

Database of migration barriers for Li/Na-containing compounds by high-throughput NEB workflow

Jingchen Lian(连景臣), Ruijuan Xiao(肖睿娟), and Hong Li(李泓)
Citation: Chin. Phys. B, 2026, 35 (**8**): 086104. DOI: 10.1088/1674-1056/ae88e2
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Unveiling stable and efficient antiperovskite semiconductors via high-throughput computation and interpretable machine learning

Hao Qu(瞿浩), Tao Hu(胡涛), Mingjun Li(李明军), Jiangyu Yang(杨江渝), Yunyi Zhou(周云逸), Shichang Li(李世长), Dengfeng Li(李登峰), Gang Tang(唐刚), and Chunbao Feng(冯春宝)
Chin. Phys. B, 2026, 35 (**4**): 046102. DOI: 10.1088/1674-1056/ae24e7

Database of superconductors with kagome lattice by high-throughput screening

Lihong Wang(王历宏), Qi Li(李琦), Ke Ma(马克), Yingpeng Yu(于英鹏), Shifeng Jin(金士锋), and Xiaolong Chen(陈小龙)
Chin. Phys. B, 2025, 34 (**10**): 106101. DOI: 10.1088/1674-1056/adf041

General-purpose moment tensor potential for Ga-In liquid alloys towards large-scale molecular dynamics with *ab initio* accuracy

Kai-Jie Zhao(赵凯杰) and Zhi-Gong Song(宋智功)
Chin. Phys. B, 2025, 34 (**6**): 066101. DOI: 10.1088/1674-1056/adc661

Database of ternary amorphous alloys based on machine learning

Xuhe Gong(巩旭菏), Ran Li(李然), Ruijuan Xiao(肖睿娟), Tao Zhang(张涛), and Hong Li(李泓)
Chin. Phys. B, 2025, 34 (**1**): 016101. DOI: 10.1088/1674-1056/ad8ec8

Databases of 2D material-substrate interfaces and 2D charged building blocks

Jun Deng(邓俊), Jinbo Pan(潘金波), and Shixuan Du(杜世萱)
Chin. Phys. B, 2024, 33 (**2**): 026101. DOI: 10.1088/1674-1056/ad0626

DATA PAPER

Database of migration barriers for Li/Na-containing compounds by high-throughput NEB workflow

Jingchen Lian(连景臣)[1,2], Ruijuan Xiao(肖睿娟)[1,2,†], and Hong Li(李泓)[1,3]

[1] *Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*

[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing* 100049, *China*

[3] *Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing* 100049, *China*

(Received 27 March 2026; revised manuscript received 28 May 2026; accepted manuscript online 10 July 2026)

Solid-state electrolytes (SSEs) are vital for advancing high-performance all-solid-state batteries. Since the energy barrier of SSEs is a main property that governs the ionic transport performance, it is important to construct a database of such barriers. While density functional theory (DFT) based nudged elastic band (NEB) method provides fundamental insights into lithium-ion migration barriers, the high computational costs make large-scale calculations challenging. Some existing migration barrier databases based on bond valence method exhibit limited accuracy and lack local descriptions of specific migration paths. To address these limitations, we have developed a high-throughput NEB workflow that integrates a fine-tuned CHGNet potential, enabling the efficient and accurate calculation of barriers across all inequivalent migration paths. This work presents a database containing 6928 materials with totally 20320 inequivalent migration paths and their corresponding energy barriers. Furthermore, by extracting local environment descriptors from these paths in lithium-containing compounds, we constructed machine learning models capable of effectively predicting energy barriers directly from the structural features of the migration pathways.

Keywords: database, high-throughput nudged elastic band (NEB) workflow, machine learning

PACS: 61.66.Fn, 61.68.+n

DOI: 10.1088/1674-1056/ae88e2 **CSTR:** 32038.14.CPB.ae88e2

1. Introduction

Solid-state electrolytes (SSEs) present a promising pathway for advancing lithium-ion battery technology by mitigating the inherent safety risks of conventional liquid electrolytes, such as leakage and flammability.[1–3] Furthermore, SSEs enable the integration of lithium metal anodes, thereby substantially enhancing energy density and extending cycle life.[4] These advantages promote SSEs as critical components for next-generation battery systems.

