JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Received 23 August 2018
Accepted 31 December 2018

Edited by Th. Proffen, Oak Ridge National Laboratory, USA

Keywords: bond-valence theory; machine learning; high throughput; Li/Na-ion conductors.

Supporting information: this article has supporting information at journals.iucr.org/j

An investigation of the structural properties of Li and Na fast ion conductors using high-throughput bond-valence calculations and machine learning

Nebil A. Katcho,[a]* Jesús Carrete,[b] Marine Reynaud,[c] Gwenaëlle Rousse,[d] Montse Casas-Cabanas,[c] Natalio Mingo,[e] Juan Rodríguez-Carvajal[a]* and Javier Carrasco[c]*

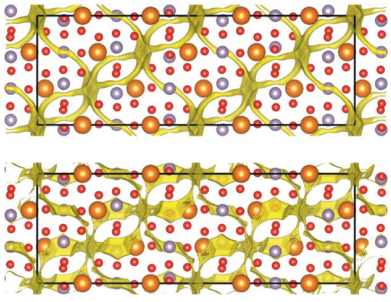
[a]Institute Laue–Langevin, 71 Avenue des Martyrs, Grenoble, 38042, France, [b]Institute of Materials Chemistry, TU Wien, Vienna, A-1060, Austria, [c]CIC EnergiGUNE, Albert Einstein 48, Miñano, Alava, 01510, Spain, [d]Chimie du Solid et de l'Energie, UMR 8260, Collège de France, Paris, 75231, France, and [e]LITEN, CEA-Grenoble, Grenoble, 38054, France. *Correspondence e-mail: katcho@ill.fr, rodriguez-carvajal@ill.fr, jcarrasco@cicenergigune.com

Progress in energy-related technologies demands new and improved materials with high ionic conductivities. Na- and Li-based compounds have high priority in this regard owing to their importance for batteries. This work presents a high-throughput exploration of the chemical space for such compounds. The results suggest that there are significantly fewer Na-based conductors with low migration energies as compared to Li-based ones. This is traced to the fact that, in contrast to Li, the low diffusion barriers hinge on unusual values of some structural properties. Crystal structures are characterized through descriptors derived from bond-valence theory, graph percolation and geometric analysis. A machine-learning analysis reveals that the ion migration energy is mainly determined by the global bottleneck for ion migration, by the coordination number of the cation and by the volume fraction of the mobile species. This workflow has been implemented in the open-source *Crystallographic Fortran Modules Library* (*CrysFML*) and the program *BondStr*. A ranking of Li- and Na-based ionic compounds with low migration energies is provided.

1. Introduction

The rechargeable Li-ion battery – with higher energy density than its predecessors – has been a key technological enabler for the rise of portable electronics. Most importantly, looking ahead, battery technology offers great potential to enable a future world powered by clean energy sources. From the electrification of transportation (*e.g.* electric vehicles) to the support of smart grids (*e.g.* storing wind and solar energy produced during off-peak hours), rechargeable batteries are called on to create a new energy paradigm. However, for this scenario to materialize, advances beyond the current state of the art in battery performance, safety and cost are necessary.

A salient goal of research in batteries is the discovery of new materials or novel properties of existing compounds. In this quest, the traditional approach is to focus on archetypal compounds in which a desirable property was first observed, stimulating further investigations. For example, many good Li-ion solid electrolyte oxides and sulfides crystallize in a few crystal structures: layered compounds (Mizushima *et al.*, 1980), olivine (Padhi *et al.*, 1997), perovskites (Stramare *et al.*, 2003), NASICON-like (Martínez-Juárez *et al.*, 1998) and LISICON-like (Rodger *et al.*, 1985) (*i.e.* resembling Na and Li super-ionic conductors), garnets (Thangadurai *et al.*, 2014), and argyrodite (Deiseroth *et al.*, 2008). Similarly, most Li-ion cathode materials belong to a handful of structural families (Whittingham,



© 2019 International Union of Crystallography

2004): layered oxides, spinel and olivine. Yet this trial-and-error exploratory research based on extrapolating known solutions to new compositional spaces has high demands in terms of synthesis and characterization times that hinder the emergence of new, compounds that do not fit the traditional model.

To mitigate this issue and accelerate the process of materials discovery, computational modelling is emerging as a powerful complementary tool. Thanks to the proven predictive power of quantum chemistry methods together with the spectacular growth of computational resources, computer simulation is nowadays able to bring valuable insight for understanding the structure, properties and function of technological materials (Hautier *et al.*, 2012). In this regard, high-throughput screening of materials databases using first-principles simulation approaches has a demonstrated successful track record of guiding advances in materials science [see Jain *et al.* (2016) for a recent review], including areas as diverse as heterogeneous catalysis (Greeley *et al.*, 2002), bulk crystal structure prediction (Curtarolo *et al.*, 2003; Meredig & Wolverton, 2010) and thermoelectricity (Carrete *et al.*, 2014). In the domain of batteries, first-principles simulations based on density functional theory (DFT) have proven to be a useful method to understand the mechanisms of electroactive materials (Meng & Dompablo, 2009; Ceder, 2010; Islam & Fisher, 2014) and, in combination with high-throughput searches, are widely viewed as a promising approach to proposing new materials (Hautier *et al.*, 2011; Cheng *et al.*, 2015; Qu *et al.*, 2015; Kirklin *et al.*, 2013; Husch & Korth, 2015; Schütter *et al.*, 2015; Borodin *et al.*, 2015; Kireeva & Pervov, 2017).

Armed with these theoretical methods, an effective assessment of candidate materials first requires choosing relevant screening criteria. Keeping in mind that the chief fundamental process that enables the functioning of most of today's rechargeable batteries is the reversible insertion of ions into solids, ionic conductivity emerges as a paramount criterion. High ionic conductivity does not guarantee high electrochemical performance of the material, but it is a necessary condition to screen out materials with unwanted slow kinetics.

The theoretical evaluation of ionic conductivities involves identifying ion migration paths and computing their corresponding energy barriers. These can be estimated from DFT energy landscapes using transition state finders (such as the nudged elastic band method) or by simulating the migration by molecular dynamics (Deng *et al.*, 2016, 2017). However, it is unfeasible to tackle the study of ionic conductivity for the high-throughput screening of tens of thousands of compounds purely with DFT owing to the high computational cost of such an endeavour. Instead, one alternative approach is to prescreen candidate materials by a less accurate but faster method before undertaking detailed DFT calculations. The bond-valence (BV) method (Brown, 2009) has become particularly appropriate for this task since Adams & Rao (2011, 2014) reworked it into an effective local force field by introducing a Morse-type potential together with a screened Coulomb potential. With this approach, it is possible to sample the bond-site energy of the mobile ion throughout the unit cell in a few seconds. This enables the fast construction of bond-valence energy landscapes (BVELs) and therefore direct exploration of ion migration pathways, which can be visualized as regions enclosed by energy isosurfaces. The threshold energy ($E_{\rm th}$) for which these isosurfaces start to percolate the unit cell in at least one dimension can be taken, in general, as a rough estimate of the activation energy for ion migration ($E_{\rm m}$).

