Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

# Machine Learning-Driven Mass Discovery and High-Throughput Screening of Fluoroether-Based Electrolytes for High-Stability Lithium Metal Batteries

*Q. Jia, H. Liu, X. Wang, Q. Tao, L. Zheng, J. Li, W. Wang, Z. Liu, X. Gu, T. Shen, S. Hou, Z. Jin*, J. Ma**

**Supporting Information**

# Machine Learning-Driven Mass Discovery and High-Throughput Screening of Fluoroether-Based Electrolytes for High-Stability Lithium Metal Batteries

**Authors:**

Qingqing Jia,[#] Hongguang Liu,[#] Xueping Wang, Qiantu Tao, Lifeng Zheng, Xu Gu, Junjie Li, Wei Wang, Ziteng Liu, Tianyu Shen, Shaoyi Hou, Zhong Jin[*], and Jing Ma[*]

**Affiliations:**

State Key Laboratory of Coordination Chemistry, MOE Key Laboratory of Mesoscopic Chemistry, Engineering Research Center of Photoresist Materials, Ministry of Education, MOE Key Laboratory of High Performance Polymer Materials and Technology, MOE Engineering Research Center of Photoresist Materials, Jiangsu Key Laboratory of Advanced Organic Materials, Tianchang New Materials and Energy Technology Research Center, Institute of Green Chemistry and Engineering, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, Jiangsu 210023, P. R. China.

[#]These authors contributed equally to this work.

E-mail addresses of corresponding authors: majing@nju.edu.cn; zhongjin@nju.edu.cn

**Contents**

## S1. Data Sets

**Table S1**. Summary of LBSMox and DN datasets, which is open for all academic usage at http://106.15.196.160:5667 upon the request for a license.

| LBSMox **Datasets** | | $N_{ele}$ | $N_{sol}$ | $N_a$ | $N_{sol}^{conf}$ | $N_{solLi}^{conf}$ | $N_{sal}$ | **Type of elements** | **Type of $E_{ox}$** |
|---|---|---|---|---|---|---|---|---|---|
| LBSMox-ML | | 289 | 251 | 5 - 125 | 584000 | 372000 | 4 | H, C, N, O, F, S, Cl, Br | Exp.[1] |
| LBSMox-ML subsets | LBSMox-I | 126 | 126 | 5-65 | 250000 | 250000 | 3 | H, C, N, O, F, S, Cl | Exp. (from literatures) |
| | LBSMox-II | 61 | 24 | 5-22 | 122000 | 122000 | 4 | H, C, N, O, F, S | Exp. (in this work) |
| | LBSMox-III | 102 | 102 | 5-125 | 212000 | 0 | 1 | H, C, N, O, F, S, Cl, Br | Exp. (from literatures) |
| LBSMox-Screen | | 5576 | 1394 | 5-65 | 2788000 | 2788000 | 4 | H, C, O, F | Virtual library |
| **DN Datasets** | | **$N_{sol}$** | **$N_a$** | **Type of elements** | | | | | **Type of DN** |
| DN-ML | | 199 | 3-77 | H, C, N, O, F, P, S, Cl, Br | | | | | From http://www.stenutz.eu/chem/solv21.php |
| **GCNN Datasets** | | **$N_{sol}$** | **$N_a$** | **Type of elements** | | | | | **Type of data** |
| DFT-GCNN | | 1622 | 4-125 | H, C, N, O, F | | | | | $\bar{I}_{min}$ and $E_S^L$, $E_S^{L+1}$ by DFT calculations |

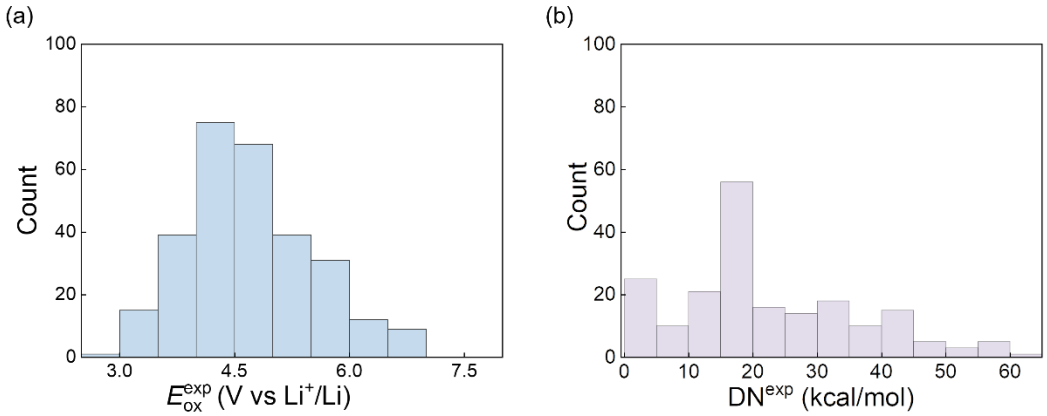



**Figure S1** Distribution of (a) the $E_{ox}^{exp}$ values for LBSMox-ML (b) and the DN$^{exp}$ values for DN-ML.

## S2. The computational details

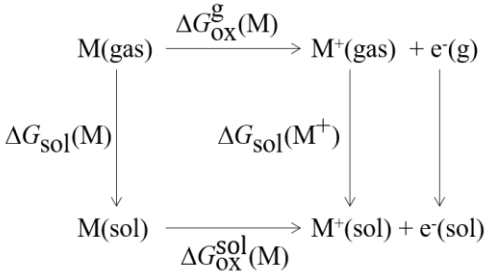
### *S2.1 Conformation sampling and calculation of descriptors*

The conformation sampling for solvent molecules in LBSMox-ML data set was performed with the molclues software. Molecular dynamics (MD) simulations of 100 ps in the canonical ensemble (NVT) were carried out at the theoretical level of GFN0-xTB using xTB software. The time step of MD simulations was fixed in 1fs and the temperature was set to 350 K to accelerate the conformational transition. All 2000 conformers extracted from a MD trajectory per 50 steps were optimized at the level of GFN1-xTB, with the GABS implicit solvent model used to consider the effect of acetone solvent on geometries. The conformational similarity analysis for the optimized conformations was done with two evaluation indexes consisting of energy and root-mean-square deviation (RMSD) between conformations. Specifically, when the differences in energy and RMSD between conformations are within 0.5 kcal/mol and 0.25 Å, respectively, these conformations are considered very similar, so only one is

retained and forms the set of representative conformers of the molecules. Then, the boltzmann distribution of the represent conformers were calculated and only the lowest-energy conformer was further performed geometries optimization and frequency calculations at M06-2X/6-31+G(d, p) level as well as the calculation of descriptors at M06-2X/6-311++G(d, p) level[2]. Solvent effects were accounted by the (Solvation Model based on Density) SMD model and employing the acetone solvent as its dielectric constant is similar to that of a typical mixture of linear and cyclic carbonates currently used in batteries[3]. The DFT-D method was used for dispersion correction[4].

### *S2.2 The theoretical calculation of oxidation potential*

The adiabatic oxidation potential ( $E_{ox}^{DFT}$ ) was calculated based on the thermodynamic cycle[5] (Scheme S1) and was converted to the $Li^+/Li$ potential scale by subtracting 1.4 V (equation 1, 2). The geometries optimization of neutral species M and oxidized species $M^+$ were finished using the Gaussian 16[6] software at the M06-2X/6-31+G(d, p) theoretical level. The frequency calculations were done at the same level with the optimization to confirm the local minimum with no imaginary frequencies. The gas Gibbs free energies of neutral and oxidation species were estimated as a sum of the correction of Gibbs free energies at M06-2X/6-31+G(d, p) level and the electronic energy at M06-2X/6-311++G(d, p) level. The solvation free energy was calculated at the M06-2X/6-31G(d) level. The SMD implicit solvation model[7] and acetone solvent was used to account the solvation effects on theoptimization and the calculation of solvation free energies.




**Scheme S1**. Thermodynamic cycle for the calculation of oxidation potential, where the subscript (gas) and (sol) denotes the gas phase and solution phase, respectively.

$$\Delta G^{sol}_{ox}(M) = \Delta G^{g}_{ox}(M) + \Delta G_{sol}(M^{+}) - \Delta G_{sol}(M) \quad (1)$$

$$E^{DFT}_{ox}(\text{vs Li/Li}^{+}) = \frac{\Delta G^{sol}_{ox}(M)}{F} - 1.4 \quad (2)$$

where F is the Faraday constant, $\Delta G^{g}_{ox}(M)$ is the gas-phase ionization free energy of M, $\Delta G_{sol}(M^{+})$ and $\Delta G_{sol}(M)$ are the solvation free energies of the oxidized species $M^{+}$ and neutral species M in solvent, respectively. All quantities were given at 298 K.

