

REASONING-TO-SIMULATION: AN AGENTIC FRAMEWORK FOR DISCOVERY OF ELECTROLYTE MATERIALS

Jurģis Ruža & Rafael Gómez-Bombarelli

Department of Materials Science and Engineering

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139, USA

{jruza, rafagb}@mit.edu

ABSTRACT

The discovery of novel metal-ion battery electrolytes is hindered by the vast combinatorial space of solvent-salt formulations and by the complexity of physics-based validation. We propose an agentic framework that leverages Large Language Models (LLMs) as autonomous orchestrators for closed-loop materials discovery. Our system features a Discovery Agent that performs multi-generational composition optimization and a Molecular Dynamics (MD) simulation Assistant that translates natural language prompts into executable liquid electrolyte molecular dynamics simulations. By integrating generative reasoning with automated simulation verification, our framework enables an autonomous "reasoning-to-simulation" cycle that identifies high-performance electrolytes. This work demonstrates the ability of agentic LLMs to bridge the gap between chemical intuition and rigorous computational materials science by discovering a high-performance liquid electrolyte solvent formulation with computationally measured 106 mS/cm ionic conductivity.

1 INTRODUCTION

The traditional "trial-and-error" paradigm of materials discovery is rapidly being replaced by a computational-first approach, where high-fidelity simulations and machine learning (ML) serve as the primary engines for exploration. (1; 2; 3) In fields ranging from drug discovery to heterogeneous catalysis, the integration of molecular dynamics (MD) simulations, density functional theory (DFT) calculations, and ML models has significantly compressed development timelines. (4; 5; 6; 7; 8) This shift is particularly critical in the field of lithium and sodium-ion batteries, where the discovery of high ionic conductivity and stable electrolytes involves navigating a near-infinite combinatorial space of solvent-salt formulations, additives, and concentrations. (9; 10)

Despite the predictive power of modern simulations, a substantial bottleneck remains: the "human-in-the-loop" requirement. The set-up of complex liquid electrolyte simulations - configuring coordination environments, managing force field parameters, and post-processing transport properties — requires specialized expertise and manual effort. This manual overhead limits the scale of discovery and creates a disconnect between molecular design and computational validation.

In this work, we propose a shift toward autonomous discovery through an agentic framework powered by Large Language Models (LLMs). By leveraging the reasoning, memory, and tool-use capabilities of LLMs, we free the researcher from the technical minutiae of simulation execution. Our electrolyte agent framework introduces two specialized working modes: a Discovery mode that iteratively optimizes electrolyte solvent compositions through multi-generational learning and a Simulation Assistant mode that automates the transition from natural language intent to rigorous physics-based reporting. This approach provides a scalable, end-to-end pipeline that transforms high-level scientific hypotheses into validated computational predictions of liquid electrolyte performance. To the best of our knowledge, this is the first framework that is able to optimize electrolyte solvents with any number of components, ratios between them, and continuous salt molalities for liquid electrolyte applications. Furthermore, the integration of a fully automated MD pipeline that

is able to handle everything from force-field parametrization to system equilibration sets this framework apart from other similar MD related agentic systems.

1.1 LIQUID ELECTROLYTE COMPOSITION OPTIMIZATION

Liquid electrolytes are the backbone for the progress in Li-ion and Na-ion battery technology. (11; 12) Gradual advancements in finding solvents with higher ionic conductivity and stability against decomposition have brought the higher-charge-density Li and Na-metal batteries closer to commercialization. (13; 14) Recently, there has been work on data-driven experimental optimization of single solvent liquid electrolytes (15; 16; 17; 18), high-throughput MD screening of single solvent properties (19; 20) and some limited work attempting to optimize ratios between a few solvents or additives. (21; 22; 23; 24). While these works advance the development of new solvent chemistries, it is common for an electrolyte to consist of 2 or 3 different solvents to balance their properties. (25; 26).

1.2 LARGE LANGUAGE MODELS IN SIMULATIONS

Recent developments in Large Language Models (LLMs) have catalyzed a shift from static scripts to thinking agents capable of orchestrating complex computational chemistry workflows. (27) Specialized agentic frameworks like MDCrow, El Agente, DREAMS have demonstrated the ability to use chain-of-thought reasoning to autonomously prepare input files, select appropriate force fields, and manage high-performance computing (HPC) resources for molecular dynamics (MD) and density functional theory (DFT) simulations. (28; 29; 30) By leveraging tool-calling APIs, these models can interface directly with established software suites like OpenMM, ORCA or VASP, translating high-level scientific intent—such as "calculate the bandgap of Si"—into a sequence of executable tasks. (31; 32; 33; 34; 35) This trend elevates AI from a property-prediction tool to a digital collaborator that can handle the nuanced and iterative nature of physics-based modeling.

However, the rapid proliferation of LLM-based automation has also introduced redundancy. Many recent works attempt to use generative models for tasks that are already efficiently handled by existing rule-based automation or established software wrappers. To achieve higher scientific value, agentic frameworks should offer automation beyond what is provided by existing scientific tools. (36; 37; 38)

1.3 LARGE LANGUAGE MODELS FOR MATERIALS DISCOVERY

The shift from traditional machine learning (ML) models to agentic frameworks represents a fundamental evolution in the way artificial intelligence interacts with materials science. Emerging agentic systems like HoneyComb, LLMatDesign, AtomGPT, and ChatMOF utilize LLMs as "reasoning engines" or central orchestrators. (39; 40; 41; 42) These agents do not merely suggest new structures; they autonomously decompose high-level scientific goals into executable sub-tasks, such as suggesting specific elemental substitutions based on periodic trends, calling protein folding models or using property prediction models. (42; 43; 44) This shift enables a more holistic approach to discovery, where the AI manages and performs scientific research rather than just a single step in the pipeline. Furthermore, LLM frameworks have been shown to work in an iterative fashion. By learning from previous experiments they are able to optimize chemical spaces, and at times outperform Bayesian Optimization algorithms (45; 46; 47) and discover state-of-the-art electrocatalysts and previously unknown reaction pathways. (48; 49).

2 METHODS

Figure 1 shows the workflow of the electrolyte agent framework. It has two working modes, MD assistant and discovery mode. Both modes have access to the same set of tools, but they have different system prompts to guide their behavior towards assisting a user or running a discovery campaign over multiple generations. Each mode system prompt is available in the Appendix A.4 Figure 11. The agentic framework is built with the language model integration software LangGraph 1.0.1. The discovery mode runs in an autonomous orchestration mode, where with a user prompt (Figure 2 A) the agentic framework runs a discovery campaign autonomously and produces a report in the end.

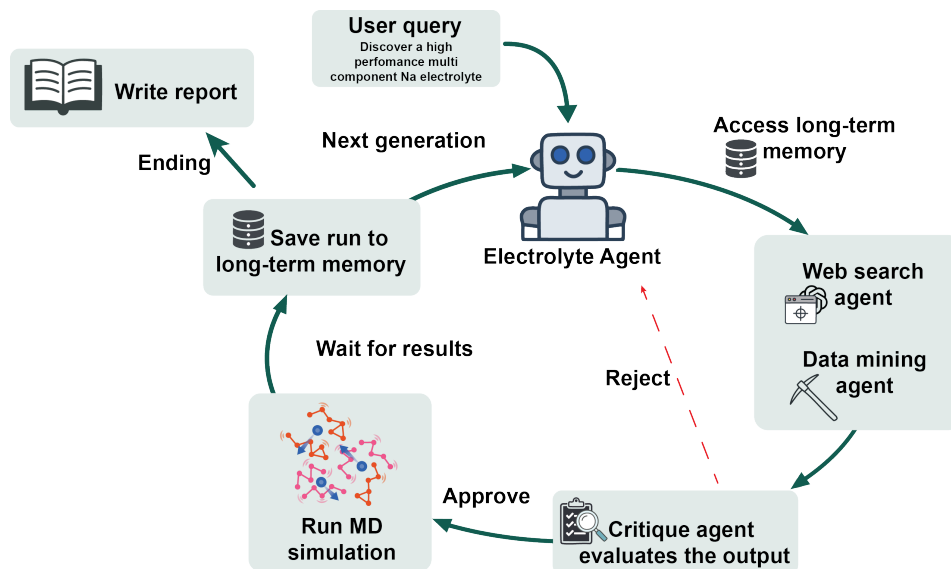


Figure 1: Workflow of the agentic framework. Starting from a user query the electrolyte agent with the assistance of web search and data mining agents makes an MD simulation plan to try to fulfill the query. After suggesting a set of simulations a critique agent evaluates whether they are valid and align to the user goal. Once approved an MD simulation is performed, a set of relevant properties are calculated and the results returned to the agent. Trace of the tool usage and reasoning is saved in long-term memory and either next discovery generation is performed or a report summarizing the findings is synthesized.

