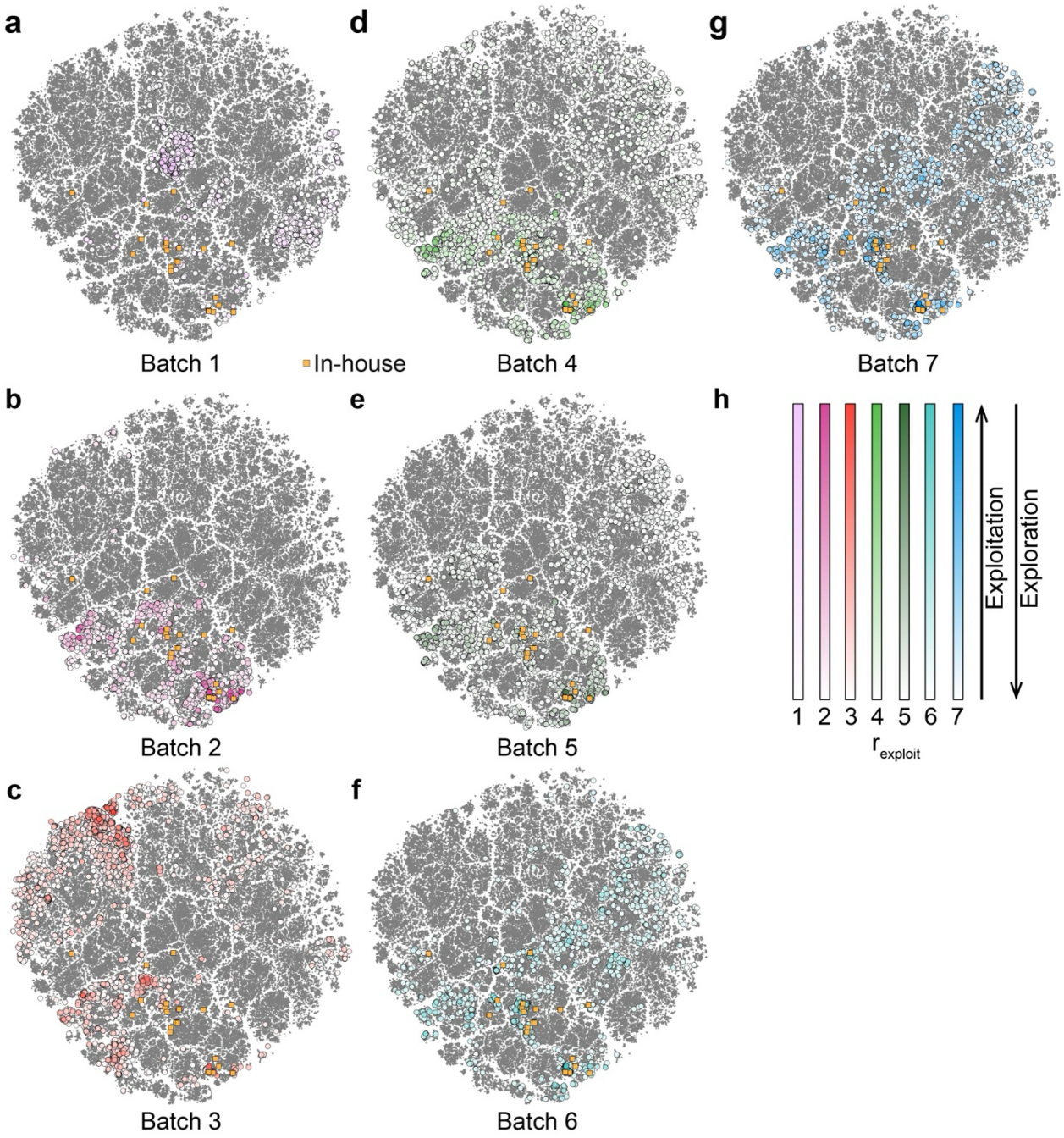


Figure S6. **Visualization of trends in exploration and exploitation ratios in different AL batches.** t-SNE plots showing exploitation ratios for **a**) batch 1; **b**) batch 2, **c**) batch 3, **d**) batch 4, **e**) batch 5, **f**) batch 6, and **g**) batch 7. Color bars depicting exploitation ratio ($r_{exploit}$) for each batch is shown in **h**).

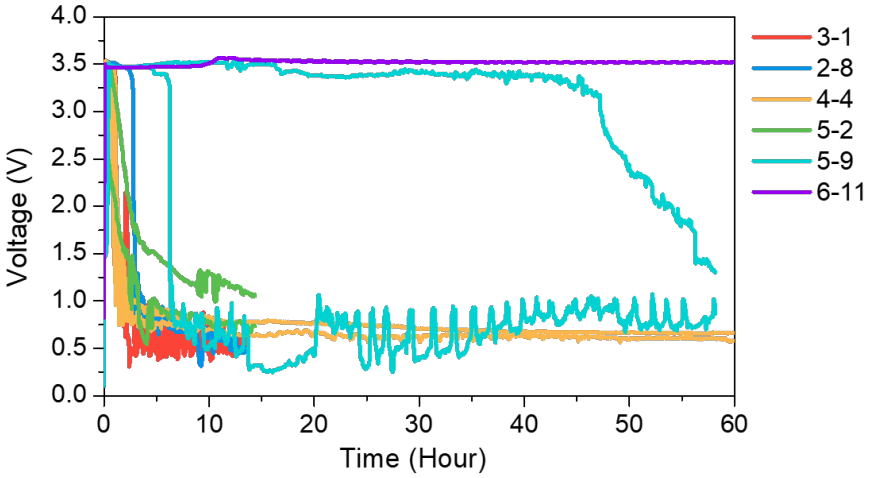


Figure S7. **Voltage profiles of some Cu||LFP cells that cannot finish the first charging.** Since such voltage profiles indicate either continuous parasitic reaction or short circuit, the performance of corresponding electrolytes is marked as 0.

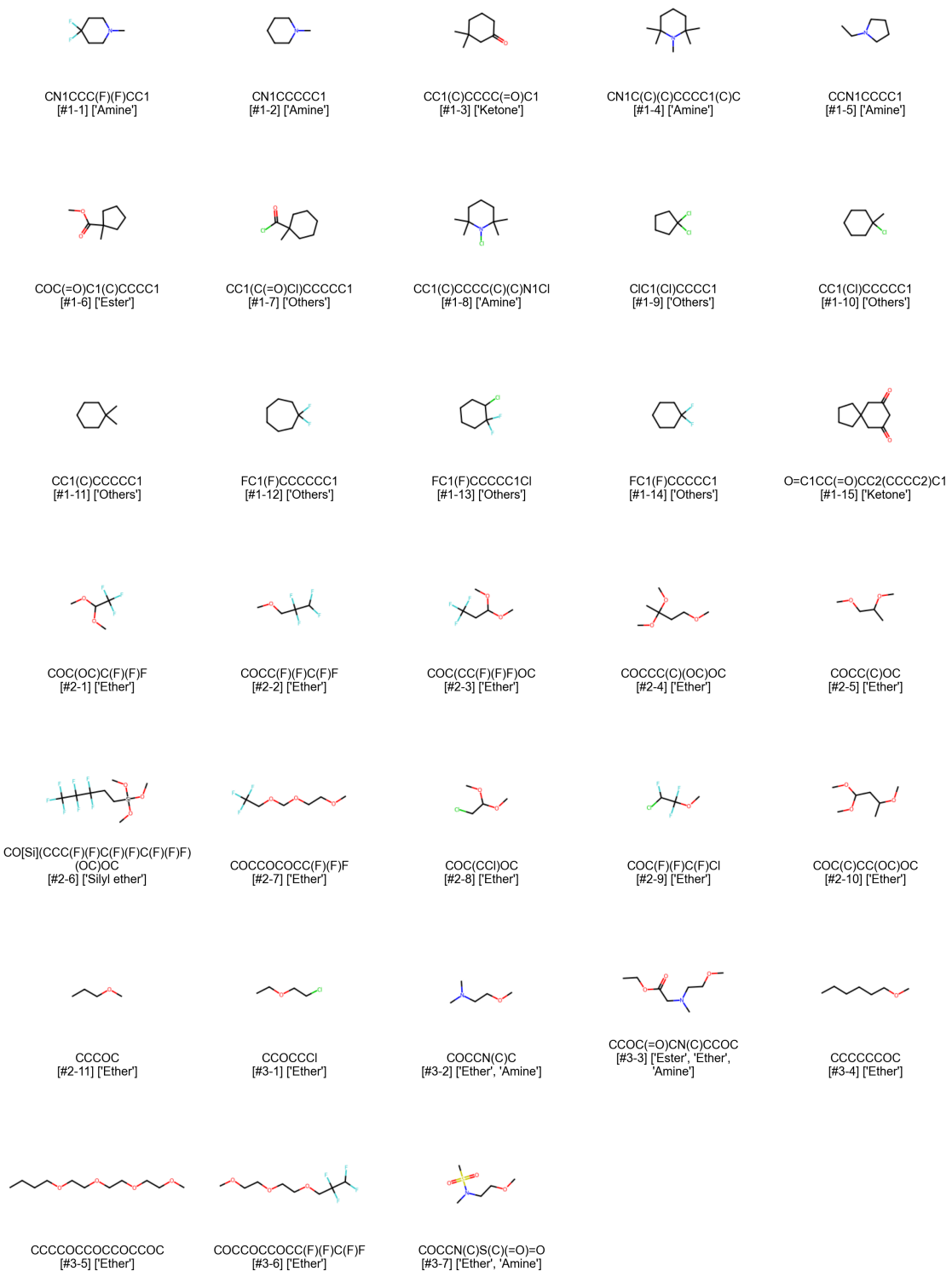


Figure S8. Visualization of molecular structures of electrolyte solvents tested in AL batches 1-3. Molecular structures of electrolyte solvent candidates selected from batches 1 to 3. First row shows SMILES representation, second row consists of ID number in the first parenthesis, and the functional groups to which the molecule belongs to, in the second parenthesis.

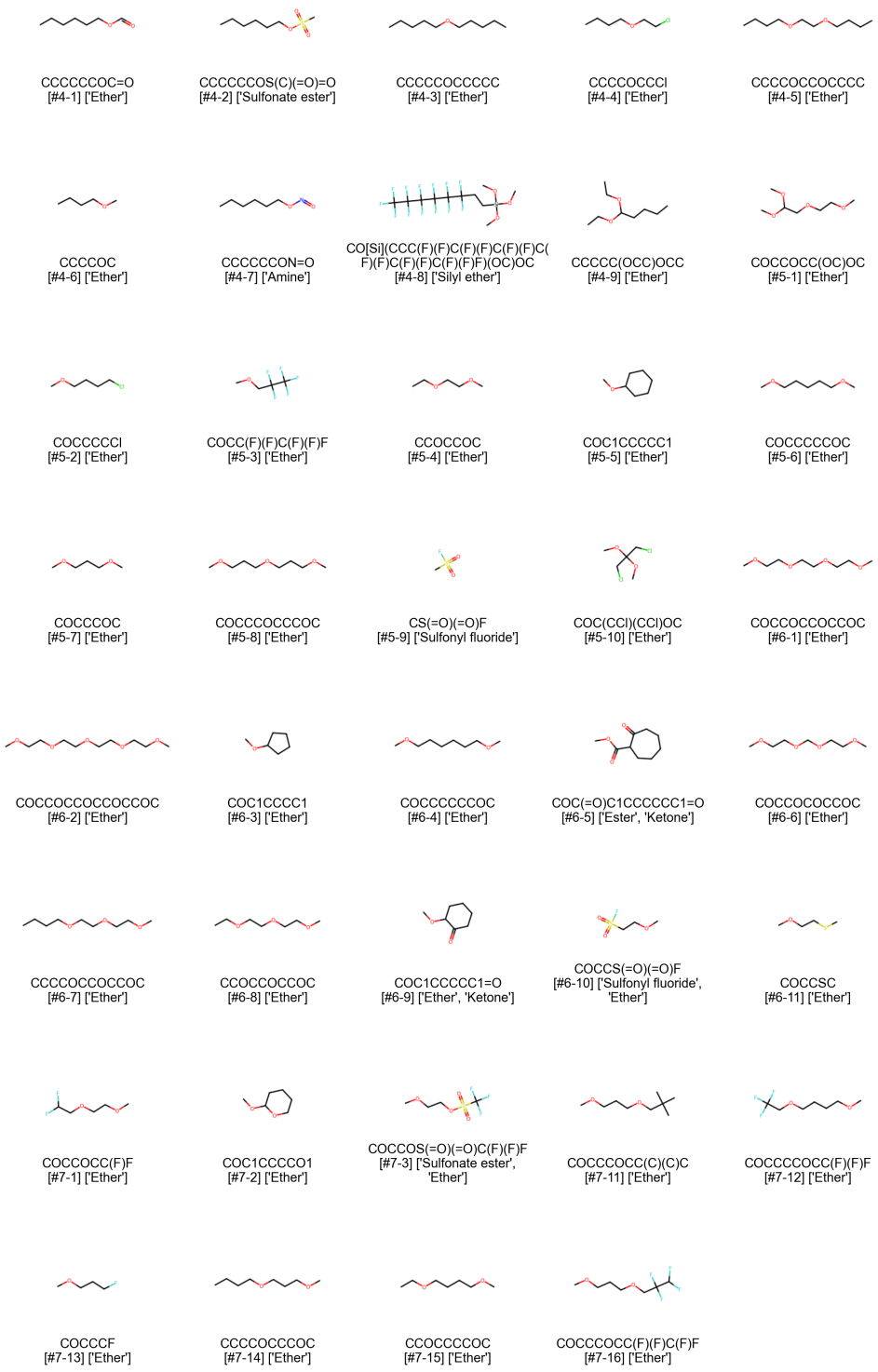


Figure S9. Visualization of molecular structures of electrolyte solvents tested in AL batches 4-7. Molecular structures of electrolyte solvent candidates selected from batches 4 to 7. First row shows SMILES representation, second row consists of ID number in the first parenthesis, and the functional groups to which the molecule belongs to, in the second parenthesis.

