



OPEN

Enhancing ReaxFF for molecular dynamics simulations of lithium-ion batteries: an interactive reparameterization protocol

Paolo De Angelis¹✉, Roberta Cappabianca¹, Matteo Fasano¹, Pietro Asinari^{1,2}✉ & Eliodoro Chiavazzo^{1,2}✉

Lithium-ion batteries (LIBs) have become an essential technology for the green economy transition, as they are widely used in portable electronics, electric vehicles, and renewable energy systems. The solid-electrolyte interphase (SEI) is a key component for the correct operation, performance, and safety of LIBs. The SEI arises from the initial thermal metastability of the anode-electrolyte interface, and the resulting electrolyte reduction products stabilize the interface by forming an electrochemical buffer window. This article aims to make a first—but important—step towards enhancing the parametrization of a widely-used reactive force field (ReaxFF) for accurate molecular dynamics (MD) simulations of SEI components in LIBs. To this end, we focus on Lithium Fluoride (LiF), an inorganic salt of great interest due to its beneficial properties in the passivation layer. The protocol relies heavily on various Python libraries designed to work with atomistic simulations allowing robust automation of all the reparameterization steps. The proposed set of configurations, and the resulting dataset, allow the new ReaxFF to recover the solid nature of the inorganic salt and improve the mass transport properties prediction from MD simulation. The optimized ReaxFF surpasses the previously available force field by accurately adjusting the diffusivity of lithium in the solid lattice, resulting in a two-order-of-magnitude improvement in its prediction at room temperature. However, our comprehensive investigation of the simulation shows the strong sensitivity of the ReaxFF to the training set, making its ability to interpolate the potential energy surface challenging. Consequently, the current formulation of ReaxFF can be effectively employed to model specific and well-defined phenomena by utilizing the proposed interactive reparameterization protocol to construct the dataset. Overall, this work represents a significant initial step towards refining ReaxFF for precise reactive MD simulations, shedding light on the challenges and limitations of ReaxFF force field parametrization. The demonstrated limitations emphasize the potential for developing more versatile and advanced force fields to upscale *ab initio* simulation through our interactive reparameterization protocol, enabling more accurate and comprehensive MD simulations in the future.

Lithium-ion batteries (LIBs) are and will continue to play an important role in the energy transition for energy storage and the fossil fuel replacement^{1–4}. In the last decades, the growing popularity of electric vehicles, and energy storage systems⁵ has fuelled the demand for batteries with greater capacity, more consistent performance over time, and improved safety⁶. While the materials used as anodes and cathodes largely determine battery capacity, the degradation phenomena that occur within the device govern battery stability and safety. One of the most important and yet poorly understood of these phenomena is the formation of a thin passivation film at the electrode-electrolyte interface, known as the solid electrolyte interphase (SEI)^{7–9}. This layer is formed by the irreversible reaction of lithium ions with the electrolyte due to the initial thermodynamic instability between the anode and electrolyte^{10,11}. The ions used to create the SEI are subtracted from the battery's capacity. Indeed the increase in the SEI over time is one of the causes of capacity fade observed during battery charging cycles¹². Additionally, the uncontrolled formation and growth of the SEI layer can limit the performance and life of current LIBs, and its degradation can lead to severe battery damage and uncontrolled exothermic reactions such as the

¹Department of Energy "Galileo Ferraris", Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy. ²Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135 Torino, Italy. ✉email: paolo.deangelis@polito.it; pietro.asinari@polito.it; eliodoro.chiavazzo@polito.it

battery thermal runaway¹³. Therefore, it is essential to model the SEI to understand and predict the behaviour of lithium-ion batteries and improve their performance and safety¹⁴. Despite the importance of SEI, it remains a conundrum due to its high reactivity and multiscale nature, which make challenging its study *in operando* and *in silico*, respectively^{3,15,16}. The theoretical understanding of the SEI may allow control of its final characteristics and it can enable the engineering of such layer, thus possibly leading to enhanced properties including better electronic insulation, prevention or alleviation of graphite anode exfoliation¹⁷ and dendrite formation in lithium metal batteries¹⁸. In addition, the modelling and subsequent theoretical understanding of SEI can contribute to significantly accelerate the discovery of new materials and electrolytes for future lithium-ion batteries^{3,16}.

In the past two decades, various atomistic techniques have been utilized to study the kinetics of electrolyte dissociation in lithium-ion batteries¹⁴. The density functional theory (DFT) method has proven particularly useful in predicting the effects of lithium salt and additives on the final components of the solid electrolyte interphase^{19–24}. Additionally, DFT has been used to investigate how the surface of the anode affects the electrolyte dissociation^{25–28} and to calculate various properties of SEI products, including ionic conductivity^{29–31}, electron-transfer properties^{28,32,33}, and elastic modulus^{34–36}. Although such simulations produce highly accurate results, they are computationally expensive. The cost of DFT simulations increases with the number of atoms in the system, typically scaling with $O(N^2)$ where N is the number of atoms³⁷, limiting the size of the system that can be simulated to a few thousand atoms and the time interval to a few picoseconds. To observe the evolution of the SEI over extended periods, one solution is to use approximate functions $V(\mathbf{r}_i)$, known as force fields (FF), to calculate the system's energy. Indeed, the physical intuition-based functions employed in many force fields enable the replacement of the expensive Kohn-Sham functional $E[n]$ ³⁸, based on electronic density n , with an approximate functional $E[V(\mathbf{r}_i)]$ that depends on the atomic configuration of the system \mathbf{r}_i , neglecting the degree of freedom of the electrons and thus being computationally less demanding³⁸. Given the reactive nature of lithium-ion batteries, traditional FF methods, which rely on harmonic laws to approximate bonded interactions, are inadequate for their study. To overcome this limitation, the ReaxFF method³⁹, which utilizes the bond-order (BO) and electron equilibration method (EEM)⁴⁰ to describe the connectivity and charges of atoms, respectively, has been suggested and utilized. This enabled the use of molecular dynamics (MD) simulations to observe the electrolyte decomposition and the formation of SEI components in LIBs. For instance, Kim et al.⁴¹ used ReaxFF optimized for C–H–O–Li species by Han et al.⁴² to study the formation of the SEI on lithium metal surfaces. They observed the production and layering of organic and inorganic components in the SEI on lithium metal surface, resulting from the decomposition of an Ethylene carbonate (EC) and Dimethyl carbonate (DMC) electrolyte mixture. Using the same force field, Guk et al.⁴³ demonstrated how the layer of SEI that forms on the surface of graphite prevents the percolation of the electrolyte and the anode exfoliation. After this initial work, Yun et al.⁴⁴ expanded the FF to include C–H–O–Li–Si–Li–F atoms, enabling the study of SEI formation in high-capacity batteries where graphite is replaced with silicon, which has a higher theoretical capacity (372 mA h g^{-1} for LiC_6 and 4212 mA h g^{-1} for $\text{Li}_{4.4}\text{Si}$)⁴⁵. This parameterization was then further enhanced by Wang et al.⁴⁶ Current available ReaxFFs have demonstrated success in reproducing dissociation energies and reaction kinetics. However, as discussed below in this work, they fall short in accurately describing the solid phase of the SEI components.

In this study, we propose a protocol for reparameterizing the Yun et al. ReaxFF for C–H–O–Li–Si–Li–F atoms. Our approach focuses on correcting parameters related to Li and F atoms to better capture the properties of lithium fluoride (LiF), which is one of the possible inorganic salts resulting from the formation of SEI. LiF stems mainly from fluoroethylene carbonate (FEC) decomposition: LiF-rich SEI is known to exhibit improved cycling stability in batteries⁴⁷. Furthermore, LiF has exceptional properties such as high ion transport, electronic insulation, and mechanical properties⁴⁸, making it a crucial element in the engineering of the solid electrolyte interface, as demonstrated by Tan et al.⁴⁹ While addressing a computational challenge, this work is also a critical step towards the broader and long-term goal of designing a multi-scale modelling framework for LIBs, which, together with additional experimental efforts, can help significantly to close some of the current technological gaps³. These methods are instrumental in comprehensively understanding the complex dynamics and energetics in LIB materials. This includes insights into electron localization in cathode materials, lithium diffusion dynamics, and the prediction of SEI formation. In this framework, computational innovation is key to integrating various scales and corresponding modelling methods. This integration aims to predict LIB performance from the material level up to the full battery pack design. For a deeper understanding of these challenges, we recommend further reading the works of Shi et al.⁵⁰, Wang et al.¹⁴, and Cappabianca et al.⁹, which provide valuable insights into the multi-scale computational approach in LIB research.

The advancements in powerful python libraries for the management of atomistic systems, as outlined in the Table 1, have enabled the automation and orchestration of atomistic simulations using Jupyter notebooks and libraries such as ASE^{51,52}, PyMatgen^{53,54}, and ParAMS^{55,56}. These tools have facilitated the parametrization process and made it more efficient. By sharing the Jupyter notebooks^{57,58} and the database used in this parametrization work, Authors hope to address a current gap in the field of ReaxFF parameterization, i.e., the lack of access to raw data and datasets, which can limit the reproducibility and validity of the results presented.

This article is structured as follows: firstly, we present the reparameterization process and results of the new ReaxFF. Afterwards, we demonstrate how the enhanced force field improves the representation of the crystal structure of the LiF inorganic salt. We then proceed to test the ability of the enhanced force field to accurately describe the lithium mobility in LiF and highlight how it provides a more realistic diffusivity value for the Li atoms. Then we present a deep investigation of the results from the new ReaxFF, discussing the challenging aspects of this kind of potential. In the subsequent section, conclusion are drawn and a possible future outlook provided. Finally, the methods used are outlined in the final section.

	Description	PreP	Run	PostP
ASE ^{51,52}	Atomic Simulation Environment (ASE) is a Python library that provides a versatile framework for configuring, running, visualizing, and analyse atomistic simulations. Thanks to the object <code>Calculator</code> ASE provides a powerful interface to different codes.	✓	✓	✓
Pymatgen ^{53,54}	Python Materials Genomics (Pymatgen) is a library for materials analysis, with a particular focus on solid-state studies and extensively used to produce and collect the data for the Materials Project (MP) ⁵⁹ database. It has functions for reading and manipulating structural, thermodynamic, and electronic properties of materials. It can also handle inputs and outputs from various DFT codes and easily access the MP crystallography database via its integrated API.	✓		✓
AiiDA ^{60,61}	Automated interactive infrastructure and Database (AiiDA) is an open-source, Python-based workflow management platform. Its main objective is to assist researchers in organizing, automating, managing, sharing, and tracking their simulations, thus enabling the reproducibility of complex workflows in computational materials science. Its plugging interface allows for the management of simulations performed with various codes, and it is designed to be used in conjunction with the Jupyter web interface, making the whole study interactive and easy to share.	✓	✓	✓
MDAnalysis ^{62,63}	It is a versatile Python library mainly focused on allowing the manipulation and analysis of MD trajectories. With support for various trajectory and system configuration formats, it simplifies the process of translating data into n-dimensional NumPy ⁶⁴ arrays for further analysis.			✓
Crystal Toolkit ⁶⁵	It is a library and web application framework for handling, manipulating, and analysing crystal structures. It is widely used in the MP database as an interactively visualizing web tool.			✓
Quippy ^{66,67}	It is the Python high-level interface to QUantum mechanics and Interatomic Potentials (QUIP) ⁶⁸ . QUIP is a set of software tools for performing MD simulations that can be used as a plugin for other programs such as LAMMPS, CP2K ⁶⁹ , and ASE.	✓	✓	
PyLammps ^{70,71}	Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) is a powerful classical molecular dynamics simulation code. The python package, which can be installed after code compilation, allows to manage the LAMMPS simulations through the <code>lammps</code> module, a wrapper for the code's C-library API.		✓	
PLAMS ^{72,73}	The Python Library for Automating Molecular Simulation (PLAMS) is the Python interface for the commercial code Amsterdam Modeling Suite (AMS) ⁷⁴ . Depending on the code used, it allows the automation of a wide range of simulations, including geometry optimization, vibrational spectroscopy, molecular dynamics, monte carlo (MC) simulations, and many other types of computational studies.		✓	✓
ParAMS ^{55,56}	Parameter optimization for Atomistic and Molecular Simulations (ParAMS) is a specialized Python library included in the AMS. Its main purpose is to facilitate the optimization workflow to search parameters for empirical energy functionals such as ReaxFF and density functional based tight binding (DFTB).	✓		

Table 1. List of the most commonly used and useful Python tools and libraries in computational materials science. These libraries can be used in one or more phases of material in silico study: PreP = pre-processing, Run = simulation running, and PostP = post-processing.

