npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-026-02093-y

Open electrolyte database generated via an automated molecular dynamics simulation framework

Kou Nakamura[1], Norio Takenaka[1], Masatoshi Hanai[2], Yuna Oikawa[3], Ryo Tamura[3,4], Koji Tsuda[3,4], Masanobu Nakayama[5], Junichiro Shiomi[6] & Atsuo Yamada[1,7] ✉

Modern battery technologies demand electrolytes that simultaneously deliver multiple functions tailored to diverse applications. Achieving such multi-objective optimization remains fundamentally challenging, as many key electrolyte properties are intrinsically in competition. Data-driven approaches provide a systematic route to navigating the vast compositional space of electrolyte systems; however, their effectiveness critically depends on the availability of comprehensive datasets that capture not only descriptors derived from isolated molecular structures and properties but also structural and physicochemical characteristics of electrolytes as integrated solutions. Here, we report an open electrolyte database comprising approximately 5600 electrolyte formulations, generated using a fully automated, high-throughput molecular dynamics simulation framework. The dataset spans diverse combinations of solvents, salts, and concentrations, and provides unified descriptions of electrolyte structures and physicochemical properties. To facilitate data exploration and utilization, we implement a web-based graphical user interface (https://oedb.jp) that enables interactive browsing and comparison of electrolyte compositions together with their associated descriptors, while also making the database accessible to LLM-based agents for data reference. This Open Electrolyte Database for Batteries (OEDB) establishes a foundation for data-driven electrolyte design grounded in structure–property relationships at the electrolyte level.

The transition toward a sustainable energy society critically relies on efficient and reliable electrochemical energy storage technologies. Rechargeable batteries underpin a wide range of applications, from electric mobility to grid-scale integration of intermittent renewable energy sources[1–3]. As these applications diversify, electrolytes are increasingly required to exhibit multiple physicochemical functionalities, including high ionic conductivity, wide electrochemical stability windows, thermal and chemical robustness, and stable interfacial compatibility with electrodes[4,5]. Simultaneous optimization of these properties remains a central challenge in electrolyte design, as many of the desired characteristics are mutually competing. Addressing this complexity necessitates data-driven approaches capable of capturing and exploiting the intricate relationships among electrolyte composition, structure, and physicochemical properties[6].

Existing databases for electrolyte-related research, such as the Electrolyte Genome Project[7], ZINC[8], PubChem[9], and Molecular Universe[10], primarily focus on molecular-level descriptors, including geometric structures, dipole moments, and HOMO/LUMO energies of isolated molecules. While such information is invaluable for characterizing individual molecular species, these databases do not provide information on electrolyte-level structures and physicochemical properties such as ionic conductivity and viscosity that emerge when molecules and ions collectively form an electrolyte solution. As a result, conventional data-driven electrolyte design often attempts to directly correlate the properties of isolated molecules with battery performance[11], effectively treating the collective properties of molecular assemblies as a black box and thereby limiting interpretability and rational design insight (Fig. 1a). A more robust design paradigm therefore requires electrolyte-level structure and physicochemical properties as

[1]Department of Chemical System Engineering, School of Engineering, The University of Tokyo, Tokyo, Japan. [2]Information Technology Center, The University of Tokyo, Tokyo, Japan. [3]Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan. [4]Center for Basic Research on Materials, National Institute for Materials Science, Ibaraki, Japan. [5]Department of Advanced Ceramics, Nagoya Institute of Technology, Nagoya, Japan. [6]Department of Mechanical Engineering, The University of Tokyo, Tokyo, Japan. [7]Sungkyunkwan University Institute of Energy Science & Technology (SIEST), Sungkyunkwan University, Suwon-si, Korea. ✉e-mail: yamada@chemsys.t.u-tokyo.ac.jp

intermediate variables, establishing a causal, stepwise link between molecular properties and battery performance (Fig. 1b).

However, acquiring electrolyte-level structural and physicochemical information in a systematic and reproducible manner across a broad compositional space remains a significant challenge. In general, direct experimental measurement of solution structure is difficult because liquid structures fluctuate dynamically over time, making large-scale acquisition impractical. Against this backdrop, molecular dynamics (MD) simulations provide a powerful methodology for evaluating both solution structure and the associated physicochemical properties[12,13]. Nevertheless, conventional MD workflows consist of multiple sequential steps, often requiring substantial manual intervention, thereby hindering scalability and reproducibility.

To address these challenges, we developed a fully automated MD framework that integrates fragmented simulation steps into a single, robust workflow and is tightly coupled with a large-scale computational environment equipped with high-capacity file storage. Using this framework, electrolyte-level structural and physicochemical data are generated under unified conditions in a high-throughput manner for a broad range of commercially relevant electrolyte compositions, and systematically curated into the Open Electrolyte Database for Batteries (OEDB) (Fig. 1c). Accordingly, in this study, we adopt classical MD simulations with widely applicable force fields to enable large-scale exploration of electrolyte composition space within feasible computational cost. To facilitate data exploration and utilization, we further develop a web-based graphical user interface (GUI) that enables interactive inspection of electrolyte compositions together with their associated structural and physicochemical descriptors, while also making the database accessible for reference by large language model (LLM)-based agents.

Results

Development of an electrolyte database

Conventional MD workflows typically treat charge assignment, force field parametrization, initial structure generation, equilibration, and subsequent analysis as separate and loosely connected steps. This fragmented workflow often requires manual intervention and ad hoc handling of data formats between individual stages, which not only increases the computational overhead of each calculation but also hinders efficient large-scale parallelization required for high-throughput studies across many electrolyte compositions (Fig. 2, top). To overcome these limitations, we constructed a fully automated MD framework in which all necessary simulation and analysis programs are deployed within a single supercomputer system and systematically integrated. In this framework, programs with different input and output formats are automatically linked, enabling seamless data transfer throughout the entire workflow (Fig. 2, bottom). As a result, thousands of electrolyte compositions can be processed end-to-end, from compositional input to the evaluation of solution structures and physicochemical properties, without manual intervention.

Operating an automated MD framework at scale requires a computational infrastructure that supports both massively parallel high-throughput calculations and large-scale data storage. In this study, we employed mdx[14,15] as the supercomputer system and carried out approximately 16 million CPU-hours of MD simulations using the latest-generation CPUs, exceeding the computational resource used in the initial construction of the Materials Project (15 million CPU-hours)[16], one of the largest first-principles databases for inorganic crystalline materials. This large-scale environment, combined with the automated workflow, enabled high-throughput generation of electrolyte structural and physicochemical property data for 5616 compositions (see Methods section for calculation details), with a throughput more than two orders of magnitude higher than that of conventional manual MD workflows. Approximately 50 GB of trajectory data were generated per composition, resulting in a total data volume of about 300 TB. The resulting dataset spans three cations (Li^+, Na^+, and K^+), nine anions, 26 solvents, and salt concentrations ranging from 0.5 to 4.0 mol kg^{-1}, with benchmark comparisons for representative systems confirming consistency with experimental values for density, ionic conductivity, and viscosity (Fig. S1).