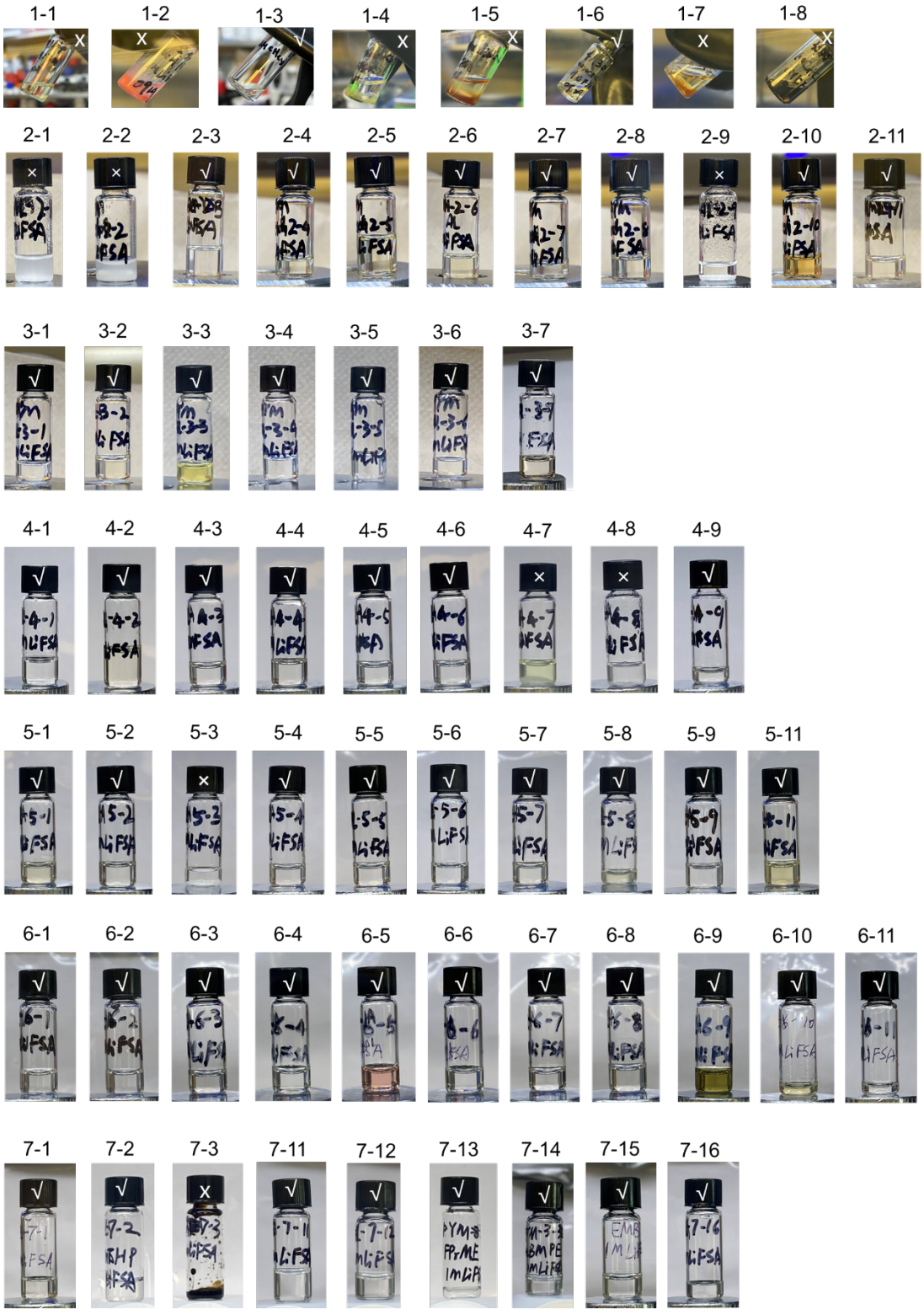


Figure S10. **Digital photos of 1 M LiFSA in tested electrolyte solvent candidates**. 5-10 is not shown because the compound itself is a solid at room temperature.

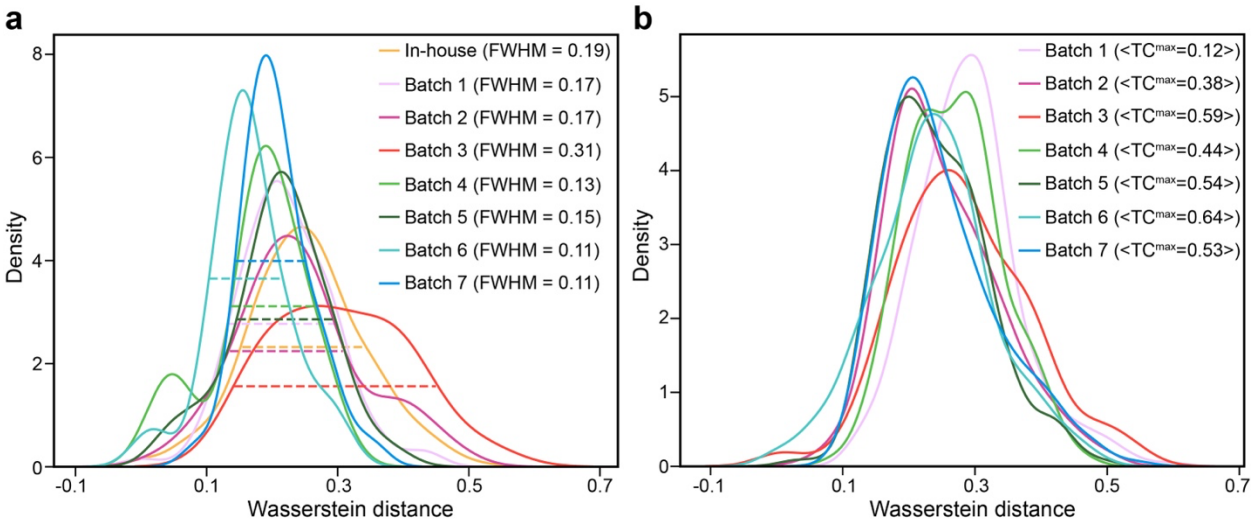


Figure S11. **Comparison of Wasserstein distance distributions among different batches of AL acquisitions.** Density of Wasserstein distance distribution for **a**) solvent molecules within each batch and **b**) solvent molecules in different batches with respect to in-house dataset. $< TC^{max} >$ = average of maximum Tanimoto coefficient values with respect to electrolyte solvent molecules in the in-house dataset.

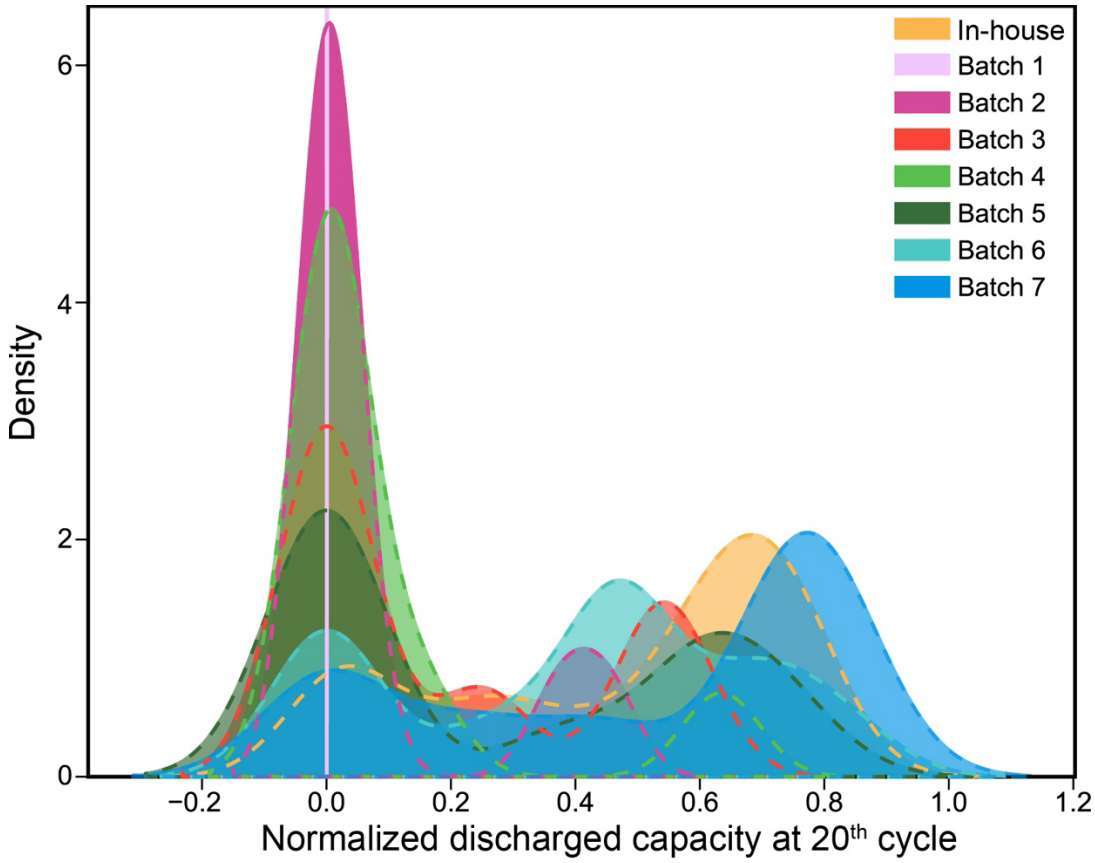


Figure S12. **Comparison of discharge capacity distribution among different batches of AL acquisitions.** Density of normalized discharge capacity (at 20th cycle) distribution of electrolytes within each acquired batch including the in-house dataset.

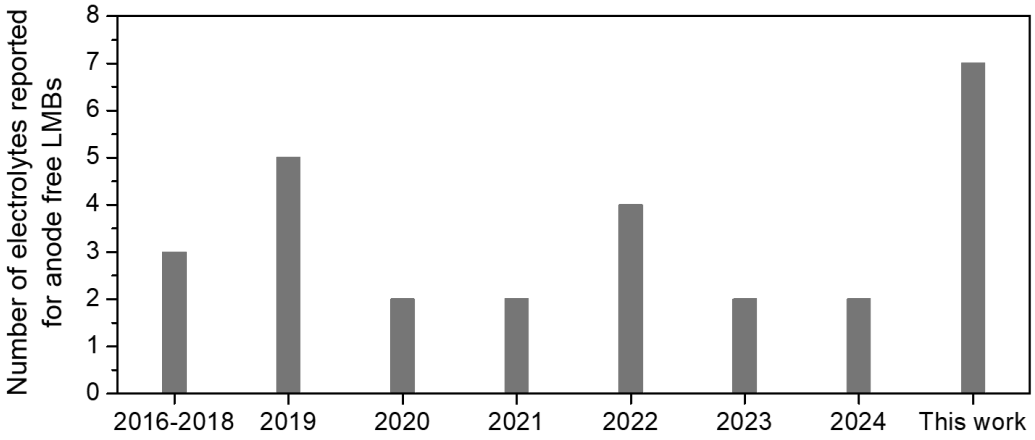


Figure S13. Number of electrolytes reported for anode-free LMBs in recent years. Seven unique new electrolytes (5-7, 6-3, 7-1, 7-12, 7-14, 7-15, 7-16) are discovered in this work to enable good performance in anode-free LMBs, which is more than multiple combined studies from any previous year.

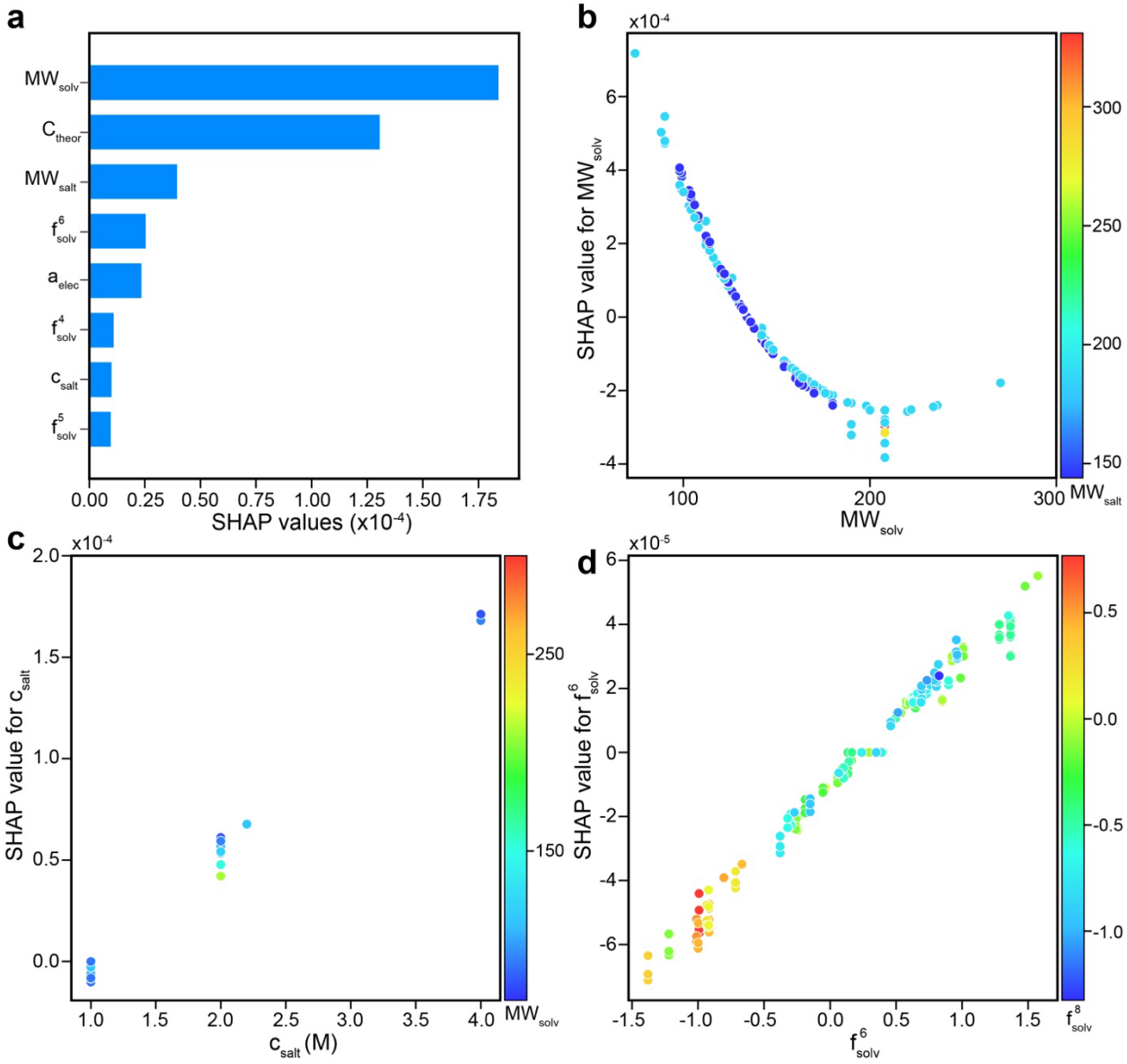


Figure S14. **SHAP analyses for Rational Quadratic surrogate model from batch 7. a**) SHAP summary bar plot for C_{norm}^{20} prediction corresponding to RQ surrogate model obtained from batch 7. SHAP dependence plots for **b**) solvent molecular weight (MW_{solv}), **c**) salt concentration (c_{salt}), and **d**) solvent fingerprint-6 (f_{solv}^{6}) features.

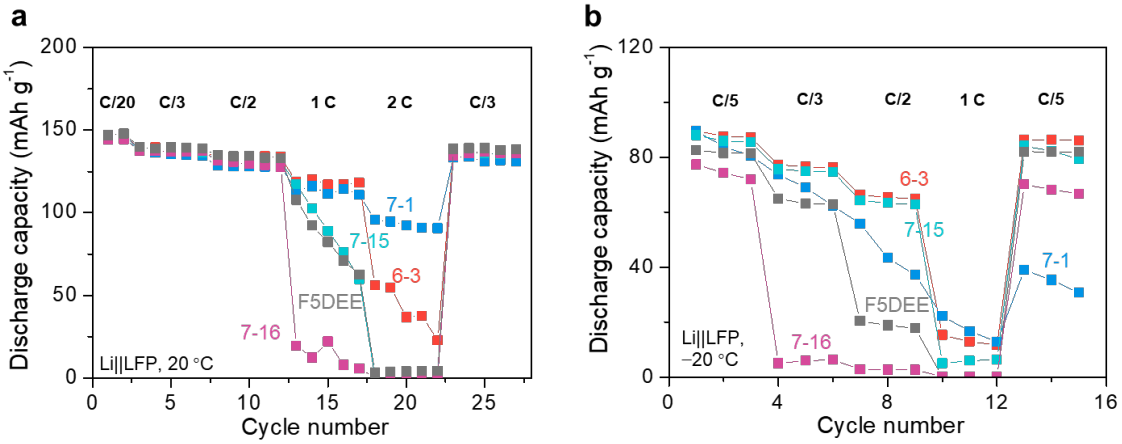


Figure S15. **Rate capability and low temperature performance. a**) Rate capability test of AL discovered electrolytes and F5DEE in Li||LFP cells at 20 °C. **b**) Low temperature rate capability test of AL discovered electrolytes and F5DEE in Li||LFP cells at −20 °C. Two formation cycles at C/10 were performed prior to low temperature test and was not shown in (b). 1 C = 150 mA/g.

