Discovery of unconventional proton-conducting inorganic solids via defect-chemistry-trained, interpretable machine learning

Susumu Fujii\textsuperscript{1,2}, Yuta Shimizu\textsuperscript{3,4,5}, Junji Hyodo\textsuperscript{3,5}, Akihide Kuwabara\textsuperscript{2*}, and Yoshihiro Yamazaki\textsuperscript{3,4,5*}

(Date: May 16, 2023)

1 Division of Materials and Manufacturing Science, Osaka University, Osaka 565-0871, Japan
2 Nanostructures Research Laboratory, Japan Fine Ceramics Center, Nagoya 456-8587, Japan
3 INAMORI Frontier Research Center, Kyushu University, Fukuoka 819-0395, Japan
4 Department of Materials, Kyushu University, Fukuoka 819-0395, Japan
5 Kyushu University Platform of Inter-/Transdisciplinary Energy Research (Q-PIT), Kyushu University, Fukuoka 819-0395, Japan

* Corresponding authors: yamazaki.yoshihiro.105@m.kyushu-u.ac.jp, kuwabara@jfcc.or.jp
High-throughput computational screening and machine learning hold significant promise for exploring diverse chemical compositions and discovering novel inorganic solids\textsuperscript{1,2}. However, the complexity of point defects, which occur in all inorganic solids and are often central to their functionality and synthesizability\textsuperscript{3–5}, presents research challenges. Here, we introduce defect-chemistry-trained, interpretable machine learning for the efficient exploration and discovery of unconventional proton-conducting inorganic solid electrolytes. By considering dopant dissolution and hydration\textsuperscript{6–9}, our machine learning models provide quantitative predictions and physical interpretations for synthesizable host–dopant combinations with hydration capabilities across various structures. Utilizing these insights, the unconventional proton conductors Pb-doped Bi\textsubscript{12}SiO\textsubscript{20} sillenite and eulytite-type Sr-doped Bi\textsubscript{4}Ge\textsubscript{3}O\textsubscript{12} are discovered in the first two synthesis trials. Notably, Pb-doped Bi\textsubscript{12}SiO\textsubscript{20} represents an unprecedented class of proton-conducting electrolyte composed exclusively of groups 14 and 15 cations and featuring a sillenite structure, exhibiting fast and unique three-dimensional proton conduction along a softly bonded BiO\textsubscript{5} network.

Broad explorations of structure and chemical compositions using high-throughput computational screening and machine learning have been carried out for numerous inorganic functional solids, including those used in batteries\textsuperscript{10–12}, fuel cells\textsuperscript{13–15}, solar cells\textsuperscript{16,17}, catalysts\textsuperscript{18–20}, semiconductors\textsuperscript{21}, and thermoelectric materials\textsuperscript{22,23}. However, these explorations have primarily focused on predicting the functionality and/or the synthesizability of the materials in a defect-free ideal state, such as phonons for alkali-ion conductors\textsuperscript{11} and bandgaps for semiconductors\textsuperscript{21}. Point defects, which play critical roles in the functionality and synthesizability in a wide variety of inorganic materials, have been largely neglected due to the explosive increase in the degrees of freedom.
describing materials with point defects\textsuperscript{24,25}. Although some studies have incorporated point defects into the screening process\textsuperscript{13,15,16,19}, their explorational scope has remained limited to specific crystal structures and compositions, which is not conducive for the discovery of novel and unconventional functional materials.

Another challenge in materials exploration using computations and machine learning is the absence of adequate explanatory capabilities\textsuperscript{1,2,26}. Explanation and interpretation are vital for making informed decisions regarding structure, composition, and subsequent synthesis and experimental evaluations. Although interpretable machine learning models trained using high-throughput computational materials data can offer quantitative predictions and general understanding, their applications have been again limited to specific crystal structures and compositions, such as the formability of perovskites\textsuperscript{27} and Currie temperatures in two-dimensional materials\textsuperscript{28}. No study has yet considered and integrated the fundamental behaviours of point defects into interpretable machine learning models for both materials’ synthesizability and functionality in extensive chemical explorations.

