Search for Active and Inactive Ion Insertion Sites in Organic Crystalline Materials

Harshan Reddy Gopidi,[1,2,3] Alae Eddine Lakraychi,[1,2,3] Abhishek A. Panchal,[1,2,3] Yiming Chen,[1,4] V. S. Chaitanya Kolluru,[1,4] Jiaqi Wang,[1,5] Ying Chen,[1,6] Liu Jue,[7] Kamila Wiaderek,[1,5] Maria K.Y. Chan,[1,4] Yan Yao,[1,2,3,*] and Pieremanuele Canepa[1,2,3,#]

[1]Energy Storage Research Alliance, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

[2]Texas Center for Superconductivity, University of Houston, Houston, TX 77204, USA

[3]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204, USA

[4]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, USA

[5]X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, USA

[6]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA

[7]Neutron Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

*email: yyao4@central.uh.edu

#Email: pcanepa@uh.edu

Abstract

The position of mobile active and inactive ions, specifically ion insertion sites, within organic crystals, significantly affects the properties of organic materials used for energy storage and ionic transport. Identifying the positions of these atom (and ion) sites in an organic crystal is difficult, especially when the element has a low X-ray scattering power, such as lithium (Li) and hydrogen, which are difficult to detect with powder X-ray diffraction (XRD) methods. First-principles calculations, exemplified by density functional theory (DFT), are very effective for confirming the relative stability of ion positions in materials. However, the lack of effective strategies to identify ion sites in these organic crystalline frameworks renders this task extremely challenging. This work presents two algorithms: (i) Efficient Location of Ion Insertion Sites from Extrema in electrostatic local potential and charge density (ELIISE), and (ii) ElectRostatic InsertioN (ERIN), which leverage charge density and electrostatic potential fields accessed from first-principles calculations, combined with the Simultaneous Ion Insertion and Evaluation (SIIE) workflow –that inserts all ions simultaneously—to determine ion positions in organic crystals. We demonstrate that these methods accurately reproduce known ion positions in 16 organic materials and also identify previously overlooked low-energy sites in tetralithium 2,6-naphthalenedicarboxylate (Li$_4$NDC), an organic electrode material, highlighting the importance of inserting all ions simultaneously as done in the SIIE workflow.

Introduction

Understanding the structure of inorganic and organic-based materials containing electroactive elements, such as Li, Na, and zinc (Zn)[1–14] is relevant for the development of the next generation of rechargeable battery technologies, which are prevalent in vehicular transportation, heavy-duty applications, and the supporting electrical grids.

With the advent of the Rietveld methods,[15–18] diffraction-based techniques using X-ray (XRD) or neutron (ND) sources are commonly employed to determine the structures of energy materials. However, structural determinations through diffraction methods can be hindered by the weak scattering of X-rays by light elements, such as hydrogen and lithium (Li), which can partially occupy multiple crystallographic sites because of their high intrinsic mobility. In some cases, significant ion mobilities can cause the blurring of Bragg intensities.[19,20] While ND experiments enhance sensitivity to specific elements, e.g., light elements H and Li, ND experiments are significantly less available than XRD, and they still face challenges related to disorder, defects, and temperature-dependent site mixing.[21]

Organic electrode materials (OEMs) are promising for inexpensive and high-energy-density rechargeable batteries. Indeed, OEMs can provide three main benefits: (i) high material-level energy densities of 900-1000 kWh kg^{-1}, (ii) a growing range of designs as shown by the variety of molecules studied so far, and (iii) potential benefits in sustainability and supply chains.[6,22–26] One of the main challenges hindering the practical implementation of OEMs is a poor understanding of their structure-property

3

relationships. The electrochemical reaction pathways of OEMs are typically interpreted based on simplistic molecular models, rather than describing these processes at the material level. These molecular-type OEM models are often paired with *ex situ* or *operando* XRD experiments, which frequently cannot reveal their underlying crystal structures and potential phase transformations.[27–30] To date, only a handful of studies have resolved structures at both charged and discharged (the ion-containing form) states of specific OEMs.[23,31–35] For example, 1,4-benzoquinone (BQ) undergoes a phase transition from monoclinic ($P2_1/c$) to orthorhombic ($P4_2/ncm$) upon lithiation or sodiation, forming Li_2BQ or Na_2BQ,[23,36] indicative of conversion-type reactions. By contrast, dilithium 2,6-naphthalene dicarboxylate (Li_2NDC) retains a monoclinic ($P2_1/c$) structure upon complete lithiation (Li_4NDC),[32,33,37] following an intercalation reaction. However, distinguishing between specific reaction mechanisms remains poorly understood in the OEM literature, mainly because of the lack of structural data for the discharged state, unlike the charged state, which is often known.

A comprehensive understanding of OEM structures and their behavior across different ion content (i.e., different states of charge) remains crucial for bridging the gap between molecular-scale insights and crystalline materials with intrinsic periodicity. In general, this knowledge is essential for elucidating the physicochemical processes that govern electrochemical behavior, including phase transitions,[23] voltage profiles,[35] and mechanical degradation.[34,38] However, accurate identification of the positions of mobile ions such as Li^+, Na^+, K^+, Zn^{2+}, and Mg^{2+}, in organic crystalline frameworks, remains a significant experimental and computational challenge.[27,31,33,34,39–47] This challenge arises primarily from the ample intermolecular space (void within the crystal

framework) caused by weak intermolecular interactions, which are responsible for molecular assembly into organic molecular crystals. This results in numerous possible cation arrangements within the organic framework. Therefore, developing computational tools that can accurately predict ion positions within the crystalline structure of organic materials is essential for making breakthroughs and unlocking hidden potential for rational materials design from a crystal-level perspective. While such approaches have been established for inorganic electrode materials, [48–51] predictive search for ion sites in OEMs remains largely confined to the molecular scale and,[44,52] to the best of our knowledge, has not yet been demonstrated for crystalline organic frameworks.

In this paper, we present two complementary computational algorithms and an ion insertion workflow for identifying the optimal active and inactive ion insertion sites within organic frameworks: (i) The Efficient Location of Ion Insertion Sites from Extrema (ELIISE) in electrostatic local potential and charge density, which identifies electrostatically stable ion positions by locating local extrema in electrostatic local potential and electronic charge density (CD) fields. (ii) The ElectRostatic InsertioN (ERIN), which, by following the electrostatic potential, iteratively places ions into electrostatically favorable regions of the unit cell until the entire cell volume is sampled. (iii) The Simultaneous Ion Insertion and Evaluation (SIIE), which inserts the ions simultaneously and evaluates to determine the correct structure. ELIISE and ERIN, in conjunction with SIIE, are applied with a library of 16 representative organic structures, illustrating their ability to automatically identify candidate ion sites within complex, flexible organic frameworks using limited computational resources. These predictions

are subsequently refined using agnostic first-principles calculations, resulting in high-confidence ion positions, which provide crucial insights into ion coordination environments and other structural evidence. These predictive workflows establish a powerful toolkit for the systematic exploration of electroactive structures and their ion transport pathways, elucidating reaction mechanisms and guiding the rational design of materials for the development of competitive organic-based electrochemical devices, as well as the exploration of soft materials for other energy-storage applications.