The development of high-performance SSEs necessitates the discovery of materials exhibiting high ionic conductivity, where high-throughput screening methods could dramatically accelerate this discovery.[5] In recent years, database-driven approaches have emerged as an effective strategy for accelerating the development of battery materials and solid-state electrolytes. Several large-scale databases have been established for ionic transport properties, electrochemical stability windows, electrolyte solvents, and composite solid-state electrolytes, providing valuable resources for data-driven materials discovery and machine-learning studies.[6–10] However, despite these advances, comprehensive databases containing migration-path-specific ion transport barriers remain limited.

Although density functional theory (DFT)-based nudged elastic band (NEB)[11] method calculates ion migration energy barriers with high precision in SSEs, the significant computational cost renders it unsuitable for large-scale calculations. Previous work[12] developed machine-learning (ML) models to learn barrier values based on the empirical bond valence method.[13] However, the models exhibit limited accuracy due to the intrinsic precision limits of the bond valence method and the absence of localized descriptors that capture path-specific migration characteristics. These models typically utilize global structural features, resulting in averaged barrier predictions that fail to reflect variations across distinct migration pathways. Recently, we developed a high-throughput NEB workflow[14] which samples all inequivalent lithium migration pathways and integrates the CHGNet[15] model to accelerate the exploration efficiency. Based on the higher accuracy of CHGNet potential and our path-specific sampling method, it is possible to construct a database containing each migration path and barrier of various materials, and extract local environment descriptors to accurately predict the barriers of specific paths.

In this study, we developed an automated high-throughput NEB workflow designed to sample all inequivalent lithium migration pathways and select the minimal-energy cyclic path spanning the crystal structure as the representative energy bar-

[†]Corresponding author. E-mail: rjxiao@iphy.ac.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

rier for each material. This workflow integrates the fine-tuned CHGNet potential to effectively compute the corresponding migration energy barriers. Utilizing this approach, we have generated a comprehensive database comprising fundamental properties and predicted migration barriers for 6928 Li-containing materials and 20320 inequivalent migration paths. Furthermore, using the Matminer[16] platform, we extracted 164 local environment descriptors (bonds related: AverageBondLength and AverageBondAngle;[17] local geometry related: OPSiteFingerprint[18] and VoronoiFingerprint;[19] interstice distribution related: IntersticeDistribution;[19] element related: LocalPropertyDifference;[20] and migrating distances related descriptors constructed manually) from the initial and final states of the migrating ion and built ML models for predicting barriers directly from migration path features. The resulting models demonstrate accurate prediction of migration energy barriers for a given ion migration pathway, thereby offering a potential application for the database generated by high-throughput NEB calculations.

2. Methods

Data acquisition The workflow for constructing the database is illustrated in Fig. 1. The workflow begins with the retrieval of candidate structures in CIF (Crystallographic Information File) format from the Materials Project[21] database. To ensure computational feasibility, a three-stage screening protocol was implemented. First, we focused on ternary, quaternary and quinary compounds containing lithium ions while excluding other alkali earth metals (e.g., sodium, potassium, rubidium). Next, structures were selected based on possessing only 1, 2, or 3 distinct lithium sites, thereby eliminating complex low-symmetry systems. This constraint is critical because the number of inequivalent migration pathways scales quadratically (N^2) with the number of ion sites (N), significantly reducing computational expense. Additionally, a thermodynamic stability criterion was enforced, requiring convex hull energy (E_{hull}) smaller than 100 meV/atom to enhance the experimental synthesizability.

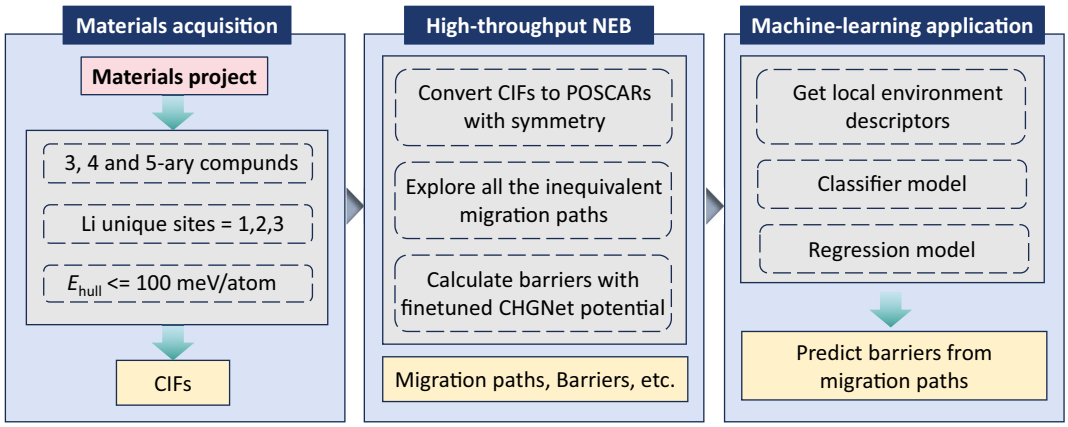


Fig. 1. Schematic of the automated workflow in this study.