However, classical force fields have inherent accuracy limitations (see Section S1 and Fig. S2 for a detailed discussion). Accuracy will be further enhanced in future work by adopting higher-fidelity simulation approaches tailored to the target electrolyte systems, using polarizable[17], machine-learning-based[18], and first-principles MD[19], which will enable more reliable evaluation of ion transport properties, including transference numbers[20,21]. Building on this strategy, the database will be frequently updated to incorporate both higher-quality data and an expanded range of electrolyte systems, including mixed-solvent and more complex formulations.

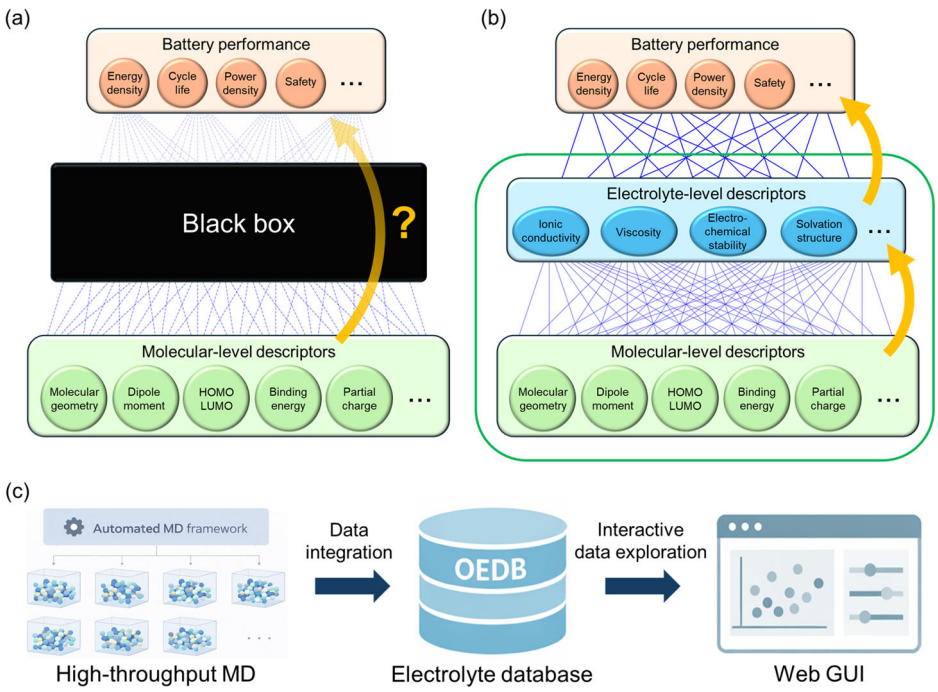

Fig. 1 | A data-driven hierarchical framework for electrolyte design. a Schematic illustration of a molecular-descriptor-based electrolyte design framework, in which molecular properties are directly linked to battery performance without explicit consideration of electrolyte-level structure or physicochemical properties. **b** Schematic illustration of a hierarchical design framework that introduces electrolyte-level structural and physicochemical properties as intermediate variables connecting molecular properties and battery performance. **c** Overview of the workflow employed in this study, including high-throughput molecular dynamics simulations, construction of the Open Electrolyte Database for Batteries (OEDB), interactive data exploration using a web-based graphical user interface, and use of the database by LLM-based agents.

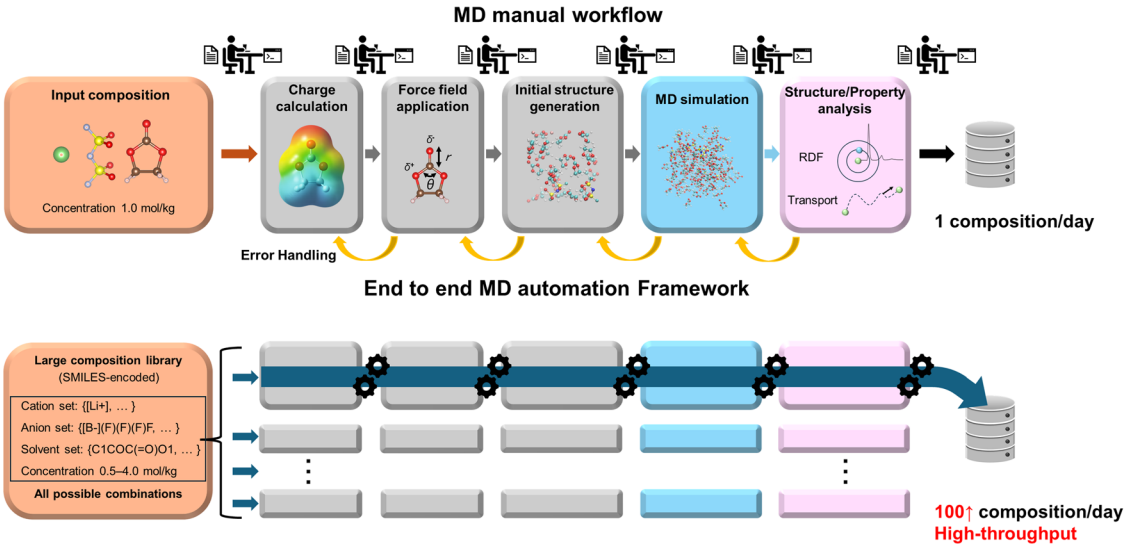

Fig. 2 | From manual molecular dynamics to scalable electrolyte data infrastructure. Top: Conventional manual MD workflow, in which an input electrolyte composition is processed sequentially through charge assignment, force field parametrization, initial structure generation, MD simulation, and subsequent structural and physicochemical property analysis, followed by database registration. Owing to the sequential and manual nature of the workflow, the achievable throughput is limited to approximately one composition per day. **Bottom:** End-to-end automated MD framework developed in this study, where all simulation and analysis steps are integrated and executed under Python-based control and a job scheduling system, enabling parallel processing of multiple electrolyte compositions. The workflow includes automatic analysis and direct registration of results into the database, achieving a throughput exceeding 100 compositions per day.

