

Supporting Information

Machine Learning Molecular Dynamics Simulations of Coordination and Diffusion Behaviors in Lithiated Gallium Electrode

Qiuyi Fu,^{1‡} Hao Yuan,^{1‡} Haitang Wang,¹ Wenbin Liu,³ Guobing Zhou,^{*,1,2} Zhen Yang^{*,1,2}

¹*School of Chemical Engineering, College of Chemistry and Materials, Jiangxi Normal University, Nanchang 330022, China*

²*Key Lab of Fluorine and Silicon for Energy Materials and Chemistry of Ministry of Education, Jiangxi Normal University, Nanchang 330022, China*

³*Jiangxi Provincial Science and Technology Affairs Center, Nanchang 330046, China*

‡These authors contributed equally

Corresponding Authors

*E-mail: gbzhou@jxnu.edu.cn (G.B. Zhou)

*E-mail: yangzhen@jxnu.edu.cn (Z. Yang)

S1. Details on simulation cell and training dataset

Table S1. The cell sizes and training datasets for four Li-Ga alloy (LGA) systems in this work.

System	No. of Li	No. of Ga	Cell sizes (Å)	Training datasets
Li₃Ga₁₄	48	224	16.94	1438
Li₂Ga₇	32	112	13.99	600
LiGa	64	64	12.39	700
Li₂Ga	144	72	14.86	546

S2. Root mean square errors of energy and force

Table S2. Root mean square error (RMSE) of the energy (meV per atom) and force (eV \AA^{-1}) for the test sets of four LGA systems.

System	Energy (meV per atom)	Force (eV \AA^{-1})
Li₃Ga₁₄	2.509	0.074
Li₂Ga₇	4.532	0.104
LiGa	1.643	0.085
Li₂Ga	1.058	0.060

S3. Analysis of training dataset diversity

To assess the structural diversity and representativeness of the MLFF training datasets, we performed an unsupervised clustering analysis based on smooth overlap of atomic positions (SOAP) descriptors. For each LGA, we randomly sampled ≥ 40000 local atomic environments from the corresponding training dataset (> 80000 local configurations) and projected their SOAP descriptors into a reduced feature space via principal component analysis (PCA). Hierarchical clustering was then applied to identify structurally distinct groups. As shown in **Fig. S1**, the PCA projections of the SOAP descriptors exhibit broad and continuous distributions for all four LGAs, indicating a wide variety of local coordination environments. The first two principal components capture ~ 41 – 67% of the total variance depending on composition ($\text{Li}_3\text{Ga}_{14}$: 65.96%, Li_2Ga_7 : 67.42%, LiGa : 40.92%, Li_2Ga : 50.24%). The coexistence of densely populated regions and widely dispersed regions in the PCA space suggests that the datasets include both frequently sampled motifs and a broad range of less common local topologies (e.g., coordination-number variations, bond-angle distortions, and different degrees of disorder). Notably, the smaller variance captured by PC1–PC2 for LiGa suggests a more complex configurational manifold, further supporting the structural richness of the dataset. This analysis demonstrates that the training datasets span a broad and representative structural space rather than being biased toward a narrow set of configurations for each LGA system. Thus, we believe that the MLFFs for the four LGAs were trained on structurally diverse datasets that adequately capture the diversity of atomic environments.

