

RESEARCH ARTICLE

Accelerating the theoretical study of Li-polysulfide adsorption on single-atom catalysts via machine learning approaches

 Eleftherios I. Andritsos¹  | Kevin Rossi² 
¹Department of Physics, King's College London, London, UK

²Laboratory of Nanochemistry for Energy, Institute of Chemistry, Ecole Polytechnique Fédérale de Lausanne, Sion, Switzerland
Correspondence

Eleftherios I. Andritsos, Department of Physics, King's College London, Strand, London WC2R 2LS, UK.

 Email: lefteri.andritsos@kcl.ac.uk

Kevin Rossi, Laboratory of Nanochemistry for Energy, Institute of Chemistry, Ecole Polytechnique Fédérale de Lausanne, Sion, Switzerland.

 Email: kevin.rossi@epfl.ch
Funding information

EPSRC, Grant/Award Numbers: EP/T022213/1, EP/P020194/1; European Research Council, Grant/Award Number: 890414

Abstract

Li-S batteries are a promising alternative to Li-ion batteries, offering large energy storage capacity and wide operating temperature range. However, their performance is heavily affected by the Li-polysulfide (LiPS) shuttling. Computational screening of LiPS adsorption on single-atom catalyst (SAC) substrates is of great aid to the design of Li-S batteries which are robust against the LiPS shuttling from the cathode to the anode and the electrolyte. To facilitate this process, we develop a machine learning (ML) protocol to accelerate the systematic mapping of dominant local energy minima found with calculations based on the density functional theory (DFT), and, in turn, fast screening of LiPS adsorption properties on SACs. We first validate the approach by probing the potential energy surface for LiPS adsorbed on graphene decorated with a Fe-N₄-C SAC. We identify minima whose binding energies are better or on par with the one previously reported in the literature. We then move to analyze the adsorption trends on Zn-N₄-C SAC and observe similar adsorption strength and behavior with the Fe-N₄-C SAC, highlighting the good predictive power of our protocol. Our approach offers a comprehensive and computationally efficient alternative to conventional approaches studying LiPS adsorption.

KEYWORDS
 adsorption, DFT, Li-S batteries, Li-polysulfide (LiPS) shuttling, machine learning, single-atom catalysts, Zn-N₄-C SAC

1 | INTRODUCTION

Lithium-sulfur (Li-S) batteries are an emerging technology in the field of energy storage. Their high-density energy storage capacity, of ~2500 Wh kg⁻¹ theoretical specific energy [1], is indeed much higher than that of Li-ion or other commercial batteries. However, in practice, the Li-S battery rate capability and life cycle degrades severely over time because of the so called “shuttle” effect. The latter is caused by lithium polysulfide (LiPS) molecules remaining in the electrolyte during the charge-discharge process, leading to loss of active material and reduced battery performance.

Controlling the LiPS rate of migration to the anode and the reaction kinetics in the cathode is important for tackling the issue of the “shuttle” effect. Several strategies have been adopted to reduce the diffusivity of LiPS to the electrolyte, mainly focusing on the development of cathode host materials that can strongly bind the LiPS while also improving the reaction kinetics [2-4]. However, regardless of the vast amount of research in the field over the recent years, prevention of battery degradation still remains a challenge.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *International Journal of Quantum Chemistry* published by Wiley Periodicals LLC.

To overcome the issue, recent studies focus on single-atom catalyst (SAC) materials. The latter provide strong polysulfide adsorption and improve the reaction kinetics, thanks to their exceptional electrocatalytic properties. Recent studies indeed suggest that SAC with atomically dispersed transition metals (TMs) on graphene lattices are particularly promising candidates. In particular, calculations on the effect of SACs with TM-N₄-C formation (TM = Co, Fe, Mn, Ru, V, W, Zn, and C as a graphene lattice) on the LiPS adsorption, the Gibbs free energy change and the reaction activation barrier [5–13], predict a remarkable improvement in the cathode properties which could potentially reduce the LiPS “shuttle” effect.

Numerical methods provide ground to screen SAC adsorption properties. The identification of the global energy minimum LiPS/SAC structure, a stepping stone to predict adsorption trends, is however rather challenging. Several different initial guess structures and structural optimizations have to be considered in order to identify the global minimum structure. For example, in high-order LiPS intermediates (Li₂S₄, Li₂S₆, and Li₂S₈) the final binding energies varies over .5 eV [14] depending on the initial LiPS structure that is used as input in the relaxation. By the same token, reported values of Li₂S₆ binding energy on Fe-based SAC, one of the most examined SAC material for Li-S batteries, vary by about .6 eV, from −.9 eV to −1.5 eV [5, 7, 9, 13]. The large discrepancy in the results, raises concerns regarding the quality of the protocols adopted to identify the minimum energy structure, and adds uncertainty when comparing SAC results from different studies. Here, we demonstrate the use of machine learning (ML) methods to accelerate the prediction of LiPS binding energies on SAC materials.

The proposed approach gauges on the recent advances in ML methods applied to the modeling of catalysts and materials for energy applications [15, 16]. While application of ML methods emerged as a new paradigm to accelerate the exploration of complex interactions between small and medium molecules with a variety of binding sites, [17–20] and to predict properties in energy storage materials [21–29] data-driven approaches in the field of Li-S batteries still remain under-exploited [30].

Data being the primary key in data-driven approaches, the construction of non-noisy databases has been the first stepping-stone toward the rapid development of ML accelerated design of Li-based materials, which exploit experimental data [31–33]. On the other hand, ML provide a platform to largely speed-up quantum mechanical calculations, which can be promptly organized in homogeneous databases, and enable in silico high-throughput screening [34]. To date, ML modeling in Li-S batteries has been used to examine the impact of materials and battery design on the battery performance [35]. Also, the LiPS adsorption on two-dimensional layer-structured MoSe₂ and WSe₂ was probed by calculating the binding energy of LiPSs for arbitrary spatial configurations at random sites on the host material [36]. Further, a recent study combined a machine learning framework, based on a modified Crystal Graph Convolutional Neural Network and DFT calculations to offer rational design of SACs for Li-S batteries [37]. The authors focus mainly on the low-order LiPS and S₈ to reveal the binding energy pattern of LiPSs on a wide range of SACs.

In this work, we develop a protocol which hinges on a supervised/unsupervised machine learning strategy, in combination with density functional theory (DFT) calculations, to efficiently screen the stable structures of Li₂S_{*n*} (*n* = 1, 2, 4, 6, 8) on TM-N₄-C (TM = Fe, Zn) anchoring materials. The Fe based system will be used to benchmark with previous results in the literature [13]. The Zn-containing material is instead selected as a focus of our investigation because it presents a similar chemistry to the one of Fe, but with a potential for improved properties when applied as a LiPS anchoring site [31, 32, 38]. To isolate a structurally diverse and energetically favorable set of structures from a database encompassing roto-translations of LiPSs on a graphene-based SAC substrate we adopt unsupervised and supervised machine learning schemes. We then show that by relaxing these structures we identify SAC/LiPS geometries whose binding energy is on par or lower with those reported in the literature. We thus demonstrate how machine learning methods of different nature can accelerate the systematic identification of the energetically favorable LiPS/SAC substrate geometries.