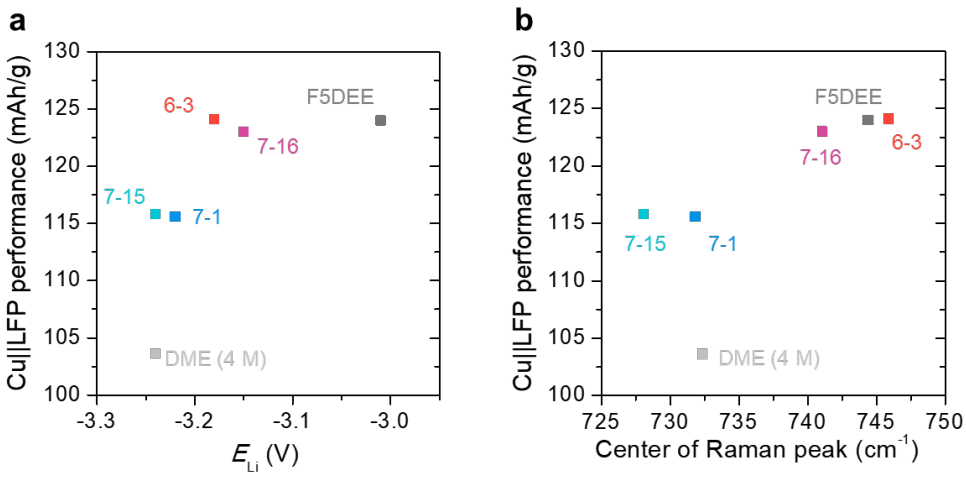


Figure S16. **Correlation of ion solvation structure to anode-free LMB cycling performance.** **a**) Discharge capacity at 20th cycle as a function of lithium redox potential (E_{Li}). **b**) Discharge capacity at 20th cycle as a function of gravity center of FSA Raman peak. Ion solvation structure does not show clear correlation among the best performing electrolytes especially for 6-3, 7-16 and F5DEE.

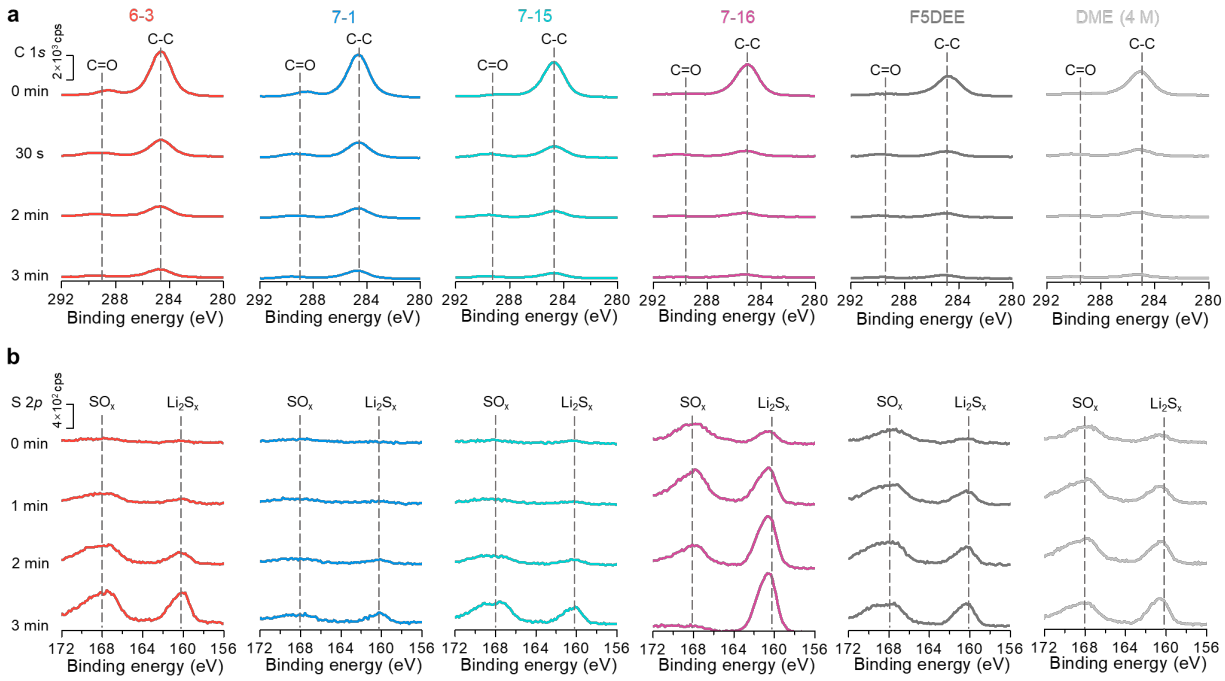


Figure S17. **XPS spectra of lithium metal SEI at different etching time. a**) C 1*s* spectra. **b**) S 2*p* spectra.

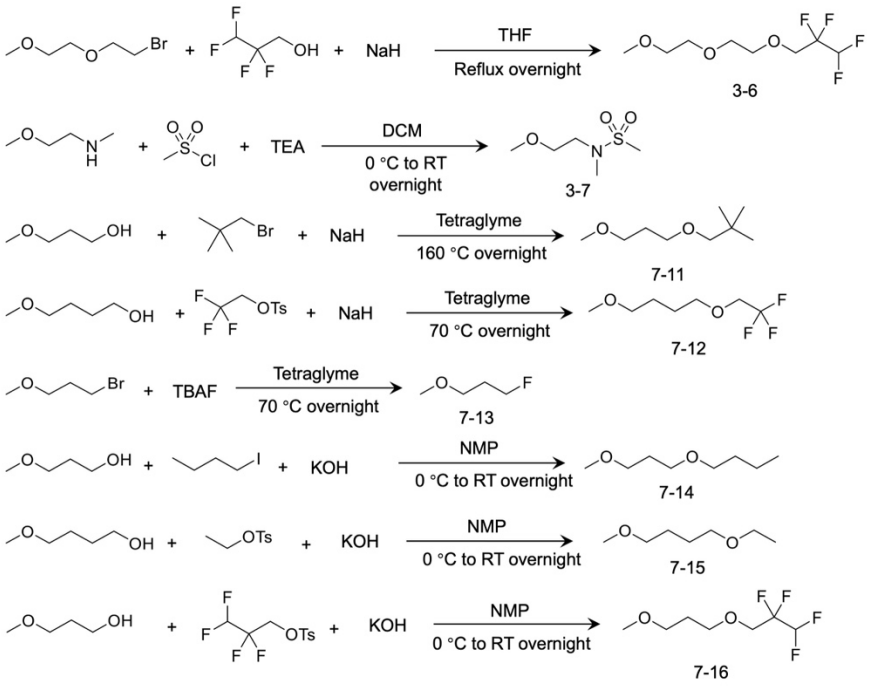
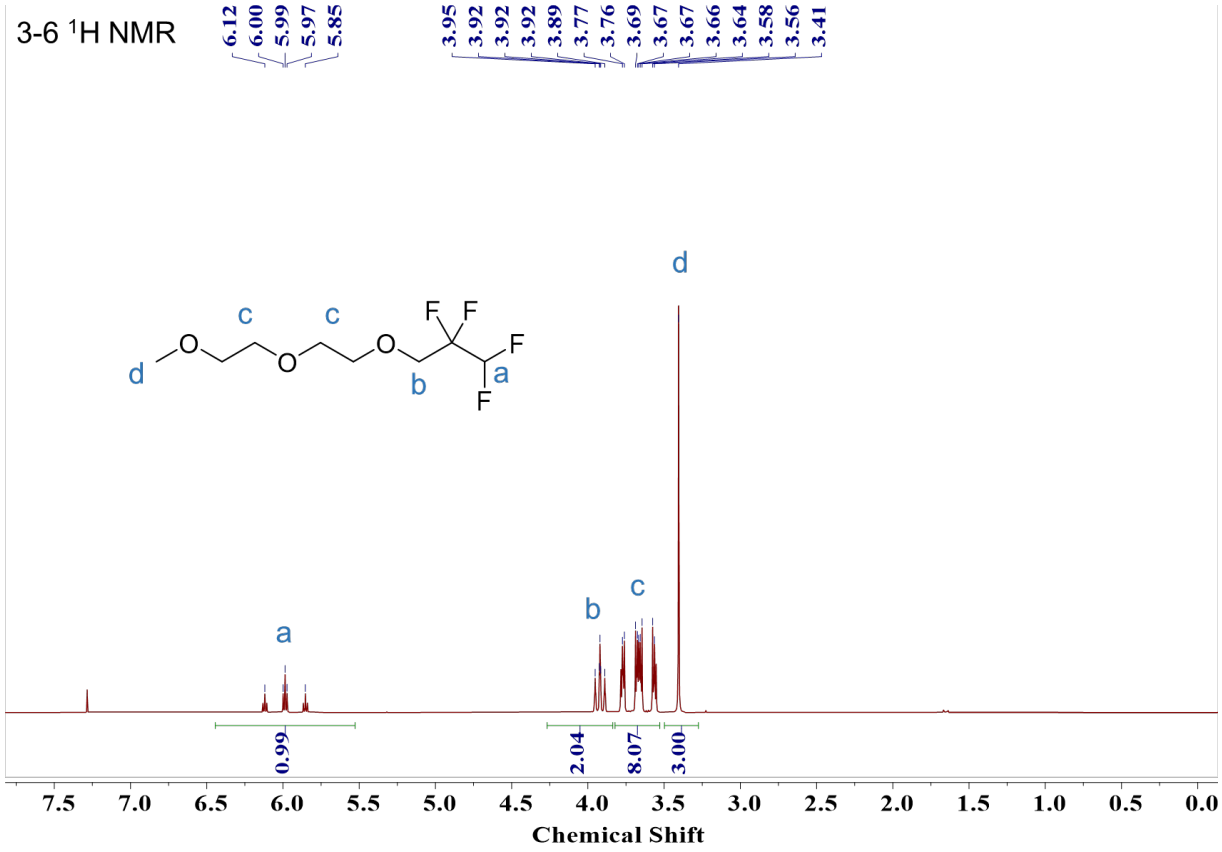
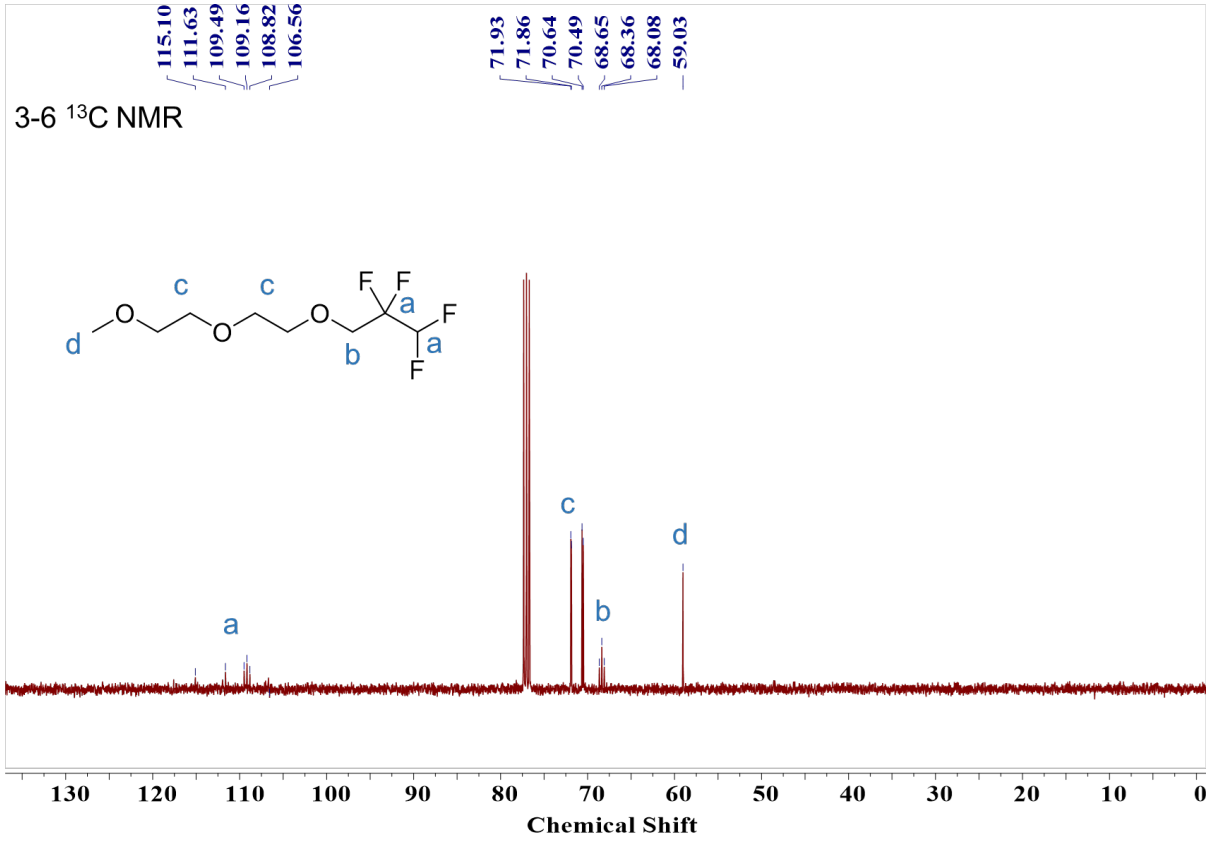
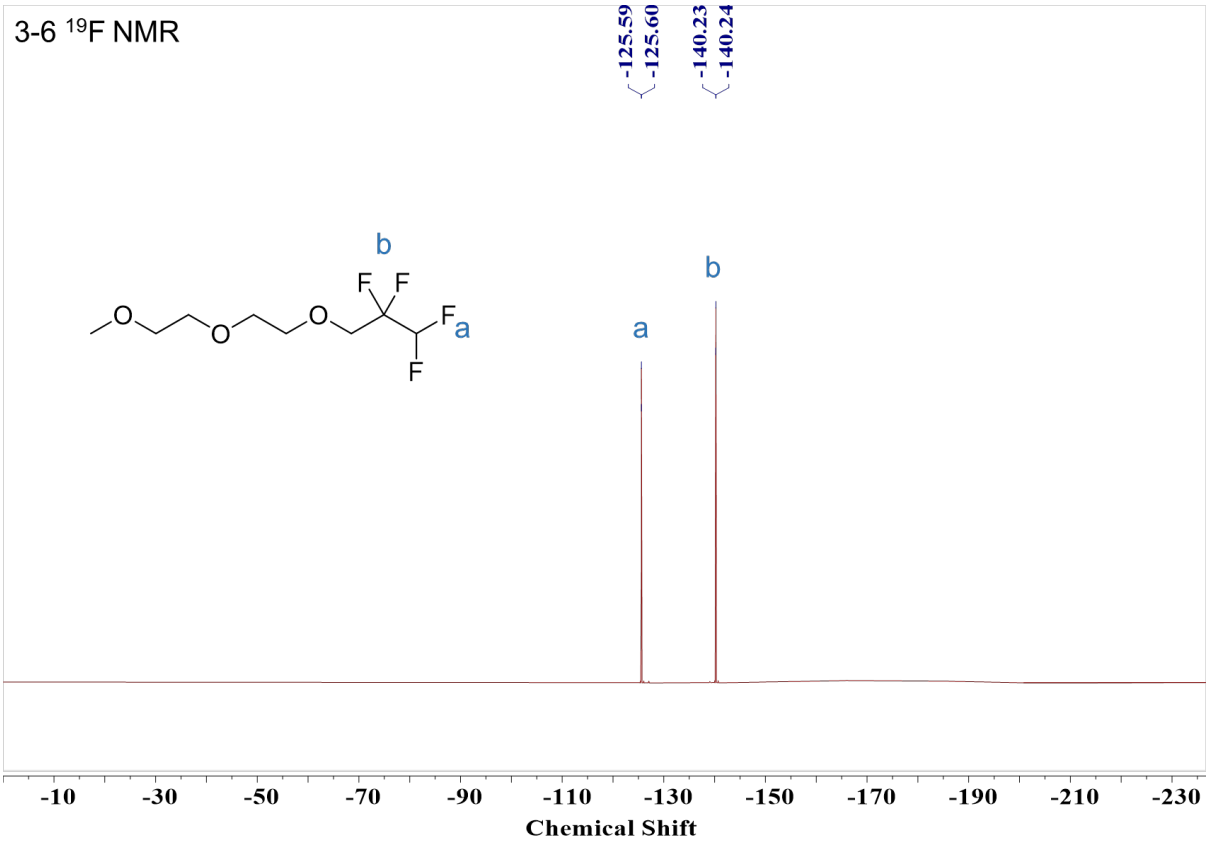


