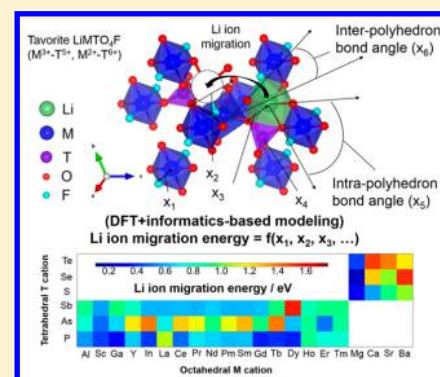


Informatics-Aided Density Functional Theory Study on the Li Ion Transport of Tavorite-Type LiMTO_4F ($\text{M}^{3+}-\text{T}^{5+}$, $\text{M}^{2+}-\text{T}^{6+}$)Randy Jalem,^{*,†,‡,⊥} Mayumi Kimura,[‡] Masanobu Nakayama,^{†,‡,§} and Toshihiro Kasuga^{||}[†]Unit of Elements Strategy Initiative for Catalysts & Batteries (ESICB), Kyoto University, Katsura, Saikyo-ku, Kyoto, 615-8520, Japan[‡]Department of Materials Science and Engineering and ^{||}Department of Frontier Materials, Nagoya Institute of Technology, Gokiso, Showa, Nagoya, Aichi 466-8555, Japan[§]Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan

Supporting Information

ABSTRACT: The ongoing search for fast Li-ion conducting solid electrolytes has driven the deployment surge on density functional theory (DFT) computation and materials informatics for exploring novel chemistries before actual experimental testing. Existing structure prototypes can now be readily evaluated beforehand not only to map out trends on target properties or for candidate composition selection but also for gaining insights on structure–property relationships. Recently, the tavorite structure has been determined to be capable of a fast Li ion insertion rate for battery cathode applications. Taking this inspiration, we surveyed the LiMTO_4F tavorite system ($\text{M}^{3+}-\text{T}^{5+}$ and $\text{M}^{2+}-\text{T}^{6+}$ pairs; M is nontransition metals) for solid electrolyte use, identifying promising compositions with enormously low Li migration energy (ME) and understanding how structure parameters affect or modulate ME. We employed a combination of DFT computation, variable interaction analysis, graph theory, and a neural network for building a crystal structure-based ME prediction model. Candidate compositions that were predicted include LiGaPO_4F (0.25 eV), LiGdPO_4F (0.30 eV), LiDyPO_4F (0.30 eV), LiMgSO_4F (0.21 eV), and $\text{LiMgSeO}_4\text{F}$ (0.11 eV). With chemical substitutions at M and T sites, competing effects among Li pathway bottleneck size, polyanion covalency, and local lattice distortion were determined to be crucial for controlling ME. A way to predict ME for multiple structure types within the neural network framework was also explored.



INTRODUCTION

Development of solid electrolytes for high performance all-solid-state rechargeable Li ion batteries has now become a major interest in the field of energy conversion devices. This is primarily driven by desires of nonflammability, operating stability, corrosion resistance, a wide operating temperature window, and longer service life. However, only a limited number of potential materials are known to date, with most of the suggested materials having various limiting factors (eg., Li diffusivity, chemical stability, and cycle performance), ultimately preventing their commercial use.^{1–7} Thus, new solid electrolytes are still continually sought, hoping to realize the next breakthrough in this technological field.

Recent studies have reported tavorite-type materials as promising candidate battery cathodes.^{8–14} This class of materials has the general formula AMTO_4X , where A is occupied by an alkali or alkaline-earth element, M is a metal, T is a p-block element, and X can be any anion species such as O, F, or OH. MO_4X_2 units are vertex-linked via X anion and form chains in one dimension (see Figure 1a). There are two Li path types in the cavities formed by MTO_4X (Figure 1b and c), either passing through trans- TO_4 units (paths 1 and 4) or trans- MO_4X_2 units (paths 2 and 3). What prompted us to explore this structure is the report on one tavorite composition,

LiFeSO_4F ,¹² which was confirmed to be capable of a high Li insertion rate. If this fast Li ionic conduction is mainly characteristic of the MTO_4X framework, then a similar behavior could be observed for analogs with nontransition metals at M sites; the latter would ensure ionic conduction, making the chemistry suitable for solid electrolyte application. Mueller et al. used high-throughput (HT) density functional theory (DFT) method to evaluate tavorite-structured materials with M redox couples, including fluorosulfates, for use as Li-based cathodes.¹⁵ However, a chemical survey geared toward solid electrolyte application has not been done yet. Looking for other isostructural compounds in this group could provide informed choices for experimentalists. So far, only the tavorite LiMgSO_4F has been experimentally tested as a solid electrolyte. Although the measured Li ion diffusion barrier was relatively high (0.94 eV),¹⁶ it can be made intrinsically low enough (0.20 eV) if Li vacancies are introduced through defect structure engineering (eg., by supervalent doping at M sites).¹⁷

DFT-based calculation with chemical substitution has now become a common practice aiding the discovery and theoretical evaluation of battery materials.^{15,18–22} Moreover, sophisticated

Received: October 14, 2014

Published: May 22, 2015

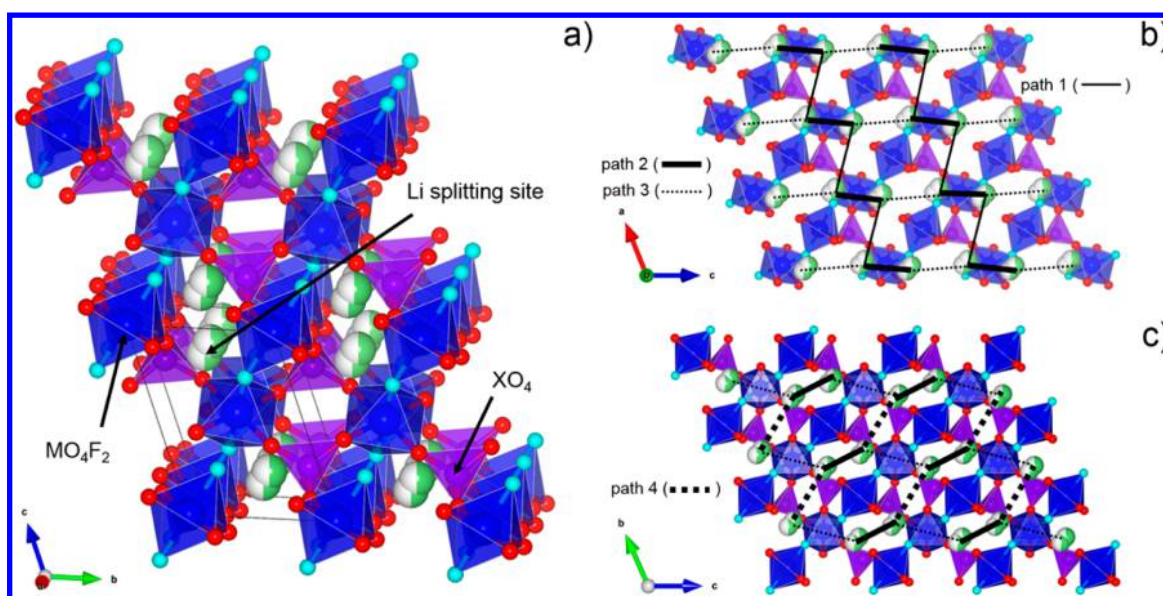


Figure 1. (a) Schematic 3D-projection, (b) *ac*-plane, and (c) *bc*-plane views of the tavorite structure (in a $2 \times 2 \times 2$ supercell). Green/white spheres are half occupied Li/vacancy atoms arranged in a zigzag manner along the *a*-channel, light gray octahedra are MO_4F_2 units, and dark gray tetrahedra are TO_4 units.

statistical and informatics-based techniques are now becoming increasingly popular in enhancing molecular simulations and providing robust platforms for establishing structural and structure–property relationships from large multivariate chemical data sets.^{23–29} Recently, we demonstrated how to efficiently screen and analyze the Li-based olivine-type materials for solid electrolyte application using a combination of DFT calculation and multivariate analysis.^{26,27} We were able to identify promising compositions and elucidate how structural features can affect a given property of interest.

In this work, we screened and evaluated the structural and Li ion transport properties of the tavorite-type LiMTO_4F with chemical substitutions at M and T sites. We employed an informatics-driven approach by combining DFT computation and neural network (NN) analysis to build a prediction model for ME. Input variable neurons were derived mainly from a database of structural information extracted from DFT-optimized crystal structures of different compositions (M–T pairings). The model building step includes input variable data preprocessing, training and validation.