Despite the simplicity of the BV method, its straightforward application to high-throughput schemes has some shortcomings. First, oxidation states, which are the main input of the BV method, are not always specified for every entry of common crystal structure databases such as the Inorganic Crystal Structure Database (ICSD; http://www2.fiz-karlsruhe.de/icsd_home.html). Second, although there is computer software available to perform BVEL calculations [*e.g.* *BondStr* within the *FullProf* suite (Rodriguez Carvajal, 1993); see the manuals of the programs *FormalCharges* and *BondStr* in the supporting information (SI), the determination of $E_{\rm th}$ for each investigated compound usually requires tedious time-consuming visual inspection of three-dimensional energy landscapes. In order to facilitate this task, Sale & Avdeev (2012) used the *Materials Studio* package (Accelrys, San Diego, California, USA) to automate the determination of the threshold energies by developing a computer program called *3DBVSMAPPER*. This implementation has some practical drawbacks though. For example, the user must specify the oxidation states, and the program provides $E_{\rm th}$ but not the ion migration energy ($E_{\rm m}$), which in some cases can differ from $E_{\rm th}$, as shown later. Finally, *3DBVSMAPPER* depends on a piece of proprietary software. *3DBVSMAPPER* has been applied to the screening of Li-, Na-, K-, Cu- and Ag-based materials (Avdeev *et al.*, 2012). More recently, some other screening studies based on BV calculations have appeared, using their own implementations and only focused on Li-based materials (Gao *et al.*, 2014; Xao *et al.*, 2015; Xiao *et al.*, 2016; Nakayama *et al.*, 2016). In all of these previous screening studies, the goal was to identify new promising ion conductor candidates. Despite the huge quantity of data that some of these studies have generated, there has been no attempt to carry out an exhaustive analysis of those data to understand correlations between BV mismatches or BV energies and the structural and chemical properties of the materials. This stands in contrast with the importance of developing a fundamental understanding to establish a universal guide for fast ion conductors, whose lack was acknowledged, for instance, in a recent review of solid electrolytes (Bachman *et al.*, 2016). As BV theory describes ion diffusion reasonably well and can generate a lot of data in a very short time, it is an ideal tool to address this question in a first approximation. The data analysis can identify and quantify some of the fundamental $E_{\rm m}$ and crystal structure correlations, as well as providing a quantitative understanding of the limitations of the BV theory in describing ion diffusion in solids.

In this work we first report the technical details of the automation of the BV calculations. We describe the two algorithms required for the automation: (i) the automatic

assignment of oxidation states given the crystal structure and (ii) the determination of the threshold value at which the energy isosurface starts to percolate the unit cell and the correction required in the cases where E_{th} does not equal E_m. These algorithms have been implemented using *CrysFML* (https://forge.epn-campus.eu/projects/crysfml), an open-source library of Fortran modules (see the program manuals in the SI) which is widely used in crystallographic applications. The methodology described is then used to scan Li and Na compounds from the ICSD. We provide a ranking of compounds classified as a function of their computed E_m. The correlation between E_m and the crystal structure is studied by computing the correlation factors between E_m and a broad range of simple structural descriptors. To this end, we show that the bottleneck of the ion diffusion path can be retrieved even more easily just by using a simple hard-sphere model, with the same methodology being used to compute E_{th}. In addition to the bottleneck, we introduce several structural descriptors derived from a Voronoi partition of the crystal space. Additionally, we use machine learning to evaluate the predictive capabilities of the structural descriptors used in this work. Finally, we focus on those descriptors that are more strongly correlated with E_m and we show that they take rather unusual values for the best Na-ion conductors, contrary to what is observed in the Li case.

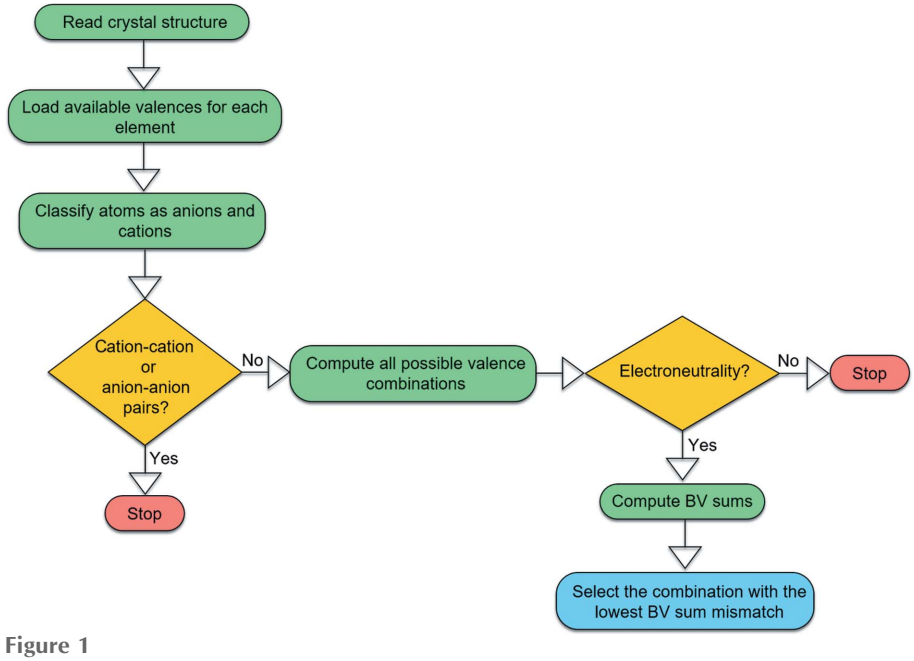


Figure 1
Flowchart of the algorithm used to assign oxidation states.

2. Methods

2.1. BV calculations

BV calculations are carried out using the program *BondStr* (https://forge.epn-campus.eu/projects/crysfml/repository/show/Program_Examples/BondStr), built with *CrysFML*. This code implements the BV-based force-field approach developed by Adams & Rao (2014). For oxide compounds, we use the Morse-type potential parameters supplied by those authors. The parameters for the remaining compounds are derived from the softBV model (Adams, 2001; Chen & Adams, 2017). BVELs are computed using a grid resolution of 0.1 Å. This choice ensures an energy convergence of a few hundredths of an electron volt. As is discussed in Section 3, this energy is small compared with the expected overestimation of the diffusion energy barrier by this approximation. The BV cut-off radius is set to 8 Å. In order to automate the BV calculations, we have implemented the following two tasks in *CrysFML*.