Figure S18. **Synthetic route of synthesized compounds.**

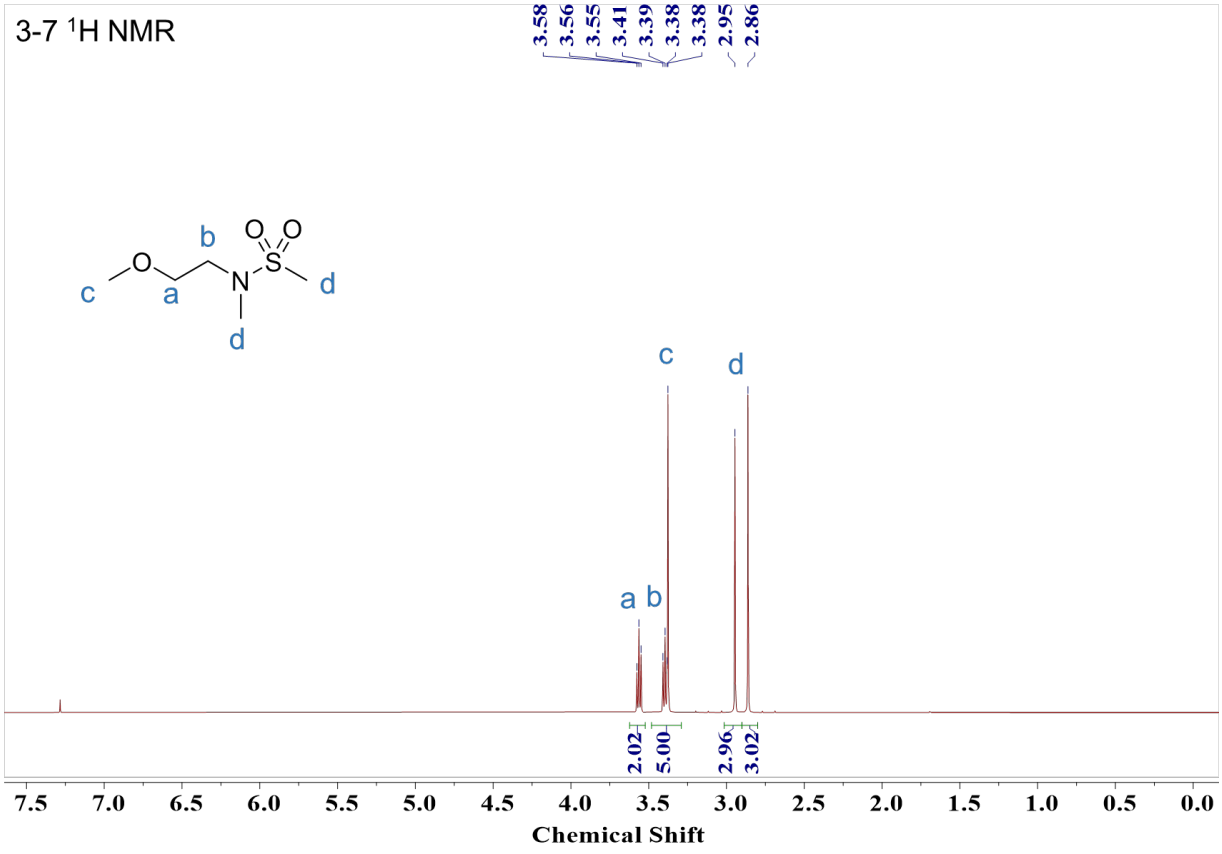

3-6 1H NMR
6.12
6.00
5.99
5.97
5.85
3.95
3.92
3.92
3.92
3.89
3.77
3.76
3.69
3.67
3.67
3.66
3.64
3.58
3.56
3.41
d
c
c
F
F
O
O
O
F
d
b
a
F
a
b
c
d
0.99
2.04
8.07
3.00
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Chemical Shift

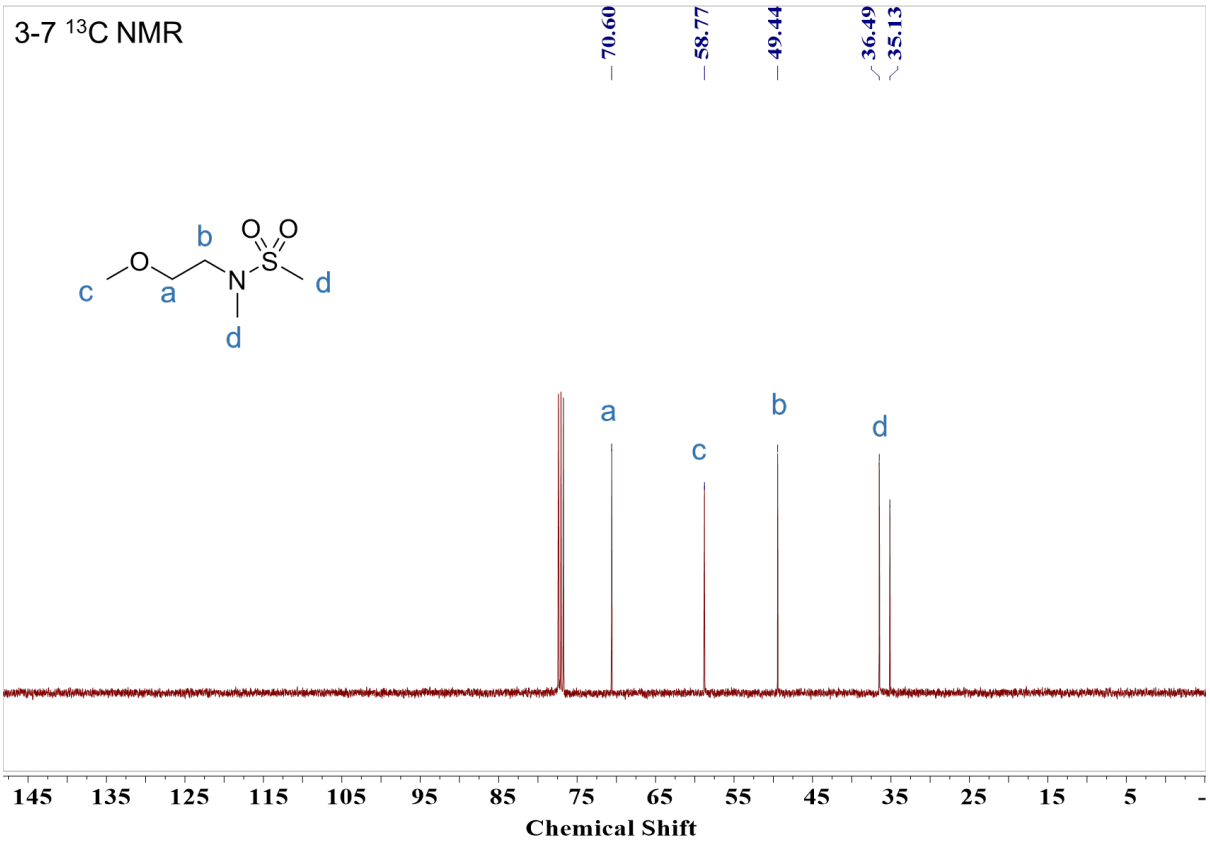

115.10
111.63
109.49
109.16
108.82
106.56
71.93
71.86
70.64
70.49
68.65
68.36
68.08
59.03
3-6 13C NMR
c
c
F
F
a
F
O
O
O
d
b
a
F
c
d
b
a
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Chemical Shift