2 | COMPUTATIONAL METHODS

The workflow we adopt in this study is visualized in Figure 1. While we refer the interested reader to the dedicated subsections (2.3–2.5) for detailed technical description on each technique, we here briefly summarize the key aspect in each step of the workflow:

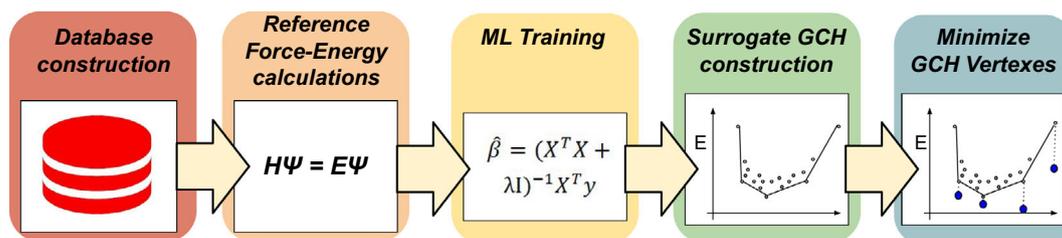


FIGURE 1 Schematic representation of the five steps describing the binding energy prediction workflow in this study

1. Database construction:

- Firstly, we define a supercell space and relax the forces on atoms in the LiPS and in the SAC, each considered as an isolated system, at DFT level. This ensures that the LiPS and SAC structure are more likely to be energetically favorable when modeling LiPS adsorption on the SAC.
- We define a 3D grid in the supercell, where translations of the LiPS will be placed. The grid spacing is .5 Å along the plane parallel to the substrate, and .25 Å along the axis perpendicular to the substrate. For each translation, we apply rotation of 10° in the solid angle centered at the center of mass of the polysulfide, and obtain the full set of structures. The final number roto-translation is reduced and differs for each LiPS, as we apply cutoff criteria so that no LiPS atom is closer than 1.5 Å to atoms in the substrate.

2. Reference force-energy calculations:

- We map each structure generated in the previous step into a high dimensional feature space. The latter is defined through a smooth overlap of atomic orbitals (SOAP) representation [39], employing seven angular and six radial components and a radial cut-off of 6.5 Å.
- We then select a maximum of 10% of the most different structures, in terms of SOAP representation, within the full set of LiPS roto-translations along the substrate, by means of a furthest point sampling (FPS) algorithm [40].
- We calculate from single-point DFT calculations the binding energy corresponding to each of these structures.

3. ML training: we train a Kernel Ridge regressor [41] to predict the binding energies of selected LiPS-SAC configuration, based on its SOAP representation. We utilize the regressor to map predicted binding energies in the full set of roto-translations generated in the initial step of this protocol.

4. Surrogate generalized convex hull (GCH) construction: We build a GCH which utilizes the first principal component analysis (PCA) [42] component projection of the SOAP features assigned to each LiPS-SAC as the free thermodynamic variable, and the predicted binding energy as the energetic coordinate.

5. Minimize GCH Vertexes: We perform DFT geometry optimization to the structures which lie at vertexes of the GCH. We make this choice under the heuristic hypothesis that vertexes of the GCH are the most likely to relax into dissimilar low-energy structures [43].

2.1 | Materials

In this study, we examine two SAC materials with TM-N₄-C formation (TM = Fe, Zn) and five LiPS intermediates (Li₂S_n, n = 1, 2, 4, 6, 8). The TM-N₄-C configurations, based on a modified 72-atom relaxed pristine graphene lattice, consist of 66 C atoms (92.96 at%), four substitutional N atoms (5.63 at%) and one interstitial TM atom (1.41 at%). The TM and two N atoms were manually shifted *out of plane* by .5 Å and .3 Å, respectively, in order to realistically represent the previously observed geometry in SAC materials, when LiPSs are present [7, 9]. The top and side views of the TM-N₄-C, and the relaxed LiPS structures are shown in Figure 2.

2.2 | DFT set up

We perform all spin-polarized DFT simulations with the CASTEP code [44]. We calculate the exchange and correlation potential using the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) functional, using plane-wave cut-off energy of 500 eV. We use energy and force convergence criteria of 10⁻⁵ eV and 5 × 10⁻⁵ eV/Å respectively for the self-consistent field (SCF) for single-point and geometry optimization simulations, and 10⁻⁵ eV and 10⁻³ eV/Å respectively, for the geometry optimization simulations. The sampling of the Brillouin zone is carried on a Monkhorst-Pack grid, using a 3 × 3 × 1 k-point mesh. We include Van der Waals dispersion corrections as described in Grimme's empirical method [45]. We consider periodic boundaries conditions in the simulation box and a vacuum separation distance of 18 Å at the direction normal to the substrate's surface to avoid replica interaction.

We define the binding energy (E_b) as:

$$E_b = E_{\text{mat}} - (E_{\text{SAC}} + E_{\text{LiPS}}), \quad (1)$$

where E_{SAC} , E_{LiPS} , and E_{mat} are the energies of the substrate, LiPS and substrate and LiPS together, respectively. For reference, we report the total energies for the SACs and LiPSs as: $E_{\text{tot}}^{\text{Fe-N}_4\text{-C}} = -12383.41$ eV, $E_{\text{tot}}^{\text{Zn-N}_4\text{-C}} = -13307.78$ eV, $E_{\text{tot}}^{\text{Li}_2\text{S}} = -708.74$ eV, $E_{\text{tot}}^{\text{Li}_2\text{S}_2} = -1013.31$ eV, $E_{\text{tot}}^{\text{Li}_2\text{S}_4} = -1621.69$ eV, $E_{\text{tot}}^{\text{Li}_2\text{S}_6} = -2229.35$ eV, $E_{\text{tot}}^{\text{Li}_2\text{S}_8} = -2836.12$ eV.

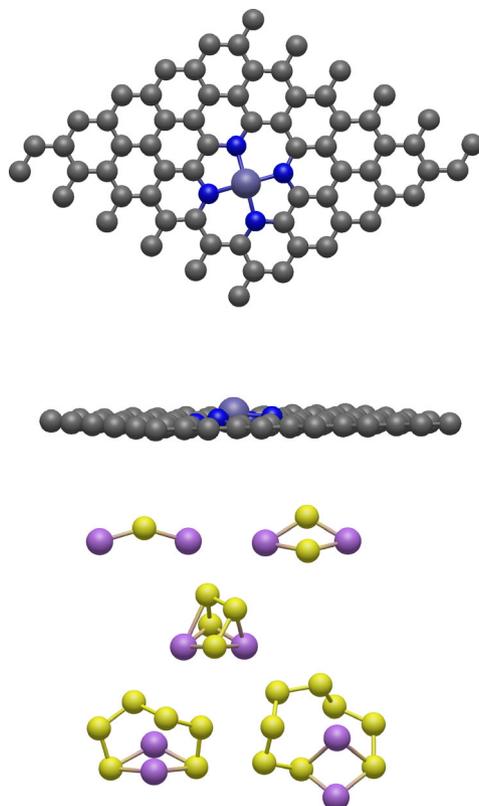


FIGURE 2 Top (A) and side (B) views of TM-N₄-C structure. Geometry optimized Li₂S_n ($n = 1, 2, 4, 6, 8$) structures. Gray, blue, purple, magenta and yellow denote the C, N, TM, Li, and S elements, respectively (C)

2.3 | SOAP representation

Each configuration consisting of a LiPS and the SAC is described by means of D features deriving from the smooth overlap of atomic orbitals representation [39]. This representation is smooth, invariant to permutations, rotations, and translations of the system. Within the SOAP representation, structures are encoded by means Gaussian smeared atomic densities expanded via orthonormal functions, based on spherical harmonics and radial basis functions, centered at atoms positions. Operatively, the SOAP power spectrum is calculated as:

$$p_{n,n',l}^{Z_1,Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m c_{n,l,m}^{Z_1} \times c_{n',l,m}^{Z_2}, \quad (2)$$

where n and n' are the indices for the radial basis functions (up to n_{\max}), l labels the degree of the spherical harmonics (up to l_{\max}) and Z_1 and Z_2 is a label associated to the atom species. The coefficients c_{nlmZ} in Equation (2) are instead defined as (inner products):

$$c_{n,l,m}^Z = \iiint_{\mathcal{R}} dV g_n(r) Y_{l,m}(\theta, \Phi) \rho_Z(r), \quad (3)$$

where $\rho_Z(r)$, is the Gaussian smoothed atomic density for atoms with atomic number Z defined as $Y_{l,m}(\theta, \Phi)$ are the real spherical harmonics, and $g_n(r)$ is the radial basis function (spherical Gaussian type orbitals here).