## S3. The development of ML models for the prediction of experimental oxidation potential of electrolytes and donor number of solvents

### *S3.1 Creation of training and testing set*

**Table S2**. Group of LBSMox-ML dataset

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $E_{ox}^{exp}$ | ≤ 3.5 | 3.5-4.0 | 4.0-4.5 | 4.5-5.0 | 5.0-5.5 | 5.5-6.0 | 6.0-6.5 | > 6.5 |
| Counts | 17 | 41 | 80 | 69 | 40 | 22 | 14 | 6 |

**Table S3**. Group of DN-ML dataset

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $DN^{exp}$ | ≤ 10 | 10-20 | 20-30 | 30-40 | 40-50 | >50 |
| Counts | 37 | 79 | 27 | 29 | 21 | 6 |

### *S3.2 Model evaluation*

Three evaluation metrics were applied to assess the performance of ML models for predicting the $E_{ox}^{exp}$ of electrolytes, namely, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Pearson's coefficients (r), as defined in equation S3, S4 and S5, respectively.

$$\mathrm{MAE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left|y_{\mathrm{i}}-y_i^{pred}\right| \tag{S3}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(y_i-y_i^{pred}\right)^2} \tag{S4}$$

$$\mathrm{r}=\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(y_i-y_{mean}\right)\times\left(y_i^{pred}-y_{mean}^{pred}\right)}{\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(y_i-y_{mean}\right)^2}\times\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(y_i^{pred}-y_{mean}^{pred}\right)^2}} \tag{S5}$$

where $y_i$ is the true value and the $y_i^{pred}$ is the predicted value for total $n$ samples. $y_{\mathrm{mean}}$ is the mean values of $y_i$.

### *S3.3 The searched hyper-parameter space*

**Table S4** The hypermeter space explored in the construction of $E_{ox}$ prediction model

Seven algorithms and hypermeters space searched for the $E_{ox}$ prediction

```
grid_param_svr = {
        'C': [int(x) for x in np.arange(2,10,1)],
        'epsilon': [float(x) for x in np.arange(0.1, 0.6, 0.05)],
        'gamma': [float(x) for x in np.arange(1.0, 4, 0.2)]
        }
grid_param_rfr = {
        'n_estimators': [int(x) for x in np.arange(60, 360, 40)],
        'max_depth': [int(x) for x in np.arange(6, 16, 2)],
        'min_samples_split': [int(x) for x in np.arange(2, 6, 1)],
        'min_samples_leaf': [int(x) for x in np.arange(2, 6, 1)]
        }
grid_param_GBRT = {
        'n_estimators': [int(x) for x in np.arange(10, 50, 20)],
        'learning_rate': [float(x) for x in np.arange(0.11, 0.5, 0.02)],
        'max_depth': [int(x) for x in np.arange(1, 4, 1)],
        'min_samples_split': [int(x) for x in np.arange(2, 6, 1)],
        'min_samples_leaf': [int(x) for x in np.arange(1, 6, 1)]
        }
grid_param_ETR = {
        'n_estimators':[int(x) for x in np.arange(5, 50, 10)],
        'max_depth': [int(x) for x in np.arange(6, 16, 2)],
        'min_samples_split': [int(x) for x in np.arange(2, 8, 1)],
        'min_samples_leaf': [int(x) for x in np.arange(1, 4, 1)],
        'max_features': [None,'sqrt', 'log2']
        }

grid_param_DTR = {
        'max_depth': [int(x) for x in np.arange(3, 6, 1)],
        'min_samples_split': [int(x) for x in np.arange(2, 8, 2)],
        'min_samples_leaf': [int(x) for x in np.arange(3, 8, 1)],
        'max_features': [None,'sqrt', 'log2']
        }

grid_param_KNR = {
        'n_neighbors': [int(x) for x in np.arange(2, 8, 2)],
        'weights': ['uniform', 'distance'],
        'p': [int(x) for x in np.arange(1, 4, 1)],
        'algorithm': [None, 'ball_tree', 'kd_tree', 'brute']
        }
```

Voting ={ 'grid_params_weights': [int(x) for x in np.arange(1,10)]

**Table S5** The hypermeter space explored in the construction of ML model for the DN prediction

Seven algorithms and hypermeters space searched for the DN prediction

```
grid_param_svr = {
      'C': [int(x) for x in np.arange(16,40,2)],
      'epsilon': [float(x) for x in np.arange(1.6, 4, 0.2)],
      'gamma': [float(x) for x in np.arange(7.0, 16, 0.5)] }
grid_param_rfr = {
      'n_estimators': [int(x) for x in np.arange(300, 500, 20)],
      'max_depth': [int(x) for x in np.arange(4, 16, 2)],
      'min_samples_split': [int(x) for x in np.arange(2, 5, 1)],
      'min_samples_leaf': [int(x) for x in np.arange(1, 5, 1)]
      }
grid_param_GBRT = {
     'n_estimators': [int(x) for x in np.arange(40, 200, 20)],
     'learning_rate': [float(x) for x in np.arange(0.3, 1.0, 0.1)],
     'max_depth': [int(x) for x in np.arange(8, 20, 2)],
     'min_samples_split': [int(x) for x in np.arange(2, 6, 1)],
     'min_samples_leaf': [int(x) for x in np.arange(1, 6, 1)]
     }
grid_param_ETR = {
      'n_estimators':[int(x) for x in np.arange(100, 150, 10)],
      'max_depth': [int(x) for x in np.arange(8, 16, 2)],
      'min_samples_split': [int(x) for x in np.arange(2, 5, 1)],
      'min_samples_leaf': [int(x) for x in np.arange(1, 5, 1)],
      'max_features': [None, 'sqrt', 'log2']
      }

grid_param_DTR = {
      'max_depth': [int(x) for x in np.arange(4, 10, 1)],
      'min_samples_split': [int(x) for x in np.arange(2, 5, 1)],
      'min_samples_leaf': [int(x) for x in np.arange(1, 4, 1)],
      'max_features': [None, 'sqrt', 'log2']
      }
```

```
grid_param_KNR = {
        'n_neighbors': [int(x) for x in np.arange(2, 5, 1)],
        'weights': ['uniform', 'distance'],
        'p': [int(x) for x in np.arange(1, 4, 1)],
        'algorithm': [None, 'ball_tree', 'kd_tree', 'brute']
        }
voting ={ 'grid_params_weights': [int(x) for x in np.arange(1,10)]
```

**Table S6.** Hyper-parameters of 7 machine learning models for the prediction of oxidation potential and Donor number (in parentheses)

| Models | Parameters |
|---|---|
| SVM | C = 2 (34) |
| | epsilon = 0.3 (1.7) |
| | Gamma = 3.79 (7.0) |
| GB | n_estimators = 30(40) |
| | learning_rate = 0.45(0.3) |
| | max_depth = 2(8) |
| | min_samples_split = (4) |
| | min_samples_leaf = 5(1) |
| ET | n_estimators = 45(140) |
| | max_depth = 12(12) |
| | min_samples_split = 2(4) |
| | min_samples_leaf = 1(1) |
| DT | max_depth = 4(4) |
| | min_samples_split = 2(2) |
| | min_samples_leaf=3(1) |

| | |
|---|---|
| RF | n_estimators =180(480)<br>max_depth = 12(10)<br>min_samples_split = 2(2)<br>min_samples_leaf = 3(1) |
| KNN | n_neighbors = 6(4)<br>p = 1(3) |
| Voting | GB:ET:DT = 6:9:1(SVM-DT-KNN=4:8:3) |

***S3.4 Feature selection***

In order to develop interpretable ML models for the prediction of oxidation potential measured by experiments, a set of descriptors which represent the oxidation stability of components of electrolytes with the solvent environment taken into consideration, were considered by chemical institution. In consideration of the calculation cost, the solvated structures were simplified into solvent $\cdots Li^+ \cdots$anion (SLiA), solvent $\cdots Li^+$ (SLi), and solvent $\cdots$anion (SA) complexes, in conjunction with isolated solvents (S) and isolated anions (A) as well as isolated salts (LiA), generating six kinds of theoretical models to simulate the complex species in the actual electrolyte. Total 66 descriptors, referring to 8 features ($E_L$, $E_{L+1}$, $E_H$, $E_{H-1}$, $\bar{I}_{min}$, $ESP^{min}$, $ESP^{max}$) directly related to the electronic structure and 4 features (*MPI, PSA, NPSA,* $\alpha$) related to the polarity, as well as 3 geometrical parameters (*TSA, V,* $N_F/N_a$) at 5 species (solvent, solvent-$Li^+$, solvent-$Li^+$-anion, anion and salts), were collected and summarized in

**Table S7**. Besides, 208 RDKit[8] descriptors, which has been successfully applied in the prediction of oxidation potential generated by DFT calculation for homobenzylic ether molecules[9], were collected. The mutual information was conducted to capture the non-linear dependencies between features and targets [10].