Methods

1. Finding candidate ion insertion sites in organic crystals: ELIISE and ERIN

Figure 1 shows the identification of candidate ion sites using ELIISE and ERIN, which rely on the electronic charge density (CD, or $\rho(r)$) and the local electrostatic potential $LP(r)$ **Eq. 1**, and the spherically averaged local potential, $ALP(r)$, of **Eq. 2**.

$$LP(r) = V_{hartree}(r) + V_{ionic}(r) \qquad \text{Eq. 1}$$

where $V_{hartree}(r)$ is the Coulomb potential set by $\rho(r)$, and $V_{ionic}(r)$ is the Coulomb potential originating from the ionic core of all atoms. From $LP(r)$, the $ALP(r)$ is defined as:

$$ALP(r) = \frac{\int_{V_s} LP(r + r')\, d^3 r'}{\frac{4}{3}\pi r_s^3} \qquad \text{Eq. 2}$$

where V_s represents the spherical region within radius r_s that is averaged at each point r, and r' is a dummy variable over which the averaging occurs. Note, in **Eq. 2**, $ALP(r)$ is the spherically averaged electrostatic potential at each point in the organic crystal framework, with constant radii (r_s) set by the user (here, fixed to the empirical covalent radii of the atom that is inserted).

Both methodologies rely on an initial organic framework to compute $LP(r)$ and CD. The choice of the initial organic framework is trivial when the organic framework of the discharged (ion-inserted) structure— i.e., the lattice parameters, the symmetry, and the atomic positions— is known. In such cases, only the precise determination of ion positions is necessary, as is often the case with lithium; here, ELIISE serves as an

appropriate tool. Conversely, the more prevalent scenario is when only the organic

framework of the charged (ion-removed) phase is known. Under these circumstances,

ERIN is recommended, as employing ELLISE may result in unreliable outcomes.

Figure 1: Workflow and schematic illustration of the ELIISE (left) and ERIN (right) algorithms. ELIISE: Process for identifying potential ion sites based on local potential and charge density. ERIN: Workflow illustrating the iterative process used to identify potential ion sites based on electrostatic stability systematically.

2. The Efficient Location of Ion Insertion Sites from Extrema in electrostatic local potential and charge density (ELIISE) Method

The ELIISE method identifies candidate ion sites using both the CD and the spherically averaged local potential $ALP(r)$, defined in **Eq. 2,** within the empty (i.e., without any mobile species) organic framework from the discharged state.

Following the ELIISE workflow in the left part of **Figure 1**, the local potential and electronic CD are initially computed by performing a first-principles calculation without changing the atom positions (or changing the volume or cell shape of the material) of the empty organic framework. Subsequently, ELIISE identifies candidate active sites in an organic framework by identifying the local maxima and the local minima in the $ALP(r)$ the CD fields, respectively. **Figure 1** shows a schematic diagram of the local minima in CD and local maxima in the $ALP(r)$, which are used to identify candidate sites. We have concluded that a spherical average of the $LP(r)$ provides the best predictions in terms of candidate ion sites, as it accounts for the finite size of inserted ions. Here, the empirical covalent radii of atoms by Slater[53] appear to be a reliable option for defining this ion size. When different types of active species are involved in ion insertion, we imposed the empirical radius of the smaller atom for spherical averaging of the local potential. The candidate sites identified by both the CD and the $ALP(r)$ descriptors are combined into a single collection. To avoid repetition of sites, spatially close sites are merged by calculating the geometric mean of their coordinates, as indicated in **Figure 1**. The resulting sites are then ordered by their electrostatic stability, from the most favorable to the least favorable, as determined by the $ALP(r)$.

Large values of $ALP(r)$ indicate regions that are electrostatically favorable for positively charged ions, typically in the vicinity of more electronegative (negatively charged) atoms within the crystal structure. Therefore, local maxima in $ALP(r)$ naturally point to potential cation sites. Similarly, local minima in the CD usually occur in voids near more electronegative atoms. It is reasonable to assume that these electronegative atoms strongly attract electron density, leading to electron-deficient regions nearby, which are likely to form local minima in the charge density. Therefore, these minima of CD often align with favorable coordination environments for cations. An exception to this trend becomes clear when local minima occur in large structural voids, which are common in organic crystals or nanoporous materials. These voids may be too far from the organic framework to provide any stabilizing interactions for inserted ions. These situations can lead to surprisingly high Coulomb energies. To address this potential issue during the site search in both ELIISE and ERIN, we set cutoff distance thresholds to exclude regions that are either closer than the specified distance (e.g., 1.5 Å) or farther than ~3 Å from atoms in the organic framework.

3. The ElectRostatic InsertioN (ERIN) Method

In contrast to ELIISE, ERIN only requires the $LP(r)$ (**Eq. 1**) of the organic framework, obtained from first-principles calculations. The ERIN approach in the right part of **Figure 1** only uses the spherically averaged local potential $\mathrm{ALP}(r)$ (**Eq. 2**) to explore electrostatic energy landscapes and systematically generates sparsely spaced candidate ion sites in a decreasing order of electrostatic favorability within the given unit cell. ERIN identifies and selects the site with the maximum $ALP(r)$ value. Subsequently, ERIN excludes a spherical region within a user-defined radius, $r_{exclude}$ (e.g., 2 Å), around this site from future searches. Then, ERIN repeats the search for

the site with the next highest value of $ALP(r)$ to select as a candidate. This process is repeated iteratively until the entire volume of the unit cell is scanned systematically and excluded. The site search is performed independently within each symmetry-distinct region of the unit cell to identify candidate sites in all possible Wyckoff positions.

4. Simultaneous Ion Insertion and Evaluation (SIIE) Workflow for Candidate Site Insertion

Once ELIISE or ERIN determines the candidate sites, the next step is to identify the optimal combination of sites yielding the expected stoichiometry of the system and energetically viable structures. For this purpose, we introduce the Simultaneous Ion Insertion and Evaluation (SIIE) workflow, in which ions are introduced simultaneously within the organic framework to achieve the desired stoichiometry, as opposed to the sequential insertion method, where ions are inserted one at a time.[48,54] In the SIIE workflow, if the organic framework of the discharged structure is known, its atoms are kept fixed during ion insertion, with only inserted ions and decorating hydrogen atoms (whose positions are typically unknown from standard XRD experiments) allowed to change. Otherwise, if only the organic framework of the charged phase is known, then both the ions and the organic framework are allowed to relax completely.

As shown in **Figure 2a**, SIIE uses the candidate ion insertion sites from the ELIISE or ERIN methods discussed in the previous sections as inputs. While ELIISE and ERIN provide sets of candidate ion positions, they do not determine the final configuration of ions in the discharged structure. In real systems, multiple symmetry-distinct sites

may be occupied simultaneously, and therefore, the correct discharged state often results from a specific combination of these candidate sites.

Finding the right combination of candidate sites is challenging due to the large number of unique possibilities. Note that the number of possible combinations of predicted ion sites that produce the target ion content can grow combinatorially, in the range $10 - 10^6$, which typically makes direct DFT evaluation of all combinations computationally infeasible.