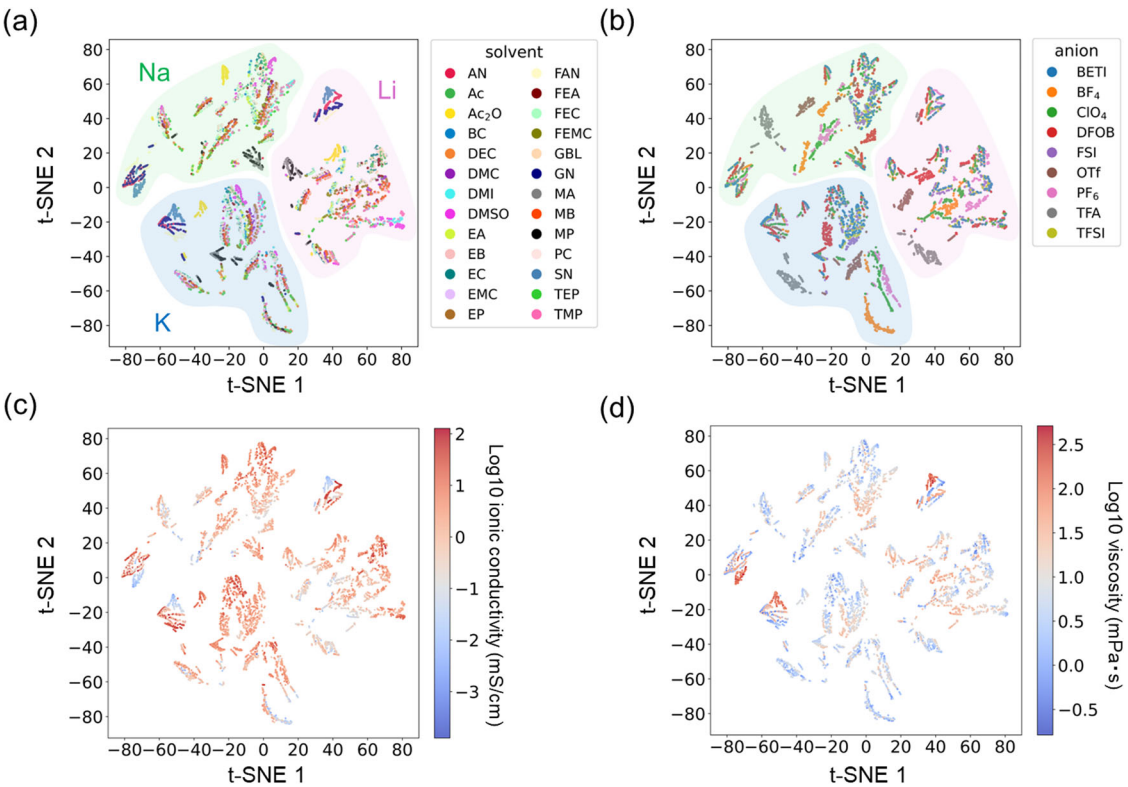

Fig. 3 | Electrolyte structure map and associated property distributions. a Two-dimensional t-distributed stochastic neighbor embedding (t-SNE) projection of MD–derived solution-structure descriptors for all electrolyte compositions, where each point represents a single composition and is colored according to the solvent category. **b** The same t-SNE projection as in (**a**), with points colored according to the anion category. **c** The same t-SNE projection, colored by the logarithm of ionic conductivity, log σ (mS cm^{-1}), showing the spatial distribution of ionic conductivity across the electrolyte structure map. **d** The same t-SNE projection, colored by the logarithm of viscosity, log η. (mPa·s), showing the corresponding distribution of viscosity across the structure map.

Electrolyte structure map

To systematically visualize the large number of electrolyte compositions contained in the database, all electrolytes were embedded into a two-dimensional structural space using t-distributed stochastic neighbor embedding (t-SNE) (Fig. 3; see Methods section for details)[22]. In this electrolyte structure map, compositions with similar solution structures are positioned in close proximity. Electrolytes separate primarily by cation species, with Li^+-, Na^+-, and K^+-based electrolytes