Here, we present a defect-chemistry-trained, interpretable machine learning approach that efficiently identifies unconventional inorganic oxides for proton-conducting electrolytes, and detail how the approach achieves two successful syntheses in the first two trials. Small amounts of dopants and oxygen vacancies are crucial for activating proton incorporation and conduction in inorganic oxides\textsuperscript{6–8}. Our strategy, depicted in Fig. 1, comprises a funnel-like sequential high-throughput computation for point defects, physically interpretable machine learning for materials’ functions and synthesizability, followed by materials synthesis and experimental validation. To accommodate the defect chemistry of proton-conducting oxides, the initial sequential
high-throughput computations were designed to efficiently screen and collect data on five
critical attributes, including hydration and dopant dissolution, across a wide variety of
potential materials. The sequence begins with computations of bandgap, hydration, and
proton conduction for potential host compounds reported in the Inorganic Crystal
Structure Database (ICSD)\textsuperscript{29}, followed by more computationally demanding calculations
related to the phase stability and dopability of oxides (see Methods and Supplementary
Figs. 1–6). Based on the collected data, interpretable machine learning models that
explicitly consider the hydration and dopant dissolution are constructed using random
forest regressors and Shapley additive explanation (SHAP) (see Methods and
Supplementary Figs. 7–11)\textsuperscript{30}. A minimum set of descriptors, four for hydration
capability and seven for synthesizability of doped materials, is employed to explain the
structural and chemical features required for proton conductors. Materials exhibiting
desirable values of moderately negative hydration energy (around $-1$ eV\textsuperscript{8,31}) and lower
dopant solution energy are selected and subjected to materials synthesis and evaluations
(see Supplementary Discussion 1, 2).

Figure 2 shows the impact of each structural and chemical descriptor on the
predicted hydration and dopant solution energies. The SHAP value $\phi_i$ quantifies the
impact of each feature $i$ on the predicted hydration and dopant solution energies as a
function of $x$ in the descriptor space, $f(x) = \phi_0 + \sum_{i=1}^{N} \phi_i(x)$, where $\phi_0$ denotes the
average value over the training dataset (base value) and $N$ is the number of features. The
predictive capabilities of the developed models are validated by their consistency with
experimentally determined hydration energies and dopant solubilities (see
Supplementary Discussion 3).
The most impactful feature on both hydration and dopant solution energies is the type of the structure, distinguished by the mean coordination number of cations that represents the shape and network of coordination polyhedra (Figs. 2a,c). The predicted outcome agrees with empirical rules7 (Supplementary Discussion 3). The second most significant factor for hydration is the openness of the structure, as indicated by the volume per atom in Fig. 2b. This is consistent with the empirical rule stating that greater open space is favourable for hydration in conventional perovskite oxides7. A plateau over 11 Å³ in Fig. 2b provides a new insight that an increase in structural openness beyond 11 Å³ is unnecessary for hydration.

Regarding dopant solution, the compatibility in ionic size and electronegativity between the host and acceptor dopant cations is the next most important factor after structure type (Figs. 2d,e and Supplementary Table 5). The new insights revealed here are, firstly, that a dopant slightly larger in size than the host cation (by up to 0.3 Å in Fig. 2d) and having an equal or smaller electronegativity (by up to −1 in Fig. 2e) enhances dopability by reducing the dopant solution energy. The latter can be interpreted as indicating that incorporating a less electron-attractive dopant neutralizes the relative negative charge of the acceptor, making the dopant solution more favourable. Additionally, subtle stability of the host compound, compared to thermodynamically competing phases (closer to 0 eV/atom in the x-axis of Fig. 2f), is also favoured for dopant solution by −0.3 eV, which is physically reasonable. The models, along with the impacts of the additive features, successfully capture the empirical rules and provide new physical insights in a quantitative manner.

Novel proton-conducting inorganic electrolyte candidates were explored using predictions from the physically interpretable machine learning models. The candidates
are unreported hosts with acceptor dopants that primarily exhibit lower solution energies and secondarily possess hydration energies close to $-1$ eV. The list of candidates sorted by the primary condition in Supplementary Fig. 6 includes Bi$_{12}$SiO$_{20}$ sillenite, Bi$_4$Ge$_3$O$_{12}$ (eulytite-type), Li$_2$NbO$_4$ (ordered rocksalt-type), Cs$_3$BiO$_3$ and Rb$_3$BiO$_3$ (Na$_3$AsS$_3$-type), K$_2$Sn$_2$O$_3$ and K$_2$Pb$_2$O$_5$ (Hg$_3$S$_2$Cl$_2$-type), and ZnGa$_2$O$_4$ (spinel-type), shown as blue symbols in Fig. 2. Here, Pb-doped Bi$_{12}$SiO$_{20}$ sillenite and Sr-doped Bi$_4$Ge$_3$O$_{12}$ eulytite were chosen for the first two synthesis trials due to them having the lowest predicted dopant solution energies of 1.06 and 1.58 eV, respectively (Supplementary Fig. 12). The model predicted hydration energies of $-0.94$ and $-1.61$ eV, respectively, close to $-1$ eV (Supplementary Fig. 12). Low energy differences between proton sites of $\Delta E_{\text{hyd}} = 0.41$ and 0.20 eV, respectively, are also favourable for facilitating proton transport (Supplementary Table 2).