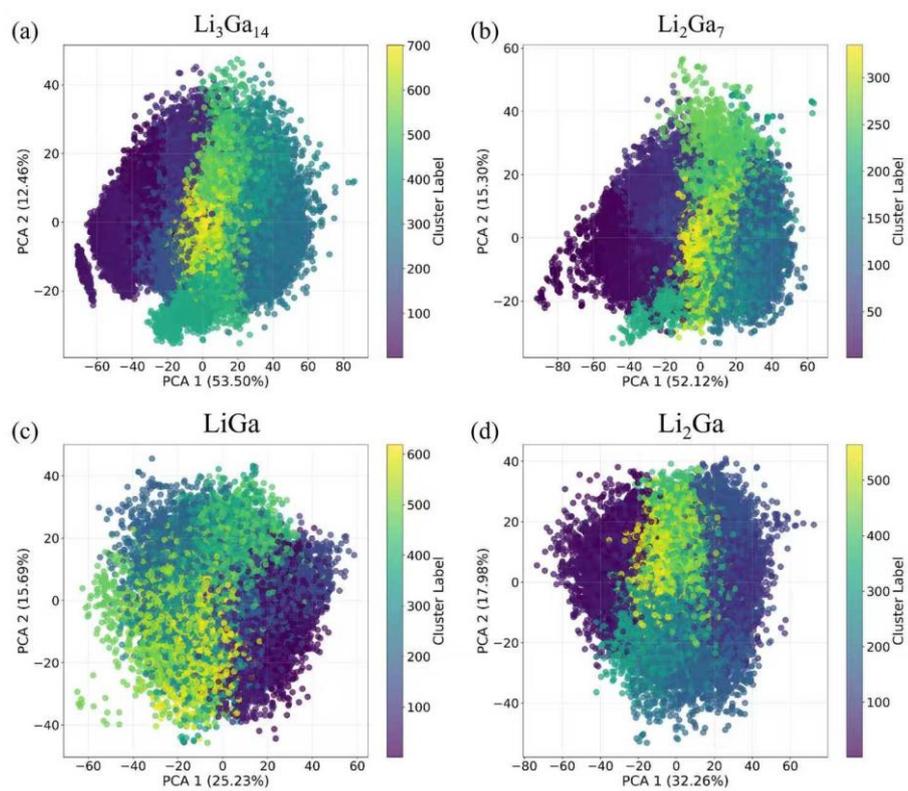


Fig. S1 PCA of local structures randomly selected from the training datasets of (a) $\text{Li}_3\text{Ga}_{14}$, (b) Li_2Ga_7 , (c) LiGa , and (d) Li_2Ga .