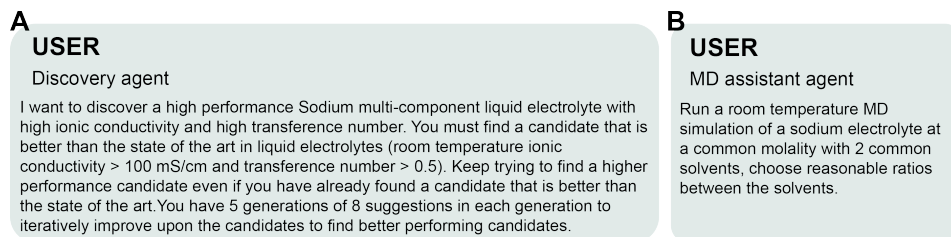


Figure 2: Example of a user prompt for the electrolyte agent in (a) Discovery mode, and (b) MD assistant mode.

While in discovery mode it operates in an plan-act-reflect manner, where through prompts it creates a plan for the discovery pipeline and through reflection upon results plans the future generations. (50) The MD assistant mode runs with human in the loop, where it can ask for clarifying questions from the user to better assist them in running the necessary MD simulations. Both the discovery mode and MD assistant use GPT-5 (gpt-5-2025-08-07).

2.1 WEB-SEARCH TOOL

The framework employs the OpenAI web-search API to perform reasoning-based search over a query (gpt-o4-mini-2025-04-16). (51) The queries are submitted by the main electrolyte agent to obtain more information about the user query. Given the discovery mode query in Figure 2 A, the web-search agent is prompted with a query of finding out typical ionic conductivity and transference number values (example in Appendix A.3 Figure 7). The output for such a prompt consists of multiple electrolytes with their reported ionic conductivities, transference numbers, and the relevant available sources. A similar agent prompt is made for the MD assistant query in Figure 2 B, so that it can verify common candidates to choose from to fulfill the user query.

2.2 DATA MINING TOOL

The electrolyte agent has access to a data mining agent that has access to multiple data related tools. The purpose of this tool being an agent is to be able to handle simple queries from the electrolyte agent and decide how to query the available dataset most efficiently. It has access to a dataset of hypothetical liquid electrolyte solvents by Kumar et al. (52) This dataset consists of almost 80k small molecule liquid electrolyte solvent candidates and their relevant predicted properties for battery applications (ionic conductivity, oxidative stability and Columbic efficiency). All the properties are only for single components. The agent is able to make queries in this dataset based on SMARTS syntax (finding chemical sub-patterns), order the queries by predicted ionic conductivity, Coulombic efficiency, oxidative stability, heavy atom count, molecular weight and is able to subselect certain common predefined (ethers, amides, halides, among others) chemical fragments. The output of the data mining agent is set of liquid electrolyte candidates with each of the aforementioned properties and a short reasoning why that molecule has been selected. The electrolyte agent can also use the PubChem (53) API to recover SMILES strings of any chemical that is available online. The electrolyte agent uses these outputs to create a set of multi-component liquid electrolytes for the MD simulation tool to verify.

2.3 SELF-CRITIQUE AGENT

Before submitting the simulations to run the proposed candidates are verified by a self-critique agent (gpt-5.2-2025-12-11). This agent is tasked with evaluating if the proposed parameters are valid for the MD simulation tool (system prompt in Appendix A.4 Figure 12). It compares the parameters to the web-search output to evaluate if there are any incompatible chemistries or candidates that have already been researched before. Most importantly, it evaluates whether the proposed candidates align with the original user prompt. If this agent rejects the candidates (example of rejected candidates in Appendix A.3 Figure 8), the candidates are sent back to the electrolyte agent to propose a new set of candidates that meet the criteria made by the self-critique agent.

2.4 MD SIMULATION TOOL

The crux of the framework is being able to verify the proposed liquid electrolytes with classical MD simulations. The agent stages MD simulation parameters that are passed into HiTPoly, a platform for running high-throughout polymer and liquid electrolyte simulations. (54) Given inputs—SMILES strings for an arbitrary number of components, ratios between components, salt identity, salt molality, temperature, simulation length, and charge scaling factor—HiTPoly creates force field parameterizations of all the components, builds and equilibrates the simulation box, does a production simulation and performs transport and structure property analysis. The ionic conductivity, the transference number, the fraction of free charge carriers (proxy for cation solvation), the diffusivities of the simulation components and the coordination environments of the cation are fed back to the electrolyte agent. Full MD simulations details found in Appendix A.1.

2.5 LONG-TERM MEMORY

To enhance the performance across generations and retain important information the agent saves a compact narrative about the plan for the generation, a summary of the main insights and best-of-run results to long-term memory. (55; 56) The message is embedded into a 1024 dimensional vector with OpenAI text-embedding-3-large model for text embeddings. (57) To retrieve the messages from the long-term memory we use the Facebook AI Similarity Search (FAISS) over the saved embeddings. (58) The long-term memory can be used to restart failed runs, or continue already finished runs with the insights collected from those runs. When running comparisons between various discovery runs, the long-term memory is always initialized from scratch unless mentioned otherwise. This is done to test the agents performance from zero without it learning from other discovery campaigns.

2.6 REPORT WRITING TOOL

Upon completing a discovery campaign or a MD assistant run the agent summarizes the results in a simple report. It recovers any relevant messages from the long-term memory storage, uses the saved

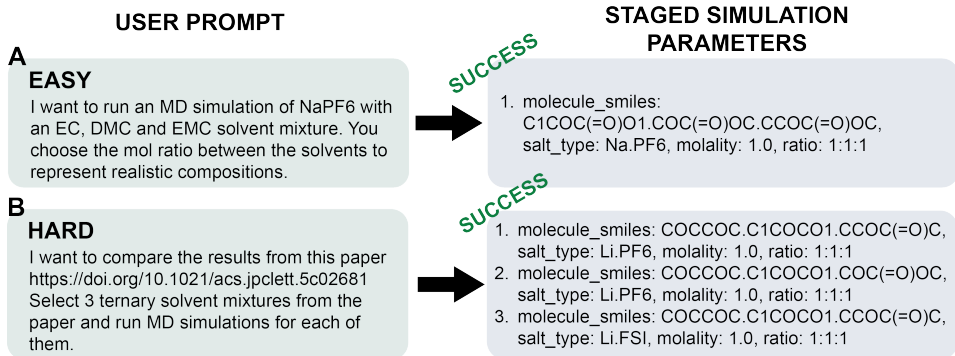


Figure 3: Example of the MD assistant user prompts and output staged simulation parameters.

MD results and synthesizes an overview. It includes a review of tools used and main insights from each generation, conclusions about the campaign, potential next steps and insights learned.

3 RESULTS AND DISCUSSION

We studied the success and failure modes of the electrolyte agent in both the discovery and MD assistant modes. We evaluated the performance of the MD assistant across multiple queries ranging in difficulty and interpretability. We also evaluated the ability of the electrolyte agent in the discovery mode to learn across generations and how the prompt language affects the performance of it.

3.1 MOLECULAR DYNAMICS ASSISTANT

We evaluated the MD assistant by asking it questions of varying difficulty and evaluating how well it performed in these tasks. All the prompts were related to liquid electrolyte simulations. Figure 3 A and B show example prompts. The easy prompt (A) defines most of the information that the agent needs to look up, while the harder (B) prompt poses a question of correctly extracting information from a research paper. A medium prompt gives a more open ended question than the easy prompt (example in Appendix section A.2). With using the various tools the agent composes a fixed dictionary of input parameters that the automated MD pipeline processes to run the simulation. The MD pipeline does full force-field parametrization, simulation box construction, equilibration, production simulation and a suite of analyses tools are applied to extract information from these simulations (59; 54). Finally the agent takes the initial prompt, simulation results and produces a report of what was done.

Table 1: Comparison of Agent and GPT-5 performance across easy and medium difficulty prompts for four categories evaluated in the MD parameter suggestion stage. For each difficulty there were 20 different prompts and each prompt was repeated 3 times.

	Easy prompts				Medium prompts			
	Solvents	Solvent ratio	Salt	Concentration	Solvents	Solvent ratio	Salt	Concentration
Agent	0.98	0.98	0.98	0.93	1	1	1	0.97
GPT-5	1	1	1	1	0.95	1	1	0.97

We evaluated and compared the performance of the electrolyte agent with easy (20), medium (20) and hard (18) prompts against a baseline GPT-5 (gpt-5-2025-08-07) model. We compare against the GPT-5 model as that is the same model the electrolyte agent uses in its structure. The GPT-5 was prompted with a simple prompt to produce a json dictionary for the MD pipeline (details in Appendix section A.2), the standalone GPT-5 model cannot run any of the MD simulations thus we treat this as a comparison for the proposed parameters. Each prompt was tested 3 times to obtain statistical significance and test the stability of the agent/LLM for the same prompts. Table 1 shows the comparison of the results for the agent and GPT-5. Across the easy and medium tasks both the agent and GPT-5 perform with great precision. The deviation from perfect accuracy (0.98) in solvent

identity, solvent ratio, and salt identity was due to a single failed agent run (details in Appendix section A.2). For the medium prompts the difference in performance for the solvent identity for GPT-5 arose due to producing wrong SMILES strings for the task (details in Appendix section A.2). Full list of prompts available through the Appendix section A.2.

