# **Supplementary Information**

for

# Machine learning-guided discovery of ionic polymer electrolytes for lithium metal batteries

Kai Li<sup>1</sup>, Jifeng Wang<sup>1</sup>, Yuanyuan Song<sup>1</sup>, Ying Wang<sup>1\*</sup>



<span id="page-1-0"></span>To investigate the solubility of Li salts, here we introduce the concept of binding energy  $(E<sub>binding</sub>)$ , which is the lowest energy required to combine the cation( $Li<sup>+</sup>$ ) and the anions to form the Li-anion pairs. The lower of the E<sub>binding</sub> means easier for the Li<sup>+</sup> and anions to stay in tightly associated pairs and pack into solid crystals. We observe that the  $E_{binding}$  of LiFSI (-512 kJ mol<sup>-1</sup>) is significantly higher than that of LiTfO  $(-603 \text{ kJ mol}^{-1})$  and LiBF<sub>4</sub>  $(-602.39 \text{ kJ mol}^{-1})$ , this explains the high solubility of LiFSI in ionic liquids compared to the other salts. The Ebinding is calculated based on Equation 1 in the manuscript. The used theory and basis set is M062X/6- 311(+)G(2d,p) with dispersion correction of DFT-D3.

#### <span id="page-1-1"></span>**Supplementary Note 2**

The Nernst-Einstein equation correlates the conductivity (*σ*) to the diffusion coeffficents of cation  $(D^+)$  and anion  $(D^-)$  in the system. The Stokes-Einstein equation displays the relationship among the viscosity (*η*) ,diffusion coefficients (*D*) and radii of gyration (*r*) of the ions or molecules. Thus, we can derive the relationship between the ionic conductivity and the viscosity of the system.

Nernst-Einstein Equation 
$$
\sigma = \frac{z_i^2 F^2}{RT} (D^+ + D^-)
$$
 Equation (1)  
\nStoke-Einstein Equation 
$$
D = \frac{k_B T}{6\pi \eta r}
$$
 Equation (2)

<span id="page-2-0"></span>As indicated in previous literature,<sup>1,2</sup> the uncertainty of the comparison for different cation and anion types is highly related to the limitation of the HOMO/LUMO theory, for example, the  $C_2$ mimBF<sub>4</sub> is usually overestimated with a "weak" cation paired with a strong anion. For the cations, the imidazolium type is also not accurate, because the description of the top of the valence band for some of the imidazolium based ILs is not very accurate using the DFT and related approximations, especially for the imidazolium ones with BF<sup>4</sup> anions. However, the overall trend of the ECW values is reliable and showed enough accuracy for screening of potential ILs in this application.

### <span id="page-2-1"></span>**Supplementary Note 4**

Here is the list of the 16 factors used for the hierachical clustering of the unsupervised learning. The top 15 factors are obtained from the XGBoosting model for the conductivity classification task. The last factor is ECW\_computed calculated based on the HOMO/LUMO theory through Psi4 as described in the manuscript.

- 1. "MaxPartialCharge\_anion",
- 2. "PMI2\_cation",
- 3. "energy\_anion",
- 4. "RadiusOfGyration\_pair",
- 5. "FpDensityMorgan2\_anion",
- 6. "Eccentricity\_pair",
- 7. "PMI2\_pair",
- 8. "Asphericity\_anion",
- 9. "volume\_pair",
- 10. "Asphericity\_pair",
- 11. "HOMO\_cation",
- 12. "volume\_cation",
- 13. "PMI3\_pair",
- 14. "FpDensityMorgan1\_pair",
- 15. "RadiusOfGyration\_anion
- 16. "'ECW\_computed'

<span id="page-3-0"></span>The Graph Convolutional Neural Network (GCN) was built upon the combination of RDkit and PyTorch Geometrics. The RDKit offers the node and edge features of cations and anions. PyTorch takes charge of the train and prediction work. The GCN model as below has two convolutional layers followed with the graph prediction layer. The iteration epoch times is 1000 to ensure stable accuracy values. The code is included in the class object published on GitHub (https://github.com/WangsGroupFDU/IL\_machine\_learning).