S4. Comparisons of RDF between DFT and MLFF

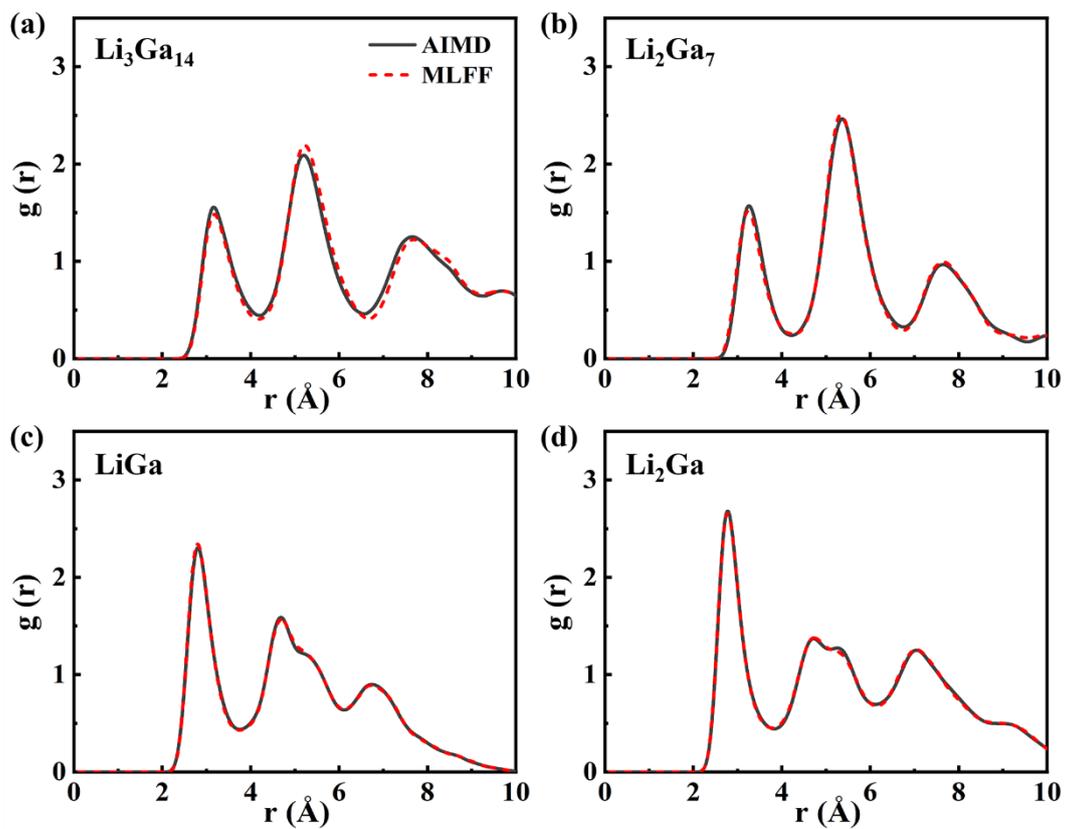


Fig. S2 Comparisons of radial distribution function from DFT and MLFF of Li-Li pair in (a) $\text{Li}_3\text{Ga}_{14}$, (b) Li_2Ga_7 , (c) LiGa , and (d) Li_2Ga systems at 300K.

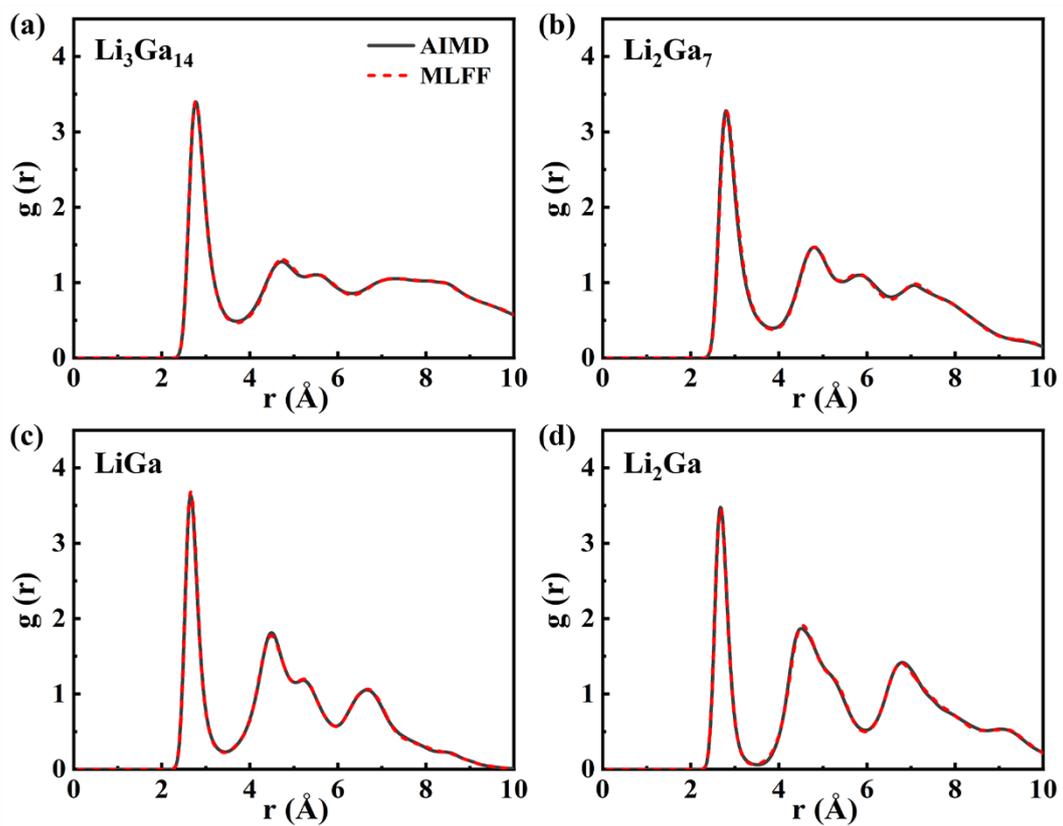


Fig. S3 Comparisons of radial distribution function from DFT and MLFF of Ga-Ga pair in (a) $\text{Li}_3\text{Ga}_{14}$, (b) Li_2Ga_7 , (c) LiGa , and (d) Li_2Ga systems at 300K.