To calculate the SOAP power spectra we adopt the DESCRIBE [46] library, and utilize the following parameters: $n_{\max} = 6$, $l_{\max} = 7$, and $r_{\text{cut}} = 6.5 \text{ \AA}$. After calculating the power spectrum associated to each atomic sites in a structure, we evaluate an average power spectrum which is associated to the atomic configuration. We utilize the elements in the averaged p power spectrum as the features in kernel ridge regression (KRR). SOAP power spectrum coefficients [47–49], and local-density expansion coefficients more in general [50–53], have been largely successful features in kernel-based and linear machine learning models for structure classification and energy regression.

SOAP representations, and in general local-density representations, when coupled with kernel and linear methods alike, are proven to be capable of accurate and fast predictions of atomistic systems properties in bulk and of molecular materials with diverse chemistries, also providing fairly accurate predictions when trained on small ($\sim 10^3$ configurations) datasets. [51, 54, 55]. Atom-density representation further encode by

design the physical symmetries (rotation, permutation, translation invariance) which rule interatomic interaction, and, though not-transparent, correlate with more interpretable “classical” descriptors, for example, coordination [56–58].

2.4 | Kernel ridge regression

KRR is a well-known and popular non-parametric regression approach. In particular, it consists of L2-norm regularized least squares fitting which exploits the kernel trick [41], where the latter allows for an efficient framework to transform and compare data in an high number dimensions. Within the framework of kernel methods, data are represented by a matrix of pairwise similarity where each (i, j) entry is defined by the kernel function $k(x_i, x_j)$. The latter returns the dot product of the features vectors associated to each i and j data point in the projected space, for any mapping $\phi : x \rightarrow R^N$. The kernel trick thus allows for operations in a high dimensional space without the explicit need of calculating $\phi(x)$. Given a set of covariate variables x_i , the KRR prediction $f(x)$ is written as a linear combination [59]:

$$f(x) = \sum_{i=1}^N \alpha_i k(x_i, x), \quad (4)$$

and the vector of α_i weights is found by solving:

$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \cdot \mathbf{y}, \quad (5)$$

where \mathbf{K} is the kernel matrix, λ tunes the regularization strength and \mathbf{y} labels response variables in the training set. In the reported case study, we choose λ to 0.025 and a radial basis function kernel (length scale $l = 1$) to evaluate the K_{ij} elements of the \mathbf{K} matrix. We do not exclude that other values of λ and l could also work well, as they have been chosen following a non-systematic trial-and-error procedure.

2.5 | Generalized convex hull

The generalized convex hull (GCH) allows to evaluate the probabilities of structures being stabilized by general thermodynamic constraints. In a nutshell, the convex hull of a finite number of points is defined as the smallest convex set that contains them. In materials science, the convex hull construction is used to identify structures that are stable. Given a set of energies, as a function of one or more thermodynamic variables, only the configurations which lie at the vertexes of the convex hull which are lower in energy will be stable. If a structure does not lie on one of such vertexes it is unstable with respect to decomposition into two or more parent structures (assuming fixed thermodynamic conditions and disregarding possible stabilization due to kinetic effects).

The generalized convex hull construction adopts a geometric fingerprints $\Phi = \{\phi_j\}$, to encode the full structural diversity of a dataset of configurations, rather than one or more explicit thermodynamic variables to predict stable phases. The GCH framework has been successfully used in the past for predicting previously unseen ice phases [56] and molecular crystals polymorphs [43]. Overall, this elegant construction provides a rigorous framework to account on the same footing for both geometrically diversity and predicted energetic stability in the choice of diverse candidates for relaxation at DFT level.

3 | WORKFLOW VALIDATION FOR Fe-N₄-C SAC

3.1 | Li₂S on Fe-N₄-C

To showcase in a detailed manner the application of the workflow we discuss the case of Li₂S binding on a Fe-based SAC. Figure 3A shows a 2D map where the two axes are the first and second principal components of the SOAP features, associated to all structures generated for this system. Each point in the map corresponds to one of those structures. The color scale refers to the order with which structures have been generated during the systematic enumeration of LiPS roto-translations. Points in black corresponds to structures selected via an heuristic FPS scheme where the SOAP features for each structure and an Euclidean metric were considered to evaluate the distance between different structures.

We repeat the same process for the other LiPS molecules and run DFT-based single point calculation on the first ~10% most different structures (i.e., ~2000% out of the total ~20 000% structures per each LiPS), and utilize the ~80% of that data as the training dataset in the Kernel Ridge regression fitting. Features associated to each data point are normalized by removing the mean and scaling to unit variance before performing the KRR training. A grand total of 7793 datapoints are used as the training test, after excluding configurations with too unfavorable

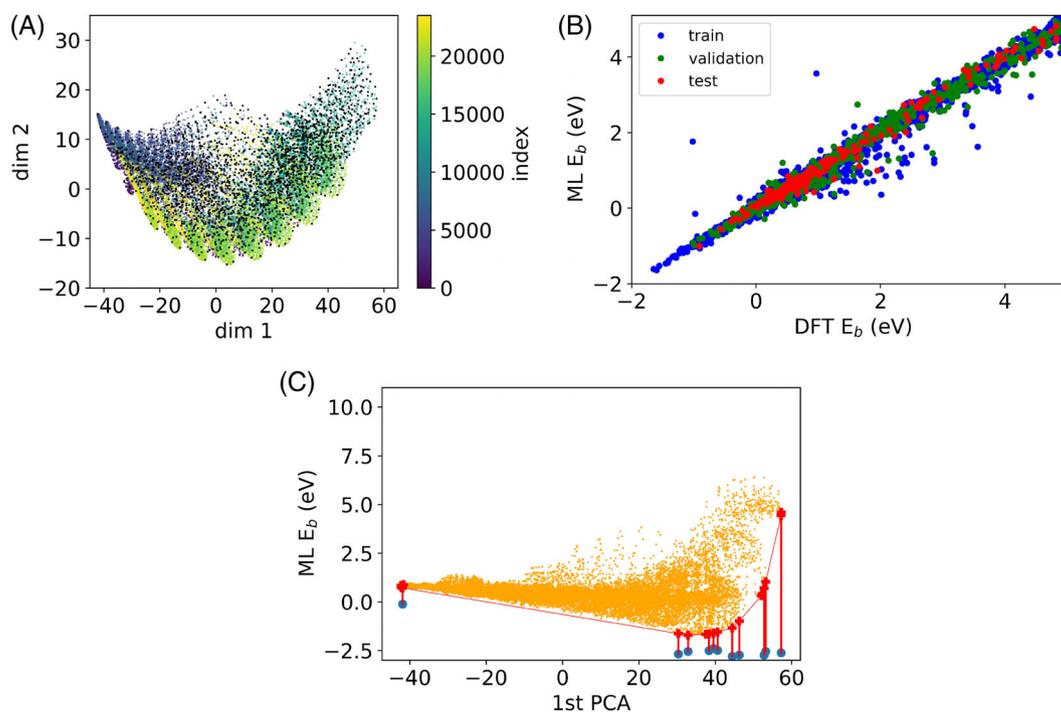


FIGURE 3 (A) Visualization of the SOAP features for initial full set of $\sim 27\,000$ structure generated for Li_2S on Fe-based SAC, when projected in the space determined by their first two PCA components (adimensional). Points in the map are color coded according to their structure index, while the ~ 2000 structures selected by the FPS procedure are shown in black. Note their heterogeneous distribution in the PCA space, which also comprises outliers. (B) Parity plot between the ML-predicted and DFT-calculated binding energies for the training, testing, and validation datapoints. (C) Visualization of the GCH construction when utilizing the first SOAP feature's PCA component as the thermodynamic variable and binding energy as the energetic variables. Vertices are highlighted with red crosses. The binding energy of structures found by relaxing each GCH vertex is shown in blue. A thick line acts as a guide to the eye in connecting each GCH vertex energy to its relaxed configuration binding energy

binding energies. The model Mean Absolute Error (MAE) and R^2 correlation coefficient for training set predictions are .08 eV and .99, respectively. The same quantities for validation set predictions are instead equal to .12 eV and .98. The models errors are well within the acceptable limit for this type of calculations. MAE and R^2 values are obtained by averaging result found when performing an 8-fold cross-validation test [60].