High-throughput NEB method The obtained CIFs were processed by the high-throughput NEB workflow. For each possible migration path, the energy barriers were evaluated by the finetuned CHGNet potential[14] and transition-state configurations would be recorded. This workflow involved converting CIF files to POSCAR format while preserving symmetry labels for distinct Li sites (e.g., Li_1, Li_2, Li_3). Next, for each structure, the workflow identified and enumerated all possible inequivalent Li migration pathways ($\text{Li}_i \rightarrow \text{Li}_j$ pairs, where $i, j \in [1, N]$; total paths $= N^2$). For each identified pathway, the migration energy barrier was calculated using the fine-tuned CHGNet potential by the Atomic Simulation Environment,[22] encompassing structural optimization of the initial, final, and intermediate images along the path followed by NEB barrier determination. The convergence criteria for the NEB calculations were set with a maximum force threshold of 0.1 eV/Å. If the optimization exceeded 70 steps or if the forces along the path surpassed 10 eV/Å, the calculation was terminated and labeled as "fail" (indicating non-convergence). In such cases, the resulting energy barriers are considered less reliable and should be interpreted with caution.

To determine a representative barrier for each material (marked as "chgnet_barrier" in the database), a specific strategy was employed: structures with only a single unique pathway utilized that barrier directly. Conversely, for materials with multiple pathways ($N > 1$), the overall migration barrier was determined by identifying the lowest-energy continuous cyclic pathway spanning the crystal structure (e.g., $\text{Li1} \rightarrow \text{Li2} \rightarrow \text{Li3} \rightarrow \text{Li1}$). The barrier for such a cyclic path is defined by the maximum segment barrier encountered within the cyclic pathway, and the final representative barrier for the material is the minimum value found among all feasible cyclic pathways spanning the structure. Finally, we obtained 6928 materials with 20320 inequivalent migration paths.

Machine-learning method After obtaining the migration pathways and corresponding energy barriers via the high-throughput NEB method, we employed lithium-containing compounds to develop machine-learning (ML) models that correlate ionic migration paths with their energy barriers. The ML workflow involved three stages: First, local environment descriptors for the migrating ions, including AverageBondLength, AverageBondAngle, OPSiteFingerprint, IntersticeDistribution, VoronoiFingerprint, LocalPropertyDifference and migrating distances, were constructed based solely on the initial and final states of the migrating ion. Next, each migration pathway was treated as an individual sample, with the CHGNet-predicted barrier as the target label. Among the total 20320 samples, 9289 pathways failed to converge, primarily due to physically unreasonable migration paths, which led to unreliable barrier estimates. To filter out these non-convergent pathways, we conducted an XGBoost classifier to predict the convergence probability of each path, and then retained samples with a predicted convergence probability exceeding 90% for subsequent regression analysis. Finally, using the refined dataset, we trained an XGBoost-based regression model and applied Bayesian optimization to tune its hyperparameters, thereby obtaining a robust model for predicting migration energy barriers. The test sets for both the classification and regression models were constructed by randomly selecting 10% of the samples from their respective datasets.

DFT calculations All the DFT calculations were conducted by the Vienna *ab initio* Simulation Package (VASP)[23] within the projector augmented wave (PAW) approach and the Perdew–Burke–Ernzerhof (PBE)[24] generalized gradient approximation. For each structure, a uniform Monkhorst–Pack k-point mesh was employed such that the spacing between neighboring k-points in reciprocal space was less than 0.05 $Å^{-1}$, ensuring uniform sampling density across different unit cell sizes. A kinetic energy cutoff of 520 eV was used for the plane-wave basis set. The convergence criteria for total free energy and forces were set as 1.0×10^{-5} eV and 0.01 eV·A^{-1}, respectively.

3. Data records

The database of 6928 materials and the dataset for ML are available at https://in.iphy.ac.cn/eln/link.html#/184/J3w9. The database is stored in the CSV format. Table 1 summarizes the columns and descriptors for the database of migration barriers.