2.1.1. Assignment of oxidation states. We have developed an algorithm based on a BV sum (BVS) approach, which is sketched in Fig. 1. The algorithm is implemented in the program *FormalCharges* (see the manual in the SI). After reading the crystal structure, the program loads the possible valences for each element of the structure. As this algorithm is devised to be used in the framework of BV theory, the available valences for each element of the periodic table are those for which a BV parametrization exists. Elements with only positive oxidation states are automatically classified as cations. If negative oxidation states are possible, they are classified as potential anions in a first stage. In a second stage, first nearest neighbours are computed. When a cation and potential anion pair is found, the potential anion is confirmed as such. If a potential cation–anion pair is found, the atom with lower (higher) electronegativity is classified as cation (anion). If anion–anion or cation–cation pairs are confirmed, the algorithm stops and no assignment of oxidation states is done. In the next step, all possible combinations of atomic valences are computed, and those, if any, which fulfil electroneutrality are selected. Finally, if more than one electroneutral combination has been found, the more plausible combination is chosen. A quantitative measurement of this plausibility is given by the absolute BV sum mismatch (ΔV) (Brown, 2009), defined as

$$|\Delta V| = \sum_M \left| \sum_X s_{M-X} - V_{id}(M) \right|, \qquad (1)$$

where M runs over all the ions of the system, X runs over all adjacent counter-ions of M, s_{M-X} is the predicted bond valence for the $M-X$ bond and V_{id} is the nominal valence of ion M. This quantity measures the deviation of the predicted valences from the nominal valences. The lower the absolute BV sum mismatch, the more plausible is the combination. Therefore, the combination with the lowest BV mismatch is selected.

2.1.2. Calculation of E_{th} and E_m. E_{th} is the energy at which the isosurface corresponding to the bond-site energy of the

mobile ion starts to percolate the unit cell. The percolation of an isosurface through a parallelepiped can be easily analysed by using graph theory. To this end, the isosurface is computed by using the marching cubes algorithm (Lorensen & Cline, 1987), which builds the isosurface as a triangular mesh. The vertices and sides of the triangular mesh are considered as the nodes and edges of the graph. The percolation algorithm proceeds in two steps. First, it extracts the connected components of the graph without applying periodic boundary conditions (PBC). A connected component is an equivalence class with respect to the relation of mutual accessibility. They are extracted using a depth-first search algorithm (Cormen *et al.*, 2009). In the second step, PBC are applied. The number of connected components usually decreases with the addition of new bonds between different images of the cell. Those final connected components are analysed to check if at least one of them percolates the unit cell. If there is percolation, then there must be a sequence of bonds connecting two nodes related by translational symmetry. Fig. 2 sketches how the algorithm works in the particular case of two dimensions. The migration energy, $E_{\rm m}$, corresponds to $E_{\rm th}$ minus the lowest energy point of the migration path. Previous screening tacitly assumed the equivalence between $E_{\rm th}$ and $E_{\rm m}$. This is generally true because the mobile ion is usually located at the absolute minimum site energy. However, there are cases where this equivalence does not hold, such as for example compounds where the mobile ion can be located at several crystallographic sites but not all of these sites participate in the ionic diffusion. $Li_{10}GeP_2S_{12}$ (Kamaya *et al.*, 2011), one of the best Li-ion conductors, is a paradigmatic example. In this compound, the Li atoms are located at 4*c*, 4*d*, 8*f* and 16*h* Wyckoff positions, with bond-site energies E_{4c}, E_{4d}, E_{8f} and E_{16h} of 0.75, 0.00, 0.71 and 0.54 eV, respectively (E_{4d} corresponds to the absolute minimum of the energy and it is taken as reference). $E_{\rm th}$ is 0.97 eV. As only the 8*f* and the 16*h* sites are involved in Li diffusion and 16*h* has lower energy, if we approximate the lowest energy of the diffusion path by the site energy of the 16*h* position, $E_{\rm m}$ is $E_{\rm th} - E_{16h} = 0.43$ eV, *i.e.* 0.59 eV lower than $E_{\rm th}$. We have therefore included this correction in our screening. We notice that in a few cases the correction is quite large, around 1 eV or even greater. In these cases, there is usually a disagreement between the reported atomic coordinates for the alkali atoms and the coordinates predicted by BV theory.

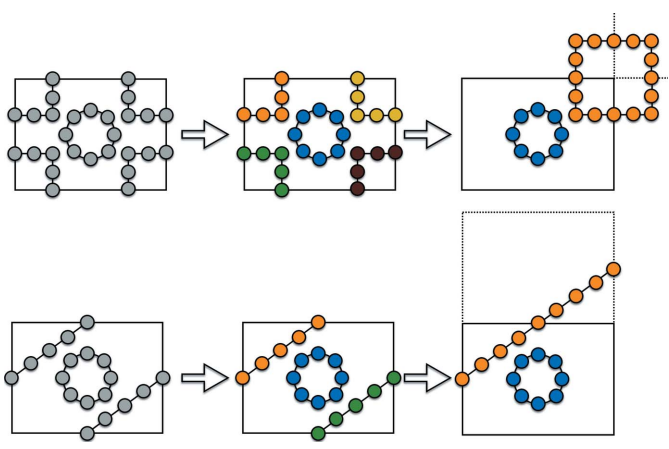


Figure 2
Percolation algorithm steps in two dimensions. Left: isoline identification. Middle: extraction of connected components (in colour) without PBC. Right: connected components after applying PBC. Top: isolines crossing unit-cell boundaries without percolation. Bottom: isolines crossing unit-cell boundaries with percolation.

2.2. Compound selection

In this work, we screen the 2013 release of the ICSD. We initially considered all Li and Na entries of the database, 7283 and 12 065, respectively. However, since we seek stable compounds under ambient conditions, high pressure and/or high temperature, structural data were disregarded. Compounds with hydrogen were also excluded because the experimental structural data are often unreliable. In addition, for the sake of simplicity, we do not consider compounds with partial occupancies on atomic sites that do not correspond to the mobile species. The assignment of oxidation states is carried out after this filtering. The number of Li and Na entries with a successful assignment of oxidation states is 2043 and 2736, respectively. The last step involves removing duplicate crystal structures. Two entries are considered to correspond to the same crystal structure if both the chemical formula and the space group are identical. In the cases where there is more than one entry for a crystal structure, we select the entry with the lowest global instability index (GII) (Salinas-Sánchez *et al.*, 1992). This index is defined in the framework of the BV theory to quantify the degree to which the atomic structure as a whole is strained. In stable, well determined structures, the GII is usually less than 0.1 valence units. Values between 0.1 and 0.2 indicate a strained structure, whereas higher values are unusual and can point to an incorrect structure determination. The GII is also computed with *BondStr*. After removal of duplicate entries, the number of Li and Na compounds is 1025 and 1477, respectively.