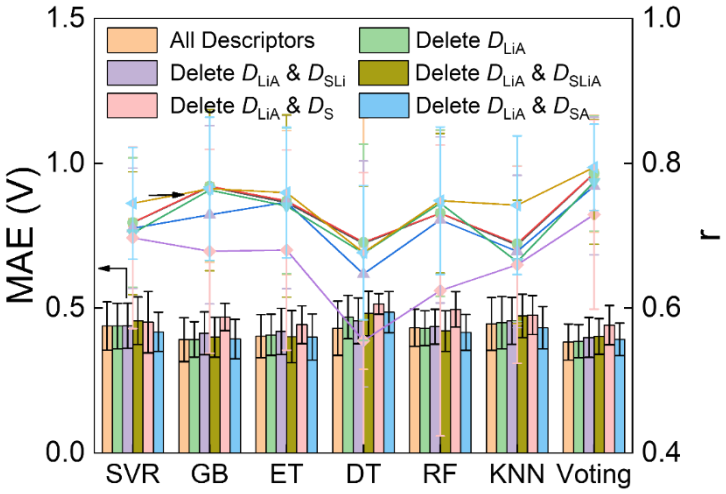
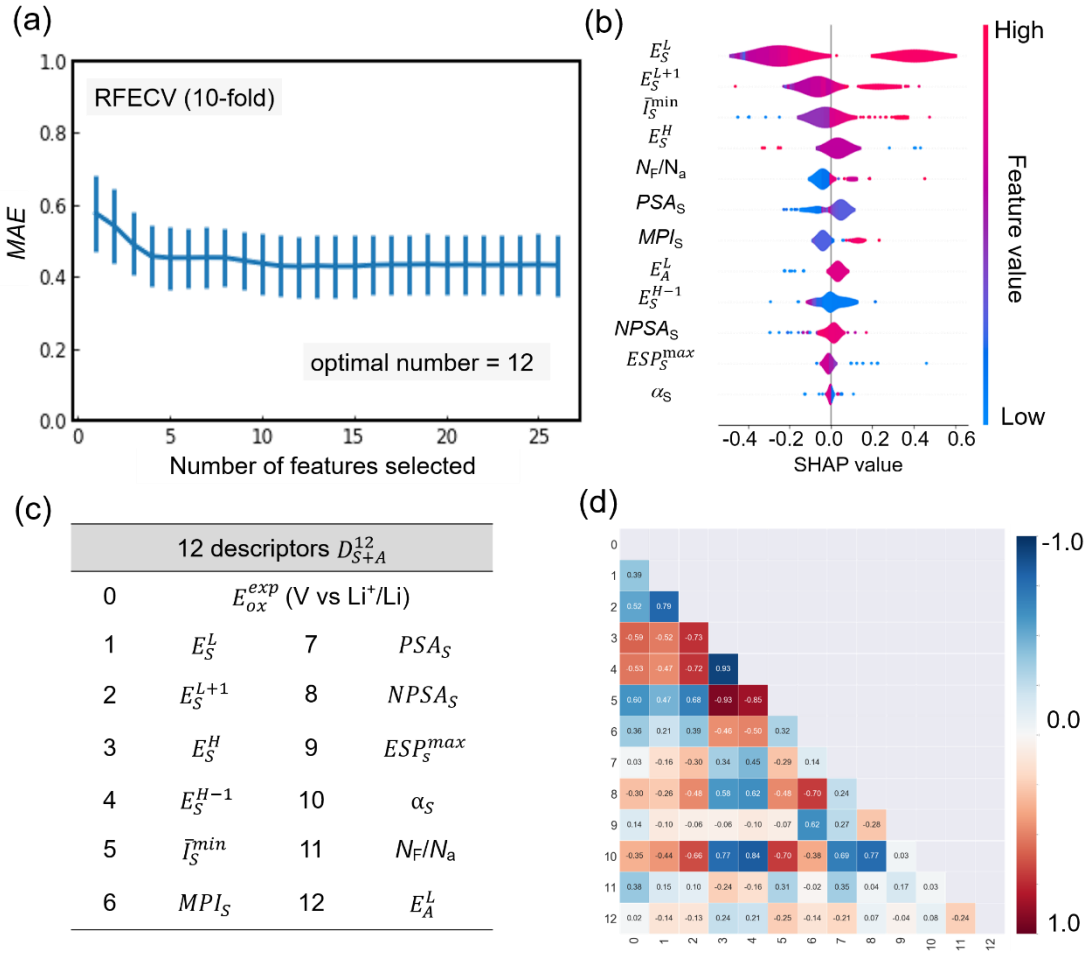



**Figure S2** The performance of 7 ML models developed based on 6 groups descriptor schemes. The error bars represent the standard deviation across the 10 folds.




| 12 descriptors $D_{S+A}^{12}$ | | | |
|---|---|---|---|
| 0 | $E_{ox}^{exp}$ (V vs Li⁺/Li) | | |
| 1 | $E_S^L$ | 7 | $PSA_S$ |
| 2 | $E_S^{L+1}$ | 8 | $NPSA_S$ |
| 3 | $E_S^H$ | 9 | $ESP_S^{max}$ |
| 4 | $E_S^{H-1}$ | 10 | $\alpha_S$ |
| 5 | $\bar{I}_S^{min}$ | 11 | $N_F/N_a$ |
| 6 | $MPI_S$ | 12 | $E_A^L$ |

**Figure S3** (a) The results of RFECV using 10-fold cross-validation (b) The SHAP values of 12 descriptors (c) The 12 descriptors selected by RFECV(10-fold) (d) Correlation between the $E_{ox}^{exp}$ and these 12 descriptors.

**Table S7.** All descriptors collected in this work

| No. | Abbrev. | Description |
|---|---|---|
| 1 | $E_{\mathrm{L}}$ | energy of LUMO |
| 2 | $E_{\mathrm{L+1}}$ | energy of LUMO+1 |
| 3 | $E_{\mathrm{H}}$ | energy of HOMO |
| 4 | $E_{\mathrm{H-1}}$ | energy of HOMO-1 |
| 5 | $\bar{I}_{min}$ | Minimum value of average local ionization energy |
| 6 | *MPI* | Molecular polarity index |
| 7 | *PSA* | Polar surface area |
| 8 | *NPSA* | Nonpolar surface area |
| 9 | *TSA* | Overall surface area |
| 10 | *V* | Molecular Volume |
| 11 | $ESP_{\mathrm{min}}$ | Minimum value of electrostatic potential |
| 12 | $ESP_{\mathrm{max}}$ | Maximum value of electrostatic potential |
| 13 | $\alpha$ | Polarizability of solvent |
| 14 | $\Delta E_{SSalt}^{int}$ | The interaction energy between solvent and salts |
| 15 | $\Delta E_{SLi}^{int}$ | The interaction energy between solvent and $Li^{+}$ |
| 16 | $N_{\mathrm{F}}/N_{\mathrm{a}}$ | The ration of the number of F atoms to total atoms |

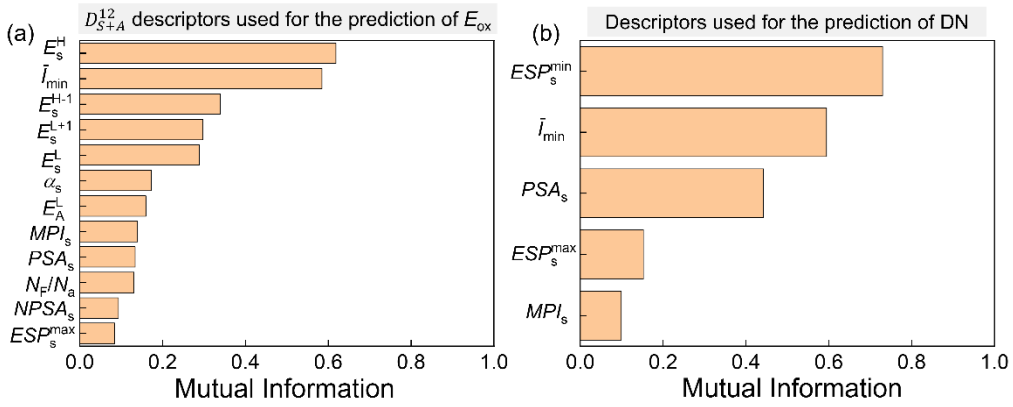



**Figure S4** The mutual information of descriptors used in the prediction of the $E_{ox}$ and the donor number, respectively.