To address this challenging task, we incorporate a machine learning interatomic potential (MLIP), such as MACE or Orb-v3,[55–57] to screen thousands of configurations efficiently. Here, we utilize MACE, which enables us to identify and remove high-energy structures. More details on how MACE and DFT are combined in the SIIE workflow are provided later (see **Computational Details**).
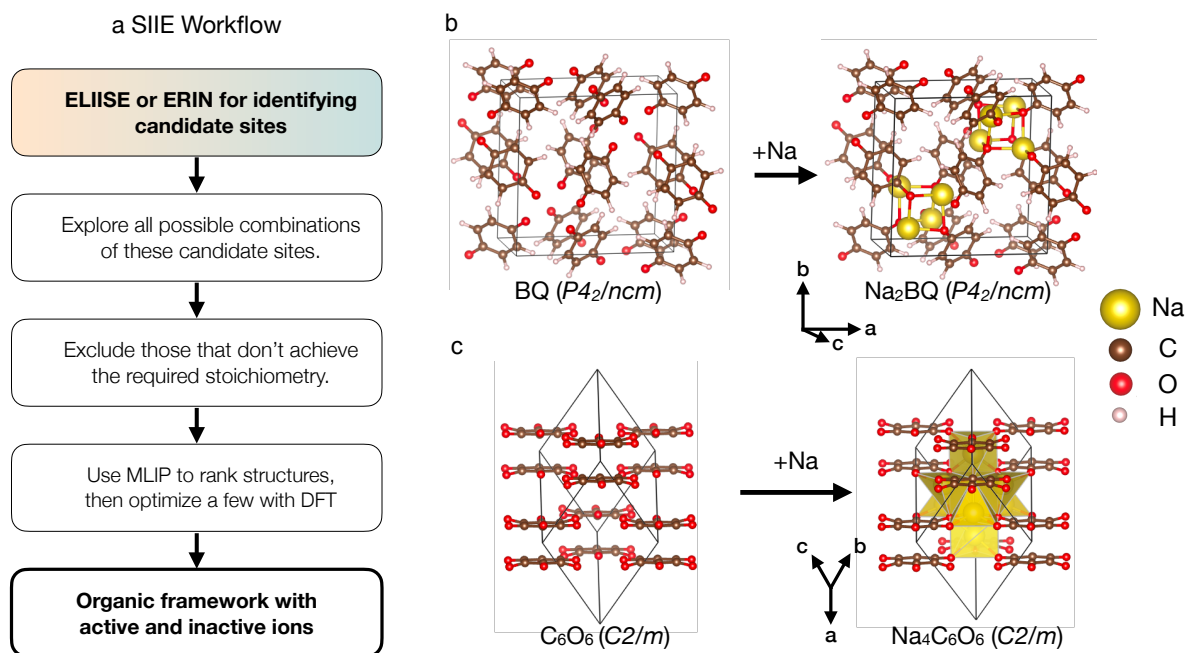
a SIIE Workflow

ELIISE or ERIN for identifying candidate sites

↓

Explore all possible combinations of these candidate sites.

↓

Exclude those that don't achieve the required stoichiometry.

↓

Use MLIP to rank structures, then optimize a few with DFT

↓

Organic framework with active and inactive ions

BQ ($P4_2/ncm$)   +Na→   Na$_2$BQ ($P4_2/ncm$)

C$_6$O$_6$ ($C2/m$)   +Na→   Na$_4$C$_6$O$_6$ ($C2/m$)

Na
C
O
H

Figure 2: Workflow for the Simultaneous Ion Insertion and Evaluation (SIIE) method (a) and its application in two organic-based materials. (b) Disodium 1,4-benzoquinone (Na$_2$BQ) and (c) Tetrasodium rhodizonate (Na$_4$C$_6$O$_6$). Combining SIIE with ELIISE, the Na-ion positions are located in two exemplary electro-active organic materials.[36,58]

Results

Case Studies: Implementing the Suggested Workflows on Example Systems

To demonstrate the effectiveness of these algorithms, we showcase the application of the ELIISE (ERIN) + SIIE workflow to several model compounds. The accuracy of ELIISE+SIIE is verified by "rediscovering" the Na-ion sites in two known fully sodiated compounds: (i) the disodium 1,4-benzoquinone ($P4_2/ncm$, Na_2BQ),[36] and (ii) the tetra-sodium rhodizonate with formula $Na_4C_6O_6$ ($C2/m$).[58] Similarly, starting from the known dilithium 2,6-naphthalenedicarboxylate ($P2_1/c$, Li_2NDC), using ERIN+SIIE, we predict Li positions in Li_4NDC (tetra-lithium 2,6-naphthalenedicarboxylate),[33,59] which contains 2 additional Li sites per formula unit (f.u.).

$Na_4C_6O_6$ ($C2/m$)[58] is an oxocarbon-salt-based layered compound; it is a planar conjugated ring with six carbonyl oxygens, known to undergo a two-to-four electron reduction with Na. To test the ELIISE method, we first remove all Na atoms from $Na_4C_6O_6$ and perform a static DFT calculation, at the experimentally determined volume and geometry, to calculate the electronic charge density and local electrostatic potential fields and subsequently apply the ELIISE algorithm to find the possible sites for Na-ions (as described in **Methods Sec. 2**).

The ELIISE algorithm yields nine symmetry-distinct Na sites with different multiplicities (**Table S2**). These candidate sites from ELIISE are then "funneled" into the SIIE workflow. We examine all possible combinations of sites from these nine symmetry-distinct candidates that produce the desired $Na_4C_6O_6$ stoichiometry. This yields only 59 unique sodium configurations, for which we perform DFT and MACE optimizations

to identify the most plausible Na-atom positions. To quantify the prediction accuracy of ELIIER+SIIE, we use a metric called the maximum atom displacement error (MADE), and defined in **Eq. 3**:

$$\text{MADE} = Max\{|r_i - r_i^{ref}|;\ i \in \textbf{all active and inactive species}\} \qquad \textbf{Eq. 3}$$

Where r_i is the predicted ion position of the ith ion, and r_i^{ref} is the position of the corresponding ion from the literature. The predicted and literature ion pairs are matched based on the pairing that minimizes the sum of their absolute deviations. Thus, MADE (**Eq. 3**) is the maximum displacement between any predicted ion position and its corresponding literature reference, i.e., the actual atom positions reported in the literature.[58] MADE is defined when the organic framework of the discharged structure is known and remains fixed during the optimization of ionic positions. MADE measures the accuracy of the predicted sites by indicating how much the ions deviate from literature positions.