S5. Discussions on the MLFF transferability limitations

To address the concern regarding thermodynamic transferability, we used the Li_2Ga system as a representative test case. We performed a 500 ps MD simulation at 400 K using the MLFF model originally trained on 300 K data (denoted as MLFF_300K). We then randomly selected 100 structures from the 400 K trajectory and benchmarked their energies and forces against DFT single-point calculations. As shown in **Fig. S4**, both the energy and force parity plots deviate markedly from the ideal $y=x$ line, yielding RMSE values of 0.594 eV/atom (energy) and 0.554 eV/Å (force). These results indicate that the MLFF_300K model encounters out-of-training-domain environments at 400 K and does not maintain DFT-level accuracy under the elevated-temperature condition without additional data. To remedy this, we refined the LiGa MLFF at 400 K (denoted as MLFF_400K) by running an additional 500 ps AIMD simulation with the same on-the-fly active-learning scheme described in the manuscript. During the active learning procedure, a total of 516 configurations were incorporated into the training dataset. After this augmentation, an independent validation at 400 K shows dramatically improved agreement with DFT, with RMSE values reduced to 0.001 eV/atom (energy) and 0.059 eV/Å (force) (see **Fig. S5**). These results imply that the MLFF can be systematically extended to new thermodynamic conditions through focused data enrichment

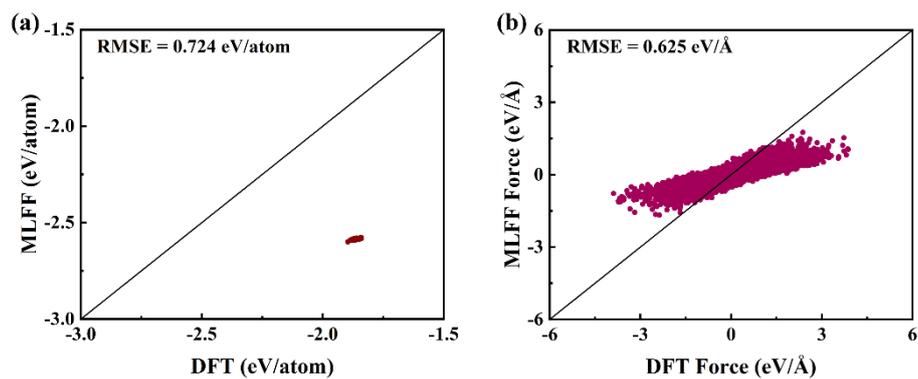


Fig. S4 Comparisons of (a) energies and (b) forces from DFT and MLFF on the test sets of Li_2Ga alloy at 400 K using the MLFF_300K model.

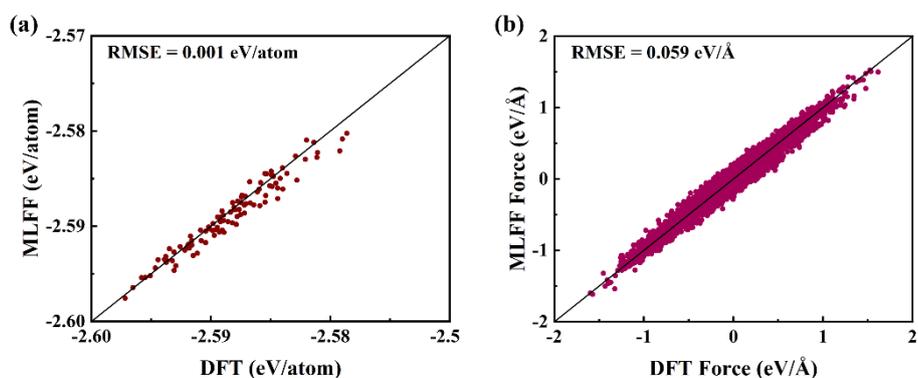


Fig. S5 Comparisons of (a) energies and (b) forces from DFT and MLFF on the test sets of Li_2Ga alloy at 400 K using the MLFF_400K model.

S6. Cutoff distances and average CNs

Table S3. Cutoff distances r_m (Å) corresponding to the first minimum of the RDFs for Li–Ga, Li–Li, and Ga–Ga pairs in four LGAs

Phase	Li–Ga	Li–Li	Ga–Ga
Li₃Ga₁₄	4.1	4.2	3.7
Li₂Ga₇	4.2	4.2	3.8
LiGa	3.7	3.8	3.4
Li₂Ga	3.7	3.8	3.5

Table S4. Average CN in the first coordination shell for Li–Ga, Li–Li, Ga–Ga and Ga–Li pairs in four LGAs, using different parts of the trajectories.