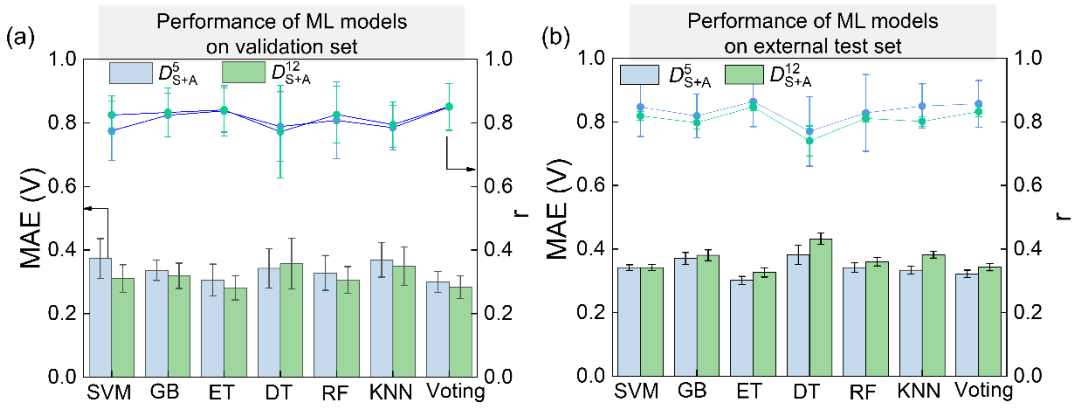



**Figure S5** The performance of 7 models developed based on the 5 descriptors scheme and the 26-descriptors scheme on the (a) validation set (n = 228 samples) and (b) external test sets (n = 61 samples) of LBSMox-ML dataset. The error bars represent the standard deviation across the 10 folds.

### *S3.5 Hyper-parameters of graph convolution neural network model*

**Table S8**. Hyper-parameters of graph convolution neural network model (GCNN)

| **Category** | **Hyper-parameter** | **Value** |
|---|---|---|
| Message Passing Function | Number of message passing steps | 6 |
| | Edge network layers | 5 |
| | Edge network hidden dim | 128 |
| Update Function | Node hidden units dim | 64 |
| | Attention layer | 3 |
| Readout Function | set2set | 6 |
| | Output NN hidden units | 64 |
| Auxiliary Target/ACSFs | Radial functions | [1, 1], [1, 2], [1, 3] |
| | Angular functions | [1, 1, 1], [1, 2, 1], [1, 1, -1], [1, 2, -1] |
| Training | Initial learning rate | 1.5e-4 |
| | Scheduler | Cosine Annealing Lr |
| | Optimizer | Adam |
| | Batch size | 8 |
| | Training epochs | 1500 |

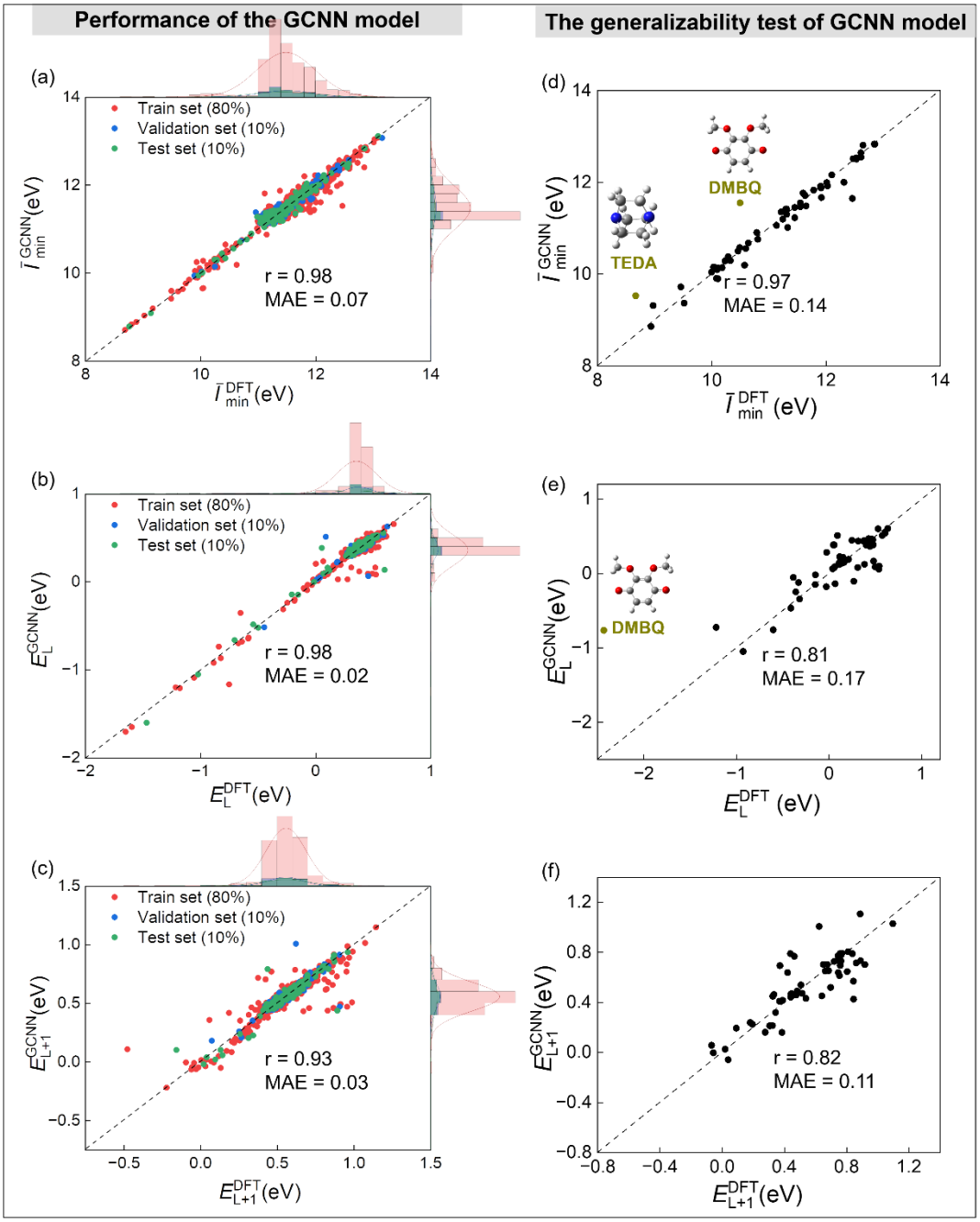



**Figure S6** Correlation of (a) the minimum value of average local ionization energy ($\bar{I}_{min}$) (b) the LUMO energy level ($E_L$) (c) the LUMO+1 energy levels of the solvent ($E_{L+1}$) calculated by DFT and predicted by GCNN on DFT-GCNN dataset (1622 samples) and generalizability of trained GCNN model on the external test set (61 samples) of LBSMox-ML data set for the prediction of (d) $\bar{I}_{min}$, (e) $E_L$, (f) $E_{L+1}$ with the outliers shown in figures, respectively.

