Energy Storage Materials 44 (2022) 313–325

Contents lists available at ScienceDirect

Energy Storage Materials

journal homepage: www.elsevier.com/locate/ensm

Artificial intelligence driven in-silico discovery of novel organic lithium-ion battery cathodes

Rodrigo P. Carvalho[a,b], Cleber F.N. Marchiori[c], Daniel Brandell[b], C. Moyses Araujo[a,c,*]

[a] Materials Theory Division, Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden
[b] Department of Chemistry - Ångström Laboratory, Uppsala University, Box 538, 75121 Uppsala, Sweden
[c] Department of Engineering and Physics, Karlstad University, 65188 Karlstad, Sweden

ARTICLE INFO

Keywords:
Battery materials
Organic electrode
Li-ion battery
Artificial intelligence
Machine learning
Materials discovery

ABSTRACT

Organic electrode materials (OEMs) combine key sustainability and versatility properties with the potential to enable the realisation of the next generation of truly green battery technologies. However, for OEMs to become a competitive alternative, challenging issues related to energy density, rate capability and cycling stability need to be overcome. In this work, we have developed and applied an alternative yet systematic methodology to accelerate the discovery of suitable cathode-active OEMs by interplaying artificial intelligence (AI) and quantum mechanics. This AI-kernel has allowed a high-throughput screening of a huge library of organic molecules, leading to the discovery of 459 novel promising OEMs with candidates offering the potential to achieve theoretical energy densities superior to 1000 W h kg^{-1}. Moreover, the machinery accurately identified common molecular functionalities that lead to such higher-voltage electrodes and pointed out an interesting donor-accepter-like effect that may drive the future design of cathode-active OEMs.

1. Introduction

It is an indisputable truth that the design of energy-efficient and environmentally friendly technology is one of the most challenging scientific topics of our century. On top of that, energy consumption on our planet is expected to face a drastic increase in the coming years, imposing a strong demand for novel energy supplies that need to be secure, clean and sustainable [1]. Moving toward more environmentally friendly options, organic materials are emerging as a promising alternative for the next generation of energy-related technologies. Ranging from energy harvesting [2] to electrical energy storage [3] (EES), organics present a combination of attractive features [4] like low cost, versatile synthesis routes, lightweight, tailorable properties and production from renewable sources [5,6]. Therefore, the proper design of novel organic materials with enhanced properties is of utmost importance for a sustainable development. Specially connected to recent debates on grid load-levelling and the revolution in portable devices, several organic electrode materials (OEMs) have been proposed as potential electrodes in Li-ion and Na-ion batteries with competitive performances [7–12]. Although these materials still do not meet all requirements on either (volumetric) energy density, lifetime, rate capability and cycling stability, they are indeed attracting attention as an alternative to the inorganic counterparts in conventional batteries, with the aim of dramatically reducing related environmental impacts.

The performance of the organic materials depends heavily on the type of electrochemical reactions at work during the battery cycling. These materials can, generally, be grouped as *n*-, p- or bipolar-type depending on their charge states in the redox reactions [13]. For instance, n-type redox units will change reversibly between the negatively charged and neutral states while the p-type compounds change between neutral and positively charged states. The latter are commonly applied as battery cathodes due to their higher redox potentials in the range 3.0–4.0 V vs. Li/Li^+. This class of materials is mainly dominated by polymeric compounds [7,14-16], which display suitable combinations of energy density and stability. However, p-type materials require the design of dual-ion batteries, and thereby going well beyond the current technologies. On the other hand, n-type materials are better compatible with the standard Li-ion battery architectures and display a wider range of redox potentials that can be applied as both anodes and cathodes. The main hurdle here, however, lies on the fact that out of almost two hundred known organic materials employed as electrodes [17–23], including polymers, macromolecular assemblies and organic crystals comprising low-M_w molecular salts, just a few examples may serve as cathode-active materials. Instead, potential anode materials dominate. The existing cathode materials are dominated by quinone based chemistries with redox potentials in the range of 2.5–3.5 vs. Li/Li^+ [19,24-26]. Recently this chemical space has been expanded by including a new class of con-

* Corresponding author.
E-mail address: Moyses.Araujo@kau.se (C.M. Araujo).

https://doi.org/10.1016/j.ensm.2021.10.029
Received 26 July 2021; Received in revised form 14 October 2021; Accepted 19 October 2021
Available online 24 October 2021

jugated sulfonamides, which has shown to display superior air-stability and redox potentials in the range 2.85–3.45 V vs. Li/Li$^+$. [23]