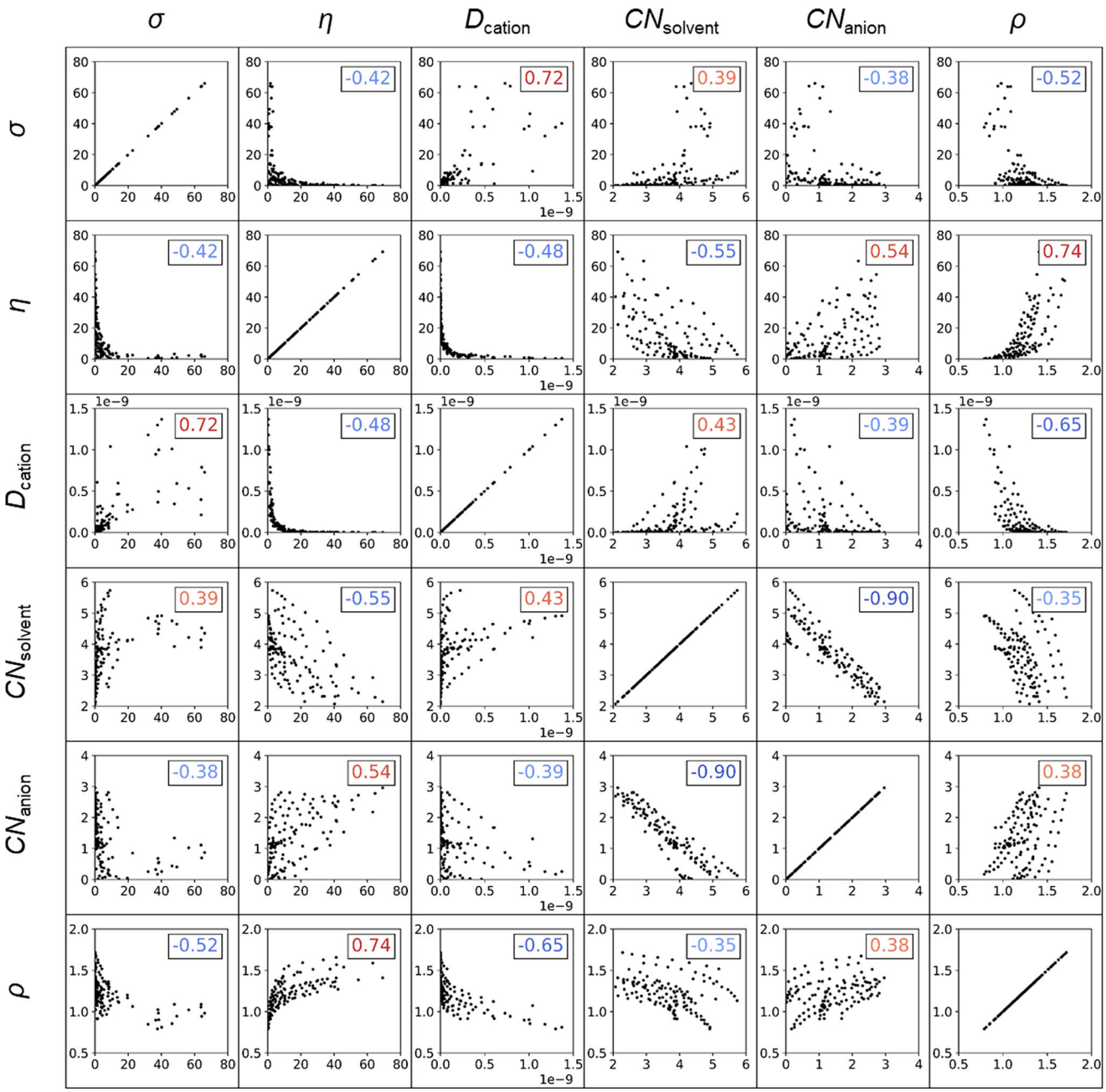

Fig. 4 | Physically consistent correlations in LiFSI-based electrolyte systems. Pairwise correlations among electrolyte physicochemical properties and coordination descriptors for LiFSI-based electrolyte systems extracted from the OEDB. The correlation matrix presents scatter plots for all pairwise combinations of ionic conductivity σ (mS cm^{-1}), viscosity η (mPa·s), Li$^+$ diffusion coefficient D_{Li} (m^2 s^{-1}), solvent and anion coordination numbers around Li$^+$ ($CN_{solvent}$ and CN_{anion}), and density ρ (g cm^{-3}). Pearson correlation coefficients are shown in each panel, with positive and negative correlations indicated in red and blue, respectively.

occupying three distinct regions (Fig. 3a, b), while further differentiation according to solvent and anion species appears within each region. To explore the relationship between electrolyte structure and physicochemical properties, ionic conductivity and viscosity were overlaid onto the same structural space (Fig. 3c, d). In these maps, the color distributions vary smoothly across the structural space rather than appearing randomly, suggesting a systematic correspondence between electrolyte structure and physicochemical properties. Notably, comparison of the two property maps shows that regions of higher ionic conductivity tend to coincide with lower viscosity, indicating a clear inverse relationship between these properties across the compositional space.

Correlations among electrolyte properties

To quantitatively assess the trends in physicochemical property distributions identified in the electrolyte structure space, we focus on a specific subset of compositions and examine pairwise relationships among key properties. In particular, lithium bis(fluorosulfonyl)imide (LiFSI)-based electrolyte systems were analyzed to evaluate correlations among ionic conductivity (σ), Li$^+$ diffusion coefficient (D_{Li}), viscosity (η), density (ρ), and the coordination numbers of solvent molecules and anions around Li$^+$ ($CN_{solvent}$ and CN_{anion}) (Fig. 4). A strong positive correlation is observed between σ and D_{Li}, reflecting the close correspondence between ion diffusion and ionic transport, in agreement with the Nernst–Einstein relation[23,24]. By contrast, the relationship between σ

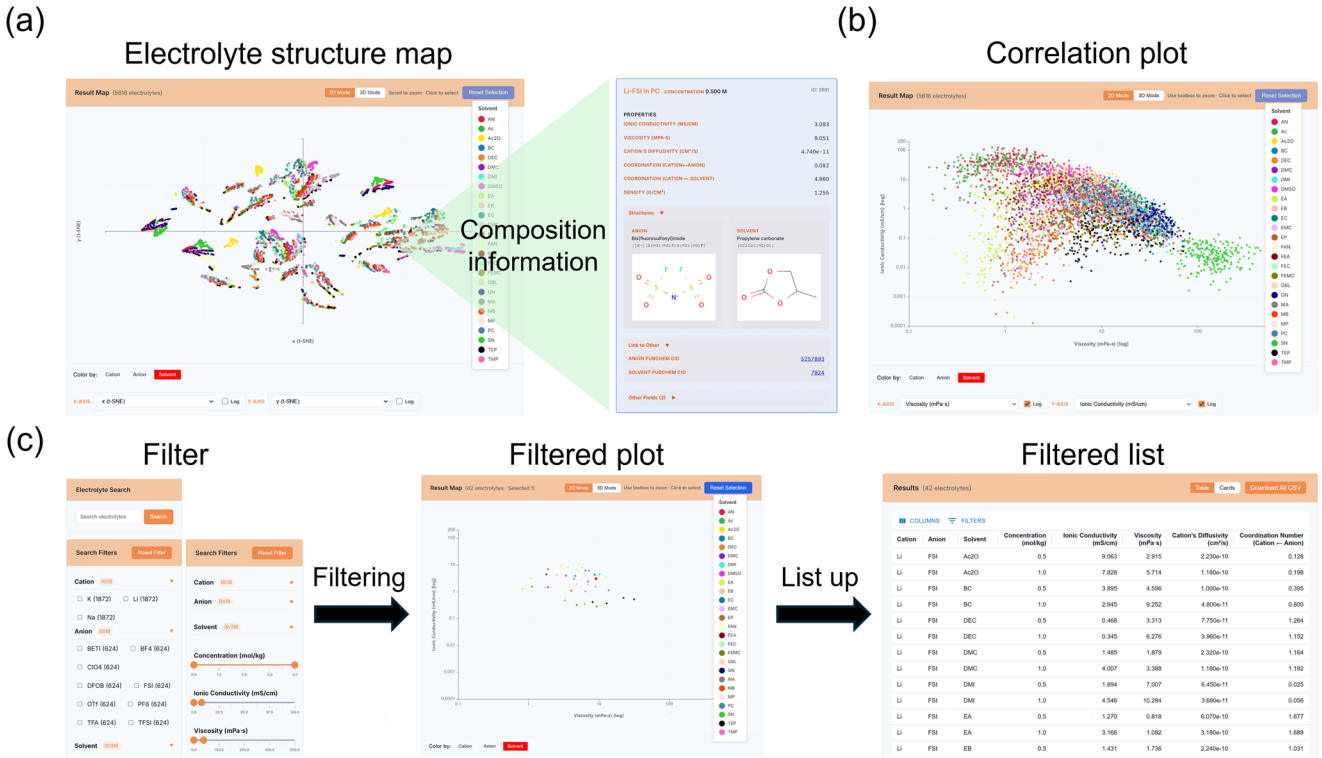

Fig. 5 | Graphical user interface for interactive exploration of electrolyte structure and properties. a Electrolyte structure map displayed as a t-SNE projection, where each point represents an electrolyte composition defined by cation, anion, solvent, and concentration. Selecting a point interactively displays the corresponding compositional information together with summaries of electrolyte-level physicochemical properties and structural descriptors. **b** Correlation plot showing electrolyte compositions in a two-property space (for example, ionic conductivity versus viscosity), constructed from the same composition-resolved dataset as in (**a**). **c** Filtering workflow of the interface, in which compositional categories and property ranges are specified via a side panel, resulting in synchronized updates of the structure map, correlation plots, and a corresponding list of electrolyte compositions that satisfy the selected conditions.

and η is consistent with the inverse dependence of diffusion on viscosity implied by the Stokes–Einstein relation[23,24]. Notably, comparison between low and high concentrations reveals that this relationship holds at 1.0 mol kg^{-1} but breaks down at 4.0 mol kg^{-1} due to stronger ion–ion correlations (Fig. S3). Taken together, these results show that physicochemical properties in electrolyte systems are linked in a physically consistent manner, connecting transport behavior with both molecular diffusion and local coordination environments. The OEDB provides a platform that enables systematic examination of such consistent relationships across a wide range of electrolyte compositions.