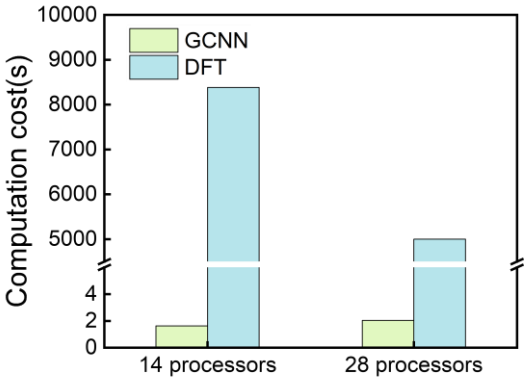



**Figure S7** The comparison of computational cost for the generation of three descriptors ($\bar{I}_{min}$ and $E_S^L$, $E_S^{L+1}$) on 61 molecules based on the GCNN and DFT, respectively.

**Table S9** The performance of ML models over the external test set (61 samples) of LBSMox-ML dataset for the prediction of $E_{ox}^{exp}$, using $\bar{I}_{min}$, $E_S^L$, $E_S^{L+1}$, $N_F/N_a$ and $E_A^L$ descriptors calculated by DFT ($\bar{I}_{min}^{DFT}$, $E_L^{DFT}$, $E_{L+1}^{DFT}$) or GCNN (($\bar{I}_{min}^{GCNN}$, $E_L^{GCNN}$, $E_{L+1}^{GCNN}$, results shown in parentheses). The error bars represent the standard deviation across the 10-folds.

| | SVM | GB | ET | DT | RF | KNN | GB-ET-DT(6:9:1) |
|---|---|---|---|---|---|---|---|
| MAE | 0.34 ± 0.01 | 0.37 ± 0.02 | 0.30± 0.01 | 0.38 ± 0.01 | 0.34 ± 0.01 | 0.33 ± 0.01 | 0.32 ± 0.01 (0.33 ± 0.01 ) |
| r | 0.85 ± 0.01 | 0.82 ± 0.02 | 0.86 ± 0.01 | 0.77 ± 0.03 | 0.83 ± 0.02 | 0.85 ± 0.01 | 0.86±0.01(0.84±0.01) |

**Table S10** The statistical analysis of the ML model performance on external test for the prediction of $E_{ox}$ (n = 61 samples) and DN (n = 43 samples) evaluated by four statistical methods

| Statistical analysis of ML models | | | | | | | |
|---|---|---|---|---|---|---|---|
| Eox(GB-ET-DT) | | | | | | | |
| KS value | 0.1148 | U statistic | 1917 | t statistic | 0.04225 | r | 0.86 |
| p-value | 0.8211 | p-value | 0.7743 | p-value | 0.9664 | p-value | 4.07E-19 |
| DN (SVM-DT-KNN) | | | | | | | |
| KS value | 0.1860 | U statistic | 882 | t statistic | 0.01725 | r | 0.90 |
| p-value | 0.4506 | p-value | 0.7168 | p-value | 0.9863 | p-value | 7.68E-17 |

**Table S11** The performance of ML models over the external test set (43 samples) of DN-ML dataset for the prediction of donor number, using descriptors of $ESP_{min}$, $ESP_{max}$, $PSA$, $\bar{I}_{min}$ and MPI descriptors calculated by DFT. The error bars represent the standard deviation across the 10-folds.

| | SVM | GB | ET | DT | RF | KNN | SVM-DT-KNN(4:8:3) |
|---|---|---|---|---|---|---|---|
| MAE | 4.01 ± 0.19 | 5.88 ± 0.31 | 4.15 ± 0.07 | 6.17 ± 0.19 | 4.88 ± 0.20 | 4.71 ± 0.15 | 4.85 ± 0.12 |
| r | 0.92 ± 0.01 | 0.83 ± 0.02 | 0.92± 0.01 | 0.83 ± 0.02 | 0.90 ± 0.01 | 0.90 ± 0.01 | 0.90 ± 0.004 |

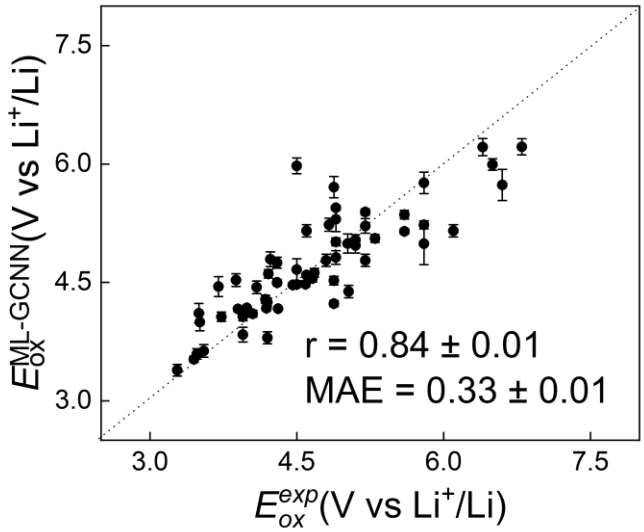



**Figure S8** The predicted oxidation potential of ML model trained by GB-ET-KNN learners versus experimental oxidation potential on external test sets (n = 61 samples), using $D_{S+A}^{5}$ calculated by GCNN ($\bar{I}_{min}^{GCNN}$, $E_{L}^{GCNN}$, $E_{L+1}^{GCNN}$). The error bars represent the standard deviation across the 10-folds.

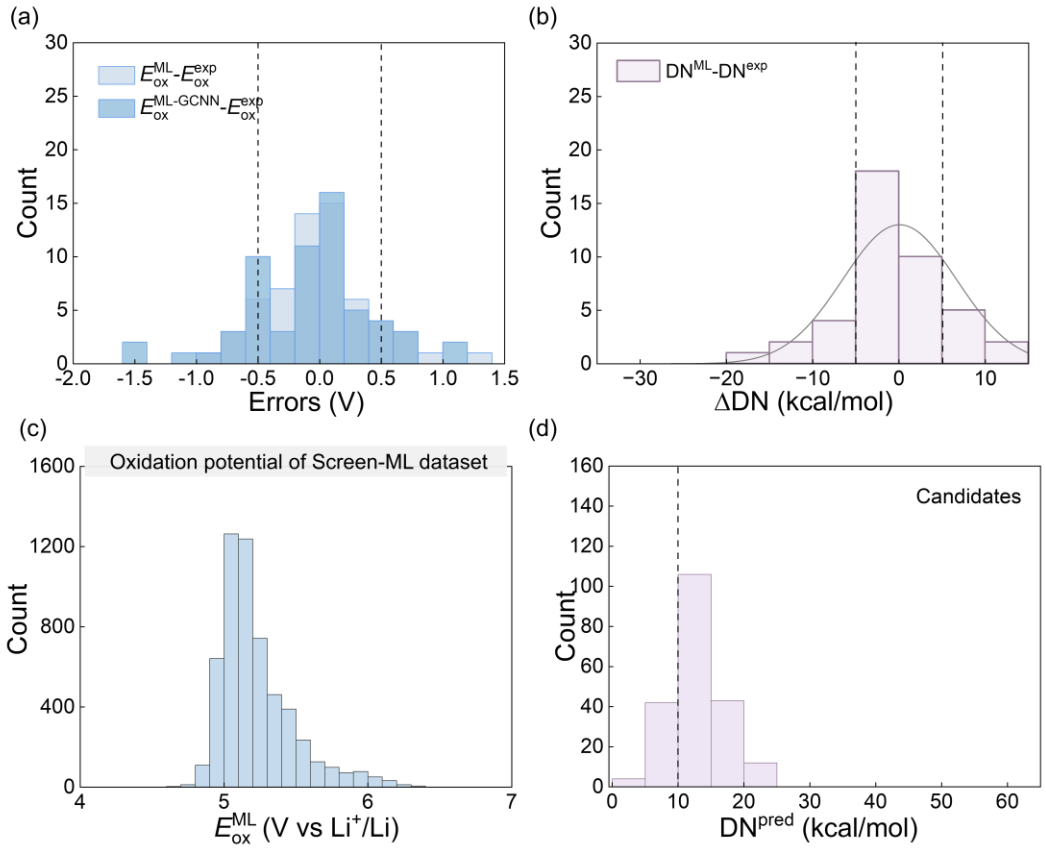



**Figure S9** (a) The prediction errors of oxidation potential at the LBSEMox-ML dataset (289 samples) using the GB-ET-DT ML model (b) the errors of DN values between the prediction values using the SVM-DT-KNN ML model and the experiment values (c) the distribution of oxidation potential predicted by GB-ET-DT ML model for the Screen-ML dataset (d) the distribution of donor number predicted by SVM-DT-KNN learner.