<span id="page-4-0"></span>The  $\mathbb{R}^2$  value is widely used to validate the performance of the models based on the predicted values and experimental values. It is very difficult to directly correlate the performance of the model with the  $\mathbb{R}^2$  determinant. The  $\mathbb{R}^2$  value is highly dependent on many factors, including the sample size, the data sources and the sample uniqueness, thus we have to evaluate case by case.

As far as we know, most of the related literature using machine learning to predict ionic conductivity are based on the ILThemro database, which contains 7234 entries of ionic conductivity values at varying temperatures for only 523 unique ILs, the unique cations and anions are high up to 244 and 109 correspondingly, indicating the database is highly sparse. In addition, we can estimate that there are  $\sim$  14 (7234/523) records for every unique ionic liquid in the dataset, thus there must be a large number of same ILs appearing both in the training and testing dataset, which will boost the  $R^2$  and lead to the overfitting of the model. Besides, the validation reported in the literature usually relays on the ILThermo database itself. There is seldom validation work using an external database. We didn't find other reported  $\mathbb{R}^2$  in the literature, which compares the unique ILs with external database at a single temperature as we report here.

When we do the comparison, 4 ILs contain more than two records as shown in Supplementary Table 6, but the variation between the two records is very high, this uncertainty from the database (experimental values) itself will further increase the fluctuation of our validation shown in Figure 2. Thus, we leave out this invalid value point for 1-butylpyridinium dicyanamide with a huge percentage of difference for the measured values (> 66%). Even though there are only 18 overlaps between the two datasets, we find that the model did a nice job to predict IL especially with high ionic conductivity, which is one of our targets.

Above all, we conclude that the model is important and the distinctive  $\mathbb{R}^2$  value is insightful to the field. This also indicates that we can pay more attention to the commonly existing bias of the database collection and management for future ML investigations.

### <span id="page-5-0"></span>**Supplementary Note 7**

The  $t_{Li}^+$  of Bruce-Vincent analysis is defined in Equation (3), where  $I_0$  is the current at the starting point of polarization,  $I_{ss}$  is the steady state current at the end of the polarization.  $R_0$  and Rss refer to the initial resistances of the first semi-circle before and after polarization, respectively. The frequency range used is 1Hz – 1MHz. The impedance spectra are attached in Supplementary Table 7.

$$
t_{\text{Li}}^{+} = \frac{I_{ss}(\Delta V - I_0 R_0)}{I_0(\Delta V - I_{ss} R_{ss})}
$$
 Equation (3)





<span id="page-6-0"></span>The conductivity and ECW boxplots for the ILs classified by the cation and anion types. (a, b) The boxplots of the ILs with known conductivity  $(\sigma)$  classified by the cation and anion types, corresponding. In terms of the cations, the average ionic conductivity (σ) follows the order of  $a$ mmonium > sulfonium > pyrrolidinium > imidazolium > pyridinium > piperidinium > phosphonium. Accordingly, the rule for the anions is nitrate > thiocyanate >  $DCA > BFA >$ triflate > sulfate > imide > sulfonate > acetate > halogen > phosph. (c, d) The boxplots of the ILs with known ECWs classified by the cation and anion types, correspondingly.



<span id="page-7-0"></span>Heatmap for the solid/liquid prediction results as shown in a permutation table between cations and anions. The black squares indicate solid IL pairs at room temperature. The pink squares refer to liquid IL pairs at room temperature. The patch highlighted with red squares are IL pairs with known state in the dataset.



<span id="page-8-0"></span>The mechanical properties of IPEs. (a) The stress-strain curve for IPE with 10% PBDT and C2mim TfO. (b) The corresponding DMA curve for IPE with 10% PBDT and C2mim TfO from −50 to 300 °C.