2.3. Structural descriptors and machine-learning analysis

Every compound has been characterized by several structural descriptors in order to establish correlations between $E_{\rm m}$ and the crystal structure. The short-range order of the crystal is characterized by structural descriptors that are derived from a Voronoi partition (Blatov, 2007) of the crystal space. This partition provides an unambiguous definition of the coordination number, free of any arbitrariness such as, for example, a cut-off bond distance. This partition encloses every atom in the convex polyhedron defined by the set of points closer to it than to any other atom. By construction, the normal of each face of the polyhedron points to a neighbouring atom. The coordination number N is then defined as

$$N = \sum_i \frac{\Omega_i}{\Omega_{\rm max}}, \qquad (2)$$

where i runs over all faces of the polyhedron and Ω_i is the solid angle subtended by the face i. The face with the largest solid

angle, Ω_{max}, is used to normalize this angle. The volume of the corresponding polyhedron gives the size of each atom. When analysing ion diffusion, it is also essential to have long-range order descriptors in order to describe the diffusion channels. For this reason, we have developed a descriptor that quantifies the size of the open channels inside the crystal structure. We call this new descriptor the percolation radius, r_p, which we computed from a hard-sphere model as follows. Firstly, we remove the mobile ion from the crystal structure. Then, the remaining ions are considered as hard spheres with a size determined by the ionic radius. We have used the ionic radii tabulated in *CrysFML*, which correspond to the Pauling ionic radius. In close analogy with BVEL calculations, the unit cell is mapped by computing, on a fine grid, the maximum radius of a hard sphere that does not overlap with any atomic sphere of the crystal (Fig. 3, left). The isosurface corresponding to an arbitrary radius, r, encloses all the points for which the point–ion distance is equal to or greater than r. If the isosurface percolates the unit cell, it means that there is a path across which a sphere of radius r can traverse the unit cell without overlapping with any atomic sphere. The threshold value at which the isosurface stops percolating is r_p, which corresponds to the maximum radius of a sphere that can percolate the crystal structure considered as an arrangement of hard spheres. r_p corresponds strictly to the bottleneck of the migration path in a hard-sphere model, and provides a quantitative characterization of the open channels through which the mobile ion can diffuse. In Fig. 3 (right), we compare the calculated energy and radius isosurfaces at their threshold values for a well known ionic conductor, NASICON (Rodger *et al.*, 1985). The computed diffusion paths and the hard-sphere model agree well. In the next section, we demonstrate that there is a very strong correlation between E_m and r_p, which suggests that r_p is a good estimation of the bottleneck of the diffusion path predicted by the BVEL calculation. However, as relaxation effects are absent in this approximation, r_p cannot be linked directly to the bottleneck of the real diffusion path, where anions must move apart from the Li (Na) diffusion ion at the bottleneck. One of the consequences of neglecting relaxation effects is the overestimation of the activation energy. These effects are taken into account by more accurate approximations based on first-principles calculations. In spite of not considering the dynamics of the crystal structure, the BVEL method is still able to provide the correct migration path in many situations. In the SI we compare *ab initio* with BVEL calculations for the diffusion of Na ions in the layered oxide $Na_{2/3}Mn_{1/3}Fe_{2/3}O_2$ (Katcho *et al.*, 2017). We show how the BVEL method predicts correctly the Na migration path, which depends on the local Na arrangement.

To build a regression model connecting our set of descriptors with the target property of ion migration energy, we resort to the machine-learning technique known as random forests (Breiman, 2001). More specifically, we employ the implementation found in *scikit-learn* (Pedregosa *et al.*, 2011). The method is based on an ensemble of decision trees, each of which is trained on a different set of compounds, sampled at random with replacement from the input set. Each tree is a complete hierarchy of splits based on the values of different sets of descriptors. While the tree is growing, the set of descriptors used for each split is sampled at random, but then the splitting criterion is optimized to minimize the mean-square error. The training phase of the method ends when the predefined number of trees have been grown. To extract a prediction for an arbitrary compound from the model, its data are run down each individual tree and the results are averaged. Among the advantages of random forests with respect to other regression techniques, crucial to us is the fact that they work well even when the input contains correlated descriptors, as in our case. Another interesting feature is that they provide a built-in method to assess the importance of each descriptor, based on the average reduction in variance. To do this, we compute the reduction in total variance between each parent node and the set of its children, and then average this

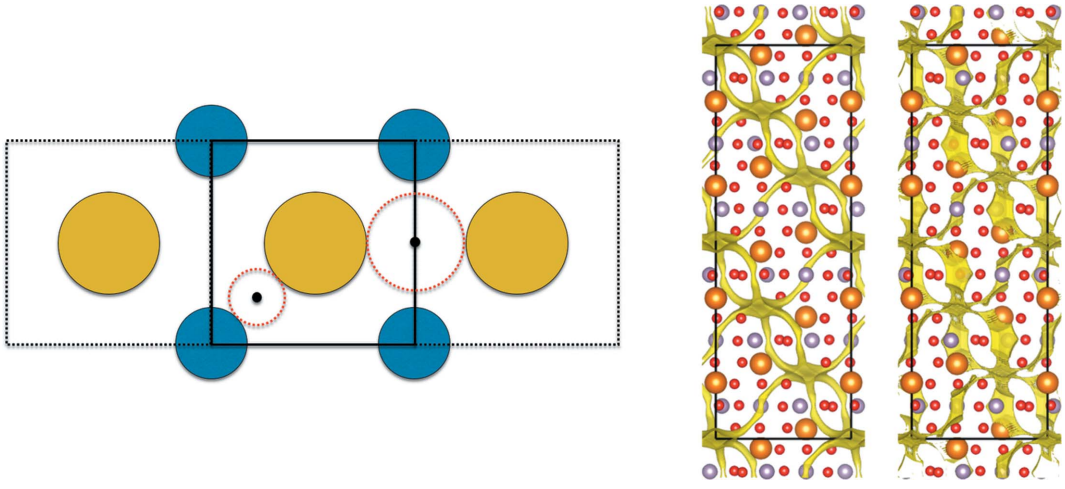


Figure 3
Left: two-dimensional hard-sphere model. Filled circles represent atoms. Empty circles define the maximum radius of a circle that does not overlap with any atom. The radius of the smallest red circle corresponds to the percolation radius. Right: BV (left) and radius (right) isosurfaces corresponding to the percolation energy and the percolation radius in NASICON.

reduction over all splits involving the descriptor under study, weighting each split by the fraction of compounds in the node. This per-tree importance is then normalized over all features and averaged over all trees in the forest. To avoid overfitting, we randomly choose 50% of the compounds for training the model and reserve the other 50% for validation. Another important detail of our specific application is our choice of $\log(E_m)$ instead of E_m itself as the target variable. The precision and accuracy of predictions in a certain subinterval is affected by the density of values from the training set in that subinterval. By applying monotonic transformations to the predicted variable, we can achieve different tradeoffs between homogeneous predictive power across the board and focus on a particular range of values. Given the very inhomogeneous distribution of our training data, we settled on using the logarithm to avoid large differences in accuracy and precision in different subintervals. We must point out, however, that this choice does not affect the results on a qualitative level.