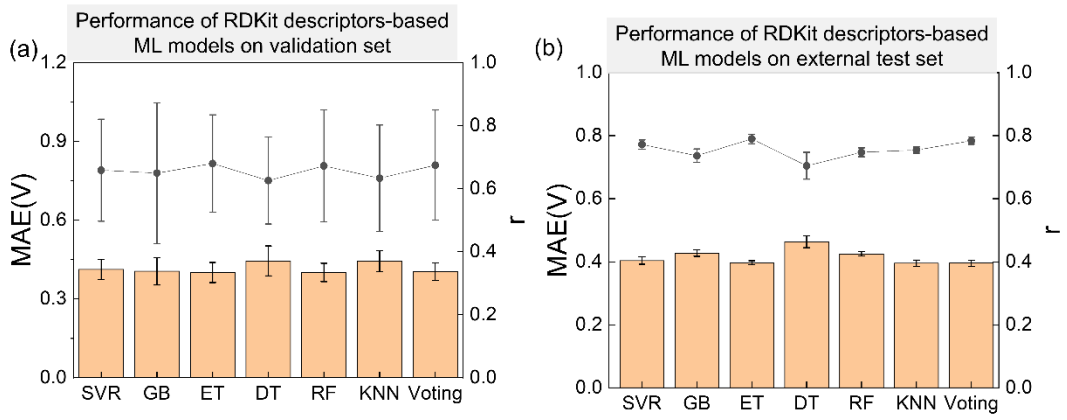



**Figure S10** The performance of 7 models developed based on the RDKit descriptors and on the (a) validation set (n = 228 samples) and (b) external test sets (n = 61 samples) of LBSMox-ML dataset. The error bars represent the standard deviation across the 10 folds.

### *S3.6 Calculation of the SAscore*

The computation of the SAscore integrates two primary components: fragment contributions and molecular complexity penalties. The process can be delineated as follows:

**Fragment Analysis**:

The molecule is decomposed into its constituent fragments using algorithms such as the Extended Connectivity Fingerprints (ECFP). Each fragment is assigned a score based on its occurrence frequency in a reference database of known compounds. Common fragments, which are presumed easier to synthesize, receive lower scores, while rare fragments are assigned higher scores.

**Fragment Score Calculation**:

The individual fragment scores are summed and normalized by the total number of fragments to yield an average fragment score for the molecule.

**Complexity Penalties**: Additional penalties are incorporated to account for structural features that complicate synthesis, including the molecular Size, the Chirality, Ring Systems and Macrocycles.

**Score Aggregation**: The normalized fragment score and the complexity penalties are combined to produce a raw SA score.

**Score Transformation**: The raw SA score is transformed onto a scale of 1 to 10, ensuring that the scores are interpretable and comparable across different molecules.

**Table S12** The properties of 162 promising solvent candidates screened candidates

| canonsmiles | SAscore | Eox | DN | logP | |
|---|---|---|---|---|---|
| COCCOC(C(F)(F)F)C(F)(F)F | 2.80 | 5.51 | 14.88 | 2.26 | Synthesized in this work |
| COC(COCC(C(COCC(OC)OC)(F)F)(F)F)OC | 3.24 | 5.62 | 21.58 | 2.96 | Synthesized in this work |
| COC[C@H](C(F)(F)F)O[C@H](C(F)(F)F)COC | 3.90 | 6.10 | 14.27 | 2.77 | |
| COCC(C(C(F)F)(F)F)(C(C(F)F)(F)F)COC | 3.56 | 5.47 | 14.77 | 3.06 | |
| COCOCOC(C(OC)(F)F)C(OC)(F)F | 3.56 | 5.52 | 15.85 | 2.62 | |
| FC(COCCCOCC(F)F)F | 2.87 | 5.73 | 13.76 | 1.86 | |
| FC(C(F)(F)F)(COCCCCOCC(C(F)(F)F)(F)F)F | 2.71 | 5.75 | 10.15 | 3.17 | |
| … | … | … | … | … | … |
| CO[C@@H](C(C(F)F)(F)F)COCC(C(F)F)(F)F | 3.94 | 6.21 | 12.95 | 2.79 | |
| COCOC(COC(F)F)COC(F)F | 3.27 | 5.44 | 10.48 | 2.34 | |
| COCCOCC(CF)(CF)OC | 3.38 | 5.48 | 15.63 | 2.09 | |

Note: To save the space, we only list part of data. If one needs the complete set of data, please write to majing@nju.edu.cn.

## S4. Electrochemical measurements

### *S4.1 Electrolyte and electrode preparation*

All the materials and solvents were purchased from commercial sources and used as received without further purification. The electrolytes were prepared by dissolving predetermined amounts of salt (LiFSI, $LiPF_6$, $LiBF_4$, and LiTFSI) into the solvents of interest and stirred until the clear solution is obtained. To prepare the NMC811 cathode, NMC811 powder, polyvinylidene fluoride (PVDF) binder, and carbon black (Super P) were mixed in N-methyl-2-pyrrolidone (NMP) with a weight ratio of 8:1:1. After magnetic stirring for 12 h, the slurry was spread onto an Al foil by doctor blading and dried under vacuum at 60 ℃ overnight.

### *S4.2 Electrochemical measurements*

The Linear Sweep Voltammetry (LSV) was conducted to measure the oxidation potential of electrolytes, using the configuration of Li||stainless steel cells with the cell voltage scanned from 2 V to 7 V at a rate of 0.5 mV/s. The oxidation potentials were obtained from the LSV curve by measuring the intersection between two extrapolated lines of the residual current line and the increasing oxidation current line.

The configuration of Li||stainless steel cells was Li||15 μL of electrolyte||Celgard 2500 separator||15 μL of electrolyte||stainless steel. The lithium electrode has a thickness of 450 μm and a diameter of 15.6 mm.

The ionic conductivities of the electrolytes were measured using electrochemical impedance spectroscopy (EIS) with a 5 mV amplitude across a frequency range of 100

kHz to 10 MHz at 25 ℃, employing a stainless steel blocking cell and Celgard 2500 separator. The $Li^+$ transference number was determined using a Li||Li symmetric cell under a polarization potential of 10 mV and calculated based on the equation:

$$t_{Li^+} = \frac{I_{ss}(\Delta V - I_0 R_0)}{I_0(\Delta V - I_{ss} R_{ss})}$$

The initial current $I_0$, steady current $I_{ss}$, initial interfacial resistance $R_0$, and steady interfacial resistance $R_{ss}$ were measured by DC polarization and AC impedance; the $\Delta V$ is the polarization voltage.

The above all electrochemical measurements were conducted by the Chenhua CHI-760e workstation at 25 ℃ and the CR-2032-type coin cells were assembled in an Ar-filled glovebox ($O_2$ < 0.5 ppm and $H_2O$ < 0.1 ppm) and rested for 8 h before measurements.

The full cell Li||NMC811 cells, consisting of Li foil and NMC811 laminate as well as electrolyte formed by 1.5 M LiTFSI in HFMEP, were cycled at various specific currents within a voltage range of 2.8–4.4 V *vs* $Li/Li^+$ after formation cycles. The full cell was galvanostatically cycled using a LAND-CT2001 8-channel battery testing system.

