Supplementary Materials for

Machine learning-guided optimization for ionic liquid-based polyethylene terephthalate waste recycling

Ji Gao[a]†, Wenbo Peng[a]†, Andres Galindo[a], Ethan Slaton[b], Jose Perez Martinez[a], Guanghui Lan[c], Zhaohui Tong[a]*

*Corresponding author. Email: Zhaohui Tong. zhaohui.tong@chbe.gatech.edu

The PDF file includes:

Supplementary Text
Table S1
Figs. S1 to S26
References

Other Supplementary Materials for this manuscript include the following:

Data S1

Supplementary Text

Synthesis of ionic liquids

In our experimentation step, we applied the following steps to mitigate impurity of water in the ILs: 1) All synthesized ILs were dried under vacuum at elevated temperatures (typically 70 °C) overnight before use. Second, we ensured minimal exposure to ambient humidity by storing them in a desiccator and minimizing air exposure during weighing and reaction setup. 2) All the reagents were dried before the synthesis of the ILs.

[N1111][$ZnCl_3$]: In a 50 mL flask, 10 mmol (1.3630 g) of $ZnCl_2$ in 10 mL of DI water and 0.05 mL of concentrated HCl, and 10 mmol (1.0960 g) of [N1111][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

$[N1111]_2[ZnCl_4]$: In a 50 mL flask, 10 mmol (1.3630 g) of $ZnCl_2$ was dissolved in 10 mL of DI water and 0.05 mL of concentrated HCl, and 20 mmol (2.1920 g) of [N1111][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

[TMG][Cl]: In a 100 mL Erlenmeyer flask, 10 mL of concentrated hydrochloric acid (~0.12 mol) was added dropwise and slowly to 12.55 mL of 1,1,3,3-tetramethylguanidine under vigorous stirring. After the addition, 10 mL of DI water was added to the mixture, and the mixture was kept under stirring for 4 h. The mixture was then dried in the vacuum oven overnight.

$[TMG]_2[ZnCl_4]$: In a 50 mL flask, 10 mmol (1.3630 g) of $ZnCl_2$ was dissolved in 10 mL of DI water and 0.05 mL of concentrated HCl, and 20 mmol (3.0328 g) of [TMG][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

[TMG][For]: In a 100 mL Erlenmeyer flask, 5 mL of formic acid was added dropwise and slowly to 12.55 mL of 1,1,3,3-tetramethylguanidine (liquid) and 20 mL of DI water under vigorous stirring. After the addition, the mixture was kept at 70℃ under stirring for 4 hours, and the mixture was dried in the vacuum oven overnight.

[Ch][$ZnCl_3$]: In a 50 mL flask, 10 mmol (1.3630g) of $ZnCl_2$ was dissolved in 10 mL of DI water and 0.05 mL of concentrated HCl, and 10 mmol (1.3962 g) of [Ch][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

$[Ch]_2[ZnCl_4]$: In a 50 mL flask, 10 mmol (1.3630g) of $ZnCl_2$ was dissolved in 10 mL of DI water and 0.05 mL of concentrated HCl, and 20 mmol (2.7924 g) of [Ch][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

[Emim][$ZnCl_3$]: In a 50 mL flask, 10 mmol (1.3630 g) of $ZnCl_2$ was dissolved in 10 mL of DI water and 0.05 mL of concentrated HCl, and 10 mmol (1.4617 g) of [Emim][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

[Emim]$_2$[$ZnCl_4$]: In a 50 mL flask, 10 mmol (1.3630 g) of $ZnCl_2$ was dissolved in 10 mL of DI water and 0.05 mL of concentrated HCl, and 20 mmol (2.9324 g) of [Emim][Cl] was slowly added to the solution. The mixture was kept at 60℃ under vigorous stirring for 4 h and dried in the vacuum oven overnight.

Bulk sale prices for ionic liquid anion and cation components

Bulk sale prices for anion components in USD:

- **$ZnCl_2$**: 1200, https://www.quheqihuo.com/news/202306272957390.html
- **$CoCl_2$**: 6800, https://www.mysteel.com/hot/1232742.html
- **$CrCl_2$**: 10000, https://www.alibaba.com/product-detail/Chromium-ii-Chloride-CAS-10049-05_1600599662519.html
- **$CuCl_2$**: 8200, https://www.alibaba.com/product-detail/Industrial-grade-Copper-II-chloride-dihydrate_1600459090540.html
- **Alanine**: 3900, https://www.chinabgao.com/jiage/1469574.html
- **Serine**: 3400, https://www.alibaba.com/product-detail/High-quality-and-Low-price-L_1601044123712.html
- **Acetic Acid**: 386, https://www.sci99.com/monitor-499-0.html
- **Glycine**: 2000, https://www.chinabgao.com/jiage/ganansuan/7.html
- **Formic Acid**: 535, https://finance.sina.cn/futuremarket/nyshzx/2023-07-05/detail-imyzqyat3239520.d.html
- **Aspartic Acid**: 6800, https://www.alibaba.com/product-detail/Factory-High-Grade-L-Aspartic-acid_1600552057471.html
- **$MnCl_2$**: 1400, https://www.alibaba.com/product-detail/Factory-Price-Sell-Industrial-Grade-Manganous_1600236042763.html
- **$Co(Ac)_2$**: 9000, https://www.alibaba.com/product-detail/High-Quality-Cobalt-II-Acetate-Tetrahydrate_1600697473625.html
- **$CoCl_2$**: 6900, https://www.mysteel.com/hot/1486340.html
- **Lysine**: 1270, **Tryptophan**: 9500, https://ncp.mysteel.com/23/0625/17/DCC221066E75203B.html
- **$FeCl_3$**: 2000, https://www.alibaba.com/product-detail/98-Ferric-Chloride-Anhydrous-IRON-III_1600281287607.html
- **$Zn(Ac)_2$**: 2100, https://www.alibaba.com/product-detail/Best-Price-Zinc-Acetate-Anhydrous-CAS_1600827762025.html

- **Proline**: 18000, https://www.100ppi.com/mprice/detail-5231231.html
- **$Cu(Ac)_2$**: 5110, https://www.100ppi.com/news/detail-20230625-2692835.html
- **H_3PO_4**: 870, https://www.100ppi.com/kx/detail-message-558--1.html
- **Histidine**: 44000, https://www.100ppi.com/mprice/detail-5231233.html
- **Leucine**: 8700, https://www.alibaba.com/product-detail/OEM-Amino-Acid-Powder-CAS-61_1600515411949.html
- **$NiCl_2$**: 6700, https://www.chinabgao.com/jiage/lvhuanie/
- **Arginine**: 9400, https://www.quheqihuo.com/news/202306282959642.html
- **$Mn(Ac)_2$**: 2600, https://www.quheqihuo.com/news/202305302938694.html
- **$Ni(Ac)_2$**: 9670, https://www.100ppi.com/mprice/detail-4747651.html
- **Butane**: 1321, https://www.100ppi.com/mprice/plist-1-1615-1.html
- **$NaHCO_3$**: 250, https://www.alibaba.com/product-detail/Bangze-Sodium-Bicarbonate-Manufacturing-nahco3-Sodium_1600534966474.html
- **$NaHSO_4$**: 250, https://www.alibaba.com/product-detail/Sodium-Hydrogen-Sulfate-sodium-bisulfite-CAS_1600063098933.html
- **Imidazole**: 8343, https://detail.1688.com/offer/667148161718.html
- **MsOH**: 1600, https://www.alibaba.com/product-detail/Factory-Supply-CAS-75-75-2_1600886557882.html

Bulk sale price for cation components in USD:

- **AMIMCl**: 2760, https://www.zhaosw.com/product/detail/258728715
- **ChCl**: 3000, https://www.alibaba.com/product-detail/Choline-Chloride-Powder-99-CAS-67_1600612814490.html
- **ChOH**: 3000, https://www.zhaosw.com/product/detail/252391238
- **TMG**: 5000, https://www.alibaba.com/product-detail/High-Quality-Tetramethylguanidine-CAS-80-70_10000012371707.html
- **N2222OH**: 8290, https://www.zhaosw.com/product/detail/254355183
- **N1111OH**: 4000, https://www.alibaba.com/product-detail/High-Quality-Tetramethylammonium-hydroxide-pentahydrate-TMAH_1600219538414.html
- **BMIMCl**: 2000, https://www.alibaba.com/product-detail/High-Quality-99-1-Butyl-3_11000002467025.html
- **HMIMCl**: 2760, https://www.zhaosw.com/product/detail/22723499
- **DMIMCl**: 2600, https://www.zhaosw.com/product/detail/243293490
- **EMIM**: 2300, https://www.alibaba.com/product-detail/1-Ethyl-3-methylimidazolium-Chloride-65039_1600278467682.html
- **KOH**: 800, https://www.alibaba.com/product-detail/Potassium-Hydroxide-KOH-90-Flakes_1600695538145.html

Emission and price data for simulation

- CO_2 emissions for ethylene glycol: **1.418** tCO2e / t
  https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R2066&rid=1#ntr5-L_2018334EN.01007901-E0005

- Price of ethylene glycol: **660** USD / t
  https://www.echemi.com/pip/ethylene-glycol-eg-pid_Seven2471/asia.html

- CO_2e calculation for electricity and steam and heat:

  CO_2e multipliers: CH_4=25, N_2O=298
  2023 eGrid US Average data for electricity emission in lb/MWh
  CO_2=852.3, CH_4=0.071, N_2O=0.010
  2023 eGrid US Average data for steam and heat emission in kg/mmBtu
  CO_2=66.33, CH_4=1.250e-3, N_2O=0.125e-3
  CO_2e electricity = **0.389** kg / kWhr
  CO_2e steam = **6.293e-05** kg / kJ
  https://www.epa.gov/climateleadership/ghg-emission-factors-hub

Conditions for screening data generation

Ionic liquids are generated using the combinations of existing cations and anions in the database. The initial conditions used for case generation are catalyst loading (mass ratio to PET): [0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.15], solvent loading (mass ratio to PET): [4, 6, 8, 10], temperature: [160, 170, 180, 190, 200], and reaction time: [30, 60, 90, 120, 150, 180]. The minimum solvent loading is further limited to 4 after experiment, as lower amounts would make stirring difficult. Reaction times below 60 minutes and catalyst loading of 0.01 were excluded due to feasibility reasons as the melting time is around 30 to 60 minutes for PET bottles and powder based on observation during experiments. The maximum temperature was limited below the boiling point of ethylene glycol under atmospheric pressure.

Machine learning implementation details

The reactant quantities are preprocessed as ratios relative to the PET amount. Thus, the PET quantity feature can be set to unity for all cases and omitted. The categorical feature, PET source, is encoded using one-hot encoding. Each atom in the molecular data is embedded with five features: atomic number, formal charge, hybridization, aromaticity (True/False), and molecular weight. Each bond is embedded with three features: bond type, stereochemistry, and conjugation (True/False). All embeddings are generated using RDKit following the Torch Geometric framework, with categorical features encoded via label encoding.

