Supplementary Information (SI) for Digital Discovery.

Supporting Information: Deep learning-enabled discovery of low-melting-point ionic liquids

Gaopeng Ren,[†] Austin Mroz,[†,‡] Frederik D. Philippi,[†] Tom Welton,[†] and Kim E. Jelfs[*,†]

[†]*Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, White City Campus, Wood Lane, London, W12 0BZ, U.K.*

[‡]*I-X Centre for AI in Science, Imperial College London, White City Campus, Wood Lane, London, W12 0BZ, U.K.*

E-mail: k.jelfs@imperial.ac.uk

Contents

S1 Melting point distribution

The distribution of melting points in the IL melting point database[S1] is shown in Figure S1. The melting points range from 177 K to 632 K, with 60% of the ILs having melting points below 373 K. The mean melting point is 361 K.

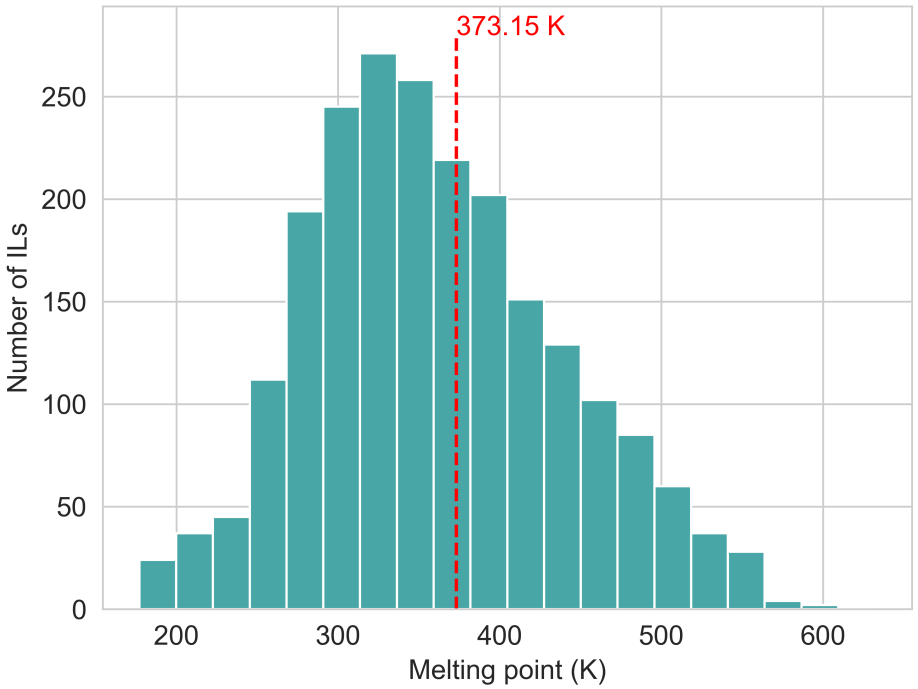

Figure S1: Melting point histogram of ILs. The red dashed line indicates the room temperature threshold of 373 K.

S2 Melting point classification model

The melting point is closely related to the Gibbs free energy of fusion (ΔG_{fus}). When $\Delta_{\text{fus}}G^T < 0$, it indicates that the melting point is lower than the given temperature T. In this work, we focus on ILs with melting points below 373 K, corresponding to $\Delta_{\text{fus}}G^{T=373K} < 0$. Direct calculation of $\Delta_{\text{fus}}G^{T=373K}$ is challenging, but fortunately, this value can be approximated using empirical equations and simple linear models.

S2.1 The estimation of lattice enthalpy

There is a Gibbs energy equation to describe the relationship between the Gibbs free energy of fusion the temperature (T), and the enthalpy ($\Delta_{\text{fus}}H$) and entropy of fusion ($\Delta_{\text{fus}}S$),

$$\Delta_{\text{fus}}G = \Delta_{\text{fus}}H - T\Delta_{\text{fus}}S. \tag{1}$$

Here, $\Delta G_{\text{fus}} = 0$ at the melting point. To calculate the Gibbs free energy of fusion, Krossing et al.[S2] proposed using the Born-Fajans-Haber cycle to correlate the energy with lattice and solvation Gibbs energies,

$$\Delta_{\text{fus}}G^T = \Delta_{\text{latt}}G^T - \Delta_{\text{solv}}G^T, \tag{2}$$

where the lattice Gibbs free energy can be calculated by:

$$\Delta_{\text{latt}}G^T = \Delta_{\text{latt}}H^T - T\Delta_{\text{latt}}S, \tag{3}$$

where the lattice enthalpy, $\Delta_{\text{latt}}H^T$, is calculated as:

$$\Delta_{\text{latt}}H^T = U_{\text{POT}} + 2RT, \tag{4}$$

where U_{POT} represents the lattice potential energy and can be estimated using the method proposed by Jenkins et al.[S3]:

$$U_{\mathrm{POT}} = 2(\frac{\alpha}{\sqrt[3]{V_m}} + \beta). \tag{5}$$

The molecular volume can be estimated using the atom-contribution method described by Hofmann[S4]:

$$V_m = \sum_{i}^{n} a_i v_i, \tag{6}$$

where n is the number of unique atom types in the molecule, a_i and v_i are the count of the atom type i and the contribution volume of the atom type i, respectively. Based on the findings of Krossing et al.[S2], this volume estimation method provides a reliable approximation of the lattice enthalpy $\Delta_{\mathrm{latt}}H^T$.

S2.2 The estimation of lattice entropy

The lattice entropy ($\Delta_{\mathrm{latt}}S$) can be determined using gas-phase entropy (S_{gas}) and solid-phase entropy (S_{solid}):

$$\Delta_{\mathrm{latt}}S = S_{\mathrm{gas}} - S_{\mathrm{solid}}, \tag{7}$$

where the solid-phase entropy can also be estimated based on the molecular volume:

$$S_{\mathrm{solid}} = kV_m + c, \tag{8}$$

where k and c are empirical parameters (1360 $\mathrm{J/(K \cdot mol \cdot nm^3)}$ and 15 $\mathrm{J/(K \cdot mol)}$, respectively).

For the gas entropy, Krossing et al.[S2] employed the TURBOMOLE software.[S5] However, since our goal is to apply the melting point classification model for high-throughput screening, we required a faster approach. Instead of following their method, we adopted an alternative strategy based on the work of Venkatraman and Roy[S6], which predicts gas-phase entropy using molecular surface curvatures. Since their code is based on JavaScript and only

applicable to websites, we rebuilt their model in Python. In the original work, van der Waals (vdW) surfaces were first generated using atom-centred Gaussian distributions:

$$G(x) = \sum_i \exp\left(-\frac{(|x - r_i| - a_i)}{\sigma}\right) \tag{9}$$

where r_i is the position of i^{th} atom center, and a_i is the corresponding vdW radius. In this work, we generated 3D molecular structures from SMILES using RDKit[S7] and obtained vdW radii from RDKit as well. The parameter σ controls surface smoothness, and we used the same value ($\sigma = 0.1$) as in the original paper. We then constructed a cube grid with a padding of 2.0 and a spacing of 0.2, computing $G(x)$ at each grid point. A vdW threshold of 0.1 was applied to define the vdW surfaces.[S6] An example of a generated vdW surface is shown in Fig. S2.