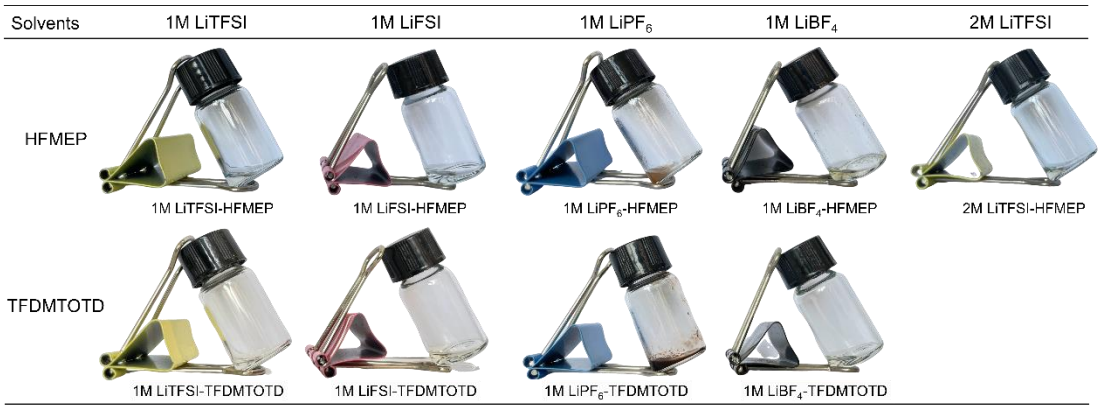


**Figure S11** The electrolytes formed by the 2 synthesized molecules paired with four lithium salts.

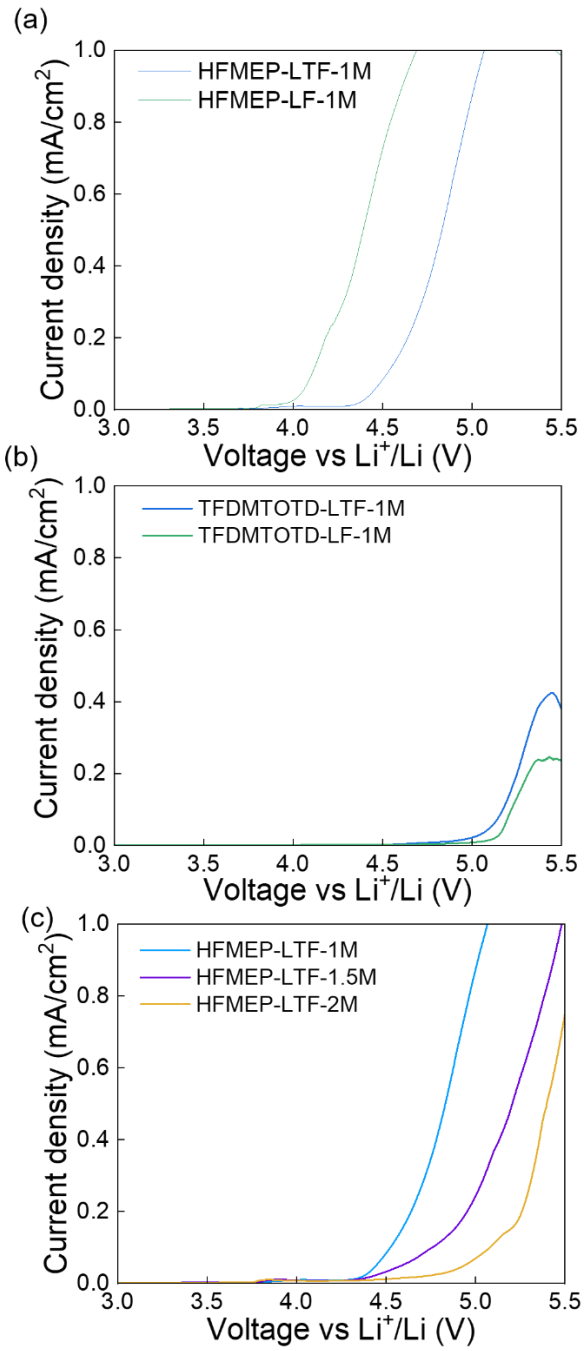



**Figure S12** Linear Sweep Voltammetry analysis for the electrolytes formed by the (a) HFMEP with LTF and LF at 1M concentration (b) TFDMTOTD with LTF and LF at 1M concentration (c) HFMEP with the LTF at 1M, 1.5M and 2M.

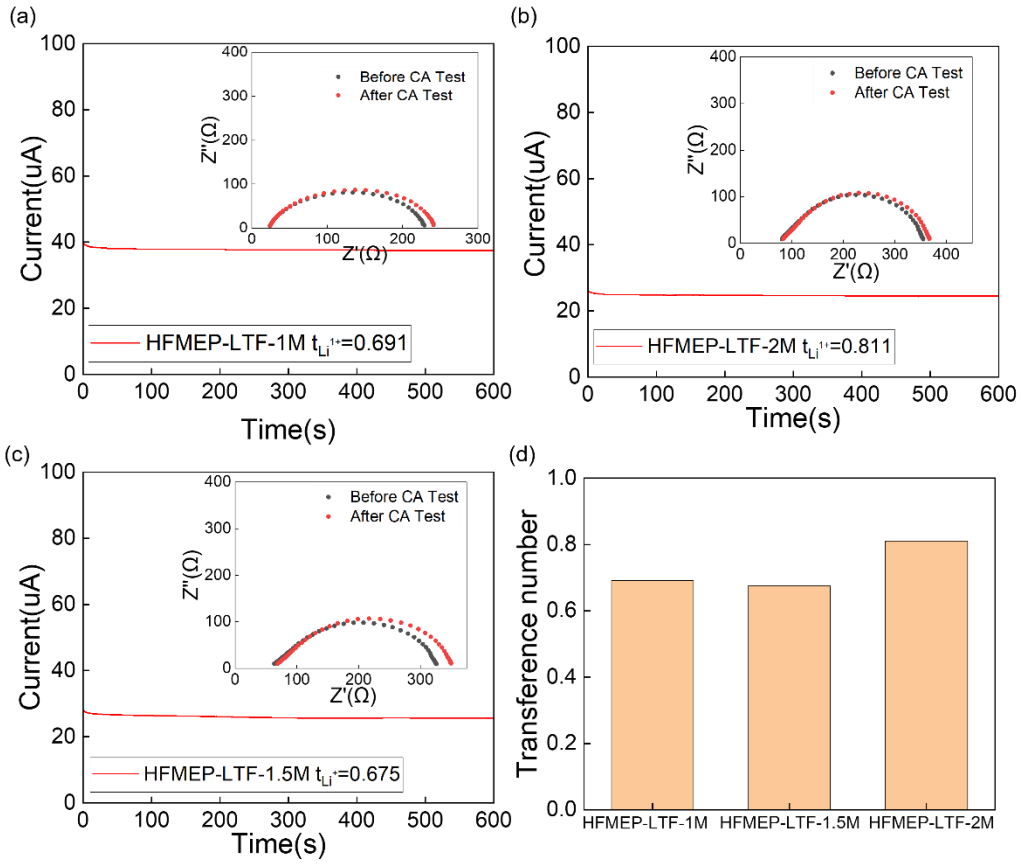



**Figure S13** The Chronoamperometry and impedance measurement for the electrolytes formed by the HFMEP and LiTFSI salts at (a) 1M, (b) 2 M (c) 1.5 M concentration, respectively. (d) lithium ion transference numbers ($t_{Li+}$) of electrolytes formed by the HFMEP with LiTFSI at three concentration of salts.

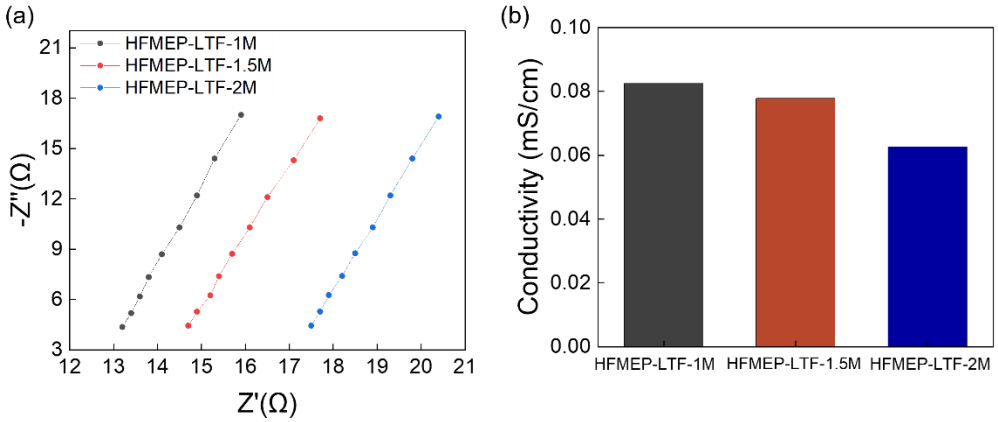



**Figure S14** (a) The electrochemical impedance spectroscopy for the electrolytes formed by the HFMEP with LiTFSI salts at 1M, 1.5 M and 2M concentration (b) The bulk conductivity with swelled Celgard separator of electrolytes formed by the HFMEP with LiTFSI at three concentrations of salts.