The molecular connectivity is represented using a tuple-based format, where ($atom_i$, $atom_j$) indicates a bond between atom i and atom j. For each IL molecule, the cation and anion are

embedded in a single graph without connections between them, reflecting their separation. Furthermore, only one cation and one anion are embedded per IL molecule, irrespective of their ratio, as this information can be inferred from the charge balance of the ions in the dataset.

For graph neural network part of the model, we used 4 graph attention layers of (input features, output features, heads) equal to [(5, 16, 4), (64, 32, 4), (128, 32, 4)*5, (128, 32, 2)] followed by a linear layer of (64, 32), so that each ionic liquid molecule is featurized into a vector of length 38. A 70% dropout of is added to the penultimate graph attention layer to avoid overfitting. The featurized ionic liquid molecule is then concatenated with 6 reaction parameters including catalyst loading, solvent loading, temperature, reaction time, source of PET, and size of PET. This concatenated feature is fed into the fully connected layer of size (38, 80, 96, 1) with dropout 5% for each intermediate layers for yield prediction. We applied RAdam optimizer with learning rate of 1e-3 and weight decay of 2e-4.

t-SNE plot for GNN-featurized IL outputs.

To analyze the feature representations learned by our GNN, we applied t-distributed stochastic neighbor embedding (t-SNE) to the GNN-featurized IL outputs (GNN output before fully connected layer). By tuning the perplexity parameter, we found that a value of 48 produced a clustered visualization, capturing relationships within the data.

To further investigate the features learned by the GNN, we generated two separate t-SNE plots (Fig. S24). In one plot, the points are labeled by their cation names and another is labeled by their anion names. These plots show that the GNN can featurize IL molecules based on structural and chemical similarities. Specific groups of ILs form distinct clusters, reinforcing the model's ability to capture meaningful molecular features. For example, quaternary ammonium compounds, zinc chloride-based ILs, and amino acid-based ILs form their own clusters, indicating that the model recognizes underlying molecular patterns.

DBSCAN clustering analysis

We carried out outlier analysis using Density-Based Spatial Clustering of Applications (DBSCAN)[1] for IL molecules with GNN featurization, shown in Fig. S26. We applied t-SNE coordinates with reduced dimensions for better visualization in the plots. Clusters are labelled using the same colors. Note that each point/circle in the figures indicates multiple data points as we have repeated IL molecules.

Fig. S26A shows the DBSCAN clustering results with raw GNN output features. Using GNN without further dimension reduction methods like t-SNE can already provide good clustering results for the ILs including both cation and anion. The algorithm identified 21 clusters with 8 outliers (including AMIM, BMIM based metal-Cl ILs, [Ch][OH], and [TMG] [ZnCl3]). Clustering after applying t-SNE to GNN features (Fig. S26B) yielded similar clusters (21) but reduced outliers (3), by unsupervised learning on the GNN outputs. Incorporating reaction conditions (Fig. S26C) along with IL features resulted in 11 clusters and 7 outliers. Importantly, as detailed in Table S1, these outliers were primarily data points associated with extreme reaction conditions, rather than unique molecular structures overlapping with the previous outliers. These points, while distinct in

feature space, represent valid experimental observations at the tail of the distribution. Such outliers often contain crucial information about system limits and should not be removed[2].

Hyperparameter tuning

We carried out hyperparameter optimization with Optuna for eight parameters including batch size, number of GNN layers, GNN output size (size of vector to embed one IL), fully connected layer dimensions, GNN dropout rate, fully connected layer dropout rate, optimizer learning rate, optimizer weight decay, and whether to use scheduler. We discretized all the continuous parameters, and define their ranges as follows:

- Batch size: 8 to 64, interval = 4
- Number of GNN layers: 0 to 5, interval = 1
- GNN output size: 4 to 64, interval = 4
- Number of fully connected layers: 1 to 5, interval = 1
- Each fully connected layers size (separately): 16 to 128, interval =16
- GNN dropout rate: 0 to 0.8, interval = 0.05
- Fully connected layer dropout rate: 0 to 0.2, interval = 0.05
- Optimizer learning rate: 1e-4, 1e-3, interval =1e-4
- Optimizer weight decay: 1e-4, 1e-3, interval =1e-4
- Scheduler: True/False

Two hundred case studies were carried out with Hyperband pruning method. In this case, a larger testing dataset is used, and the dataset was separated randomly in 75% to 25% ratio for training and testing. We defined a customized optimization objective using the weight sum of exponentially smoothed MAE and R^2, which helps to optimize both loss and goodness of fit. We applied an exponential smoothing factor of 0.05 to smooth the fluctuations. We define the optimization objective as: MAE - 20*R^2. The weighting factors are chosen using the value range of MAE (~5-30) and R^2 (0-1).

Fig. S25A showed the sampled points and the progression of the optimization process. Fig. S25B showed the importance of the hyperparameters based on the samples. Weight decay for the optimizer is the most important factor, as it controls the regularization of the optimization, hence affects the degree of overfitting. The GNN featurization size is the second important factor, and it controls the complexity of the latent representation of IL molecules. Fig. S25C, showed a combined visualization of objective value, trials, and each of the parameters explored.

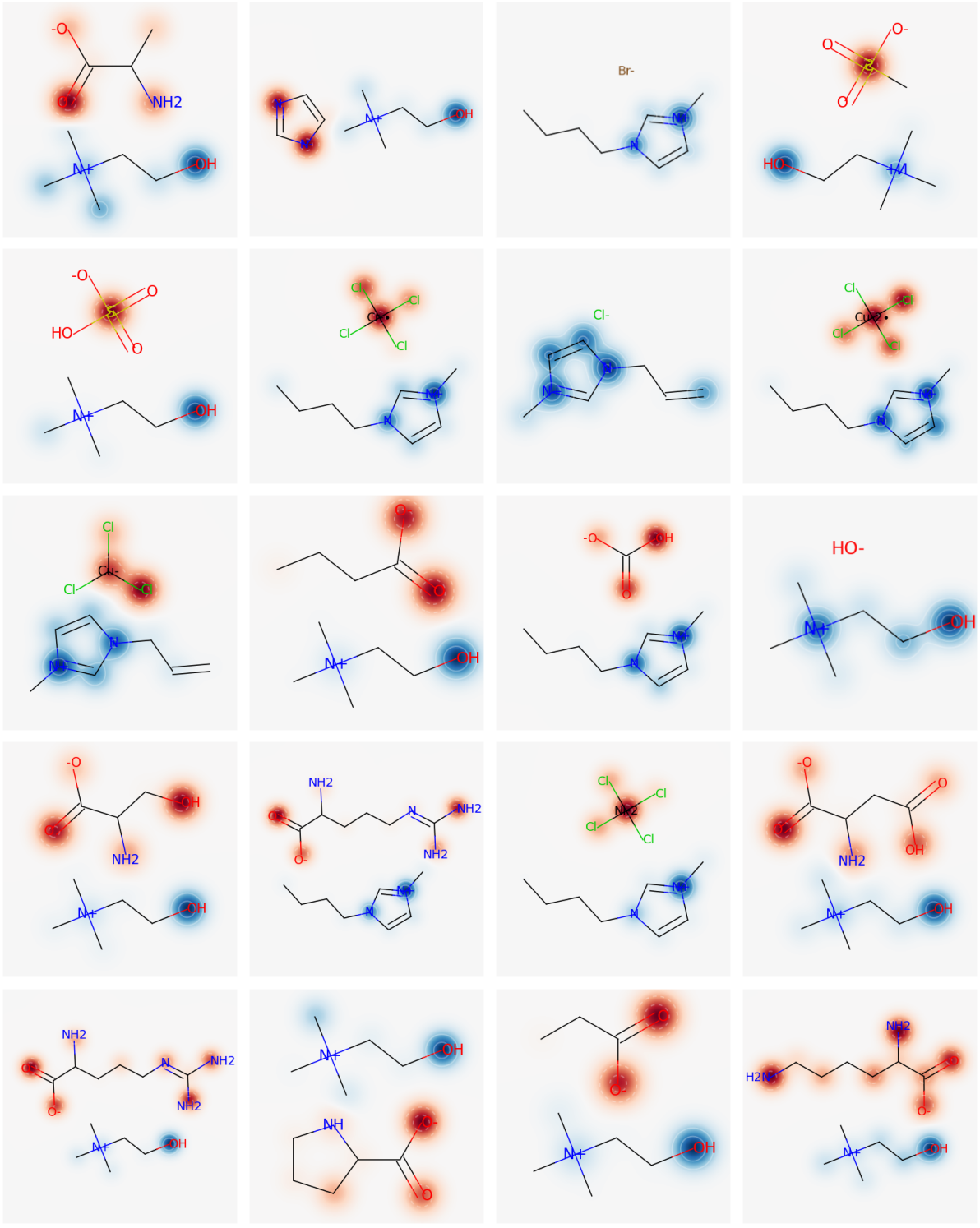
We then applied stratified 10-fold cross validation on the best hyperparameter set to determine the training epoch. The stratification criterion was based on reaction yield, dividing the data into intervals of (0, 20, 40, 60, 80, 100) to ensure balanced representation across different yield ranges. In the Figure on the left, we plotted the difference of average cross-validation MAE loss (MAE_{t+1}-MAE_t). The curve shows an approximation of speed of decrease of loss. We observe that the delta validation loss curve has three stages, where the gradient clearly changes (Fig. S4). We then select epoch 2000 as the training epochs, from the right figure, we can see that it gives a reasonable loss without a strong difference in training and testing loss, to deal with the potential overfitting issue.