■ COMPUTATIONAL DETAILS

All DFT calculations were performed using projected augmented wave (PAW) method³⁰ as implemented in Vienna Ab Initio Simulation Package (VASP).³¹ Standard generalized gradient approximation (GGA) using the parametrization of Perdew–Burke–Ernzerhof for solids (PBEsol) was used to describe the exchange correlation energy.³² Published structural information on experimentally synthesized tavorite compounds were taken from the Inorganic Crystal Structure Database (ICSD).³³ Virtual compositions were calculated using experimental coordinate set. Spin-polarized calculation with an energy cutoff of 500 eV was carried out and allowed the total energies to converge within 3 meV per formula unit (f.u.). A $5 \times 4 \times 3$ Monkhorst–Pack kpoint grid in a 16-atom unit cell showed a reasonable convergence and an error of <1 meV/f.u. in the total energy. Density functional perturbation theory (DFPT) was used for the estimation of the Born effective

charges, for evaluating ionicity/covalency of cation bonds.³⁴ The nudged-elastic-band (NEB) method³⁵ was used to determine migration pathway and energy; the calculation was done with a $1 \times 2 \times 2$ supercell. Numerical integration was performed over the Brillouin zone by sampling the Γ -point. Convergence was determined to be within 30 meV per formula unit.

The LiMTO_4F formula has 72 conceivable M–T pairings ($\text{M}^{3+}-\text{T}^{5+}$ and $\text{M}^{2+}-\text{T}^{6+}$), but some M cations were excluded (Eu^{3+} , Yb^{3+} , Lu^{3+}) due to convergence issues, leaving only a total of 63 compositions for the chemical space to be evaluated. For supervision, 37 compositions were sampled (randomly) with their ME by DFT.

For the selection of ME descriptor variables, we systematically collected a large number of geometric- and electronic-based parameters from our database of DFT-optimized structures; these variables can be categorized as lattice parameters, ion effective charges, intrapolyhedron parameters, and interpolyhedron parameters.^{26,27} In total, 85 variables were selected (see Table S1 and Figure S1 in the Supporting Information). We then constructed a matrix with 63 rows (of M–T pairs) and 85 columns (of variables), which hereafter will be referred to as the initial input variable space *X*. Although redundant variables are likely to be included in *X* with the random choice of variables, this also allows us to pick up important variables that could be overlooked when simply relying on chemical intuition. One of the most common approaches to deal with multicollinearity is by principal component analysis (PCA). However, when we tested our multivariate data set, the first principal component (PC1) from PCA can only weakly explain for the data variability, that is, only 32% of the variance has been explained; PC2 and PC3 each both have 20% of the variance explained (see Figure S1 in the Supporting Information). So instead, co-occurrence patterns were analyzed using graph theory³⁶ and a selection scheme based on Pearson product-moment correlation coefficient *R*.³⁷ We set a condition for determining the presence of strong collinearity: for $|R| > 0.5$, two distinct variables are considered strongly collinear; otherwise, they are

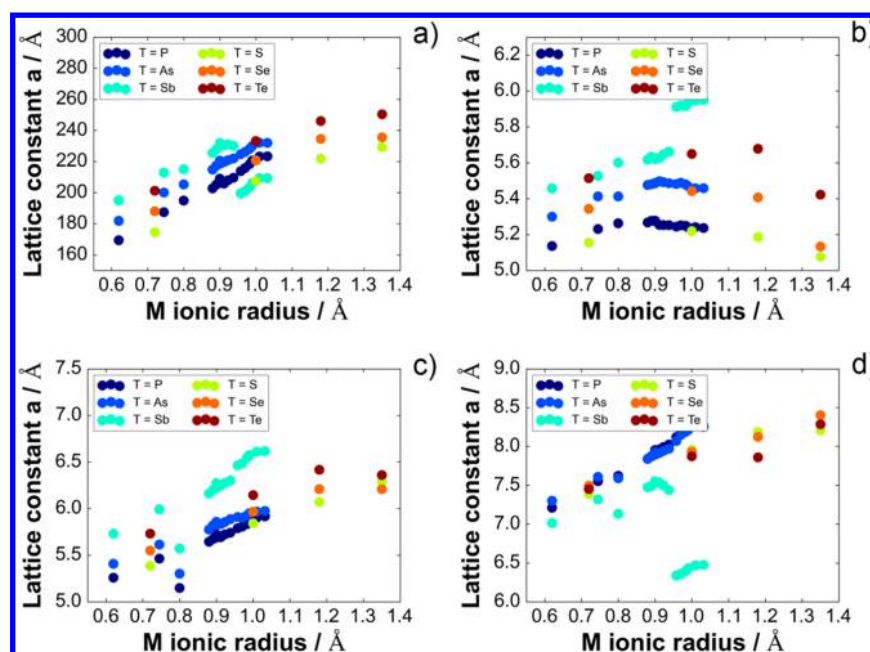


Figure 2. Structural data across the olivine LiMTO_4F search space: (a) cell volume, (b) lattice parameter a , (c) lattice parameter b , and (d) lattice parameter c of DFT-optimized structures.

treated to be independent from each other. These constraints allows us to effectively remove strictly collinear variables. This collinearity flag was then incorporated into the graph model by forming connection edges between vertices (variables). The cluster/class structure in the graph topology was determined using the modularity score, with a resolution limit of 0.75.³⁸ The graph model was analyzed and visualized using the Python package NetworkX³⁹ and interactive platform Gephi.⁴⁰

Similar with our previous work, the ME prediction model was constructed based on the neural network (NN) architecture using DFT-based structure data.²⁶ The model is described by the equation:

$$\text{ME}_l = \sigma^{(2)}\left(\sum_{k=1}^K w_{2kl}\sigma^{(1)}\left(\sum_{j=1}^J w_{1jk}X_j + \theta_j\right) + \theta_k\right) \quad (1)$$

where σ is the activation function, w is the neuron weight, and θ is the applied bias at a layer. In this study, a single hidden layer network with four hidden nodes and a single external attribute (ME) is evolved. Training was performed using the back-propagation algorithm⁴¹ with learning rate η fixed to 0.1⁴² and with a hyperbolic tangent function for activating the network neurons.⁴³ Overfitting mitigation was made by adding an absolute penalty term for the error function during ME model training.^{44,45} The convergence criterion for the error function was set 10^{-8} eV. Given that no prior information is available on the actual number of local minima of the error landscape,⁴⁶ the training process was performed by varying combinations on training and validation, each combination was repeated with different random starting node weight values 10 000 times. The number of random combinations was set to 30, leading to 300 000 models from which the best fit was selected. The sample ME data set was partitioned into 80% training set and 20% validation set.

For the analysis of structure–property relationship from the final ME model, the causal index (CI) method was used:⁴⁷

$$\text{CI} = \sum_{k=1}^h w_{lk}w_{kj} \quad (7)$$

where h is the number of hidden nodes, w_{lk} are weights linking the hidden node k to the output node l , and w_{kj} are the weights linking input node j to hidden node k . The indices reveal the direction (positive or negative) and relative contribution of each input variable to the ME attribute.

For the predictor variables of ME, structure-related data was extracted from DFT-optimized LiMTO_4F model cells. Descriptions of these variables are found in Table S1 and Figure S2 in the Supporting Information section. The JMP Statistical Software was used for NN model building as well as for other analyses (PCA and PLS).⁴⁸

RESULTS AND DISCUSSION

Structural Information. We followed a basic causation principle that ME can be explained either by how atoms are arranged to form a crystal structure or by how it responds to variations in crystal structure-inherent parameters. Typical parameters are shown in Figure 2 for DFT-optimized favorite LiMTO_4F cells. With M and T substitutions, extreme terminals in the trend are readily determined such as for cell volume; the largest volume is $\text{LiBaTeO}_4\text{F}$ (250.37 \AA^3) while the smallest one is LiAlPO_4F (159.04 \AA^3), with a difference of 44.6% on per atom basis. Understandably enough, the observed volume trend positively correlates with M and T cation sizes. Meanwhile, more primitive parameters can also be used as ME descriptors such as lattice constants a , b , and c are as shown in Figure 2b–d. Except for lattice constant a , a strong positive correlation exists for cell volume, lattice constant b , and lattice constant c with M cationic size. The main cluster of data points from 0.88–1.03 Å M size, which belongs to rare-earth-containing samples, shows discontinuity associated with Sb cations (in aqua). This can be attributed to the severe distortion caused by the large-sized Sb cation that is supposed to occupy the T site. But instead of a tetrahedral environment, large-sized M cations