Results and discussion

ReaxFF reparameterization

In the currently available ReaxFF force fields, all the energy contributions, including the Li-F interactions, were parameterized using a database of *ab initio* simulations specifically designed to capture the dissociation energy for potential reactions in the system. As a result, both the force field of Yun et al.⁴⁴ and that of Wang et al.⁴⁶ can predict the reaction products in Si-Li batteries with acceptable accuracy. However, the insufficient data on the crystal properties of inorganic salts restricts their ability to describe LiF aggregation and solid-phase transitions accurately. This limitation underscores the need for a new parameterization. In parametrizing the ReaxFF, the methodologies employed for data collection are essential. Tracing back to the foundational work by Van Duin et al.³⁹, followed by the significant contributions of Chenoweth et al.⁷⁷ in the context of hydrocarbons FF, the prevailing approach has been to leverage the functional forms of the reactive force field. This is typically achieved by utilizing DFT data from exploring diverse reaction pathways, varying environmental conditions, and frames captured from *ab initio* MD simulations. Such strategies ensure an accurate representation of molecular dissociation. However, prior studies have not specifically emphasized the quality assurance and correlation analysis of the collected data. Given our objective to integrate solid phase information into the ReaxFF, we developed a new database following the methodology proposed by LaBrosse et al.⁷⁸. This methodology, initially tailored for developing force fields specifically for cobalt crystals, provides a robust framework for refining ReaxFF to precisely model solid-state phenomena. Furthermore, in this study, we have adopted a stochastic approach to select configurations. This strategy aims at minimizing autocorrelation and potential biases within the database, ensuring a more representative and accurate dataset for ReaxFF enhancement, and efficiently reduces the number of DFT calculations required for the database. A more exhaustive exploration, in contrast, might question the advantages of ReaxFF's functional form even if it ensures a wider potential energy surface exploration. As a result, it is not always obvious if this method is more appropriate as compared to fully data-driven force field approach. Namely, the new database⁷⁹ was created using stable and metastable crystalline unit cells from the Material Project database⁵⁹ (Fig. 1). Several initial configurations were created using the Pymatgen and ASE libraries. Over 300 DFT simulations were performed using those systems, and all relevant quantities, such as energies, forces, and partial charges, were extracted for the objective function. In accordance with FAIR (Findable, Accessible, Interoperable, and Reusable) principles⁸⁰, we stored all system and simulation data in an SQLite3 database using the ASE library. This database, along with metadata and clear usage instruction, has been shared on Zenodo (<https://doi.org/10.5281/zenodo.7959121>) and Github (https://github.com/paolodeangelis/Enhancing_ReaxFF_DFT_database) for easy accessibility and usability.

The ReaxFF coefficients were optimized using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES)⁸¹ genetic algorithm. Generally speaking, this evolution algorithm searches for the optimal solution by sampling the population from a multivariate normal distribution (MVN). The generated samples are ranked based on the chosen loss function, and the best points, representing lower values for the loss function, are utilized to update the covariance and mean of the MVN. Then the updated MVN is used to obtain the next generation of

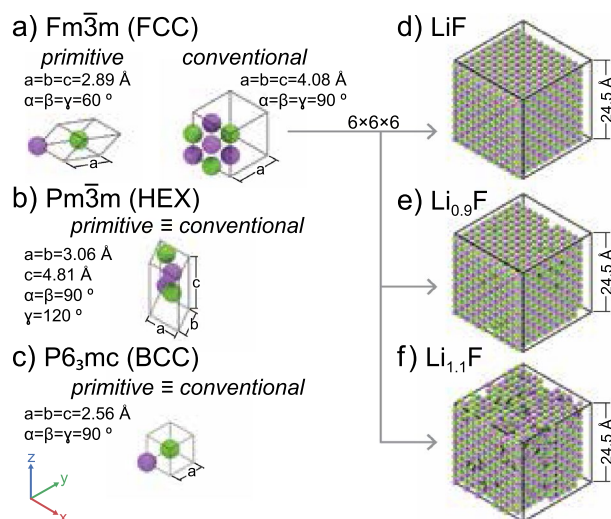


Figure 1. Isometric representation of unit cells used for ReaxFF reparameterization database generation and validation simulation boxes. The left column shows unit cells of LiF stable (a) and metastable phases (b,c) corresponding to face-centered cubic (FCC), hexagonal (HEX), and body-centered cubic (BCC) lattice types with space groups $Fm\bar{3}m$, $Pm\bar{3}m$, and $P6_3mc$ respectively⁵⁹. The right column displays atomistic systems used for testing the improved force field built as supercells of the stable unit cell (a): equal Li-F atoms (d), 10% vacancies at Li sites (e), and 10% interstitial Li atoms (f). The shown renderings were generated using the Visual Molecular Dynamics (VMD) code⁷⁵ and followed the Corey–Pauling–Koltun (CPK) colouring scheme⁷⁶.

samples, and the process continues iteratively until convergence or other predefined stopping criteria are met. In our case, the loss function is the sum of square errors (SSE) between values obtained from density functional theory calculations and molecular dynamics simulations. The algorithm converged after 1.3×10^4 iterations as shown in Fig. 2, thanks to the use of the step-wise optimization technique already employed by Wang et al.⁸². Indeed, similar to the study conducted by Wang et al.⁸², we capitalized on the physical-inspired structure of the ReaxFF force field, where parameters are categorized based on interaction and atom type, as outlined in Table 2. The approach consisted in gradually optimizing a selection of parameters relevant to particular types of interactions, instead than letting the algorithm operate in a very high-dimensional search space. Hence, we proceeded by optimizing initially the values associated with the bond interactions, then the values related to the dispersion energy, as evidenced by the leap in the loss function in Fig. 2a. However, after these first two subsets, further optimization of angle and torsional interactions was not feasible because all the initial guess values for their coefficients produced similar loss functions, which prevented the calculation of the second generation

Type	No. of parameters	Note
General	41	
Atoms	32	For each atom type N_A
Bonds	16	For each bond $N_{Bond} = \frac{(N_A + 1)!}{(N_A - 1)!2!}$
Off-diagonal	6	For each heterogeneous pair $N_{vdW} = \frac{N_A!}{(N_A - 2)!2!}$
Angles	7	For each possible angle in the system $N_{Ang} \leq \frac{(N_A + 2)!}{(N_A - 1)!3!}$
Dihedral	7	For each possible dihedral angle in the system $N_{Tors} \leq \frac{(N_A + 3)!}{(N_A - 1)!4!}$
Hydrogen bonds	4	Usually only for O, C and N atoms

Table 2. Listing of ReaxFF coefficients for each section in the force field file, organized by interaction type. Please note that the numbers in the second column represent the number of coefficients per single entry (e.g., for a single atom type or a single bond). However, the total number of parameters required in the force field depends on the total number of atom types (N_A), as indicated in the third column.

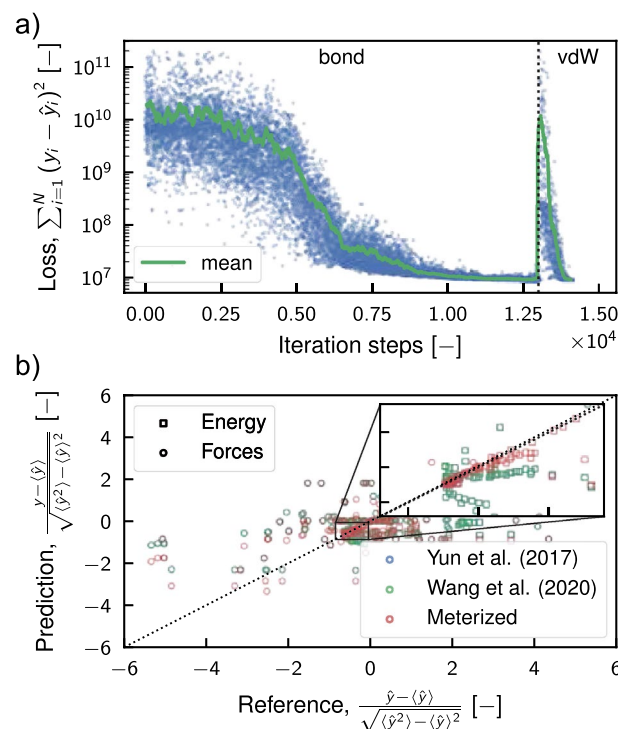


Figure 2. Evolution of the loss function during optimization of parameters and regression analysis. Panel (a) displays the loss function at each optimization iteration, represented by blue points. A green line indicates the moving average of the loss function, calculated using a 100-iteration window. The vertical dashed line marks the transition between optimizing different subsets of parameters, specifically the bond interactions (bond) and the van der Waals interactions (vdW). Panel (b) presents the regression results in a scatter plot format, contrasting reference values with predicted values for energy (square markers) and forces (circle markers). Predictions were computed using ReaxFF parameterizations by Yun et al.⁴⁴, Wang et al.⁴⁴, and our novel reparameterization (represented in blue, green, and red, respectively). Due to the distinct nature and wide range of these quantities, standardization was performed using the reference mean, $\langle \hat{y} \rangle$, and standard deviation, $\sqrt{\langle \hat{y}^2 \rangle - \langle \hat{y} \rangle^2}$, enabling legible visualization on the plot.

FF	Energy		Forces	
	R ²	RMSE (eV)	R ²	RMSE eV Å ⁻¹)
Yun et al.	−0.093	0.562	0.227	5.1 × 10 ^{−3}
Wang et al.	−0.093	0.562	0.227	5.1 × 10 ^{−3}
This work	0.293	0.452	0.377	4.6 × 10 ^{−3}

Table 3. Comparative performance metrics for energy and force FF predictions on the training set. To evaluate the efficacy of the FFs we compute the Coefficient of Determination (R²) and the Root Mean Square Error (RMSE) metrics.

of parameters and covariance matrix for the CMA-ES algorithm. More information about the full protocol is provided in the methods section. The new parameterization has markedly enhanced the prediction accuracy for energy and forces within the database configurations, as evidenced by the regression plot in Fig. 2b, and the metrics detailed in Table 3 compared to the previously proposed ReaxFFs. This improvement has positively influenced the model's capability to accurately describe the solid state of the inorganic salt, as demonstrated in Fig. 3. The figure includes an energy-strain curve (Fig. 3a) and Murnaghan equation of state (EOS) (Fig. 3b). The first plot was obtained by applying a shear strain ϵ_{xy} to a $2 \times 2 \times 2$ supercell of the stable LiF crystal. While the EOS was computed by expanding and contracting the system using a three-dimensional volumetric strain, since the two are connected by the following relation $\Delta V/V = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}$. At a glance, it is clear that the previous parameterizations were unable to predict the condensed state of LiF accurately. Indeed, neither the force field developed by Yun et al.⁴⁴, nor that of Wang et al.⁴⁶ showed the presence of a minimum configuration in both the elastic strain and EOS cases. The shear deformation case was particularly problematic because the inverted concavity of the curves led to a nonphysical negative value for the elastic tensor component c_{12} , where the stress