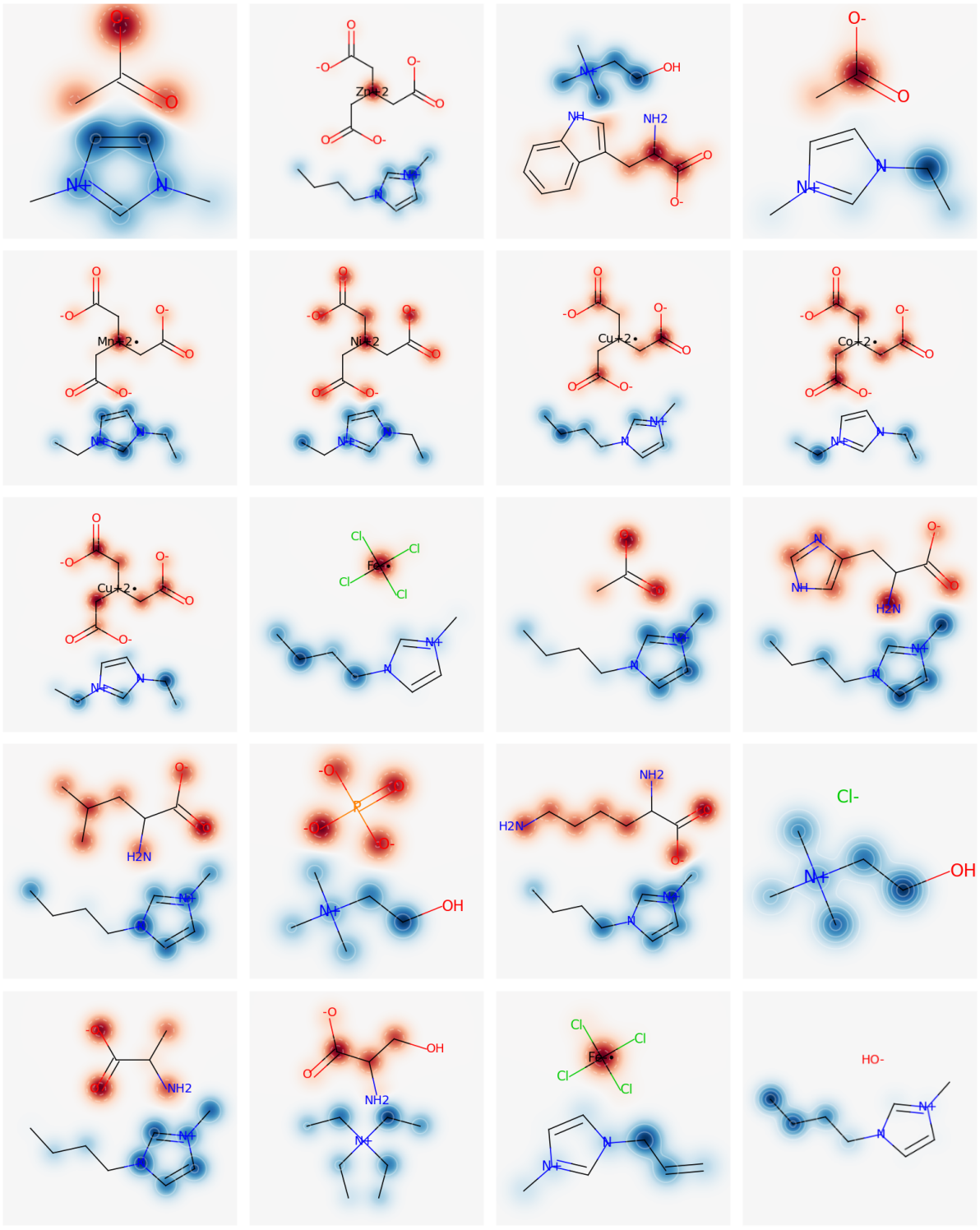
Finally, we applied the optimized training epoch and hyperparameters for 10-fold cross validation one more time to obtain the performance. We achieved an MAE of 0.12 and R^2 of 0.64.

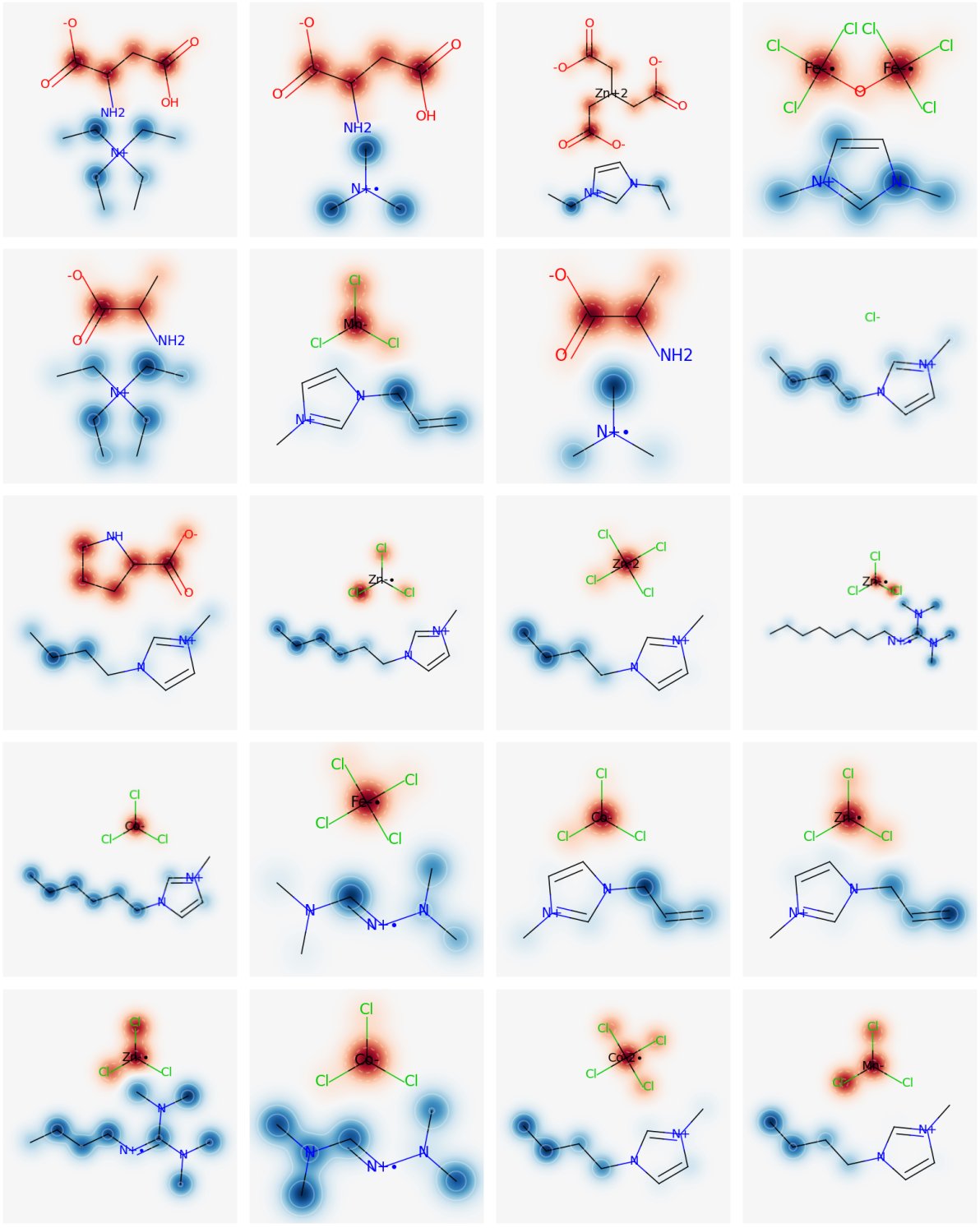
To assess overfitting and model performance limits, we conducted a bias/variance analysis based on definitions (Bias = training error; Variance = validation error – Bias)[3]. As detailed in Figure S4, the model achieved a bias of approximately 9.4% and a variance of around 2.8% near epoch 2000. The low variance suggests that overfitting was effectively controlled. This control was achieved through a combination of strong regularization techniques: a large dropout[4] value of 70% applied to the GNN part of the model, 5% dropout applied to each fully connected layer, optimizer regularization, and early stopping. These measures collectively prevent overfitting with limited data and enhance model robustness. We also note that a significant portion of the bias likely originates from unavoidable experimental error (estimated at ~5% or higher for reaction yields under varying conditions), implying a smaller avoidable bias.

Table S1. Outliers from DBSCAN clustering. Values at the tails of the population are highlighted.

	cation_name	anion_name	catalyst_amount	solvent_amount	temperature_c	reaction_time_min
93	BMIM	Cl	0.80	4.0	170	480
167	BMIM	Cl	0.05	11.0	190	480
236	DEIM	Zn(Ac)3	0.67	6.7	180	150
254	DEIM	Zn(Ac)3	0.50	20.0	180	150
281	Ch	Gly	0.19	4.0	120	240
293	Ch	PO4	0.21	4.0	150	360
298	Ch	Cl	0.10	4.0	180	840


-O
NH2
OH
Br-
O-
HO
Cl-
Cl
Cu-
HO-


Zn+2
Mn+2
Ni+2
Cu+2
Co+2
Cu+2
Fe
Cl
Cl
Cl
Cl
Fe
Cl
Cl
Cl
Cl-
HO-
OH
NH2
H2N



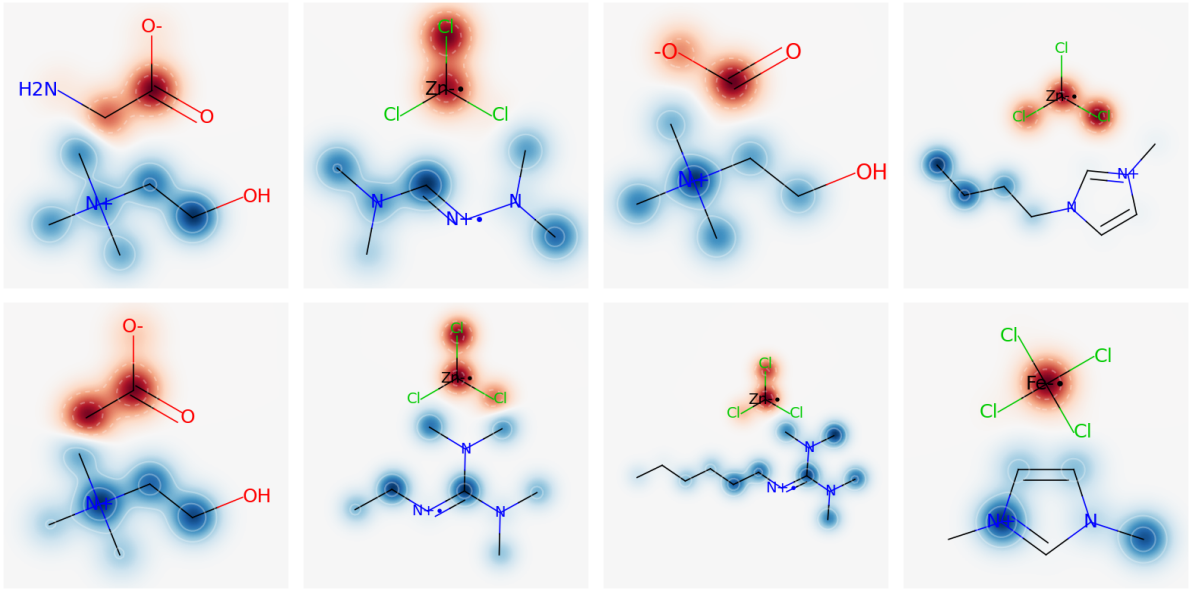


Fig. S1. Node importance for literature ionic liquid cations and anions.
Darker color indicates comparatively stronger contribution to increasing yield for PET glycolysis reaction.