3. Results and discussion

Initially we compute E_m for every selected Li and Na compound. We start a coarse search for E_m by applying the percolation algorithm to the isosurface with $E = 1$ eV. If there is no percolation, we increase E by steps of 1 eV until percolation is found. We do not search for percolation at energies larger than 5 eV. Then, E_m is refined by using energy steps of 0.1 and 0.01 eV consecutively. The results are summarized in Fig. 4, which shows the histograms of the number of scanned compounds as a function of E_m. The histograms for both Li and Na compounds have a similar shape. At low energies, the density of compounds increases with energy. There is a maximum around 1 eV, followed by a monotonic decrease of the density of compounds. It is well known that the BVEL method overestimates E_m. Scaling factors of about 0.9 for network structures and 0.45 for interlayer paths in layered structures have been proposed (Adams & Rao, 2014). Xao *et al.* (2015) report an extensive comparison between *ab initio* and BVEL migration energies for a wide range of families of compounds, such as olivine, spinel, layered, borates/carbonates, antiperovskite, phosphates, sulfates and tavorite. In general, the scaling factor ranges from 0.3 to 0.7. Even inside a given family, the dispersion of the scaling factors can be quite large. For Li compounds, a reasonable choice for an upper limit of the activation energy is 0.5 eV (Bachman *et al.*, 2016). In the Na case, β-alumina and NASICON also have activation energies well below 0.5 eV (Lu *et al.*, 2010; Park *et al.*, 2016). If we combine this upper limit with the expected scaling factors for the energies predicted by the BVEL method, it should be expected that BVEL calculations will provide $E_m < 1.5$ eV for any fast ion conductor. A complete ranking of Li and Na compounds with E_m less than 5 eV is provided in the SI. Compounds with large E_m are impermeable to Li (Na) diffusion and could serve as blocking electrodes.

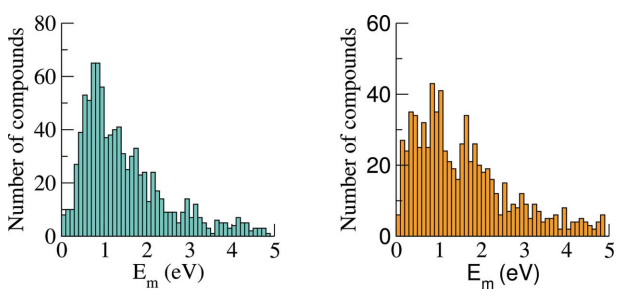


Figure 4
Number of scanned compounds as a function of E_m. Left: Li compounds. Right: Na compounds.

In order to go beyond the simple generation of a ranking of compounds, we have analysed the correlations between E_m and the crystal structure as well. To this end, we compare compounds with the same type of anion, since the nature of the anion is an important parameter affecting E_m. We focus on lithium and sodium oxides because they provide good statistics as there are a large number of compounds belonging to this family in the ICSD. Each compound is characterized by the following descriptors: the cation site energy (E_{site}), the cation BVS, the cation–cation and cation–anion coordination numbers (N) calculated following equation (2), the migration energy (E_m), the Voronoi polyhedron volume (V) of the cation and the anion, the cation and anion volume fractions (fV), the density of the compound (ρ), and the computed percolation radius (r_p). The BVS is the predicted valence for the ion [first term of the right-hand side of equation (1)]. Correlations are quantified by computing the Spearman correlation coefficient between each pair of descriptors. Fig. 5 displays the correlograms for lithium and sodium oxides. The two correlograms are qualitatively similar. There is a clear negative correlation between E_m and r_p, $N_{cation–cation}$, fV_{cation}, V_{cation} and E_{site}, and a positive correlation with $N_{cation–O}$, fV_O and BVS. The negative correlation between E_m and r_p was expected. As we have described in Section 2, r_p measures the bottleneck of the diffusion channel. An increase of r_p reduces the repulsive forces at the bottleneck and, therefore, this reduces E_m. The negative correlation with V_{cation} has the same origin. Indeed, it can be seen that V_{cation} and r_p are strongly and positively correlated. $N_{cation–cation}$ is greater than zero when the polyhedra of the two cations share one face. This situation should be expected when fV_{cation} is large. This is confirmed by the correlation coefficient between fV_{cation} and $N_{cation–cation}$, which is large and positive. Increasing fV_{cation} implies a larger accessible volume for the diffusion of the cation and, in general, this is accompanied by a decrease of E_m. This explains the negative correlation coefficient between E_m and fV_{cation} and $N_{cation–cation}$. The last descriptor with a clear negative correlation with E_m is E_{site}. Long (strained) bonds are expected to have smaller potential wells, lowering the diffusion energy barrier. Especially in the case of Na compounds, this is reflected by the strong positive correlation with BVS. Na ions with predicted valences lower than 1, pointing to long bond distances (*i.e.* local tensile strain), have lower E_m in general. Finally, $N_{cation–O}$ is positively correlated with E_m, *i.e.* structures with low coordination numbers have generally lower ion diffusion barriers. Interestingly, we show at the end