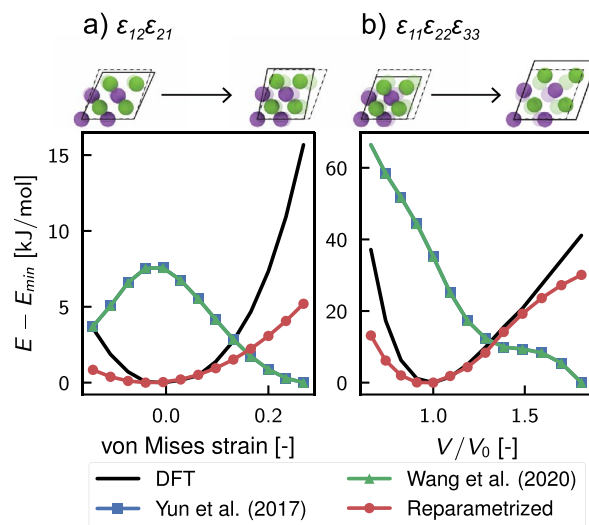


Figure 3. Exploring the stability of LiF under stress: A dual analysis reveals the material's response to mechanical manipulation. In (a), the energy-strain plot displays the total energy changes with applied shear strain $\varepsilon_{12} = \varepsilon_{21}$, while (b) depicts the equation of state, i.e., the energy variations versus volume deformation ratio V/V_0 (V_0 is the volume at the equilibrium). The black line represents DFT results used for training, while the blue, green, and red lines illustrate the energy predictions made by the ReaxFF from Yun et al.⁴⁴, Wang et al.⁴⁶, and the proposed new reparameterization, respectively. The simulation snapshot above each plot illustrates the frontal view of the crystal at two extreme deformation cases, with the dashed black line indicating the minimal energy unit cell. Notice that green and blue lines are mostly overlapped in the reported pictures.

tensor is defined as $c_{ij} = \sigma_{ij}\varepsilon_{ij}^{-1}$ with ε_{ij} representing the strain tensor and σ_{ij} the stress tensor. The concavity of the energy-strain plot determines this component since it is the function of the second partial derivative of the energy with respect to strain, i.e., $c_{12} = V_0^{-1} \cdot \partial^2 E / \partial \varepsilon_1 \partial \varepsilon_2$. On the other side, thanks to the newly designed database, the reoptimized ReaxFF force field accurately predicts equilibrium configurations for both EOS and elastic strain curves. Specifically, the prediction of the shear strain closely matches *ab initio* results in the region close to the equilibrium value, but deviates substantially for high shear strain values, indicating its limitations in predicting material stiffness accurately. Similarly, the ReaxFF prediction of the equation of state agrees well with the reference values, even for significant volume changes. This remarkable improvement of the proposed reparameterization over previous ones is also demonstrated for the one-dimensional deformation of the crystal and for the metastable crystals (as shown in supplementary Fig. S18). It is important to note, however, that the new force field predictions are less accurate for metastable cases with hexagonal and body-centered cubic lattice (Fig. 1), suggesting that the functionals used to describe various interactions may have limitations in generalizing the energies from the DFT simulations.

ReaxFF prediction of bulk LiF

To validate the new parameterization, as first test case, we conducted reactive MD simulations using all the discussed ReaxFFs for systems with Li and F atoms to simulate bulk LiF at ambient conditions (300 K, 1 bar). The simulation was performed on a relatively big system with respect to the ones used for the training. Indeed, a supercell of the LiF crystal created by replicating the conventional unit cell five times in all directions was placed in a simulation box of size $20.4 \times 20.4 \times 20.4 \text{ \AA}^3$, which is analogous to the system in Fig. 1d. After the initial relaxation of the system, including box size adjustment, the pure LiF crystal was simulated for 500 ps under an NVT ensemble (constant number of atoms, volume, and temperature) set at 300 K and with a time-step of 0.25 fs, and to attenuate possible artefacts due to the thermostat we have adjusted the relaxation time to be 1000 times the time-step (250 fs). In Fig. 4 we report the total energy, temperature, and radial distribution function (RDF) between the Li and F atoms, $g_{Li,F}(r)$, evaluated during the reactive MD simulations with different potentials. Comparing the results from the simulations using the Yun et al.⁴⁴ and Wang et al.⁴⁶ force fields (Fig. 4a and b), we observed similar behavior, indicating minimal differences in the parameters for Li and F used in both ReaxFF. Furthermore, we can observe that both force fields failed to accurately describe the simulated system, as the thermal agitation overpowered the binding energy, resulting in the amorphization of the inorganic salt when it reached the external heat bath temperature. This resulted in a sudden drop in the total energy at 2 ps that results into an artificial phase change. Indeed, since the melting temperature of LiF is 1121.35 K (848.2 °C)⁸³, the system should ideally exist in a solid phase in the simulated virtual conditions. This remarkable and unexpected result is even more evident looking at the two radial distribution functions in Fig. 4a and b, where the initial peaks due to the ordered structure of the system and immobility of the atoms disappear, and the curves become liquid-like RDF as the temperature rises. Moreover, we can notice a shift in the initial RDF peaks with respect to the theoretical position represented by the vertical dotted line, indicating an anomaly in the LiF solid state description of the two FFs. In contrast, our new parameterization performed well in simulating a simple bulk LiF system.

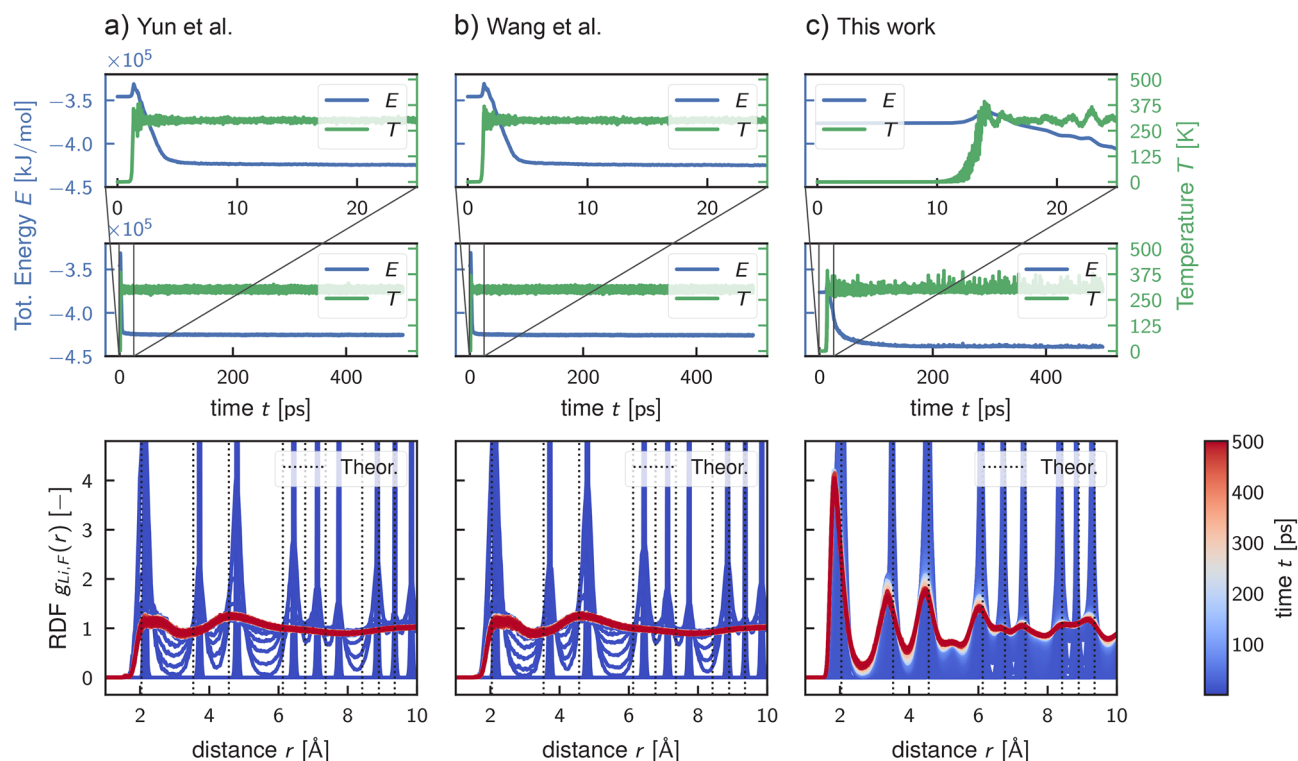


Figure 4. The evolution of bulk LiF simulations using reactive molecular dynamics at 300 K using three ReaxFF: Yun et al.⁴⁴ (a), Wang et al.⁴⁶ (b), and the new reparameterization (c). The top graphs depict total energy trends (solid blue line) alongside temperature trends (solid green line), with inset plots providing detailed views of the early simulation trajectories. Conversely, the bottom panels show the time evolution of the Li-F radial distribution function (RDF). Each RDF curve is obtained by averaging the trajectory every 0.1 ps and then coloured according to the simulation time indicated by the colour bar on the right. Notably, RDF curves from the literature force field resemble more systems in a liquid phase rather than a solid phase.

Indeed, Fig. 4c shows that the radial distribution function has initial peaks that align with theoretical values and remain consistent even at high temperatures. In this case, the thermal agitation of the atoms resulted only in the smoothening of the peaks and the emergence of secondary peaks due to the oscillation of some atoms between the lattice and interstitial positions. This important improvement is also evident by visualizing the temperature growth, which gradually reaches 300 K and is held constant as the system converges to an equilibrium state, as visible from the total energy (Fig. 4c).

Diffusion of Li in LiF prediction from ReaxFF

In designing the training set, we aimed to enhance the accuracy of predicting both the mechanical and thermophysical properties of the inorganic material through ReaxFF simulation. To achieve this, we included configurations with point defects like vacancies, interstitials, and substitutions and incorporated data from *ab initio* simulations conducted at different temperatures. Therefore, the second test case for the new ReaxFF was to compute the lithium transport properties with a campaign of MD simulations. We evaluated the mass diffusivity of Li atoms in three systems illustrated on the right side of Fig. 1, specifically: one with 10 % interstitial lithium atoms ($\text{Li}_{1.1}\text{F}$), one with 10 % lithium vacancies ($\text{Li}_{0.9}\text{F}$), and one with no defects (LiF). These systems were simulated for 500 ps at three different NVT temperatures (300 K, 400 K, and 500 K). The lithium diffusion coefficients were extracted by measuring the mean square displacement (MSD) during the simulation and by using the Einstein-Smoluchowski diffusion equation^{84–86}:

$$\lim_{t \rightarrow \infty} \langle \|\mathbf{r}_i(t) - \mathbf{r}_i(0)\|^2 \rangle_{i \in \text{Li}} = c_d D t, \quad (1)$$

where the limit argument is the MSD definition, t is time, D is the diffusion coefficient, and c_d is a constant indicating dimensionality, with values of 2 for one-dimensional, 4 for two-dimensional, and 6 for three-dimensional diffusion. We obtained the diffusion coefficient values by determining the slope of the straight line that best fits the MSD evolution for each simulation (as seen in the supplementary Figs. S19–S21). In Fig. 5, we plotted them on a logarithmic scale and used the inverse of the temperature as abscissa. This enabled us to extract the activation energy that appears in the Arrhenius equation:

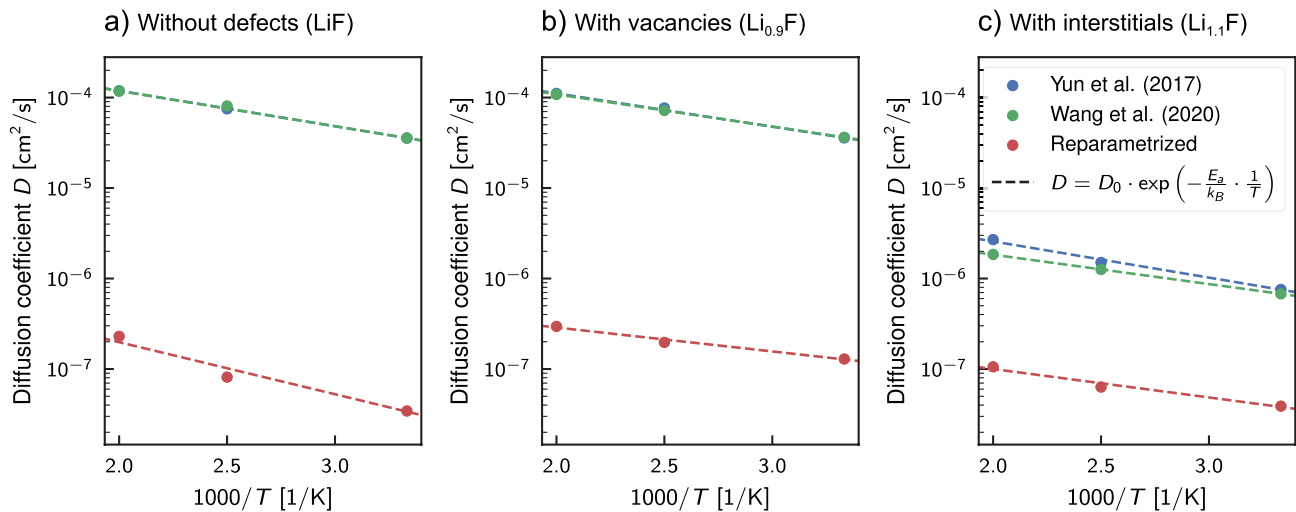


Figure 5. Arrhenius plot displaying the diffusion coefficient of Li (D) in pure bulk LiF (LiF) and with point defects of vacancy ($\text{Li}_{0.9}\text{F}$) and interstitial ($\text{Li}_{1.1}\text{F}$) types, obtained through the molecular dynamics simulations. The blue, green, and red dots represent the D values obtained using the Yun et al.⁴⁴, Wang et al.⁴⁶, and the proposed new ReaxFF parameterization, respectively. The drawn dashed line results from a linear regression of the coefficients using the least square method.