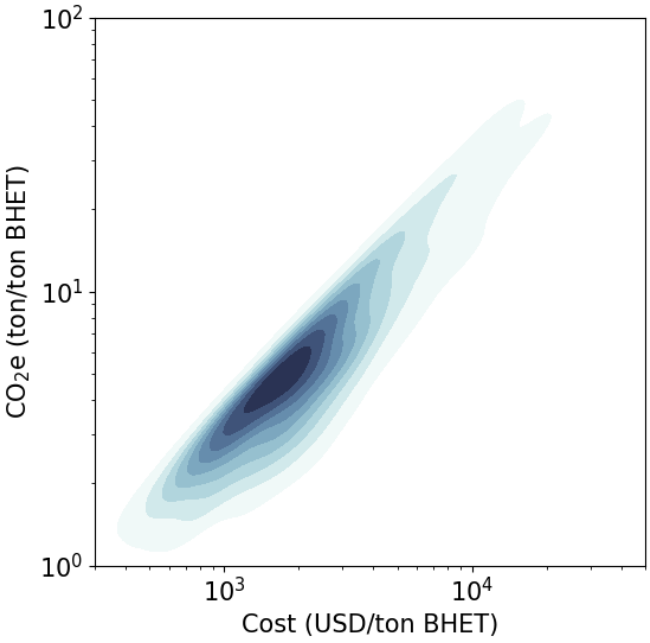


Fig. S2. Full distribution plot of costs and emissions for process simulation

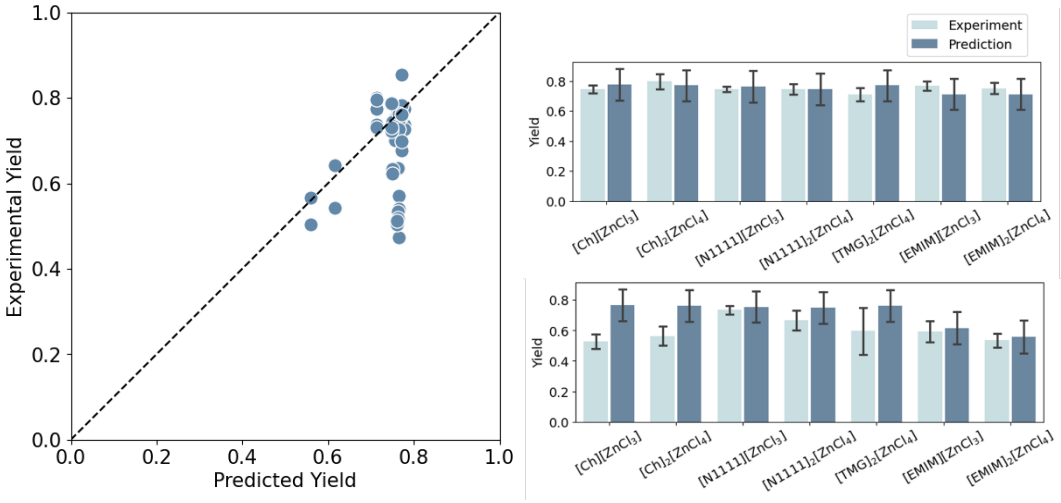


Fig. S3. Scattered and bar plots of predicted and experimental data.

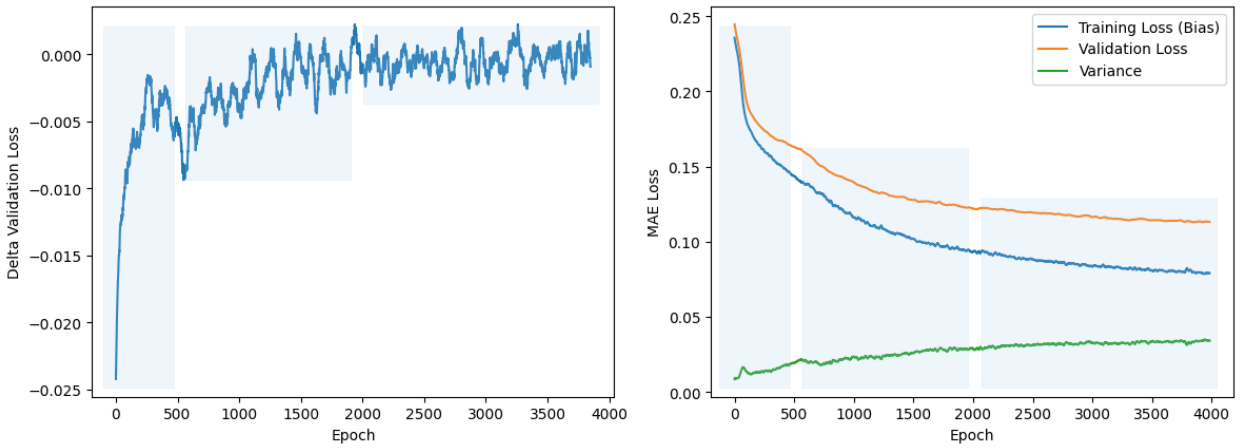


Fig. S4. Average metrics of Stratified 10-Fold cross validation for the optimized hyperparameters.

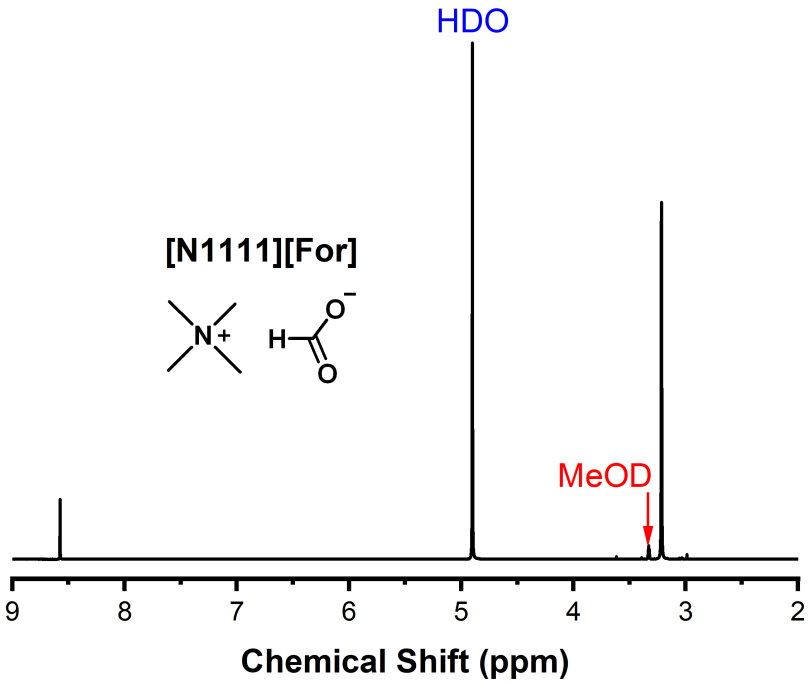


Fig. S5. NMR spectrum of [N1111][For], ^{1}H NMR (400 MHz, MeOD) δ 8.58 (s, 1H), 3.24 – 3.19 (m, 12H).

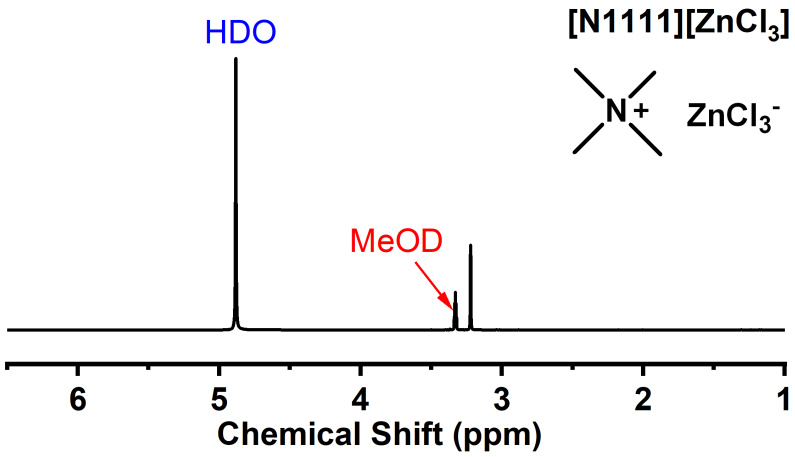


Fig. S6. NMR spectrum of [N1111][$ZnCl_3$], ^{1}H NMR (400 MHz, MeOD) δ 3.25 – 3.20 (m, 12H).

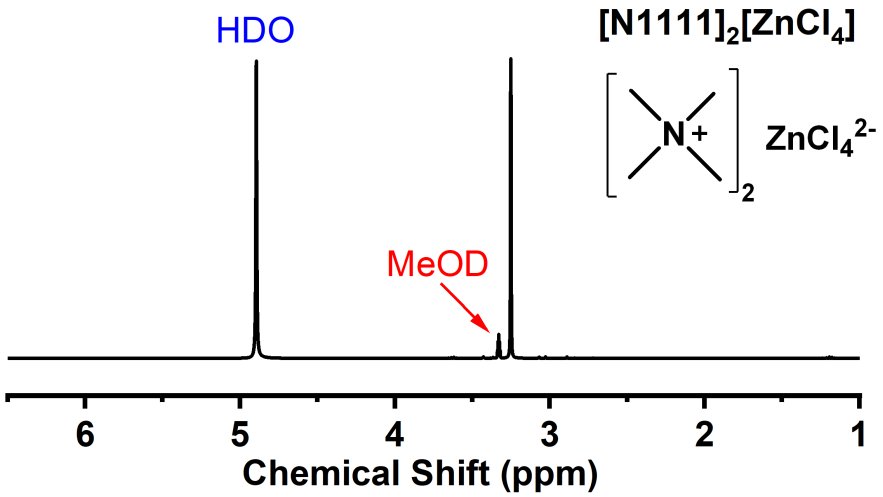


Fig. S7. NMR spectrum of $[N1111]_2[ZnCl_4]$, ^1H NMR (400 MHz, MeOD) δ 3.25 (d, 12H).

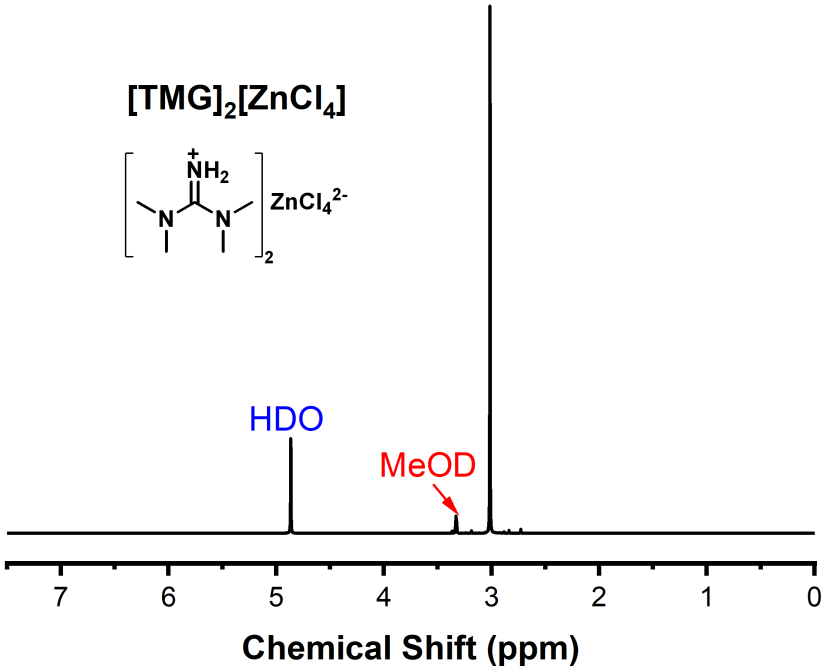


Fig. S8. NMR spectrum of $[TMG]_2[ZnCl_4]$, ^1H NMR (400 MHz, MeOD) δ 3.01 (s, 12H).