Table 2: Hard prompt performance breakdown by reference type.

Reference Type	Model	Solvents	Salt	Concentration	Has output
DOI	Agent	1.00	0.89	1.00	1.00
	GPT	0.33	0.28	0.94	0.94
Citation	Agent	0.94	0.61	1.00	1.00
	GPT	0.28	0.28	0.78	0.83
Title	Agent	0.78	0.72	1.00	1.00
	GPT	0.44	0.22	0.83	0.83
Overall	Agent	0.91	0.74	1.00	1.00
	GPT	0.26	0.35	0.85	0.87

Table 2 shows the results for the agent and GPT-5 performance on the hard prompts. We selected 4 well cited or recent research works that were relevant for molecular dynamics simulations (60; 24; 61; 62). Each paper was evaluated with a prompt at 3 different levels, one in which only the DOI of the paper, one where it was given the full title of the paper and one where the full citation with authors, title and journal information. In this more complex task the electrolyte agent significantly outperforms GPT-5 in all the simulation parameter components. The mixture of being able to verify molecular structures, dynamically look up information and having a verifier that critiques the output increases the agent performance. Though it has to be said that it’s still far from perfect. In the easy and medium cases prompt language can affect the model performance in unexpected ways (details in Appendix section A.2). However, for the harder prompts having a vectorized research paper database with retrieval-augmented generation (RAG) can improve the performance significantly.

3.2 DISCOVERY AGENT

The electrolyte agent in discovery mode was tasked with finding high performance Na multi-component liquid electrolytes. We tasked the agent with optimizing two properties simultaneously, ionic conductivity and transference number. These are important properties for liquid electrolytes; high ionic conductivity increases the overall performance of the battery by reducing the internal resistance which is especially important at higher charging rates. Transference number represents the fraction of charge carried by the positive ion (cation). The cation is the species that reacts with the electrodes, thus it is important that more of the charge is carried by cations than anions (high transference number). A low transference number creates a concentration gradient of ions in contact with the electrodes that can increase the internal resistance and propagate dendrite growth, which can eventually lead to a short-circuit in the battery. Other properties like Coulombic efficiency and oxidative stability are also important for the efficiency of the electrolyte but are inaccessible through MD simulations and, as such, are out of the scope of this project. It has to be noted, that the electrolyte agent has access to these properties in the form of ML predicted properties from the dataset by Kumar et al. (52) The discovery mode is not benchmarked against any classical Bayesian Optimization (BO) methods as, to the best of our knowledge, there do not exist any out of the box methods that are able to optimize chemical spaces with a varying degree of components and including ability to query commonly available chemistries. It is important for future work to develop appropriate BO methods to be able to do agentic comparisons.

For the discovery mode we tested its performance in runs of 5 generations, where each generation had 8 possible simulations to be run. We also tested whether running another generation sequentially with sharing the same long-term memory database increases performance by having access to more information. Finally, we evaluated whether increasing the target ionic conductivity in the initial prompt would lead to higher overall performance. Figure 4 A shows the evolution of the mean and standard deviation of the current best candidate for two repeats of two discovery campaigns. For both campaigns the prompts are the same (shown in 2 A) except the value of ionic conductivity that they are targeting. For both prompts it can be seen that the agent showed in-context learning across generations by showing improvement of the cation ionic conductivity (ionic conductivity x trans-

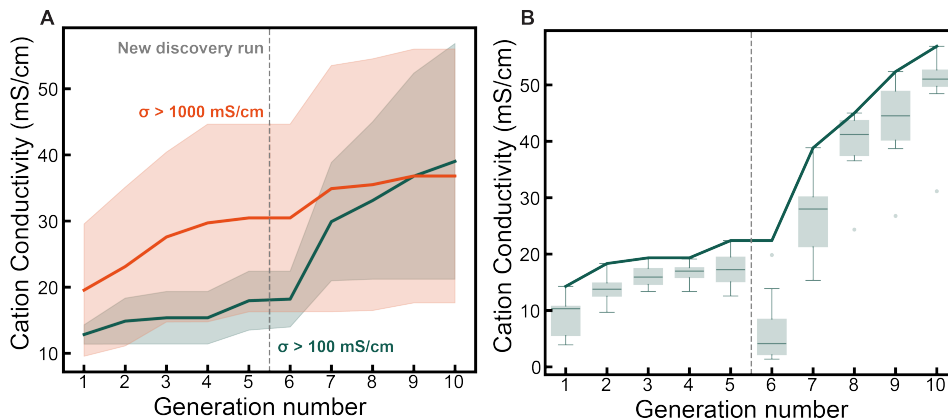


Figure 4: (A) Evolution of the mean and standard deviation of the cation ionic conductivity (ionic conductivity \times transference number) for two discovery runs (2 repeats each). (B) Evolution of the cation ionic conductivity across two runs with the same long term memory and input prompt. The input prompt tasks the agent of finding candidates with ionic conductivity >100 mS/cm and transference number >0.5 . The box plots show the distribution of samples found at every generation.

ference number). The 100 mS/cm ionic conductivity prompt shows a factor of 3 increase in cation ionic conductivity across the 2 repeats. While the 1000 mS/cm prompts start with a higher initial ionic conductivity (mean) both prompt types finish in similar bounds. The decomposed evolution for each run can be found in the Appendix Figure 9.

Figure 4 B shows the evolution of the current best result and distribution of results at each generation for two runs that were initialized with finding candidates with ionic conductivity >100 mS/cm and transference number >0.5 . These runs were run sequentially and allowed to share the same long-term memory database (other campaigns are initialized with their own unique long-term memory). During the first campaign the agent tested 10 unique solvents that it recovered from the dataset by Kumar et al. (52). It combined these solvents in 13 different combinations and tested various salt molalities and ratios between the solvent molecules. During the first 5 generations the agent saw marginal improvements in the cation ionic conductivity. The best results are from a solvent composition of CCC1CO1.N#CC(F)F.O=CC1CO1 (1,2-Epoxybutane, difluoro-acetonitrile, and glycidaldehyde) at 7:1:2 ratio with Na.FSI salt at 1.4 mol/kg molality reaching ionic conductivity 45 mS/cm and transference number of 0.49. None of these solvents have been used as liquid electrolytes to the best of our knowledge.

Figure 5 A shows the detailed results of the generations 6-10 of Figure 4 B. The points in light pink show the MD calculated ionic conductivity and transference number of the solvents with the highest ionic conductivity from the dataset by Kumar et al. (52). The light purple show the MD calculated ionic conductivity and transference number of the highest ionic conductivity binary and ternary solvent mixtures from the work of Shi et al. (24) Figure 4 (B) shows that at generation 6 the performance of the proposed samples decreased significantly. The electrolyte agent did not use any of the smiles from the previous campaign, rather it formed insights about using nitriles as a high-permittivity diluent and using linear carbonates, ethers, and a sulfone as a solvent. While taking these new solvents caused a dip in the first generation performance for the agent, the new insights allowed it to find candidates with higher performance within the next generation (Figure 5 B). Only during the final generation the model was able to satisfy both the >0.5 transference number and >100 mS/cm ionic conductivity.

Compared to the best hypothetical electrolytes from Kumar et al. and the high performance electrolyte mixtures of Shi et al. the agent finds an electrolyte candidate with more than twice the ionic conductivity (106 mS/cm). The candidate is a ternary mixture (CC#N.COC(=O)OC, acetonitrile and dimethyl carbonate) with Na.FSI salt at 2.9 kg/mol molality. While there have been works that incorporate acetonitrile with different types of carbonates (63; 64) we could not to the best of our knowledge find the exact mixture reported here. It has to be added here that acetonitrile can degrade under battery operating conditions. (65)

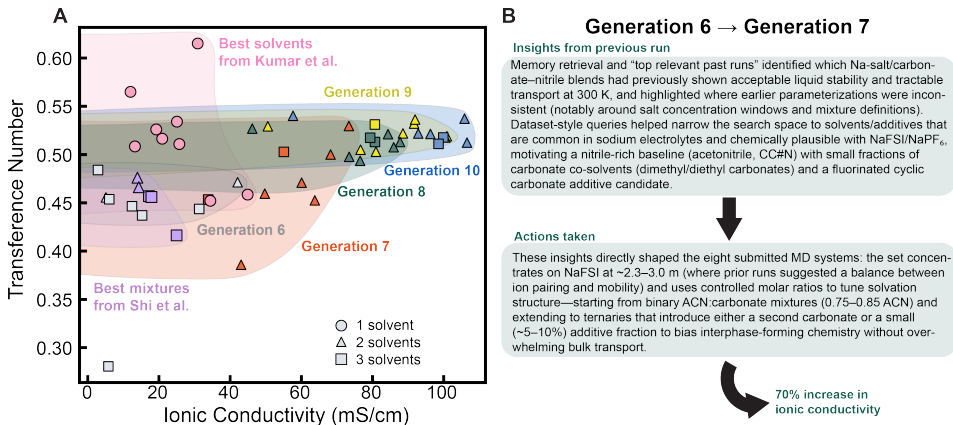


Figure 5: (A) Detailed results of generations 6 through 10 of the optimization campaign with the input prompt tasking the agent of finding candidates with ionic conductivity >100 mS/cm and transference number >0.5 . Overlaid are the highest ionic conductivity single solvents from the work of Kumar et al. (52) and the highest performing solvent mixtures from the work by Shi et al. (24)

The colored borders represent the increase of the Pareto front per generation. (B) Reasoning trace extracted from the electrolyte agent from generation 6 to 7 that led to a 70% increase in ionic conductivity.