This illustrates how challenging it is to achieve organic lithium reservoirs (n-type) with high redox potentials to be applied as positive electrodes, which in turn is imperative to realize organic batteries with higher energy densities. In fact, rational design strategies need to be developed to tap into the almost limitless chemical universe available for organic compounds and accelerate the discovery of such materials. To give an idea of the extension of such a universe, 166 billion organic moieties have recently been enumerated by considering only small molecules with up to 17 non-hydrogen atoms like C, N, O, S and halogens [27]. Thus, there is an immeasurable number of possible materials yet to be discovered. The synthesis and characterization of such a huge amount of materials, to fully explore all possibilities, would clearly be unfeasible. At the same time, conventional computational approaches based on quantum mechanics also suffers from technical limitations due to the computationally demanding task of solving the associated equations.

An alternative pathway has recently been getting attention to potentially accelerate the discovery of novel materials within the framework of Machine Learning (ML) techniques. ML has proven to be a useful tool in proposing novelties for batteries [28–32], predicting band-gap of semiconductors [33–36] and general properties of small molecules [37–40], accelerating crystal structure prediction [41] and molecular dynamics [42–44], predicting new perovskites [45] and even helping to solve quantum mechanics equations [46]. More connected with lithium-ion batteries, recent works have highlighted how ML can be useful in obtaining insights to optimise the battery's operation and guide the rational design of electrodes and electrolytes [47–49]. However, some approaches [29,32] still rely heavily on information that needs to be extracted from expensive quantum mechanical calculations, which may be a hindrance in the materials discovery process. In this context, exploring the chemical universe presented by organics through fast techniques like machine learning may aid an intelligent design of OEMs. To successfully achieve this goal, some requirements need to be addressed, e.g., knowledge of the solid-state structure of these molecular systems and their electrochemical properties, unveiling the structure-property relationship to be accessed by the ML framework.

In this work, we have developed an efficient and elegant workflow combining density functional theory (DFT) and machine learning to accelerate the discovery of novel organic electroactive materials. The built-up tools compose a package for materials design aided by artificial intelligence (AI) that is capable of directly predicting the Li-ion insertion voltages, completely by-passing the demanding computational efforts from traditional quantum mechanics-based methods. The AI-kernel relies solely on structural information of the active molecule as input, encoded here as a simple structure-string, to assess their reduction potentials and the expected battery open-circuit voltage (vs. Li/Li$^+$). Thereafter, this machinery has been employed in high-throughput screening 20 million organic moieties from the GDB17 database [27] in order to identify new cathode-active compounds. Following a 2-tier filtering system of voltages and capacities, we have encountered 1001 n-type cathode candidates. This amount represents only 0.0050% of the total screened moieties, which further demonstrates the challenges in finding n-type organic positive electrodes. The molecular properties of the selected candidates have been re-calculated using high-level DFT methods as a final part of the high-throughput screening and with a threefold objective: (i) evaluate the accuracy of the AI-kernel; (ii) shorten the candidates list by removing possible unphysical results due to statistical noise and (iii) feedback the molecular database. From this step, the cathode candidates have been further narrowed down to a shorter list of 459 promising compounds. A closer look at the selected molecules revealed an interesting donor-acceptor-like configuration that may be explored in future tailoring the electronic structure of novel electrodes. A list of the most prominent molecules with their respective predicted lithiation voltages is, thereby, proposed with the potential to offer theoretical energy densities higher than 1000 W h kg^{-1}.

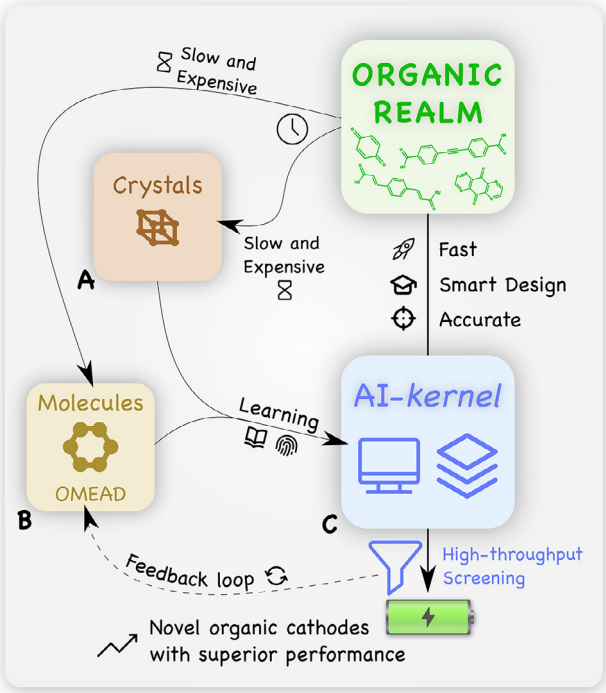


Fig. 1. Flowchart illustrating the whole workflow of the developed framework and how the AI-kernel enables a fast access to the world of organic materials after the learning step. OMEAD stands for "Organic Materials for Energy Applications Database".