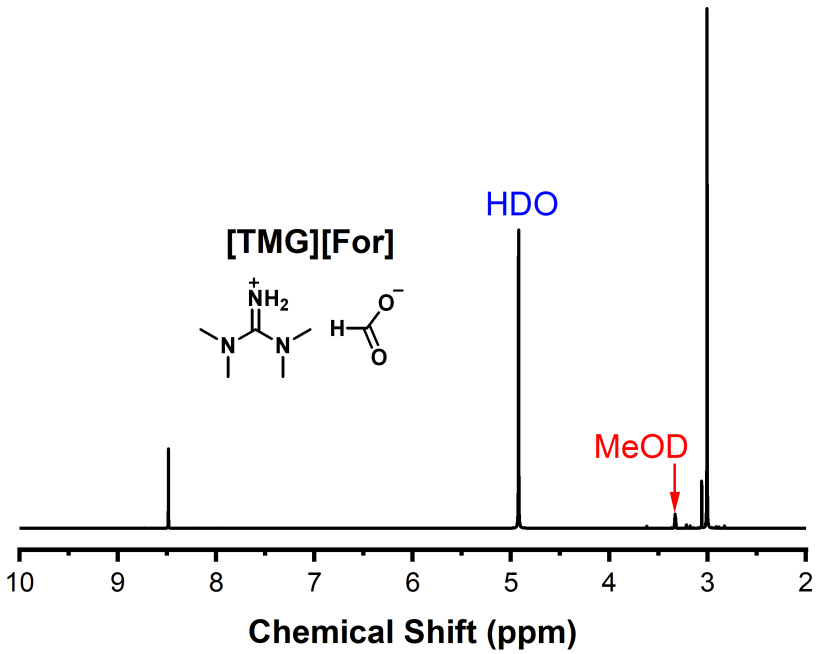


Fig. S9. NMR spectrum of [TMG][For], ^{1}H NMR (400 MHz, MeOD) δ 8.49 (s, 1H), 3.06-3.00 (m, 12H).

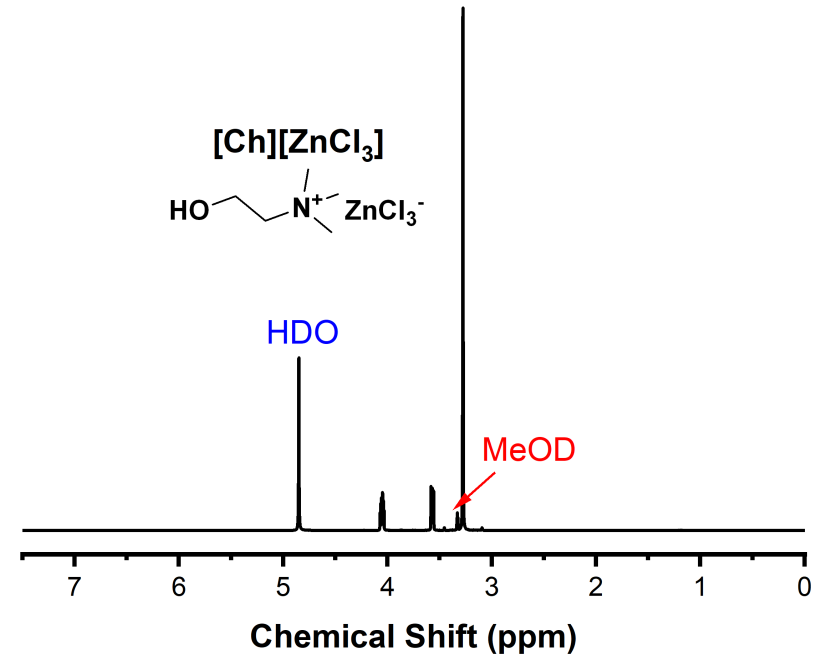


Fig. S10. NMR spectrum of [Ch][$ZnCl_3$], ^{1}H NMR (400 MHz, MeOD) δ 4.09 – 4.01 (m, 2H), 3.61 – 3.54 (m, 2H), 3.28 (d, J = 0.9 Hz, 9H).

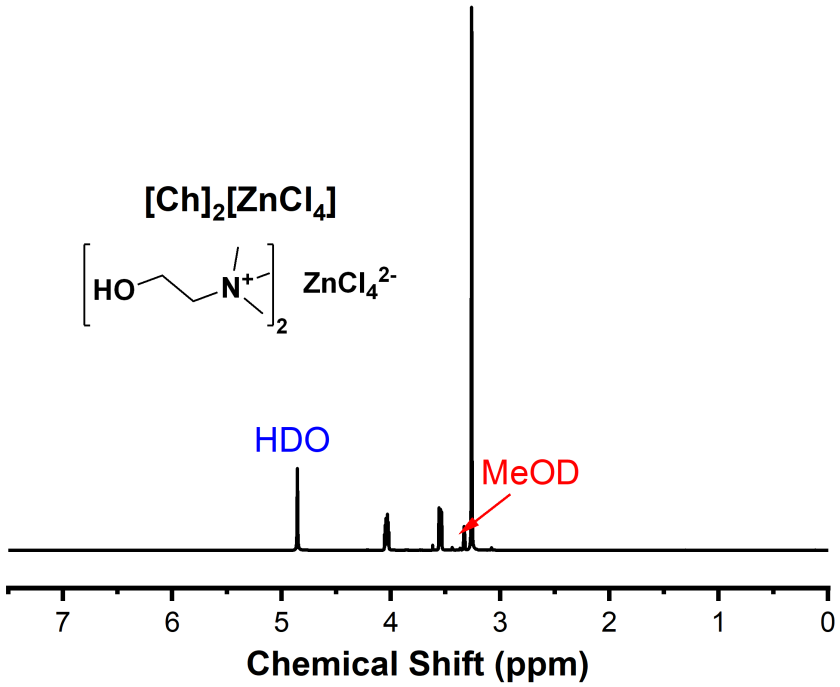


Fig. S11. NMR spectrum of $[Ch]_2[ZnCl_4]$, ^1H NMR (400 MHz, MeOD) δ 4.12 – 3.92 (m, 2H), 3.59 – 3.43 (m, 2H), 3.26 (d, J = 0.9 Hz, 9H).

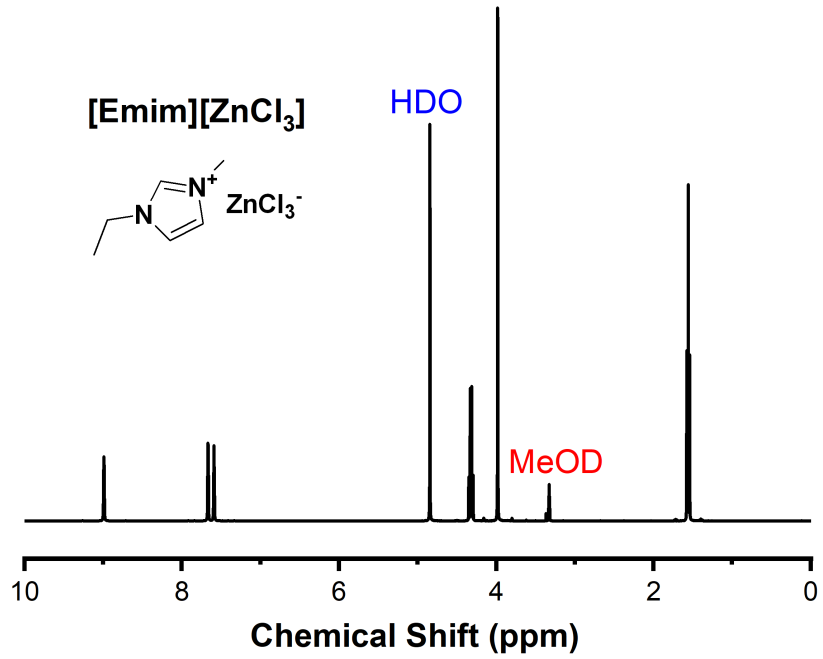


Fig. S12. NMR spectrum of $[Emim][ZnCl_3]$, ^1H NMR (400 MHz, MeOD) δ 8.99 (d, J = 1.9 Hz, 1H), 7.63 (dt, J = 30.4, 1.9 Hz, 2H), 4.32 (q, J = 7.3 Hz, 2H), 3.98 (d, J = 0.6 Hz, 3H), 1.56 (t, J = 7.3 Hz, 3H).

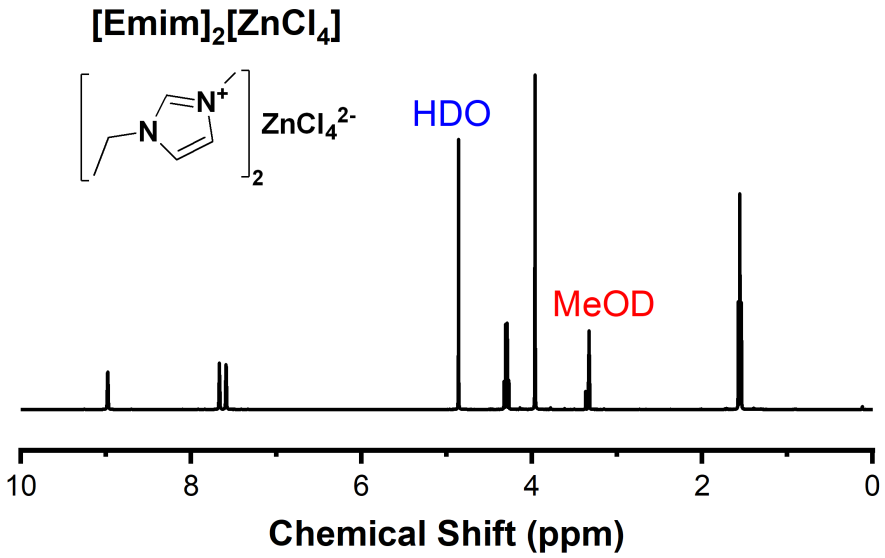


Fig. S13. NMR spectrum of [Emim][$ZnCl_4$], ^{1}H NMR (400 MHz, MeOD) δ 8.98 (d, J = 1.9 Hz, 1H), 7.63 (dt, J = 31.3, 1.8 Hz, 2H), 4.30 (q, J = 7.3 Hz, 2H), 3.96 (d, J = 0.5 Hz, 3H), 1.56 (t, J = 7.4 Hz, 3H).