In practice, we interpret MADE values ≤~0.4 Å as indicating correct predictions for all ions in the structure, while MADE values ≥~0.4 Å suggest that at least one ion site has been predicted incorrectly. Among the predicted structures for $Na_4C_6O_6$, the lowest-energy structure of **Figure 2c** displays a MADE of ~0.026 Å, indicating excellent agreement with the structure previously reported in the literature.[58]

In the second example, we investigate the structure of disodium 1,4-benzoquinone, Na_2BQ^{36} (**Figure 2b**), and the fully reduced (sodiated) form of a para-benzoquinone,

i.e., BQ. In Na_2BQ, the Na atom positions are available from the literature.[36] In **Figure 2b**, the Na_2BQ features a highly open organic framework with large voids, making a brute-force search of the potential Na positions computationally expensive. Using ELIISE, we identify four symmetrically distinct candidate sites. Applying the SIIE workflow, we evaluate all combinations of these four symmetry-distinct ions that produce the stoichiometry of Na_2BQ. This step yields only three valid configurations, which we used for further DFT structural relaxations (keeping the volume and shape fixed to the experimental structure of Na_2BQ). The most stable structure with the lowest DFT total energy singled out with ELIISE+SIIE, shown in **Figure 2b**, exhibits a MADE of ~0.024 Å, demonstrating excellent agreement with the experimental Na positions.[36]
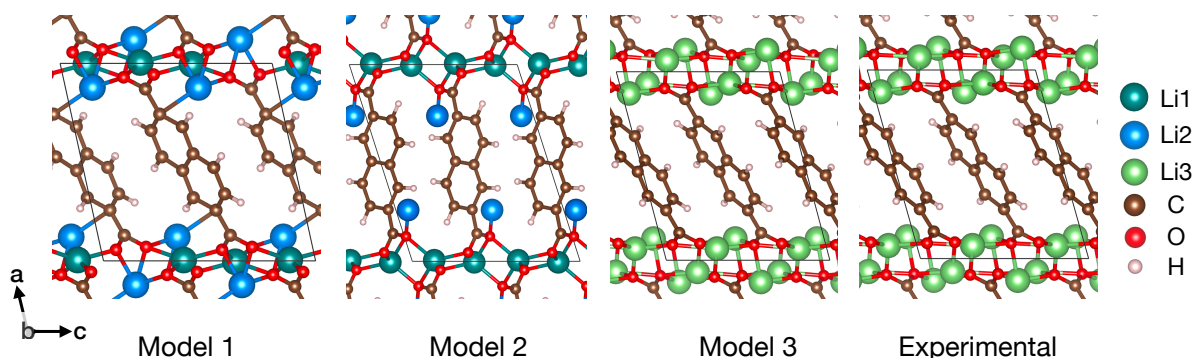
Figure 3: **Li_4NDC structures:** Model 1, Model 2, Model 3, and experimentally refined from single-crystal XRD.[33] Different crystallographic Li sites are shown.

We tested our methods on a third structure, Li_4NDC, which shares structural similarities to its precursor Li_2NDC,[60] with 2 Li atoms per molecule. Prior experimental studies indicated that the NDC organic framework largely remains unchanged upon reaction with Li. Two structural Li arrangements, Model 1 and Model 2 (**Figure 3**), have been proposed for Li_4NDC.[59] Models 1 and 2 preserve Li-ions in tetrahedral coordination sites as observed in Li_2NDC[60] (shown as Li1 in **Figure 3**) and introduce

new Li positions (shown as Li2 in **Figure 3**), approaching composition Li_4NDC (four Li atoms per molecule).

Using ERIN with SIIE workflow, we predict the Li positions in Li_4NDC based on the organic framework of Li_2NDC as input. We recovered Models 1 and 2, described in the literature. A close inspection of Model 1 and Model 2 reveals that these two structures reported in Ref. [59] are different in the relative orientation of NDC molecules, as displayed in **Figure 3**. However, this striking contrast is not highlighted in Ref. [59], which seems to suggest that Model 1 and Model 2 differ only by Li distribution but maintain the same molecular framework.

Along with Models 1 and 2, we identified a different model as the lowest-energy structure, as determined by DFT minimization (and MACE), which we named Model 3. In Model 3, the tetrahedral sites reported for Li_2NDC are unoccupied, and all the Li atoms assume new crystallographic positions (collectively indicated as Li3 in **Figure 3**), which are localized near the inorganic layer of the organic framework, consisting of oxygen atoms in NDC. These newly identified Li-atom positions of Model 3 provide an electrostatically favorable arrangement of the Li atoms compared to Models 1 and 2, at the same composition. Quantitatively, DFT total energy comparisons show that Model 3 is energetically favored over both Model 1 and Model 2 by ~100 meV/$Li_4H_6C_{12}O_4$. and ~440 meV/$Li_4H_6C_{12}O_4$, respectively.

Indeed, Model 3 is in excellent agreement with a recent study (appeared while drafting this paper) that used single-crystal XRD to refine the structure of Li_4NDC. [33] Because

ERIN uses the organic framework of the charged state as input and optimizes all atomic positions, the MADE metric—measuring the prediction error in active species positions—is not suitable. Instead, a simple qualitative measure of similarity between two structures can be obtained using, for example, StructureMatcher (using default parameters) as implemented in Pymatgen.[61] Here, StructureMatcher indicated that model 3 qualitatively matches the recently reported experimental structure of Li_4NDC.[33]

These three case studies, discussed in the previous paragraphs, collectively demonstrate the robustness of the ELIISE and ERIN methods when integrated with the SIIE workflow in accurately identifying ion positions across various organic electrode materials.

Assessing the accuracy of the ELIISE+SIIE approach

To assess the accuracy of the ELIISE + SIIE approach, we apply it to a diverse set of organic electrode materials (16 in total) for which the organic framework and ion positions of the crystal structure are known from literature. Using MADE (**Eq. 3**) as a metric for similarity, we evaluate the performance of the ELIISE method with the SIIE workflow across 16 organic systems. These structure organic electrode materials, such as Li_2NDC and Li_4NDC (di and tetra lithium 2,6-naphthalene dicarboxylate),[60] Li_2BDC and Li_4BDC (di and tetra lithium 1,4-benzene dicarboxylate),[62] Li_2-BPDC (dilithium biphenyl dicarboxylate),[63] Li_2Mg-p-DHT (magnesium(2,5-dilithium-oxy)-terephthalate),[31] Na_2BQ (disodium hydroquinone),[36] $Na_2C_6O_6$ (disodium rhodizonate),[64] and $Na_4C_6O_6$ (tetra sodium rhodizonate),[58] as well as select structures

(see **Table S1**) obtained from crystallographic databases, the Cambridge Crystallographic Data Centre (CCDC)[65] and the Inorganic Crystal Structure Database (ICSD).[66]
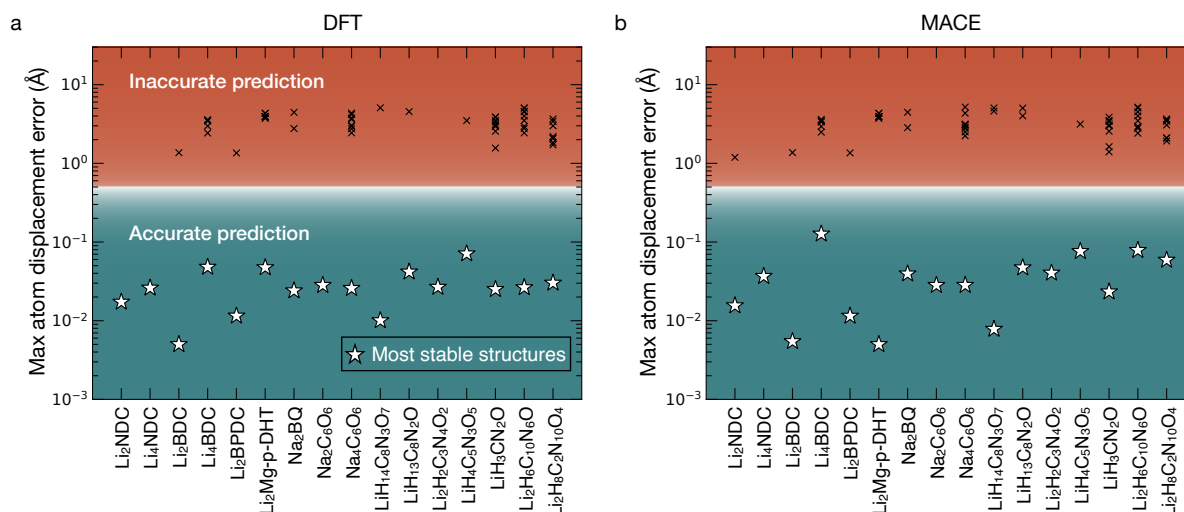

Figure 4: Accuracy of the predicted active ion positions with the proposed SIIE workflow and ELIISE. MADE (Eq. 3) values determined after optimization of ionic positions with (a) first-principles DFT and (b) the machine learned potential MACE. Stars indicate the most stable (lowest total energy) ion arrangements in each organic structure, and crosses indicate the thermodynamically unfavorable (high energy) structures. Details of reference structures reported in the literature are in **Table S1**.