To further evaluate our model's accuracy, we select the $\sim 2\%$ most different structures with respect to the ones in the training set as our test set (i.e., ~ 50 structures per each LiPS). More in detail, if the index 1...20000 list the structures ranked via a furthest point sampling starting from structure 1, the training set comprises a random selection of 80% of structures with index between 1 and 2000, the validation set contains the remaining 20% of structures with index ranging from 1 to 2000, while the test set includes structures 2001 to 2050. We then calculate the binding energy in the test set configurations via single point DFT calculations. The parity plot between the ML predicted and DFT true binding energies for this test is illustrated in Figure 3B. A MAE of .11 eV is registered for the test set, along with a very good correlation ($R^2 = .99$) between the single-point predicted binding energy and the ground truth energy.

We utilize the KRR regressor to evaluate the binding energies for the whole set of Li_2S structures, generated at the first stage of our minima exploration protocol. We then map them on an Energy–First SOAP PCA dimension map, Figure 3C, where each LiPS roto-translated configuration corresponds to one orange dot. The thin red edges and the 11 bold red crosses respectively identify the edges and vertexes of the convex hull construction associated to this dataset. While a number of CH vertexes appear at very negative first PCA component values, we consider only one because of their small structural and energetic diversity. The blue dots label the DFT-level Binding Energy found when relaxing each of the structure lying at a CH vertexes. We then analyze the intermolecular distances in the LiPS molecule and the intramolecular one between the LiPS and the Fe and N atoms in the support, and that all lie in a distinct local minima.

3.2 | LiPSs on Fe-N₄-C

We follow the protocol described for Li_2S_2 to calculate the binding energy for the rest of the examined LiPSs (Li_2S_n , $n = 1, 4, 6, 8$) when interacting with Fe-N₄-C. Our training set initially comprises of approximately 10 000 structures, that is, the first most diverse ~ 2000 Li_2S_n , $n = 1$,

2, 4, 6, 8) structures and associated binding-energies. Configurations with too unfavorable energies (above 5 eV) are then discarded. The training test thus consists of 7793 datapoints. A single ML model to learn predict binding energies regardless of the LiS stoichiometry is finally trained on this dataset. Figure 4 illustrates the GCH construction resulting from the SOAP-KRR model. Each LiPS configuration is reported as an orange dot,

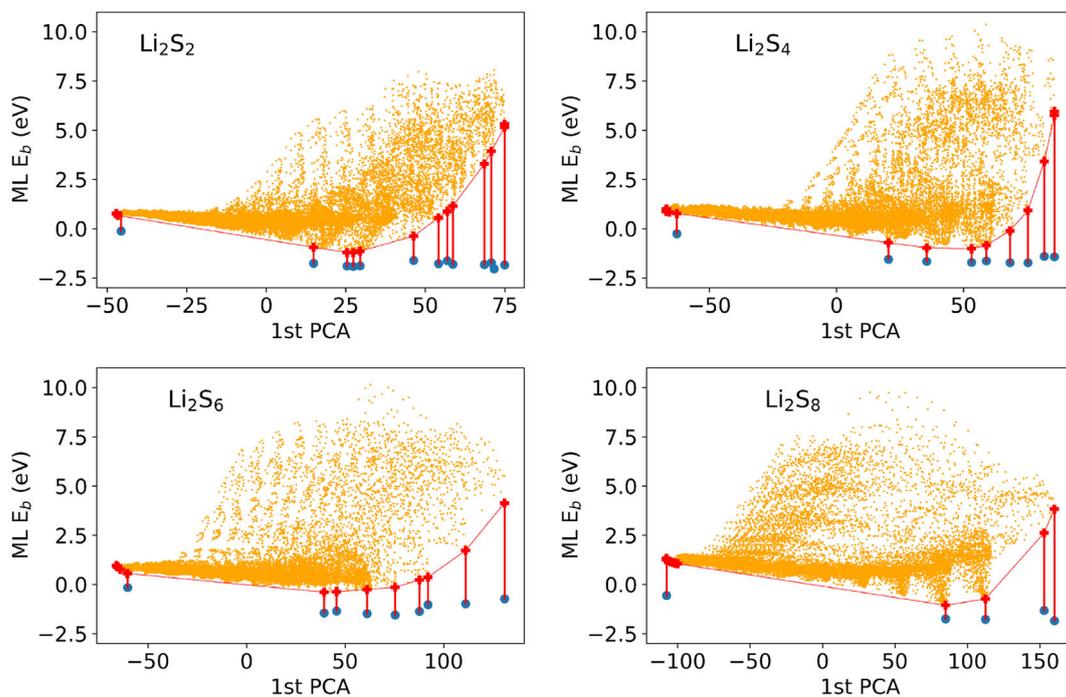


FIGURE 4 The surrogate GCH construction for Li_2S_n ($n = 1, 4, 6, 8$) structures adsorbed on $\text{Fe-N}_4\text{-C}$ is illustrated with orange points. GCH vertices are shown with red crosses. Blue dots show the binding energy corresponding to structures sampled after relaxing the GCH vertices

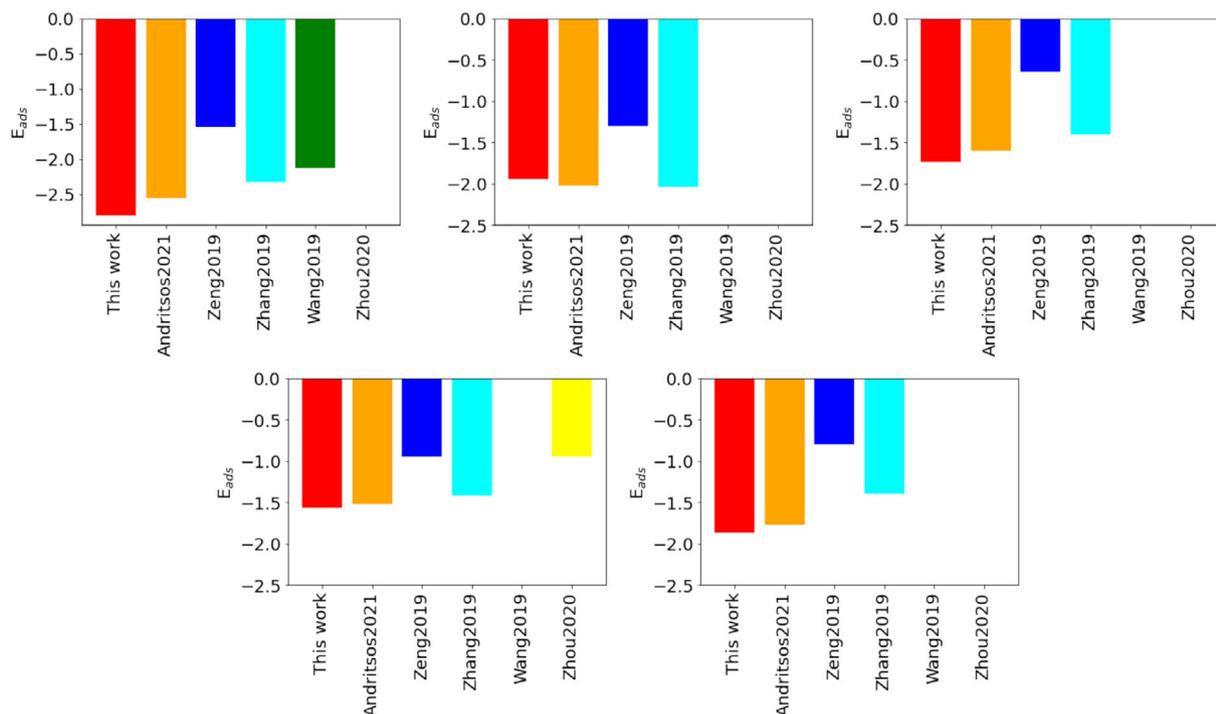


FIGURE 5 Summary of the highest LiPS binding energies for $\text{Fe-N}_4\text{-C}$ SAC calculated by geometry optimized DFT simulations found in “This work” or found in the literature [5–7, 9, 13]. All datapoints are in eV. If datapoint are not available, the bar chart is not shown