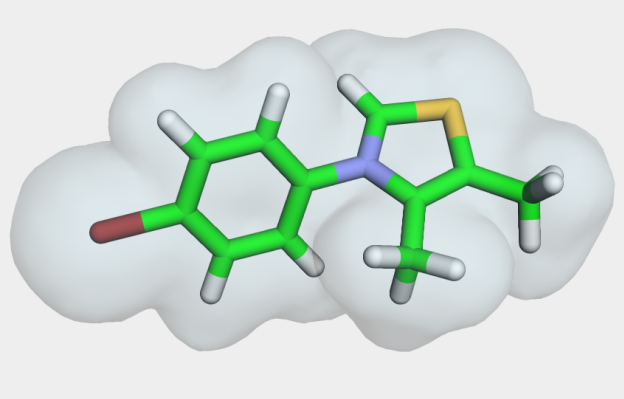

Figure S2: vdW surface of the cation Cc1sc[n+](-c2ccc(Br)cc2)c1C.

After obtaining vdW surfaces, the curvedness $\mathcal{C}$ for each triangle in the surface using the principal curvatures κ_1 and κ_2 was calculated by:

$$\mathcal{C} = \frac{2}{\pi} \ln\left[\frac{\kappa_1^2 + \kappa_2^2}{2}\right]. \tag{10}$$

Next, the Shannon shape entropy was computed as a molecular descriptor to estimate gas

entropy:

$$H^{\mathcal{C}}_{\text{shape}} = -\sum_{j=1}^{ntri} \left[\rho\left(\mathcal{C}_j\right) \log_2\left(\rho\left(\mathcal{C}_j\right)\right)\right] \cdot A_j, \tag{11}$$

where the A_j is the area of the j^{th} triangle and $\rho\left(\mathcal{C}_j\right)$ is the probability distribution of curvedness values $\mathcal{C}_j$. In this work, we constructed a curvedness histogram with 64 bins to estimate these probabilities.

Finally, the gas-phase entropy was estimated by fitting a linear function:

$$S_{\text{gasphase}} = m \cdot H^{\mathcal{C}}_{\text{shape}} + c. \tag{12}$$

We used the same database (the gas-phase entropy database) as the original paper[S6] to train the linear model. The dataset was split into a training set and a test set with a ratio of 80:20. The fitting results were $m = 8.703$ and $c = 70.93$. The plot of the predicted gas-phase entropy versus the experimental values is shown in Figure S3. The results indicate that this simple linear model achieves good performance on the gas-phase entropy prediction with an R^2 value of 0.87 and a root mean square error (RMSE) of 32.42 J/(mol · K).

Since ILs are mixtures of ions, generating their geometries is challenging. In this work, we computed the gas-phase entropy separately for cations and anions and then summed their contributions to estimate the total gas-phase entropy of ILs. To validate the gas-phase entropy prediction model for ions, we compared our predicted values with experimentally measured gas-phase entropies of several common IL cations and anions reported by Krossing et al.[S2]. The results, presented in Table S1, show that the predicted gas-phase entropies closely match the experimental values.

Table S1: Performance of the gas-phase entropy prediction model on ions of ILs (kJ/(mol · K)

ion	experimental[S2]	predicted
$[EMIM]^+$	0.3807	0.3992
$[BF4]^-$	0.2700	0.2822
$[TfO]^-$	0.3632	0.3952
$[BMIM]^+$	0.4719	0.4422
$[PF6]^-$	0.3498	0.3130

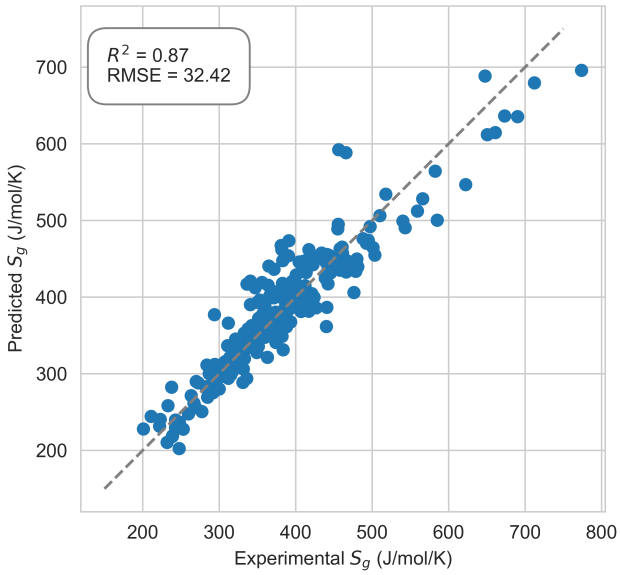

Figure S3: Predicted gas-phase entropy versus experimental gas-phase entropy. The dashed line represents perfect predictions. The predicted gas-phase entropies closely match the experimental values, with an R^2 of 0.87 and an RMSE of 32.42 J/(mol · K).

S2.3 Flexibility and SASA calculation

In this work, we adopted the approach proposed by Dannenfelser and Yalkowsky [S8] to estimate molecular flexibility. They defined the number of effective torsional angles, τ, as an indicator of molecular flexibility, calculated using the following formula:

$$\tau = \mathrm{SP3} + 0.5\mathrm{SP2} + 0.5\mathrm{RING} - 1 \tag{13}$$

where $\mathrm{SP3}, \mathrm{SP2}, \mathrm{RING}$ represent the number of SP3-hybridised, SP2-hybridised, and ring groups, respectively. In this work, we used SMARTS patterns to identify the SP2 and SP3 groups. Specifically, we included SP3-hybridised C, N, O, and S-related groups for SP3 groups, while SP2-hybridised C, N, and O-related groups for SP2 groups. To determine the number of rings, we first identified ring structures and their corresponding atom indices using RDKit. We then merged rings that shared common atoms to obtain the final ring count. The SASA values were also calculated using RDKit.

S3 MD for melting point estimation

In this work, we modified the MD workflow from Karu et al.[S9] to enable the automated estimation of IL melting points. As illustrated in Fig. S4, the workflow consists of three main steps: (1) preparing GROMACS input files, (2) running MD in different configurations, and (3) analysing diffusion coefficients to estimate the melting point. The code for this workflow is available at https://github.com/fate1997/MD4IL.