As shown in **Figure 4a**, the lowest-energy structure identified through the SIIE workflow using ELIISE correctly reproduces the known ion positions in all tested cases, as measured by MADE, whose value is always ≤ 0.4 Å in all the cases considered. Furthermore, **Figure 4b** demonstrates that using a pre-trained machine learning interatomic potential (MLIP), specifically MACE (only used for optimizing ion positions), yields the correct prediction of ion positions, with MADE values consistently lower than 0.4 Å. This suggests that DFT-level structural relaxation may not be strictly necessary for determining the general location of inserted ions, leading to a substantial reduction in computational cost. **Figure S1** presents a similar assessment, but with predictions generated using ERIN+SIIE instead of ELIISE+SIIE, and only MACE was employed for optimizing ion positions. The MADE values in **Figure S1** show that

ERIN+SIIE also accurately predicted the positions of all inserted ion sites for all 16 tested organic systems.

Discussion

Predicting ion positions in "soft" crystalline materials, such as organic electrode materials, remains a significant challenge due to the combination of large free volumes, weak intermolecular forces, and the dynamic nature of ion coordination. In this paper, we have developed two new methodologies, ELIISE and ERIN, along with an ion insertion workflow, SIIE, designed to identify ion insertion sites in organic structures. These include metal atoms, such as Na, Zn, and Mg, as well as light elements, for example, Li. Here, we discuss the pros and cons of implementing the ELIISE and ERIN algorithms for resolving unknown ion positions in crystal structures, where traditional structural techniques fail.

ELIISE uses both the electronic charge density (CD) and the local electrostatic potential $LP(r)$ derived from first-principles calculations of the empty organic framework. In combination, the CD and the $LP(r)$ are used to find candidate crystallographic sites for light elements in these materials. However, since the CD and the $LP(r)$ are intrinsically linked (as the latter is derived from atomic positions and the electronic charge density distribution), the independent use of the CD and the $LP(r)$ often arrives at similar predictions for candidate ion sites. Since ELIISE directly identifies coordinating environments—regions surrounded by electronegative atoms of the organic frameworks, ELIISE appears most effective for systems where the organic framework structure of the discharged (ion-inserted) phase is known. This

signifies that prior experiments determining the space group, lattice constants, and atom positions of the organic framework (excluding the mobile ions) must be available. In some sense, this limits the capabilities of the ELIISE strategy. For example, if only the charged-phase structure of an organic system is available and significant structural changes and/or phase transitions are expected upon electrochemical reactions, using ELIISE may yield unreliable results.

To address the apparent limitation of ELIISE, we have developed an alternative method called ERIN. ERIN generates a sparse set of points that cover all possible Wyckoff sites, ranked from the electrostatically lowest-energy sites to the highest-energy sites. Therefore, ERIN should be applied when the structure of the organic framework, whose ion-inserted/discharged phases remain unknown. In such cases, the charged (empty) structure of the materials can be used as a starting point, assuming the unit cell of the pristine organic crystal shape and most of the symmetry remains unchanged or structural modifications occur gradually (for example, a displacive phase transition) upon reaction with Li (or sodium).

ELIISE identifies the stationary sites, i.e., the local minima, by populating the energy landscape; these sites are expected to be stationary points for ions in the potential created by the anion framework. In contrast, in ERIN, the identified sites are not necessarily located at energy minima –non-stationary sites, such as saddle points, on the potential energy surface, and may be mobile during structural relaxation. These non-stationary sites can perturb the organic framework during their relaxation and

promote structural changes that lead the organic framework toward a more stable discharged structure.

Even if the discharge phase does not involve phase transitions, the specific coordination environments and lattice parameters required to host mobile metal atoms in the discharged organic framework may not be available in the charged organic framework. ERIN systematic electrostatic screening helps identify plausible insertion sites that would be hard to locate with ELIISE alone. Note, however, that ERIN predicts more candidate sites than ELIISE, resulting in a greater number of possible ion-vacancy arrangements to evaluate in the SIIE workflow.

Our study has focused on identifying plausible sites for cation insertion; however, with suitable modifications, these algorithms can also be used to locate anion positions in a crystal structure. For example, in the case of the ELIISE method, local minima in the $ALP(r)$ should be used instead of local maxima. The approach remains the same when using CD. In ERIN, instead of selecting the maximum in the $\mathrm{ALP}(\mathrm{r})$ as a candidate site, the minimum in the $\mathrm{ALP}(\mathrm{r})$ should be chosen.

At its core, ERIN uses the Coulomb electrostatic potential to model the energy landscape and identify sites generated from electronic charge densities obtained from first-principles calculations. Other methods, such as machine learning interatomic potentials [57,67,68] or Bond Valence Sums[69] appear viable for modeling and developing the potential energy landscape for ion insertion.

In general, when testing the ELIISE and ERIN algorithms on 16 previously known structures (**Figure 4** and **Figure S1**), we demonstrated how these algorithms, in combination with the SIIE workflow, offer a reliable and systematic approach for accurately predicting unknown ion positions in organic frameworks.

Another important aspect is the integration of machine learning interatomic potentials (MLIPs), such as MACE, into the SIIE workflow. In principle, ERIN and ELIISE are applicable together with any foundational MLIP model, including ME3GNET, CHGNET, and ORB-3, among others.[57,67,68] This hybrid MLIP–DFT approach appears to be viable, significantly reducing the computational burden of exploring vast combinatorial site occupancies while providing sufficient accuracy in predicting ion positions. Furthermore, the ability of MLIPs to screen thousands of candidate configurations before subsequent, more accurate but computationally intensive refinements using first-principles calculations highlights a scalable pathway for high-throughput exploration of ion-insertion chemistries in soft systems, such as organic crystals, porous systems, metal-organic frameworks, and larger molecules with biological relevance.

In this vein, we have tested MACE predictions against those obtained by DFT on the same systems of **Figure 4**. We have observed that when the organic framework is kept fixed at the known discharged structure, identifying the structure with the lowest energy and correct ion arrangements using (MADE) prediction error within a tolerable range of 0.4 does not require DFT; foundational MLIPs without re-training appear suited for this task. Here, we have only tested MACE.[55,56]

A limitation exists within the SIIE workflow when the number of possible ion combinations generated is in the order of 10^4 to 10^6, as this can still require significant computational resources to evaluate total energies across all configurations, even with MLIP such as MACE. In this context, we had relied on the fact that candidate sites are ranked in decreasing order of electrostatic favorability, as indicated by the averaged local potential. Unfavorable sites are systematically excluded during the generation of ion combinations; this removal process continues until the number of valid ion combinations falls below the preset threshold ($\sim 10^3$).