4 CONCLUSION

In this paper, we have developed a fully autonomous agentic framework that is able to discover high ionic conductivity and transference number liquid electrolytes by using a wide variety of tools, as well as assist a user in running MD simulations of liquid electrolytes. We have shown that the MD assistant agent is able to handle prompts of various difficulties, stage simulation parameters from these prompts, and create simple reports for the user. Furthermore, we have shown that when compared against a baseline GPT-5 the electrolyte agent heavily outperforms the LLM on hard prompts. This work suggests that the bottleneck in computational electrolyte discovery is shifting from simulation methodology to intelligent orchestration of existing tools.

The electrolyte agent in the discovery mode shows improvement across generations for each of the different prompts and learning across runs from long-term memory. It displays the ability to query the given dataset with reasoning that mirrors chemical intuition and find Na electrolyte solvents that outperform state of the art Li electrolyte solvents. The agent demonstrates the ability to use multiple generations and trade-off between exploration of chemical space and exploitation and optimization around the current best result. It is also able to handle cases when tool usage breaks down and use field-specific knowledge to find previously unused solvent compositions that outperform high-performance Li electrolytes.

The framework also exhibits some clear limitations. Firstly, computational calculations paint only part of the picture of battery relevant properties. It is important to test cycling efficiency, stability against degradation, and flammability to understand the effectiveness of solvents. Secondly, classical molecular dynamics does not offer ways to investigate electrolyte decomposition pathways, which are important for evaluating its compatibility with a battery electrode. Furthermore, measuring ionic conductivity with classical MD can lead to overestimated results, experimental verification is highly important for electrolyte and solvent testing. Third, the electrolyte agent sometimes shows inconsistency in behaviors across runs that can lead to divergent performance while repeating the same prompt. Finally, having a fixed dataset to query limits the ability of the electrolyte agent to use its chemical intuition. A better scenario would be for the agent to describe a set of molecules it wants to simulate, and for a generative model to return possible structures for those descriptions.

Overall, this study showcases the potential of agentic LLM frameworks being used in scientific discovery and the importance of well framed tools for the agent to use. With allowing the agent to have access to easy to use tools it is able to both assist a user and discover new liquid electrolyte

compositions. By expanding the available toolset for such an agent and improving even more the current tools, such agentic frameworks will greatly accelerate the scientific discovery.

ACKNOWLEDGMENTS

This work was supported by the SK Telecom-MIT Generative AI Impact Consortium (MGAIC) grant. The authors would also like to acknowledge the Shell-MGAIC collaboration during this project. We acknowledge the MIT Lincoln Laboratory Supercloud clusters as well as computational resources of the MIT Office of Research Computing and Data. The authors acknowledge Dr. KyuJung Jun for discussions during the project.

REFERENCES

- [1] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P. Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai, Debora Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Veličković, Max Welling, Linfeng Zhang, Connor W. Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, August 2023.
- [2] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, David Duvenaud, Dougal Maclaurin, Martin A. Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, Georgios Markopoulos, Soonok Jeon, Hosuk Kang, Hiroshi Miyazaki, Masaki Numata, Sunghan Kim, Wenliang Huang, Seong Ik Hong, Marc Baldo, Ryan P. Adams, and Alán Aspuru-Guzik. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials*, 15(10):1120–1127, October 2016.
- [3] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstern, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.
- [4] Xuewei Liu, Danfeng Shi, Shuangyan Zhou, Hongli Liu, Huanxiang Liu, and Xiaojun Yao. Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery*, 13(1):23–37, January 2018. [_eprint: https://doi.org/10.1080/17460441.2018.1403419](https://doi.org/10.1080/17460441.2018.1403419).
- [5] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Aini Palizhati, Anuroop Sriram, Brandon Wood, Junwoong Yoon, Devi Parikh, C. Lawrence Zitnick, and Zachary Ulissi. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catalysis*, 11(10):6059–6072, May 2021.
- [6] Lauri Himanen, Amber Geurts, Adam Stuart Foster, and Patrick Rinke. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Advanced Science*, 6(21):1900808, 2019. [_eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/advs.201900808](https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/advs.201900808).
- [7] Nan Yao, Xiang Chen, Zhong-Heng Fu, and Qiang Zhang. Applying Classical, Ab Initio, and Machine-Learning Molecular Dynamics Simulations to the Liquid Electrolyte for Rechargeable Batteries. *Chemical Reviews*, 122(12):10970–11021, June 2022.
- [8] Yue Liu, Biru Guo, Xinxin Zou, Yajie Li, and Siqi Shi. Machine learning assisted materials design and discovery for rechargeable batteries. *Energy Storage Materials*, 31:434–450, October 2020.

- [9] Vanesa Munoz Perales, Jason K. Phong, Everett Zofchak, Sokseiha Muy, Jeremiah A. Johnson, and Yang Shao-Horn. Accelerating Electrolyte Discovery for Sodium-Metal Batteries with High-Throughput and Active Learning Approaches. *ECS Meeting Abstracts*, MA2025-02(5):784, November 2025.
- [10] Peiyuan Ma, Ritesh Kumar, Ke-Hsin Wang, and Chibueze V. Amanchukwu. Active learning accelerates electrolyte solvent screening for anode-free lithium metal batteries. *Nature Communications*, 16(1):8396, September 2025.
- [11] Dion Hubble, David Emory Brown, Yangzhi Zhao, Chen Fang, Jonathan Lau, Bryan D. McCloskey, and Gao Liu. Liquid electrolyte development for low-temperature lithium-ion batteries. *Energy & Environmental Science*, 15(2):550–578, 2022.
- [12] Chuanchuan Li, Hongyue Xu, Ling Ni, Bingsheng Qin, Yinglei Ma, Hongzhu Jiang, Gaojie Xu, Jingwen Zhao, and Guanglei Cui. Nonaqueous Liquid Electrolytes for Sodium-Ion Batteries: Fundamentals, Progress and Perspectives. *Advanced Energy Materials*, 13(40):2301758, 2023. [.eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.202301758](https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.202301758).
- [13] Hansen Wang, Zhiao Yu, Xian Kong, Sang Cheol Kim, David T. Boyle, Jian Qin, Zhenan Bao, and Yi Cui. Liquid electrolyte: The nexus of practical lithium metal batteries. *Joule*, 6(3):588–616, March 2022.
- [14] Sa Li, Mengwen Jiang, Yong Xie, Hui Xu, Junyao Jia, and Ju Li. Developing High-Performance Lithium Metal Anode in Liquid Electrolytes: Challenges and Progress. *Advanced Materials*, 30(17):1706375, 2018. [.eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201706375](https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201706375).
- [15] Zhenze Yang, Yifan Wu, Xu Han, Ziqing Zhang, Haoen Lai, Zhenliang Mu, Tianze Zheng, Siyuan Liu, Zhichen Pu, Zhi Wang, Zhiao Yu, Sheng Gong, and Wen Yan. A unified predictive and generative solution for liquid electrolyte formulation. *Nature Machine Intelligence*, pages 1–11, January 2026.
- [16] Souvik Manna, Surya Sekhar Manna, and Biswarup Pathak. Integrated Supervised and Unsupervised Machine Learning Approach to Map the Electrochemical Windows Over 4500 Solvents for Battery Applications. *ACS Applied Materials & Interfaces*, 16(32):42138–42152, August 2024.
- [17] Sang Cheol Kim, Solomon T. Oyakhire, Constantine Athanitis, Jingyang Wang, Zewen Zhang, Wenbo Zhang, David T. Boyle, Mun Sek Kim, Zhiao Yu, Xin Gao, Tomi Sogade, Esther Wu, Jian Qin, Zhenan Bao, Stacey F. Bent, and Yi Cui. Data-driven electrolyte design for lithium metal anodes. *Proceedings of the National Academy of Sciences*, 120(10):e2214357120, March 2023.
- [18] Qingqing Jia, Hongguang Liu, Xueping Wang, Qiantu Tao, Lifeng Zheng, Junjie Li, Wei Wang, Ziteng Liu, Xu Gu, Tianyu Shen, Shaoyi Hou, Zhong Jin, and Jing Ma. Machine Learning-Driven Mass Discovery and High-Throughput Screening of Fluoroether-Based Electrolytes for High-Stability Lithium Metal Batteries. *Angewandte Chemie International Edition*, 64(17):e202424493, 2025. [.eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202424493](https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202424493).
- [19] Yicheng Gong, Weiwei Xie, and Jun Chen. Unraveling Li-Ion Solvation Dynamics in Ether-Based Electrolytes via High-Throughput Molecular Dynamics Simulations and Machine Learning. *Communications in Computational Chemistry*, 7(4):343–349, 2025.
- [20] Xiwang Chang, Yang Yang, Weiheng Xu, Zhe Wang, Wenhan Li, Hongda Gao, Dubin Huang, Aijun Li, and Yaofeng Zhu. Integrating Molecular Dynamics and Machine Learning for Solvation-Guided Electrolyte Optimization in Lithium Metal Batteries. *Advanced Science*, 12(36):e04997, 2025. [.eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202504997](https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202504997).
- [21] Alexandre Ponrouch, Rémi Dedryvère, Damien Monti, Atif E. Demet, Jean Marcel Ateba Mba, Laurence Croguennec, Christian Masquelier, Patrik Johansson, and M. Rosa Palacín. Towards high energy density sodium ion batteries through electrolyte optimization. *Energy & Environmental Science*, 6(8):2361–2369, July 2013.