The structural disadvantages and chemical advantages of the nominated Pb-doped Bi$_{12}$SiO$_{20}$ sillenite and Sr-doped Bi$_4$Ge$_3$O$_{12}$ eulytite are quantitatively captured in Fig. 2 (see also Supplementary Figs. 12 and 13–16 for comparison by compound/structure). The drawbacks of Bi$_{12}$SiO$_{20}$ sillenite and Bi$_4$Ge$_3$O$_{12}$ eulytite as host structures lie in their large positive dopant solution energies of $+0.26$ and $+0.61$ eV, respectively, compared to a negative value of $-0.38$ eV for well-known BaZrO$_3$ perovskite (Fig. 2c and Supplementary Fig. 12). However, the strong compatibilities between the Bi$^{3+}$ host and Pb$^{2+}$ and Sr$^{2+}$ dopant cations in terms of ionic radius, electronegativity, and reaction energy counterbalance the structural disadvantages, reducing the solution energies of Bi$_{12}$SiO$_{20}$ sillenite by 0.20, 0.06, and 0.26 eV, respectively, and those of Bi$_4$Ge$_3$O$_{12}$ eulytite by 0.12, 0.06, and 0.20 eV (Figs. 2d–f and Supplementary Fig. 12). As a result,
the dopant solution energies of 1.06 and 1.58 eV become the lowest among the potential new proton-conducting oxides (Supplementary Fig. 6).

X-ray diffractometry confirmed the predictive capability of the models for the nominated Bi$_{12}$SiO$_{20}$ sillenite and Bi$_4$Ge$_3$O$_{12}$ eulytite in terms of dopant solution. The Rietveld refinements of the diffraction patterns indicate that the primary phases are sillenite (space group $I23$) and eulytite (space group $I4_3d$) with minor secondary phases of PbO and Bi$_2$O$_3$ in Bi$_{12-x}$Pb$_x$SiO$_{20-\delta}$ and of SrGeO$_3$ and GeO$_2$ in Bi$_{4-x}$Sr$_x$Ge$_3$O$_{12-\delta}$, along with traces of unreacted raw materials (Supplementary Figs. 17, 18; see Supplementary Tables 7, 8 for all the structural parameters). The presence of these secondary phases agrees with the computed phase diagram for the Bi-rich conditions of Bi$_{12-x}$Pb$_x$SiO$_{20-\delta}$ and Ge-rich conditions of Bi$_{4-x}$Sr$_x$Ge$_3$O$_{12-\delta}$ (Supplementary Fig. 5).

The lattice constant of Bi$_{12-x}$Pb$_x$SiO$_{20-\delta}$ decreases with increasing Pb dopant content up to $x = 0.6$ and becomes constant beyond that (Fig. 3a), indicating Pb dissolution at Bi sites with a solubility limit of around 5 at%. A monotonic decrease in lattice constant against Sr content up to $x = 0.4$ is observed for Bi$_{4-x}$Sr$_x$Ge$_3$O$_{12-\delta}$, also indicating the solubility limit of Sr at Bi sites of around 10 at%. These experiments validate the model predictions that Pb and Sr can be doped into the sillenite and eulytite oxides, respectively.

Proton conduction in the Bi$_{10.8}$Pb$_{1.2}$SiO$_{20-\delta}$ sillenite and Bi$_{3.6}$Sr$_{0.4}$Ge$_3$O$_{12-\delta}$ eulytite bulk solids is validated by AC impedance spectroscopy. Bulk proton conductivities in both materials show linear Arrhenius plots, with activation energies of 0.42 and 0.98 eV for the sillenite and eulytite, respectively (Fig. 3b and the Nyquist plots in Supplementary Figs. 19, 20). The activation energies are consistent with the calculated migration barriers of 0.63 and 1.16 eV, respectively, considering proton–dopant
association (Fig. 4 and Supplementary Figs. 21–23). Isotope exchange measurements from protons to deuterium at 485 °C reveal a decrease in conductivity by a factor of 1.2 (Fig. 3c), providing strong evidence of proton conduction (see Supplementary Fig. 24 for the Nyquist plots). Pb-doped Bi$_{12}$SiO$_{20}$ is the first proton-conducting oxide consisting solely of groups 14 and 15 cations and based on a sillenite structure.