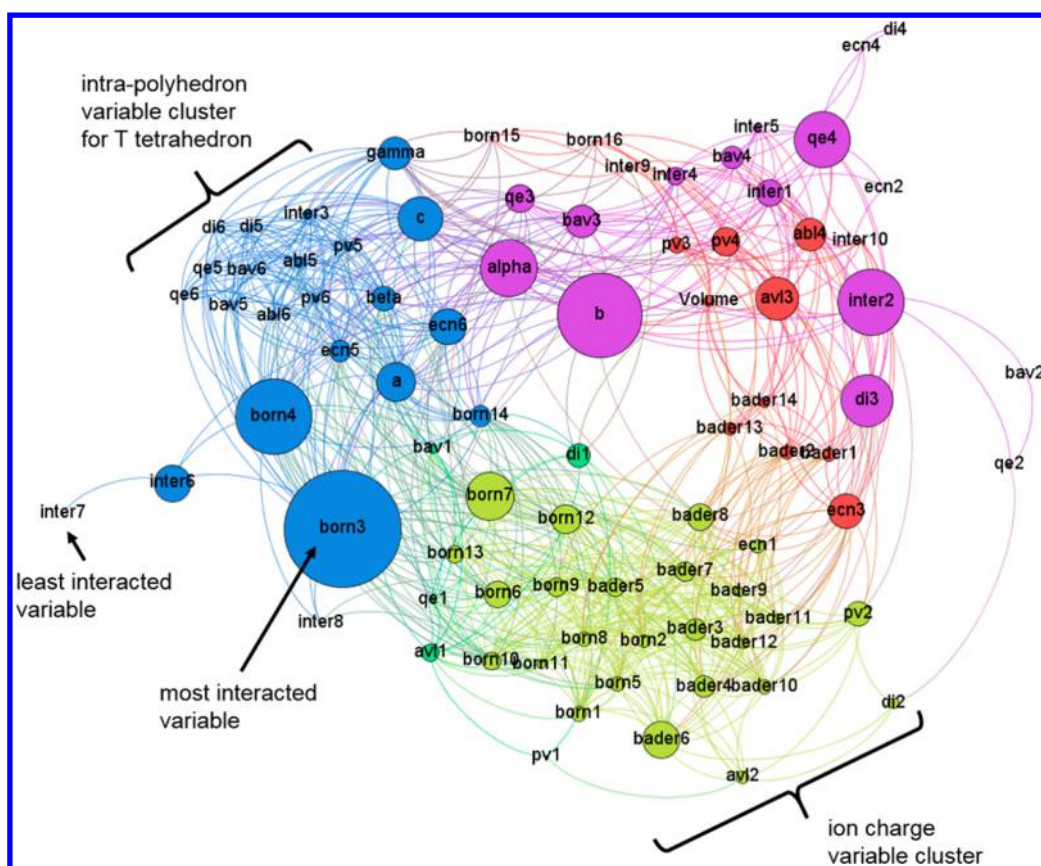


Figure 3. Graph theory-based network constructed from variable interaction effects according to Pearson's product-moment correlation coefficient R . The network nodes (structure-related variables) are linked by node edges to other nodes if collinearity is present ($|R| > 0.5$). Relatively larger nodes have more edges formed than smaller ones. Node clustering indicates strong in-group collinearity; four clusters/classes are determined (modularity class score = 0.376, resolution limit = 0.75).

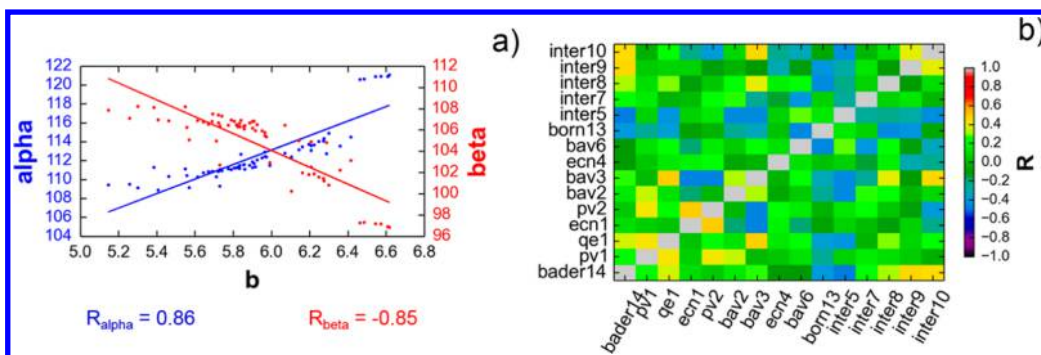


Figure 4. (a) Examples of input variables (alpha angle α , beta angle β) with high collinearity (positive and negative correlation, respectively) with another input variable (b lattice constant b). (b) R table of reduced input variable data set (with correlation criterion $|R| > 0.5$).

tends to assume higher coordination. For example, $\text{LiLaSbO}_4\text{F}$ (a terminal point for $0.6 \text{ \AA} < r_{\text{M}} < 1.0 \text{ \AA}$ for $\text{M}^{3+}-\text{T}^{5+}$ pairing) has Sb^{5+} occupying a distorted 5-fold coordination after structure relaxation. This observation underpins the instability when forcing Sb^{5+} in a tetrahedral site, this is consistent with those observed for several known Sb-containing compounds.^{49–52} Although unlikely to be synthesized, we nevertheless included Sb^{5+} -containing samples in our subsequent ME modeling in order to capture information and insights about the extent of cation size to be accommodated in the favorable structure.

Input Variable Space Analysis. Prior to model building, we preprocessed the initial X matrix to reduce model

complexity at the onset and eliminate early on redundant variables. Results of correlation coefficient, R , evaluation are visualized in the graph theory-based network shown in Figure 3. This network of variable interaction consists of vertices representing the variables themselves and connection edges which shows whether two variables are highly collinear or not. An example of strong collinearity to multiple variables is displayed in Figure 4a for variable b (lattice parameter b), with variables alpha (angle α) and beta (angle β). Upon inspection, we can qualitatively check the number of collinearities (the larger the vertex size, the more edges are formed with it) and variables' in-group collinearity (as indicated by their color clustering and close proximity in the network). Variables

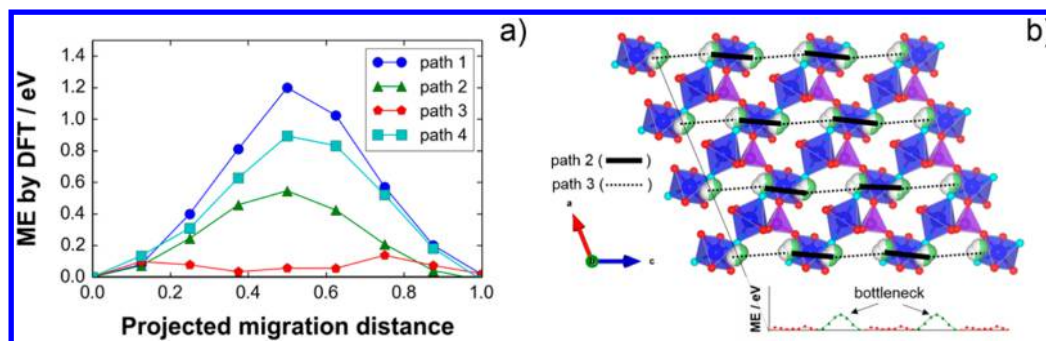


Figure 5. (a) Evaluated Li ion migration energy (ME) profiles of pathways necessary for long-range Li ion transport in all three major directions (test case: LiAlPO_4F) as reported in our previous work.¹⁷ (b) Predicted 1D conduction channel formed by energetically favorable pathways (paths 2–path 3 linkages) in part a.