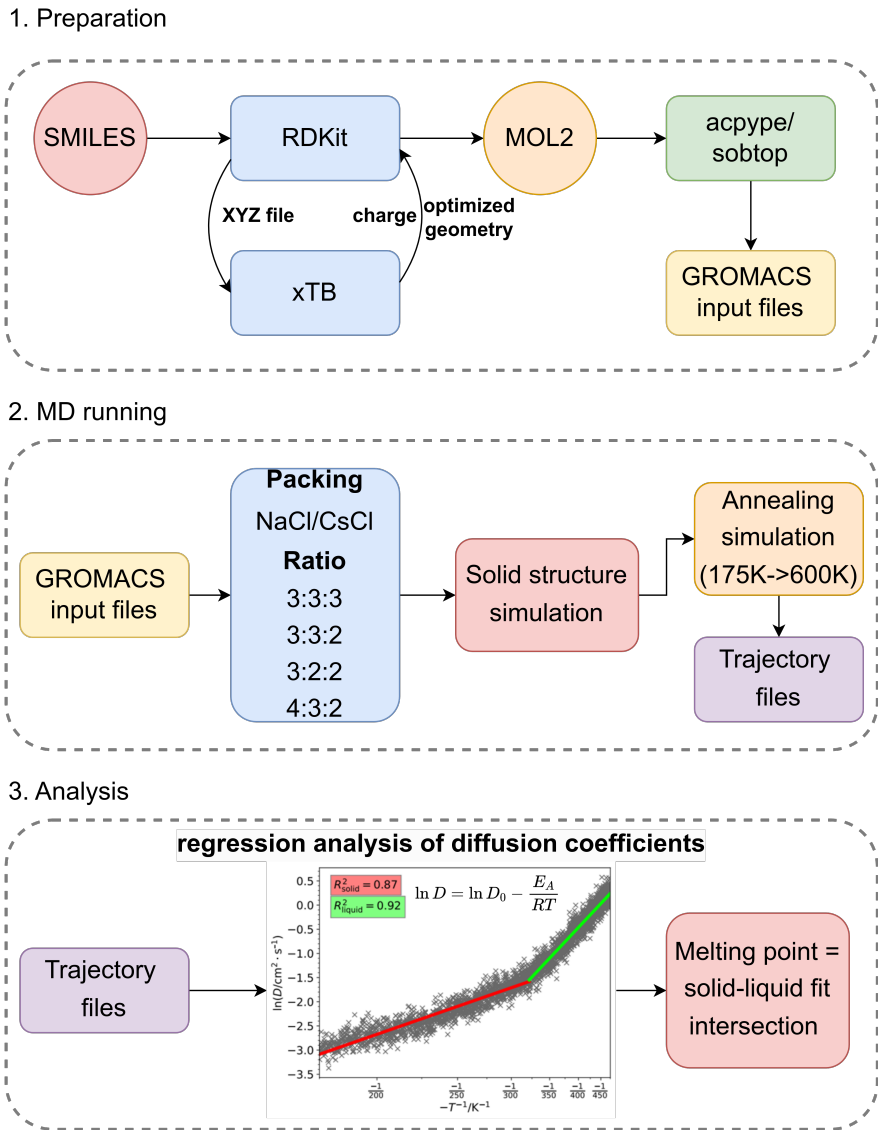

Figure S4: MD workflow for melting point estimation of ILs

S3.1 MD preparation

In the original study, GROMACS input files were prepared manually. To minimise user input and enable an automated workflow, we chose to start from SMILES representations of ILs. Given an IL SMILES string, we first split it into separate SMILES strings for the cation and anion. For each ion, we generated an initial 3D geometry using the ETKDGv3 distance geometry method in RDKit and saved it as an XYZ file. We then optimised the geometry and computed atomic charges using the xTB software.[S10] The results were saved in MOL2 format. Using the MOL2 file as input, we assigned force field parameters to each atom with the Sobtop software.[S11] The GAFF force field[S12] was used by default to generate GROMACS input files for each ion.

S3.2 MD simulation

The MD simulation process consists of two main steps: (1) simulating the solid-phase structure and (2) performing a production run to analyse the relationship between diffusion coefficients and temperature. The initial molecular systems were randomly generated using Packmol.[S13] These structures were first equilibrated through a short MD run using the steepest descent integrator. The solid-phase structures were then obtained via a 200 ps NVT simulation, during which the temperature was gradually reduced from 1000 K to 1 K over 100 ps and subsequently maintained at 1 K for another 100 ps. Since the exact arrangement of ions in the solid phase is unknown, Karu et al.[S9] proposed constructing different potential well lattices by combining CsCl-type and NaCl-type ionic lattices with lattice vector ratios of 3:2:2, 3:3:2, 3:3:3, and 4:3:2. After this initial structuring, six additional 500-ps-long NVT simulations were performed using the potential wells. In simulations 1, 3, and 5, only the anions were allowed to move, while in simulations 2, 4, and 6, only the cations were allowed to move.

Following the solid-phase simulations, the system was first equilibrated at 175 K for 2.5 ns using an NPT ensemble. The temperature was then gradually increased to 600 K over 42.5

ns. The diffusion coefficients were subsequently calculated to estimate the melting point. The cut-off for van der Waals and electrostatic interactions was 1.3 nm. The particle-mesh Ewald model was used to account for the long-range electrostatic interactions with 0.10 nm Fourier spacing.

S3.3 Melting point estimation

Melting points are estimated by identifying the intersection of the diffusion coefficients for the solid and liquid phases. The relationship between the diffusion coefficient (D) and temperature (T) can be described by the Arrhenius equation:

$$\ln D = \ln D_0 - \frac{E_A}{RT}, \tag{14}$$

where E_A denotes the diffusion activation energy and D_0 is the pre-exponent factor. Based on this equation, the relationship between the diffusion coefficient and the reciprocal of temperature is represented by two linear segments. The melting point is determined by finding the intersection point of these two lines, which can be identified by maximising the coefficient of determination (R^2) for the two fitted lines. For each IL, we conducted MD simulations in eight configurations with different ratios and packages. The final melting point was determined by selecting the highest estimated melting point.

S3.4 Performance

To assess the accuracy of the MD workflow in estimating the melting points of ILs, we randomly selected 20 ILs from the melting point database and computed their melting points using the workflow. As shown in Figure S5, the estimated values align well with experimental data, yielding an RMSE of 39.6 K. The performance is comparable, and in some cases better than, the performance reported by existing melting point prediction models.[S1,S14]

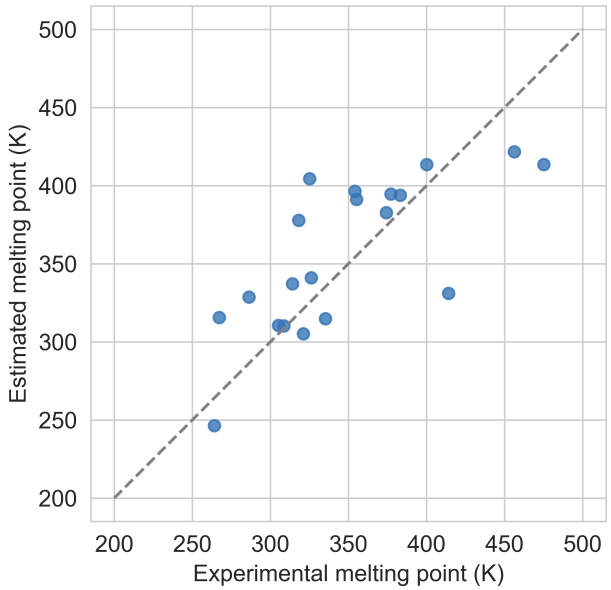

Figure S5: MD-estimated melting point (K) vs. experimental melting point (K)

S4 Generated ILs from link prediction

To assess whether the ILs generated through link prediction have relatively low melting points, we conducted MD simulations on 10 randomly selected ILs. As shown in Figure S6, most of the generated ILs exhibit low-melting points, with 6 out of 10 having melting points below 373 K. Since the collected IL database includes many ILs that are not liquids below 373 K, it is reasonable that some ILs generated through link prediction also have high melting points. The corresponding plots to determine melting points are shown in Figure S7 and Figure S8.