Figure 2 illustrates the elemental distribution across the 6928 compounds in the database, encompassing 66 distinct elements. To better represent the heterogeneous occurrence of elements, the colormap is scaled according to the natural logarithm (ln) of the percentage frequency of each element. In this representation, deeper red shades indicate higher elemental abundances. For instance, lithium, present in 100% of the compounds, defines the upper limit of the color bar at 4.61 (ln(100)). Similarly, Oxygen appears in 79.4% of the materials, yielding a value of 4.37 (ln(79.4)).

Table 1. Columns and descriptions for the database of migration barriers.

Content	Type	Comment
mp_id	str	Unique id of the compounds in Materials Project
formula	str	Formula of the compounds
space_group	str	Space group
crystal_system	str	Crystal system
band_gap	float	Band gap (eV) calculated by VASP using GGA function
ion_type	str	Migration ions type, 'Li' or 'Na'
ion_sites	int	Numbers of unique migration ion sites
volume_conv	float	Volume of the conventional cell ($Å^3$)
ion_num_conv	int	Numbers of migration ions in the conventional cell
ion_conc_conv	float	Migration ion concentration in the conventional cell (10^{27} m^{-3})
volumn_sp	float	Volume of the relaxed supercell ($Å^3$)
ion_num_sp	int	Numbers of migration ions in the supercell
ion_conc_sp	float	Migration ion concentration in the supercell (10^{27} m^{-3})
ion_num_neb	int	Numbers of migration ions in the NEB cell
ion_conc_neb	float	Migration ion concentration in the NEB cell (10^{27} m^{-3})
POSCAR_sp	str	Pymatgen structure information of super cell
ion_defect_energy	float	Defect energy calculated by VASP of ion (similarly ion1 and ion2)
min_defect_energy	float	Minimum defect energy
chgnet_barrier	float	Minimum chgnet energy barrier among all cyclic migration paths that start and end at the same site (eV)
ion0_to_ion0_chgnet_barrier	float	CHGNet energy barrier from site ion to site ion (eV), similarly ion0_to_ion1, ion0_to_ion2, etc.
ion0_to_ion_convergence	str	Success or fail of the convergence.
ion0_to_ion0_POSCARs	str	List of Pymatgen structures of the migration path from site ion to site ion, similarly ion0_to_ion1, ion0_ to_ion2, etc.
dft_barrier	float	Minimum DFT energy barrier among all cyclic migration paths that start and end at the same site (eV)
ion0_to_ion0_dft_barrier	float	DFT static energy barrier from site ion to site ion (eV), similarly ion0_to_ion1, ion0_to_ion2, etc.

In Fig. 3, we present the distributions of crystal systems, band gap (calculated by VASP using the GGA functional), CHGNet predicted overall migration barriers, and the concentrations of migrating ions. The band gap is a key descriptor of a material's electronic conductivity. A wider band gap (> 3 eV) suppresses electron excitation from the valence band to the conduction band, thereby minimizing electronic conductivity. This characteristic is essential for solid-state electrolytes to mitigate battery self-discharge and performance decay. To manage computational expense, the band gaps were evaluated using the GGA functional, a method recognized for generally yielding values lower than those obtained experimentally.[25]

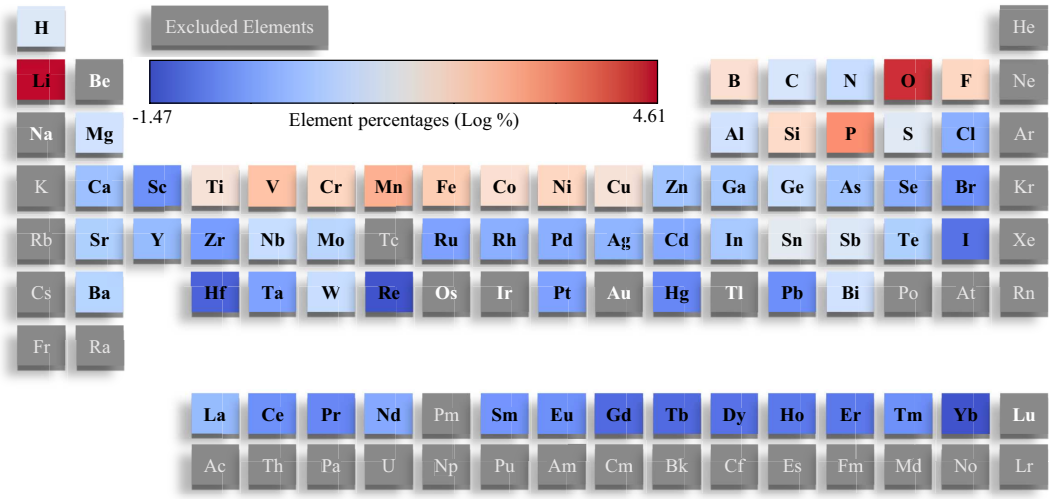


Fig. 2. Elemental distribution in the database. The color intensity is mapped to the natural logarithm of the element's percentage occurrence frequency (ln(f%)) to effectively visualize the non-uniform distribution.