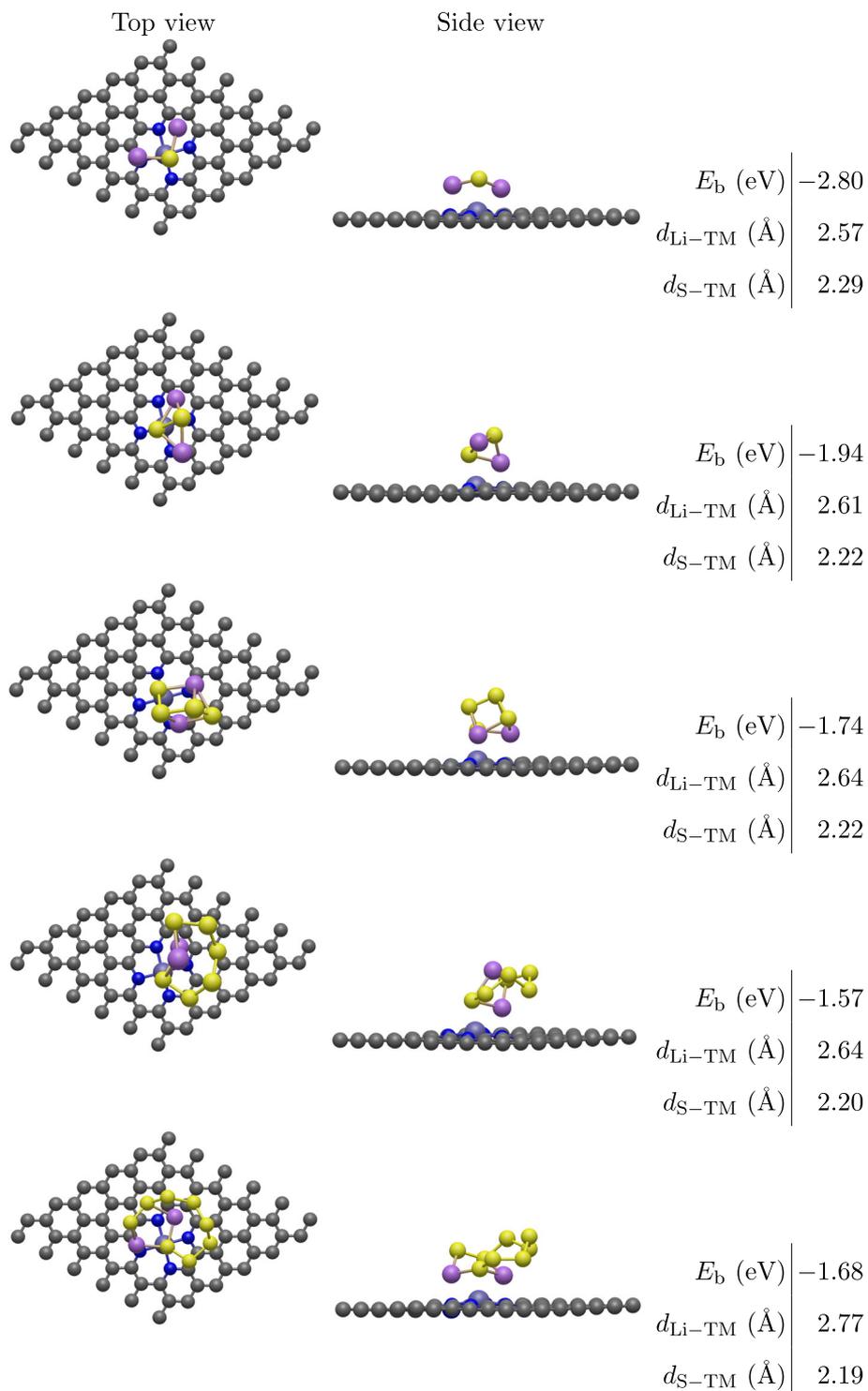


FIGURE 6 Top and side views, binding energy, and structural descriptors of the lowest energy configurations for each LiPS (Li_2S_n , $n = 1, 2, 4, 6, 8$) on $\text{Fe-N}_4\text{-C}$. Color code follows Figure 2

the GCH vertexes are highlighted with red crosses, and the edge connecting to vertexes is visualized with a red line. The blue dots in Figure 4 instead report the DFT-level Binding Energy found when relaxing each GCH vertex. For the four examined cases, the LiPS isomers found to give the lowest binding energies after geometry optimization are GCH vertex with positive 1st PCA component and low single-point binding energy (lower right-hand side region of the plots).

First, we compare the binding energies found with our protocol with respect to those reported in the literature, and gather them in Figure 5. The reported binding energies are in good agreement with reported values from the literature, showing similar or stronger binding,

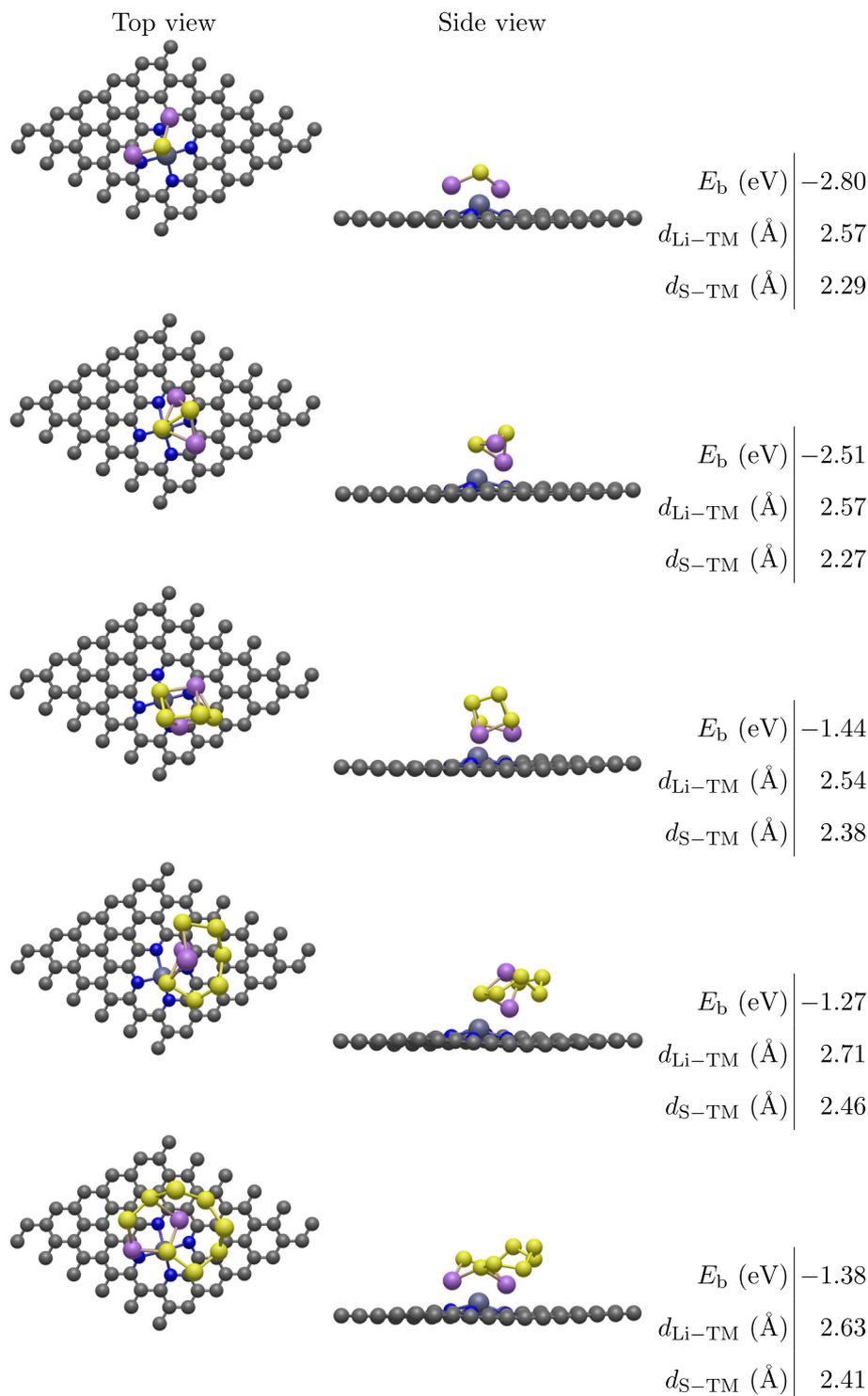


FIGURE 7 Top and side views, binding energy, and structural descriptors of the lowest energy configurations for each LiPS (Li_2S_n , $n = 1, 2, 4, 6, 8$) on $\text{Zn-N}_4\text{-C}$. Color code follows Figure 2