Development of a graphical user interface

To enable direct and integrated exploration of electrolyte-level compositional, structural, and physicochemical data, we developed a web-based graphical user interface (GUI), publicly accessible at https://oedb.jp. The interface integrates the electrolyte structure map, physicochemical property distributions, and compositional metadata within a single interactive environment (Fig. 5a, b). Individual electrolyte compositions can be selected directly from the structure map, while subsets of electrolytes can be dynamically extracted using filters based on structural similarity or target physicochemical properties (Fig. 5c). Selected compositions are visualized consistently across the structure map, correlation plots, and tabulated composition lists, enabling seamless comparison across multiple representations. By linking electrolyte-level composition, structure, and properties within a unified workflow, the GUI facilitates intuitive identification and comparison of electrolyte behaviors that are difficult to infer when considering individual molecular components in isolation. The interface is designed for future expansion, including integration of experimental datasets, incorporation of higher-accuracy simulation results, and extension to more complex electrolyte systems such as mixed-solvent and multi-salt formulations.

LLM-assisted electrolyte design grounded in OEDB

While the GUI provides systematic access to the OEDB for researchers familiar with electrolyte chemistry, translating database queries into actionable design insights often requires integration of multiple information sources, including database property records, knowledge from the literature, and application-specific constraints. To address this challenge, we implemented a large language model (LLM)[25,26]–assisted framework[27,28] that organizes electrolyte-related information through a natural-language-based interface (Fig. 6a). In this framework, user inputs specifying application context, target properties, and design constraints are processed by an LLM that integrates pretrained general knowledge with structured electrolyte data retrieved from the OEDB via the Model Context Protocol[29], together with publicly available web information. This integration enables candidate solvent–salt compositions and associated property trends to be organized within a unified language-based workflow.

Further specialization of the LLM for electrolyte exploration requires incorporation of domain-specific knowledge through additional training. As an initial step, we fine-tuned an open-source LLM[30–33], Mistral[34], using a question–answer dataset derived from the OEDB (Fig. 6b, left; see Methods section and Table S1 for training details). After fine-tuning, the model demonstrated improved accuracy in answering questions related to electrolyte composition and physicochemical properties contained in the OEDB, compared with the baseline model (Fig. 6b, right). These results suggest that training LLMs with OEDB-based data provides a practical pathway toward adapting language models for electrolyte research. Future integration of experimental datasets and additional simulation results is expected to further expand the scope and utility of this framework.

Discussion

In this work, we developed the Open Electrolyte Database for Batteries (OEDB), a first-of-its-kind open electrolyte-level database, by deploying a

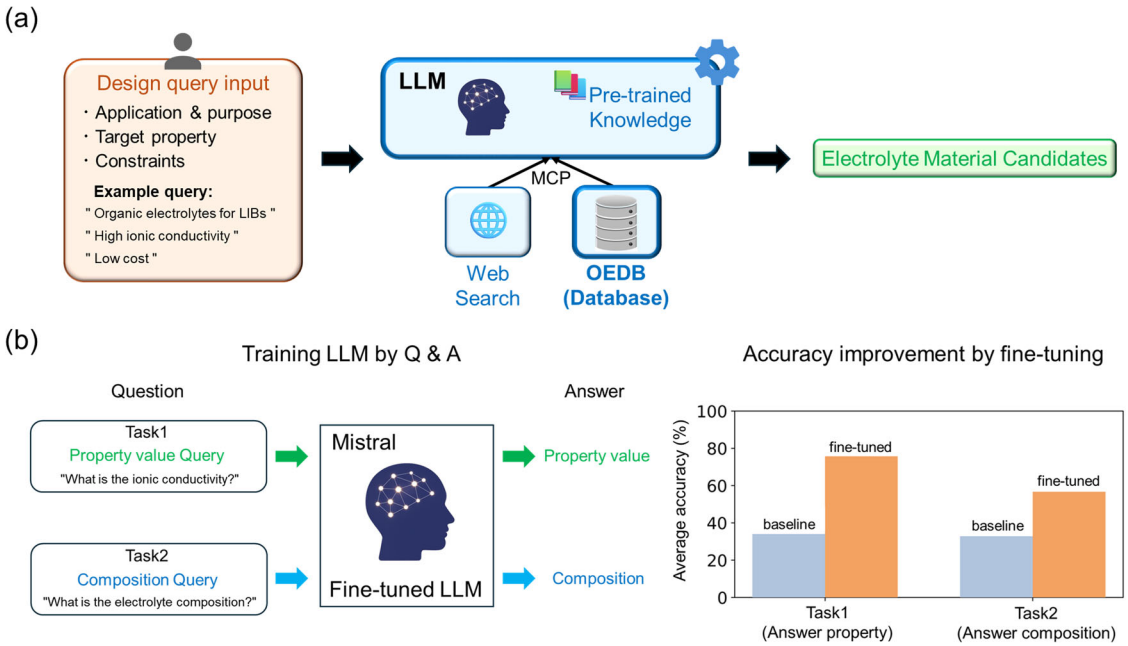

Fig. 6 | OEDB-grounded LLM frameworks for electrolyte design and knowledge adaptation. a Schematic overview of the LLM-assisted framework, in which a user query specifying application context, target properties, and design constraints is processed by a LLM connected to the OEDB via the Model Context Protocol. The model integrates pretrained general knowledge with structured OEDB data and publicly available web information to organize candidate electrolyte compositions and associated property trends. **b** Training and evaluation of a domain-specific knowledge using question-and-answer data derived from OEDB, illustrating representative property-value and composition queries and the corresponding improvement in average accuracy after fine-tuning compared with the baseline model.

fully automated, high-throughput MD framework on the large-scale computational environment mdx. This approach enables the generation and aggregation of solution structures and physicochemical properties for a wide range of electrolyte compositions under unified conditions. By organizing these data consistently at the level of electrolyte composition and providing integrated access through a graphical user interface, electrolyte structures, properties, and compositions can be explored and compared within a single research environment. In addition, leveraging the OEDB as a structured contextual reference, we introduced an LLM-based framework for organizing and interpreting electrolyte-related information in a natural-language setting, supporting application-driven exploration of candidate electrolyte systems. We are committed to the continuous development of the OEDB through frequent updates that simultaneously expand its coverage and improve data quality, by incorporating data generated from higher-fidelity MD simulations, experimental measurements, and increasingly complex electrolyte formulations.