The unconventional sillenite structure forming proton-conducting inorganic solids provides fast and unique three-dimensional proton conduction pathways. Bi$_{12}$SiO$_{20}$ sillenite consists of a network of soft and distorted BiO$_5$ polyhedra stabilized by a few rigid SiO$_4$ tetrahedra$^{32}$. Protons occupy and conduct near the corner-shared oxygen of the soft BiO$_5$ polyhedra while avoiding the rigid SiO$_4$ tetrahedra (Figs. 4b, 4d). The proton conduction pathways formed in the sillenite are distinct from known compounds since protons in conventional proton-conducting inorganic solids, including perovskite ABO$_3$, scheelite AMO$_4$$^{33}$, and solid acid CsH$_2$PO$_4$$^{34}$ migrate along the rigid BO$_6$ octahedra or the MO$_4$ or PO$_4$ tetrahedra$^{7,35}$. The Bi$^{3+}$ cation has a lone pair, high electronegativity (polarizability), and diverse coordination environments$^{36}$, offering a softly bonded BiO$_5$ network for proton conduction with a low migration barrier of 0.3 eV (Fig. 4c). The similarity in electronic configuration between Pb$^{2+}$ dopant and Bi$^{3+}$ host cation reduces the proton–dopant association energy to 0.3 eV (Fig. 4c and Supplementary Fig. 21), in contrast to the 0.7 eV value in Sr-doped Bi$_4$Ge$_4$O$_{12}$ (Supplementary Figs. 22, 23). Utilizing the network of softly bonded BiO$_5$, the p-block cation-based material discovered in this study exhibits fast proton conduction comparable to well-known perovskites.

In summary, we have demonstrated defect-chemistry-trained, interpretable machine learning models that can efficiently explore and discover novel proton-conducting inorganic solid electrolytes. The models provide a priori quantitative
predictions and physical interpretations for selecting synthesizable host–dopant combinations that can be hydrated from a wide range of possible structures and compositions. The insights gained in this study provide clarity in materials synthesis and experimental evaluation, and their application lead to the discovery of Pb-doped Bi$_{12}$SiO$_{20}$ sillenite and Sr-doped Bi$_4$Ge$_3$O$_{12}$ eulytite in the first two synthesis trials. Notably, the former is an unprecedented proton-conducting inorganic material solely composed of groups 14 and 15 cations and featuring a sillenite structure. The combination of the p-block Bi$^{3+}$ host and Pb$^{2+}$ dopant cations in the sillenite provides fast and unique three-dimensional proton conduction pathways along a softly bonded BiO$_5$ network, avoiding the rigid SiO$_4$ tetrahedra. This work demonstrates an efficient approach for exploring novel inorganic materials by explicitly considering the defect chemistry.
References


Data and code availability

The data that support the findings of this study and the computational codes used in this study are available from the corresponding authors upon reasonable request.

Acknowledgements

This work was supported by the Japan Science and Technology Agency CREST under Grant Number JPMJCR18J3. Y.Y. acknowledges financial support from the Kyushu University Platform of Inter-/Transdisciplinary Energy Research (Q-PIT) through “Module-Research Program”. The computations were performed in part using the Supercomputer Fugaku (Fugaku Battery & Fuel Cell Project) and the Supercomputer Center in the Institute for Solid State Physics, the University of Tokyo. We wish to thank Prof. Y. Okuyama, Prof. S. Kasamatsu, Dr. K. Shitara, and Prof. M. Shiga for fruitful discussions and Ms. M. Yoshitake for her assistance in the oxide synthesis.

Author contributions

Y.Y. and A.K. conceived the project. A.K. and S.F. designed the computational screening. S.F. developed the computational codes, performed theoretical calculations, and carried out the machine learning. J.H. compiled the literature for non-perovskite proton conductors. Y.S. synthesized the oxides and carried out the structural and electrochemical evaluations under the supervision of J.H. and Y.Y. S.F., J.H., and Y.Y. wrote the paper with contributions from all the authors.

Competing interests

The authors declare no competing interests.