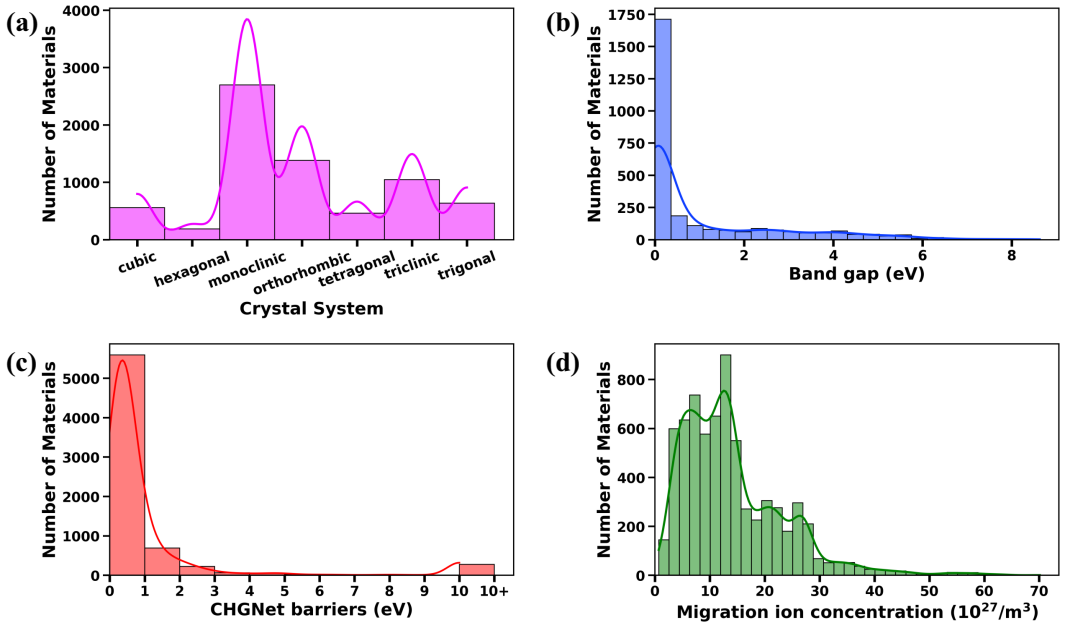


Fig. 3. The distribution of (a) crystal system; (b) band gap calculated by GGA; (c) CHGNet barriers; (d) migration ion concentrations in the database.

4. Technical validation

Comparison of CHGNet barriers with DFT barriers Due to the computational cost of the DFT-based NEB, we did not perform full NEB optimizations for all migration paths using DFT. Instead, we initially performed single-point DFT calculations on the transition states generated from the high-throughput NEB method and evaluated DFT barriers from static energies. Figure 4 displays a direct comparison between these DFT static energy barriers and the corresponding barriers predicted by the fine-tuned CHGNet model across 208 migration paths. The regression analysis yields an R^2 of 0.98 and a mean absolute error (MAE) of 0.10 eV, similar to that reported in our previous work,[14] indicating a strong agreement between the CHGNet predictions and the DFT reference values. To further validate the accuracy of the fine-tuned model,

we performed full DFT-NEB optimizations on a subset of eight representative materials. These selected systems encompass a wide range of structural and compositional complexities, spanning orthorhombic, tetragonal, trigonal, hexagonal, and cubic crystal systems, as well as ternary, quaternary, and quinary compositions. As shown in Fig. 5, the migration barriers derived from these complete DFT-NEB profiles exhibit excellent agreement with the CHGNet predictions, demonstrating the reliability of our model for high-throughput NEB calculations.