but the Li_2S_2 case were the binding energy is underestimated by .1 eV. This highlights the predictive power of the framework under investigation, which is generally able to identify structures offering the strongest binding. We note that when we include higher energy Li_2S_2 CH vertices at the relaxation stage, we identify a configuration with binding energy of -2.04 eV, in line with what reported in the literature. In this regards, the binding energy change trend among the various LiPSs is also in good agreement with our previous study [13], which adopts the exact same DFT calculation set up.

TABLE 1 Reported relaxation steps for each material. “Single-point (S-P) number” refers to the calculations required during the “Reference Force-Energy calculations” to train the ML model, the label “Vertexes” refers the total number of geometry optimizations performed, and “Total steps” refers to the sum of the geometry optimization steps for all vertexes

Material	S-P number	Vertexes	Total steps
Zn-N ₄ -C/Li ₂ S	537	9	3962
Zn-N ₄ -C/Li ₂ S ₂	496	9	3922
Zn-N ₄ -C/Li ₂ S ₄	505	8	3139
Zn-N ₄ -C/Li ₂ S ₆	404	7	8831
Zn-N ₄ -C/Li ₂ S ₈	408	8	7700

Figure 6 shows the top and side views of LiPS structures lowest energy structures. All relaxed LiPS structures have a S atom located on/near the top of the TM atom, while the Li atoms are located closer to the substrate. This is verified by an atomic distance analysis, reported in Figure 6, describing the shortest Li-TM and S-TM distances for each LiPS molecule corresponding to the lowest energy structure. The distance between Li atoms and the TM is always smaller than the one between the S atoms and the TM. The closest S-TM distance decreases for high order LiPSs, while the opposite is observed for the closest Li-TM distance. The tendency of the TM in “pulling” the S atoms closer while “pushing” the Li atoms away is in agreement with our previous report in the literature [13].

4 | WORKFLOW EXPLOITATION: ZN-N₄-C SAC

We move on to exploit the binding energy prediction workflow described in this study and identify the low-lying minima for the shuttle-effect cycle on a Zn-based SAC. The system configuration and geometry of the Zn-based SAC structures were designed in accordance to the previously described Fe-based SAC, and are illustrated in Figure 7. We examine the binding energies for five LiPSs (Li₂S_n, n = 1, 2, 4, 6, 8), and compare LiPS energetics and geometries against the one found for Fe-based SAC.

Figure 5 reports on the lowest identified binding energies for all examined structures lying on a GCH vertex. Although the Zn-N₄-C has not been extensively studied previously, the observed binding energies are in good agreement with previously reported values ($E_b^{Zn-N_4-C/Li_2S_6} = 1.02$ eV [9]). Atomic binding on Zn-based SAC seems stronger than on Fe-based SAC, by .45-.64 eV depending on the LiPS order. The presence of Zn rather than Fe in the SAC tends to affect more the lower order LiPSs comparing, a property that can be of help toward the reduction of the “shuttle” effect at the final catalytic steps of a LiPS. The shortest Li-TM distances behave similarly for either Zn and Fe-based SACs, although these are generally shorter when Zn is present. In contrast, the shortest reported S-TM distances for the two systems follow an opposite trend. When Zn is present the S-TM distance tends to increase as the LiPS order increases, and, generally, the distances are larger when comparing to the Fe-based SAC. Based on the above, we can conclude that Zn tends to prefer Li bonding for high order LiPSs. This can potentially improve the conversion rate at that stage.

As a next step, we compare the lowest found energy configurations for the two different SACs, as reported in Figures 6 and 7. We note that some of the LiPSs which end up in the most competitive minima find the same, or almost the same, initial vertex in the GCH, comparing the Fe and Zn SAC cases. For Li₂S, Li₂S₂, and Li₂S₈, where the initial unrelaxed configurations are not lying in the same vertex for the two different SACs, the final relaxed configurations are minimizing into similar configurations, as shown in Figures 6 and 7.

It is a matter of debate whether best LiPS configurations found for a given substrate could be translated to other systems of similar nature. Our results suggest that relaxing a LiPS/SAC configuration which resulted in lower binding energy for the case of a given TM, is likely, yet not always, to lead to favorable configurations when substituting the TM itself. More robust approaches to identify putative minimum energy configurations, like the one here proposed, are thus necessary toward more accurate predictions.

For completeness in the discussion on the computational costs associated with the discussed protocol, we display the number of DFT relaxation steps necessary in our workflow for each Zn-LiPS system in Table 1. In particular, we report the total number of relaxation steps required in this workflow, including the single-point calculations necessary during the “Reference Force-Energy calculations” step, and the structure relaxation calculations of all GCH vertexes during the “Minimize GCH Vertexes” step. From that table, it is clear that the computational complexity increases for the high-order LiPSs (Li₂S₆ and Li₂S₈) as the number of possible local energy minima increases for larger molecules. We recognize that the computational cost of DFT-based calculations widely depends on multiple system and software dependent factors, hence effective cross-reference comparison is not trivial. Additionally, the computational cost per DFT relaxation step varies when using self-consistent functionals, hence time comparison would not be immediately insightful.

The standard approach of manually, and often arbitrarily, constructing the LiPS/SAC configuration is known to be ineffective and often might not correspond to the most energetically favorable structure, unless excessive testing is performed. On one hand we cannot exclude that other minima searches methods, for example, via genetic algorithm approaches, will find minima structures competitive with those found in this study.

On the other we note that these more extensive searches, involving hundreds or thousands structural relaxations, rather than tens of structural relaxations as in the proposed protocol, would result in much larger computational costs and practically unusable. However, we argue that the workflow under examination proposes a good set of candidate structures and offers a fair balance between accuracy and computational efficiency.

5 | DISCUSSION AND CONCLUSIONS

The binding energy of LiPSs on TM–N–C substrates is a useful descriptor to infer which SAC material is likely to provide an optimal performance in preventing shuttle effects in Li–S batteries. Here, we propose an efficient protocol to identify a likely minimum energy configuration candidate, and thus the most informative binding energy prediction, via a combination of DFT single-point calculations, unsupervised and supervised data-driven algorithms, and DFT structure optimizations. The protocol hinges on: I) the construction of a database of roto-translation of LiPS molecules deposited on a rigid TM–N–C substrate, II) the calculation of the binding energy for a carefully selected number of these, III) the training of a kernel ridge regressor to chart structure-binding energy relationships systematically, IV) the use of a generalized convex hull construction to isolate likely-stable and structurally diverse geometries, V) their final optimization at DFT level.

We benchmark the protocol robustness for the case of LiPS on a Fe–N₄–C SAC and recover binding energies on-par or lower than the ones reported in the literature. We move to analyze the LiPS adsorption properties on a Zn–N₄–C SAC and find that the latter systematically displays a strong interaction (i.e. lower binding energies). From a structural perspective, we confirm that the minimum S–TM distances are always lower than the minimum Li–TM distances, for both systems. We further find that, regardless of the TM nature, an increase in the nuclearity of the LiPS molecule corresponds to a decrease in the S–TM minimum distance and an increase in the Li–TM minimum distance. Interestingly, we note that notwithstanding a change in the transition metal single atom catalysts, the lowest energy relaxed structure often, but not always, originates from a LiPS located and oriented in similar initial position and orientation w.r.t. the substrate TM.