Supplementary information

Methods
Supplementary Discussions 1–3
Tables S1 to S10
Figures S1 to S24
References (1–39)
Fig. 1 | Defect-chemistry-trained, interpretable machine learning for exploring proton-conducting inorganic solids. The proposed approach incorporates defect chemistry, which is essential for the synthesis and activation of proton conduction in inorganic solids. The initial sequential high-throughput computations efficiently screen and collect data on a broad range of candidate materials while considering the defect chemistry, namely hydration and dopant solution (bottom left). Interpretable machine learning models provide physical interpretations and new insights for selecting host–dopant combinations for discovering novel inorganic proton conductors (bottom middle and right). The models’ predictions regarding materials’ synthesizability and functionality are validated through materials synthesis and experimental evaluations (top right). MCN, mean coordination number; DIR, difference in ionic radius; DIE, difference in electronegativity; MIR, mean ionic radius; RE, reaction energy; VPA, volume per atom; ME, mean electronegativity.
Fig. 2 | Structural and chemical features and new insights into hydration and synthesizability of doped materials. a, Impact of mean coordination number of host cations on hydration energy. b, Impact of volume per atom on hydration energy. c, Impact of mean coordination number of host cations on solution energy. d, e, Impact on solution energy of differences in (d) ionic radius and (e) electronegativity between dopant and substituted host cation. f, Impact on solution energy of reaction energy of host compound from the neighbouring equilibrium phases (inverse distance to hull). Perovskites, reported proton-conducting non-perovskites, promising non-perovskites, and other non-perovskite compounds are indicated by yellow, green, blue, and grey circles, respectively. The promising non-perovskite compounds are the top eight unreported candidates, (Pb-doped) Bi$_2$SiO$_{20}$ sillenite, (Ti-doped) Li$_3$NbO$_4$ ordered rocksalt, (Sr-doped) Bi$_4$Ge$_2$O$_{12}$ eulytite, (Pb-doped) Cs$_3$BiO$_3$, (Pb-doped) Rb$_3$BiO$_3$, (Li-doped) K$_2$Sn$_2$O$_7$, (Li-doped) K$_2$Pb$_2$O$_5$, and (Mg-doped) ZnGa$_2$O$_4$ spinel. The total amount of hydration-energy data for training and testing is 80, which includes 8 perovskites, 12 double-perovskites, and 60 non-perovskites, with an averaged hydration energy ($\Phi_0$) of $-1.42$ eV. The total amount of solution-energy data is 200, including 26 perovskites, 94 double-perovskites, and 80 non-perovskites, with $\Phi_0$ of 1.36 eV. Pyrochlores composed of trivalent and tetravalent cations are located at the top of (a) whereas those composed of divalent and pentavalent cations are located at the bottom. The models provide physical interpretations and new insights into selecting a suitable host and dopant combination for materials discovery in a quantitative manner.
Fig. 3 | Synthesis and proton conductivities of nominated Bi$_{12-x}$Pb$_x$SiO$_{20-\delta}$ sillenite and Bi$_{4-x}$Sr$_x$Ge$_3$O$_{12-\delta}$ eulytite. a, Lattice constants. b, Arrhenius plots of bulk proton conductivities. c, H-D isotope effects of bulk proton conductivities. The activation energies of 0.42 and 0.98 eV for bulk proton conductivities in Bi$_{10.8}$Pb$_{1.2}$SiO$_{20-\delta}$ and Bi$_{3.6}$Sr$_{0.4}$Ge$_3$O$_{12-\delta}$, respectively, in (b) agree with the calculated values of 0.63 and 1.16 eV. Water vapour was changed from H$_2$O to D$_2$O at $x = 0$ in (c). The isotope effect of 1.2 is a strong indication of proton conduction.
Fig. 4 | Proton conduction through softly bonded BiO$_5$ network in Pb-doped Bi$_{12}$SiO$_{20}$ sillenite. a, Proton sites in Bi$_{12}$SiO$_{20}$. The larger white spheres represent more stable proton sites. b, Unique proton conduction pathways through softly bonded BiO$_5$ network. Non-equivalent proton sites are labelled by numbers. c, Migration energy barriers of protons in Bi$_{12}$SiO$_{20}$ and Bi$_{11.5}$Pb$_{0.5}$SiO$_{20}$ (without and with proton–dopant association, respectively). The energy barriers reach a maximum in the transition between positions 5 and 5’, yielding a migration energy barrier of 0.31 eV without proton–dopant association. The association between a proton and Pb$^{2+}$ is largest at position 1, with an energy of ~0.3 eV. The total migration energy barrier considering the association is about 0.6 eV, comparable to conventional proton-conducting perovskite oxides. d, Two examples of long-range proton diffusion. Only the most stable proton sites and SiO$_4$ tetrahedra are shown. Protons migrate from the most stable site in three directions, reflecting its cubic symmetry, while avoiding SiO$_4$ tetrahedra. The area enclosed by red dotted lines is shown in (b).