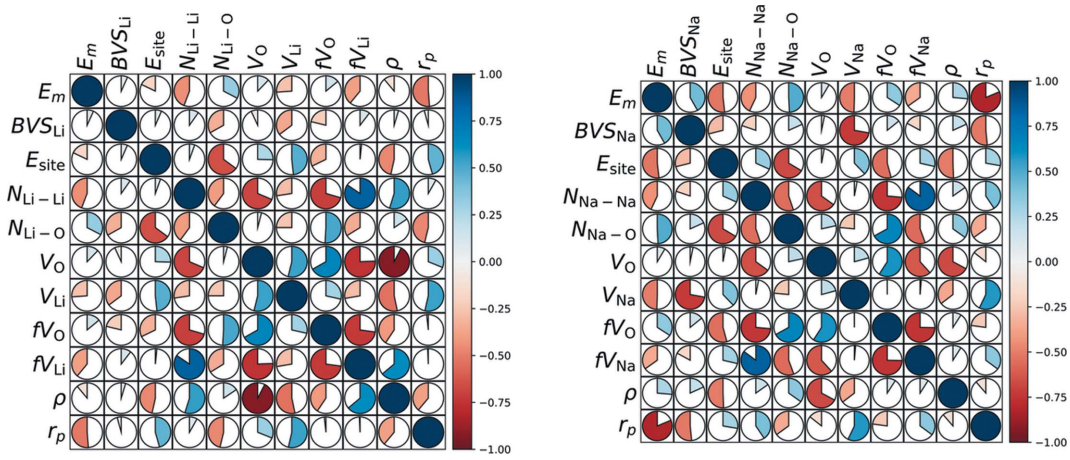


Figure 5
Correlograms between the ion migration energy E_m, the cation bond-valence sum BVS, the cation site energy E_{site}, the cation–cation and cation–anion coordination numbers $N_{cation-cation}$ and $N_{cation-O}$, the Voronoi polyhedra of the anion and cation, V_O and V_{cation}, the anion and cation volume fractions, fV_O and fV_{cation}, the density ρ, and the percolation radius r_p. Left: Li compounds. Right: Na compounds. The data corresponding to this figure are provided in the SI.

of this section that one of the best Na-ion conductors, β-alumina, deviates from this rule by compensating a large $N_{cation-O}$ with a very large r_p.

We also found that, from a quantitative point of view, the influence of the crystal structure on E_m seems to be larger in the case of Na compounds, since the correlations are clearly stronger. However, as many of the structural data of the ICSD come from X-ray experiments, which are not well suited for locating light elements such as Li, this conclusion should be treated with caution. The weaker correlations found in the Li case could also be explained by a loss in accuracy of the atomic coordinates of Li compared with Na, which can blur the correlations. To minimize the risk of dealing with not well determined crystal structures, we have only considered those crystal structures with GII less than 0.5.

In Fig. 6, we again summarize the results of the scan of the ICSD, but now by plotting E_m as a function of the descriptor with the strongest correlation, *i.e.* r_p. Compounds are classified according to the nature of the anion: oxides (O^{2-}), sulfides (S^{2-}), nitrides (N^{3-}), chalcogenides (Se^{2-}, Te^{2-}), halides (F^-, Cl^-, Br^-, I^-) and oxyhalides (O^{2-} + $F^-/Cl^-/Br^-/I^-$). The negative correlation between E_m and r_p is clearly shown, although in the lithium case, especially for oxides, we observe a larger dispersion of the data. As a consequence of this negative correlation, r_p imposes a lower limit to E_m for a given r_p. In the lithium case, it can be observed that up to *ca* 0.5 Å the lowest value taken by E_m as a function of r_p decreases monotonically. In this region, the channel size is too small for an Li ion, and an increase in r_p decreases E_m because it reduces the repulsive forces at the bottleneck. For r_p greater than 0.5 Å, there is no further decrease of E_m. The channel size starts to become too large for the Li ion, increasing the potential well. In the sodium case, trends are similar although some differences can be observed. Obviously, the data are shifted to larger r_p values since the Na ion is larger than the Li ion. In addition, the data can easily be clustered by families. For oxides, the monotonic decrease of E_m stops around 0.9 Å, whereas for sulfides, chalcogenides and halides it stops around 0.7 Å. Oxyhalides are located between these two sets.

A machine-learning regression of E_m results for lithium and sodium oxides was carried out following a strategy of recursive feature elimination. We find that reasonably good predictions are obtained by using only the following three structural descriptors: r_p, $N_{cation-O}$ and fV_{cation} (Fig. S1 in the SI). As the correlation between the crystal structure and E_m is clearly

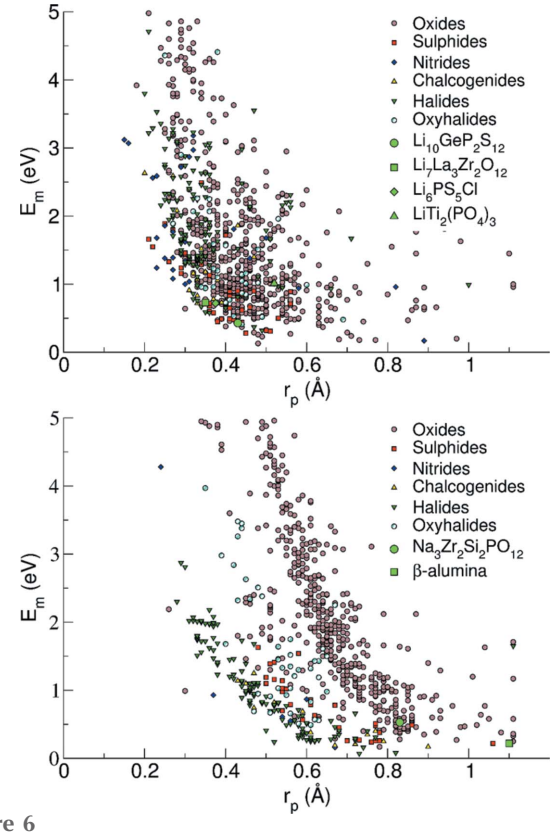


Figure 6
Migration energy as a function of the percolation radius. Top: Li compounds. Bottom: Na compounds.

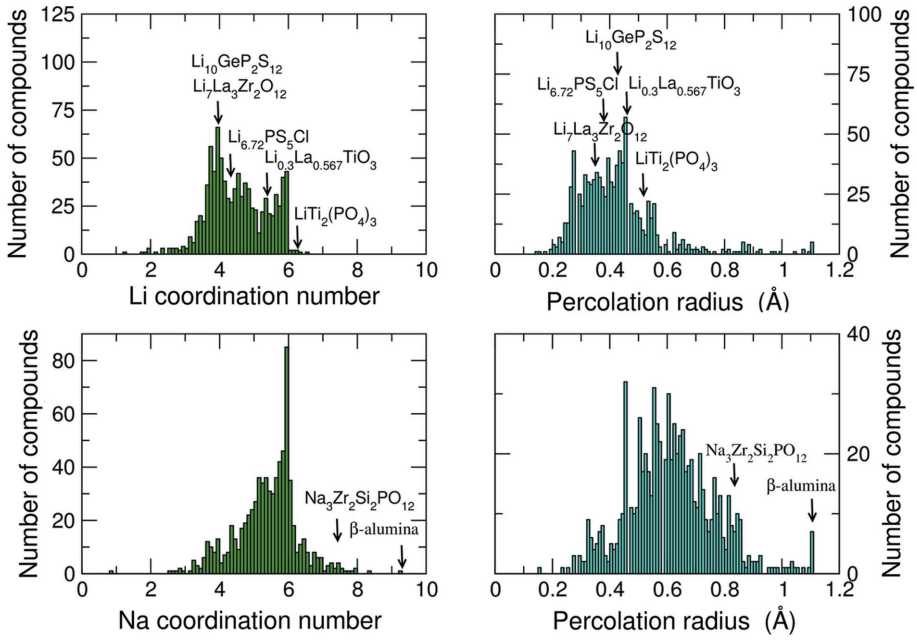


Figure 7
Number of compounds as a function of the cation coordination number and percolation radius. Top: Li compounds. Bottom: Na compounds.