Methods

Automated high-throughput MD simulations

The electrolyte composition space comprised a total of 5616 distinct compositions, spanning three alkali metal cations (Li^+, Na^+ and K^+), nine anions including $BETI^-$ (bis(pentafluoroethanesulfonyl)imide), BF_4^- (tetrafluoroborate), ClO_4^- (perchlorate), $DFOB^-$ (difluoro(oxalato)borate), FSI^- (bis(fluorosulfonyl)imide), OTf^- (trifluoromethanesulfonate), PF_6^- (hexafluorophosphate), TFA^- (trifluoroacetate), and $TFSI^-$ (bis(trifluoromethanesulfonyl)imide), as well as 26 organic solvents. The solvent set included acetone (Ac), acetic anhydride (Ac_2O), acetonitrile (AN), butylene carbonate (BC), diethyl carbonate (DEC), dimethyl carbonate (DMC), 1,3-dimethyl-2-imidazolidinone (DMI), dimethyl sulfoxide (DMSO), ethyl acetate (EA), ethyl butyrate (EB), ethylene carbonate (EC), ethyl methyl carbonate (EMC), ethyl propionate (EP), fluoroacetonitrile (FAN), ethyl fluoroacetate (FEA), fluoroethylene carbonate (FEC), methyl (2,2,2-trifluoroethyl) carbonate (FEMC), γ-butyrolactone (GBL), glutaronitrile (GN), methyl acetoacetate (MA), methyl butyrate (MB), methyl pyruvate (MP), propylene carbonate (PC), succinonitrile (SN), triethyl phosphate (TEP), and trimethyl phosphate (TMP). Salt concentrations were varied from 0.5 to 4.0 mol kg^{-1} in increments of 0.5 mol kg^{-1}.

All MD simulations were carried out using the LAMMPS package[35] with periodic boundary conditions applied in all three spatial dimensions. Interatomic interactions were described using classical force fields: cation parameters were taken from ref. 36, solvent molecules were parameterized using the General AMBER Force Field (GAFF)[37], and anion force field parameters were taken from OPLS-AA[38–42]. Atomic partial charges were derived using the restrained electrostatic potential (RESP) method[43] based on density functional theory calculations performed with Psi4[44] at the B3LYP/aug-cc-pVDZ level. Following established practices for concentrated (or ionic liquid) electrolytes[45–49], ionic charges were uniformly scaled by a factor of 0.8. The complete set of atom types and partial charges is provided in Table S2.

Initial simulation cells were constructed as cubic boxes with side lengths of approximately 80 Å, containing roughly 30,000 atoms. Molecular compositions were determined to reproduce the target salt concentrations for each system. Initial configurations were generated using PACKMOL[50]. Each system was first relaxed for 1 ns under *NPT* conditions at 1 atm and 298 K, followed by a 1 ns equilibration period and a 20 ns production run under *NVT* conditions at 298 K. A time step of 1 fs was employed throughout all simulations, and SHAKE constraints[51] were applied to all bonds involving hydrogen atoms. Long-range electrostatic interactions were treated using the particle–particle particle–mesh (PPPM) method[52] with a real-space cutoff of 10 Å.

The structural and transport properties of each electrolyte composition were evaluated from the MD trajectories, including radial distribution functions (RDFs), coordination numbers, density, ionic conductivity, and viscosity; detailed definitions and computational procedures are provided in the following sections. Among these, density, ionic conductivity, and viscosity were further compared with experimental data for selected electrolyte systems (LiFSI/AN, LiFSI/DMC, LiFSI/EMC, LiFSI/PC, and LiFSI/FEC systems at 0.5 and 1.0 mol kg^{-1}); experimental measurement procedures are described below. For these systems, five independent production runs were performed, and the averaged values with error bars representing the

standard error are shown in Fig. S1. In addition, for the LiFSI/EC system, ionic conductivity was further analyzed as a function of salt concentration, and the averaged values from five independent production runs with error bars are shown in Fig. S2.

Electrolyte solution structures were characterized using RDFs[53] describing atomic pair correlations between cations and solvent molecules, as well as between cations and anions. RDFs were evaluated with a bin width of 0.05 Å up to a maximum distance of 30 Å. For each atomic pair *i*–*j*, the RDF $g_{ij}(r)$ was defined as

$$g_{ij}(r) = \frac{V}{N_i N_j 4\pi r^2 d} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \delta\left(r - r_{ij}\right) \quad (1)$$

where V is the system volume, N_i and N_j are the numbers of atoms of types i and j, respectively, d is the bin width, and r_{ij} is the interatomic distance.

Coordination numbers were obtained by integrating the RDF from $r = 0$ to the first minimum r_c of each pair distribution function,

$$N_c = 4\pi\rho_j \int_0^{r_c} r^2 g_{ij}(r) dr \quad (2)$$

where $\rho_j = N_j/V$ is the number density of species j. The cutoff distance r_c was determined automatically for each system.

Self-diffusion coefficients were calculated from the mean square displacement (MSD)[53] of individual species using the Einstein relation. The MSD was defined as

$$\mathrm{MSD}(t) = \frac{1}{n} \left\langle \sum_i \left| \boldsymbol{R}_i(t) - \boldsymbol{R}_i(0) \right|^2 \right\rangle \quad (3)$$

and diffusion coefficients were obtained from the long-time limit,

$$D_i = \lim_{t\to\infty} \frac{\mathrm{MSD}(t)}{6t} \quad (4)$$

MSDs were evaluated over the full 20 ns production trajectories, and diffusion coefficients were extracted from the linear regime at long time scales.

Ionic conductivity was computed using the Einstein formulation of the Green–Kubo relation[54–56] by averaging over 10,000 time windows spanning lag times from 1 ps to 10 ns. In this formulation, the ionic conductivity σ was calculated as

$$\sigma = \lim_{t\to\infty} \frac{e^2}{6Vtk_BT} \sum_{i,j} z_i z_j \left\langle \left[r_i(t) - r_i(0)\right] \cdot \left[r_j(t) - r_j(0)\right] \right\rangle \quad (5)$$

where z_i and z_j are the charge numbers of ion i and j, $\mathbf{r}_i(t)$, $\mathbf{r}_j(t)$ are their positions at time t, k_B is the Boltzmann constant, and T is the temperature. Final conductivity values were obtained from the slope of the mean charge displacement in the 1–2 ns time window.