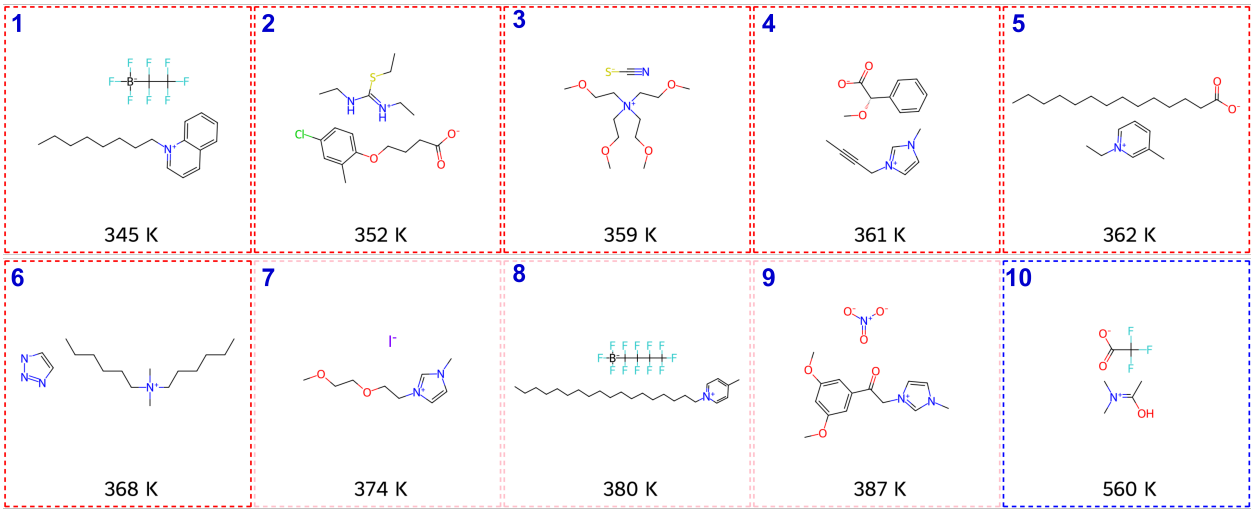

Figure S6: MD-estimated melting points (K) of expanded ILs from link prediction. The ILs are sorted by their MD-estimated melting points. ILs outlined in red have melting points below 373 K; those in pink are slightly above the threshold (within 15 K); and those in blue exhibit significantly higher melting points.

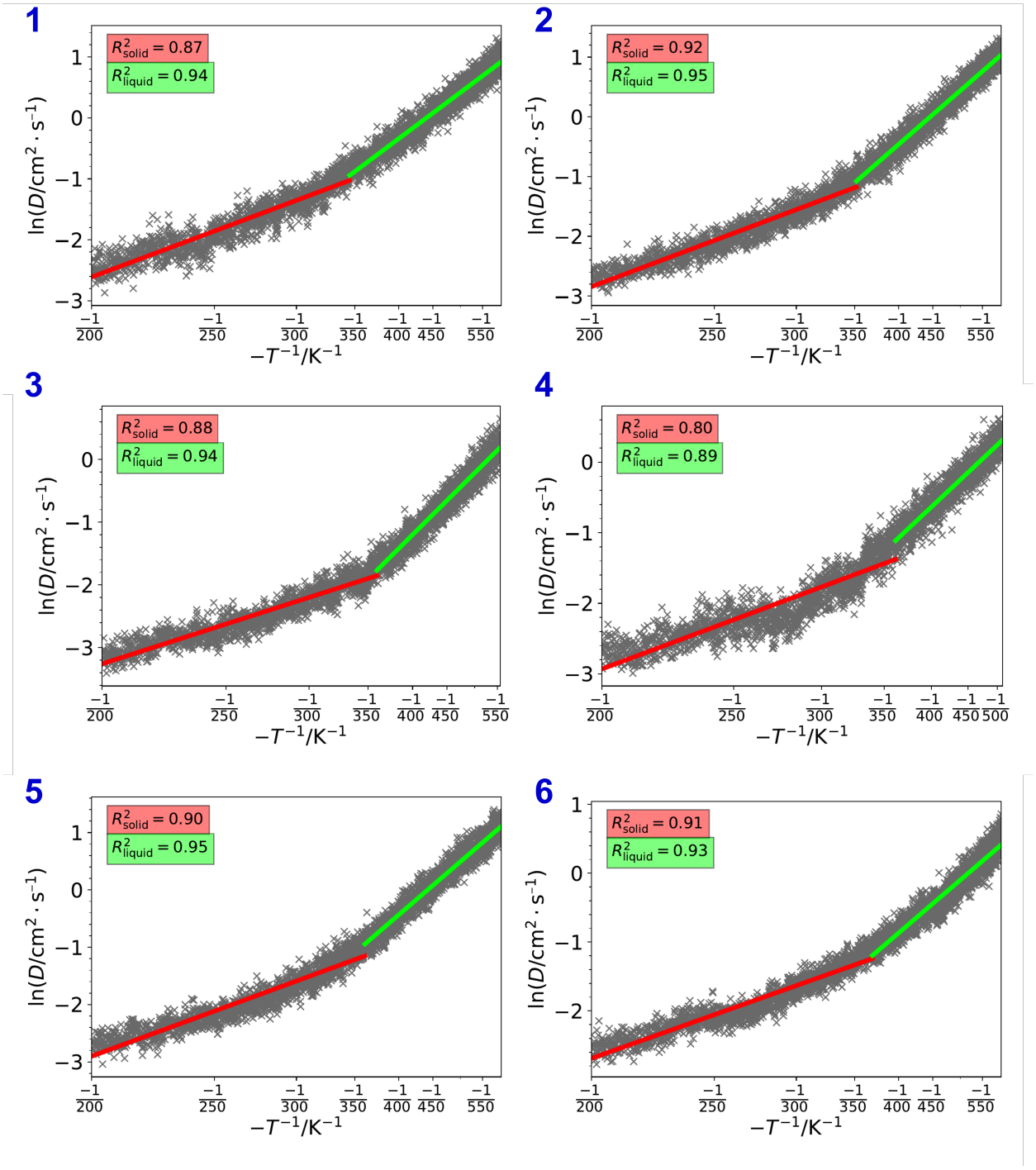

Figure S7: The diffusion coefficient (D) dependence on temperature (T) during annealing simulation for expanded ILs 1-6. The left-top number indicates the corresponding molecules in Figure S6.