Few upgrades could further benefit the efficiency and accuracy of the search strategy we propose. If deemed necessary, the GCH vertex search can encompass not the first PCA component but also high order ones. Further, adopting more complex statistical learning methods (e.g., Gaussian Process regression or Neural Network based) or other more recent local-density featurization approaches (e.g., Atomic Cluster Expansion based [52, 61, 62] approaches) could enable models which retain the same target accuracy but demand for a lesser number of training structures. Second, co-regionalized [63] or transfer learning [64] approaches may further benefit the making of more efficient and accurate models when tackling the study of LiPS adsorption on a substrates with many and diverse SACs transition metals. Third, ML approaches could be also adopted to accelerate the minimization [65–67] and reduce drastically the number of DFT force-and-energy calculations which are necessary at the “structure optimization” stage following the surrogate generalized convex hull construction. Currently, the proposed strategy does not aim to replace DFT-based calculations, but rather to guide and speed-up DFT-based searches of LiPS configurations for strong adsorption on SACs. The ML acceleration of the structural relaxation step could further diminish the need of DFT single-point calculations by a factor of 2 or larger.

Finally, we remark that the method is general and transferable to study not only other LiPS-substrate interactions, but, in principle, also other molecular system–substrate interactions. Here, we demonstrate our protocol's transferability by examining the Zn–N₄–C SAC, and observing consistent results as in the Fe–N₄–C SAC. Conversely we believe this work will nurture the advancement in minima searches protocols that investigate medium and large molecules adsorption on rigid and soft substrates of diverse nature.

From an application standpoint, the synthetic process and performance of single atom catalysts have witnessed ground-breaking advancement since the emergence of this class of materials, almost 15 year ago [68, 69]. We hope our work aids and contributes to raise interest in the integration of that category of materials in the next generation of Li–S-based energy storage device.

AUTHOR CONTRIBUTIONS

Eleftherios Andritsos: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; visualization; writing – original draft; writing – review and editing. **Kevin Rossi:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGMENTS

K.R. acknowledges funding from European Research Council (H2020 Marie Curie Individual Fellowship grant NANOCO2RE with agreement number 890414). We are grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1 and EP/T022213/1).

DATA AVAILABILITY STATEMENT

The databases, python notebooks, and CASTEP input files used in this publication for carrying out predictions are available at DOI: 10.24435/materialscloud:5a-vm. The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Eleftherios I. Andritsos  <https://orcid.org/0000-0002-3289-266X>

Kevin Rossi  <https://orcid.org/0000-0001-8428-5127>

REFERENCES

- [1] X. Ji, L. F. Nazar, *J. Mater. Chem.* **2010**, *20*, 9821.
- [2] X. Tao, J. Wan, C. Liu, H. Wang, H. Yao, G. Zheng, Z. W. Seh, Q. Cai, W. Li, G. Zhou, C. Zu, Y. Cui, *Nat. Commun.* **2016**, *7*, 1.
- [3] W. Ren, W. Ma, S. Zhang, B. Tang, *Energy Storage Mater.* **2019**, *23*, 707.
- [4] L. Zhou, D. L. Danilov, R.-A. Eichel, P. H. L. Notten, *Adv. Energy Mater.* **2020**, *11*, 2001304.
- [5] L. Zhang, P. Liang, X. lei Man, D. Wang, J. Huang, H. Bo Shu, Z. gang Liu, L. Wang, *J. Phys. Chem. Solids* **2019**, *126*, 280.
- [6] J. Wang, L. Jia, J. Zhong, Q. Xiao, C. Wang, K. Zang, H. Liu, H. Zheng, J. Luo, J. Yang, H. Fan, W. Duan, Y. Wu, H. Lin, Y. Zhang, *Energy Storage Mater.* **2019**, *18*, 246.
- [7] Q.-W. Zeng, R.-M. Hu, Z.-B. Chen, J.-X. Shang, *Mater. Res. Express* **2019**, *6*, 095620.
- [8] Z. Du, X. Chen, W. Hu, C. Chuang, S. Xie, A. Hu, W. Yan, X. Kong, X. Wu, H. Ji, L.-J. Wan, *J. Am. Chem. Soc.* **2019**, *141*, 3977.
- [9] G. Zhou, S. Zhao, T. Wang, S.-Z. Yang, B. Johannessen, H. Chen, C. Liu, Y. Ye, Y. Wu, Y. Peng, C. Liu, S. P. Jiang, Q. Zhang, Y. Cui, *Nano Lett.* **2020**, *20*, 1252.
- [10] C. Wang, H. Song, C. Yu, Z. Ullah, Z. Guan, R. Chu, Y. Zhang, L. Zhao, Q. Li, L. Liu, *J. Mater. Chem.* **2020**, *A8*, 3421.
- [11] Y. Li, J. Wu, B. Zhang, W. Wang, G. Zhang, Z. W. Seh, N. Zhang, J. Sun, L. Huang, J. Jiang, J. Zhou, Y. Sun, *Energy Storage Mater.* **2020**, *30*, 250.
- [12] Q. Shao, L. Xu, D. Guo, Y. Su, J. Chen, *J. Mater. Chem. A* **2020**, *8*, 23772.
- [13] E. I. Andritsos, C. Lekakou, Q. Cai, *J. Phys. Chem. C* **2021**, *125*, 18108.
- [14] S. P. Jand, Y. Chen, P. Kaghazchi, *J. Power Sources* **2016**, *308*, 166.
- [15] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S. P. Ong, *Adv. Energy Mater.* **2020**, *10*, 1903242.
- [16] Y. Liu, O. C. Esan, Z. Pan, L. An, *Energy AI* **2021**, *3*, 100049.
- [17] K. Rossi, G. Asara, F. Baletto, *ACS Catal.* **2020**, *10*(6), 3911.
- [18] K. Tran, Z. W. Ulissi, *Nat. Catal.* **2018**, *1*(9), 696.
- [19] G. H. Gu, P. Plechac, D. G. Vlachos, *React. Chem. Eng.* **2018**, *3*, 454.
- [20] A. Wang, Z. Zou, D. Wang, Y. Liu, Y. Li, J. Wu, M. Avdeev, S. Shi, *Energy Storage Mater.* **2021**, *35*, 595. <https://doi.org/10.1016/j.ensm.2020.10.022>
- [21] B. Liu, J. Yang, H. Yang, C. Ye, Y. Mao, J. Wang, S. Shi, J. Yang, W. Zhang, *J. Mater. Chem.* **2019**, *A7*, 19961. <https://doi.org/10.1039/c9ta06748e>
- [22] Q. Zhao, L. Zhang, B. He, A. Ye, M. Avdeev, L. Chen, S. Shi, *Energy Storage Mater.* **2021**, *40*, 386. <https://doi.org/10.1016/j.ensm.2021.05.033>
- [23] Q. Zhao, M. Avdeev, L. Chen, S. Shi, *Sci. Bull.* **2021**, *66*, 1401. <https://doi.org/10.1016/j.scib.2021.04.029>
- [24] Y. Liu, T. Zhao, W. Ju, S. Shi, S. Shi, S. Shi, *J. Materiomics* **2017**, *3*, 159.
- [25] Y. Liu, B. Guo, X. Zou, Y. Li, S. Shi, *Energy Storage Mater.* **2020**, *31*, 434.
- [26] V. L. Deringer, *J. Phys. Energy* **2020**, *2*, 041003. <https://doi.org/10.1088/2515-7655/abb011>
- [27] H. Guo, Q. Wang, A. Stuke, A. Urban, N. Artrith, *Front. Energy Res.* **2021**, *9*, 695902.
- [28] N. Artrith, *J. Phys. Energy* **2019**, *1*, 032002. <https://doi.org/10.1088/2515-7655/ab2060>
- [29] T. Lombardo, M. Duquesnoy, H. El-Boussidy, F. Árén, A. Gallo-Bueno, P. B. Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide, M. Reynaud, J. Carrasco, A. Grimaud, C. Zhang, T. Vegge, P. Johansson, A. A. Franco, *Chem. Rev.* **2021**. <https://doi.org/10.1021/acs.chemrev.1c00108>
- [30] J. B. Robinson, K. Xi, R. V. Kumar, A. C. Ferrari, H. Au, M.-M. Titirici, A. Parra-Puerto, A. Kucernak, S. D. S. Fitch, N. Garcia-Araez, Z. L. Brown, M. Pasta, L. Furness, A. J. Kibler, D. A. Walsh, L. R. Johnson, C. Holc, G. N. Newton, N. R. Champness, F. Markoulidis, C. Crean, R. C. T. Slade, E. I. Andritsos, Q. Cai, S. Babar, T. Zhang, C. Lekakou, N. Kulkarni, A. J. E. Rettie, R. Jervis, M. Cornish, M. Marinescu, G. Offer, Z. Li, L. Bird, C. P. Grey, M. Chowalla, D. D. Lecce, R. E. Owen, T. S. Miller, D. J. L. Brett, S. Liatard, D. Ainsworth, P. R. Shearing, *J. Phys. Energy* **2021**, *3*, 031501.
- [31] Y. Wu, J. Jiang, Z. Weng, M. Wang, D. L. J. Broere, Y. Zhong, G. W. Brudvig, Z. Feng, H. Wang, *ACS Cent. Sci.* **2017**, *3*, 847.
- [32] J. Xu, W. Zhang, H. Fan, F. Cheng, D. Su, G. Wang, *Nano Energy* **2018**, *51*, 73.
- [33] S. Clark, F. L. Bleken, S. Stier, E. Flores, C. W. Andersen, M. Marcinek, A. Szczesna-Chrzan, M. Gaberscek, M. R. Palacin, M. Uhrin, J. Friis, *Adv. Energy Mater.* **2021**, *12*, 2102702. <https://doi.org/10.1002/aenm.202102702>
- [34] I. E. Castelli, D. J. Arismendi-Arrieta, A. Bhowmik, I. Cekic-Laskovic, S. Clark, R. Dominko, E. Flores, J. Flowers, K. U. Frederiksen, J. Friis, A. Grimaud, K. V. Hansen, L. J. Hardwick, K. Hermansson, L. Königer, H. Lauritzen, F. Le Cras, H. Li, S. Lyonnard, H. Lorrman, N. Marzari, L. Niedzicki, G. Pizzi, F. Rahmanian, H. Stein, M. Uhrin, W. Wenzel, M. Winter, C. Wölke, T. Vegge, *Batter. Supercaps* **2021**, *4*, 1803. <https://doi.org/10.1002/batt.202100117>
- [35] A. Kilic, R. Çağla Odabaşı, Yildirim, D. Eroglu, *Chem. Eng. J.* **2020**, *390*, 124117.
- [36] H. Zhang, Z. Wang, J. Ren, J. Liu, J. Li, *Energy Storage Mater.* **2021**, *35*, 88.
- [37] Z. Lian, M. Yang, F. Jan, B. Li, *J. Phys. Chem. Lett.* **2021**, *12*, 7053.
- [38] Z. Ma, Z. Zuo, Y. Li, *ACS Appl. Mater. Interfaces* **2021**, *13*, 23936.
- [39] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115.
- [40] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, M. Ceriotti, *J. Chem. Phys.* **2018**, *148*, 241730. <https://doi.org/10.1063/1.5024611>
- [41] V. Vovk, Kernel ridge regression. in *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, Springer, Berlin, Heidelberg **2013**, p. 105.
- [42] K. Pearson, *London Edinburgh Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559.
- [43] A. Anelli, E. A. Engel, C. J. Pickard, M. Ceriotti, *Phys. Rev. Mater.* **2018**, *2*, 103804.
- [44] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, M. C. Payne, *Z. Kristallogr. Cryst. Mater.* **2005**, *220*, 567.
- [45] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.
- [46] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, A. S. Foster, *Comput. Phys. Commun.* **2020**, *247*, 106949.
- [47] S. De, A. P. Bartók, G. Csányi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754.
- [48] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.* **2017**, *3*(12), e1701816.
- [49] V. L. Deringer, M. A. Caro, G. Csányi, *Nat. Commun.* **2020**, *11*(1), 1.