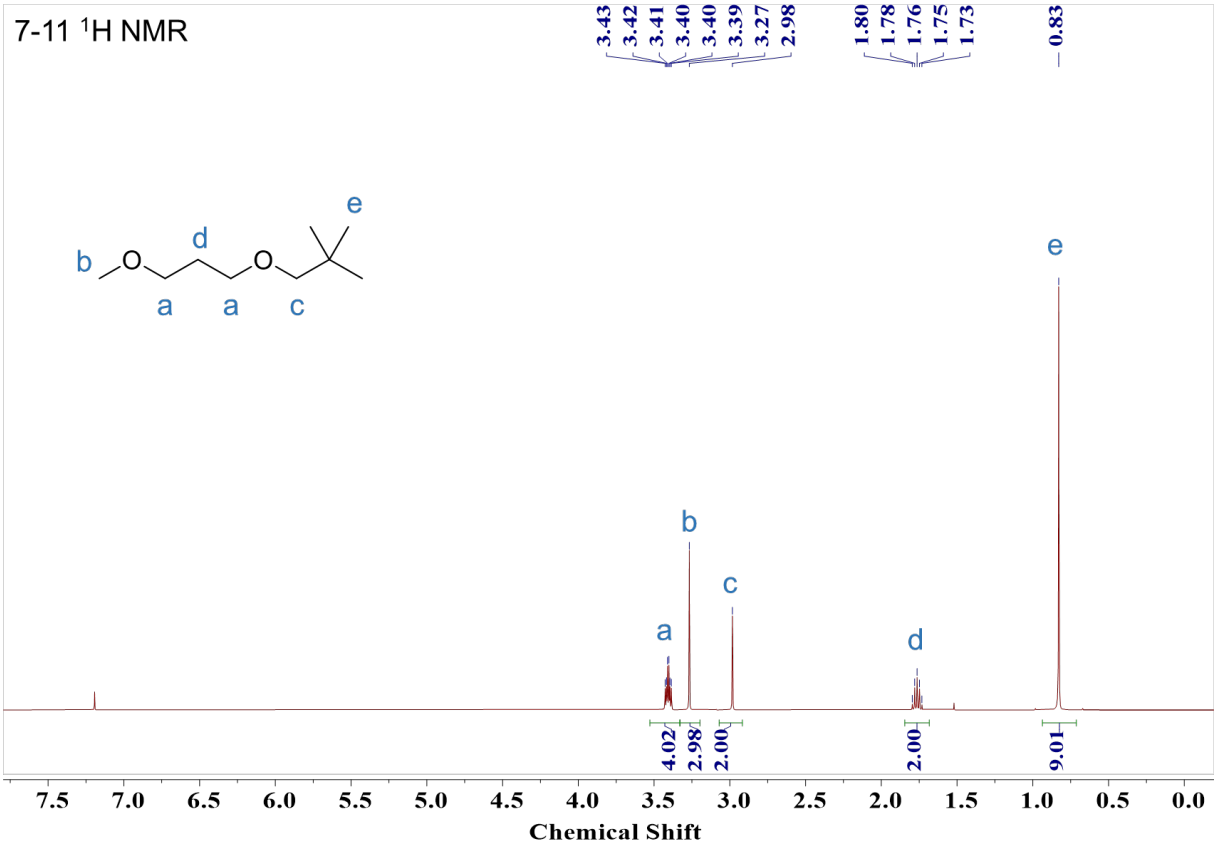
3-6 ^{19}F NMR

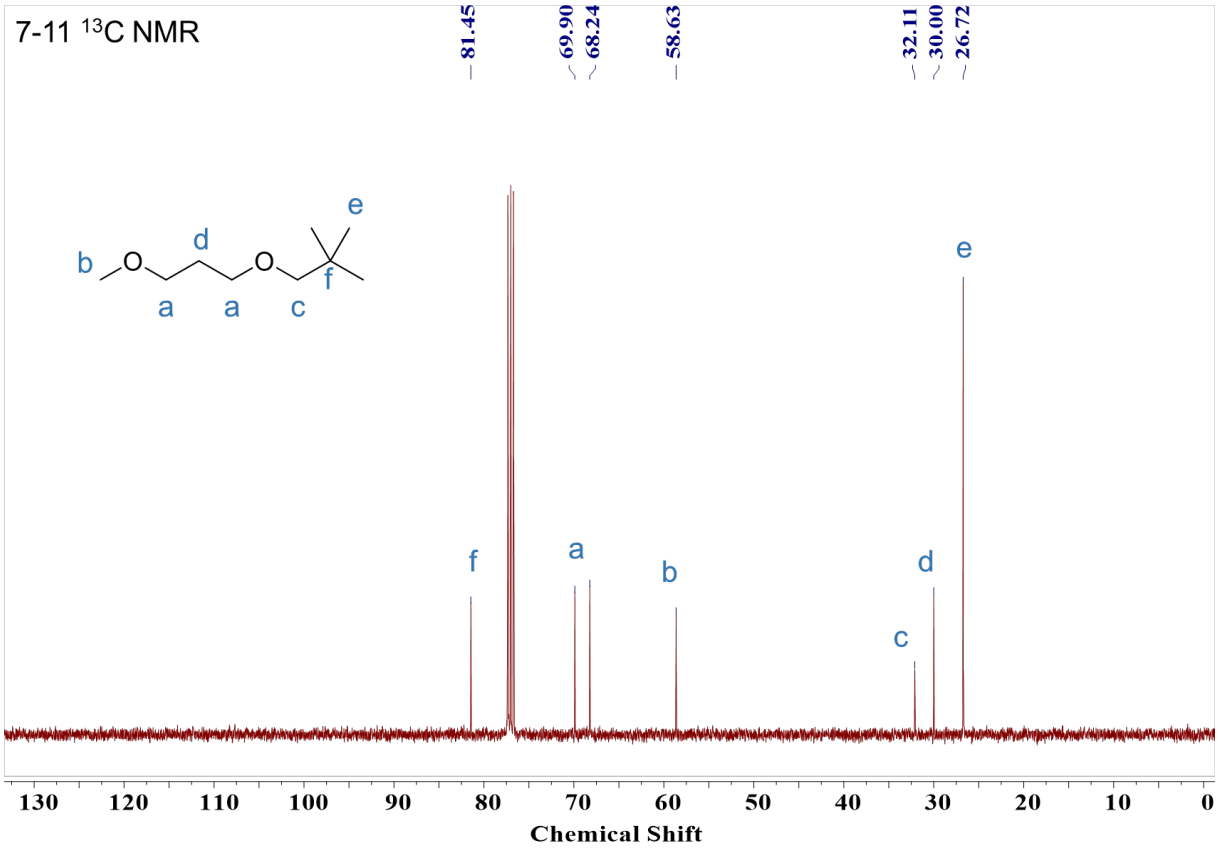


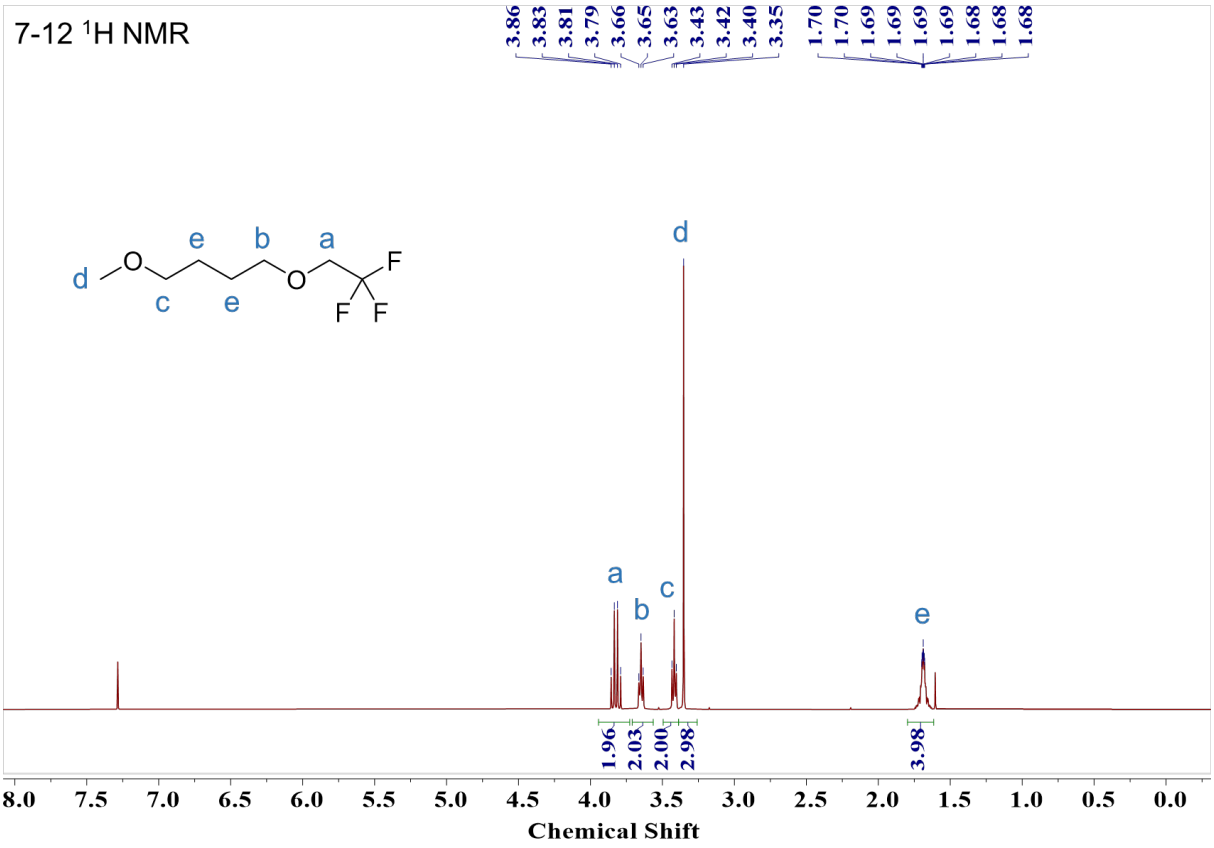
3-7 ^{1}H NMR

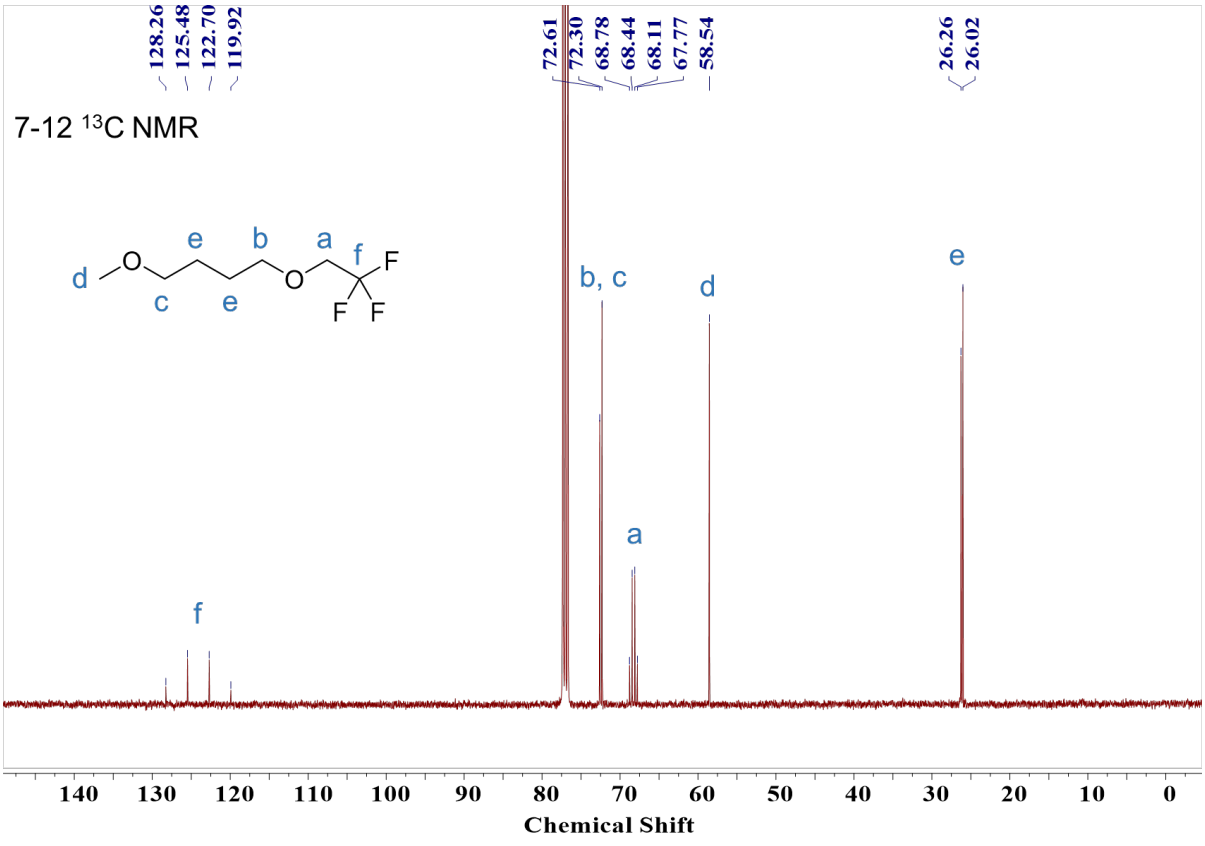



3-7 13C NMR
70.60
58.77
49.44
36.49
35.13
a
b
c
d
Chemical Shift

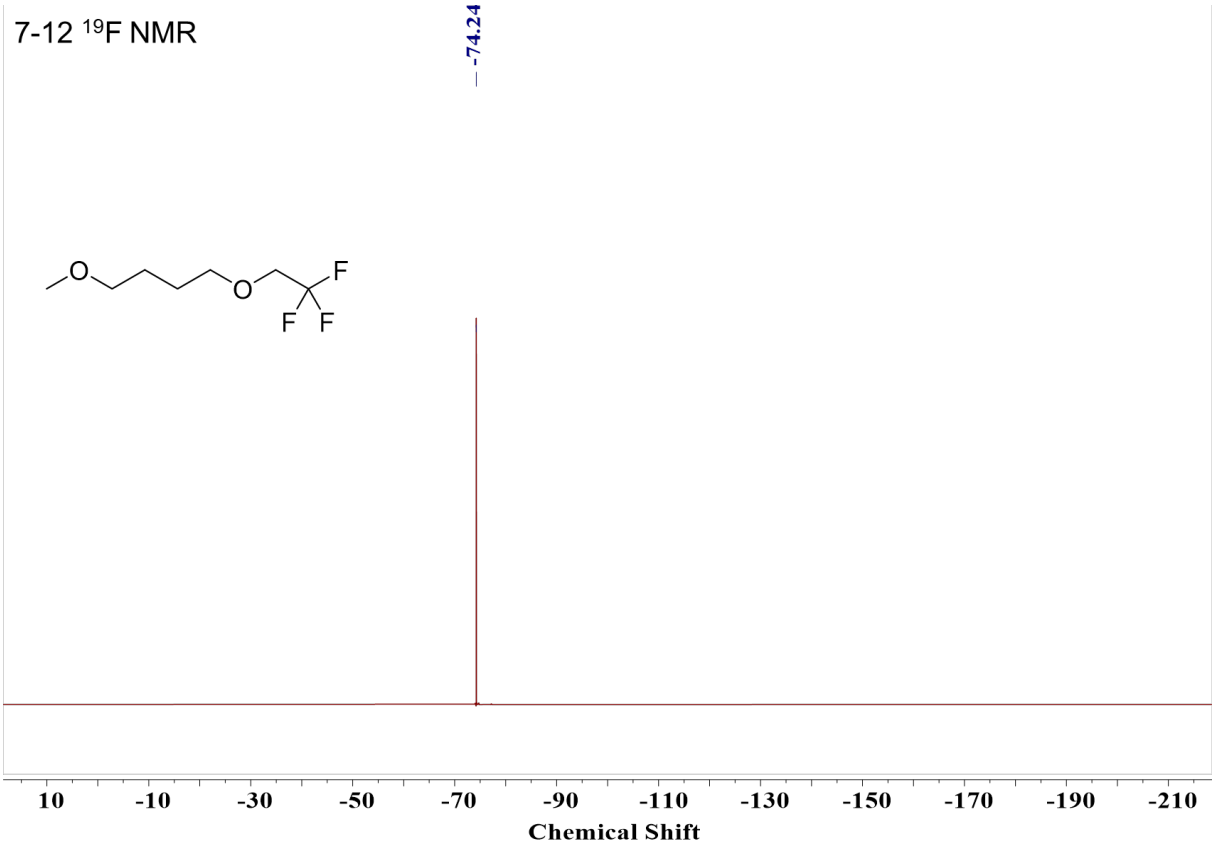

7-11 1H NMR
3.43
3.42
3.41
3.40
3.40
3.39
3.27
2.98
1.80
1.78
1.76
1.75
1.73
0.83
4.02
2.98
2.00
2.00
9.01
Chemical Shift


7-11 13C NMR
81.45
69.90
68.24
58.63
32.11
30.00
26.72
b
O
d
O
e
f
a
a
c
f
a
b
c
d
e
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Chemical Shift

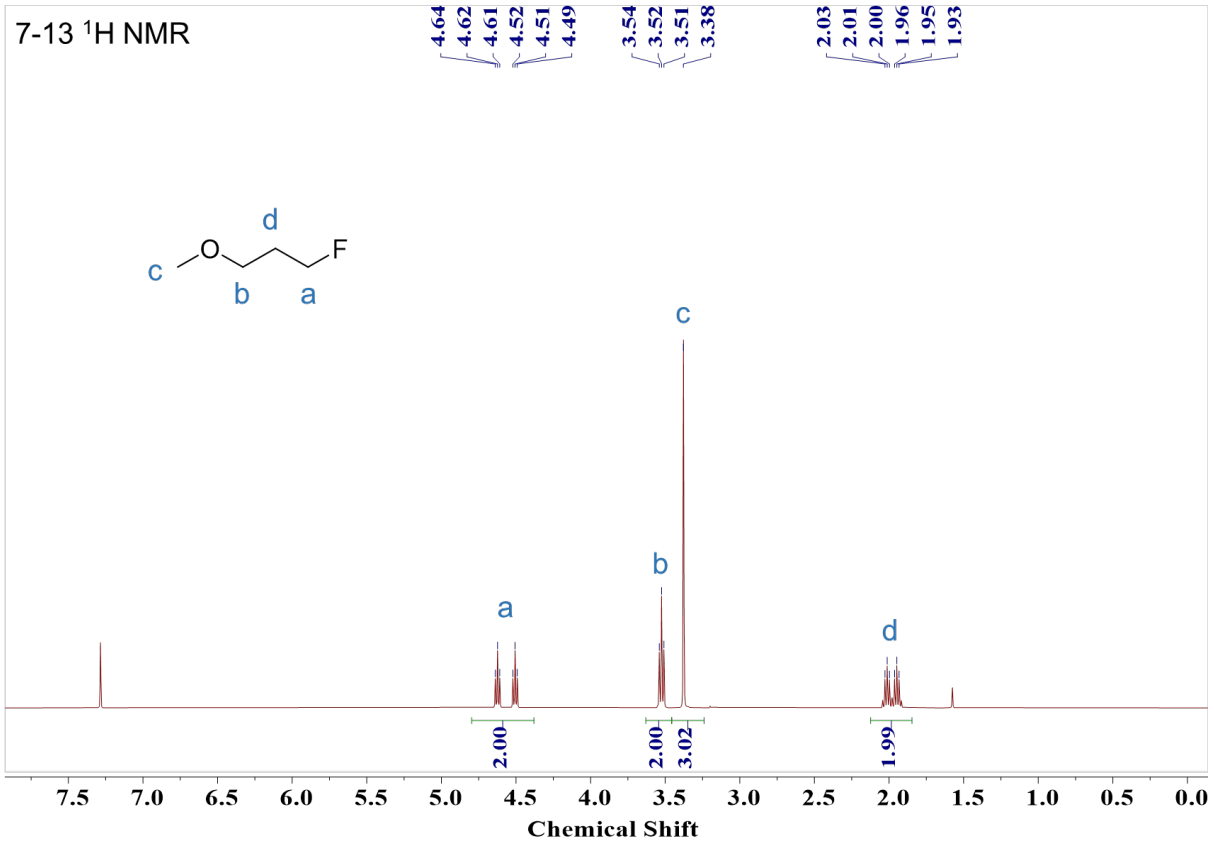

7-12 1H NMR
3.86
3.83
3.81
3.79
3.66
3.65
3.63
3.43
3.42
3.40
3.35
1.70
1.70
1.69
1.69
1.69
1.69
1.68
1.68
1.68
d
e
b
a
F
O
c
e
F
F
d
a
b
c
e
1.96
2.03
2.00
2.98
3.98
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Chemical Shift