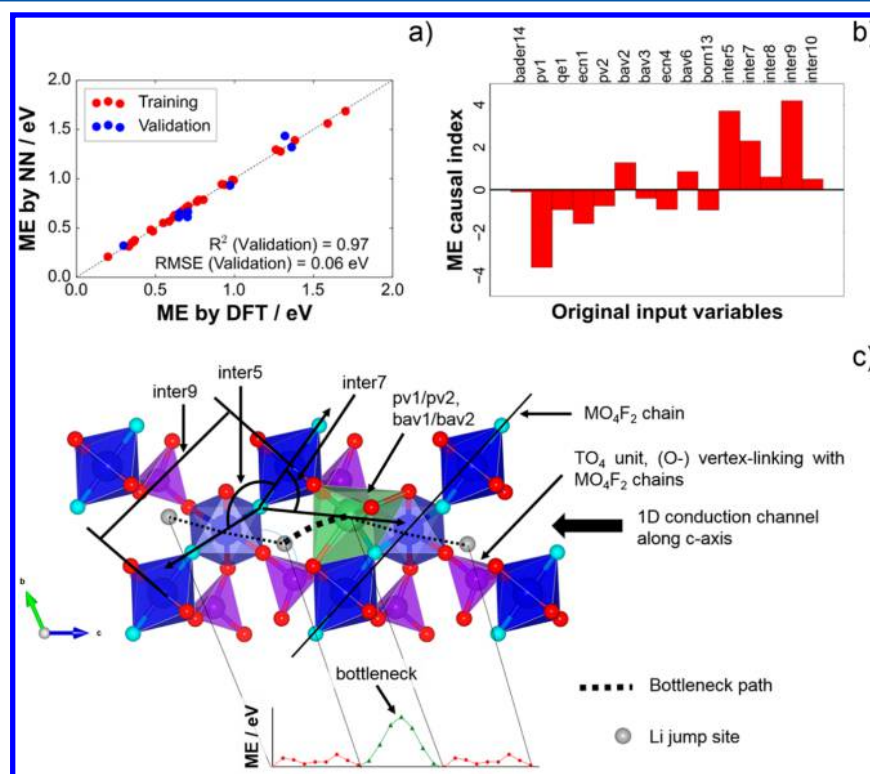


Figure 6. (a) Fitting quality of the final NN model for ME prediction, (b) calculated ME causal index (CI) values of input neuron variables, and (c) schematic representation of selected variables with high CI values and their spatial relationship with the Li ion bottleneck pathway (path 2 of Figure 4a).

appearing as large circles in Figure 3 have the most edges formed and were subsequently removed from X : examples are born3, b , born4, inter2, alpha, qe4, di3, c , gamma, and inter6. The least correlated variable is inter7 ($M(1)\text{--}F(2)\text{--}Li(2)$ angle) with only one edge formed, with inter6 ($M(2)\text{--}F(2)\text{--}Li(2)$ angle). Vertices are noted as well to fall in any of the four identified clusters, (in-group collinearity, modularity class score = 0.376). Interestingly, variables which are physically linked tend to group among themselves, an example is the clustering of ion charge parameters (green vertices) for Bader charge (e.g., bader1, bader2, ..., bader14) and Born effective ion charge (eg, born5, born6, ..., born14); mostly have $|r| > 0.8$. Another example is the clustering of intrapolyhedron parameters (blue vertices) associated with the T tetrahedron unit which, again, suggests redundancy in the inclusion of either parameter groups (T(1) tetrahedron, abl5, pv5, di5, qe5, bav5, and ecn5, and T(2) tetrahedron, abl6, pv6, di6, qe6, bav6, ecn6). Thus, the

graph network can reveal variable interactions and clustering effect that are consistent when viewed physically with respect to the favorite structure.

For the dimensionality reduction of the initial X matrix, since the graph model reveals some degree of hierarchy with respect to the number of node connections, we successively remove variables with the highest number of collinearities (i.e., remove the largest node first in Figure 3) and update the correlation R table. From the updated R table, the new largest node variable is removed next. This procedure is repeated until only variables without node edges remain.

After dimensionality reduction, X is reduced to 15 variables as indicated in the R table in Figure 4b. All variables in the lattice parameter group were eventually removed. In the charge parameter group, two variables were left (bader14, born13). For intra- and interpolyhedron parameters, eight (pv1, qe1, ecn1, pv2, bav2, bav3, ecn4, bav6) and five variables (inter5,

inter7, inter8, inter9, inter10) were left, respectively. We then fed the reduced X into input neurons of the NN network for model training.

Li Ion Diffusion within the Tavorite Framework. In any regression modeling, an external (set of) attribute(s) is (are) used for supervised learning. For practical reasons, we opted to train the ME model against a single attribute. However, there are different conceivable Li pathways within the tavorite framework,^{14,15,17,53} and any of these pathways can be essentially assigned as the external attribute. But to enable the ME model to properly discriminate compositions for solid electrolyte use, the representative attribute has to be at least a bottleneck pathway that determines full connectivity in at least one cell direction. Analyzing the ME profiles in the four known Li jump paths as shown in Figure 5a (test case: LiAlPO_4F , a material reported in the ICSD), it can be deduced that only paths passing between trans-positioned MO_4F_2 octahedra (paths 2 and 3) have low ME values (<0.55 eV), in contrast with paths passing between trans-positioned TO_4 tetrahedra (paths 1 and 4, 0.89–1.34 eV). This means that Li ion transport proceeds mainly in 1D along *c*-direction, composed of alternating path 2–path 3 linkage (Figure 1b and c and 5b). The relatively larger ME value of this 1D pathway (path 2) was chosen as the representative attribute for the NN model.

Li Ion Migration Energy (ME) Model Building. Results on ME model training and validation are shown in Figure 6a. The validation metrics are $R^2 = 0.97$ and $\text{RMSE} = 0.06$ eV which suggests a reasonable predictive power. The final NN formula is given by

$$\text{ME} = 0.800 - 0.213 \tanh(v_1) - 0.580 \tanh(v_2) - 0.295 \tanh(v_3) - 0.264 \tanh(v_4) \quad (2)$$

where v_1 , v_2 , v_3 , and v_4 contain linear combinations of the input neuron variables (with trained coefficients) in the hidden layer (see Figure S3 and Table S2 for details in the Supporting Information). The model is relatively simple, mainly with only 6 of the 15 input variables having relatively significant CI values (Figure 6b); these are inter9 (M(2)–M(2) distance, third nearest neighbor M(2)), pv1 (polyhedral volume of Li(1) octahedron), inter5 (F(1)–F(2)–F(1') angle, F(1') in mirror unit cell), inter7 (M(1)–F(2)–Li(2) angle), bav2 (bond angle variance of Li(2) octahedron), and ecn1 (effective coordination number of Li(1) ion); illustration is shown in Figure 6c. Other variables only have considerably minimal contribution or small weight values in the final model (CI values close to zero). Increasing the number of hidden neurons did not lead to any significant improvement in the fitting quality of the data (see Figure S4 in the Supporting Information section). Meanwhile, to justify the use of nonlinear fitting by NN in this work, we also fitted our data set using the partial least-squares (PLS) method,⁵⁴ a linear regression technique that can address multicollinearity while finding uncorrelated latent variables; the latent variables are linear combination of original predictor variables, similar to the principal components in PCA. To ensure fair comparison between of the PLS and the present NN results, we used the original 85 predictor variables for PLS fitting and then performed the model building step with the same training set and validation set. The resulting training and validation metrics for the PLS fitting are as follows: $R^2(\text{Training}) = 0.65$, $\text{RMSE}(\text{Training}) = 0.22$ eV, $R^2(\text{Validation}) = 0.76$, and $\text{RMSE}(\text{Validation}) = 0.19$ eV (see

Figure S5 in the Supporting Information section). Clearly, linear regression by PLS predicts ME poorly.

Using the validated NN model, ME values for the rest of M–T pairings were then predicted at a stroke. Figure 7 maps the

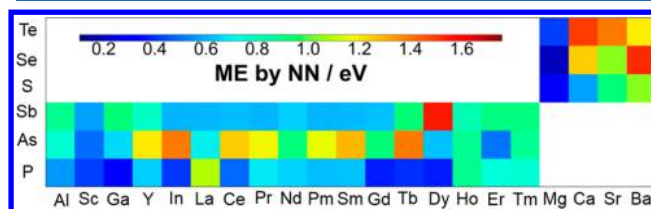


Figure 7. NN model-predicted ME for different LiMTO_4F compositions sorted with respect to M and T cations.

predicted EA values for the LiMTO_4F space. For solid electrolyte application, candidate compositions are screened based on a selection limit of ≤ 0.30 eV; this cutoff is referred from the activation energy of perovskite-type (Li, La) TiO_3 , which is a known fast Li ionic conducting oxide (1×10^{-3} S/cm at 300 K).⁴ Promising compositions, as predicted by the NN model, include LiGaPO_4F (0.25 eV), LiGdPO_4F (0.30 eV), LiDyPO_4F (0.30 eV), LiMgSO_4F (0.21 eV), and $\text{LiMgSeO}_4\text{F}$ (0.11 eV). Among these, only the composition LiMgSO_4F (no. 281119) is recorded in ICSD while the rest are novel compositions. Low ME values are clustered in P-containing samples for M^{3+} – T^{5+} pairing and Mg-containing samples for M^{2+} – T^{6+} pairing. For Sb^{5+} -containing samples which are relatively unstable according to the discontinuity in the lattice parameter trends shown in Figure 2, relatively high ME values were predicted (>0.60 eV).