	ReaxFF	$D_0[\text{cm}^2/\text{s}]$	$E_a[\text{kJ mol}^{-1}]$
LiF	Yun et al.	$(7.23 \pm 0.05) \times 10^{-4}$	7.50 ± 0.02
	Wang et al.	$(7.2 \pm 0.5) \times 10^{-4}$	7.0 ± 0.2
	This work	$(3 \pm 2) \times 10^{-6}$	11.0 ± 1.6
$\text{Li}_{0.9}\text{F}$	Yun et al.	$(6.1 \pm 0.3) \times 10^{-4}$	7.1 ± 1.4
	Wang et al.	$(5.5 \pm 0.2) \times 10^{-4}$	6.810 ± 0.013
	This work	$(9.0 \pm 1.7) \times 10^{-7}$	5.1 ± 0.5
$\text{Li}_{1.1}\text{F}$	Yun et al.	$(1.5 \pm 0.3) \times 10^{-5}$	7.6 ± 0.5
	Wang et al.	$(8.2 \pm 1.3) \times 10^{-6}$	6.22 ± 0.04
	This work	$(4.0 \pm 1.2) \times 10^{-7}$	6.0 ± 0.8

Table 4. The values of the activation energy, E_a , and the maximum value of the lithium self-diffusion coefficient, D_0 , for the three simulated systems and for each ReaxFF used. The errors were estimated based on the inference for the linear regression model coefficients (least squares method), and assuming a 95 % confidence interval.

$$D = D_0 \cdot \exp\left(-\frac{E_a}{k_B T}\right). \tag{2}$$

Here, D is the lithium diffusion coefficient from the MD simulations, D_0 is the maximum diffusivity value (the limit value at infinite temperatures), k_B is the Boltzmann constant, and E_a is the activation energy. We repeated this process for all three ReaxFFs discussed, and the resulting values of D_0 and E_a are reported in Table 4. The new force field provides an improved description of bonding forces, which results in a significant reduction in the mobility of lithium atoms within the crystal lattice, as expected. For example, in the case study of defect-free LiF supercell at 300 K, the diffusion coefficient predicted by both the Yun et al.⁴⁴ and Wang et al.⁴⁶ force fields is $3.56 \times 10^{-5} \text{ cm}^2/\text{s}$, which is significantly closer to the diffusivity of lithium ions in the electrolyte ($0.5 \times 10^{-5} - 1.4 \times 10^{-5} \text{ cm}^2/\text{s}$ ⁸⁷) rather than in the solid. In contrast, the new parameterization predicts a diffusivity of $3.44 \times 10^{-8} \text{ cm}^2/\text{s}$, which is a reduction of 3 orders of magnitude. Similar reductions in diffusivity are seen in cases involving interstitials and vacancies.

We compared the obtained values with *ab initio* values calculated by Zheng et al.³¹. They studied lithium diffusion in various organic components of the SEI by exploring the energy surface using the surface energy Climbing Image Nudged Elastic Band (CI-NEB) method⁸⁸. The exploration of the energy surface by Zheng et al. confirmed that the diffusion of lithium in lithium fluoride occurs via three distinct possible mechanisms. These mechanisms include: vacancy movement, where lithium atoms jump to the adjacent empty lattice site (vacancy); direct-hopping, where lithium moves directly from one interstitial site to another; and the knock-off mechanism, in which an interstitial lithium atom replaces a lithium atom in the crystal lattice, thus causing the

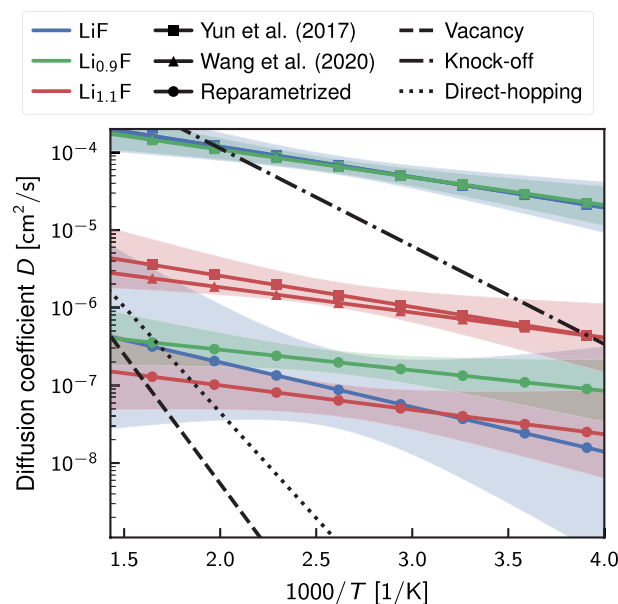


Figure 6. Arrhenius plot of diffusion comparing the predictions from Reactive MD and DFT simulations. The interpolation from MD results is shown using continuous lines and markers (square for Yun et al.⁴⁴, triangular for Wang et al.⁴⁶, and circular for the new reparameterization), and the confidence interval of 95% is shown with the shaded areas. The effect of defects on diffusion is indicated by different colors: blue color for the defects-free case, green for the cases with vacancies, and red for the case with interstitial lithium. The black dashed, dash-dot, and dotted lines represent the Arrhenius curves estimated by Zheng et al.³¹ for three potential transport mechanisms (vacancy, knock-off, and direct-hopping) from Climbing Image Nudged Elastic Band (CI-NEB) studies.

displaced lithium atom to move into another interstitial site. Figure 6 shows the Arrhenius curves for the three diffusion mechanisms calculated using the values of activation energy and maximum diffusion obtained from the *ab initio* study, as well as the curves obtained from the reactive molecular dynamics simulations. Regarding the maximum diffusion coefficient $D_0 = 3\alpha^2\nu$, Zheng et al. approximate its value using the estimated phonon frequency in the crystal $\nu = 1 \times 10^{13} \text{ s}^{-1}$ and the migration distance for the lithium ions α observed during the CI-NEB simulations³¹. Molecular dynamics simulations can exhibit various diffusion mechanisms, leading to the calculated curves for the new force field and each studied system falling within the region defined by the three diffusion mechanisms. For the calculated temperature values, the diffusion coefficients turn out to be intermediate values between pure knock-off diffusion and pure direct-hopping diffusion. However, looking at the slope of the Arrhenius curves predicted by molecular dynamics simulations in Fig. 6 it is clear that even the new ReaxFF underestimated the activation energy. Indeed, even for the most favorable diffusion mechanism (knock-off), the activation energy obtained from the *ab initio* energy profile is 24.1 kJ mol^{-1} , which is more than double the value calculated with the new force field for the defect-free case (11.0 kJ mol^{-1}). The discrepancies are even more significant for all other cases listed in Table 4.

To address this discrepancy, we conducted further analysis of the ReaxFF by reproducing the energy curves for lithium diffusion by vacancies and by direct-hopping using the CI-NEB method. Initial and final configurations were made using a $2 \times 2 \times 2$ LiF supercell and inserting the lithium atoms into two adjacent sites identified with the Voronoi analysis from the pymatgen python library⁵³. While for the vacancy diffusion case, one lithium was removed from the crystal lattice from two adjacent unit cells. The energy profile was then obtained using the CI-NEB algorithm with the same parameters used in the DFT simulation for training the new ReaxFF. To ensure convergence, we used 23 images for direct-hopping diffusion and 17 images for vacancy diffusion. Our simulations yielded activation energies of 64.3 kJ mol^{-1} for vacancy diffusion and 87.5 kJ mol^{-1} for direct-hopping. The value for vacancy diffusion was comparable to previous work (63.7 kJ mol^{-1}). Still, the value for direct-hopping is higher than that calculated by Zheng et al. (52.1 kJ mol^{-1}) due to the use of a smaller system to reduce the computational cost, resulting in a higher defect density in our case. Subsequently, from the result obtained from the DFT simulations, we proceeded to calculate each image's energy using the various ReaxFF discussed. In Fig. 7 we show the comparison of the energy profile predicted by the various methods.

The additional investigation of ReaxFF presented in Fig. 7 reveals significant discrepancies between previous parameterizations, DFT prediction, and the reoptimized force field. Specifically, the energy barrier value prediction was found to be incorrect for all the ReaxFF studied, and the new force field exhibited unrealistic behavior in the case of interstitial diffusion. Indeed, the maximum energy values corresponded to the initial and final configurations, while the minimum value corresponded to the transition configuration. This behavior contradicts the physical nature of the phenomena, and the predictions made by the DFT calculation and even

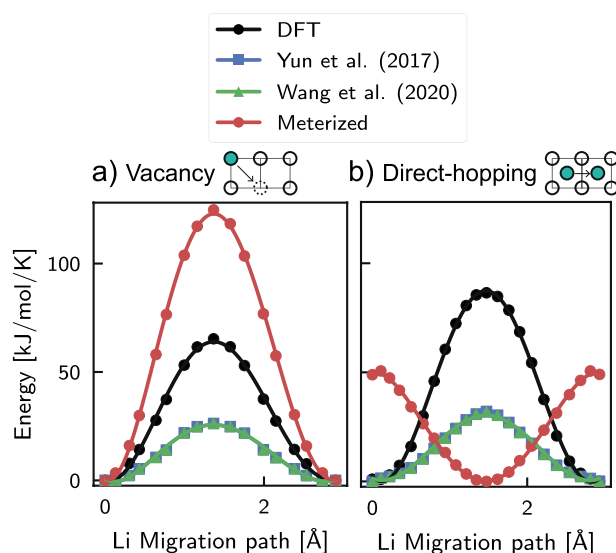


Figure 7. Graphical visualization of the energy profile for lithium migration by vacancy (a) and direct-hopping (b) calculated by CI-NEB is depicted by the black line and circular markers. The markers indicate the exact energy value of each distinct image obtained from the convergence of the CI-NEB algorithm and were used to calculate the energy using three different methods: ReaxFF by Yun et al.⁴⁴, Wang et al.⁴⁶, and a new reparameterization, represented by blue, green, and red colored markers and line, respectively.

by previous parameterizations by Yun et al.⁴⁴ and Wang et al.⁴⁶. In order to be able to explain this numerical artifact, it is important to consider the general structure of ReaxFF, which is described by the provided equation:

$$\begin{aligned}
 V_{\text{reaxFF}} = & \underbrace{V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}}}_{\text{bonded interaction}} \\
 & + \underbrace{V_{\text{over}} + V_{\text{under}}}_{\text{bonded interaction (coordination)}} + \underbrace{V_{\text{vdW}} + V_q}_{\text{non-bonded interaction}} \\
 & + \underbrace{V_{\text{conj}} + V_{\text{trip}} + V_{C_2} + V_{H\text{-bond}} + V_{lp}}_{\text{system specific}}.
 \end{aligned} \quad (3)$$

V_{reaxFF} represents the total reactive potential of the particle i interacting with neighboring atoms, which can be divided into three contributions: bonded, non-bonded, and “system specific”. The first describes the bonded interaction between atoms and is formed by the bond energy V_{bond} , the angle energy V_{angle} and the dihedral (or torsional) energy V_{dihedral} and the corrective terms over-coordination V_{over} , and under-coordination V_{under} energies. The non-bonded energy, which is analogous to classical molecular dynamics, is given by the sum of the van der Waals potential V_{vdW} and Coulomb potential V_q . The last terms, on the other hand, are a set of corrective energies that accounts for specific phenomena in certain specific systems. For additional details about the mathematical and physical implications of each term, the reader can refer to the works by Van Duin et al.³⁹ and Chenoweth et al.⁷⁷.