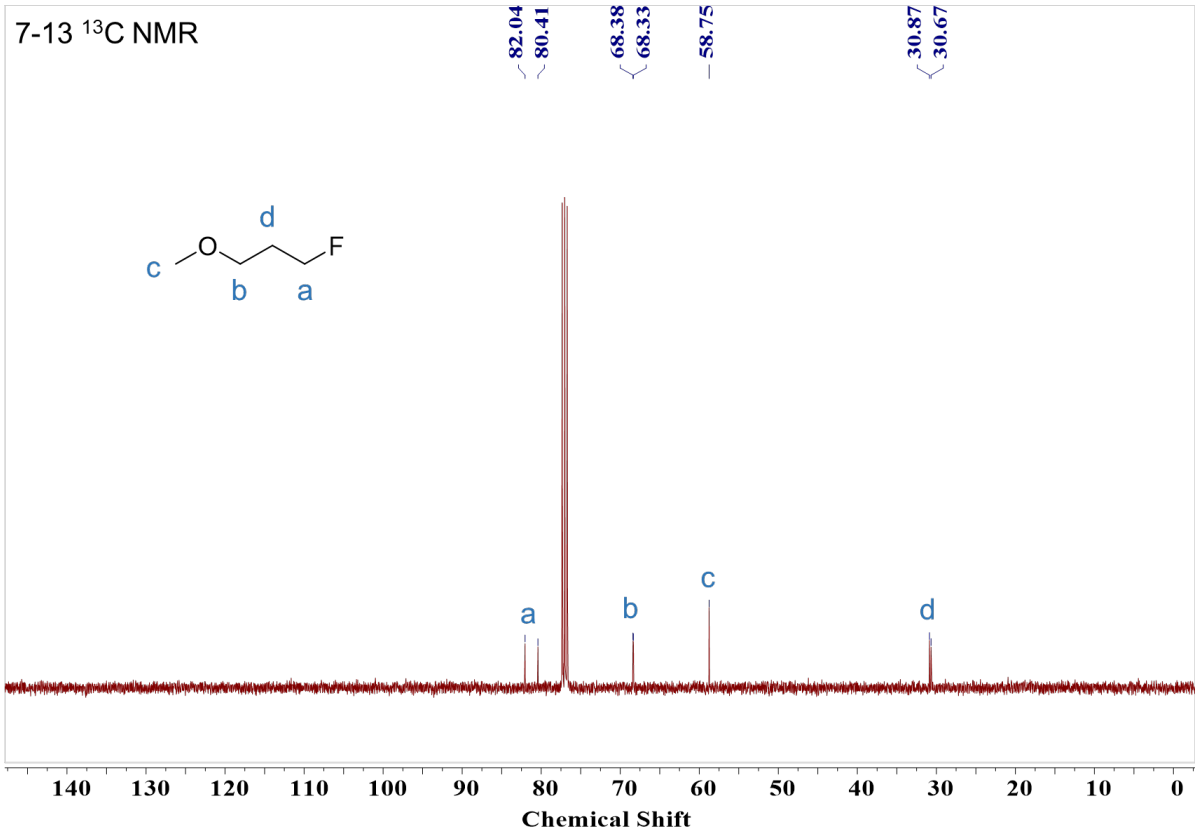

128.26
125.48
122.70
119.92
72.61
72.30
68.78
68.44
68.11
67.77
58.54
26.26
26.02
7-12 13C NMR
d
e
b
a
f
F
O
c
e
F
F
f
a
b, c
d
e
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Chemical Shift