Systems	Atom pair	CN					Average CN
		Trajectory 1	Trajectory 2	Trajectory 3	Trajectory 4	Trajectory 5	
Li ₃ Ga ₁₄	Li–Ga	13.745	13.749	13.777	13.789	13.785	13.769±0.021
	Li–Li	1.624	1.608	1.593	1.598	1.603	1.605±0.012
	Ga–Ga	9.150	9.158	9.147	9.147	9.156	9.151±0.005
	Ga–Li	2.945	2.946	2.952	2.955	2.954	2.950±0.004
Li ₂ Ga ₇	Li–Ga	13.526	13.514	13.469	13.523	13.541	13.515±0.027
	Li–Li	1.783	1.784	1.821	1.814	1.786	1.798±0.019
	Ga–Ga	8.415	8.420	8.412	8.398	8.398	8.408±0.010
	Ga–Li	3.864	3.861	3.848	3.864	3.869	3.861±0.008
LiGa	Li–Ga	7.880	7.894	7.914	7.927	7.923	7.908±0.020
	Li–Li	6.303	6.295	6.288	6.258	6.260	6.281±0.021
	Ga–Ga	4.491	4.490	4.467	4.456	4.464	4.474±0.016
	Ga–Li	7.880	7.894	7.914	7.927	7.923	7.908±0.020
Li ₂ Ga	Li–Ga	5.123	5.130	5.126	5.124	5.133	5.127±0.004
	Li–Li	10.181	10.169	10.182	10.181	10.165	10.176±0.008
	Ga–Ga	2.718	2.715	2.695	2.699	2.689	2.703±0.013
	Ga–Li	10.246	10.261	10.251	10.249	10.267	10.255±0.009

S7. Probability distributions of Ga atoms with different number of Ga and Li atoms

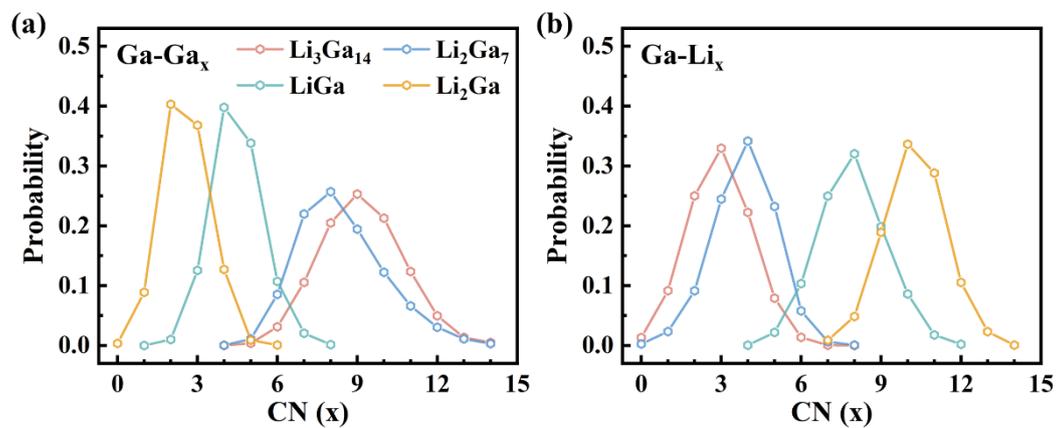


Fig. S6 Probability distribution of different coordination numbers (x) for (a) Ga-Ga_x, and (b) Ga-Li_x in four LGAs at 300K.