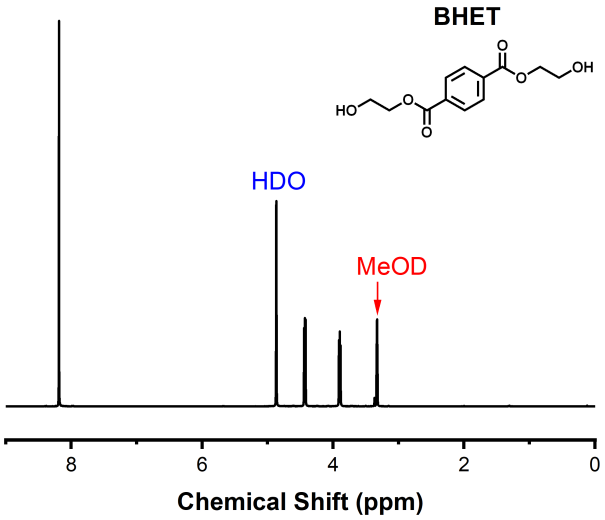


Fig. S14. NMR spectrum of bis(2-hydroxyethyl) terephthalate (BHET), ^{1}H NMR (400 MHz, MeOD) δ 8.19 (s, 4H), 4.51 – 4.32 (m, 4H), 4.00 – 3.79 (m, 4H).

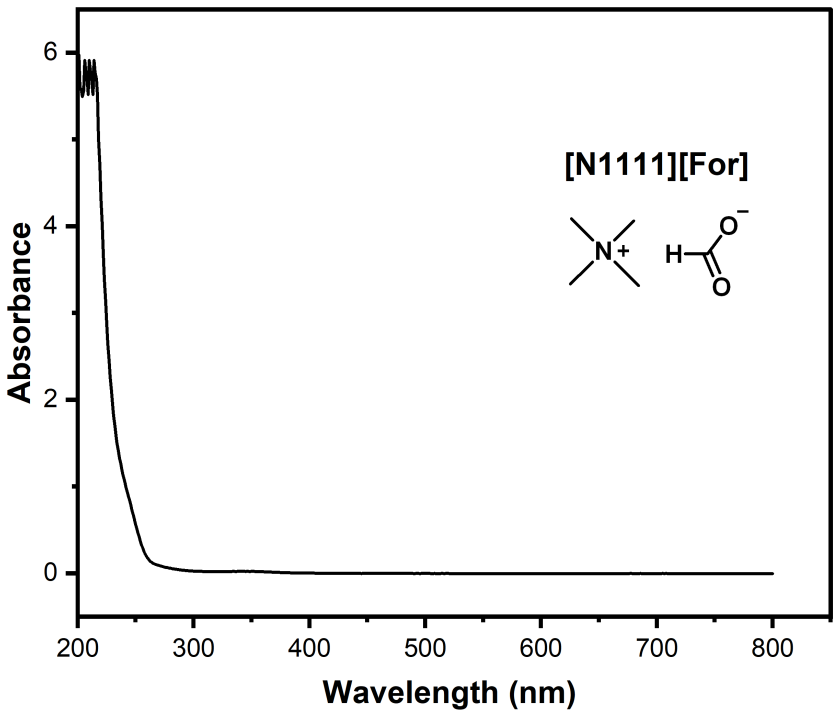


Fig. S15. UV-Vis spectrum of [N1111][For].

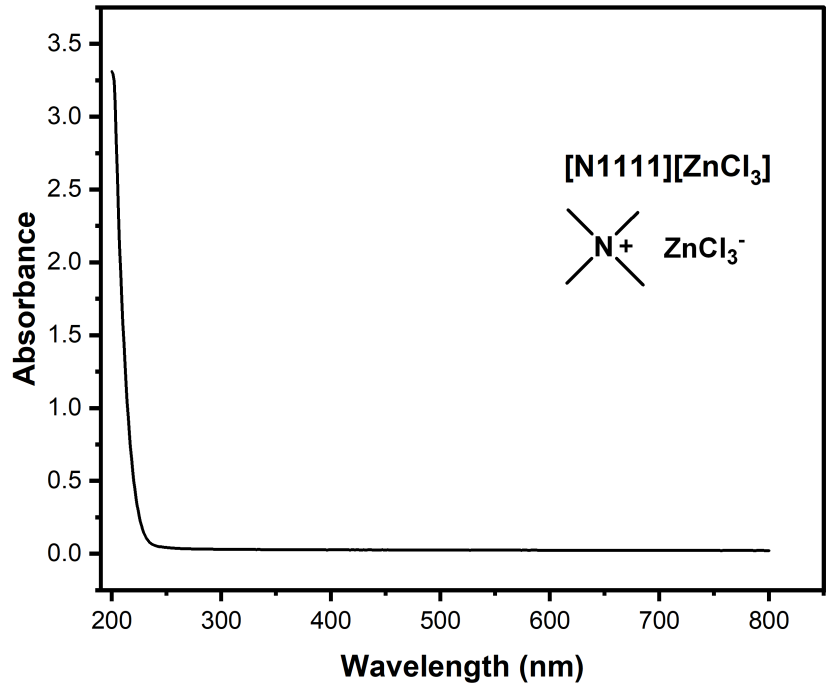


Fig. S16. UV-Vis spectrum of [N1111][$ZnCl_3$].

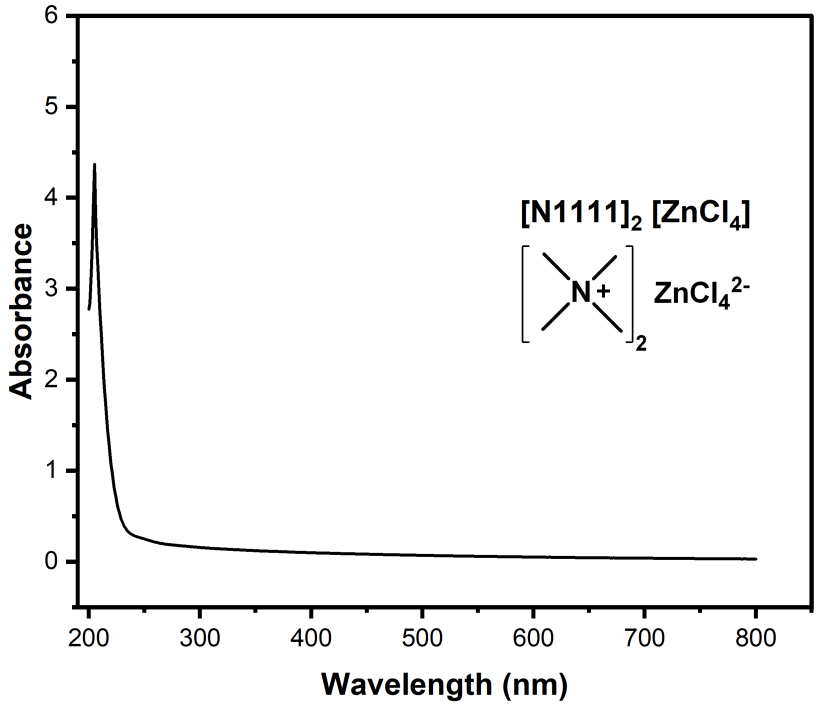


Fig. S17. UV-Vis spectrum of $[N1111]_2[ZnCl_4]$.

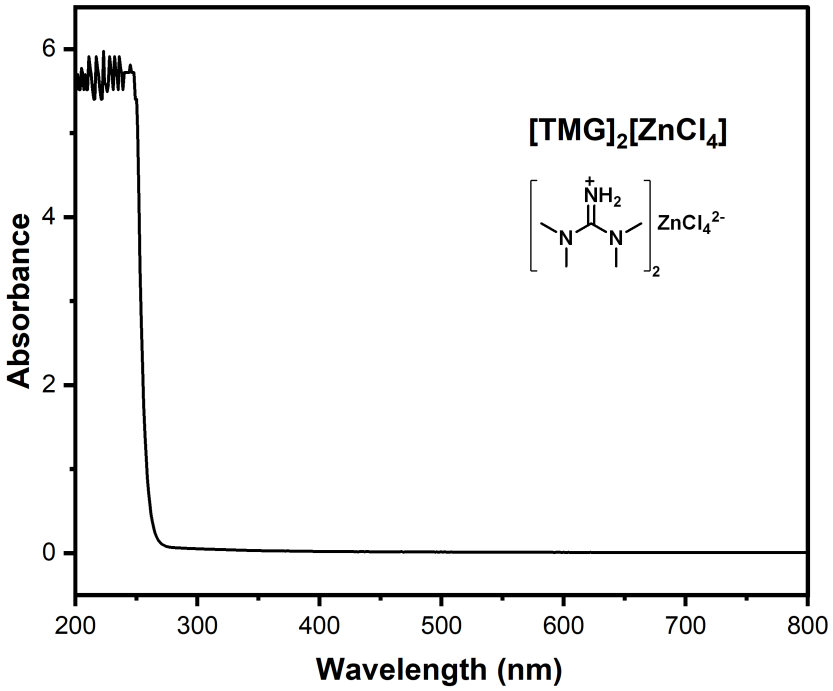


Fig. S18. UV-Vis spectrum of $[TMG]_2[ZnCl_4]$.

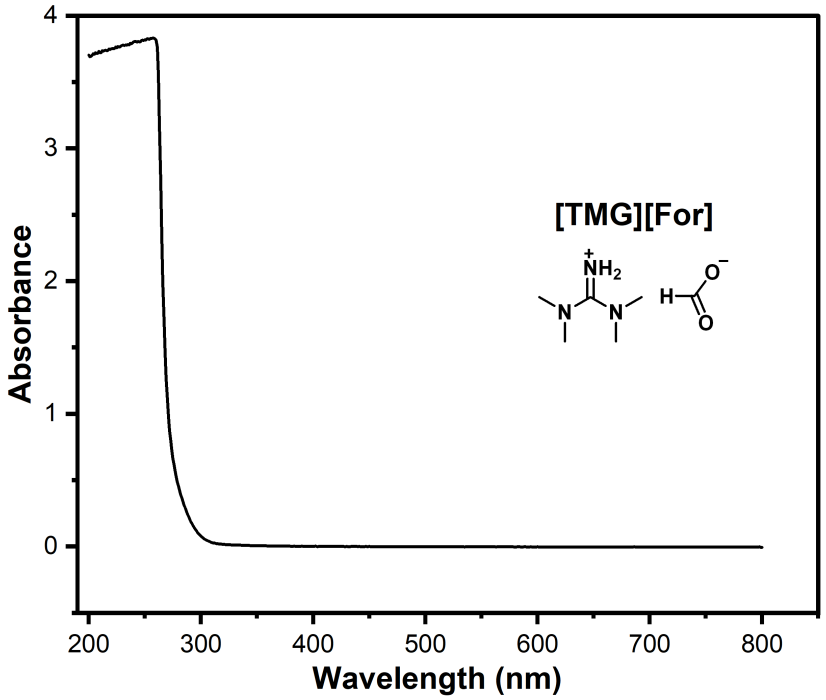


Fig. S19. UV-Vis spectrum of [TMG][For].