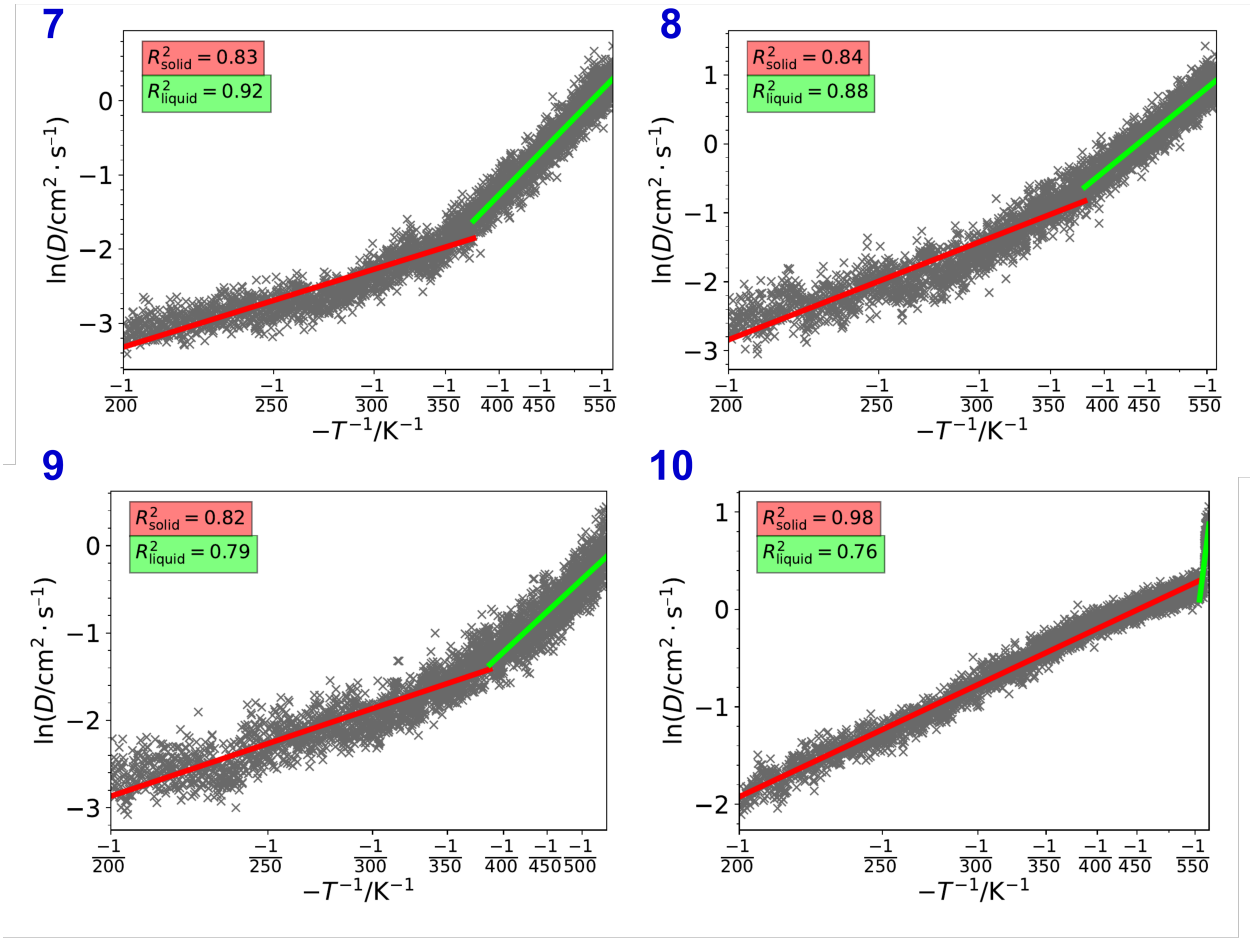

Figure S8: The diffusion coefficient (D) dependence on temperature (T) during annealing simulation for expanded ILs 7-10. The left-top number indicates the corresponding molecules in Figure S6.

S5 Generated ILs from VAE

To assess whether the ILs generated by the VAE have relatively low-melting points, we conducted MD simulations on 20 randomly selected ILs. Notably, these ILs had already passed the post-filtering module. As shown in Figure S9, most of the generated ILs exhibit low-melting points, with 18 out of 20 having melting points below 373 K. The corresponding plots to determine melting points are shown in Figure S10, S11, S12, and S13.

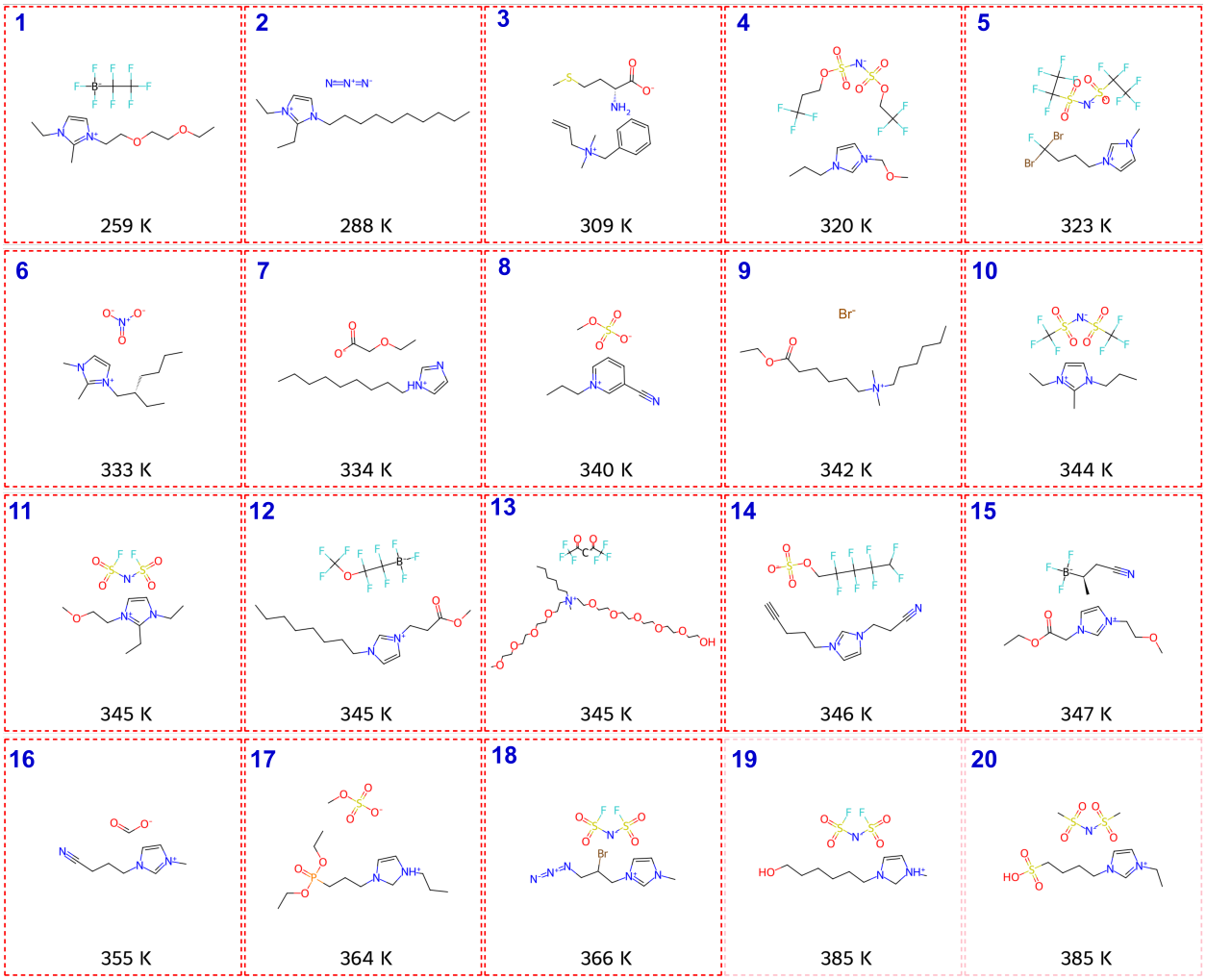

Figure S9: MD-estimated melting points (K) of generated ILs from VAE. The ILs are sorted by their MD-estimated melting points. ILs outlined in red have melting points below 373 K, and those in pink are slightly above the threshold (within 15 K).