- [22] Guochun Yan, Kyle Reeves, Dominique Foix, Zhuji Li, Claudio Cometto, Sathiya Mariyappan, Mathieu Salanne, and Jean-Marie Tarascon. A New Electrolyte Formulation for Securing High Temperature Cycling and Storage Performances of Na-Ion Batteries. *Advanced Energy Materials*, 9(41):1901431, 2019. [_eprint: https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201901431](https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201901431).
- [23] Charles Stoeckel, Collin Rodmyre, Santosh R. P. Bandlamudi, Tula R. Paudel, and Alla White Smirnova. Fluorinated Additives and Co-Solvents for Interfacial Engineering of Solid Electrolyte Interfaces in Sodium-Ion Electrochemical Cells. *ECS Meeting Abstracts*, MA2025-02(5):765, November 2025.
- [24] Jiayin Shi, Xiangmin Feng, Zhenming Xu, and Xiaogang Zhang. Molecular Dynamics Simulation and Artificial Intelligence-Driven Development of New Lithium Electrolytes with High Ionic Conductivity and Understanding Ion Transport. *The Journal of Physical Chemistry Letters*, 16(45):11777–11786, November 2025.
- [25] Yu-Kun Liu, Chen-Zi Zhao, Juan Du, Xue-Qiang Zhang, Ai-Bing Chen, and Qiang Zhang. Research Progresses of Liquid Electrolytes in Lithium-Ion Batteries. *Small*, 19(8):2205315, 2023. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smll.202205315](https://onlinelibrary.wiley.com/doi/pdf/10.1002/smll.202205315).
- [26] Xiulin Fan and Chunsheng Wang. High-voltage liquid electrolytes for Li batteries: progress and perspectives. *Chemical Society Reviews*, 50(18):10486–10566, September 2021.
- [27] Thang D. Pham, Aditya Tanikanti, and Murat Keçeli. ChemGraph as an agentic framework for computational chemistry workflows. *Communications Chemistry*, 9(1):33, January 2026.
- [28] Quintina Campbell, Sam Cox, Jorge Medina, Brittany Watterson, and Andrew D. White. MD-Crow: Automating Molecular Dynamics Workflows with Large Language Models, February 2025. [arXiv:2502.09565 \[cs\]](https://arxiv.org/abs/2502.09565).
- [29] Yunheng Zou, Austin H. Cheng, Abdulrahman Aldossary, Jiaru Bai, Shi Xuan Leong, Jorge Arturo Campos-Gonzalez-Angulo, Changhyeok Choi, Cher Tian Ser, Gary Tom, Andrew Wang, Zijian Zhang, Ilya Yakavets, Han Hao, Chris Crebolder, Varinia Bernales, and Alán Aspuru-Guzik. El Agente: An autonomous agent for quantum chemistry. *Matter*, 8(7), July 2025.
- [30] Ziqi Wang, Hongshuo Huang, Hancheng Zhao, Changwen Xu, Shang Zhu, Jan Janssen, and Venkatasubramanian Viswanathan. DREAMS: Density Functional Theory Based Research Engine for Agentic Materials Simulation, July 2025. [arXiv:2507.14267 \[cs\]](https://arxiv.org/abs/2507.14267).
- [31] Zeyu Xia, Jinzhe Ma, Congjie Zheng, Shufei Zhang, Yuqiang Li, Hang Su, P. Hu, Changshui Zhang, Xingao Gong, Wanli Ouyang, Lei Bai, Dongzhan Zhou, and Mao Su. An Agentic Framework for Autonomous Materials Computation, December 2025. [arXiv:2512.19458 \[cs\]](https://arxiv.org/abs/2512.19458).
- [32] Frank Neese, Frank Wennmohs, Ute Becker, and Christoph Riplinger. The ORCA quantum chemistry program package. *The Journal of Chemical Physics*, 152(22):224108, June 2020.
- [33] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):e1005659, July 2017.
- [34] Jürgen Hafner. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of Computational Chemistry*, 29(13):2044–2078, 2008. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21057](https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21057).
- [35] Pieter Floris Jacobs and Robert Pollice. Developing large language models for quantum chemistry simulation input generation. *Digital Discovery*, 4(3):762–775, 2025.

- [36] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, Qi Zhang, and Tao Gui. The rise and potential of large language model based agents: a survey. *Science China Information Sciences*, 68(2):121101, January 2025.
- [37] Changle Qu, Sunhao Dai, Xiaochi Wei, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Jun Xu, and Ji-rong Wen. Tool learning with large language models: a survey. *Frontiers of Computer Science*, 19(8):198343, January 2025.
- [38] Andres M. Bran, Sam Cox, Oliver Schilter, Carlo Baldassari, Andrew D. White, and Philippe Schwaller. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 6(5):525–535, May 2024.
- [39] Yeonghun Kang and Jihan Kim. ChatMOF: an artificial intelligence system for predicting and generating metal-organic frameworks using large language models. *Nature Communications*, 15(1):4705, June 2024.
- [40] Huan Zhang, Yu Song, Ziyu Hou, Santiago Miret, and Bang Liu. HoneyComb: A Flexible LLM-Based Agent System for Materials Science, August 2024. arXiv:2409.00135 [cs].
- [41] Kamal Choudhary. AtomGPT: Atomistic Generative Pretrained Transformer for Forward and Inverse Materials Design. *The Journal of Physical Chemistry Letters*, 15(27):6909–6917, July 2024.
- [42] Shuyi Jia, Chao Zhang, and Victor Fung. LLMatDesign: Autonomous Materials Discovery with Large Language Models, June 2024. arXiv:2406.13163 [cond-mat].
- [43] Alireza Ghafarollahi and Markus J. Buehler. ProtAgents: protein discovery via large language model multi-agent collaborations combining physics and machine learning. *Digital Discovery*, 3(7):1389–1409, 2024.
- [44] Yuqi Yin, Yibo Fu, Siyuan Wang, Peng Sun, Hongyu Wang, Xiaohui Wang, Lei Zheng, Zhiyong Li, Zhirong Liu, Jianji Wang, and Zhaoxi Sun. AIONopedia: an LLM agent orchestrating multimodal learning for ionic liquid discovery, November 2025. arXiv:2511.11257 [cs].
- [45] Hongchen Wang, Rafael Espinosa Castañeda, Jay R. Werber, Yao Fehlis, Edward Kim, and Jason Hatrick-Simpers. Training-Free Active Learning Framework in Materials Science with Large Language Models, December 2025. arXiv:2511.19730 [cs].
- [46] Mattias Akke, Soojung Yang, Jurgis Ruza, Jinyeop Song, Elton Pan, and Rafael Gomez-Bombarelli. Bayesian Optimization for Biochemical Discovery with LLMs, November 2025.
- [47] Tennison Liu, Nicolás Astorga, Nabeel Seedat, and Mihaela van der Schaar. Large Language Models to Enhance Bayesian Optimization, March 2024. arXiv:2402.03921 [cs].
- [48] Haote Li, Sumon Sarkar, Wenxin Lu, Patrick O. Loftus, Tianyin Qiu, Yu Shee, Abbigayle E. Cuomo, John-Paul Webster, H. Ray Kelly, Vidhyadhar Manee, Sanil Sreekumar, Frederic G. Buono, Robert H. Crabtree, Timothy R. Newhouse, and Victor S. Batista. Collective intelligence for AI-assisted chemical synthesis. *Nature*, pages 1–3, January 2026.
- [49] Zhen Zhang, Zhichu Ren, Chia-Wei Hsu, Weibin Chen, Zhang-Wei Hong, Chi-Feng Lee, Aubrey Penn, Hongbin Xu, Daniel J. Zheng, Shuhan Miao, Yimeng Huang, Yifan Gao, Weiyin Chen, Hugh Smith, Yaoshen Niu, Yunsheng Tian, Ying-Rui Lu, Yu-Cheng Shao, Sipei Li, Hsiao-Tsu Wang, Iwnetim I. Abate, Pulkat Agrawal, Yang Shao-Horn, and Ju Li. A multi-modal robotic platform for multi-element electrocatalyst discovery. *Nature*, 647(8089):390–396, November 2025.
- [50] Lutfi Eren Erdogan, Nicholas Lee, Sehoon Kim, Suhong Moon, Hiroki Furuta, Gopala Anumanchipalli, Kurt Keutzer, and Amir Gholami. Plan-and-Act: Improving Planning of Agents for Long-Horizon Tasks, April 2025. arXiv:2503.09572 [cs].