<span id="page-9-0"></span>**Supplementary Fig. 4**

Quantification of H<sub>2</sub>O amount in the IPEs using NMR and DSC. (a) <sup>1</sup>H NMR spectra for the membrane. The green dashed line shows the regular position  $(4.9 \text{ ppm})$  of  $H_2O$  peak.  $H_2O$  does have a detrimental effect on the performance of Li metal batteries. For this reason, the developed composites membranes of PBDT and ILs formed in the first step were placed in a vacuum oven at 80℃ for more than 24h to adequately remove water before assembled in the batteries. Here we mainly need to measure the  $H_2O$  in the dried membrane (10% PBDT C2mim TfO). We observe no distinct signal that belongs to H2O, which usually appears around 4.9 ppm, for the membrane. (b) DSC curve for the membrane. Notably, we observe no apparent heat absorption peaks above 100  $\degree$ C, which indicates that H<sub>2</sub>O molecules were successfully removed after the vacuum drying step. The excellent battery cycling performance in the manuscript also confirms that the effect of H2O can be neglected in the IPEs. The second ion exchange process was finished in an Ar-filled glove box  $(< 0.01$  ppm H<sub>2</sub>O).



<span id="page-10-0"></span>The ECW values for selected pure ILs. The electrochemical window of neat ILs using three electrodes design with Ag|AgCl (3M KCl) as the reference electrode and Pt as both the working and counter electrodes with diameter of 3 mm. The voltage has been converted based on the Li|Li+ potential.



<span id="page-11-0"></span>**Supplementary Fig. 6**

The cycling performance of IPEs based on DemsTFSI. (a) Cell voltage versus time for a symmetric Li|IPEs-DemsTFSI|Li cell at current densities (J) from 0.1 to 6 mA·cm<sup>-2</sup> with changes in *J* every 10 cycles at room temperature (each cycle lasts 1 h). (b) Cycling performance of Li|IPEs-DemsTFSI|LiFePO4 cell at 0.5C (0.83 mA cm<sup>-2</sup>) at RT. The blue circles show the specific discharge capacity as a function of the increasing cycle number. The black circles display the CE for each cycle correspondingly. (c) The voltage-capacity profiles for the main cycles in (d).



<span id="page-12-0"></span>SEM image of the cycled Li metal surface. The scale bar is 100 μm. We observe no Li dendrite formed on the Li metal surface for the Li||Cu cell, which indicates the formation of stable SEI in Li metal surface based on the developed IPEs.

<span id="page-13-0"></span>The calculated binding energies for selected cations and anions pairs (19) labeled with phase at RT from IoLiTec.



<span id="page-14-0"></span>The calculated binding energies for selected cations and anions pairs (72) labeled with predicted phases at RT.









<span id="page-17-0"></span>ANOVA results for the hypothesis testing.



### <span id="page-17-1"></span>**Supplementary Table 4**

T-test with unequal variance results for the hypothesis testing.



<span id="page-18-0"></span>The final recommendation list with 49 ILs for IPEs.





<span id="page-20-0"></span>The conductivity records stored in ILThermo for the 18 overlapped ILs.



### <span id="page-20-1"></span>**Supplementary Table 7**

The t<sub>Li</sub><sup>+</sup> of Bruce-Vincent analysis of the IPEs developed from different ILs.



1 Ong, S. P., Andreussi, O., Wu, Y., Marzari, N. & Ceder, G. Electrochemical Windows of Room-Temperature Ionic Liquids from Molecular Dynamics and Density Functional Theory Calculations. *Chem Mater* **23**, 2979-2986, doi:10.1021/cm200679y (2011).

2 Kuusik, I., Kook, M., Pärna, R. & Kisand, V. Ionic Liquid Vapors in Vacuum: Possibility to Derive Anodic Stabilities from DFT and UPS. *ACS omega* **6**, 5255-5265, doi:10.1021/acsomega.0c05369 (2021).