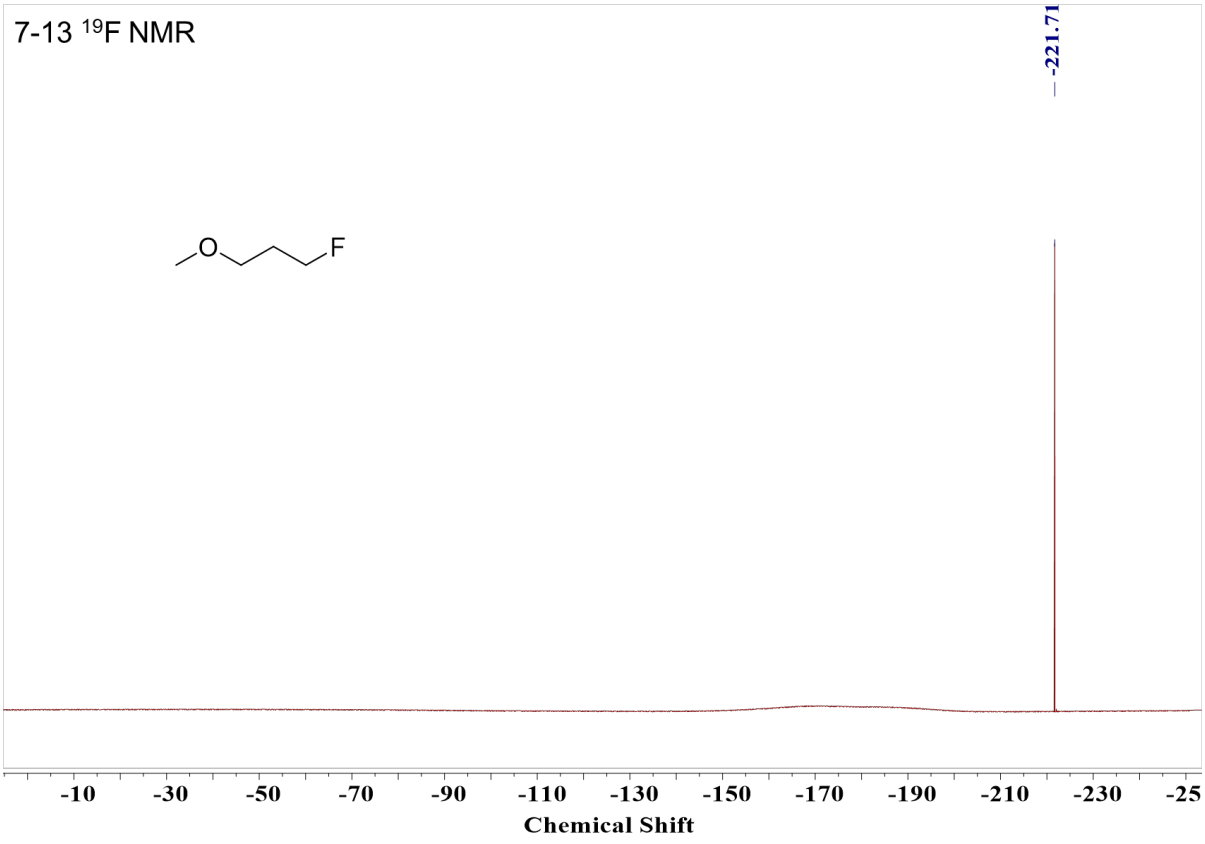
7-12 ^{19}F NMR

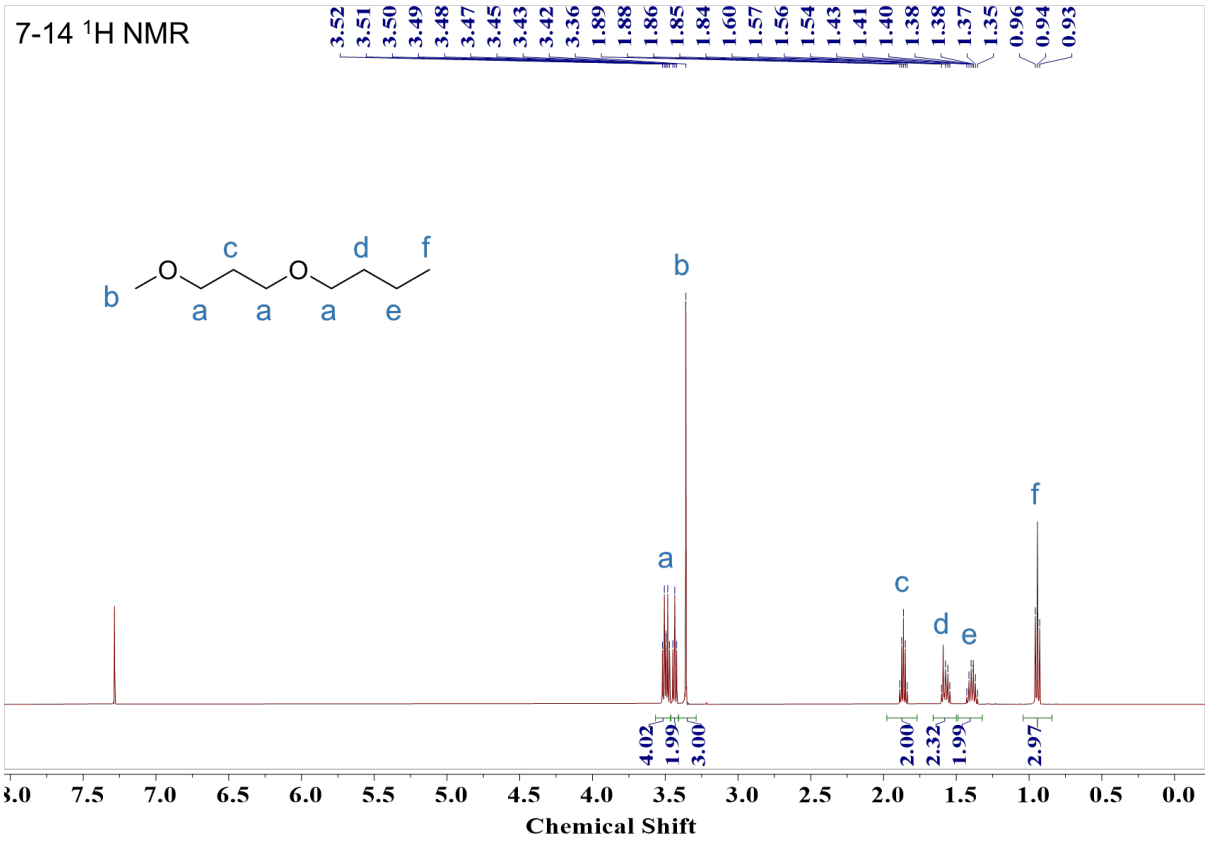


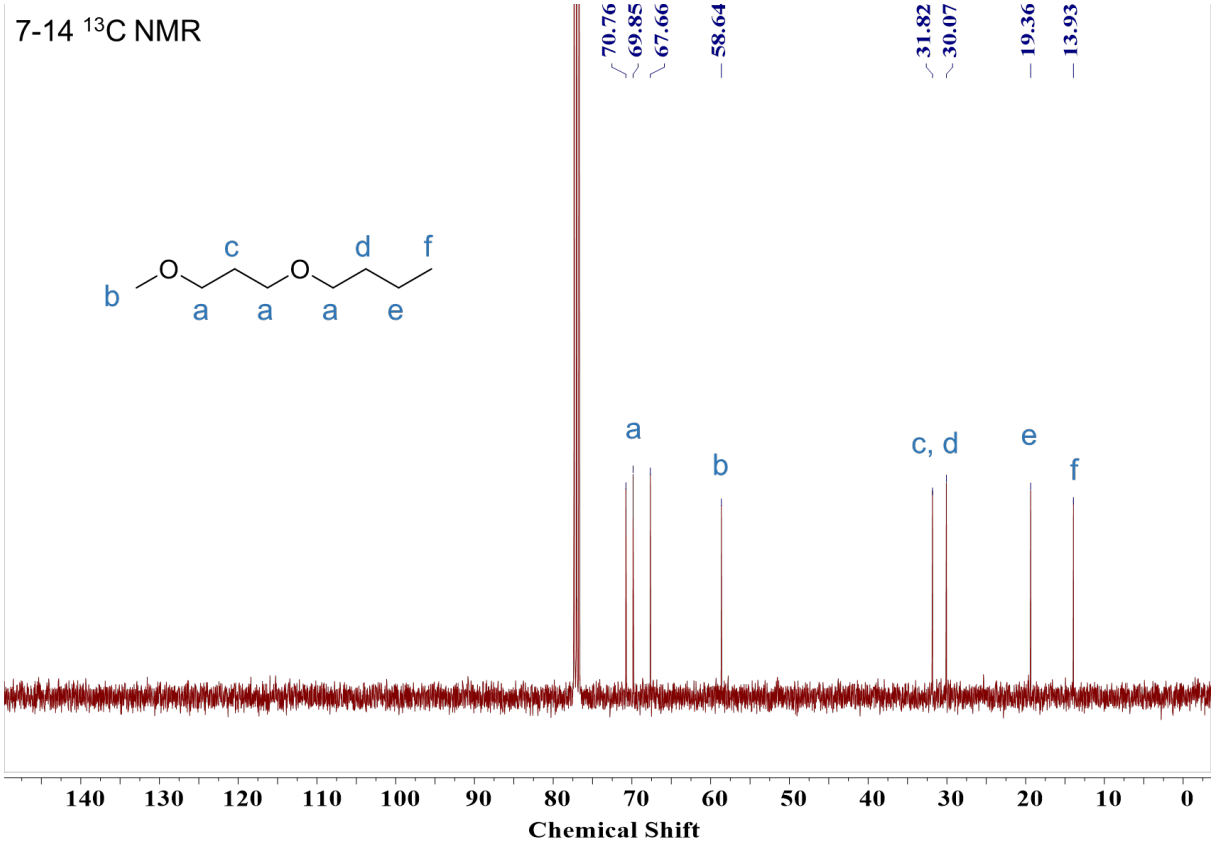
7-13 ^{1}H NMR

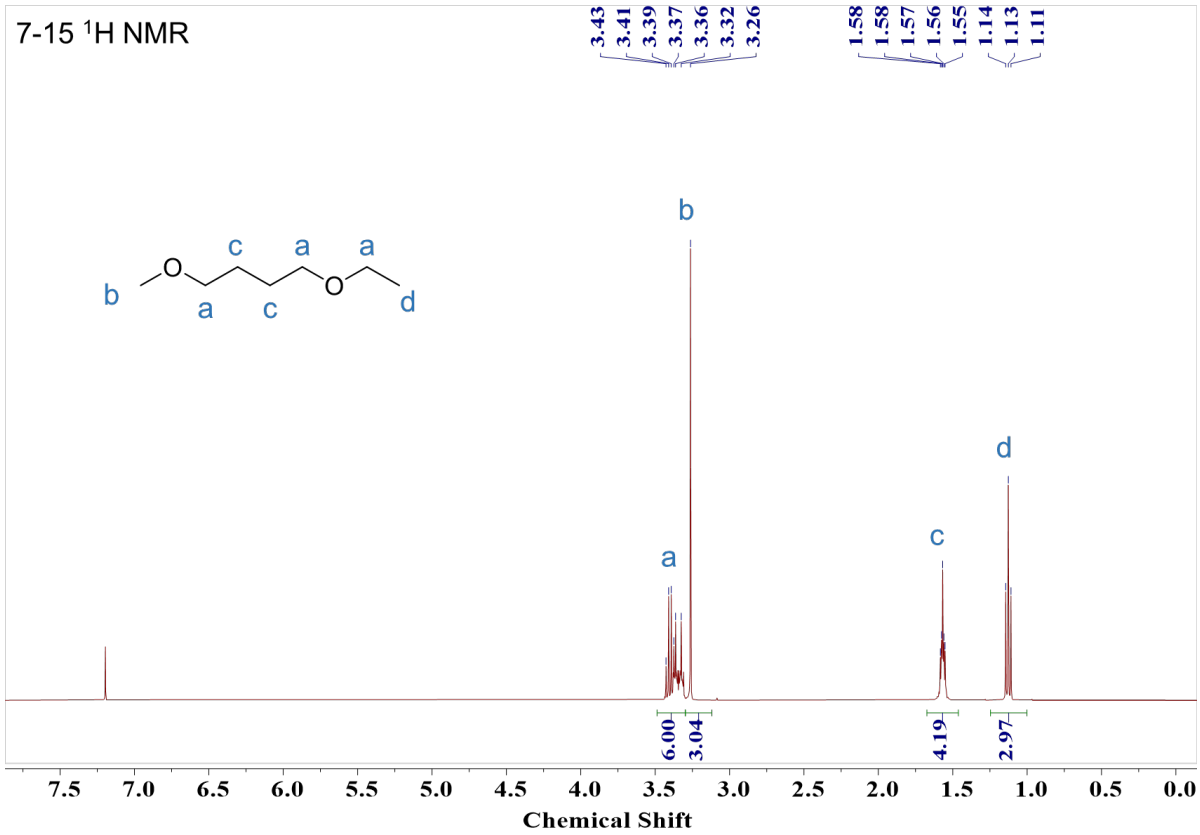


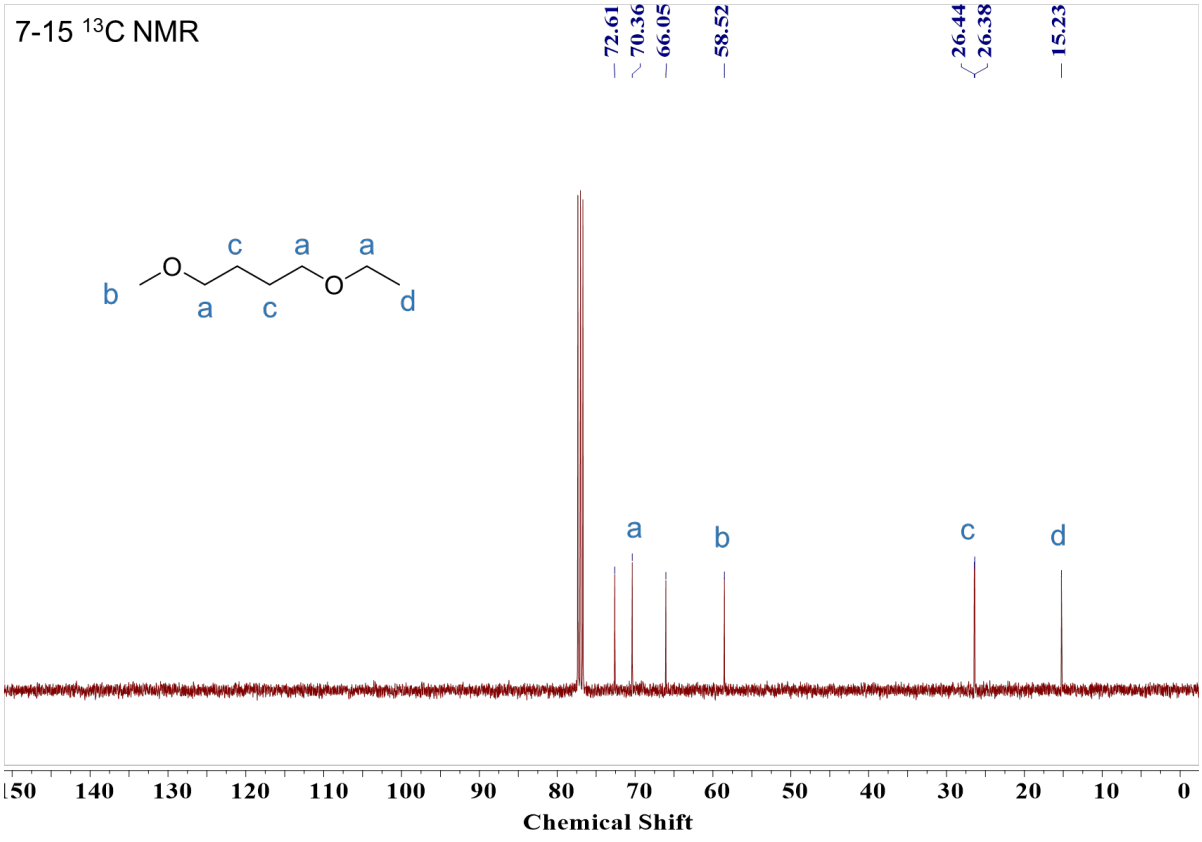

7-13 13C NMR
82.04
80.41
68.38
68.33
58.75
30.87
30.67
d
c
O
F
b
a
a
b
c
d
140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
Chemical Shift

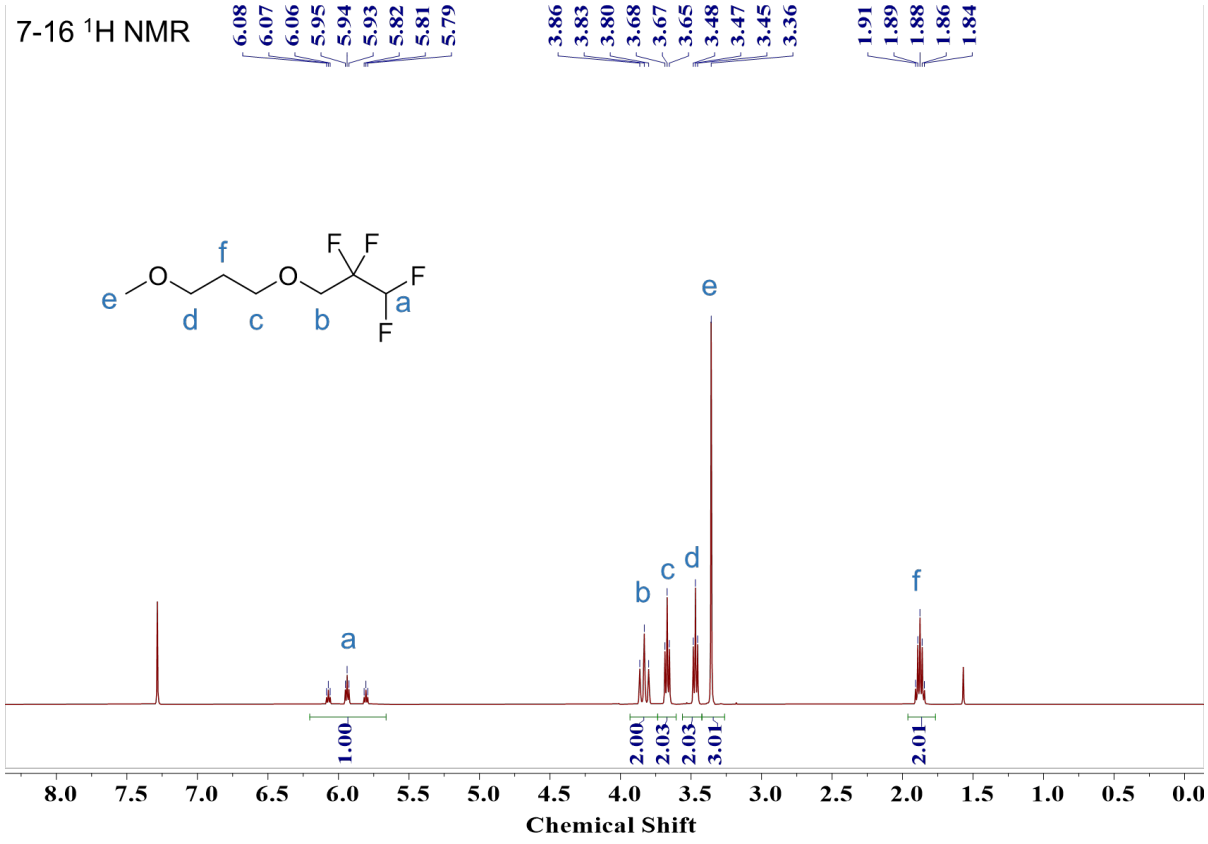

7-13 19F NMR
-221.71
O
F
-10 -30 -50 -70 -90 -110 -130 -150 -170 -190 -210 -230 -25
Chemical Shift

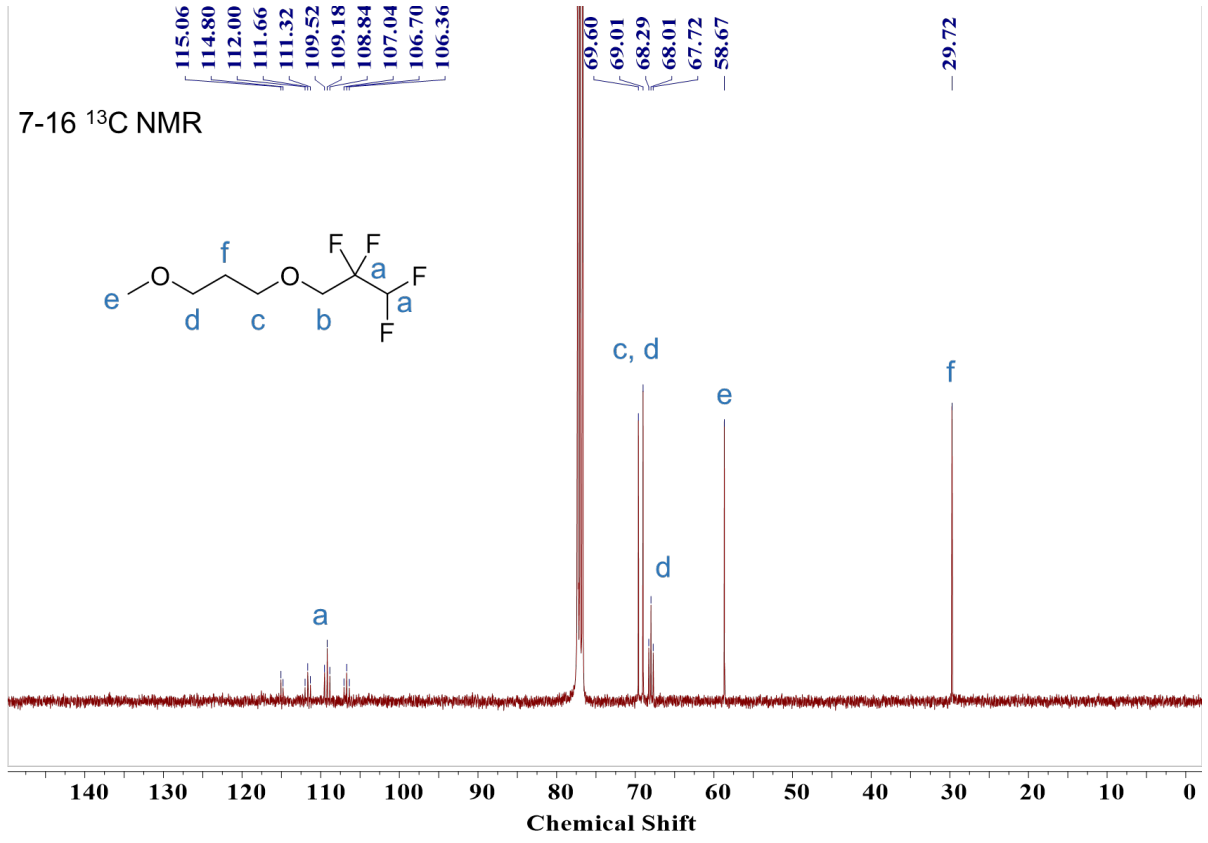

7-14 1H NMR
3.52
3.51
3.50
3.49
3.48
3.47
3.45
3.43
3.42
3.36
1.89
1.88
1.86
1.85
1.84
1.60
1.57
1.56
1.54
1.43
1.41
1.40
1.38
1.38
1.37
1.35
0.96
0.94
0.93
b
c
d
f
O
O
a
a
a
e
a
b
c
d
e
f
4.02
1.99
3.00
2.00
2.32
1.99
2.97
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Chemical Shift


7-14 13C NMR
70.76
69.85
67.66
58.64
31.82
30.07
19.36
13.93
b
c
d
f
O
O
a
a
a
e
a
b
c, d
e
f
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Chemical Shift


7-15 1H NMR
3.43
3.41
3.39
3.37
3.36
3.32
3.26
1.58
1.58
1.57
1.56
1.55
1.14
1.13
1.11
b
c
a
a
b
a
c
d
O
O
a
b
c
d
6.00
3.04
4.19
2.97
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Chemical Shift


7-15 13C NMR
72.61
70.36
66.05
58.52
26.44
26.38
15.23
b
c
a
a
b
a
c
d
O
O
a
b
c
d
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Chemical Shift


7-16 1H NMR
6.08
6.07
6.06
5.95
5.94
5.93
5.82
5.81
5.79
3.86
3.83
3.80
3.68
3.67
3.65
3.48
3.47
3.45
3.36
1.91
1.89
1.88
1.86
1.84
e
f
d
c
b
a
F
F
F
F
O
O
a
b
c
d
e
f
1.00
2.00
2.03
2.03
3.01
2.01
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Chemical Shift


115.06
114.80
112.00
111.66
111.32
109.52
109.18
108.84
107.04
106.70
106.36
69.60
69.01
68.29
68.01
67.72
58.67
29.72
7-16 13C NMR
e
f
d
c
b
a
F
F
F
F
O
O
a
c, d
e
f
d
a
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Chemical Shift

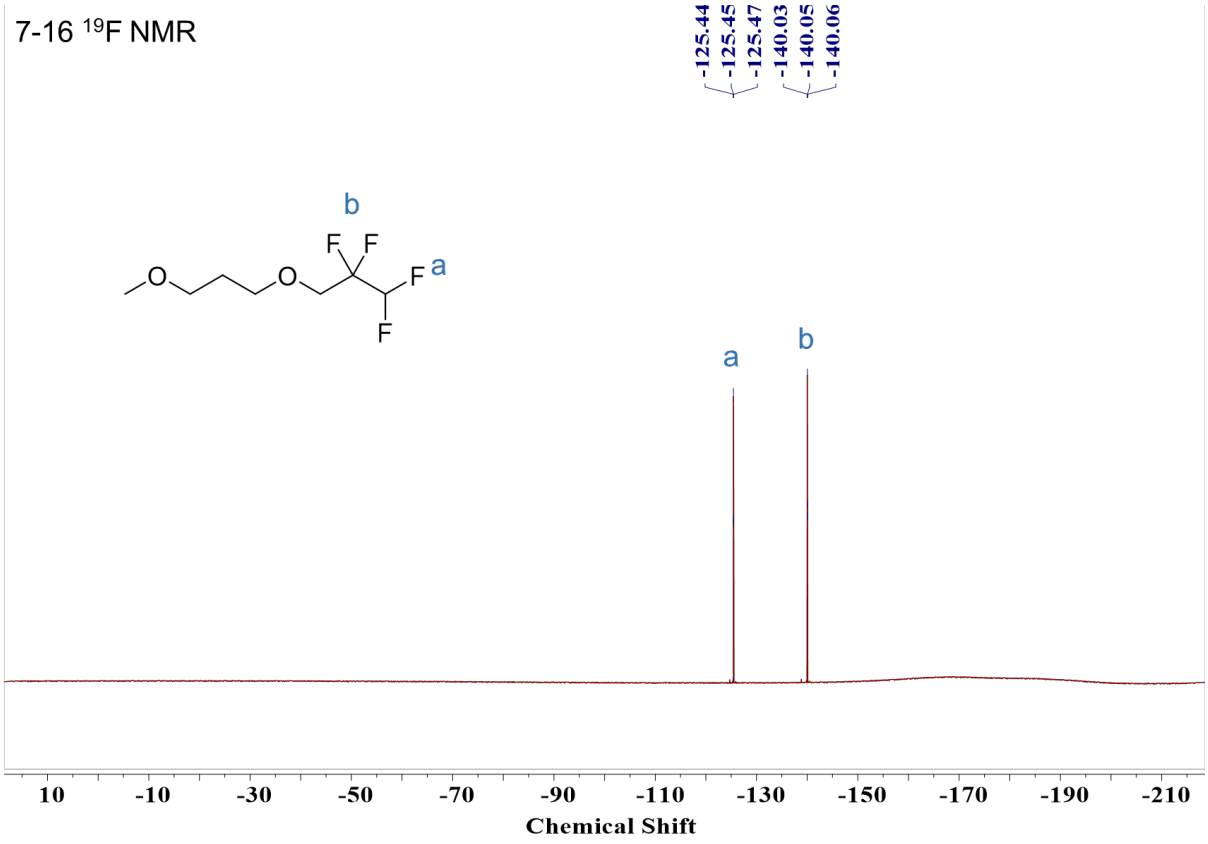


Figure S19. ^{1}H, ^{13}C and ^{19}F NMR spectra of synthesized products.