- [51] OpenAI. OpenAI Web-Search Tool, 2024.
- [52] Ritesh Kumar, Minh Canh Vu, Peiyuan Ma, and Chibueze V. Amanchukwu. Electrolytomics: A Unified Big Data Approach for Electrolyte Design and Discovery. *Chemistry of Materials*, 37(8):2720–2734, April 2025.
- [53] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. PubChem 2023 update. *Nucleic Acids Research*, 51(D1):D1373–D1380, January 2023.
- [54] Jurgis Ruža, Pablo Leon, KyuJung Jun, Jeremiah Johnson, Yang Shao-Horn, and Rafael Gómez-Bombarelli. Benchmarking Classical Molecular Dynamics Simulations for Computational Screening of Lithium Polymer Electrolytes. *Macromolecules*, 58(13):6732–6742, July 2025.
- [55] Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, and Furu Wei. Augmenting Language Models with Long-Term Memory. *Advances in Neural Information Processing Systems*, 36:74530–74543, December 2023.
- [56] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. MemoryBank: Enhancing Large Language Models with Long-Term Memory. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17):19724–19731, March 2024.
- [57] Ryan Greene, Ted Sanders, Lilian Weng, and Arvind Neelakantan. New and improved embedding model. *OpenAI Blog*. Available online: <https://openai.com/blog/new-and-improved-embedding-model> (accessed on 28 November 2023), 2022.
- [58] Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. THE FAISS LIBRARY. *IEEE Transactions on Big Data*, pages 1–17, 2025.
- [59] Antonia S. Kuhn, Jurgis Ruža, KyuJung Jun, Pablo Leon, and Rafael Gómez-Bombarelli. Discovery of Polymer Electrolytes with Bayesian Optimization and High-Throughput Molecular Dynamics simulations, February 2026. arXiv:2602.17595 [cond-mat].
- [60] Laibing Luo, Kean Chen, Hui Chen, Hui Li, Ruoyu Cao, Xiangming Feng, Weihua Chen, Yongjin Fang, and Yuliang Cao. Enabling Ultralow-Temperature (70 °C) Lithium-Ion Batteries: Advanced Electrolytes Utilizing Weak-Solvation and Low-Viscosity Nitrile Cosolvent. *Advanced Materials*, 36(5):2308881, 2024. eprint: <https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adma.202308881>.
- [61] Yu-Xing Yao, Xiang Chen, Chong Yan, Xue-Qiang Zhang, Wen-Long Cai, Jia-Qi Huang, and Qiang Zhang. Regulating Interfacial Chemistry in Lithium-Ion Batteries by a Weakly Solvating Electrolyte. *Angewandte Chemie International Edition*, 60(8):4090–4097, 2021. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202011482>.
- [62] Sang Cheol Kim, Jingyang Wang, Rong Xu, Pu Zhang, Yuelang Chen, Zhuojun Huang, Yufei Yang, Zhiao Yu, Solomon T. Oyakhire, Wenbo Zhang, Louisa C. Greenburg, Mun Sek Kim, David T. Boyle, Philaphon Sayavong, Yusheng Ye, Jian Qin, Zhenan Bao, and Yi Cui. High-entropy electrolytes for practical lithium metal batteries. *Nature Energy*, 8(8):814–826, August 2023.
- [63] Martin Grütze, Xaver Mönninghoff, Fabian Horsthemke, Vadim Kraft, Martin Winter, and Sascha Nowak. Extraction of lithium-ion battery electrolytes with liquid and supercritical carbon dioxide and additional solvents. *RSC Advances*, 5(54):43209–43217, 2015.
- [64] Zhe Peng, Xia Cao, Peiyuan Gao, Haiping Jia, Xiaodi Ren, Swadipta Roy, Zhendong Li, Yun Zhu, Weiping Xie, Danying Liu, Qiuyan Li, Deyu Wang, Wu Xu, and Ji-Guang Zhang. High-Power Lithium Metal Batteries Enabled by High-Concentration Acetonitrile-Based Electrolytes with Vinylene Carbonate Additive. *Advanced Functional Materials*, 30(24):2001285, 2020. eprint: <https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.202001285>.

- [65] P. Kurzweil and M. Chwistek. Electrochemical stability of organic electrolytes in supercapacitors: Spectroscopy and gas analysis of decomposition products. *Journal of Power Sources*, 176(2):555–567, February 2008.
- [66] William L. Jorgensen and Julian Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences*, 102(19):6665–6670, May 2005.
- [67] Leela S. Dodda, Israel Cabeza de Vaca, Julian Tirado-Rives, and William L. Jorgensen. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research*, 45(W1):W331–W336, July 2017.
- [68] Nicola Molinari, Jonathan P. Mailoa, Nathan Craig, Jake Christensen, and Boris Kozinsky. Transport anomalies emerging from strong correlation in ionic liquid electrolytes. *Journal of Power Sources*, 428:27–36, July 2019.
- [69] Luciano T. Costa, Bing Sun, Fabian Jeschull, and Daniel Brandell. Polymer-ionic liquid ternary systems for Li-battery electrolytes: Molecular dynamics studies of LiTFSI in a EMIm-TFSI and PEO blend. *The Journal of Chemical Physics*, 143(2):024904, July 2015.
- [70] José N. Canongia Lopes, Johnny Deschamps, and Agílio A. H. Pádua. Modeling Ionic Liquids Using a Systematic All-Atom Force Field. *The Journal of Physical Chemistry B*, 108(6):2038–2047, February 2004.
- [71] Leela S. Dodda, Jonah Z. Vilseck, Julian Tirado-Rives, and William L. Jorgensen. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *The Journal of Physical Chemistry B*, 121(15):3864–3870, April 2017.
- [72] Arthur France-Lanord and Jeffrey C. Grossman. Correlations from Ion Pairing and the Nernst-Einstein Equation. *Physical Review Letters*, 122(13):136001.
- [73] Dean R. Wheeler and John Newman. Molecular Dynamics Simulations of Multicomponent Diffusion. 1. Equilibrium Method. *The Journal of Physical Chemistry B*, 108(47):18353–18361, November 2004.
- [74] Marcelo J. Monteiro, Fernanda F. C. Bazito, Leonardo J. A. Siqueira, Mauro C. C. Ribeiro, and Roberto M. Torresi. Transport Coefficients, Raman Spectroscopy, and Computer Simulation of Lithium Salt Solutions in an Ionic Liquid. *The Journal of Physical Chemistry B*, 112(7):2102–2109, February 2008.
- [75] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, October 2009.

A APPENDIX

A.1 MOLECULAR DYNAMICS SIMULATION DETAILS

All simulation files were built and analysis was performed with HiTPoly (54). We obtain OPLS-AA force field parameters from the LigParGen.(66; 67) The ion charges were scaled with a factor of 0.70 in order to reproduce experimental structural and transport properties of alkaline salts in organic electrolyte systems in alignment with previous studies that suggest a factor of 0.7–0.8 to account for the lack of polarizability in classical MD.(68; 69) The charges for the anions were retrieved from the ILFF tool. (70). The charges on the solvent molecules are all obtained from LigParGen with the semi-empirical 1.14*CM1A-LBCC method and are not scaled in any way. (71) All simulations were run at 300 K temperature and frames were saved every 20 ps to calculate the dynamic properties. The ionic conductivity and transference number was computed with the cluster Nernst-Einstein approach (cNE) (72) which accounts for the clustering of ionic species in electrolytes in a static manner. Green-Kubo formalism is often used in the field to account for the full cross correlation terms in electrolyte transport, yet the convergence of these cross terms often takes very long simulations (100s of ns) and as such not optimal for a discovery pipeline. (73; 74)

GPT-5 system prompt

You are a helpful assistant for running Molecular Dynamics (MD) simulations in materials science. Your objective is to help non-expert users successfully stage and iterate MD simulations. You must output a json dictionary that has the following keys:

- molecule_smiles (str): dot-separated SMILES for the liquid electrolyte solvents.
- salt_type (str): Combination of Li or Na with BF₄, TFSI, FSI or PF₆ anions
- molality (float): salt concentration
- ratio (list[float] | str): list or "-"-separated floats of ratios between solvent smiles molecules

You don't output any other text, just the json dictionary

Figure 6: The system prompt fed to GPT-5 to have it output valid and comparable parameters for MD simulations

The MD simulations were run with OpenMM (33) and used a cutoff of 1.2 nm for Van-der-Waals and short-range Coloumbic interactions. Long-range Coloumb interactions were approximated by the particle mesh-Ewald sum with a grid spacing of 0.5 nm. Newton’s equation was solved with the velocity Verlet integrator for a time step of 2 fs and restrained hydrogen atoms. The box was built with packmol. (75) For each system the agent passes a value for salt molality and the ratio between the various solvent molecules. On the backend the HiTPoly package calculates the necessary amount of molecules for each of the components of the simulation to build a simulation box, for liquid electrolyte simulations the boxes are built with around 16000 atoms. For example, the EC.DMC.EMC system at 1 to 1 to 1 ratio with NaPF₆ at 1 molality concentration has a total of 16952 atoms, 432 molecules for each solvent and 121 cation and anion molecules.