The present NN approach that we employed has greatly aided us in significantly reducing the actual calculation time for acquiring a practical material property information (i.e., the ME values) that is directly useful for experimentalists. We would like to point out that although we can get the total energy, the optimized geometry, and some bulk properties (eg., born effective charges, band gaps, phonon frequencies, etc.) within a reasonable amount of time under the DFT approach, this is not the case when obtaining DFT activation energy values. Calculation of activation energy by NEB method (Figure 5a) needs a sufficiently expanded model cell (a $1 \times 2 \times 2$ supercell or 127-atoms cell) in order to significantly reduce the unphysical electrostatic interaction between periodic images. In addition, calculation of the saddle point of the ME profile by NEB involves bottleneck processes owing to its inherent sequential computation scheme; properly converged initial and final states for the mobile Li ion are always required as inputs in order to proceed and calculate for the energy of the intermediate state. At every NEB step, recurrent failed convergence can be encountered and this means that recalculations with modified convergence parameters will be necessary. Depending on the chemistry, calculations can also require large SCF iterations for electronic energy convergence. The NN modeling technique offers us a way to minimize, if not totally avoid, these issues with a reasonable trade-off in accuracy. To put it in perspective, the elapsed time needed just to generate the ME profile for path 1 in Figure 5a (127 atoms-cell) by NEB method is about ~ 38 h in a 2-core computation (in dedicated 3.10-GHz Intel Xeon E5–2687W 0 CPUs). Assuming similar energy convergence behavior for all 63 compositions, a complete ME evaluation would need ~ 2400

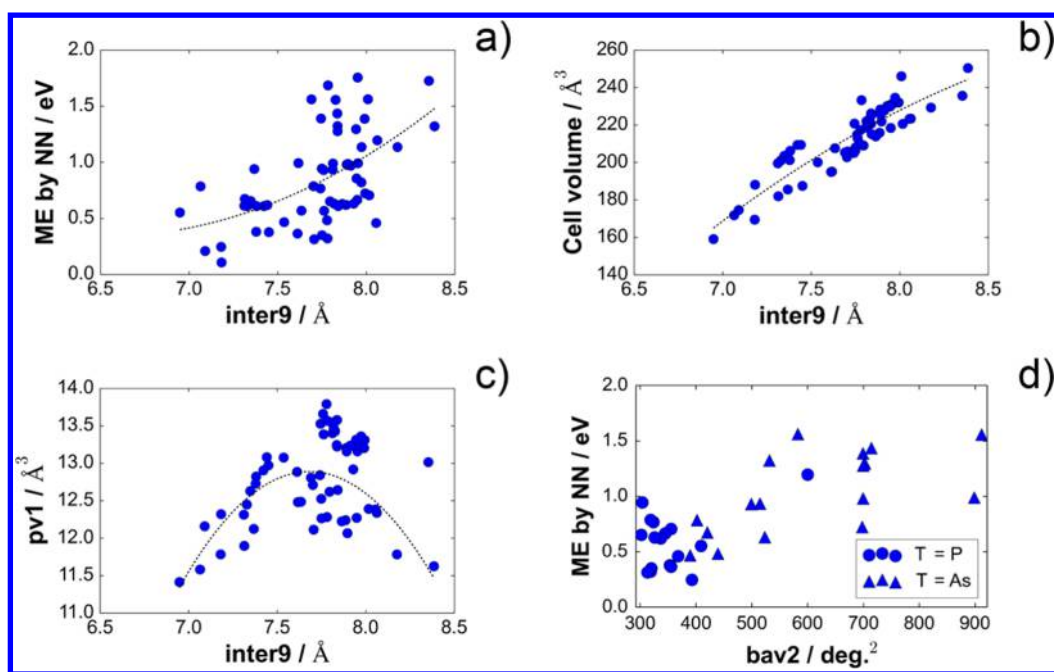


Figure 8. (a) ME trend vs the input variable inter9 (highest CI value), (b) cell volume trend vs inter9, (c) pv1 (second highest CI value) trend vs inter9, and (d) ME trend vs T cation type (as a bond covalency metric) and bav2 (as a local lattice distortion metric).

h (excluding the recalculation time expended for several compositions which encountered failed convergence). In the case of NN-based ME modeling, the total time needed from ME sampling by NEB up to model building is only ~ 1400 h (with 37 NEB-based ME values used for supervised learning). In this, model building only requires ~ 3.30 h (in a 3.40-GHz Intel Core i7-4770 CPU). In addition, the trained and validated NN model in our study also provides us with a self-contained platform for exploring explanatory variables for ME.

Although we focused mainly on one structure type (leading to a total of 63 compounds) due to our special interest with the favorite structure for battery application, we can also apply the NN concept to include many structure-types and thus, readily extending the search space that can be explored. This can be made possible if we use primitive, literature-based, and/or structure-independent predictors, eg., ionic radius, atomic mass, electronegativity, bond enthalpy, composition, etc. This modeling approach has been demonstrated successfully in one of our latest papers²⁶ and also in other high throughput studies such as by Meredith et al.²⁵ and Seko et al.⁵⁵

Model Interpretation and Variable Analysis. Aside from screening the LiMTO₄F search space, we attempted to rationalize the final ME prediction model and physically interpret the contributions of input variables.²⁶ First, we checked the highest CI-valued variable that led to ME increase, parameter inter9 (CI: +4.21). Its relationship with ME is shown in Figure 8a (the trend is fitted quadratically for visual guide, shown in a dashed curve). This variable represents the distance between end-member M cations of a cell-diagonal octahedral chain with three MO₄F₂ units vertex-linked via F ions (see Figure 6c). Its interaction with ME is determined to be nonlinear positive. From graph model analysis in Figure 3, we found out that this parameter is strictly increasing with cell volume ($R = 0.91$) as shown in Figure 8b, thus we can assign inter9 as a metric for cell size (cell volume was one of the original input variables in X but was removed due to multicollinearity). This appears to contradict with general

expectation, in which an increase in cell size should have led to a more open structure, so ME should have consequently followed a negative correlation with inter9. Instead, ME increases and rises sharply especially at inter9 $> \sim 7.7$ Å. To explain this ambiguity, the interaction of inter9 was analyzed together with the second highest CI variable that has a negative ME contribution, pv1 (CI: -3.64). The latter measures the polyhedral volume of the cage occupied by a Li ion, the trend is noted to be concave down as shown in Figure 8c (note that pv1 has a negative CI value so ME tends to decrease when it increases). It is observed that the onset of sharp ME increase with inter9 matches very well with the maxima of the pv1–inter9 trend. Also, M–T pairs with relatively larger cell do not necessarily lead to larger Li ion polyhedral cages. This could partly explain the ME trend for the second half of the inter9 data range. In addition, the trend of pv1 in the first half of the inter9 data range suggests that it has weak effect on ME for relatively small-volume cells. Only for inter9 $> \sim 7.7$ Å (or for relatively large-volume cells) that pv1 becomes increasingly important for ME lowering. This provides us a clue that for ME to decrease, pv1 should at least proportionately increase with inter9 (or cell size); we propose that the ratio between pv1 and inter9 could be related to the bottleneck size along the Li pathway. Looking at other variables to describe ME, we turned to previous DFT computational works related with polyanion groups, proposing bond covalency at the tetrahedral site and local lattice distortion as parameters that could directly affect ME.^{17,26,27,51} A strong covalency in the polyanion tends to minimize interaction between a mobile Li ion and the lattice, while a large local distortion around the Li transport pathway often adds an additional energy penalty to be surmounted by a moving Li ion. Based on these, we sorted the M–T data points with respect to T cation type and plotted their trend vs local distortion. Based from electronegativity difference, bond covalency in the TO₄ unit for M³⁺–T⁵⁺ and M²⁺–T⁶⁺ pairings generally follows the order P $>$ As $>$ Sb and Te $>$ Se $>$ Se, respectively.⁵⁶ For the local lattice distortion, we used the

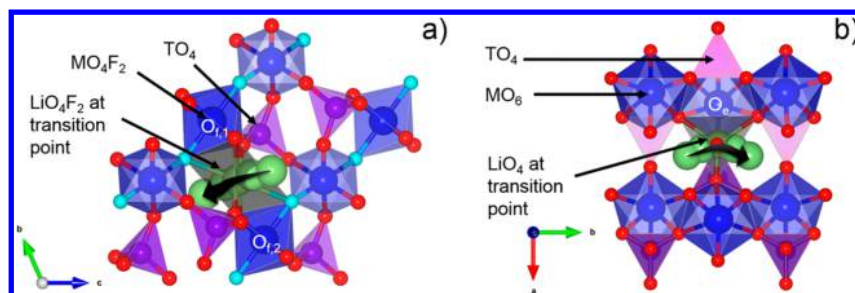


Figure 9. Local environments of the Li pathway bottleneck for the (a) tavorite-type (*P1*) LiMTO_4F ($M = \text{Al}^{3+}$, $T = \text{P}^{5+}$) and (b) olivine-type (*Pnma*) LiMTO_4 ($M = \text{Mg}^{2+}$, $T = \text{P}^{5+}$) structure. The transition point for the tavorite structure (along $[001]$) is an octahedron which is sandwiched in a face-sharing manner between two MO_4F_2 octahedra (O_{f1} , O_{f2}) while for the olivine structure (along $[010]$) is a tetrahedron which is edge-shared with an MO_6 octahedron (O_e).