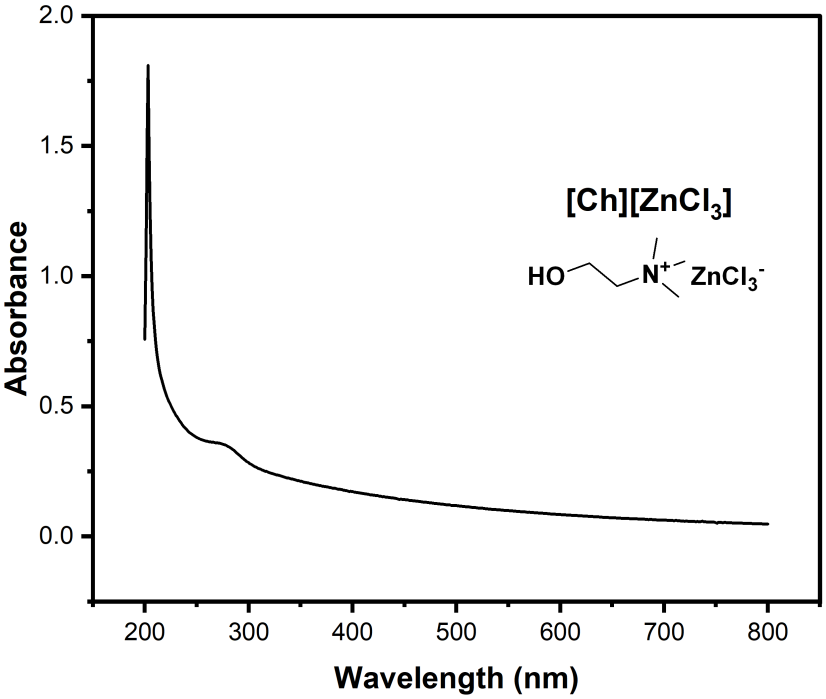


Fig. S20. UV-Vis spectrum of [Ch][$ZnCl_3$].

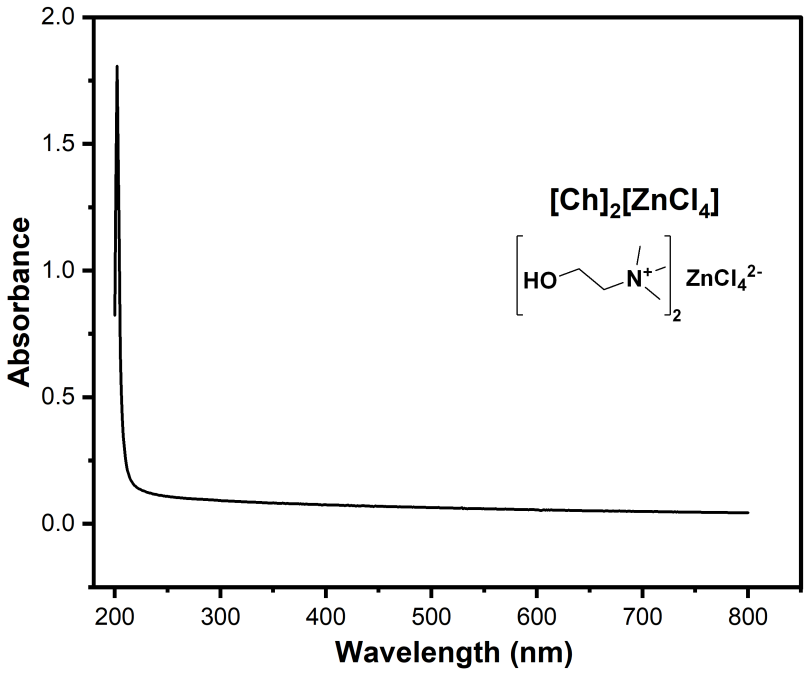


Fig. S21. UV-Vis spectrum of $[Ch]_2[ZnCl_4]$.

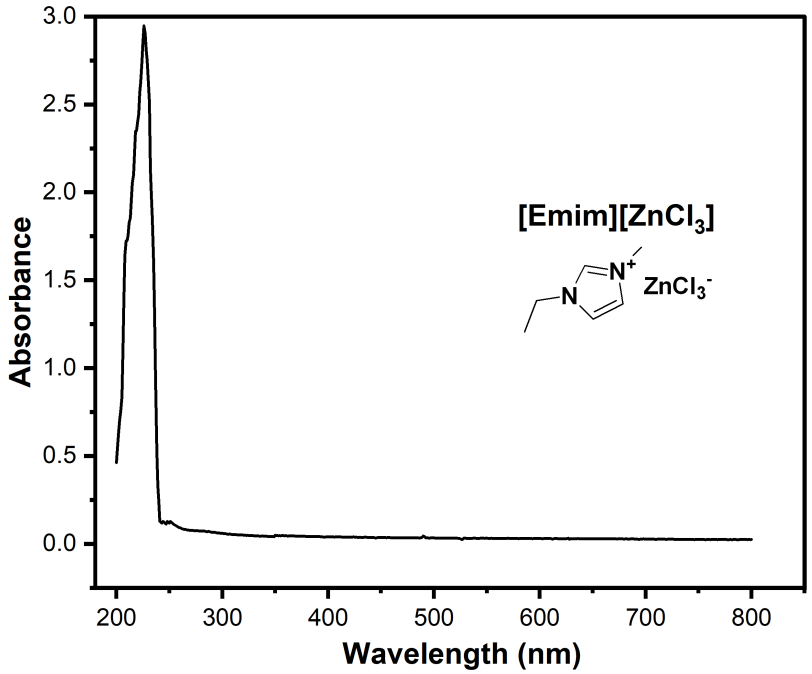


Fig. S22. UV-Vis spectrum of $[Emim][ZnCl_3]$.

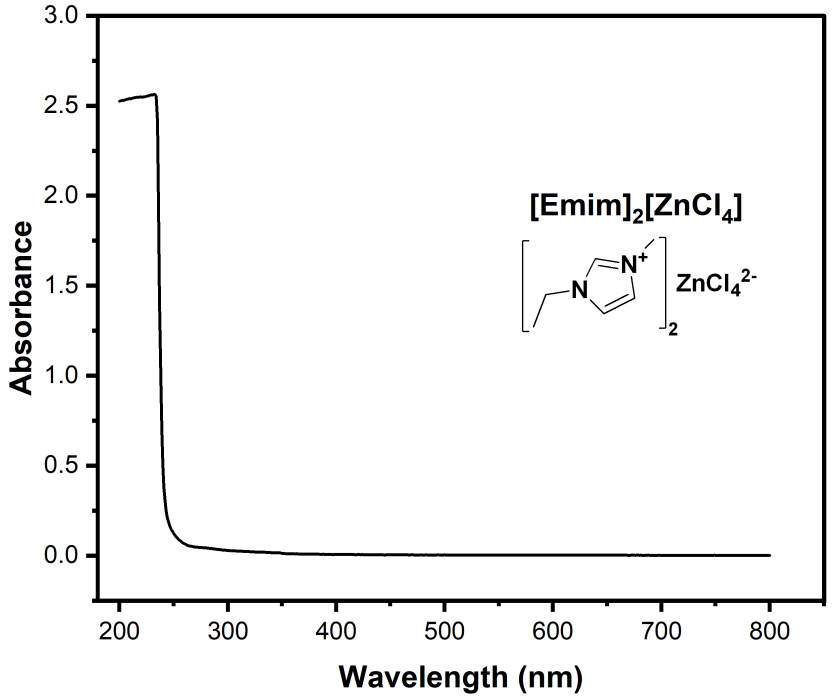


Fig. S23. UV-Vis spectrum of $[Emim]_2[ZnCl_4]$.

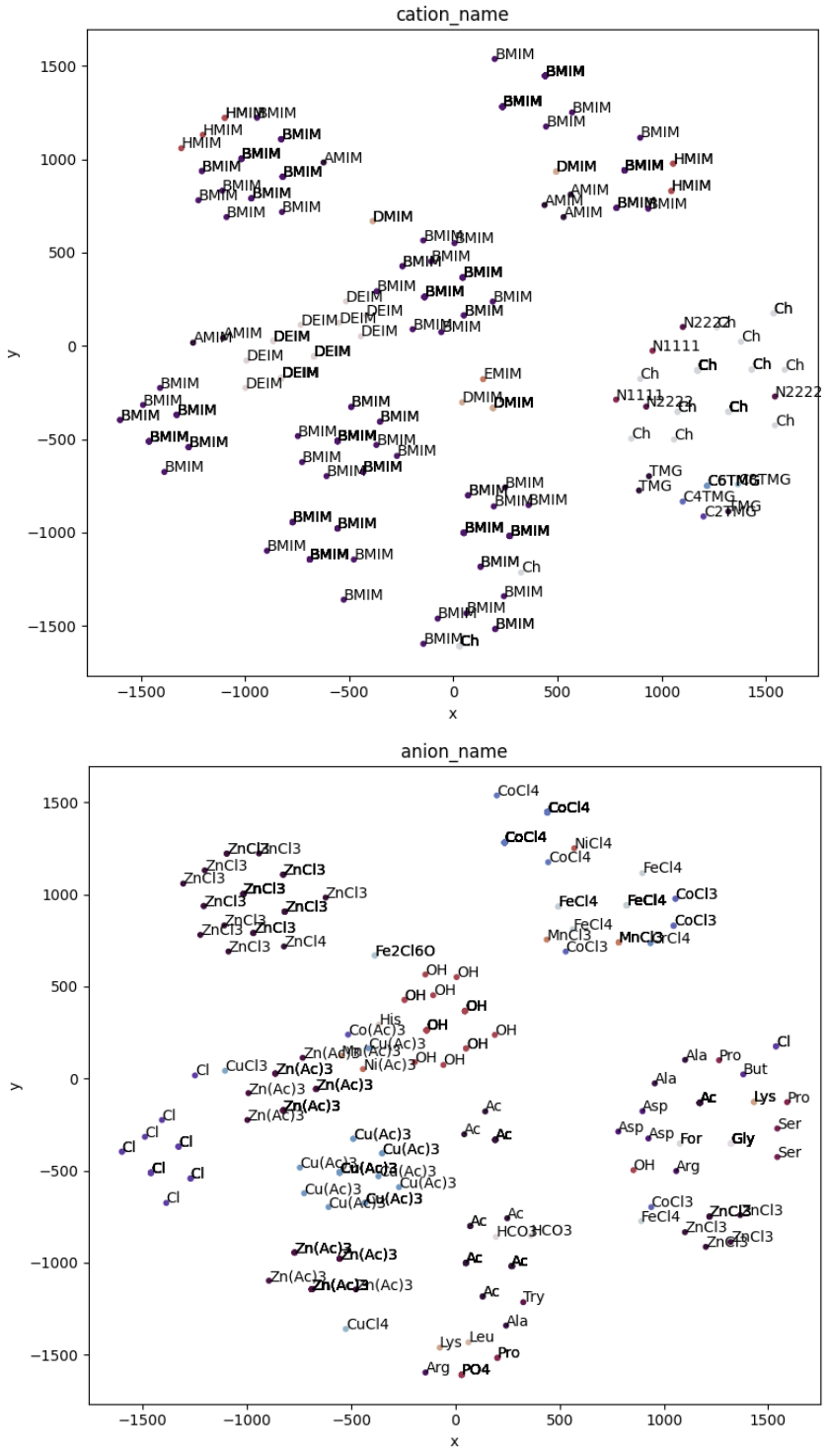


Fig. S24. t-SNE plot for GNN-featurized IL outputs.