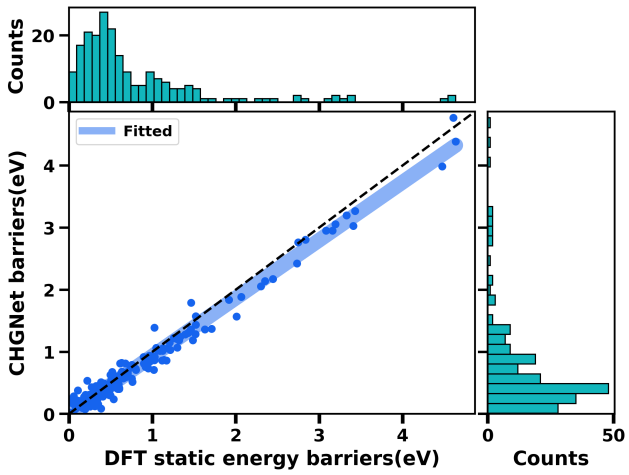


Fig. 4. Validation of CHGNet-based NEB barrier prediction comparing with DFT static energy barriers.

5. Data application

Beyond providing high-quality energy barriers, the database is designed as an AI-ready platform for the rapid screening of solid-state electrolytes. Our database offers path-specific information for all 20320 inequivalent migration pathways. For each migration path, we have extracted critical structural fingerprints, such as hopping distances and local geometric information, which could govern ionic transport. By transforming structural information into machine learning input, we can build models that can quickly predict the migration barriers of new structures.

Machine learning application We extracted 164 local environment descriptors from the initial and final states of the migrating ions and used them to train ML models to predict energy barriers of migration paths. The total dataset used in the ML is stored in npy format. As shown in Fig. 6(a), among the total 20320 pathway samples, 9289 pathways failed to converge, primarily due to physically unreasonable migration paths, which led to unreliable barrier estimates. Feature analysis reveals that convergence is predominantly governed by the total migration distance and the magnitude of structural deviation between the initial and final states. The structures with large migration distance or structural deviation are fundamentally incompatible with conventional vacancy-mechanism NEB analysis. To filter out these non-convergent pathways, we conducted an XGBoost classifier, which has an accuracy of 80.2% (Fig. 6(b)). Then we predict the convergence probability of each path, and retained 9821 samples with a predicted convergence probability exceeding 90% for subsequent regression analysis. Finally, using the refined dataset, we trained an XGBoost-based regression model. We adopted a material-based grouping strategy, ensuring that all migration paths belonging to a given material are allocated to either the training or the test set. We targeted a split ratio of approximately 10% for the test set, allowing for minor fluctuations. Specifically, from a total dataset of 9821 path samples across 4476 distinct materials, the resulting training set comprises 8830 samples from 4028 materials, while the test set comprises 991 samples from 448 materials. The regression performance on the test set is presented in Fig. 6(d). The model yields an R^2 of 0.72 and a mean absolute error (MAE) of 0.12 eV, and thereby obtains a robust model for predicting migration energy barriers.

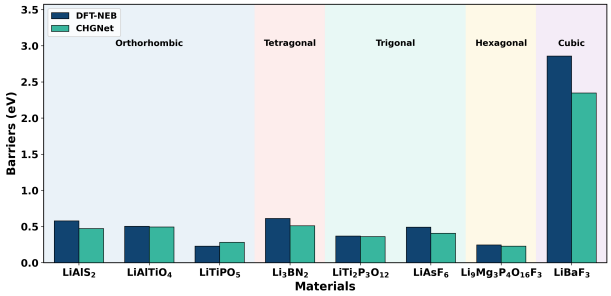


Fig. 5. Comparison of NEB barriers in representative structures of different crystal systems and compositions.

To identify the main sources of model error, we conducted error analysis in two ways: by crystal system and by chemical composition. The results are shown in Fig. 7. The model achieves the best performance on the trigonal system, with an R^2 of 0.86 and an MAE of 0.10 eV. The worst MAE (0.34 eV) occurs for the cubic system, while the lowest R^2 (0.6) is observed for the tetragonal system. When analyzed by chemical composition, oxides show the best performance.

Limitations and future perspectives First, it is important to recognize that our current database provides static barriers at 0 K calculated via NEB, neglecting factors such as finite-temperature effects, lattice vibrations (phonons), and configurational entropy. Therefore, applications of this database should be combined with molecular dynamics (MD) simulations for more accurate assessments. Second, our current NEB workflow assumes single-ion hopping transport. While interstitial or cooperative multi-ion mechanisms can in principle be studied via NEB with carefully designed initial/final states, such calculations are not yet feasible in high-throughput mode for disordered structures or events involving a large number of atoms.