From Eq. (3) and the previous studies, we can attribute the unrealistic behavior observed with the new force field to the increased energy contribution from bond energy, which improves the description of system connectivity. However, this results in an increase in the effect of correction energy terms, specifically over-coordination energy V_{over} , and under-coordination energy, V_{under} , in Eq. (3), which are functions of the difference between the total number of bonds of an atom and its valence number³⁹. Indeed, in the initial and final conditions, the lithium atom has the same number of neighboring atoms, which is higher than in the transition state, where space is created by moving the lattice atoms, and the increase of the lithium distance respects the atoms at the extremes of the unit cell. This unexpected artifact highlights the critical importance of carefully considering energy contributions in force fields to ensure accurate predictions, as observed in the case study using ReaxFF. Consequently, it is likely that additional configurations are needed to be included in the training set. However, it is important to note that this alone does not guarantee the resolution of all the aforementioned issues. In fact, due to the reliance on a fixed functional shape in the force field, a more deep and sophisticated alteration in the mathematical expression of the energy terms might be required. Implementing such changes would involve modifying the code, which exceeds the scope of this work. This also highlights that a critical challenge here is the limited transferability of the parameterization beyond the specific database, emphasizing the necessity of a comprehensive database of *ab initio* simulations that is specifically tailored and adequately representative for studying the targeted system at an atomistic scale.

Conclusion

In conclusion, the proposed partial reparameterization methodology for the ReaxFF has significantly improved the description of lithium fluoride in the solid state, leading to better predictions of its solid phase properties and lithium mobility in LiF crystal using reactive MD simulations. Indeed, implementing the new parameters of the ReaxFF has demonstrated its ability to predict the stable unit cell under mechanical deformation accurately, exhibit typical solid-state RDF, and notably reduce lithium diffusivity in LiF by at least two orders of magnitude. The automation and interactivity of the protocol, achieved by leveraging Python libraries for atomistic simulations, made it possible to construct and simulate various configurations of the LiF crystal needed to build a database for correcting bond and van der Waals interactions of the ReaxFF. The new force field obtained from the re-optimization not only improved the behavior of the crystal in the solid phase but also partially corrected the description of lithium transport phenomena in LiF.

However, the in-depth investigation carried out revealed that the diffusion activation energies predicted by the new force field are still underestimated. This limitation may be due to the method used to construct the database, which did not directly sample lithium transport, but focused more on local or global deformation of the crystal lattice. This study highlights the strong dependence of the ReaxFF on the configurations included in the database. Hence, future effort should focus in addressing the challenges associated to the effective interpolation of the energy surface in unexplored or underrepresented regions in the adopted training set. The need to tailor ReaxFF for each specific scenario presents a significant challenge, especially when considering the upscaling of electronic simulations, such as DFT, to the molecular level. A common mitigation strategy involves utilizing previous parametrizations to define the boundaries for parameter optimization. However, for less commonly explored elements, like Li and F in our case, other parametrizations might be nearly identical and not particularly useful. Therefore, in the absence of disparate parametrizations in the literature to effectively constrain the search space, a convenient approach might be to rely on an extensive database. Nevertheless, based on our experience, as far as the explored case study is concerned, the need for ever-larger databases poses serious questions on the cost-benefit ratio of this method. Although the ReaxFF MD simulations are much faster than *ab initio* simulations, given all the functionals to be computed and the need to update the charges at each step reduces by at least a factor of 100 the speed of reactive MD simulations as compared to simulations with Lennard-Jones fluid⁸⁹.

In addition to the sensitivity of the ReaxFF to the training set used, the difficulty in accurately describing the system studied in this work may also be attributed to an intrinsic bias of the force field. In particular, it is worth noting that the original formulation of the ReaxFF potentials proposed by Van Duin in 2001 was validated primarily on organic systems³⁹, and various new functionals have been added to improve the accuracy and applicability of ReaxFF in other systems over the past few decades^{90–92}. These include the “lone-pair” energy term (V_{lp} in Eq. (3)) for hydrocarbon combustion^{77,90,93}, a three-body functional (V_{trip} in Eq. (3)) term for $-\text{NO}_2$ group chemistry^{90,94,95}, and an energy term (V_{H-bond} in Eq. (3)) to account for hydrogen bonds in aqueous systems^{77,96}, among others. Therefore, it may be necessary to introduce new correction energy terms to improve the accuracy of ReaxFF in strongly inorganic systems such as LiF.

On the other hand, the rapid development of machine learning-based (MLFF) and neural network-based force fields^{97–101} (NNFF) may provide alternative and accurate approaches to the ReaxFF. The database constructed using the proposed protocol (or even further automatized by algorithmic orchestration¹⁰²) could be used for training these new force fields, which may provide superior performance, possibly at the cost of compromising on the physical insight into the parameters obtained from the training. With their flexibility and high interpolation capabilities, these newly developed force fields^{101,103,104} offer promising solutions to the challenges identified in this work and with the ReaxFF. These force fields describe the chemical local environment using a feature transformations (“descriptors”) algorithm^{99,105,106} instead of the bond order and functional parameters utilized in ReaxFF. However, they confront a limitation common in machine learning applications for materials science, i.e. the high dimensionality of features relative to the small sample size. This issue has been thoroughly reviewed by Liu et al.¹⁰⁷, who also suggest approaches like feature reduction and sample augmentation to enhance model performance and accuracy. In summary, the proposed methodology, coupled with the synergistic approach to data quantity governance¹⁰⁷, could be extended to the parameterization of various other potentials, and by increasing the number of initial configurations, it may also be possible to proceed with the parameterization of neural network potentials. Moreover, the guided reparameterization method could incorporate other frequently encountered inorganic compounds in the SEI within future ReaxFF or NNFF. This expansion offers an exciting opportunity to explore increasingly intricate and realistic systems resembling mosaic structures¹⁰⁸ including also other SEI compounds (e.g. Li_2CO_3 , Li_2O) thus emulating the Peled model⁷.

Methods

Interactive reparameterization protocol

The protocol for calculating the new ReaxFF parameters is presented in a flowchart shown in Fig. 8. The procedure is carried out using four Jupyter Notebooks (JNBs)^{57,58} that facilitate the automatic construction, visualization, and simulation of the necessary configurations for database construction and optimization of the reactive potential. Python libraries ASE^{51,52}, pymatgen^{53,54}, PLAMS^{72,73} and ParAMS^{55,56}, designed for atomistic systems manipulation and simulation are utilized throughout the entire process, allowing for streamlined and efficient handling of the various steps.

The first JNB initiates the protocol by defining the chemical formula of the component to study. Then, using the API from the Materials Project database, the available crystalline units are downloaded. In the case of LiF, there are three configurations available, corresponding to different lattices (Fig. 1a–c), one stable ($\text{Fm}\bar{3}\text{m}$) and two metastable ($\text{P6}_3\text{mc}$, $\text{Pm}\bar{3}\text{m}$). These crystals are imported into the interactive work environment as virtual objects from the ASE (Atoms) and pymatgen (Structure) libraries. Within this virtual environment, the

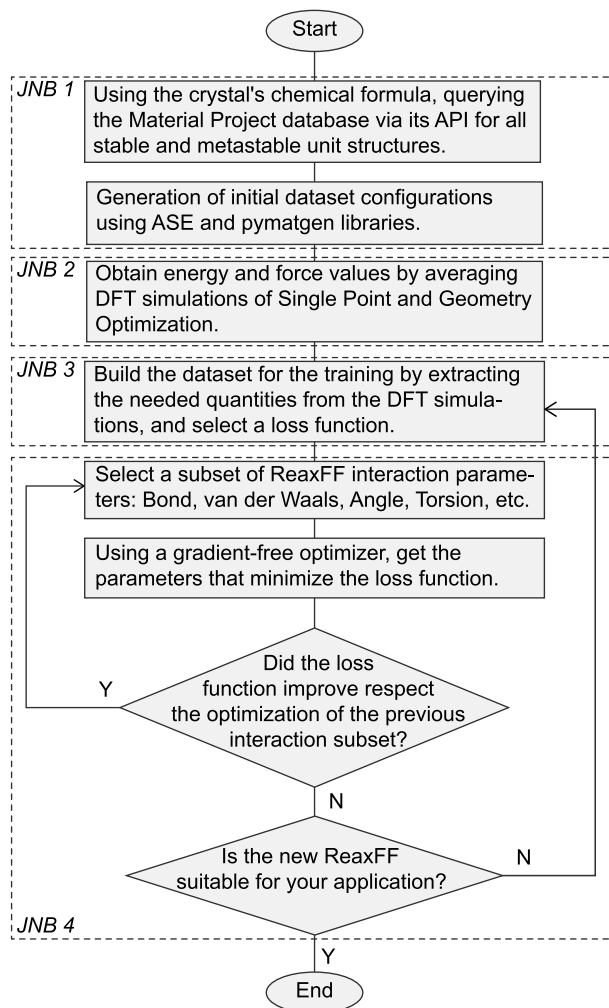


Figure 8. Flowchart followed for the reparameterization of ReaxFF. The dotted boxes indicate the Jupyter Notebook (JNB) specific to that part of the workflow, all available at <https://doi.org/10.5281/zenodo.8036775> and in the repository https://github.com/paolodeangelis/Enhancing_ReaxFF.

crystals were manipulated to generate initial configurations for the DFT simulations. Specifically, we build different defected or deformed crystals following the procedure by LaBrosse et al.⁷⁸, namely:

- Supercells:** To capture the effect of the atoms' coordination in the training set, several supercells of size $2 \times 1 \times 1$, $2 \times 2 \times 1$, $2 \times 2 \times 2$, $3 \times 2 \times 2$, $3 \times 3 \times 2$, and $3 \times 3 \times 3$ were constructed.
- Vacancies:** We randomly removed a lithium or fluorine atom from the $3 \times 2 \times 2$ supercell, up to a maximum of five vacancies.
- Strain:** We applied different types of strain to each LiF crystal, including a normal strain ε_{11} , a shear strain $\varepsilon_{12} = \varepsilon_{21}$, and a homogeneous deformation in all directions $\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{33}$ (needed for computing the equation of state), resulting in 13 configurations with strain values ranging from -12.5% to 23.5% .
- Substitution:** We randomly substituted a lithium or fluorine atom with the opposite species, with a maximum of five possible defects for each crystal.
- Interstitial:** We inserted an interstitial atom at the center of Voronoi volumes obtained from the $3 \times 2 \times 2$ supercell for each LiF crystal. We repeated this procedure five times to create structures with 1, 2, ..., and 5 interstitial atoms.
- Slab:** To include surface energy in the ReaxFF, we generated crystal slabs for the surfaces (100), (110), (111), and (210). We repeated each slab 2, 3, or 4 times to account for different thicknesses and we introduced sufficient empty space along the normal direction to avoid numerical artifacts due to the periodic boundary conditions (PBCs).

Bulk at 300 K and 500 K: We included frames of *ab initio* MD simulations at 300 K and 500 K for the $3 \times 2 \times 2$ supercell to predict possible energy fluctuations during ReaxFF MD.