Aside from ELIISE and ERIN, the simultaneous insertion of ions, implemented in the SIIE workflow, is also essential for accurately finding discharged structures, as illustrated by the case study of Li_4NDC. In Li_4NDC, we proposed Model 3, which yields the lowest DFT total energy (**Figure 3**), and is lower than that of the previously proposed Models 1 and 2.[59] Before this work,[51] other strategies, such as sequential insertion methods, have been proposed.[48,54] In the sequential approach, a new model is created by inserting an atom in a new position in the unit cell, and the structure positions are optimized. This procedure is performed multiple times until the expected stoichiometry is achieved, in the hope of finding a representative global minimum. The sequential approach can lead to incorrect selection of stable candidates because ions may become trapped in local minima of the potential energy surface.

In the case of Li_4NDC, determining the optimal ion positions was uncomplicated because the organic structure remained essentially unchanged during lithiation.

However, this is not the case for most systems. For example, when Li_2BDC is lithiated, forming Li_4BDC, a symmetry reduction is observed.[33] It involves doubling the unit cell along the stacking direction of the organic layers, accompanied by changes in the organic framework.[33] Consequently, attempting to predict ion positions in Li_4BDC by starting directly from the Li_2BDC structure would produce an inaccurate model.

These cases highlight a broader challenge: it is often unclear whether a significant phase transition occurs during the lithiation process. When this uncertainty arises, a practical approach is first to predict ion positions based on the charged structure (i.e., the host in its oxidized or delithiated form). The resulting model can then be compared with experimental data such as XRD or ND; if the simulated patterns show similarity to the experimental patterns, the predicted structure can serve as a solid starting point for refinement.

However, if the comparison shows significant differences, it becomes necessary to explore alternative structural hypotheses. In such cases, one must generate and assess a set of candidate organic framework motifs to serve as initial models in the ERIN+SIIE workflow. Developing systematic methods for constructing and testing these organic framework variations appears crucial for reliably understanding the structural complexity of ion insertion in molecular crystals.

Besides XRD and ND for determining the structures of organic materials, solid-state nuclear magnetic resonance (ss-NMR) crystallography is another valuable tool that offers structural insights.[70,71] The nascent application of ss-NMR in conjunction with

first-principles calculations, termed as NMR crystallography, has been used to establish the structure of functional materials.[72]

Conclusions

In summary, we have developed a computational framework for accurately predicting ion positions in organic materials, utilizing DFT-derived electrostatic and charge density fields with the ELIISE and ERIN algorithms and the SIIE workflow.

Through case studies of $Na_4C_6O_6$, Na_2BQ, and Li_4NDC, we showed that ELIISE and ERIN methodologies can reproduce ion positions known from the literature with sub-angstrom accuracy. Additionally, we demonstrate that these methodologies can uncover new, energetically favorable structural models that can improve our understanding of experimental observations.

We have demonstrated that computationally intensive DFT predictions are not a strong requirement for accurately identifying ion positions, and we have shown that foundational machine-learned potentials provide sufficient accuracy for large-scale screening tasks.

Our results demonstrate that ELIISE and ERIN, combined with SIIE, are powerful and can significantly improve, if not augment, the limits of traditional structural characterization techniques. The framework proposed in this paper offers a pathway toward deeper mechanistic insights and the rational design of organic materials for next-generation functional materials. Moreover, MLIPs can efficiently screen thousands of candidate configurations initially, before performing more precise but computationally expensive refinements with first-principles methods. This approach

offers a scalable pathway for high-throughput exploration of ion-insertion chemistries in soft systems, including organic crystals, porous materials, metal-organic frameworks, and larger biologically relevant molecules.

Computational Details

All first-principles calculations presented in this work were performed using the density functional theory (DFT) formalism, as implemented in the Vienna Ab initio Simulation Package (VASP).[73–75] The PAW potentials describe the core electrons. Perdew, Burke, and Ernzerhof (PBE)[76] was used to approximate the unknown DFT exchange and correlation XC functional. To account for Van der Waals interactions, the empirical D3 method proposed by Grimme and collaborators, with Becke-Johnson (BJ) damping, was used.[77,78]

VASP inputs prepared for geometry optimization and energy calculations closely followed the MITRelaxSet[79] as available in pymatgen.[61] The kinetic energy cutoff for the plane waves was set to 520 eV, and the total energy was converged to 10^{-5} eV per cell. Geometries (coordinates, volumes, and cell shapes) were considered converged when the forces on all atoms are lower than 0.05 eV/Å. A Γ-centered Monkhorst–Pack[80] grid with a density of 25 k-points per Å^{-1} for all systems. With these DFT settings, local potentials are computed as the sum of Ewald and Hartree parts, excluding the Exchange and Correlation (XC) contributions.

To reduce the number of structures for DFT optimization, a MACE-MP-0 machine-learned foundational model, based on MACE,[55,56] with PFP-based PyTorch implementation of DFT-D3,[77,78,81] was used to evaluate static total energy and/or optimize ion positions. Structural relaxations are carried out until the forces are converged below 0.05 eV/Å.

When MACE is implemented in the SIIE workflow is used as follows:

1. Generate all possible ion combinations. If the number of combinations exceeds a preset limit ($\sim 10^3$), they can be reduced systematically by removing electrostatically unfavorable sites in ELIISE or ERIN before generating possible ion combinations.

2. Use MACE to evaluate all the generated combinations.

3. Select approximately $\sim 10^2$ of the lowest-energy candidate structures for optimization of ion positions using MACE.

4. From these structures, we choose approximately 10 unique lowest-energy structures for further accurate DFT optimization of the atomic positions.

This hybrid approach, which combines the computational speed of MACE-based MLIPs with the accuracy of DFT, enables us to successfully identify the correct atomic configurations in organic frameworks with substantial structural complexity.

Data Availability

All the computational data associated with this study, including the input and output files of the simulations, are available on GitHub at https://github.com/caneparesearch/data_site_finder

Author contributions

H.R.G.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **A.E.L.:** Data curation, Discussion, Writing – original draft, Writing – review & editing. **A.A.P.:** Discussion, Writing – review & editing. **Y.C.:** Discussion, Writing – review & editing. **V. S.C.K.:** Discussion, Writing – review & editing. **L.J.:** Discussion, Writing – review & editing. **J.W.:** Discussion, Writing – review & editing. **Y.C.:**

Discussion, Writing – review & editing. **K.W.:** Discussion, Writing – review & editing. **M.K.Y.C.:** Discussion, Writing – review & editing. **Y.Y.:** Funding acquisition, Supervision, Discussion, Writing – review & editing. **P.C.:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