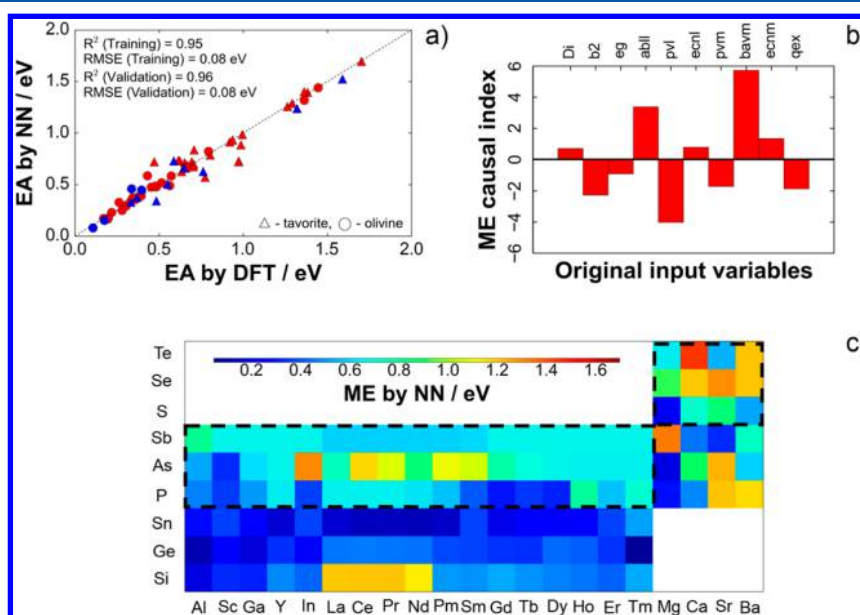


Figure 10. (a) Fitting quality of the final NN model for ME prediction for the tavorite LiMTO_4F + olivine LiMTO_4 system (for supervision, 60 NEB ME values were used and partitioned as 80% training and 20% validation). (b) Calculated causal index (CI) values of input neuron variables. (c) NN model-predicted ME sorted with respect to M and T cation. Enclosed within the dashed rectangle and square are tavorite compositions.

variation of bond angles formed by a cation with its surrounding anions. A large difference of bond angles within a cation-center polyhedron should lead to a more distorted local environment. Among the variables in the final model, we propose bav2 (CI: +1.2) as the suitable measure for local lattice distortion; bav2 gives the bond angle variance in the Li octahedron (see again Figure 6c). Figure 8d displays and isolates the effect of covalency in the T tetrahedron (for P^{5+} -based and As^{5+} -based compositions) and local lattice distortion with ME. The plot hints that a relatively stronger polyanion covalency and lower local distortion (for P^{5+}) favor lower ME. Moreover, a relatively weaker covalency (such as in As^{5+}) and a larger local distortion (i.e., large bav2) promote ME increase. Some compositions though demonstrate relatively large ME even with relatively small local distortion. Examples are $\text{Ho}^{3+}-\text{P}^{5+}$, $\text{Er}^{3+}-\text{P}^{5+}$, and $\text{Tm}^{3+}-\text{P}^{5+}$ pairs with ME values of 0.94, 0.77, and 0.79 eV, respectively. These compositions fall within the decreasing trend of pv1 in Figure 7c (with inter9 values of 7.75, 7.74, and 7.7 Å, respectively; pv1 values of 12.52, 12.84, and 12.71 Å³, respectively; and bav2 values of 304.2°, 324.2°, and 319.1°, respectively), we argue that the path bottleneck size in these compounds are not optimized enough for a minimized electron cloud overlap vis-à-vis the mobile Li ion.

Collectively, the final ME model suggests that in order achieve low ME for the tavorite system: (i) Li polyhedral cage size should be increased proportionately with cell volume (or inter9) for path bottleneck size optimization, (ii) the T cation should be strongly covalent (like P^{5+}), and (iii) the local distortion around the Li migration pathway should be minimized. These design guidelines should also be applicable to other structure-types with polyanion groups.

Activation Energy Screening with More than One Structure Type. As previously mentioned, when dealing with more than one structure type, the prediction model for a given target property (such as ME) would require an appropriate data representation for inferencing. In the present study, we can demonstrate this by using predictors that are independent not only of atom and cell axis ordering but also of rotations, translations and other symmetry operations. Some of these were, in fact, already included in the NN modeling approach discussed in the previous section—the intrapolyhedron parameters and the born effective charge parameters. Other predictors include void space parameters, average volume per oxygen atom, average atomic mass, and parameters extracted from the cumulative overall radial distribution function (27 variables; see Table S3 in the Supporting Information for the

complete list). We combined the present favorite (LiMTO_4F) data set with our previous olivine (LiMTO_4 , space group $Pnma$) data set, resulting to a total of 126 samples. Both structures obey 1D Li ion conduction which means that there is only one relevant path bottleneck or one characteristic ME attribute to consider. The difference in the local environment for the actual Li ion jump for both structures is displayed in Figure 9. The transition point in favorite (along [001]) is an octahedron which is sandwiched in a face-sharing manner between two MO_4F_2 octahedra (Figure 9a) while in olivine (along [010]) is a tetrahedron which is edge-shared with an MO_6 octahedron. The sample ME data set contains 60 DFT-calculated values (48% of the favorite+olivine chemical space; 37 ME values for favorite and 23 ME values for olivine).

Using the same data preprocessing and fitting procedure as with the single-structure-type NN fitting, the new ME prediction model for two structure types was successfully built with 10 input and 4 hidden layer neurons:

$$\begin{aligned} \text{ME} = & 0.568 - 0.496\tanh(v_1) - 0.371\tanh(v_2) \\ & - 0.415\tanh(v_3) - 0.450\tanh(v_4) \end{aligned} \quad (3)$$

where v_1 , v_2 , v_3 , and v_4 are the linear combinations of the input neuron variables (with trained coefficients) in the hidden layer (see Figure S6 and Table S4 in the Supporting Information). The estimated total computation time is ~ 1870 h (with 60 NEB calculations ME values for supervision). In this, only ~ 2.60 h was needed for the model building phase (also with 300 000 model evaluations and using the same CPU as with the previous single structure NN fitting). For the full ME evaluation by NEB method for both structures, at least ~ 3650 h would have been needed. The fitting quality is shown in Figure 10a with metrics as follows: $R^2(\text{Training}) = 0.95$, $\text{RMSE}(\text{Training}) = 0.08$ eV, $R^2(\text{Validation}) = 0.96$, and $\text{RMSE}(\text{Validation}) = 0.08$ eV. The corresponding causal indices of the predictors for the generalized ME prediction model is shown in Figure 10b. It is evident in the plot that ME generally increases when bavm (bond angle variance of the M octahedron, CI: +5.73) and abl1 (average bond length of the Li octahedron, CI: +3.38) increase. In contrast, ME decreases when pvl (polyhedral volume of the Li octahedron, CI: -4.03) and b2 (M cation Born effective charge, CI: -2.28) increase. The highest CI valued variable bavm , which represents the degree of octahedral distortion within the immediate vicinity of the Li pathway for both structures, has been determined to be a measure for local distortion which, when increased, leads to an energy penalty for Li ion migration for the olivine structure.²⁷ Interestingly as well, pvl (volume of Li octahedron) has the same causal index sign (negative) as in the final model for single-structure-type NN fitting (see pv1 in Figure 6b). Predictors Di (largest included sphere in the structure void space with Li removed), eg (electronic band gap), and ecn1 (effective coordination number of Li ion) show very small contributions. Figure 10c shows the ME mapping with respect to M–T pairings for the favorite LiMTO_4F (enclosed within dashed rectangle and square) + olivine LiMTO_4 system. For the favorite space, the ME map is in good agreement with the results based from single-structure-type fitting (see Figure 7). The predicted low ME compositions for the olivine structure are also very close to our previous results,^{26,27} particularly for Ge- and Sn-containing compounds. Based on these results, intrapolyhedron parameters and effective ion charges can be regarded as good descriptors for structure-unconstrained target

property modeling. This harmonizes well with the findings of Daams and Villars in which they found out that already 80% of 5000 structural prototypes (in ICSD,³³ 9136 structure prototypes are available as of April 2015) can already be readily accounted for by using only few atomic environment types (in our study, we derived the predictor variables from octahedral and tetrahedral environment types, both in their regular and distorted shapes).⁵⁷ Willighagen et al. have also included the atomic charge information to describe crystal structures.⁵⁸

In principle, the selected predictor variables in the favorite + olivine NN model building have enabled for data representation which can be applied to a much larger pool of structure types. Once the model has learned many local atomic environments, predictions for structure types that are not explicitly included in the training step should be possible. We intend to focus on this point in our future study. For example, in ICSD (as of April 2015), there are ~ 5000 Li-containing compounds that we can start the material property screening with and within this data set alone, there is a large number of structure prototypes (a subset of ~ 9100 structure prototypes) and a vast number of compositions that can be evaluated (especially with the inclusion of ionic substitutions). By training an NN model from a random subset of structure types, prediction for “excluded” structure types should be possible since the latter are expected (in most cases) to be located at intermediate points in the feature (predictor variable) space defined by the “included” structure types. In this way, new training and new DFT ME calculations are not necessary if one is faced with a completely new structure type.