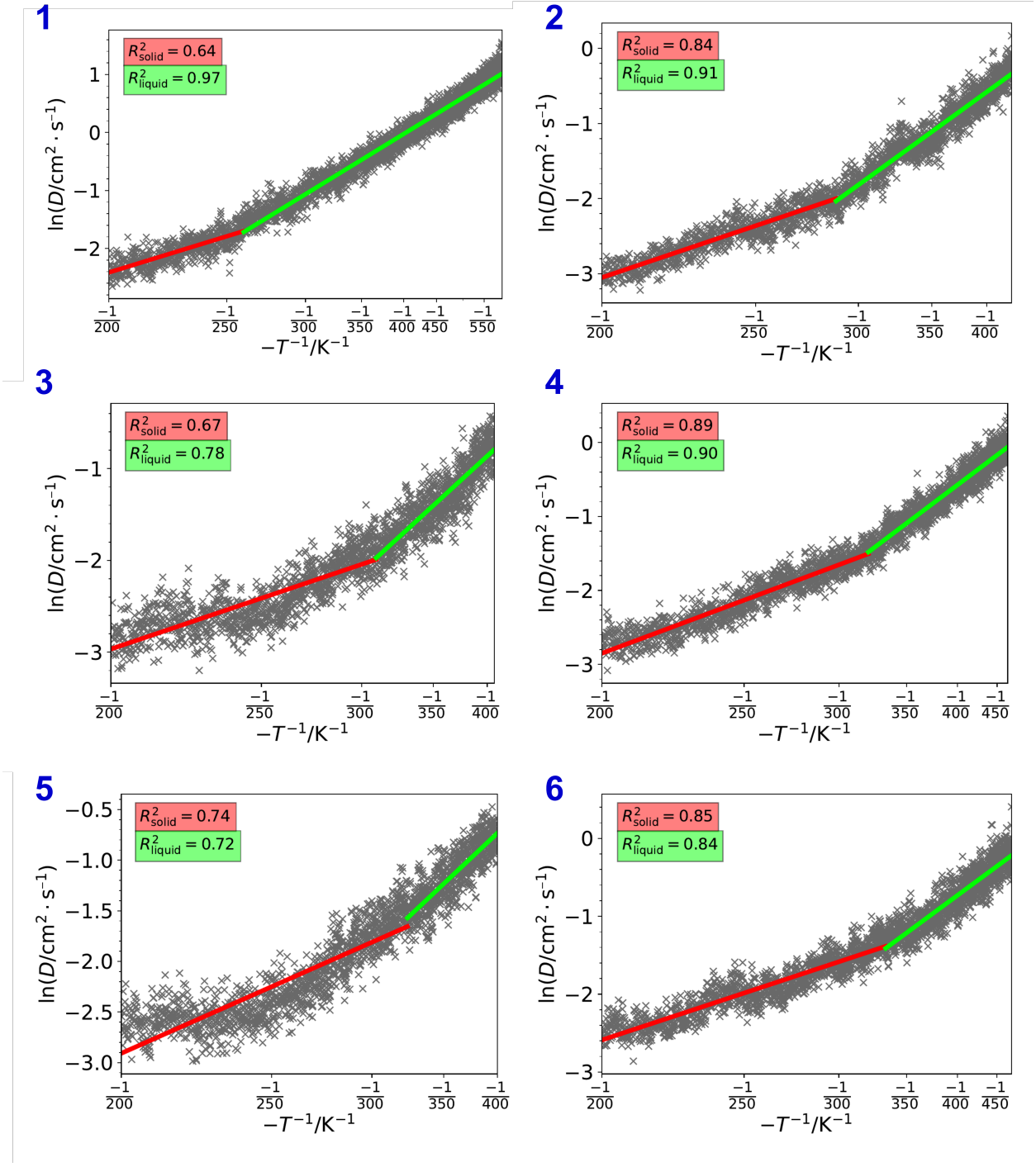

Figure S10: The diffusion coefficient (D) dependence on temperature (T) during annealing simulation for generated ILs 1-6. The left-top number indicates the corresponding molecules in Figure S9.

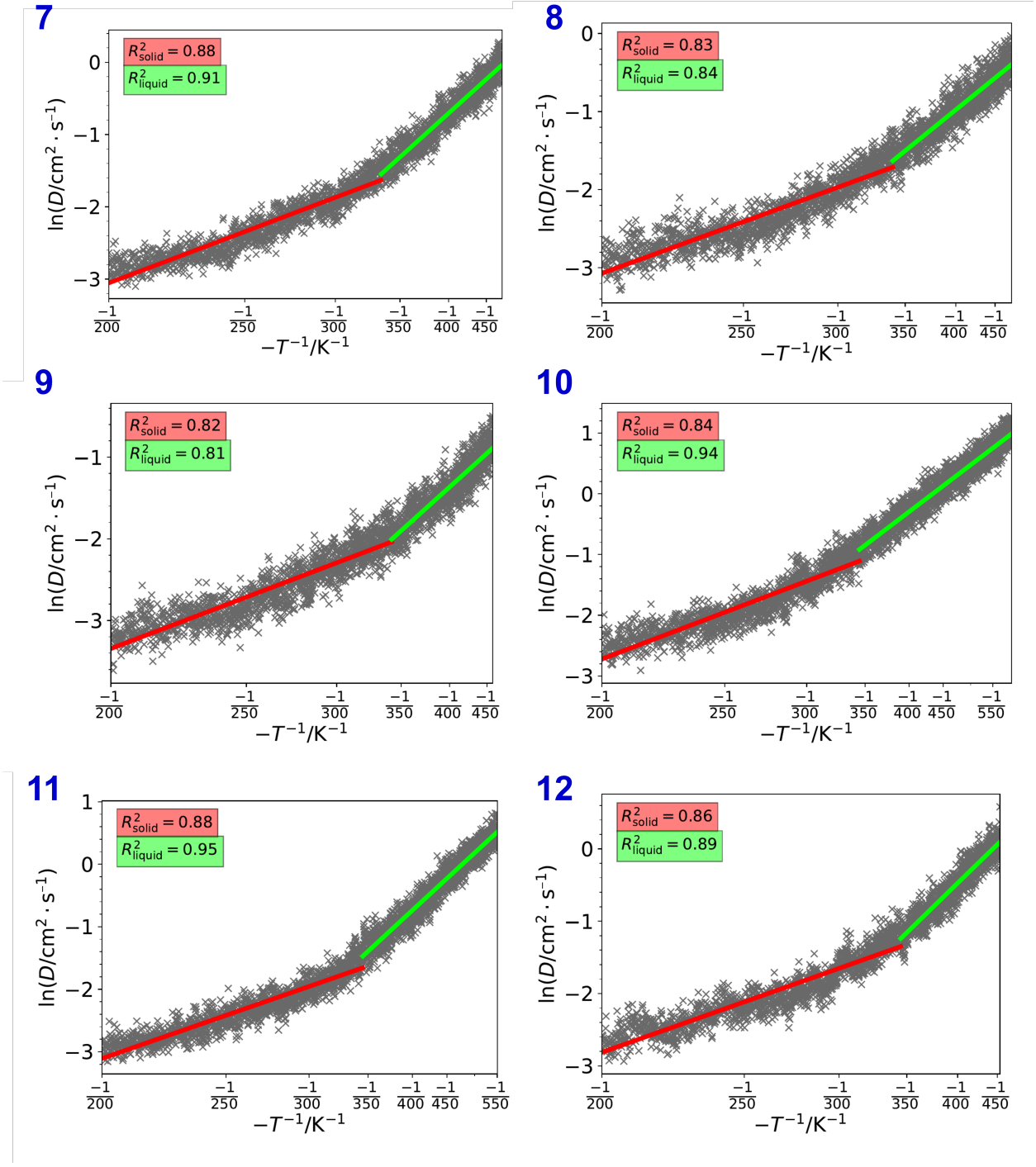

Figure S11: The diffusion coefficient (D) dependence on temperature (T) during annealing simulation for generated ILs 7-12. The left-top number indicates the corresponding molecules in Figure S9.