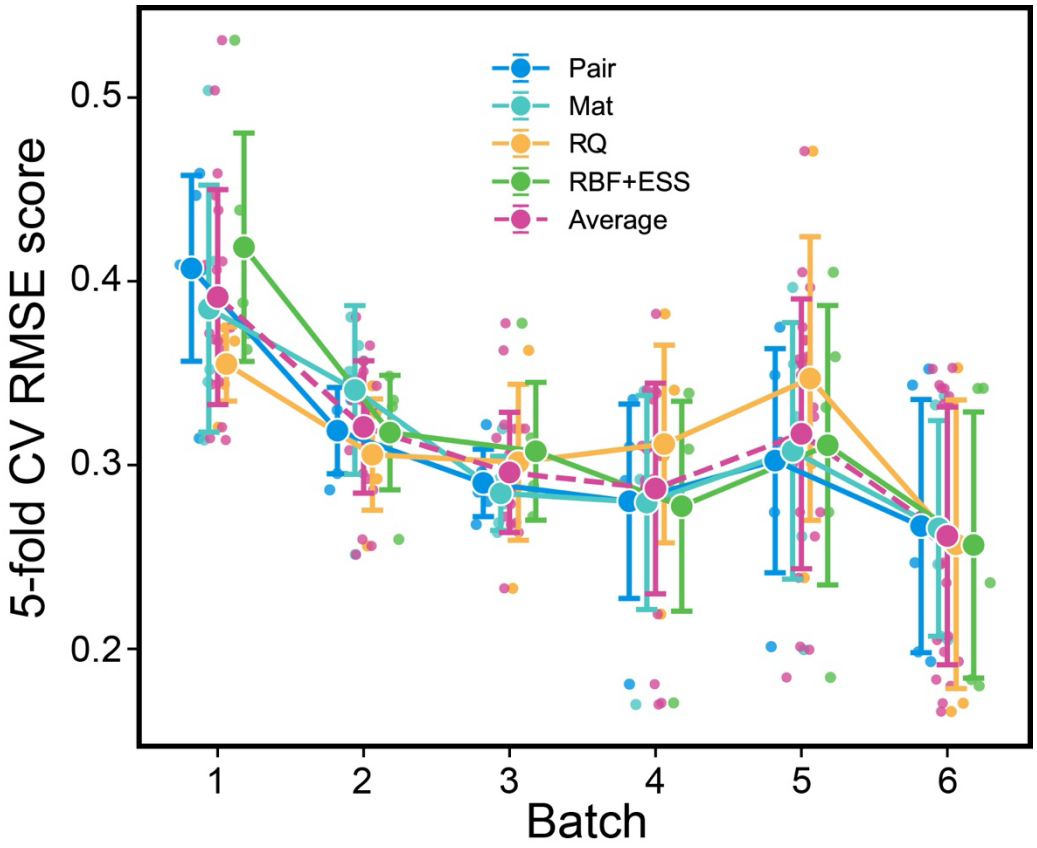


Figure S20. **Surrogate model performance comparison on only LFP data.** 5-fold cross-validation (CV) out-of-batch performance of GP surrogate models as a function of batches for data containing only LFP labels (excluding data with NMC811 positive electrode). The error bars for each kernel (solid lines) denote standard deviation (s.d.) in root mean squared error (RMSE) of model predictions on 5-fold CV splits, and error bars for the average performance (dashed purple line) denote s.d. in RMSE of all 5-fold CV models. Since there are 5 models, one for each of the folds, the performance from each individual fold models are also shown as smaller scatter points for each batch. Pair = pairwise kernel, Mat = matern-3/2 kernel, RQ = rational quadratic kernel, RBF = radial basis function, ESS = exponential-sine squared kernel.

Supplementary references

(1) Thakkar, A.; Chadimová, V.; Bjerrum, E. J.; Engkvist, O.; Reymond, J.-L. Retrosynthetic Accessibility Score (RAscore) – Rapid Machine Learned Synthesizability Classification from AI Driven Retrosynthetic Planning. *Chem. Sci.* **2021**, *12* (9), 3339–3349. https://doi.org/10.1039/D0SC05401A.

(2) 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.

(3) Winter, M.; Barnett, B.; Xu, K. Before Li Ion Batteries. *Chem. Rev.* **2018**, *118* (23), 11433–11456. https://doi.org/10.1021/acs.chemrev.8b00422.

(4) Ein-Eli, Y.; Thomas, S. R.; Chadha, R.; Blakley, T. J.; Koch, V. R. Li-Ion Battery Electrolyte Formulated for Low-Temperature Applications. *J. Electrochem. Soc.* **1997**, *144* (3), 823. https://doi.org/10.1149/1.1837495.

(5) Rustomji, C. S.; Yang, Y.; Kim, T. K.; Mac, J.; Kim, Y. J.; Caldwell, E.; Chung, H.; Meng, Y. S. Liquefied Gas Electrolytes for Electrochemical Energy Storage Devices. *Science (80-.).* **2017**, *356* (6345), eaal4263. https://doi.org/10.1126/science.aal4263.

(6) Zhang, Y.; Lee, A. A. Bayesian Semi-Supervised Learning for Uncertainty-Calibrated Prediction of Molecular Properties and Active Learning. *Chem. Sci.* **2019**, *10* (35), 8154–8163. https://doi.org/10.1039/C9SC00616H.

(7) Pan, S.-Y.; Yang, X.-R.; Zhou, Y.; Lv, C.; Deng, H.-T.; Guo, M.-J.; Chen, S.-X.; Hu, Y.-Y.; Deng, L.; Qiao, Y.; Li, J.-T.; Huang, L.; Yang, Y.; Sun, S.-G. Formulating a New Electrolyte: Synergy between Low-Polar and Non-Polar Solvents in Tailoring the Solid Electrolyte Interface for the Silicon Anode. *ACS Appl. Mater. Interfaces* **2021**, *13* (46), 55700–55711. https://doi.org/10.1021/acsami.1c16827.

(8) Talapatra, A.; Boluki, S.; Duong, T.; Qian, X.; Dougherty, E.; Arróyave, R. Autonomous Efficient Experiment Design for Materials Discovery with Bayesian Model Averaging. *Phys. Rev. Mater.* **2018**, *2* (11), 113803. https://doi.org/10.1103/PhysRevMaterials.2.113803.

(9) Doan, H. A.; Agarwal, G.; Qian, H.; Counihan, M. J.; Rodríguez-López, J.; Moore, J. S.; Assary, R. S. Quantum Chemistry-Informed Active Learning to Accelerate the Design and Discovery of Sustainable Energy Storage Materials. *Chem. Mater.* **2020**, *32* (15), 6338–6346. https://doi.org/10.1021/acs.chemmater.0c00768.

(10) Agarwal, G.; Doan, H. A.; Robertson, L. A.; Zhang, L.; Assary, R. S. Discovery of Energy Storage Molecular Materials Using Quantum Chemistry-Guided Multiobjective Bayesian Optimization. *Chem. Mater.* **2021**, *33* (20), 8133–8144. https://doi.org/10.1021/acs.chemmater.1c02040.

(11) Cao, X.; Jia, H.; Xu, W.; Zhang, J.-G. Review—Localized High-Concentration Electrolytes for Lithium Batteries. *J. Electrochem. Soc.* **2021**, *168* (1), 010522. https://doi.org/10.1149/1945-7111/abd60e.

(12) Qian, J.; Adams, B. D.; Zheng, J.; Xu, W.; Henderson, W. A.; Wang, J.; Bowden, M. E.; Xu, S.; Hu, J.; Zhang, J.-G. Anode-Free Rechargeable Lithium Metal Batteries. *Adv. Funct. Mater.* **2016**, *26* (39), 7094–7102. https://doi.org/https://doi.org/10.1002/adfm.201602353.

(13) Qiu, F.; Li, X.; Deng, H.; Wang, D.; Mu, X.; He, P.; Zhou, H. A Concentrated Ternary-Salts Electrolyte for High Reversible Li Metal Battery with Slight Excess Li. *Adv. Energy Mater.* **2019**, *9* (6), 1803372. https://doi.org/https://doi.org/10.1002/aenm.201803372.

(14) Xu, R.; Ding, J.-F.; Ma, X.-X.; Yan, C.; Yao, Y.-X.; Huang, J.-Q. Designing and Demystifying the Lithium Metal Interface toward Highly Reversible Batteries. *Adv. Mater.* **2021**, *33* (52), 2105962. https://doi.org/https://doi.org/10.1002/adma.202105962.

(15) Suo, L.; Xue, W.; Gobet, M.; Greenbaum, S. G.; Wang, C.; Chen, Y.; Yang, W.; Li, Y.; Li, J. Fluorine-Donating Electrolytes Enable Highly Reversible 5-V-Class Li Metal Batteries. *Proc. Natl. Acad. Sci.* **2018**, *115* (6), 1156–1161. https://doi.org/10.1073/pnas.1712895115.

(16) Hagos, T. T.; Thirumalraj, B.; Huang, C.-J.; Abrha, L. H.; Hagos, T. M.; Berhe, G. B.; Bezabh, H. K.; Cherng, J.; Chiu, S.-F.; Su, W.-N.; Hwang, B.-J. Locally Concentrated LiPF6 in a Carbonate-Based Electrolyte with Fluoroethylene Carbonate as a Diluent for

Anode-Free Lithium Metal Batteries. *ACS Appl. Mater. Interfaces* **2019**, *11* (10), 9955–9963. https://doi.org/10.1021/acsami.8b21052.

(17) Fan, X.; Chen, L.; Borodin, O.; Ji, X.; Chen, J.; Hou, S.; Deng, T.; Zheng, J.; Yang, C.; Liou, S.-C.; Amine, K.; Xu, K.; Wang, C. Non-Flammable Electrolyte Enables Li-Metal Batteries with Aggressive Cathode Chemistries. *Nat. Nanotechnol.* **2018**, *13* (8), 715–722. https://doi.org/10.1038/s41565-018-0183-2.

(18) Ren, X.; Zou, L.; Cao, X.; Engelhard, M. H.; Liu, W.; Burton, S. D.; Lee, H.; Niu, C.; Matthews, B. E.; Zhu, Z.; Wang, C.; Arey, B. W.; Xiao, J.; Liu, J.; Zhang, J. G.; Xu, W. Enabling High-Voltage Lithium-Metal Batteries under Practical Conditions. *Joule* **2019**, *3* (7), 1662–1676. https://doi.org/10.1016/j.joule.2019.05.006.

(19) Weber, R.; Genovese, M.; Louli, A. J.; Hames, S.; Martin, C.; Hill, I. G.; Dahn, J. R. Long Cycle Life and Dendrite-Free Lithium Morphology in Anode-Free Lithium Pouch Cells Enabled by a Dual-Salt Liquid Electrolyte. *Nat. Energy* **2019**, *4* (8), 683–689. https://doi.org/10.1038/s41560-019-0428-9.

(20) Genovese, M.; Louli, A. J.; Weber, R.; Martin, C.; Taskovic, T.; Dahn, J. R. Hot Formation for Improved Low Temperature Cycling of Anode-Free Lithium Metal Batteries. *J. Electrochem. Soc.* **2019**, *166* (14), A3342. https://doi.org/10.1149/2.0661914jes.

(21) Louli, A. J.; Eldesoky, A.; Weber, R.; Genovese, M.; Coon, M.; deGooyer, J.; Deng, Z.; White, R. T.; Lee, J.; Rodgers, T.; Petibon, R.; Hy, S.; Cheng, S. J. H.; Dahn, J. R. Diagnosing and Correcting Anode-Free Cell Failure via Electrolyte and Morphological Analysis. *Nat. Energy* **2020**, *5* (9), 693–702. https://doi.org/10.1038/s41560-020-0668-8.

(22) Xue, W.; Huang, M.; Li, Y.; Zhu, Y. G.; Gao, R.; Xiao, X.; Zhang, W.; Li, S.; Xu, G.; Yu, Y.; Li, P.; Lopez, J.; Yu, D.; Dong, Y.; Fan, W.; Shi, Z.; Xiong, R.; Sun, C. J.; Hwang, I.; Lee, W. K.; Shao-Horn, Y.; Johnson, J. A.; Li, J. Ultra-High-Voltage Ni-Rich Layered Cathodes in Practical Li Metal Batteries Enabled by a Sulfonamide-Based Electrolyte. *Nat. Energy* **2021**, *6* (5), 495–505. https://doi.org/10.1038/s41560-021-00792-y.

(23) Yu, Z.; Wang, H.; Kong, X.; Huang, W.; Tsao, Y.; Mackanic, D. G.; Wang, K.; Wang, X.;

Huang, W.; Choudhury, S.; Zheng, Y.; Amanchukwu, C. V; Hung, S. T.; Ma, Y.; Lomeli, E. G.; Qin, J.; Cui, Y.; Bao, Z. Molecular Design for Electrolyte Solvents Enabling Energy-Dense and Long-Cycling Lithium Metal Batteries. *Nat. Energy* **2020**, *5*, 526–533. https://doi.org/10.1038/s41560-020-0634-5.

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

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

(26) Zhao, Y.; Zhou, T.; Jeurgens, L. P. H.; Kong, X.; Choi, J. W.; Coskun, A. Electrolyte Engineering for Highly Inorganic Solid Electrolyte Interphase in High-Performance Lithium Metal Batteries. *Chem* **2023**. https://doi.org/https://doi.org/10.1016/j.chempr.2022.12.005.

(27) Wu, L.-Q.; Li, Z.; Fan, Z.-Y.; Li, K.; Li, J.; Huang, D.; Li, A.; Yang, Y.; Xie, W.; Zhao, Q. Unveiling the Role of Fluorination in Hexacyclic Coordinated Ether Electrolytes for High-Voltage Lithium Metal Batteries. *J. Am. Chem. Soc.* **2024**, *146* (9), 5964–5976. https://doi.org/10.1021/jacs.3c11798.

(28) Li, A.-M.; Borodin, O.; Pollard, T. P.; Zhang, W.; Zhang, N.; Tan, S.; Chen, F.; Jayawardana, C.; Lucht, B. L.; Hu, E.; Yang, X.-Q.; Wang, C. Methylation Enables the Use of Fluorine-Free Ether Electrolytes in High-Voltage Lithium Metal Batteries. *Nat. Chem.* **2024**, *16*, 922–929. https://doi.org/10.1038/s41557-024-01497-x.

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

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

(31) Ma, P.; Xue, C.; Wang, K.-H.; Mirmira, P.; Vu, M. C.; Rivera, O.; Amanchukwu, C. V. Molecular Structure Optimization of Fluorinated Ether Electrolyte for All Temperature Fast Charging Lithium-Ion Battery. *ACS Energy Lett.* **2024**, 6144–6152. https://doi.org/10.1021/acsenergylett.4c01999.

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

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