stronger for Na compounds, the corresponding prediction results are also better. The inclusion of more descriptors does not improve the prediction. Therefore, a significant improvement requires the introduction of new structural descriptors. The descriptors derived from a Voronoi partition of the crystal space can describe well the initial and final states of the hopping, but not the transition state. Notice that r_p provides some information about the transition state since it provides the size of the bottleneck, which is usually located at the transition state. However, information about the coordination of the ion at the bottleneck is missing. We therefore argue that an improvement of the description of the transition state should lead to improved predictions.

Finally, we focus on the structural properties of some of the best Li- and Na-ion conductors. For Li-based compounds, we have selected one member of each representative family of fast Li-ion conductors: LISICON-like ($Li_{10}GeP_2S_{12}$), NASICON [$LiTi_2(PO_4)_3$], garnet ($Li_7La_3Zr_2O_{12}$), perovskite ($Li_{0.3}La_{0.567}TiO_3$) and argyrodite ($Li_{6.72}PS_5Cl$). For Na-based compounds, we have taken as representative examples NASICON ($Na_3Zr_2Si_2PO_{12}$) and β-alumina [$Na_2O(Al_2O_3)_{11}$]. We investigate how some of their structural descriptors compare with the rest of the Li- and Na-based compounds. Fig. 7 displays the number of scanned compounds as a function of $N_{cation-O}$ and r_p. The values of the reference compounds are also included. The distribution corresponding to $N_{cation-O}$ is bimodal in the lithium case, with peaks at $N_{Li-O} = 4$ (tetrahedral coordination) and $N_{Li-O} = 6$ (octahedral coordination). In the sodium case, there is only one clear peak, located at $N_{Na-O} = 6$. Yet the most important difference between lithium and sodium plots is observed when we look at the reference compounds. Fast Li-ion conductors have the typical coordination numbers of lithium compounds, between 4 (LISICON-like, garnet, argyrodite) and 6 (perovskite, NASICON). However, the best two Na-ion conductors – NASICON and β-alumina – have much larger coordination numbers than the usual values found for sodium compounds. In particular, β-alumina has the largest coordination number of all scanned sodium compounds. It is important to keep in mind that these coordination numbers are derived from a Voronoi partition of the crystal space, so they may not coincide with more traditional coordination numbers derived from more arbitrary coordination polyhedra. In β-alumina, sodium ions are located at the centre of an antiprism-shaped first coordination shell. Therefore, the traditional coordination number is six (six Na—O distances of 2.83 Å). However, there are another three close oxygen anions – the second coordination shell – located at a distance of 3.23 Å which are still 'visible' (*i.e.* subtend a nonzero solid angle) in the Voronoi sense and that we therefore consider as neighbours. The same trends are observed when looking at r_p. Again, fast Na-ion conductors have unusually large r_p values, whereas in the lithium case the reference values are located in a region of high density of compounds. The r_p values for the reference Li-ion conductors with lithium in a tetrahedral environment are well below 0.5 Å, the channel size from which BV theory predicts the lowest energy diffusion barriers (see Fig. 6). This discrepancy can be understood by the absence of relaxation effects of the crystal structure in our BV approach, which favours more open crystal structures. On the sodium side, the very large r_p values of fast Na-ion conductors may be surprising since one could expect large potential wells arising from large channel sizes and, in consequence, high energy barriers. However, these large sizes are accompanied by unusually large coordination numbers, which limits the depth of the potential well. Indeed, when we look at the number of compounds as a function of the $N_{cation-O}/r_p$ ratio (Fig. S2 in the SI), we observe that the best ionic conductors for Na, as well as Li, lie in the region where the density of compounds is largest.

4. Summary and conclusions

In this work, we have reported an automated scheme to calculate E_m using BV theory in its version based on Morse and screened Coulomb potentials. This methodology has been implemented in the program *BondStr*, based on the open-source library *CrysFML*. In addition, we have shown that by using a simple hard-sphere model the same mathematical approach used to compute E_m can be used to retrieve the bottleneck of the diffusion path. This can find application in the analysis of the influence of topology in atomic transport in

solid-state problems in ordered and, in principle, also disordered systems.

We have scanned 1025 Li and 1477 Na compounds of the ICSD. We report a ranking of Li and Na compounds with E_m lower than 1.5 eV, which can stimulate further experimental studies and theoretical investigations based on approximations that are more accurate. We have analysed the correlations between E_m and the crystal structure. A machine-learning analysis reveals that E_m is mainly determined by the bottleneck and the coordination polyhedron and volume fraction of the mobile species. However, our machine-learning prediction is not accurate enough to yield quantitative accuracy. We attribute this fact to an incomplete description of the transition state by the structural descriptors used in this work. Descriptors accounting for the relative orientation and distance between the coordination polyhedra involved in the hopping of the Li (Na) ion would improve the description of the transition state. On the other hand, our set of descriptors is mainly intended to provide a good description of the local atomic environment of the mobile species. In the case of multivalent mobile species, long-range electrostatic cation–cation repulsions could be more relevant. In this case, our short-range description would probably be insufficient, and additional descriptors characterizing interactions on a longer range would be required.

The correlation between E_m and the crystal structure is stronger for Na compounds. In particular, for Na oxides we find that the correlation factor between E_m and r_p is clearly larger than for Li oxides, which suggests that the size of the diffusion channel plays a more decisive role in Na-ion conductors. This is consistent with experimental observations. In contrast to what is observed for Li compounds, the best Na-ion conductors present unusually large diffusion channels, which are compensated by large coordination numbers. This suggests that fast Na-ion conductors could be outliers whose favourable properties stem from an unusual combination of factors, which would imply that the number of fast Na-ion compounds is significantly lower than the number of fast Li-ion conductors.

Funding information

This work was supported by the Ministerio de Economía y Competitividad (MINECO) of the Spanish Government through the Proyectos I + D Retos 2016 program (ION-STORE project, reference number: ENE2016- 81020-R). NM acknowledges support from ANR through project Carnot MAPPE.