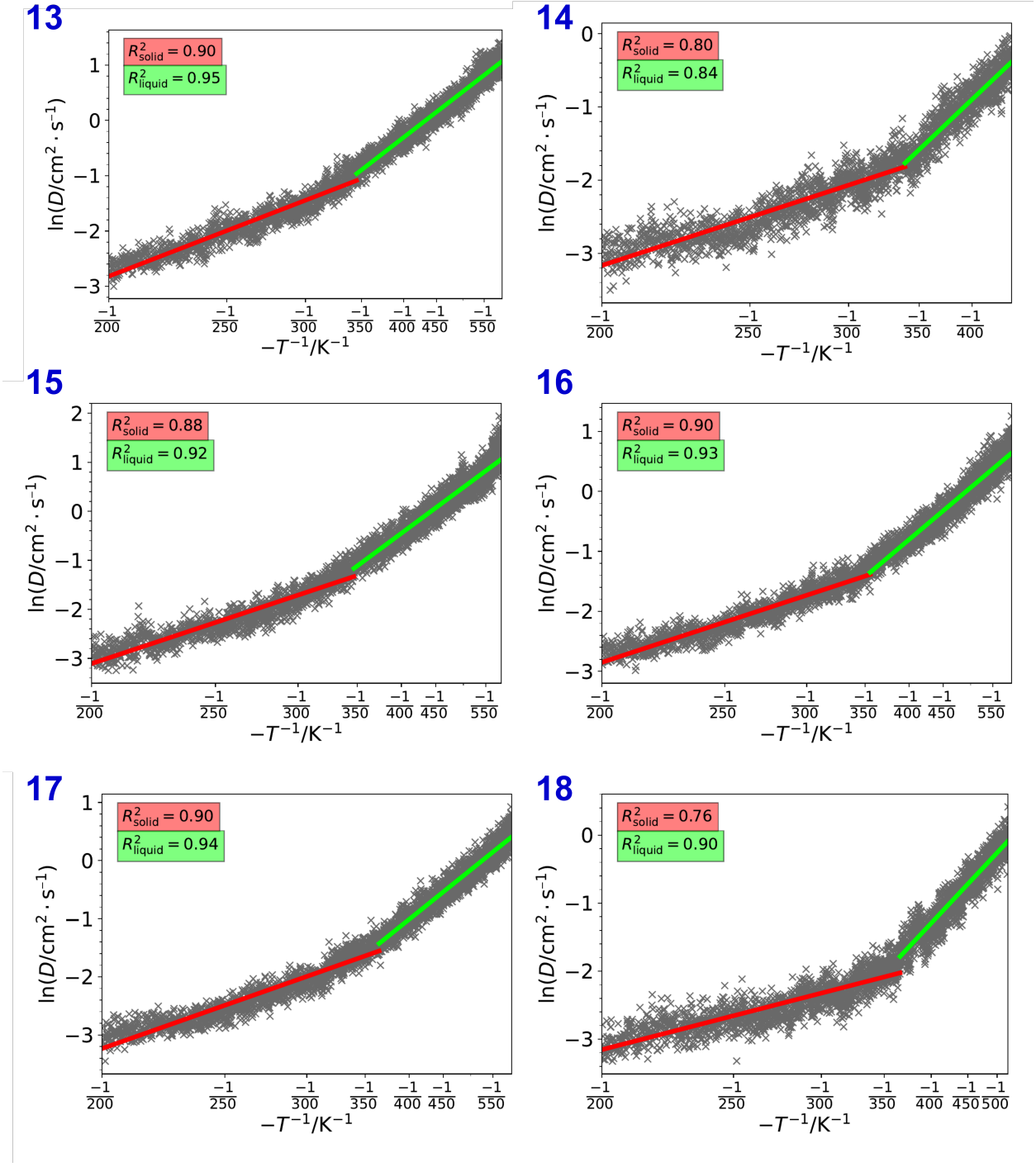

Figure S12: The diffusion coefficient (D) dependence on temperature (T) during annealing simulation for generated ILs 13-18. The left-top number indicates the corresponding molecules in Figure S9.

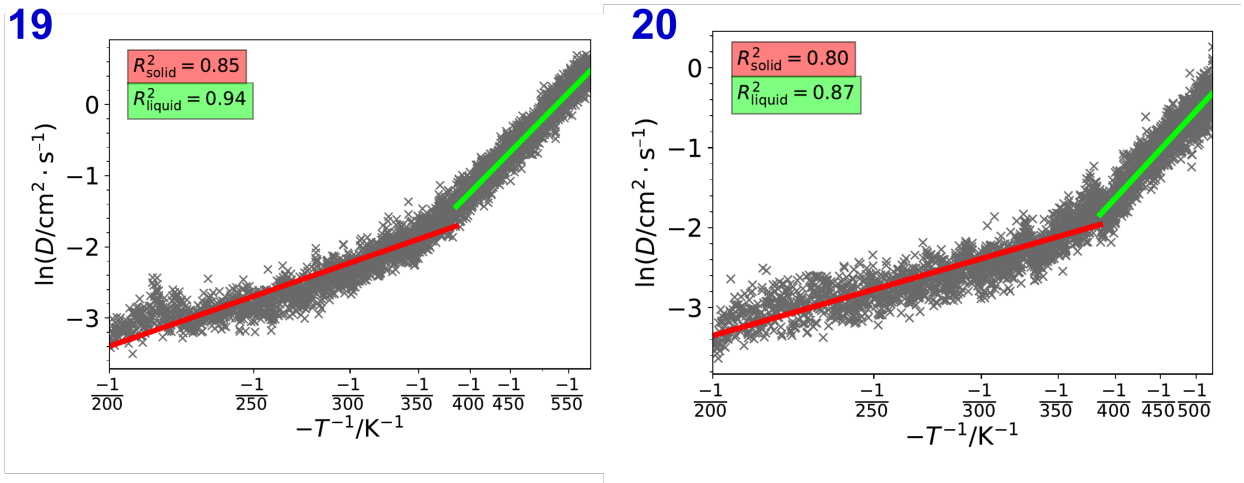

Figure S13: The diffusion coefficient (D) dependence on temperature (T) during annealing simulation for generated ILs 19-20. The left-top number indicates the corresponding molecules in Figure S9.