- [50] C. Zeni, K. Rossi, A. Glielmo, S. De Gironcoli, *J. Chem. Phys.* **2021**, *154*(22), 224112.
- [51] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglu, S. DeGironcoli, R. E. Palmer, F. Baletto, *Nat. Commun.* **2021**, *12*, 6065.
- [52] Y. Lysogorskiy, C. van der Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, R. Drautz, *npj Comput. Mater.* **2021**, *7*, 97. <https://doi.org/10.1038/s41524-021-00559-9>
- [53] D. P. Kovács, C. van der Oord, J. Kucera, A. E. A. Allen, D. J. Cole, C. Ortner, G. Csányi, *J. Chem. Theory Comput.* **2021**, *17*, 7696.
- [54] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, S. P. Ong, *J. Phys. Chem. A* **2020**, *124*, 731.
- [55] K. Rossi, J. Cumby, *Int. J. Quantum Chem.* **2020**, *120*, e26151.
- [56] E. A. Engel, A. Anelli, M. Ceriotti, C. J. Pickard, R. J. Needs, *Nat. Commun.* **2018**, *9*, 2173.
- [57] K. Konstantinou, F. C. Mocanu, T. H. Lee, S. R. Elliott, *Nat. Commun.* **2019**, *10*, 3065. <https://doi.org/10.1038/s41467-019-10980-w>
- [58] N. Bernstein, G. Csányi, V. L. Deringer, *npj Comput. Mater.* **2019**, *5*, 99. <https://doi.org/10.1038/s41524-019-0236-6>
- [59] C. Rasmussen, C. Williams, M. Press, F. Bach, P. Fierro, Gaussian processes for machine learning. in *Adaptive Computation and Machine Learning*, MIT Press, Cambridge MA, USA **2006**.
- [60] M. Stone, *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*(2), 111.
- [61] R. Drautz, *Phys. Rev. B* **2020**, *102*, 024104.
- [62] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglou, S. de Gironcoli, R. E. Palmer, F. Baletto, *Nat. Commun.* **2021**, *12*(1), 1.
- [63] E. V. Bonilla, K. M. A. Chai, C. K. Williams, *NeurIPS Proc.* **2009**, 20.
- [64] J. West, D. Ventura, S. Warnick, *Spring Research Presentation: A Theoretical Foundation for Inductive Transfer*, Brigham Young University, College of Physical and Mathematical Sciences, Provo UT, USA **2007**.
- [65] T. L. Jacobsen, M. S. Jørgensen, B. Hammer, *Phys. Rev. Lett.* **2018**, *120*, 026102. <https://doi.org/10.1103/PhysRevLett.120.026102>
- [66] E. G. D. Río, J. J. Mortensen, K. W. Jacobsen, *Phys. Rev. B* **2019**, *100*, 104103. <https://doi.org/10.1103/PhysRevB.100.104103>
- [67] Y. Yang, O. A. Jiménez-Negrón, J. R. Kitchin, *J. Chem. Phys.* **2021**, *154*, 234704. <https://doi.org/10.1063/5.0049665>
- [68] S. Mitchell, J. Pérez-Ramírez, *Nat. Commun.* **2020**, *11*, 4302.
- [69] H.-Y. Zhuo, X. Zhang, J.-X. Liang, Q. Yu, H. Xiao, J. Li, *Chem. Rev.* **2020**, *120*, 12315.

How to cite this article: E. I. Andritsos, K. Rossi, *Int. J. Quantum Chem.* **2022**, *122*(17), e26956. <https://doi.org/10.1002/qua.26956>