References

Adams, St. (2001). *Acta Cryst.* B**57**, 278–287.
Adams, S. & Rao, R. P. (2011). *Phys. Status Solidi A*, **208**, 1746–1753.
Adams, S. & Rao, R. P. (2014). *Struct. Bond.* **158**, 129–160.
Avdeev, M., Sale, M., Adams, S. & Rao, R. P. (2012). *Solid State Ionics*, **225**, 43–46.
Bachman, J. C., Muy, S., Grimaud, A., Chang, H., Pour, N., Lux, S. F., Paschos, O., Maglia, F., Lupart, S., Lamp, P., Giordano, L. & Shao-Horn, Y. (2016). *Chem. Rev.* **116**, 140–162.
Blatov, V. A. (2007). *Crystallogr. Rev.* **10**, 249–318.
Borodin, O., Olguin, M., Spear, C. E., Leiter, K. W. & Knap, J. (2015). *Nanotechnology*, **26**, 354003.
Breiman, L. (2001). *Mach. Learn.* **45**, 5–32.
Brown, I. D. (2009). *Chem. Rev.* **109**, 6858–6919.
Carrete, J., Li, W., Mingo, N., Wang, S. & Curtarolo, S. (2014). *Phys. Rev. X*, **4**, 011019.
Ceder, G. (2010). *MRS Bull.* **35**, 693–701.
Chen, H. & Adams, S. (2017). *IUCrJ*, **4**, 614–625.
Cheng, L., Assary, R. S., Qu, X., Jain, A., Ong, S. P., Rajput, N. N., Persson, K. & Curtiss, L. A. (2015). *J. Phys. Chem. Lett.* **6**, 283–291.
Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2009). *Introduction to Algorithms*. MIT Press.
Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. (2003). *Phys. Rev. Lett.* **91**, 135503.
Deiseroth, H.-J., Kong, S.-T., Eckert, H., Vannahme, J., Reiner, C., Zaiß, T. & Schlosser, M. (2008). *Angew. Chem. Int. Ed.* **47**, 755–758.
Deng, Z., Mo, Y. & Ong, S. P. (2016). *NPG Asia Mater.* **8**, e254.
Deng, Z., Zhu, Z., Chu, I.-H. & Ong, S. P. (2017). *Chem. Mater.* **29**, 281–288.
Gao, J., Chu, G., He, M., Zhang, S., Xiao, R., Li, H. & Chen, L. (2014). *Sci. China Phys. Mech. Astron.* **57**, 1526–1536.
Greeley, J., Nørskov, J. K. & Mavrikakis, M. (2002). *Annu. Rev. Phys. Chem.* **53**, 319–348.
Hautier, G., Jain, A. & Ong, S. P. (2012). *J. Mater. Sci.* **47**, 7317–7340.
Hautier, G., Jain, A., Ong, S. P., Kang, B., Moore, C., Doe, R. & Ceder, G. (2011). *Chem. Mater.* **23**, 3495–3508.
Husch, T. & Korth, M. (2015). *Phys. Chem. Chem. Phys.* **17**, 22596–22603.
Islam, M. S. & Fisher, C. A. J. (2014). *Chem. Soc. Rev.* **43**, 185–204.
Jain, A., Shin, Y. & Persson, K. A. (2016). *Nat. Rev. Mater.* **1**, 15004.
Kamaya, N., Homma, K., Yamakawa, Y., Hirayama, M., Kanno, R., Yonemura, M., Kamiyama, T., Kato, Y., Hama, S., Kawamoto, K. & Mitsui, A. (2011). *Nat. Mater.* **10**, 682–686.
Katcho, N. A., Carrasco, J., Saurel, D., Gonzalo, E., Han, M., Aguesse, F. & Rojo, T. (2017). *Adv. Energy Mater.* **7**, 1601477.
Kireeva, N. & Pervov, V. S. (2017). *Phys. Chem. Chem. Phys.* **19**, 20904–20918.
Kirklin, S., Meredig, B. & Wolverton, C. (2013). *Adv. Energy Mater.* **3**, 252–262.
Lorensen, W. E. & Cline, H. E. (1987). *Comput. Graph.* **21**, 163–169.
Lu, X., Xia, G., Lemmon, J. & Yang, Z. (2010). *J. Power Sources*, **195**, 2431–2442.
Martínez-Juárez, A., Pecharromán, C., Iglesias, J. E. & Rojo, J. M. (1998). *J. Phys. Chem. B*, **102**, 372–375.
Meng, Y. S. & Arroyo-de Dompablo, M. E. (2009). *Energy Environ. Sci.* **2**, 589.
Meredig, B. & Wolverton, C. (2010). *Nat. Mater.* **12**, 123–127.
Mizushima, K., Jones, P. C., Wiseman, P. J. & Goodenough, J. B. (1980). *Mater. Res. Bull.* **15**, 783–789.
Nakayama, M., Kimura, M., Jalem, R. & Kasuga, T. (2016). *Jpn. J. Appl. Phys.* **55**, 01AH05.
Padhi, A. K., Nanjundaswamy, K. S. & Goodenough, J. B. (1997). *J. Electrochem. Soc.* **144**, 1188.
Park, H., Jung, K., Nezafati, M., Kim, C. & Kang, B. (2016). *Appl. Mater. Interfaces*, **8**, 27814–27824.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.
Qu, X., Jain, A., Rajput, N. N., Cheng, L., Zhang, Y., Ong, S. P., Brafman, M., Maginn, E., Curtiss, L. A. & Persson, K. A. (2015). *Comput. Mater. Sci.* **103**, 56–67.
Rodger, A. R., Kuwano, J. & West, A. R. (1985). *Solid State Ionics*, **15**, 185–198.

Rodríguez Carvajal, J. (1993). *Physica B*, **192**, 55–69.
Sale, M. & Avdeev, M. (2012). *J. Appl. Cryst.* **45**, 1054–1056.
Salinas-Sánchez, A., García-Muñoz, J. L., Rodríguez-Carvajal, J., Saez-Puche, R. & Martínez, J. L. (1992). *J. Solid State Chem.* **100**, 201–211.
Schütter, C., Husch, T., Korth, M. & Balducci, A. (2015). *J. Phys. Chem. C*, **119**, 13413–13424.
Stramare, S., Thangadurai, V. & Weppner, W. (2003). *Chem. Mater.* **15**, 3974–3990.
Thangadurai, V., Narayanan, S. & Pinzaru, D. (2014). *Chem. Soc. Rev.* **43**, 4714–4727.
Whittingham, M. S. (2004). *Chem. Rev.* **104**, 4271–4301.
Xiao, R., Li, H. & Chen, L. (2015). *Sci. Rep.* **5**, 14227.
Xiao, R., Li, H. & Chen, L. (2016). *J. Materiomics*, **1**, 325–332.