S6 Melting point distribution

To further validate the melting points of the generated ILs, we predicted the melting points for ILs from the collected dataset, link prediction, and VAE. We adopted the GNN model from Feng et al.[S14], which has demonstrated strong performance in IL melting point prediction. The resulting melting point distributions are shown in Figure S14. Most ILs from the collected database, link prediction, and VAE are classified as low-melting-point ILs. Moreover, the proportion of low-melting-point ILs in the ILs generated by VAE is higher than that in the collected database, indicating an increased tendency to generate ILs with lower melting points.

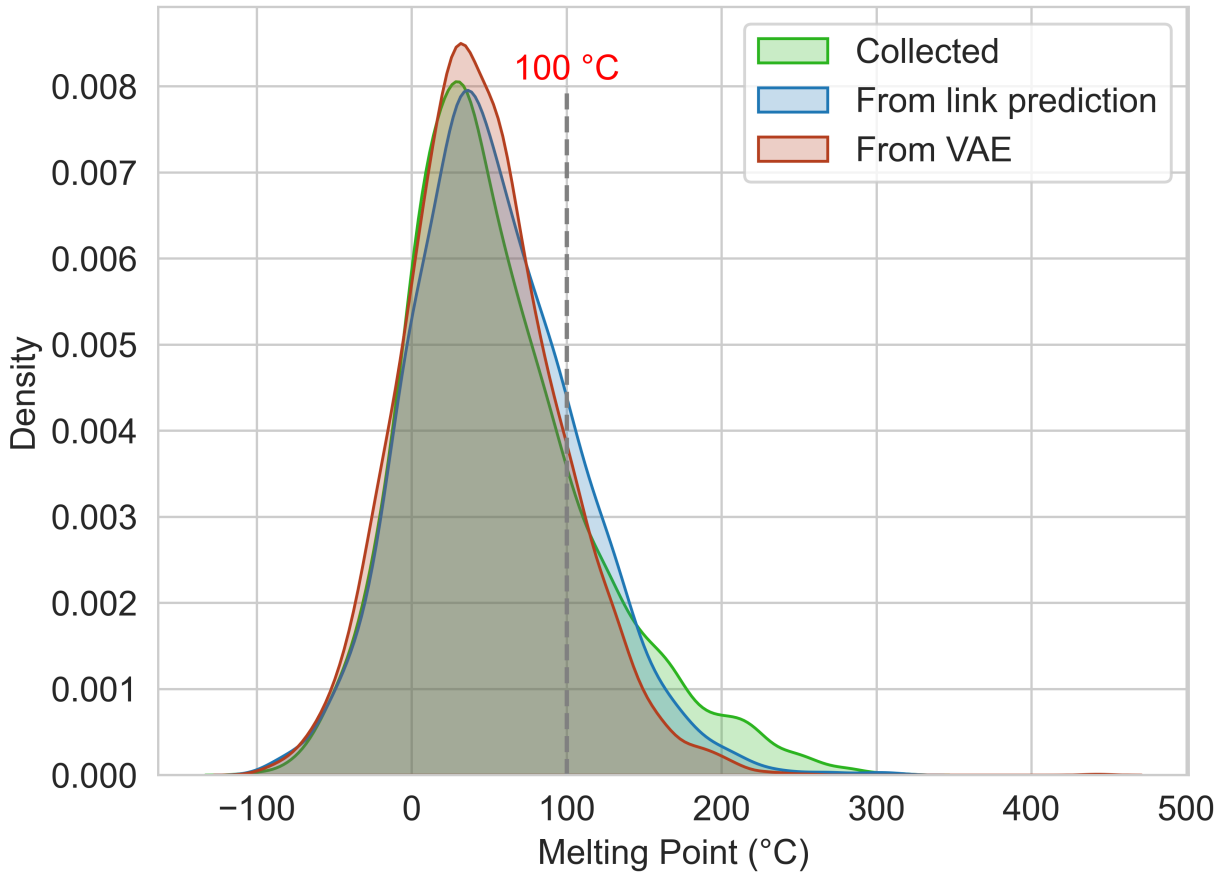

Figure S14: MD Estimated melting points (K) of generated ILs from VAE

References

(S1) Venkatraman, V.; Evjen, S.; Knuutila, H. K.; Fiksdahl, A.; Alsberg, B. K. Predicting ionic liquid melting points using machine learning. *J. Mol. Liq.* **2018**, *264*, 318–326.

(S2) Krossing, I.; Slattery, J. M.; Daguenet, C.; Dyson, P. J.; Oleinikova, A.; Weingärtner, H. Why Are Ionic Liquids Liquid? A Simple Explanation Based on Lattice and Solvation Energies. *J. Am. Chem. Soc.* **2006**, *128*, 13427–13434.

(S3) Jenkins, H. D. B.; Roobottom, H. K.; Passmore, J.; Glasser, L. Relationships among Ionic Lattice Energies, Molecular (Formula Unit) Volumes, and Thermochemical Radii. *Inorg. Chem.* **1999**, *38*, 3609–3620.

(S4) Hofmann, D. W. M. Fast estimation of crystal densities. *Acta Crystallogr. B* **2002**, *58*, 489–493.

(S5) Franzke, Y. J.; Holzer, C.; Andersen, J. H.; Begušić, T.; Bruder, F.; Coriani, S.; Della Sala, F.; Fabiano, E.; Fedotov, D. A.; Fürst, S.; Gillhuber, S.; Grotjahn, R.; Kaupp, M.; Kehry, M.; Krstić, M.; Mack, F.; Majumdar, S.; Nguyen, B. D.; Parker, S. M.; Pauly, F.; Pausch, A.; Perlt, E.; Phun, G. S.; Rajabi, A.; Rappoport, D.; Samal, B.; Schrader, T.; Sharma, M.; Tapavicza, E.; Treß, R. S.; Voora, V.; Wodyński, A.; Yu, J. M.; Zerulla, B.; Furche, F.; Hättig, C.; Sierka, M.; Tew, D. P.; Weigend, F. TURBOMOLE: Today and Tomorrow. *J. Chem. Theory Comput.* **2023**, *19*, 6859–6890.

(S6) Venkatraman, V.; Roy, A. Fast and Accurate Estimation of Gas-Phase Entropy from the Molecular Surface Curvature. *bioRxiv* **2021**, 2021.05.26.445640.

(S7) RDKit: Open-source cheminformatics. `http://www.rdkit.org`, [Online; accessed 11-April-2024].

(S8) Dannenfelser, R.-M.; Yalkowsky, S. H. Estimation of Entropy of Melting from Molecular Structure: A Non-Group Contribution Method. *Ind. Eng. Chem. Res.* **1996**, *35*, 1483–1486.

(S9) Karu, K.; Elhi, F.; Põhako-Esko, K.; Ivaništšev, V. Predicting melting points of biofriendly choline-based ionic liquids with molecular dynamics. *Appl. Sci.* **2019**, *9*, 5367.

(S10) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(S11) Sobtop. `http://sobereva.com/soft/Sobtop`, [Online; accessed 26-November-2023].

(S12) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(S13) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(S14) Feng, H.; Qin, L.; Zhang, B.; Zhou, J. Prediction and Interpretability of Melting Points of Ionic Liquids Using Graph Neural Networks. *ACS Omega* **2024**, *9*, 16016–16025.