Amorphous LiF: To account for the high-energy state of the crystal, we included possible amorphous LiF with an *ab initio* MD simulation at $T = 2500$ K.

In the second JNB, previously constructed configurations are used to perform *ab initio* simulations to obtain the numerical values needed for the databases. The ASE and pymatgen objects of the various systems were used as input to the PLAMS library of the AMS⁷⁴ commercial code, enabling the running, control, and processing of DFT simulations. For LiF, over 300 DFT simulations were carried out using the BAND plane-wave DFT code available in AMS, as listed in the supplementary Table S1. Single-point (SP) simulations were used to calculate the energy and force values of each atom in each configuration, by solving the Kohn-Sham equations while considering the atom cores as fixed. For systems with defects, geometry optimization (GO) simulations were initially performed to determine the equilibrium configuration. *Ab initio* MD simulations were performed using Grimme's extended density functional based on tight-binding (DFTB)¹⁰⁹ due to their low numerical cost. From the resulting trajectories, 10 frames were selected, and their force and energy values were refined using SP-DFT simulations.

The third and fourth JNBs heavily rely on the AMS ParAMS^{55,56} library, which is specifically designed for potential optimization. In the third JNB, all quantities, such as energy, forces, and charges needed for the database, are extracted, resulting in more than 3000 entries. Regarding the potential energy, because it is a state function, the database does not include the absolute values obtained from the DFT simulations but the formation energy approximated by the relative energy obtained with respect to the defect-free configuration of the LiF crystal, namely:

$$\Delta H_i \approx \Delta E_i = E(X_i^{\text{LiF}}) - [E(X_{\text{bulk}}^{\text{LiF}}) + n_i \cdot \mu_{\text{Li}} + m_i \cdot \mu_{\text{F}}]. \quad (4)$$

Where ΔH_i is the formation energy of i -th configuration, $E(X_i^{\text{LiF}})$ represents the energy of the i -th configuration, $E(X_{\text{bulk}}^{\text{LiF}})$ is the energy of the bulk LiF configuration, n_i and m_i are integer numbers of Li and F atoms in the i -th compared to the bulk and their sign depend on whether we removed or added atoms, while μ_{Li} and μ_{F} chemical potential, approximated by the energy per atoms obtained simulating pure Li and F unit cells.

This database is then used to optimize the ReaxFF in the fourth and final JNB. The optimization process begins by selecting the group of parameters that will form the search space, starting with those that significantly influence the behavior of the ReaxFF, such as the bond energy, followed by the van der Waals energy and angular energy terms. Various gradient-free optimization algorithms are available in the ParAMS library, and for the ReaxFF optimization for LiF, we chose the genetic algorithm Covariance Matrix Adaptation Evolution Strategy algorithm (CMA-ES)⁸¹ to minimize the objective function represented by the sum of squared errors (SSE) between the DFT values and the ReaxFF values.

Each group or subset of parameters is optimized sequentially, and the ReaxFF optimization process is considered complete when further improvements in the loss function become negligible.

Density functional theory calculations

The dataset values (energies, forces, charges, etc.) for the ReaxFF optimization were obtained from DFT simulations performed with the commercial BAND^{110,111} code available in the AMS suite. We numerically solved the Kohn-Sham equations using the Perdew-Burke-Ernzerhof (PBE)¹¹² functional and the polarized double zeta (DZ) numerical atomic orbitals (NAOs) basis set for the calculation of the s , p , and d orbitals. The software automatically chose the values of k -point and frozen electrons depending on the desired accuracy, and we selected a high accuracy value that guaranteed an error of less than 0.01 eV per atom, and by comparing the formation energy values obtained for each crystalline unit of LiF studied with the values reported on the Material Project online database (see Fig. 1a–c). For very inhomogeneous systems, such as those with a large number of interstitial atoms or surfaces, we calculated forces without frozen atoms and using a single zeta (SZ) type basis set to speed up the calculation of the equilibrium configuration. We then refined the resulting configuration using the settings described above.

For detailed instructions on installing and utilizing the protocol and database repository, please refer to the supplementary material provided.

Molecular Dynamics calculations

All the reactive molecular dynamics simulations were performed using the open-access code LAMMPS⁷⁰ with the ReaxFF package¹¹³. The initial configuration for the diffusion of Li in bulk LiF was obtained by starting from the primitive unit cell of the stable crystal obtained from the Material Project database. The unit cell was then converted into the conventional unit cell using the pymatgen routine `ConventionalCellTransformation`⁵³ and replicated six times along all directions to obtain a $6 \times 6 \times 6$ supercell with final dimensions of $24.5 \text{ \AA} \times 24.5 \text{ \AA} \times 24.5 \text{ \AA}$. To create the system with vacancies (i.e. $\text{Li}_{0.9}\text{F}$), 86 randomly selected lithium atoms were removed. While to create the system with interstitial atoms (i.e. $\text{Li}_{1.1}\text{F}$), 86 lithium atoms were placed inside the LiF supercell using the PACKMOL code¹¹⁴. After the initial energy minimization, the system was simulated for 0.5 ns using an NVT ensemble at three different temperatures (300 K, 400 K, and 500 K) with a Nose-Hoover thermostat¹¹⁵ and a relaxation time of $\tau_T = 25$ fs. The integration time step was set to $\delta t = 0.25$ fs. Thermodynamic properties, including the instantaneous mean square displacement for all lithium atoms, were sampled every 50 simulation steps to compute the diffusivity as discussed later.

Diffusion energy barrier calculations (DFT)

To determine the energy barrier of Li ion diffusion, we used the climbing image-nudged elastic band (CI-NEB) method⁸⁸. To employ this method, we built a $2 \times 2 \times 2$ supercell of the primitive cell of the stable LiF crystal and placed a Li atom in the interstitial site found with the Voronoi analysis¹¹⁶ in two adjacent cells to study the diffusion by interstitials. While, for the vacancy case, we removed a Li atom from two adjacent cells of the supercell. After geometry optimization of the initial and final states, we performed the CI-NEB calculation using 23 images for the direct-hopping diffusion case and 17 images for the vacancy case. We set the maximum perpendicular force component for the transitional state to be $2.5 \text{ eV } \text{\AA}^{-1}$ as the climbing convergence criterion. Finally, we obtained the activation energy by averaging the energy differences between the initial and transitional state and the final and transitional state.

Diffusion energy barrier calculations (MD)

To compute the diffusion coefficient D for each combination of temperature, system, and potential, we employed the Einstein-Smoluchowski diffusion equation^{84–86}, Eq. (1), that requires the time derivative of the MSD obtained from ReaxFF MD simulations. We obtained the numerical MSD from the MD trajectories and fitted it to a linear model of the form $\langle \mathbf{r}^2 \rangle = \alpha_0 + \alpha_1 \cdot t + \varepsilon$, where $\langle \mathbf{r}^2 \rangle$ represents the mean-square displacement, ε is the statistical error, and t denotes the elapsed simulation time (Figs. S19–S21). To determine the coefficients α_0 and α_1 , we used the least squares method (LSM)¹¹⁷. Consequently, the diffusivity can be evaluated as $D = \alpha_1/6$.

An analogous procedure was followed to compute the activation energy from the Arrhenius law, Eq. (2). To linearize the equation, we took the logarithm of both sides and applied the variable substitution $x = T^{-1}$. This manipulation resulted in the equation taking the form $\ln(D) = \ln(D_0) - E_a/k_B \cdot x$ (Fig. 5). Using the diffusion coefficients obtained from the ReaxFF MD simulations, we fitted a linear model of the form $\ln(D) = \beta_0 + \beta_1 \cdot x + \varepsilon$ to obtain the activation energy as function slope of the line, i.e., $E_a = -\beta_1 \cdot k_B$.

To obtain the reported 95% confidence intervals for the diffusivity and activation energy in Table 4, we assumed that the statistical error ε of the linear model follows a Student's t -distribution T_ν , where ν represents the degrees of freedom of the distribution¹¹⁷. In our case, ν equals the number of sampled data n minus the constraints of our model, which are the intercept and slope of the model ($\nu = n - 2$). Under these reasonable hypotheses, the uncertainty for the diffusivity, δD , and activation energy, δE_a , is estimated as follows¹¹⁷:

$$\begin{aligned} \delta D &= t_{n-2, 0.025} \cdot \frac{\sqrt{\frac{1}{n-2} \sum_{i=1}^n (\langle \mathbf{r}^2 \rangle_i - \alpha_0 - \alpha_1 t_i)^2}}{6 \cdot \sqrt{\sum_{j=1}^n (t_j - \langle t \rangle)^2}}, \\ \delta E_a &= t_{n-2, 0.025} \cdot \frac{\sqrt{\frac{1}{n-2} \sum_{i=1}^n \left(\ln(D_i) - \beta_0 - \beta_1 \frac{1}{T_i} \right)^2}}{k_B \cdot \sqrt{\sum_{j=1}^n \left(\frac{1}{T_j} - \langle \frac{1}{T} \rangle \right)^2}}. \end{aligned} \quad (5)$$

$t_{n-2, 0.025}$ is the cuts probability 0.025 from the upper tail of Student's t distribution T_{n-2} , with $n - 2$ degrees of freedom. α_0 , α_1 , β_0 , and β_1 are the coefficients of the two linear models determined from the linear regression. $\langle \mathbf{r}^2 \rangle_i$ represents the MSD at the i -th time t_i of the simulation, k_B is the Boltzmann constant, and D_i and T_i are the diffusivity and corresponding simulation temperature values, respectively. We use the notation $\langle \cdot \rangle$ to indicate the mean value of the independent sampled variable of each model, i.e., the time, t , for the first model, and the reciprocal of the temperature, T^{-1} , for the second. The confidence interval for the entire linear model depicted in Fig. 6 was instead obtained with the Eq. (6), which provides the uncertainty for the logarithm of the diffusivity, $\ln(D)$, as a function of the reciprocal of the temperature, T^{-1} , which is the independent variable of the linear model¹¹⁷:

$$\begin{aligned} \delta[\ln(D)] \left(\frac{1}{T} \right) &= t_{n-2, 0.025} \cdot \\ &\sqrt{\frac{1}{n-2} \sum_{i=1}^n \left(\ln(D_i) - \beta_0 - \beta_1 \frac{1}{T_i} \right)^2} \cdot \\ &\sqrt{\frac{1}{n} + \frac{\left(\frac{1}{T} - \langle \frac{1}{T} \rangle \right)^2}{\sum_{j=1}^n \left(\frac{1}{T_j} - \langle \frac{1}{T} \rangle \right)^2}}. \end{aligned} \quad (6)$$

Data availability

Furthermore, the energies, forces, settings, and all other calculated properties obtained from DFT simulation are stored as ASE SQLite3 and can be accessed in the repository available at https://github.com/paolodeangelis/Enhancing_ReaxFF_DFT_database or through the Zenodo permalink <https://doi.org/10.5281/zenodo.7959121>.

Code availability

In this study, we employed several computational tools to carry out our simulations. We used the open-source LAMMPS code for the ReaxFF MD simulations, while the commercial codes BAND and DFTB were utilized for

the DFT and ab initio MD simulations, respectively. These last two codes are all available in the Amsterdam Modeling Suite by SCM. In addition, we employed several Python libraries, including ASE, Pymatgen, and Packmol, to arrange the initial configurations, as well as the PLAMS library and an in-house Python code for automatically managing the simulation campaign. Finally, we utilized the ParAMS library to perform the ReaxFF parametrization. The repository containing all the Jupyter notebooks required for running the ReaxFF reparametrization process, which extensively utilized the aforementioned codes, can be accessed at https://github.com/paolodeangelis/Enhancing_ReaxFF or via Zenodo permalink <https://doi.org/10.5281/zenodo.8036775>.