In future work, we could utilize the migration pathways in this database and combine them with enhanced sampling MD to construct free energy barriers along the pathways to bridge

the gap between static NEB barriers and the actual dynamic behaviors. Also, we will extend the workflow to incorporate disordered structures generated via graph-theory-based Monte Carlo sampling.[26] This would enable the evaluation of ion migration barriers in complex materials with partial mobile-ion sublattice occupancy or mixed-framework configurations, thereby improving the relevance of predicted barriers to realistic, experimentally observed structures.

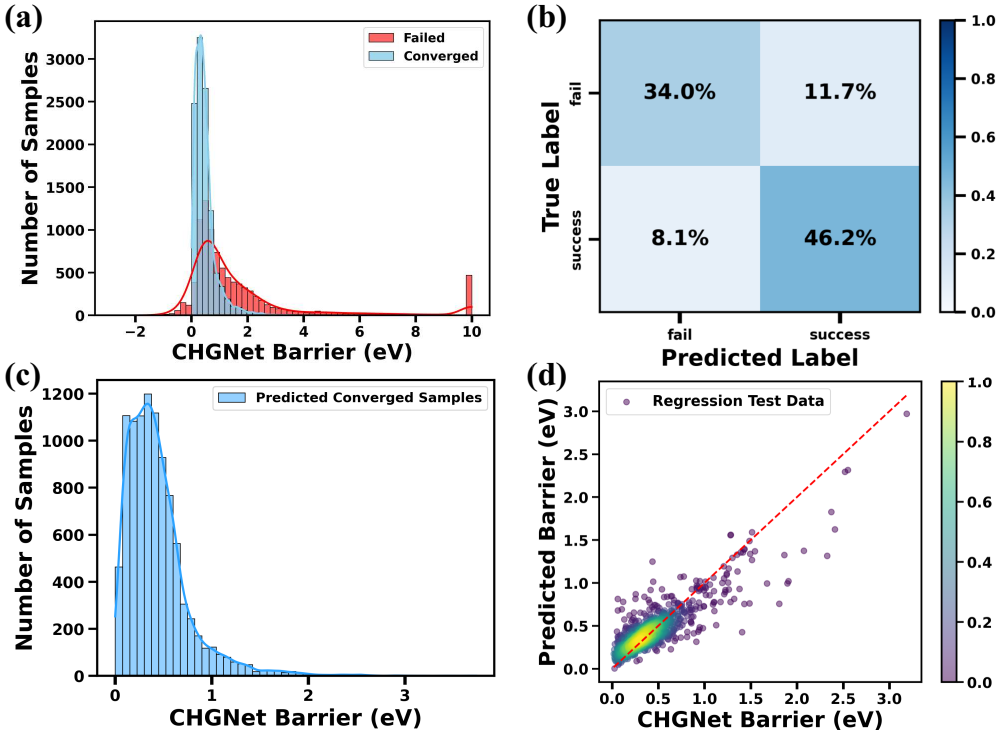


Fig. 6. (a) The distribution of converged and failed CHGNet barriers. Among the total 20320 samples, 9289 samples failed to converge. (b) Confuse matrix plot of the classifier model. The accuracy reaches 80.2%. (c) The distribution of CHGNet barriers among 9821 samples with the convergence probability exceeding 90%. (d) Comparison of predicted barriers and CHGNet barriers in the test set, R^2: 0.72, MAE: 0.12 eV.

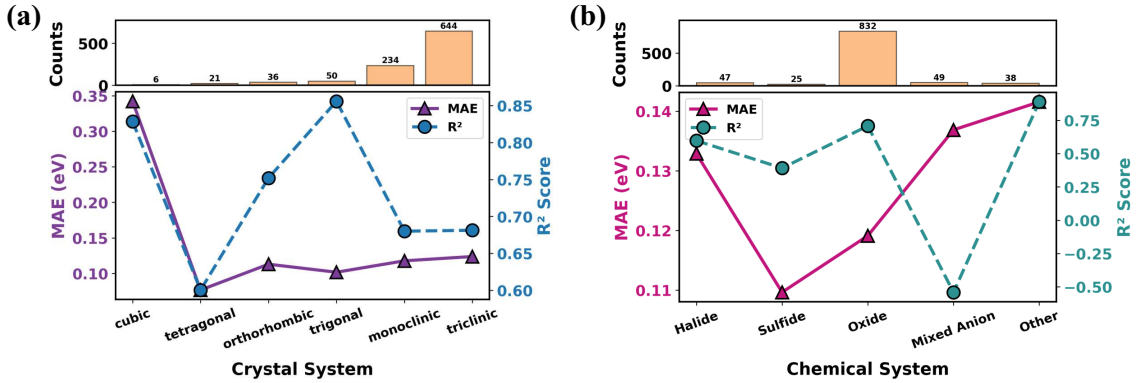


Fig. 7. The R^2 and MAE of the test set grouped by (a) crystal system and (b) chemical system.