CONCLUSIONS

In summary, we succeeded in formulating an efficient and informative NN-based functional for ME for the favorite LiMTO_4F composition space. Candidate solid electrolytes with $\text{ME} \leq 0.30$ eV are LiGaPO_4F (0.25 eV), LiGdPO_4F (0.30 eV), LiDyPO_4F (0.30 eV), LiMgSO_4F (0.21 eV), and $\text{LiMgSeO}_4\text{F}$ (0.11 eV). The use of an informatics-based model can be employed, together with chemical intuition, to elucidate complex (and usually competing) effects of different parameters controlling a target property. Important variables that were determined are Li pathway bottleneck size, polyanion covalency and local lattice distortion. A means to predict a given target property (eg., ME) for more than one structure type built mainly using intrapolyhedron parameters was also demonstrated.

ASSOCIATED CONTENT

Supporting Information

Description and code assignments for original predictors taken from DFT-relaxed favorite-type LiMTO_4F and olivine-type LiMTO_4 , loading plot from PCA of the original predictor set, final NN ME model structure and weights for one structure type and two structure types, fitting quality vs number of hidden layer neurons, fitting results of ME model based from PLS regression. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/ci500752n.

AUTHOR INFORMATION

Corresponding Author

*E-mail: JALEM.Randy@nims.go.jp. Tel.: +81-29-860-4953. Fax: +81-29-860-4981.

Present Address

[†]R.J.: National Institute for Materials Science (NIMS), Global Research Center for Environment and Energy based on Nanomaterials Science (GREEN), Namiki 1-1, Tsukuba, Ibaraki, Japan, 305-0044.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The present work was partially supported by JST, PRESTO program, and Ministry of Education Culture, Sports, Science and Technology, Japan (MEXT), program “Elements Strategy Initiative to Form Core Research Center” (since 2012). Crystal structures were drawn with VESTA.⁵⁹

REFERENCES

- (1) Kamaya, N.; Homma, K.; Yamakawa, Y.; Hirayama, M.; Kanno, R.; Yonemura, M.; Kamimaya, T.; Kato, Y.; Hama, S.; Kawamoto, K.; Mitsui, A. A Lithium Superionic Conductor. *Nat. Mater.* **2011**, *10*, 682–686.
- (2) Murugan, R.; Thangadurai, V.; Weppner, W. Fast Lithium Ion Conduction in Garnet-Type $\text{Li}_7\text{La}_3\text{Zr}_2\text{O}_{12}$. *Angew. Chem., Int. Ed.* **2007**, *46*, 7778–7781.
- (3) Thangadurai, V.; Weppner, W. $\text{Li}_6\text{Ala}_2\text{Ta}_2\text{O}_{12}$ (A = Sr, Ba): Novel Garnet-Like Oxides for Fast Lithium Ion Conduction. *Adv. Funct. Mater.* **2005**, *15*, 107–112.
- (4) Inaguma, Y.; Liqun, C.; Itoh, M.; Nakamura, T.; Uchida, T.; Ikuta, H.; Wakihara, M. High Ionic Conductivity in Lithium Lanthanum Titanate. *Solid State Commun.* **1993**, *86*, 689–693.
- (5) Alpen, U. V.; Schulz, H.; Talat, G. H.; Böhm, H. One-Dimensional Cooperative Li-Diffusion in β -Eucryptite. *Solid State Commun.* **1977**, *23*, 911–914.
- (6) Hong, H. Y.-P. Crystal Structure and Ionic Conductivity of $\text{Li}_{14}\text{Zn}(\text{GeO}_4)_4$ and Other New Li^+ Superionic Conductors. *Mater. Res. Bull.* **1978**, *13*, 117–124.
- (7) Goodenough, J. B.; Hong, H. Y.-P.; Kafalas, J. A. Fast Na^+ -Ion Transport in Skeleton Structures. *Mater. Res. Bull.* **1976**, *11*, 203–220.
- (8) Kerr, T. A.; Gaubicher, J.; Nazar, L. F. Highly Reversible Li Insertion at 4 V in ϵ - VOPO_4/α - LiVOPO_4 Cathodes. *Electrochem. Solid State Lett.* **2000**, *3*, 460–462.
- (9) Marx, N.; Croguennec, L.; Carlier, D.; Wattiaux, A.; Le Cras, F.; Suard, E.; Delmas, C. The Structure of Tavorite $\text{LiFePO}_4(\text{OH})$ from Diffraction and GGA + U Studies and Its Preliminary Electrochemical Characterization. *Dalton Trans.* **2010**, *39*, 5108–5116.
- (10) Gover, R. K. B.; Burns, P.; Bryan, A.; Saidi, M. Y.; Swoyer, J. L.; Barker, J. LiVOPO_4F : A New Active Material for Safe Lithium-Ion Batteries. *Solid State Ionics* **2006**, *177*, 2635–2638.
- (11) Zhou, F.; Zhao, X.; Dahn, J. R. Reactivity of Charged LiVOPO_4F with 1 M LiPF_6 EC:DEC Electrolyte at High Temperature as Studied by Accelerating Rate Calorimetry. *Electrochem. Commun.* **2009**, *11*, 589–591.
- (12) Recham, N.; Chotard, J. N.; Dupont, L.; Delacourt, C.; Walker, W.; Armand, M.; Tarascon, J. M. A 3.6 V Lithium-Based Fluorosulphate Insertion Positive Electrode for Lithium-Ion Batteries. *Nat. Mater.* **2010**, *9*, 68–74.
- (13) Recham, N.; Chotard, J. N.; Jumas, J. C.; Laffont, L.; Armand, M.; Tarascon, J. M. Ionothermal Synthesis of Li-Based Fluorophosphates Electrodes. *Chem. Mater.* **2010**, *22*, 1142–1148.
- (14) Tripathi, R.; Gardiner, G. R.; Islam, M. S.; Nazar, L. F. Alkali-ion Conduction Paths in LiFeSO_4F and NaFeSO_4F Tavorite-Type Cathode Materials. *Chem. Mater.* **2011**, *23*, 2278–2284.
- (15) Mueller, T.; Hautier, G.; Jain, A.; Ceder, G. Evaluation of Tavorite-Structured Cathode Materials for Lithium-Ion Batteries Using High-Throughput Computing. *Chem. Mater.* **2011**, *23*, 3854–3862.
- (16) Sebastian, L.; Gopalakrishnan, J.; Piffard, Y. Synthesis, Crystal Structure and Lithium Ion Conductivity of LiMgFSO_4 . *J. Mater. Chem.* **2002**, *12*, 374–377.
- (17) Jaleel, R.; Nakayama, M.; Kasuga, T. Lithium Ion Conduction in Tavorite-Type LiMXO_4F (M = Al, P, Mg, S) Candidate Solid Electrolyte Materials. *Solid State Ionics* **2014**, *262*, 589–592.
- (18) Ong, S. P.; Mo, Y.; Richards, W. D.; Miara, L.; Lee, H. S.; Ceder, G. Phase Stability, Electrochemical Stability and Ionic Conductivity of the $\text{Li}_{10\pm1}\text{MP}_2\text{X}_{12}$ (M = Ge, Si, Sn, Al or P, and X = O, S or Se) Family of Superionic Conductors. *Energy Environ. Sci.* **2013**, *6*, 148–156.
- (19) Hautier, G.; Jain, A.; Chen, H.; Moore, C.; Ong, S. P.; Ceder, G. Novel Mixed Polyanions Lithium-Ion Battery Cathode Materials Predicted by High-Throughput *Ab Initio* Computations. *J. Mater. Chem.* **2011**, *21*, 17147–17153.
- (20) Clark, J. M.; Eames, C.; Reynaud, M.; Rousse, G.; Chotard, J. N.; Tarascon, J. M.; Islam, M. S. High Voltage Sulphate Cathodes $\text{Li}_2\text{M}(\text{SO}_4)_2$ (M = Fe, Mn, Co): Atomic-Scale Studies of Lithium Diffusion, Surfaces and Voltage Trends. *J. Mater. Chem. A* **2014**, *2*, 7446–7453.
- (21) Nakayama, M.; Kotobuki, M.; Munakata, H.; Nogami, M.; Kanamura, K. First-Principles Density Functional Calculation of Electrochemical Stability of Fast Li Ion Conducting Garnet-Type Oxides. *Phys. Chem. Chem. Phys.* **2012**, *14*, 10008–10014.
- (22) Kirklin, S.; Meredig, B.; Wolverton, C. High-Throughput Computational Screening of New Li-Ion Battery Anode Materials. *Adv. Energy Mater.* **2013**, *3*, 252–262.
- (23) Stecher, T.; Bernstein, N.; Csányi, G. Free Energy Surface Reconstruction from Umbrella Samples Using Gaussian Process Regression. *J. Chem. Theory Comput.* **2014**, *10*, 4079–4097.
- (24) Bartók, A. P.; Gillan, M. J.; Manby, F. R.; Csányi, G. Machine-Learning Approach for One- and Two-Body Corrections to Density Functional Theory: Applications to Molecular and Condensed Water. *Phys. Rev. B* **2013**, *88*, 054104.
- (25) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial Screening for New Materials in Unconstrained Composition Space with Machine Learning. *Phys. Rev. B* **2014**, *89*, 094104.
- (26) Jaleel, R.; Nakayama, M.; Kasuga, T. An Efficient Rule-Based Screening Approach for Discovering Fast Lithium Ion Conductors Using Density Functional Theory and Artificial Neural Networks. *J. Mater. Chem. A* **2014**, *2*, 720–734.
- (27) Jaleel, R.; Aoyama, T.; Nakayama, M.; Nogami, M. Multivariate Method-Assisted *Ab Initio* Study of Olivine-Type LiMXO_4 (Main Group M^{2+} – X^{5+} and M^{3+} – X^{4+}) Compositions as Potential Solid Electrolytes. *Chem. Mater.* **2012**, *24*, 1357–1364.
- (28) Hautier, G.; Fischer, C.; Ehrlacher, V.; Jain, A.; Ceder, G. Data Mined Ionic Substitutions for the Discovery of New Compounds. *Inorg. Chem.* **2011**, *50*, 656–663.
- (29) Fujimura, K.; Seko, A.; Koyama, Y.; Kuwabara, A.; Kishida, I.; Shitara, K.; Fisher, C. A. J.; Moriwake, H.; Tanaka, I. Accelerated Materials Design of Lithium Superionic Conductors Based on First-Principles Calculations and Machine Learning Algorithms. *Adv. Energy Mater.* **2013**, *3*, 980–985.
- (30) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1998**, *59*, 1758–1775.
- (31) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (32) Csonka, G. I.; Perdew, J. P.; Ruzsinszky, A.; Philipsen, P. H. T.; Lebègue, S.; Paier, J.; Vydrov, O. A.; Ángyán, J. G. Assessing the Performance of Recent Density Functionals for Bulk Solids. *Phys. Rev. B* **2009**, *79*, 155107.
- (33) ICSD, Inorganic Crystal Structure Database; <http://www.fiz-karlsruhe.de/icsd.html> (accessed April 30, 2015).
- (34) Gonze, X.; Lee, C. Dynamical Matrices, Born Effective Charges, Dielectric Permittivity Tensors, and Interatomic Force Constants from Density-Functional Perturbation Theory. *Phys. Rev. B* **1997**, *55*, 10355–10368.