References

1. Lakraychi, A. E., Dolhem, F., Vlad, A. & Becuwe, M. Organic Negative Electrode Materials for Metal-Ion and Molecular-Ion Batteries: Progress and Challenges from a Molecular Engineering Perspective. *Advanced Energy Materials* **11**, 2101562 (2021).
2. Zhang, Y. *et al.* Ionically conducting Li- and Na-phosphonates as organic electrode materials for rechargeable batteries. *Chem. Sci.* **16**, 1819–1825 (2025).
3. Lu, Y. & Chen, J. Prospects of organic electrode materials for practical lithium batteries. *Nat Rev Chem* **4**, 127–142 (2020).
4. Dai, H., Guan, L., Mao, M. & Wang, C. Evaluating the present and future of organic batteries. *Nat. Rev. Clean Technol.* **1**, 493–510 (2025).
5. Zhao, L., Lakraychi, A. E., Chen, Z., Liang, Y. & Yao, Y. Roadmap of Solid-State Lithium-Organic Batteries toward 500 Wh kg^{-1}. *ACS Energy Lett.* **6**, 3287–3306 (2021).
6. Poizot, P. *et al.* Opportunities and Challenges for Organic Electrodes in Electrochemical Energy Storage. *Chem. Rev.* **120**, 6490–6557 (2020).
7. Blatov, V. A., Carlucci, L., Ciani, G. & Proserpio, D. M. Interpenetrating metal–organic and inorganic 3D networks: a computer-aided systematic investigation. Part I. Analysis of the Cambridge structural database. *CrystEngComm* **6**, 377–395 (2004).

8. Blatov, V. A., Shevchenko, A. P. & Serezhkin, V. N. *TOPOS* 3.2: a new version of the program package for multipurpose crystal-chemical analysis. *J Appl Crystallogr* **33**, 1193–1193 (2000).

9. Blatov, V. A., Shevchenko, A. P. & Proserpio, D. M. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro. *Crystal Growth & Design* **14**, 3576–3586 (2014).

10. Canepa, P. *et al.* Odyssey of Multivalent Cathode Materials: Open Questions and Future Challenges. *Chem. Rev.* **117**, 4287–4341 (2017).

11. Tian, Y. *et al.* Promises and Challenges of Next-Generation "Beyond Li-ion" Batteries for Electric Vehicles and Grid Decarbonization. *Chem. Rev.* **121**, 1623–1669 (2021).

12. Yabuuchi, N., Kubota, K., Dahbi, M. & Komaba, S. Research Development on Sodium-Ion Batteries. *Chem. Rev.* **114**, 11636–11682 (2014).

13. Delmas, C. Sodium and Sodium-Ion Batteries: 50 Years of Research. *Advanced Energy Materials* **8**, 1703137 (2018).

14. Slater, M. D., Kim, D., Lee, E. & Johnson, C. S. Sodium-Ion Batteries. *Adv Funct Materials* **23**, 947–958 (2013).

15. Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *J Appl Crystallogr* **2**, 65–71 (1969).

16. The structures of some titanium-niobium oxides by powder neutron diffraction. *Proc. R. Soc. Lond. A* **338**, 311–326 (1974).

17. Cheetham, A. K. & Von Dreele, R. B. Cation Distributions in Niobium Oxide Block Structures. *Nature Physical Science* **244**, 139–140 (1973).

18. Runčevski, T. & Brown, C. M. The Rietveld Refinement Method: Half of a Century Anniversary. *Crystal Growth & Design* **21**, 4821–4822 (2021).

19. Wimmer, D. S. *et al.* The crystal structure and luminescence properties of the first lithium oxonitridolithosilicate $Li_3 SiNO_2 :Eu^{2+}$. *Dalton Trans.* **51**, 16465–16478 (2022).

20. Buschmann, H. *et al.* Structure and dynamics of the fast lithium ion conductor "Li7La3Zr2O12". *Phys. Chem. Chem. Phys.* **13**, 19378 (2011).

21. Kisi, E. H. & Howard, C. J. *Applications of Neutron Powder Diffraction*. (Oxford University Press, 2008). doi:10.1093/acprof:oso/9780198515944.001.0001.

22. Chen, T. *et al.* High-rate, high-capacity electrochemical energy storage in hydrogen-bonded fused aromatics. *Joule* **7**, 986–1002 (2023).

23. Lu, Y. *et al.* High-capacity dilithium hydroquinone cathode material for lithium-ion batteries. *National Science Review* **11**, nwae146 (2024).

24. Chen, T. *et al.* High-Energy, High-Power Sodium-Ion Batteries from a Layered Organic Cathode. *J. Am. Chem. Soc.* **147**, 6181–6192 (2025).

25. Chen, T. *et al.* A Layered Organic Cathode for High-Energy, Fast-Charging, and Long-Lasting Li-Ion Batteries. *ACS Cent. Sci.* **10**, 569–578 (2024).

26. Zhang, H. *et al.* Organic Cathode Materials for Sodium-Ion Batteries: From Fundamental Research to Potential Commercial Application. *Adv Funct Materials* **32**, 2107718 (2022).

27. Wang, J. *et al.* Conjugated sulfonamides as a class of organic lithium-ion positive electrodes. *Nat. Mater.* **20**, 665–673 (2021).

28. Jouhara, A. *et al.* Raising the redox potential in carboxyphenolate-based positive organic materials via cation substitution. *Nat Commun* **9**, 4401 (2018).

29. Sieuw, L. *et al.* Through-Space Charge Modulation Overriding Substituent Effect: Rise of the Redox Potential at 3.35 V in a Lithium-Phenolate Stereoelectronic Isomer. *Chem. Mater.* **32**, 9996–10006 (2020).

30. Sieuw, L. *et al.* A H-bond stabilized quinone electrode material for Li–organic batteries: the strength of weak bonds. *Chem. Sci.* **10**, 418–426 (2019).

31. Jouhara, A. *et al.* Unveiling the Peierls effect during electrochemical delithiation/lithiation process in magnesium (2,5-dilithium-oxy)-terephthalate using combined experimental and computational studies. *Energy Storage Materials* **81**, 104456 (2025).

32. Ogihara, N., Ohba, N. & Kishida, Y. On/off switchable electronic conduction in intercalated metal-organic frameworks. *Sci. Adv.* **3**, e1603103 (2017).

33. Griffiths, K. *et al.* Chemical lithiation reveals the structures of conjugated tetralithium dicarboxylates. *Chem. Commun.* **61**, 11991–11994 (2025).

34. Lee, M. *et al.* High-performance sodium–organic battery by realizing four-sodium storage in disodium rhodizonate. *Nat Energy* **2**, 861–868 (2017).

35. Wu, X. *et al.* Unraveling the storage mechanism in organic carbonyl electrodes for sodium-ion batteries. *Sci. Adv.* **1**, e1500330 (2015).

36. Couhorn, U. & Dronskowski, R. Alkali Metal Compounds of Hydroquinone: Synthesis and Crystal Structure. *Zeitschrift anorg allge chemie* **629**, 647–652 (2003).

37. Fédèle, L. *et al.* 2D-Layered Lithium Carboxylate Based on Biphenyl Core as Negative Electrode for Organic Lithium-Ion Batteries. *Chem. Mater.* **29**, 546–554 (2017).

38. Kim, H. *et al.* The Reaction Mechanism and Capacity Degradation Model in Lithium Insertion Organic Cathodes, $Li_2C_6O_6$, Using Combined Experimental and First Principle Studies. *J. Phys. Chem. Lett.* **5**, 3086–3092 (2014).