Data availability statement

The data that support the findings of this study are openly available in Science Data Bank at the following link https://doi.org/10.57760/sciencedb.j00113.00321.

Acknowledgments

Project supported by Beijing Natural Science Foundation (Grant No. Z250005) and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB0500201). The governance of data was performed on the MatElab platform, developed by the Condensed Matter Physics Data Center of Chinese Academy of Sciences.

References

[1] Zhang S, Li S and Lu Y 2021 *eScience* **1** 163
[2] Yang H 2022 *Energy Science & Engineering* **10** 1643
[3] Xue N 2025 *Journal of Solid State Electrochemistry* **29** 4133
[4] Lin D, Liu Y and Cui Y 2017 *Nat. Nanotechnol.* **12** 194
[5] Benayad A, Diddens D, Heuer A, Krishnamoorthy A N, Maiti M, Cras F L, Legallais M, Rahmanian F, Shin Y, Stein H, Winter M, Wölke C, Yan P and Cekic-Laskovic I 2022 *Adv. Energy Mater.* **12** 2102678
[6] Zhang L, He B, Zhao Q, Zou Z, Chi S, Mi P, Ye A, Li Y, Wang D, Avdeev M, Adams S and Shi S 2020 *Adv. Funct. Mater.* **30** 2003087

[7] Lin S, Lin Y, He B, Pu B, Ren Y, Wang G, Luo Y and Shi S 2022 *Adv. Energy Mater.* **12** 2201808
[8] Wang D, He T, Wang A, Guo K, Avdeev M, Ouyang C, Chen L and Shi S 2023 *Adv. Funct. Mater.* **33** 2212342
[9] Wang X, He B, Liu B, Avdeev M and Shi S 2024 *Adv. Funct. Mater.* **34** 2406146
[10] Huang P, Yang Z, Liu Y, Yu P, Liu B, Xu M, Avdeev M and Shi S 2025 *Adv. Theory Simul.* **8** e01125
[11] Henkelman G and Jónsson H 2000 *J. Chem. Phys.* **113** 9978
[12] Wang Y, Wu S, Shao W, Sun X, Wang Q, Xiao R and Li H 2022 *J. Materiomics* **8** 1038
[13] Xiao R, Li H and Chen L 2015 *Sci. Rep.* **5** 14227
[14] Lian J, Fu X, Gong X, Xiao R and Li H 2025 *J. Mater. Chem. A* **13** 34918
[15] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J and Ceder G 2023 *Nat. Mach. Intell.* **5** 1031
[16] Ward L, Dunn A, Faghaninia A, Zimmermann N E R, Bajaj S, Wang Q, Montoya J, Chen J, Bystrom K, Dylla M, Chard K, Asta M, Persson K A, Snyder G J, Foster I and Jain A 2018 *Comput. Mater. Sci.* **152** 60
[17] De Jong M, Chen W, Notestine R, Persson K, Ceder G, Jain A, Asta M and Gamst A 2016 *Sci. Rep.* **6** 34256
[18] Zimmermann N E R and Jain A 2020 *RSC Adv.* **10** 6063
[19] Wang Q and Jain A 2019 *Nat. Commun.* **10** 5537
[20] Ward L, Liu R, Krishna A, Hegde V I, Agrawal A, Choudhary A and Wolverton C 2017 *Phys. Rev. B* **96** 024104
[21] Jain A, Ong S P, Hautier G, Chen W, Richards W D, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G and others 2013 *APL Mater.* **1** 011002
[22] Larsen A H, Mortensen J J, *et al.* 2017 *J. Phys. Condens. Matter* **29** 273002
[23] Kresse G and Furthüller J 1996 *Comput. Mater. Sci.* **6** 15
[24] Perdew J P, Ernzerhof M and Burke K 1996 *J. Chem. Phys.* **105** 9982
[25] Perdew J P and Levy M 1983 *Phys. Rev. Lett.* **51** 1884
[26] He B, Lai Z, Wang D, Liu X, Liu Y, Xu M, Pu B, Wang Q, Wang R, Avdeev M and Shi S 2026 *Acta Mater.* **302** 121628