- (35) Jonsson, H.; Mills, G.; Jacobsen, K. M. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific: Singapore, 1998.
- (36) Agnarsson, G.; Greenlaw, R. *Graph Theory: Modeling, Applications, and Algorithms*, Pearson International ed., Prentice Hall, 2007.
- (37) Pearson, K. Notes on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* **1895**, 58, 240–242.
- (38) Newman, M. E. J. Modularity and Community Structure in Networks. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, 103, 8577–8582.
- (39) Hagberg, A. A.; Schult, D. A.; Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Pasadena, CA, Aug 19–24, 2008; pp 11–15.
- (40) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*, San Jose, CA, May 17–20, 2009.
- (41) LeCun, Y. A Theoretical Framework for Back-Propagation. In *Proceedings of the 1988 Connectionist Models Summer School*; Touretzky, D., Hinton, G., Sejnowski, T., Eds.; CMU: Pittsburgh, Pa, 1988; pp 21–28.
- (42) The learning rate η is used to compute the direction in which each node weight is changed to arrive at a more correct output. A value closer to 1 can lead to faster convergence, but it can also increase the risk of getting trapped at a local minimum. In this study, $\eta = 0.1$ was used.
- (43) Hassoun, M. *Fundamentals of Artificial Neural Networks*; M.I.T. Press: Cambridge, MA, 1995.
- (44) Zhang, H.; Wu, W.; Liu, F.; Yao, M. Boundedness and Convergence of Online Gradient Method with Penalty for Feedforward Neural Networks. *IEEE T. Neural Networ.* **2009**, 20, 1050–1054.
- (45) Kim, S.-J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An Interior-Point Method for Large-Scale l_1 -Regularized Least Squares. *IEEE J. Sel. Topics Signal Process.* **2007**, 1, 606–617.
- (46) Fukumizu, K.; Amari, S. Local Minima and Plateaus in Hierarchical Structures of Multilayer Perceptrons. *Neural Networks* **2000**, 13, 317–327.
- (47) Baba, K.; Enbutu, I.; Yoda, M. Explicit Representation of Knowledge Acquired from Plant Historical Data Using Neural Network. *Proceedings of the International Joint Conference on Neural Networks*, San Diego, CA, June 17–21, 1990; Vol. 3, pp 155–160.
- (48) *New Features in JMP 11*; SAS Institute Inc.: Cary, NC, 2013.
- (49) Thornton, G. A Neutron Diffraction Study of α -Sb₂O₄. *Acta Crystallogr.* **1977**, B33, 1271–1273.
- (50) Hong, H. Y. -P. The Two-Dimensional Tunnel Structures of K₃Sb₅O₁₄ and K₂Sb₄O₁₁. *Acta Crystallogr.* **1974**, B30, 945–952.
- (51) Mercier, R.; Bernard, J. D.; et al. Structure Cristalline de Sb₂O₃·3SO₃. *Acta Crystallogr.* **1976**, B32, 2787–2791.
- (52) Michel, C.; Groult, D.; Raveau, B. Sur de Nouveaux Pyrochlores ASbWO₆ (A = K, Rb, Cs, Tl). *Mater. Res. Bull.* **1973**, 8, 201–210.
- (53) Rajkumar, T.; Nakayama, M.; Nogami, M. *Ab Initio* Prediction for the Ionic Conduction of Lithium in LiInSiO₄ and LiInGeO₄ Olivine Materials. *Solid State Commun.* **2010**, 150, 693–696.
- (54) Wentzell, P. D.; Montoto, L. V. Comparison of Principal Components Regression and Partial Least Squares Regression Through Generic Simulations of Complex Mixtures. *Chemometr. Intell. Lab.* **2003**, 65, 257–279.
- (55) Seko, A.; Maekawa, T.; Tsuda, K.; Tanaka, I. Machine Learning with Systematic Density-Functional Theory Calculations: Application to Melting Temperatures of Single- and Binary-Component Solids. *Phys. Rev. B* **2014**, 89, 054303.
- (56) Alfred, A. L. Electronegativity Values from Thermochemical Data. *J. Inorg. Nucl. Chem.* **1961**, 17, 215–221.
- (57) Daams, J.; Villars, P. Atomic environments in relation to compound prediction. *Eng. Appl. Artif. Intel.* **2000**, 13, 507–511.
- (58) Willighagen, E. L.; Wehrens, R.; Verwer, P.; de Gelder, R.; Buydens, L. M. C. Method for the computational comparison of crystal structures. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, 61, 29–36.
- (59) Momma, K.; Izumi, F. VESTA: A Three-Dimensional Visualization System for Electronic and Structural Analysis. *J. Appl. Crystallogr.* **2008**, 41, 653–658.