Received: 6 October 2023; Accepted: 28 December 2023

Published online: 10 January 2024

References

- Wen, J., Zhao, D. & Zhang, C. An overview of electricity powered vehicles: Lithium-ion battery energy storage density and energy conversion efficiency. *Renew. Energy* **162**, 1629–1648. <https://doi.org/10.1016/j.renene.2020.09.055> (2020).
- Kalair, A., Abas, N., Saleem, M. S., Kalair, A. R. & Khan, N. Role of energy storage systems in energy transition from fossil fuels to renewables. *Energy Storage* **3**, e135. <https://doi.org/10.1002/est2.135> (2021).
- Amici, J. *et al.* A roadmap for transforming research to invent the batteries of the future designed within the European large scale research initiative battery 2030+. *Adv. Energy Mater.* **12**, 2102785. <https://doi.org/10.1002/aenm.202102785> (2022).
- De Angelis, P. *et al.* Data-driven appraisal of renewable energy potentials for sustainable freshwater production in Africa. *Renew. Sustain. Energy Rev.* **149**, 111414. <https://doi.org/10.1016/j.rser.2021.111414> (2021).
- Kim, T.-H. *et al.* The current move of lithium ion batteries towards the next phase. *Adv. Energy Mater.* **2**, 860–872. <https://doi.org/10.1002/aenm.201200028> (2012).
- Blomgren, G. E. The development and future of lithium ion batteries. *J. Electrochem. Soc.* **164**, A5019. <https://doi.org/10.1149/2.0251701jes> (2016).
- Peled, E. The electrochemical behavior of alkali and alkaline earth metals in nonaqueous battery systems—the solid electrolyte interphase model. *J. Electrochem. Soc.* **126**, 2047. <https://doi.org/10.1149/1.2128859> (1979).
- Winter, M. The solid electrolyte interphase—the most important and the least understood solid electrolyte in rechargeable Li batteries. *Z. Phys. Chem.* **223**, 1395–1406. <https://doi.org/10.1524/zpch.2009.6086> (2009).
- Cappabianca, R., De Angelis, P., Fasano, M., Chiavazzo, E. & Asinari, P. An overview on transport phenomena within solid electrolyte interphase and their impact on the performance and durability of lithium-ion batteries. *Energies* <https://doi.org/10.3390/en16135003> (2023).
- Goodenough, J. B. & Kim, Y. Challenges for rechargeable Li batteries. *Chem. Mater.* **22**, 587–603. <https://doi.org/10.1021/cm901452z> (2010).
- Lee, H. G., Kim, S. Y. & Lee, J. S. Dynamic observation of dendrite growth on lithium metal anode during battery charging/discharging cycles. *npj Comput. Mater.* **8**, 103. <https://doi.org/10.1038/s41524-022-00788-6> (2022).
- Single, F., Latz, A. & Horstmann, B. Identifying the mechanism of continued growth of the solid-electrolyte interphase. *Chem-suschem* **11**, 1950–1955. <https://doi.org/10.1002/cssc.201800077> (2018).
- Liu, J., Wang, Z. & Bai, J. Influences of multi factors on thermal runaway induced by overcharging of lithium-ion battery. *J. Energy Chem.* **70**, 531–541. <https://doi.org/10.1016/j.jechem.2022.03.011> (2022).
- Wang, A., Kadam, S., Li, H., Shi, S. & Qi, Y. Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries. *npj Comput. Mater.* **4**, 1–26. <https://doi.org/10.1038/s41524-018-0064-0> (2018).
- Shan, X. *et al.* A brief review on solid electrolyte interphase composition characterization technology for lithium metal batteries: challenges and perspectives. *J. Phys. Chem. C* **125**, 19060–19080. <https://doi.org/10.1021/acs.jpcc.1c06277> (2021).
- Bhowmik, A. *et al.* A perspective on inverse design of battery interphases using multi-scale modelling, experiments and generative deep learning. *Energy Storage Mater.* **21**, 446–456. <https://doi.org/10.1016/j.ensm.2019.06.011> (2019).
- Zhou, J. *et al.* Eliminating graphite exfoliation with an artificial solid electrolyte interphase for stable lithium-ion batteries. *Small* **18**, 2107460. <https://doi.org/10.1002/smll.202107460> (2022).
- Aslam, M. K. *et al.* How to avoid dendrite formation in metal batteries: innovative strategies for dendrite suppression. *Nano Energy* **86**, 106142. <https://doi.org/10.1016/j.nanoen.2021.106142> (2021).
- Wang, Y., Nakamura, S., Ue, M. & Balbuena, P. B. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: reduction mechanisms of ethylene carbonate. *J. Am. Chem. Soc.* **123**, 11708–11718. <https://doi.org/10.1021/ja0164529> (2001).
- Tasaki, K., Kanda, K., Nakamura, S. & Ue, M. Decomposition of LiPF₆ and stability of PF₅ in Li-ion battery electrolytes: Density functional theory and molecular dynamics studies. *J. Electrochem. Soc.* **150**, A1628. <https://doi.org/10.1149/1.1622406> (2003).
- Tasaki, K. Solvent decompositions and physical properties of decomposition compounds in Li-ion battery electrolytes studied by DFT calculations and molecular dynamics simulations. *J. Phys. Chem. B* **109**, 2920–2933. <https://doi.org/10.1021/jp047240b> (2005).
- Leung, K. & Budzien, J. L. Ab initio molecular dynamics simulations of the initial stages of solid-electrolyte interphase formation on lithium ion battery graphitic anodes. *Phys. Chem. Chem. Phys.* **12**, 6583–6586. <https://doi.org/10.1039/B925853A> (2010).
- Leung, K. Electronic structure modeling of electrochemical reactions at electrode/electrolyte interfaces in lithium ion batteries. *J. Phys. Chem. C* **117**, 1539–1547. <https://doi.org/10.1021/jp308929a> (2013).
- Okuno, Y., Ushirogata, K., Sodeyama, K. & Tateyama, Y. Decomposition of the fluoroethylene carbonate additive and the glue effect of lithium fluoride products for the solid electrolyte interphase: an ab initio study. *Phys. Chem. Chem. Phys.* **18**, 8643–8653. <https://doi.org/10.1039/C5CP07583A> (2016).
- Wang, Y., Nakamura, S., Tasaki, K. & Balbuena, P. B. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: how does vinylene carbonate play its role as an electrolyte additive? *J. Am. Chem. Soc.* **124**, 4408–4421. <https://doi.org/10.1021/ja017073i> (2002).
- Leung, K. Predicting the voltage dependence of interfacial electrochemical processes at lithium-intercalated graphite edge planes. *Phys. Chem. Chem. Phys.* **17**, 1637–1643. <https://doi.org/10.1039/C4CP04494K> (2015).
- Yu, J., Balbuena, P. B., Budzien, J. & Leung, K. Hybrid DFT functional-based static and molecular dynamics studies of excess electron in liquid ethylene carbonate. *J. Electrochem. Soc.* **158**, A400. <https://doi.org/10.1149/1.3545977> (2011).
- Benitez, L. & Seminario, J. M. Electron transport and electrolyte reduction in the solid-electrolyte interphase of rechargeable lithium ion batteries with silicon anodes. *J. Phys. Chem. C* **120**, 17978–17988. <https://doi.org/10.1021/acs.jpcc.6b06446> (2016).
- Borodin, O., Zhuang, G. V., Ross, P. N. & Xu, K. Molecular dynamics simulations and experimental study of lithium ion transport in dilithium ethylene dicarbonate. *J. Phys. Chem. C* **117**, 7433–7444. <https://doi.org/10.1021/jp4000494> (2013).
- Benitez, L., Cristancho, D., Seminario, J., De La Hoz, J. M. & Balbuena, P. Electron transfer through solid-electrolyte-interphase layers formed on Si anodes of Li-ion batteries. *Electrochim. Acta* **140**, 250–257. <https://doi.org/10.1016/j.electacta.2014.05.018> (2014).

31. Zheng, J. *et al.* Lithium ion diffusion mechanism on the inorganic components of the solid-electrolyte interphase. *J. Mater. Chem. A* **9**, 10251–10259. <https://doi.org/10.1039/D0TA11444H> (2021).
32. Garcia-Lastra, J. M., Myrdal, J. S., Christensen, R., Thygesen, K. S. & Vegge, T. Dft+ u study of polaronic conduction in li2o2 and li2co3: implications for li-air batteries. *J. Phys. Chem. C* **117**, 5568–5577. <https://doi.org/10.1021/jp3107809> (2013).
33. Chen, Y., Ouyang, C., Song, L. & Sun, Z. Electrical and lithium ion dynamics in three main components of solid electrolyte interphase from density functional theory study. *J. Phys. Chem. C* **115**, 7044–7049. <https://doi.org/10.1021/jp112202s> (2011).
34. Shang, S.-L. *et al.* Lattice dynamics, thermodynamics and elastic properties of monoclinic li2co3 from density functional theory. *Acta Mater.* **60**, 5204–5216. <https://doi.org/10.1016/j.actamat.2012.06.006> (2012).
35. Shin, H., Park, J., Han, S., Sastry, A. M. & Lu, W. Component-/structure-dependent elasticity of solid electrolyte interphase layer in li-ion batteries: Experimental and computational studies. *J. Power Sour.* **277**, 169–179. <https://doi.org/10.1016/j.jpowsour.2014.11.120> (2015).
36. Bedrov, D., Borodin, O. & Hooper, J. B. Li+ transport and mechanical properties of model solid electrolyte interphases (sei): insight from atomistic molecular dynamics simulations. *J. Phys. Chem. C* **121**, 16098–16109. <https://doi.org/10.1021/acs.jpcc.7b04247> (2017).
37. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50. [https://doi.org/10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0) (1996).
38. Shaidu, Y. *et al.* A systematic approach to generating accurate neural network potentials: The case of carbon. *npj Comput. Mater.* **7**, 1–13. <https://doi.org/10.1038/s41524-021-00508-6> (2021).
39. Van Duin, A. C., Dasgupta, S., Lorant, F. & Goddard, W. A. Reaxff: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409. <https://doi.org/10.1021/jp004368u> (2001).
40. Mortier, W. J., Ghosh, S. K. & Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **108**, 4315–4320. <https://doi.org/10.1021/ja00275a013> (1986).
41. Kim, S.-P., Van Duin, A. C. & Shenoy, V. B. Effect of electrolytes on the structure and evolution of the solid electrolyte interphase (sei) in li-ion batteries: A molecular dynamics study. *J. Power Sour.* **196**, 8590–8597. <https://doi.org/10.1016/j.jpowsour.2011.05.061> (2011).
42. Han, S. S., Van Duin, A. C., Goddard, W. A. & Lee, H. M. Optimization and application of lithium parameters for the reactive force field, reaxff. *J. Phys. Chem. A* **109**, 4575–4582. <https://doi.org/10.1021/jp051450m> (2005).
43. Guk, H., Kim, D., Choi, S.-H., Chung, D. H. & Han, S. S. Thermostable artificial solid-electrolyte interface layer covalently linked to graphite for lithium ion battery: molecular dynamics simulations. *J. Electrochem. Soc.* **163**, A917. <https://doi.org/10.1149/2.0611606jes> (2016).
44. Yun, K.-S. *et al.* Simulation protocol for prediction of a solid-electrolyte interphase on the silicon-based anodes of a lithium-ion battery: Reaxff reactive force field. *J. Phys. Chem. Lett.* **8**, 2812–2818. <https://doi.org/10.1021/acs.jpclett.7b00898> (2017).
45. Szczech, J. R. & Jin, S. Nanostructured silicon for high capacity lithium battery anodes. *Energy Environ. Sci.* **4**, 56–72. <https://doi.org/10.1039/c0ee00281j> (2011).
46. Wang, Y., Liu, Y., Tu, Y. & Wang, Q. Reductive decomposition of solvents and additives toward solid-electrolyte interphase formation in lithium-ion battery. *J. Phys. Chem. C* **124**, 9099–9108. <https://doi.org/10.1021/acs.jpcc.9b10535> (2020).
47. Choudhury, S. & Choudhury, S. Lithium fluoride additives for stable cycling of lithium batteries at high current densities. *Rational Design of Nanostructured Polymer Electrolytes and Solid-Liquid Interphases for Lithium Batteries* 81–94, https://doi.org/10.1007/978-3-030-28943-0_5 (2019).
48. Ko, J. & Yoon, Y. S. Recent progress in lif materials for safe lithium metal anode of rechargeable batteries: Is lif the key to commercializing li metal batteries?. *Ceram. Int.* **45**, 30–49. <https://doi.org/10.1016/j.ceramint.2018.09.287> (2019).
49. Tan, Y.-H. *et al.* Lithium fluoride in electrolyte for stable and safe lithium-metal batteries (adv. mater. 42/2021). *Adv. Mater.* **33**, 2170331. <https://doi.org/10.1002/adma.202102134> (2021).
50. Shi, S. *et al.* Multi-scale computation methods: Their applications in lithium-ion battery research and development. *Chin. Phys. B* **25**, 018212. <https://doi.org/10.1088/1674-1056/25/1/018212> (2015).
51. Larsen, A. H. *et al.* The atomic simulation environment—a python library for working with atoms. *J. Phys.: Condens. Matter* **29**, 273002. <https://doi.org/10.1088/1361-648X/aa680e> (2017).
52. Atomic Simulation Environment (ASE) website.
53. Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319. <https://doi.org/10.1016/j.commatsci.2012.10.028> (2013).
54. Python Materials Genomics (pymatgen) website.
55. Komissarov, L., Ruger, R., Hellstrom, M. & Verstraelen, T. Params: parameter optimization for atomistic and molecular simulations. *J. Chem. Inf. Model.* **61**, 3737–3743. <https://doi.org/10.1021/acs.jcim.1c00333> (2021).
56. Parameterization Tools for AMS (ParAMS).
57. Mendez, K. M., Pritchard, L., Reinke, S. N. & Broadhurst, D. I. Toward collaborative open data science in metabolomics using jupyter notebooks and cloud computing. *Metabolomics* **15**, 1–16. <https://doi.org/10.1007/s11306-019-1588-0> (2019).
58. Loizides, F. & Schmidt, B. (eds.) *Positioning and power in academic publishing: players, agents and agendas: Proceedings of the 20th International Conference on Electronic Publishing* (Ios Press, Inc, 2016). OCLC: ocn951955627.
59. Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002. <https://doi.org/10.1063/1.4812323> (2013).
60. Huber, S. P. *et al.* Aiida 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* **7**, 300. <https://doi.org/10.1038/s41597-020-00638-4> (2020).
61. AiiDA website.
62. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327. <https://doi.org/10.1002/jcc.21787> (2011).
63. MDAnalysis website.
64. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362. <https://doi.org/10.1038/s41586-020-2649-2> (2020).
65. Crystal Toolkit.
66. Kermode, J. R. f90wrap: an automated tool for constructing deep python interfaces to modern fortran codes. *J. Phys.: Condens. Matter* <https://doi.org/10.1088/1361-648X/ab82d2> (2020).
67. QUIP and quippy documentation.
68. Csányi, G. *et al.* Expressive programming for computational physics in fortran 95+. *IoP Comput. Phys. Newsletter Spring 2007* (2007).
69. Kühne, T. D. *et al.* Cp2k: An electronic structure and molecular dynamics software package-quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103. <https://doi.org/10.1063/5.0007045> (2020).
70. Thompson, A. P. *et al.* LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171. <https://doi.org/10.1016/j.cpc.2021.108171> (2022).
71. Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS).
72. Michał, H. PLAMS.
73. PLAMS, SCM website.
74. Amsterdam Modeling Suite, SCM website.