S8. Self-part van Hove function of Ga atoms

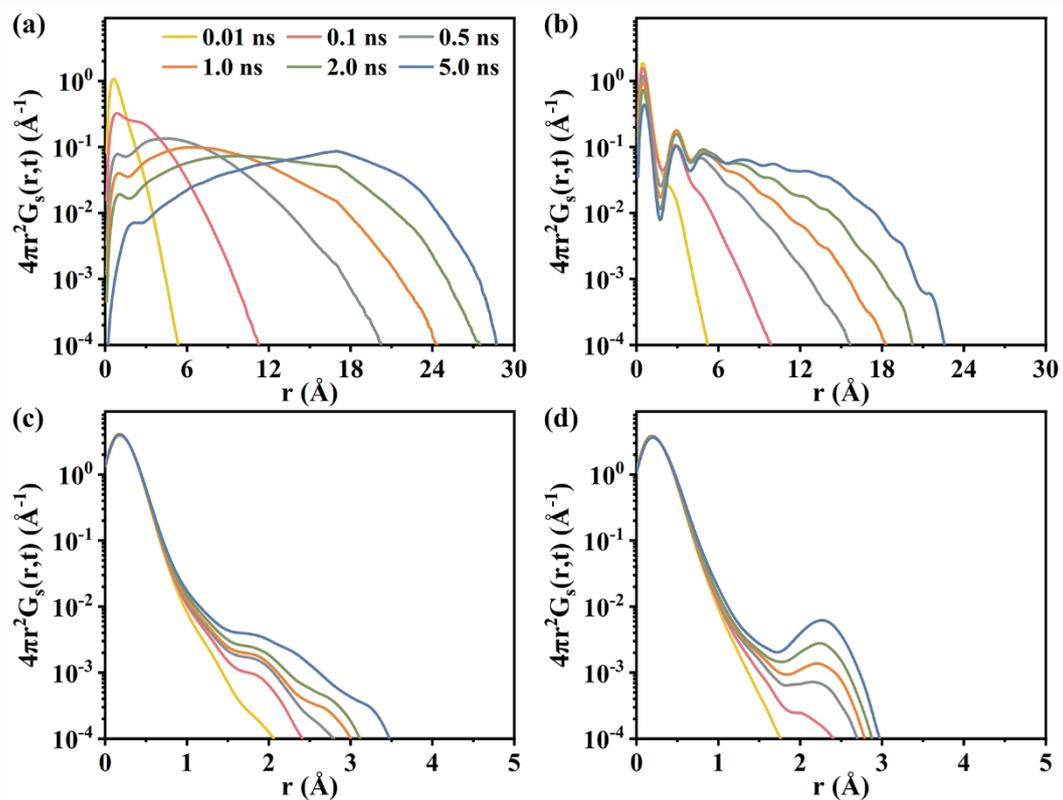


Fig. S7 Self-part $G_S(r, t)$ of the Ga atoms in the systems of (a) $\text{Li}_3\text{Ga}_{14}$, (b) Li_2Ga_7 , (c) LiGa , and (d) Li_2Ga at 300K.

S9. Atom-resolved MSD and atomic trajectories

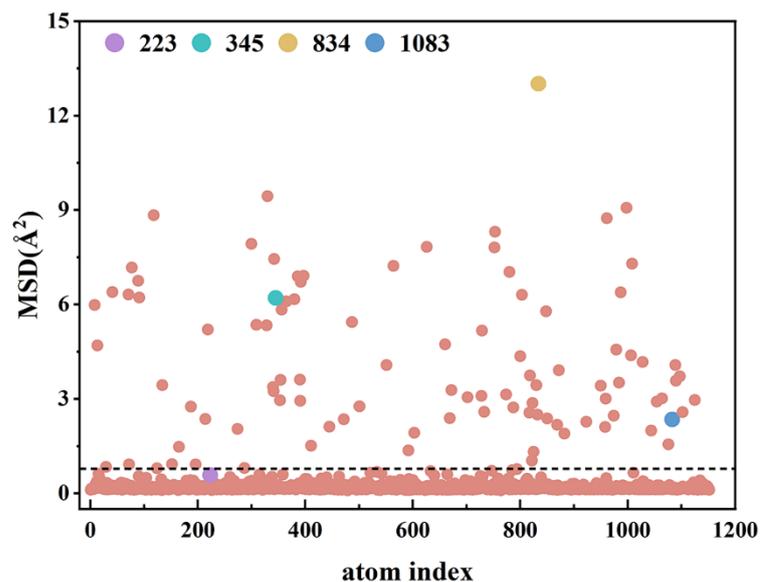


Fig. S8 Distribution of atom-resolved MSDs of Li atoms in the Li_2Ga system. Note: the horizontal black dash line denotes the ensemble-averaged MSD over all Li atoms, and the four highlighted atom indices (223, 345, 834, and 1083) indicate representative Li atoms selected for the trajectory visualizations in **Fig. 7** of the main text.