Viscosity was evaluated using the Einstein form of the Green–Kubo relation[54,55,57] based on stress tensor fluctuations. In this formulation, the viscosity η was calculated as

$$\eta = \lim_{t\to\infty} \frac{1}{2Vk_BT} \frac{d}{dt} \left\langle \left[\int_0^t P_{xy}(t')dt' \right]^2 \right\rangle \quad (6)$$

where $P_{xy}(t)$ is an off-diagonal component of the stress tensor. Production trajectories were divided into 100 segments, and viscosity values were calculated from time windows ranging from 5 fs to 50 ps. Outlier values exceeding approximately 10% of the standard deviation were excluded, and the median of the remaining values was reported.

Experimental measurement of electrolyte properties

The ionic conductivity of electrolyte solutions was measured by electrochemical impedance spectroscopy using a two-electrode cell equipped with platinum electrodes. Alternating-current impedance spectra were recorded over a frequency range from 100 mHz to 10 kHz, and ionic conductivities were determined from the bulk resistance extracted from the impedance response. The viscosity of the electrolyte solutions was measured using a rotational viscometer (EMS-1000S, Kyoto Electronics Manufacturing Co., Ltd.). The density of the electrolyte solutions was measured using a portable density meter (DMA 35, Anton Paar GmbH). All measurements were performed at room temperature. Electrolyte samples were prepared using the same compositional definitions as those employed in the molecular dynamics simulations, ensuring direct consistency between the experimental and computational datasets.

Electrolyte structure mapping using t-SNE

High-dimensional structural descriptors derived from RDFs were mapped onto a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE)[22]. For each electrolyte composition, the input descriptor was constructed by concatenating RDFs corresponding to two types of atomic pair correlations: (i) anion atoms coordinating to the cation and (ii) solvent atoms coordinating to the cation. For each atomic pair type, RDFs were sampled over the distance range from 0 to 10 Å with a bin width of 0.05 Å, resulting in fixed-length descriptor vectors. Prior to dimensionality reduction, all descriptor vectors were standardized to have zero mean and unit variance. The t-SNE analysis was performed using the default hyperparameter settings of the Python implementation, without manual optimization.

LLM-assisted database interaction and fine-tuning

A large language model (LLM) was fine-tuned using question–answer (QA) data generated from the OEDB, following the methodology established in the aLLoyM framework[30]. The pretrained Mistral-Nemo-Instruct-2407 model was employed as the base model. For each electrolyte entry in the OEDB, structured database records were converted into pairs of natural-language questions and their corresponding correct answers, with each pair constituting a single training sample. The questions were designed to emulate typical database query scenarios and were categorized into two task types. Task 1, referred to as a property query, required the model to return a specific physicochemical property value for a given electrolyte composition, such as a diffusion coefficient or ionic conductivity. Task 2, referred to as a composition query, required the model to identify a composition element, such as the cation or solvent species, that satisfies a specified set of conditions. Representative examples of the QA formats used for fine-tuning are provided in Supplementary Table S1.

Fine-tuning was performed using low-rank adaptation (LoRA)[32] with rank 16 and alpha 16, applied to both attention and feed-forward projection layers. The model was trained for 15,000 optimization steps using the AdamW optimizer with a learning rate of 2×10^{-4}. Training was conducted with a batch size of 16 per device, four gradient accumulation steps, and bfloat16 precision. For model evaluation, multiple-choice questions were generated using the same two query types. Each question consisted of one correct answer and two incorrect alternatives, and the model was required to select the correct option. Electrolyte compositions were partitioned into training and evaluation sets such that no composition appeared in both sets. Evaluation was carried out for both the fine-tuned model and the baseline model under identical conditions. Model performance was quantified as the classification accuracy, defined as the fraction of questions for which the correct option was selected.

Data availability

The electrolyte database generated in this study is publicly available at https://oedb.jp. This database contains all data underlying the results reported in this work. All information necessary to reproduce the

calculations, including computational details and protocols, is provided in the main text, Methods, and Supplementary Information.

Received: 28 January 2026; Accepted: 10 April 2026;
Published online: 20 April 2026