75. Humphrey, W., Dalke, A. & Schulten, K. Vmd: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) (1996).
76. Corey, R. B. & Pauling, L. Molecular models of amino acids, peptides, and proteins. *Rev. Sci. Instrum.* **24**, 621–627. <https://doi.org/10.1063/1.1770803> (1953).
77. Chenoweth, K., Van Duin, A. C. & Goddard, W. A. Reaxff reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *J. Phys. Chem. A* **112**, 1040–1053. <https://doi.org/10.1021/jp709896w> (2008).
78. LaBrosse, M. R., Johnson, J. K. & Van Duin, A. C. Development of a transferable reactive force field for cobalt. *J. Phys. Chem. A* **114**, 5855–5861. <https://doi.org/10.1021/jp911867r> (2010).
79. De Angelis, P., Cappabianca, R., Fasano, M., Asianri, P. & Chiavazzo, E. *Enhancing the ReaxFF DFT database* <https://doi.org/10.5072/zenodo.1204707> (2023).
80. Wilkinson, M. D. *et al.* The fair guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9. <https://doi.org/10.1038/sdata.2016.18> (2016).
81. Hansen, N. & Ostermeier, A. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.* **9**, 159–195. <https://doi.org/10.1162/106365601750190398> (2001).
82. Wang, Y. *et al.* Development of a transferable reaxff parameter set for carbon-and silicon-based solid systems. *J. Phys. Chem. C* **124**, 10007–10015. <https://doi.org/10.1021/acs.jpcc.0c01645> (2020).
83. Lide, D. R. & CRC Press (eds.) *CRC handbook of chemistry and physics: A ready-reference book of chemical and physical data* (CRC, Taylor & Francis, 2006), 87. ed., 2006–2007 edn.
84. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322**, 549–560. <https://doi.org/10.1002/andp.19053220806> (1905).
85. Von Smoluchowski, M. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys.* **326**, 756–780. <https://doi.org/10.1002/andp.19063261405> (1906).
86. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* 2nd edn. (Oxford University Press, 2017).
87. Lee, S.-I. *et al.* A study of electrochemical kinetics of lithium ion in organic electrolytes. *Korean J. Chem. Eng.* **19**, 638–644. <https://doi.org/10.1007/BF02699310> (2002).
88. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904. <https://doi.org/10.1063/1.1329672> (2000).
89. LAMMPS developers. LAMMPS benchmark: Interatomic potential comparisons.
90. Senftle, T. P. *et al.* The reaxff reactive force-field: Development, applications and future directions. *npj Comput. Mater.* **2**, 1–14. <https://doi.org/10.1038/npjcompumats.2015.11> (2016).
91. Fiesinger, F. *et al.* Development of a mg/o reaxff potential to describe the passivation processes in magnesium-ion batteries. *ChemSuschem* **16**, e202201821. <https://doi.org/10.1002/cssc.202201821> (2023).
92. Olou’ou Guifo, S. B. *et al.* Development and validation of a reaxff reactive force field for modeling silicon–carbon composite anode materials in lithium-ion batteries. *J. Phys. Chem. C* **127**, 2818–2834. <https://doi.org/10.1021/acs.jpcc.2c07773> (2023).
93. Castro-Marciano, F., Kamat, A. M., Russo, M. F. Jr., van Duin, A. C. & Mathews, J. P. Combustion of an illinois no. 6 coal char simulated using an atomistic char representation and the reaxff reactive force field. *Combust. Flame* **159**, 1272–1285. <https://doi.org/10.1016/j.combustflame.2011.10.022> (2012).
94. Strachan, A., van Duin, A. C., Chakraborty, D., Dasgupta, S. & Goddard, W. A. III. Shock waves in high-energy materials: The initial chemical events in nitramine rdx. *Phys. Rev. Lett.* **91**, 098301. <https://doi.org/10.1103/PhysRevLett.91.098301> (2003).
95. Weismiller, M. R., Van Duin, A. C., Lee, J. & Yetter, R. A. Reaxff reactive force field development and applications for molecular dynamics simulations of ammonia borane dehydrogenation and combustion. *J. Phys. Chem. A* **114**, 5485–5492. <https://doi.org/10.1021/jp100136c> (2010).
96. Agrawalla, S. & Van Duin, A. C. Development and application of a reaxff reactive force field for hydrogen combustion. *J. Phys. Chem. A* **115**, 960–972. <https://doi.org/10.1021/jp108325e> (2011).
97. Wang, H., Zhang, L., Han, J. & Weinan, E. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178–184. <https://doi.org/10.1016/j.cpc.2018.03.016> (2018).
98. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401. <https://doi.org/10.1103/PhysRevLett.98.146401> (2007).
99. Unke, O. T. *et al.* Machine learning force fields. *Chem. Rev.* **121**, 10142–10186. <https://doi.org/10.1021/acs.chemrev.0c01111> (2021).
100. Yang, M., Bonati, L., Polino, D. & Parrinello, M. Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water. *Catal. Today* **387**, 143–149. <https://doi.org/10.1016/j.cattod.2021.03.018> (2022).
101. Yang, M., Raucci, U. & Parrinello, M. Reactant-induced dynamics of lithium imide surfaces during the ammonia decomposition process. *Nature Catal.* <https://doi.org/10.1038/s41929-023-01006-2> (2023).
102. Trezza, G., Bergamasco, L., Fasano, M. & Chiavazzo, E. Minimal crystallographic descriptors of sorption properties in hypothetical mofs and role in sequential learning optimization. *npj Comput. Mater.* **8**, 123. <https://doi.org/10.1038/s41524-022-00806-7> (2022).
103. Wang, Z. *et al.* Heterogeneous relational message passing networks for molecular dynamics simulations. *npj Comput. Mater.* **8**, 53. <https://doi.org/10.1038/s41524-022-00739-1> (2022).
104. Jussupow, A. & Kaila, V. R. Effective molecular dynamics from neural network-based structure prediction models. *J. Chem. Theory Comput.* **19**, 1965–1975. <https://doi.org/10.1021/acs.jctc.2c01027> (2023).
105. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115. <https://doi.org/10.1103/PhysRevB.87.184115> (2013).
106. Himanen, L. *et al.* Dscribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949. <https://doi.org/10.1016/j.cpc.2019.106949> (2020).
107. Liu, Y. *et al.* Data quantity governance for machine learning in materials science. *Natl. Sci. Rev.* <https://doi.org/10.1093/nsr/nwad125> (2023).
108. De Angelis, P., Cappabianca, R., Asinari, P. & Chiavazzo, E. Sei builder - jupyter notebooks to build initial sei morphology. <https://doi.org/10.5281/zenodo.6470784> (2022).
109. Grimme, S., Bannwarth, C. & Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (z = 1–86). *J. Chem. Theory Comput.* **13**, 1989–2009. <https://doi.org/10.1021/acs.jctc.7b00118> (2017).
110. Te Velde, G. & Baerends, E. Precise density-functional method for periodic structures. *Phys. Rev. B* **44**, 7888 (1991).
111. SCM. BAND.
112. Perdew, J. P. *et al.* Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406. <https://doi.org/10.1103/physrevlett.100.136406> (2008).
113. Aktulga, H. M., Fogarty, J. C., Pandit, S. A. & Grama, A. Y. Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques. *Parallel Comput.* **38**, 245–259. <https://doi.org/10.1016/j.parco.2011.08.005> (2012).
114. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. Packmol: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164. <https://doi.org/10.1002/jcc.21224> (2009).

115. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268. <https://doi.org/10.1080/00268978400101201> (1984).
116. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die reine und angewandte Mathematik (Crelles Journal)* **1908**, 97–102. <https://doi.org/10.1515/crll.1908.134.198> (1908).
117. Navidi, W. C. *Statistics for Engineers and Scientists* 5th edn. (McGraw-Hill Education, 2020).

Acknowledgements

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189 (BIG-MAP project). The authors also acknowledge that the results of this research have been achieved using the DECI resource ARCHER2 based in UK at EPCC with support from the PRACE aisbl.

Author contributions

The research project was conceptualized and designed collaboratively by P.D.A., M.F., P.A., and E.C. In particular, they collectively discussed the feasibility and suitability of using ReaxFF for investigating the SEI formation. E.C. and P.D.A. subsequently decided to focus on the LiF compound. P.D.A. developed, tested, and refined the protocol for reoptimizing the ReaxFF, and conducted the simulations and postprocessing of the data with the assistance of R.C. P.D.A., R.C. and M.F. worked on analyzing and interpreting the results, and P.D.A. wrote the manuscript with input from R.C., M.F., E.C. and P.A. The project was supervised by E.C. and P.A. All the authors have read and approved the final manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-50978-5>.

Correspondence and requests for materials should be addressed to P.D.A., P.A. or E.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024