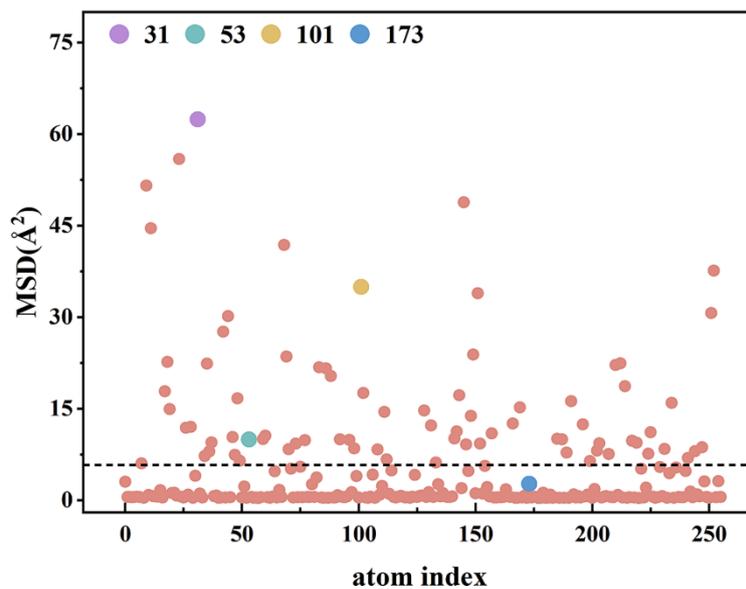


Fig. S9 Distribution of atom-resolved MSDs of Li atoms in the Li_2Ga_7 system. Note: the horizontal black dash line denotes the ensemble-averaged MSD over all Li atoms, and the four highlighted atom indices (31, 53, 101, and 173) indicate representative Li atoms selected for the trajectory visualizations in **Fig. S10**.

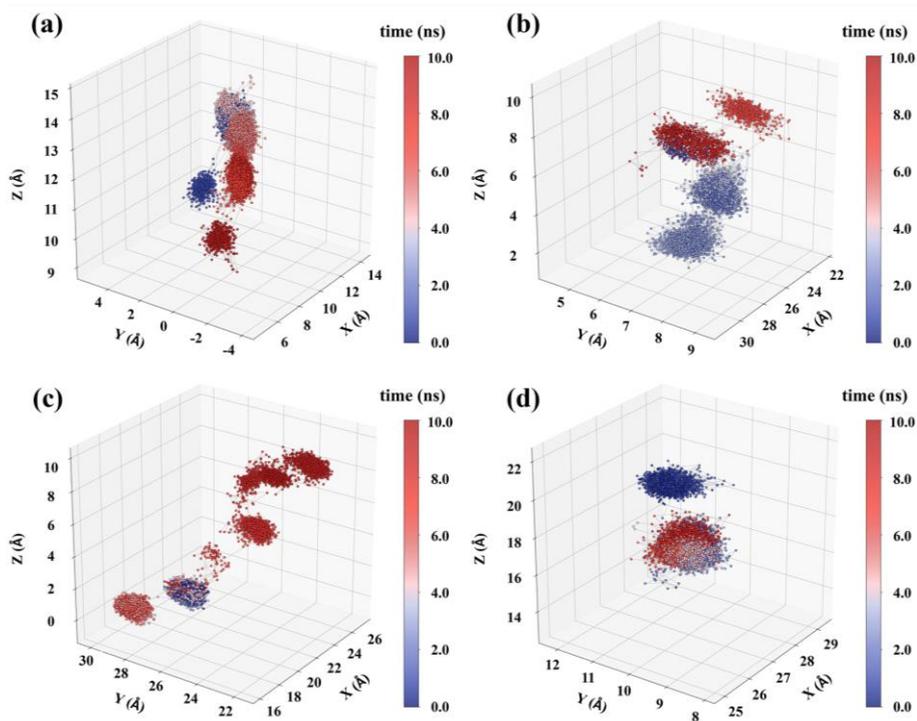


Fig. S10 3D trajectory of Li atom with indices (a) 31, (b) 53, (c) 101, and (d) 173 in the Li_2Ga_7 system. Note: the color scale represents the simulation time, and successive trajectory points are sampled every 0.5 ps.