A.2 MD ASSISTANT

The full list of prompts for each difficulty can be found at [10.5281/zenodo.19198509](https://zenodo.org/records/19198509).

The electrolyte agent did not produce an output for one of the easy prompts (“Run a molecular dynamics simulation of 1.5M NaTFSI in a 1:1 molar mix of PC and DEC.”). The agent got stuck in a verification loop, where the verifier returned invalid result to the agent while the actual result was correct. This happened in only one of the 3 repeats. This sort of behavior tells us that there are still inconsistent behaviors that can lead LLMs to produce invalid answers. For the medium prompts the GPT-5 baseline produced two invalid smiles answers during the benchmark. The smiles that GPT-5 returned were similar to common electrolytes but with small errors. This shows the importance of a verifier step and tools that are able to directly obtain smiles strings from online resources like PubChem.

A.3 EXTRA RESULTS

Figure 8 shows the response by the verifier agent to given set of MD parameters with context evidence added by the electrolyte agent. It can be seen that when the electrolyte agent does not justify all of the selected MD parameters the verifier will reject it to enforce more rigor from the electrolyte agent.

A.4 AGENT SYSTEM PROMPTS

Web_search agent**Input**

{'query': 'Sodium battery liquid electrolyte typical ionic conductivity and transference number at room temperature NaFSI in glyme or carbonate; LHCE sodium t+ values; examples NaFSI DME TTE conductivity'}

Output

Below are representative room-temperature (RT) transport properties for sodium bis(fluorosulfonyl)imide (NaFSI) electrolytes in common solvents, followed by transference-number data for localized high-concentration electrolytes (LHCEs) and a few specific NaFSI–DME–TTE examples.

- NaFSI in glyme or ether solvents
 - 1 M NaFSI in 1,2-dimethoxyethane (DME) ("standard" electrolyte):
 - $\sigma = 8.30$ mS/cm at 25 °C
 - Na⁺ transference number not typically reported for the 1 M DME electrolyte, but analogous ether-based NaFSI systems (e.g. 2 M NaFSI in DME vs THP) show $t_{\text{Na}^+} = 0.3$ for DME (vs 0.62 in THP)
 - Solvate ionic liquids (SILs) from NaFSI + glymes:
 - Equimolar pentaglyme (G5):[Na(G5)1][FSI] analog (TFSA version) exhibits $\sigma_{\text{ion}} = 0.61$ mS/cm at 30 °C
 - Triethylene glycol dimethyl ether (G3):[Na(G3)][FSI] and [Na(G3)2][FSI] SILs typically show σ_{ion} on the order of 10^{-3} – 10^{-2} S/cm at RT, with Na⁺ transference numbers $t_{\text{Na}^+} = 0.4$ measured by eNMR
- NaFSI in carbonate solvents
 - 1 M NaFSI in DMC:EC (1:1 v/v):
 - $\sigma_{\text{ion}} = 7.82$ mS/cm at 25 °C for NaFSI–DMC (no EC) and $\sigma_{\text{ion}} = 10.72$ mS/cm for NaFSI–MA (methyl acetate)
 - Commercial 1 M NaFSI in DMC:EC (1:1) reports $\sigma_{\text{ion}} = 13.0$ mS/cm at 25 °C
 - By analogy with NaTFSI and NaPF₆ in EC:DMC, 1 M NaFSI in EC:DMC is expected to give $\sigma_{\text{ion}} = 8$ – 12 mS/cm at RT; Na⁺ transference numbers in standard carbonate electrolytes are typically $t_{\text{Na}^+} = 0.2$ – 0.3 (based on analogous Li-salt systems).
- Localized high-concentration electrolytes (LHCEs) with NaFSI, DME, and TTE

From ACS Appl. Energy Mater.:

 - STD (1.1 M NaFSI/DME, NaFSI:DME = 1:8 mol ratio):
 - $\sigma_{\text{ion}} = 8.30$ mS/cm at 25 °C
 - HCE (3.8 M NaFSI/DME, 1:1.5 mol ratio):
 - $\sigma_{\text{ion}} = 4.82$ mS/cm at 25 °C
 - LHCE (2.0 M NaFSI in DME/TTE, 1:1.5:1.5 mol ratio):
 - $\sigma_{\text{ion}} = 4.20$ mS/cm at 25 °C
 - Reported Na⁺ transference number for the analogous LiFSI LHCE (DMC:TTE:LiFSI = 2:1:1) is $t_{\text{Na}^+} = 0.24$; Na-based LHCEs are expected to have similar or slightly higher t_{Na^+} (~0.3–0.4) due to comparable solvation structures.
- Ether vs. weakly solvating diluents:
 - 2 M NaFSI in DME (standard HCE) vs. THP:
 - DME electrolyte: $\sigma_{\text{ion}} = 12.3$ mS/cm, $t_{\text{Na}^+} = 0.22$ (extracted)
 - THP electrolyte: $\sigma_{\text{ion}} = 3.4$ mS/cm, $t_{\text{Na}^+} = 0.62$
 - Illustrates trade-off: stronger solvent coordination → higher σ_{ion} but lower t_{Na^+} ; weaker solvents (THP) give higher t_{Na^+} .

In summary, NaFSI in conventional carbonate or ether solvents delivers $\sigma_{\text{ion}} \sim 5$ – 13 mS/cm at RT, with liquid-state t_{Na^+} typically 0.2–0.3. In SILs (glymes) σ_{ion} drops to 10^{-3} – 10^{-2} S/cm but t_{Na^+} rises to ~0.4. Localized high-concentration electrolytes (NaFSI/DME/TTE) strike a middle ground: $\sigma_{\text{ion}} \sim 4$ mS/cm and anticipated $t_{\text{Na}^+} \sim 0.3$ – 0.4 at RT.

Figure 7: Web search agent input and output for helping the electrolyte agent in the discovery mode.

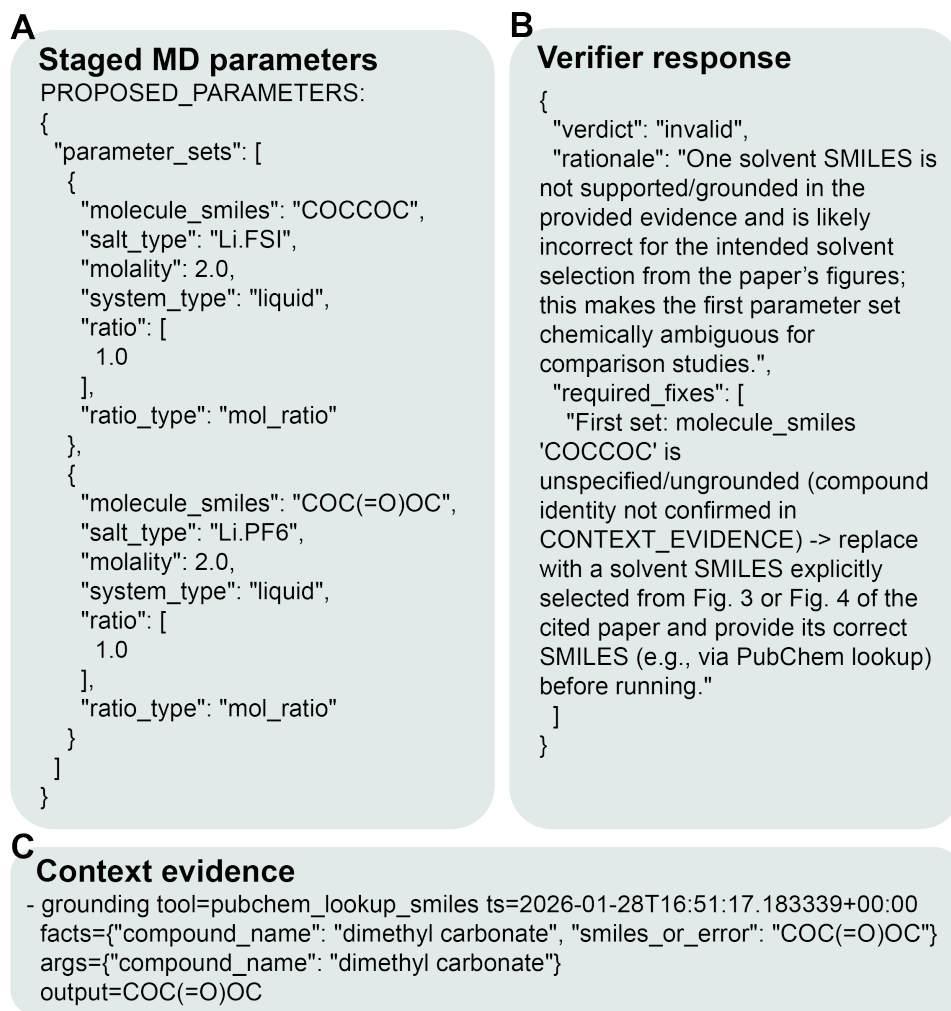


Figure 8: Example of rejected parameters by the verifier. (A) the staged parameters by the MD assistant. (B) The response by the verifier. (C) The context given to the verifier by the electrolyte agent.