39. Rambabu, D. *et al.* Mixed proton-electron conductivity in a dynamic 3D metal-organic framework. *Chem* 102590 (2025) doi:10.1016/j.chempr.2025.102590.

40. Rambabu, D. *et al.* An Electrically Conducting Li-Ion Metal–Organic Framework. *J. Am. Chem. Soc.* **143**, 11641–11650 (2021).

41. Bernard, L. *et al.* Influence of Polymorphism on the Electrochemical Behavior of Dilithium (2,3-Dilithium-oxy)-terephthalate vs. Li. *Inorganics* **10**, 62 (2022).

42. Marchiori, C. F. N., Brandell, D. & Araujo, C. M. Predicting Structure and Electrochemistry of Dilithium Thiophene-2,5-Dicarboxylate Electrodes by Density Functional Theory and Evolutionary Algorithms. *J. Phys. Chem. C* **123**, 4691–4700 (2019).

43. Lu, Z. & Zhu, B. Crystal Structure Prediction for Battery Materials. in *Computational Design of Battery Materials* (ed. Hanaor, D. A. H.) vol. 150 187–210 (Springer International Publishing, Cham, 2024).

44. Liu, L. *et al.* Molecular Electrostatic Potential: A New Tool to Predict the Lithiation Process of Organic Battery Materials. *J. Phys. Chem. Lett.* **9**, 3573–3579 (2018).

45. Quarez, É. *et al.* From partial to complete neutralization of 2,5-dihydroxyterephthalic acid in the Li–Na system: crystal chemistry and electrochemical behavior of $Na_2Li_2C_8H_2O_6$ *vs.* Li. *CrystEngComm* **22**, 1653–1663 (2020).

46. Wang, J. *et al.* Revealing the reversible solid-state electrochemistry of lithium-containing conjugated oximates for organic batteries. *Sci. Adv.* **9**, eadg6079 (2023).

47. Guo, X. *et al.* Towards the 4 V-class n-type organic lithium-ion positive electrode materials: the case of conjugated triflimides and cyanamides. *Energy Environ. Sci.* **17**, 173–182 (2024).

48. Shen, J.-X., Horton, M. & Persson, K. A. A charge-density-based general cation insertion algorithm for generating new Li-ion cathode materials. *npj Comput Mater* **6**, 161 (2020).

49. Rong, Z., Kitchaev, D., Canepa, P., Huang, W. & Ceder, G. An efficient algorithm for finding the minimum energy path for cation migration in ionic materials. *The Journal of Chemical Physics* **145**, 074112 (2016).

50. Zimmermann, N. E. R. *et al.* Electrostatic Estimation of Intercalant Jump-Diffusion Barriers Using Finite-Size Ion Models. *J. Phys. Chem. Lett.* **9**, 628–634 (2018).

51. Pascazio, R., Chen, Q., Li, H. H., Kaplan, A. D. & Persson, K. A. Toward High-Voltage Cathodes for Zinc-Ion Batteries: Discovery Pipeline and Material Design Rules. *Chem. Mater.* **37**, 6213–6226 (2025).

52. Zhang, Y. Y., Sun, Y. Y., Du, S. X., Gao, H.-J. & Zhang, S. B. Organic salts as super-high rate capability materials for lithium-ion batteries. *Applied Physics Letters* **100**, 091905 (2012).

53. Slater, J. C. Atomic Radii in Crystals. *The Journal of Chemical Physics* **41**, 3199–3204 (1964).

54. Tateyama, Y., Kagatsume, A., Yao, M., Matsuda, S. & Uosaki, K. Exploration of Organic Cathode Active Materials with High Energy Densities for Li-Ion Batteries via First-Principles Calculations. *J. Phys. Chem. C* **127**, 12867–12873 (2023).

55. Batatia, I. *et al.* The design space of E(3)-equivariant atom-centred interatomic potentials. *Nat Mach Intell* **7**, 56–67 (2025).

56. Batatia, I., Kovács, D., Simm, G., Ortner, C. & Csányi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. Preprint at https://doi.org/10.17863/CAM.111586 (2022).

57. Rhodes, B. *et al.* Orb-v3: atomistic simulation at scale. Preprint at https://doi.org/10.48550/ARXIV.2504.06231 (2025).

58. Yamashita, T., Momida, H. & Oguchi, T. First-Principles Investigation of a Phase Transition in Na $_x$ C_6 O_6 as an Organic Cathode Material for Na-ion Batteries: Role of Intermolecule Bonding of C_6 O_6. *J. Phys. Soc. Jpn.* **84**, 074703 (2015).

59. Ogihara, N. *et al.* Organic Dicarboxylate Negative Electrode Materials with Remarkably Small Strain for High-Voltage Bipolar Batteries. *Angew Chem Int Ed* **53**, 11467–11472 (2014).

60. Banerjee, D., Kim, S. J. & Parise, J. B. Lithium Based Metal–Organic Framework with Exceptional Stability. *Crystal Growth & Design* **9**, 2500–2503 (2009).

61. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).

62. Kaduk, J. A. Terephthalate salts: salts of monopositive cations. *Acta Crystallogr B Struct Sci* **56**, 474–485 (2000).

63. Banerjee, D., Borkowski, L. A., Kim, S. J. & Parise, J. B. Synthesis and Structural Characterization of Lithium-Based Metal–Organic Frameworks. *Crystal Growth & Design* **9**, 4922–4926 (2009).

64. Dinnebier, R. E., Nuss, H. & Jansen, M. Disodium rhodizonate: a powder diffraction study. *Acta Crystallogr E Struct Rep Online* **61**, m2148–m2150 (2005).

65. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **72**, 171–179 (2016).

66. Rühl, S. The Inorganic Crystal Structure Database (ICSD): A Tool for Materials Sciences. in *Materials Informatics* (eds Isayev, O., Tropsha, A. & Curtarolo, S.) 41–54 (Wiley, 2019). doi:10.1002/9783527802265.ch2.

67. Deng, B. *et al.* CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat Mach Intell* **5**, 1031–1041 (2023).

68. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat Comput Sci* **2**, 718–728 (2022).

69. Chen, H., Wong, L. L. & Adams, S. *SoftBV* – a software tool for screening the materials genome of inorganic fast ion conductors. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **75**, 18–33 (2019).

70. Bryce, D. L. NMR crystallography: structure and properties of materials from solid-state nuclear magnetic resonance observables. *IUCrJ* **4**, 350–359 (2017).

71. Harris, K. D. M. NMR Crystallography as a Vital Tool in Assisting Crystal Structure Determination from Powder XRD Data. *Crystals* **12**, 1277 (2022).

72. Whewell, T. *et al.* A structural investigation of organic battery anode materials by NMR crystallography. *Magnetic Reson in Chemistry* **60**, 489–503 (2022).

73. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **6**, 15–50 (1996).

74. Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).

75. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

76. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

77. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132**, 154104 (2010).

78. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **32**, 1456–1465 (2011).

79. Jain, A. *et al.* A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science* **50**, 2295–2310 (2011).

80. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).

81. Friede, M., Hölzer, C., Ehlert, S. & Grimme, S. *dxtb* —An efficient and fully differentiable framework for extended tight-binding. *The Journal of Chemical Physics* **161**, 062501 (2024).