References

1. Goodenough, J. B. & Kim, Y. Challenges for rechargeable Li batteries. *Chem. Mater.* **22**, 587–603 (2010).
2. Tarascon, J.-M. & Armand, M. Issues and challenges facing rechargeable lithium batteries. *Nature* **414**, 359–367 (2001).
3. Goodenough, J. B. & Park, K.-S. The Li-ion rechargeable battery: a perspective. *J. Am. Chem. Soc.* **135**, 1167–1176 (2013).
4. Xu, K. Electrolytes and interphases in Li-ion batteries and beyond. *Chem. Rev.* **114**, 11503–11618 (2014).
5. Yamada, Y., Wang, J., Ko, S., Watanabe, E. & Yamada, A. Advances and issues in developing salt-concentrated battery electrolytes. *Nat. Energy* **4**, 269–280 (2019).
6. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
7. Qu, X. et al. The Electrolyte Genome project: a big data approach in battery materials discovery. *Comput. Mater. Sci.* **103**, 56–67 (2015).
8. Irwin, J. J. & Shoichet, B. K. ZINC – a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
9. Kim, S. et al. PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373–D1380 (2023).
10. Hannah, D. et al. Searching for ideal electrolytes in the molecular universe. *Electrochem. Soc. Interface* **34**, 35–38 (2025).
11. Wang, B. et al. Data-driven design of electrolyte additives supporting high-performance 5 V $LiNi_{0.5}Mn_{1.5}O_4$ positive electrodes. *Nat. Commun.* **16**, 3413 (2025).
12. Borodin, O. & Smith, G. D. Quantum chemistry and molecular dynamics simulation study of dimethyl carbonate: ethylene carbonate electrolytes. *J. Phys. Chem. B* **113**, 1763–1776 (2009).
13. Borodin, O. & Smith, G. D. Mechanism of ion transport in amorphous poly(ethylene oxide)/LiTFSI from molecular dynamics simulations. *Macromolecules* **39**, 1620–1629 (2006).
14. Suzumura, T. et al. mdx: a cloud platform for supporting data science and cross-disciplinary research collaborations. In *Proceedings of 2022 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* 1–7 (IEEE, Falerna, 2022).
15. Hanai, M. et al. ARIM-mdx data system: towards a nationwide data platform for materials science. In *Proceedings of 2024 IEEE International Conference on Big Data* 2326–2333 (IEEE, Washington, DC, 2024).
16. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
17. Bedrov, D. et al. Molecular dynamics simulations of ionic liquids and electrolytes using polarizable force fields. *Chem. Rev.* **119**, 7940–7995 (2019).
18. Wu, S. et al. Applications and advances in machine learning force fields. *J. Chem. Inf. Model.* **63**, 6972–6985 (2023).
19. Car, R., De Angelis, F., Giannozzi, P. & Marzari, N. First-principles molecular dynamics. in *Handbook of Materials Modeling* (ed. Yip, S.) 59–76 (Springer, 2005).
20. Doyle, M., Fuller, T. F. & Newman, J. The importance of the lithium ion transference number in lithium/polymer cells. *Electrochim. Acta* **39**, 2073–2081 (1994).
21. Dong, D. & Bedrov, D. Charge transport in [Li(tetraglyme)][bis(trifluoromethane)sulfonimide] solvate ionic liquids: insight from molecular dynamics simulations. *J. Phys. Chem. B* **122**, 9994–10004 (2018).
22. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
23. Atkins, P., de Paula, J. & Keeler, J. *Atkins' Physical Chemistry* 11th edn (Oxford Univ. Press, Oxford, 2018).
24. McQuarrie, D. A. *Statistical Mechanics* (University Science Books, Sausalito, 2000).
25. Brown, T. B. et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
26. Bai, Y. et al. Constitutional AI: harmlessness from AI feedback. Preprint at https://arxiv.org/abs/2212.08073 (2022).
27. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
28. Mishra, V. et al. Foundational large language models for materials research. Preprint at https://arxiv.org/abs/2412.09560 (2024).
29. Hou, X., Zhao, Y., Wang, S. & Wang, H. Model Context Protocol (MCP): landscape, security threats, and future research directions. Preprint at https://arxiv.org/abs/2503.23278 (2025).
30. Oikawa, Y., Deffrennes, G., Shimayoshi, R., Abe, T., Tamura, R. & Tsuda, K. aLLoyM: a large language model for alloy phase diagram prediction. *npj Comput. Mater.* **12**, 97 (2026).
31. Wei, J. et al. Finetuned language models are zero-shot learners. In *Proc. 10th International Conference on Learning Representations* (ICLR, 2022).
32. Hu, E. J. et al. LoRA: low-rank adaptation of large language models. In *Proc. 10th International Conference on Learning Representations* (ICLR, 2022).
33. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at https://arxiv.org/abs/2108.07258 (2021).
34. Jiang, A. Q. et al. Mistral 7B. Preprint at https://arxiv.org/abs/2310.06825 (2023).
35. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
36. Jensen, K. P. & Jorgensen, W. L. Halide, ammonium, and alkali metal ion parameters for modeling aqueous solutions. *J. Chem. Theory Comput.* **2**, 1499–1509 (2006).
37. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
38. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
39. Doherty, B., Zhong, X., Gathiaka, S., Li, B. & Acevedo, O. Revisiting OPLS force field parameters for ionic liquid simulations. *J. Chem. Theory Comput.* **13**, 6131–6145 (2017).
40. Shimizu, K., Almantariotis, D., Costa Gomes, M. F., Pádua, A. A. H. & Canongia Lopes, J. N. Molecular force field for ionic liquids V: hydroxyethylimidazolium, dimethoxy-2-methylimidazolium, and fluoroalkylimidazolium cations and bis(fluorosulfonyl)amide, perfluoroalkanesulfonylamide, and fluoroalkylfluorophosphate anions. *J. Phys. Chem. B* **114**, 3592–3600 (2010).
41. Han, S.-D. et al. Electrolyte solvation and ionic association: IV. acetonitrile–lithium difluoro(oxalato)borate (LiDFOB) mixtures. *J. Electrochem. Soc.* **160**, A2100–A2110 (2013).
42. Xing, H. et al. Molecular dynamics simulation study on the absorption of ethylene and acetylene in ionic liquids. *Ind. Eng. Chem. Res.* **52**, 9308–9316 (2013).
43. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for

deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
44. Smith, D. G. A. et al. PSI4 1.4: open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **152**, 184108 (2020).
45. Zhang, Y. & Maginn, E. J. A simple AIMD approach to derive atomic charges for condensed phase simulation of ionic liquids. *J. Phys. Chem. B* **116**, 10036–10048 (2012).
46. Ray, P., Vogl, T., Balducci, A. & Kirchner, B. Structural investigations on lithium-doped protic and aprotic ionic liquids. *J. Phys. Chem. B* **121**, 5279–5292 (2017).
47. Molinari, N., Mailoa, J. P. & Kozinsky, B. General trend of a negative Li effective charge in ionic liquid electrolytes. *J. Phys. Chem. Lett.* **10**, 2313–2319 (2019).
48. Zhang, Y. et al. Water or anion? Uncovering the Zn^{2+} solvation environment in mixed $Zn(TFSI)_2$ and LiTFSI water-in-salt electrolytes. *ACS Energy Lett.* **6**, 3458–3463 (2021).
49. Wettstein, A., Diddens, D. & Heuer, A. Controlling Li^+ transport in ionic liquid electrolytes through salt content and anion asymmetry: a mechanistic understanding gained from molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **24**, 6072–6086 (2022).
50. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).
51. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
52. Hockney, R. W. & Eastwood, J. W. *Computer Simulation Using Particles* (Adam Hilger, Bristol, 1988).
53. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
54. Green, M. S. Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids. *J. Chem. Phys.* **22**, 398–413 (1954).
55. Kubo, R. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.* **12**, 570–586 (1957).
56. Verma, A. K., Thorat, A. S. & Shah, J. K. Estimating ionic conductivity of ionic liquids: Nernst–Einstein and Einstein formalisms. *J. Ion. Liq.* **4**, 100089 (2024).
57. Yao, N. et al. Probing the origin of viscosity of liquid electrolytes for lithium batteries. *Angew. Chem. Int. Ed.* **62**, e202305331 (2023).

Acknowledgements

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Programme: Data Creation and Utilization Type Materials Research and Development Project (A.Y.; Grant Number JPMXP1121467561). In this research work, we used the "mdx: a platform for building data-empowered society". The authors are grateful to Shoichi Matsuda from National Institute for Materials Science for insightful and helpful discussions.

Author contributions

A.Y. proposed the concept and initiated the project. A.Y. and J.S. co-directed the project. K.N. and N.T. performed computational simulations and analyzed the data. M.H. implemented the graphical user interface and the LLM-assisted framework. K.N., Y.O., and R.T. fine-tuned the LLM. K.N., N.T., and A.Y. wrote the manuscript. K.N., N.T., M.H., R.T., K.T., M.N., J.S., and A.Y. contributed to editing the manuscript.

Competing interests

Koji Tsuda is an Editorial Board Member of npj Computational Materials. He was not involved in the journal's review of, or decisions related to, this manuscript. The other authors declare no competing financial or non-financial interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-026-02093-y.

Correspondence and requests for materials should be addressed to Atsuo Yamada.