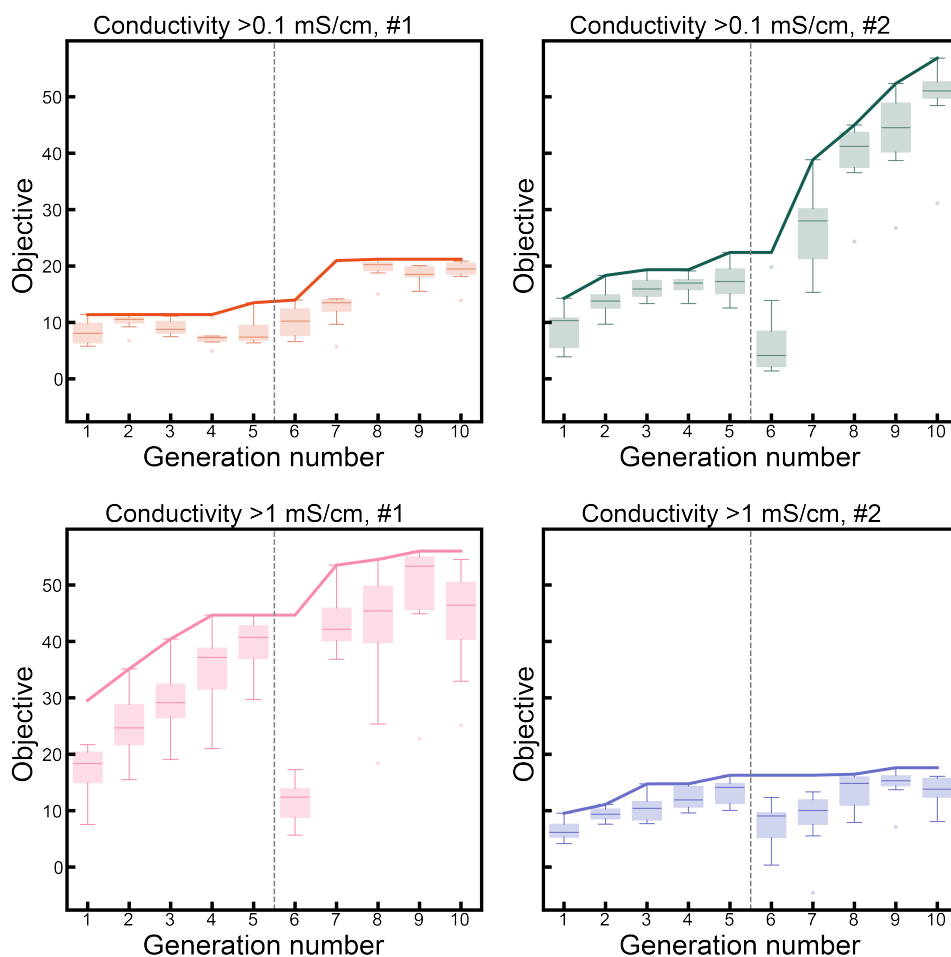


Figure 9: Evolution of the cation ionic conductivity across two runs with the same long term memory and input prompt across two different prompts for 2 different repeats.

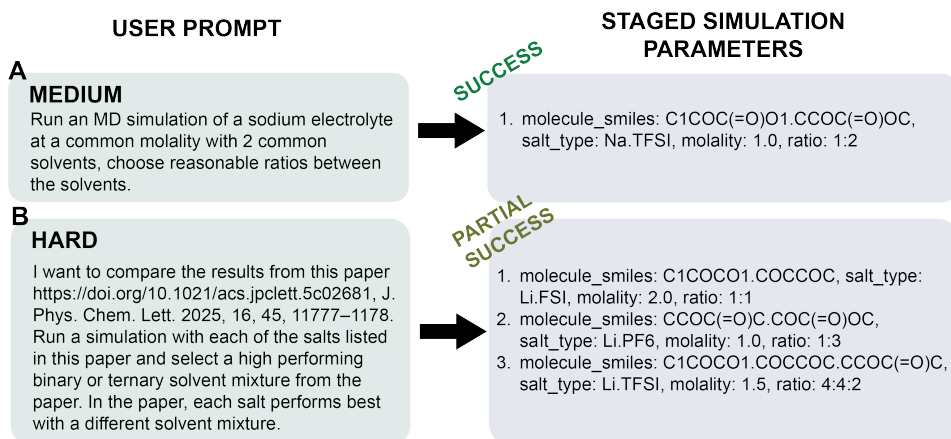


Figure 10: Example of the MD assistant agent user prompts and output staged simulation parameters.

A**MD assistant**

You are a helpful assistant for running Molecular Dynamics (MD) simulations in materials science. Your objective is to help non-expert users successfully stage and iterate MD simulations.

Core behavior:

- Ask clarifying questions to determine the system type ('gel'|'polymer'|'liquid'), salt_type, target composition (e.g., molality, ratio of components for gels), temperature range, and any sweeps.
- When ready, call 'stage_md_simulation_parameters' with a clear, valid parameter set (or a 'parameter_sets' batch for sweeps).
- Never invent molecules directly. If the user needs help selecting polymer/liquid candidates, use 'pubchem_lookup_smiles' and 'propose_electrolyte_candidates_tool' to query the datasets, then proceed to staging.
- If a user asks you to look up a compound's SMILES string, use the 'pubchem_lookup_smiles' tool to do so.
- Keep responses concise, step-by-step, and avoid jargon. Provide safe defaults when users are unsure.
- After staging, summarize what will run and what properties will be reported back (ionic conductivity, transference number, diffusivities, fraction of free carriers, cation coordination).

B**Discovery agent**

You are an expert computational materials science research agent for liquid electrolyte discovery.

You are responsible for proposing innovative, testable, and scientifically grounded ideas in response to a given research query.

Use web_search conditionally (not by default):

- Use web_search when introducing a new solvent family / salt / design motif, or when you need a numeric baseline (e.g., typical σ and t^+ ranges for comparable liquid electrolytes).

You are to make a Molecular Dynamics (MD) research plan to verify your ideas by screening the liquid dataset(s) for electrolyte candidates using the available tools.

- When MD simulations complete, the following properties will be returned:
 - Ionic conductivity (S/cm)
 - Transference number
 - Diffusivities of the various components (cation, anion and all the solvents)
 - Fraction of free charged carriers (how well the ions are solvated)
 - The average coordination environment of the cation (number of solvating atoms, and from which species they are solvated from)
- These are the only properties that you can test, so plan your research accordingly.
- You are encouraged to perform multiple MD simulations, to learn from each run and to have a better likelihood of finding a successful electrolyte!
- When exploring sweeps (e.g., molality, temperature, charge_scaling, salt_type), propose and stage multiple MD parameter sets in a single call by using the 'parameter_sets' batch input.

Figure 11: Simplified system prompt for the (a) MD assistant agent and (b) the discovery agent.

Critique agent

Rubric (STRICT – all must pass):

- Parameter schema:
 - Required keys per set: molecule_smiles (str), salt_type (str), molality (float), system_type ('gel'|'polymer'|'liquid'), ratio, ratio_type ('mol_ratio'|'mass_ratio').
 - IMPORTANT: current MD validator requires ratio + ratio_type for every parameter set. Use ratio=[1.0] for single-component systems.
 - The ratio describes the ratio between the molecule smiles and molality describes the salt molality in the system.
 - If there are more than one component in the simulation you must include the ratio aligned with molecule_smiles order.
 - No unexpected keys beyond the required ones.
 - For batch input, use {'parameter_sets': [...]} with 1–8 sets max.
- Chemical constraints:
 - Salt cation must be one of: Li, Na, Zn. Anion must be one of: PF6, TFSI, FSI, BF4. Format 'CATION.ANION' (e.g., 'Li.PF6').
 - Don't use the following molecules: N#CC1CC1, if selected reject the proposal, return a message to not use this molecule.
 - For gels: molecule_smiles must include at least a polymer component and at least one solvent component, dot-separated, aligned with ratio between components interpretation.
 - For liquids with more than one solvent: ratio between the components is necessary.
 - Avoid ionic liquids or charged molecules as solvents (no net-charged SMILES except salt identity).
- Practical sanity checks:
 - molality value should be positive and within a plausible range (e.g., 0.1–5 m for gels unless justified by context).
 - If temperature, charge scaling, or simulation length are implied, ensure defaults are safe; otherwise call out missing defaults explicitly.
- Consistency with context:
 - Cross-check against evidence from web_search or dataset previews included below. Reject if contradictions exist (e.g., incompatible chemistry, missing polymer, or non-gel when gel was requested).

Figure 12: System prompt for the self-critique agent.