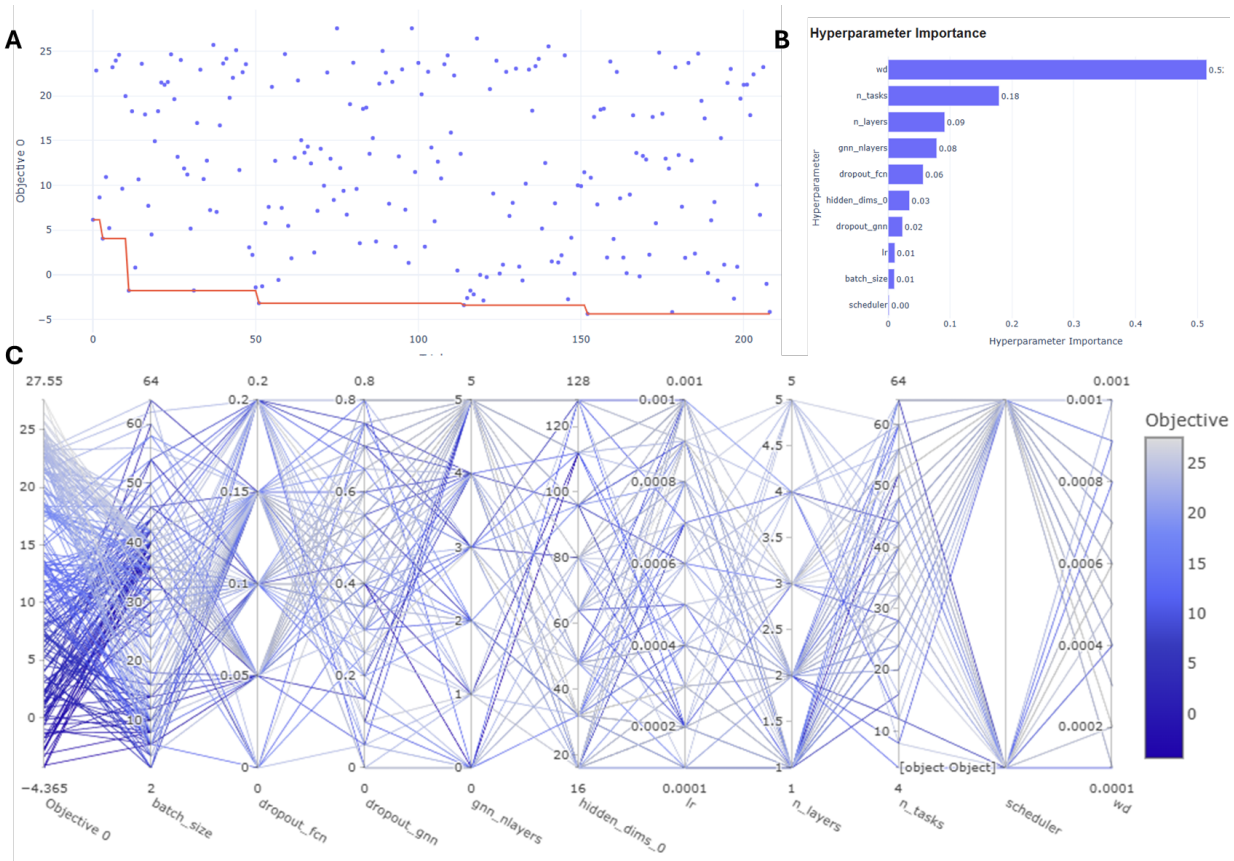


Fig. S25. Hyperparameter tuning via Optuna.

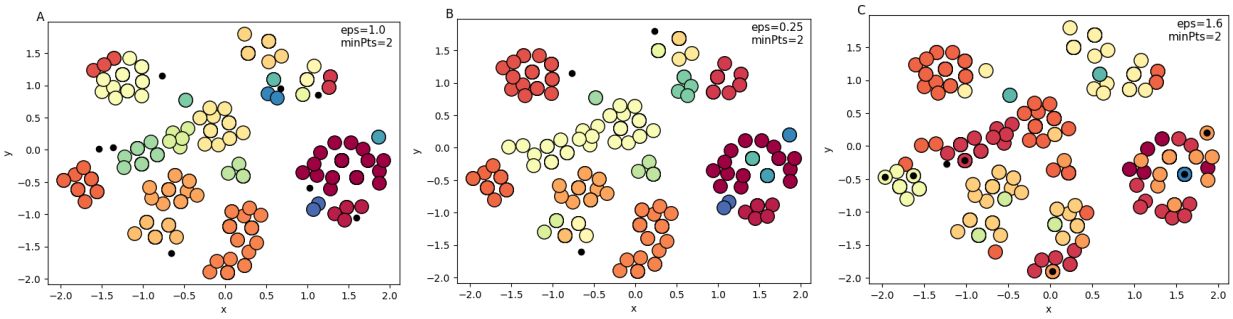


Fig. S26. (A) DBSCAN clustering results with raw GNN output features. (B) DBSCAN clustering results with t-SNE on GNN outputs. (C) DBSCAN clustering results with t-SNE on GNN outputs along with reaction conditions.

Data S1. (separate file)

The PET glycolysis data in the database is collected from paper in the literature[5–24]. The raw dataset consists of entries related to the IL-catalyzed PET glycolysis, structured across the following columns:

- **paper**: Citations or references for the source of the experimental data.
- **cation_name**: Name of the cation associated with the ionic liquid.
- **anion_name**: Name of the anion associated with the ionic liquid.
- **cation_smiles**: SMILES representation of the cation.
- **anion_smiles**: SMILES representation of the anion.
- **catalyst_amount**: Quantity of the IL in grams.
- **solvent**: Type of solvent, which is EG for all cases.
- **solvent_amount**: Quantity of the solvent in grams.
- **PET_amount**: Amount of PET in grams.
- **conversion**: The percentage of PET converted.
- **selectivity**: Percentage of BHET in all converted products.
- **temperature_c**: The reaction temperature in degrees Celsius.
- **reaction_time_min**: Duration of the reaction in minutes.
- **PET_source**: The origin or type of PET used (bottle, powder, or pellet).
- **PET_size_mm**: The size of PET particles in millimeters before the reaction.
- yield: The final yield of BHET in percentage. Numbers are present if the yield is reported in the paper. Otherwise, the entries will be calculated by using conversion multiplied by selectivity.

1 M. Ester, H.-P. Kriegel, J. Sander and X. Xu, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 1996, pp. 226–231.

2 J. Osborne and A. Overbay, *Pract. Assess. Res. Eval.*

3 Andrew Ng, *Machine Learning Yearning*, deeplearning.ai, 2018.

4 N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Journal of Machine Learning Research*, 2014, **15**, 1929–1958.

5 Q. Yue, L. Xiao, M. Zhang and X. Bai, *Polymers (Basel)*, 2013, **5**, 1258–1271.

6 A. M. Al-Sabagh, F. Z. Yehia, A. M. F. Eissa, M. E. Moustafa, Gh. Eshaq, A. M. Rabie and A. E. ElMetwally, *Polym Degrad Stab*, 2014, **110**, 364–377.

7 A. M. Al-Sabagh, F. Z. Yehia, A.-M. M. F. Eissa, M. E. Moustafa, G. Eshaq, A.-R. M. Rabie and A. E. ElMetwally, *Ind Eng Chem Res*, 2014, **53**, 18443–18451.

8 Y. Liu, X. Yao, H. Yao, Q. Zhou, J. Xin, X. Lu and S. Zhang, *Green Chemistry*, 2020, **22**, 3122–3131.

9 L. Liu, H. Yao, Q. Zhou, D. Yan, J. Xu and X. Lu, *ACS Engineering Au*, 2022, **2**, 350–359.

10 S. Marullo, C. Rizzo, N. T. Dintcheva and F. D'Anna, *ACS Sustain Chem Eng*, 2021, **9**, 15157–15165.

11 F. Scé, I. Cano, C. Martin, G. Beobide, Ó. Castillo and I. de Pedro, *New Journal of Chemistry*, 2019, **43**, 3476–3485.

12 C. Shuangjun, S. Weihe, C. Haidong, Z. Hao, Z. Zhenwei and F. Chaonan, *J Therm Anal Calorim*, 2021, **143**, 3489–3497.
13 C. V. G. Silva, E. A. da Silva Filho, F. Uliana, L. F. R. de Jesus, C. V. P. de Melo, R. C. Barthus, J. G. A. Rodrigues and G. Vanini, *Polímeros*, 2018, **28**, 450–459.
14 J. Sun, D. Liu, R. P. Young, A. G. Cruz, N. G. Isern, T. Schuerg, J. R. Cort, B. A. Simmons and S. Singh, *ChemSusChem*, 2018, **11**, 781–792.
15 H. Wang, Y. Liu, Z. Li, X. Zhang, S. Zhang and Y. Zhang, *Eur Polym J*, 2009, **45**, 1535–1544.
16 H. Wang, R. Yan, Z. Li, X. Zhang and S. Zhang, *Catal Commun*, 2010, **11**, 763–767.
17 Q. Wang, X. Yao, S. Tang, X. Lu, X. Zhang and S. Zhang, *Green Chemistry*, 2012, **14**, 2559–2566.
18 Q. Wang, X. Lu, X. Zhou, M. Zhu, H. He and X. Zhang, *J Appl Polym Sci*, 2013, **129**, 3574–3581.
19 Q. Wang, Y. Geng, X. Lu and S. Zhang, *ACS Sustain Chem Eng*, 2015, **3**, 340–348.
20 L. Wang, G. A. Nelson, J. Toland and J. D. Holbrey, *ACS Sustain Chem Eng*, 2020, **8**, 13362–13368.
21 T. Wang, X. Gong, C. Shen, G. Yu and X. Chen, *Polym Degrad Stab*, 2021, **190**, 109601.
22 H. Yao, X. Lu, L. Ji, X. Tan and S. Zhang, *Ind Eng Chem Res*, 2021, **60**, 4180–4188.
23 X. Zhou, X. Lu, Q. Wang, M. Zhu and Z. Li, *Pure and Applied Chemistry*, 2012, **84**, 789–801.
24 Q. F. Yue, C. X. Wang, L. N. Zhang, Y. Ni and Y. X. Jin, *Polym Degrad Stab*, 2011, **96**, 399–403.