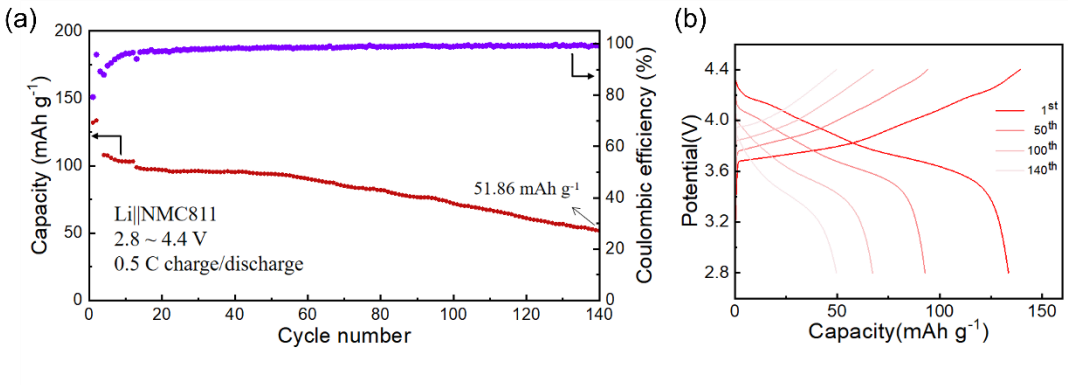



**Figure S15** (a) Discharge capacity and Coulombic efficiency as a function of cycle number for the 1.5 M LiTFSI in HFMEP electrolyte (b) Galvanostatic cycling of 1.5 M LiTFSI in HFMEP.

**Table S13** The cost of some starting materials

| Starting materials | Price |
|---|---|
| 1,1,1,3,3,3-hexafluoropropan-2-ol | 5.38 $/100 g |
| 1-bromo-2-methoxyethane | 12.15 $/100 g |
| 2,2,3,3-tetrafluorobutane-1,4-diol | 780 $/100 g |
| 2-bromo-1,1-dimethoxyethane | 3.17 $/100 g |
| 2,2,3,3-tetrafluoropropan-1-ol | 4.55 $/100 g |
| 1,4-dibromobutane | 3.31 $/100 g |
| 2,2,3,3-tetrafluoropropan-1-ol | 4.55 $/100 g |
| 1-bromo-2-(2-bromoethoxy)ethane | 16.40 $/100 g |

**Table S14** Environment toxicity analysis of the two novel fluoroethers solvents

| Name | SMILES | Organism | Duration | Toxicity indicators | Concentration (mg/L) | Max Log Kow | Log Kow |
|---|---|---|---|---|---|---|---|
| HFMEP | COCCOC(C(F)(F)F)C(F)(F)F | Fish | 96h | LC50 | 488.23 | 5 | 1.53 |
| | | Daphnid | 48h | LC50 | 265.51 | 5 | 1.53 |
| | | Green Algae | 96h | EC50 | 165.46 | 6.4 | 1.53 |
| | | Fish | | ChV | 45.35 | 8 | 1.53 |
| | | Daphnid | | ChV | 22.96 | 8 | 1.53 |
| | | Green Algae | | ChV | 39.36 | 8 | 1.53 |
| | | Fish (SW) | 96h | LC50 | 612.92 | 5 | 1.53 |
| | | Mysid | 96h | LC50 | 625.15 | 5 | 1.53 |
| | | Fish (SW) | | ChV | 50.01 | 8 | 1.53 |
| | | Mysid (SW) | | ChV | 62.06 | 8 | 1.53 |
| TFDMTOTD | COC(COCC(C(COCC(OC)OC)(F)F)(F)F)OC | Fish | 96h | LC50 | 6114.78 | 5 | 0.51 |
| | | Daphnid | 48h | LC50 | 3024.40 | 5 | 0.51 |
| | | Green Algae | 96h | EC50 | 1273.21 | 6.4 | 0.51 |
| | | Fish | | ChV | 507.90 | 8 | 0.51 |
| | | Daphnid | | ChV | 200.87 | 8 | 0.51 |
| | | Green Algae | | ChV | 245.20 | 8 | 0.51 |
| | | Fish (SW) | 96h | LC50 | 7629.47 | 5 | 0.51 |
| | | Mysid | 96h | LC50 | 15608.63 | 5 | 0.51 |
| | | Fish (SW) | | ChV | 332.41 | 8 | 0.51 |
| | | Mysid (SW) | | ChV | 2101.09 | 8 | 0.51 |

## References


[S1] a) A. Abouimrane, I. Belharouak, K. Amine, *Electrochem. Commun.* **2009**, *11*, 1073-1076; b) P. Y. Ma, P. Mirmira, C. V. Amanchukwu, *ACS Cent. Sci.* **2021**, *7*, 1232-1244; c) C. V. Amanchukwu, Z. Yu, X. Kong, J. Qin, Y. Cui, Z. Bao, *J. Am. Chem. Soc.* **2020**, *142*, 7393-7403; d) B. Flamme, M. Haddad, P. Phansavath, V. Ratovelomanana-Vidal, A. Chagnes, *Chemelectrochem* **2018**, *5*, 2279-2287; e) C. C. Su, M. He, P. C. Redfern, L. A. Curtiss, I. A. Shkrob, Z. C. Zhang, *Energy Environ. Sci.* **2017**, *10*, 900-904; f) B. Flamme, G. R. Garcia, M. Weil, M. Haddad, P. Phansavath, V. Ratovelomanana-Vidal, A. Chagnes, *Green Chem.* **2017**, *19*, 1828-1849; g) W. Xu, X. L. Chen, F. Ding, J. Xiao, D. Y. Wang, A. Q. Pan, J. M. Zheng, X. H. S. Li, A. B. Padmaperuma, J. G. Zhang, *J. Power Sources* **2012**, *213*, 304-316; h) N. Shao, X. G. Sun, S. Dai, D. E. Jiang, *J. Phys. Chem. B* **2011**, *115*, 12120-12125.

[S2] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 291-309.

[S3] L. Xing, O. Borodin, G. D. Smith, W. Li, *J. Phys. Chem. A* **2011**, *115*, 13896-13905.

[S4] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*.

[S5] P. Winget, C. J. Cramer, D. G. Truhlar, *Theor. Chem. Acc.* **2004**, *112*, 217–227.

[S6] M. J. T. Frisch, G. W. Schlegel, H. B. Scuseria, G. E. Robb, M. A. Cheeseman, J. R. Scalmani, G. Barone, V. Petersson, G. A. Nakatsuji, H. Li, X. Caricato, M. Marenich, A. V. Bloino, J. Janesko, B. G. Gomperts, R. Mennucci, B. Hratchian, H. P. Ortiz, J. V. Izmaylov, A. F. Sonnenberg, J. L. Williams Ding, F. Lipparini, F. Egidi, F. Goings, J. Peng, B. Petrone, A. Henderson, T. Ranasinghe, D. Zakrzewski, V. G. Gao, J. Rega, N. Zheng, G. Liang, W. Hada, M. Ehara, M. Toyota, K. Fukuda, R. Hasegawa, J. Ishida, M. Nakajima, T. Honda, Y. Kitao, O. Nakai, H. Vreven, T. Throssell, K. Montgomery Jr., J. A. Peralta, J. E. Ogliaro, F. Bearpark, M. J. Heyd, J. J. Brothers, E. N. Kudin, K. N. Staroverov, V. N. Keith, T. A. Kobayashi, R. Normand, J. Raghavachari, K. Rendell, A. P. Burant, J. C. Iyengar, S. S. Tomasi, J. Cossi, M. Millam, J. M. Klene, M. Adamo,

C. Cammi, R. Ochterski, J. W. Martin, R. L. Morokuma, K. Farkas, O. Foresman, J. B. Fox, D. J. Gaussian 16 Rev. A.03, Wallingford, CT, 2016.

[S7] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

[S8] G. Landrum, RDKit: Open-Source Cheminformatics Software, https://www.rdkit.org (accessed 1 August 2023).

[S9] H. A. Doan, G. Agarwal, H. Qian, M. J. Counihan, J. Rodriguez-Lopez, J. S. Moore, R. S. Assary, *Chem. Mater.* **2020**, *32*, 6338-6346.

[S10] J. Chen, Y. Gu, Q. Zhu, Y. Gu, X. Liang, J. Ma, Langmuir. 10.1021/acs.angmuir.4c04638.