2. Methods

The developed framework is divided into three main steps. In the first step (A), the crystal structures for a limited set of 28 electrode-candidates and their corresponding lithiated phases have been resolved by combining DFT and an evolutionary algorithm [50–52]. This methodology has been successfully implemented in some of our previous studies, leading to good agreements with the experimental outcomes [50–52]. The purpose of this step is to achieve an accurate description of the material's electronic structure and electrochemical lithiation mechanism. As a second step (B), we have developed a database containing structural information and properties of 26,218 organic molecules extracted from high-level DFT calculations. Most of the organic moieties that have been recently proposed for energy conversion and storage applications have been included, plus an additional set extracted from the database GDB17, proposed by Ruddigkeit et al. [27]. The purpose of this step is to access a significantly larger chemical space than the one available in step A. Nevertheless, this is still far from covering all the possible organic molecular configurations and a relatively small sampling space when aiming to find novel cathodic materials. Therefore, in the third step (C) we have developed models based on machine learning methodologies to significantly speed up the assessment of the OEMs' electrochemical properties. By combining data from A and B, an efficient AI-kernel with good statistical fidelity has been designed, which relies only on the knowledge of molecular structure as input to predict the battery open-circuit voltages, completely by-passing the time demanding ab-initio calculations. This kernel enabled a high-throughput screening of a large materials library (millions of molecules). This workflow is illustrated in Fig. 1, summarising the final design pathway followed by

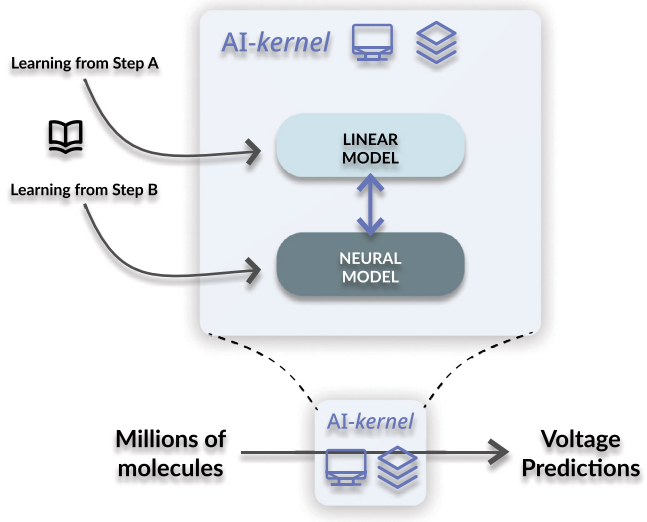


Fig. 2. A schematic of the AI-kernel operational principle.

the AI-kernel in which molecular units can now be extracted from the huge organic realm while simultaneously being quickly and accurately evaluated.

2.1. Step A – electrode crystal structure prediction

The crystal structure prediction was performed using an interplay between DFT calculations and a so called 'evolutionary algorithm'; the former as implemented in the Vienna Ab-initio Simulation Package (VASP) [53–55] and the latter as implemented in the USPEX code [56–59]. The evolutionary algorithm is capable of predicting the crystal structure of a compound by evolving a set of possible initial structure candidates through a successive process of common genetic operations, such as heredity, mutation and permutation. The heredity, as in nature, is responsible to create a new structure by merging two parenting candidates amongst the best-fitted group; mutation transforms a selected aspect of a single parenting structure to create a new one, e.g., randomly displacing an atom or applying a strain on the lattice; permutation exchanges different group of atoms or molecules. The initial set of candidates is randomly generated by combining the given units – molecules and Li atoms in this work – following different space group symmetries. Afterwards, these structures are optimised through a sequential batch of DFT calculations performed in VASP by using the projector-augmented wave [53] method and the Perdew, Burke, and Ernzerhof (GGA-PBE) [60] as exchange-correlation functional. These calculations have been done with increasing cutoff energy (400 - 550 eV) and k-space sampling resolution (0.12 - 0.08 2π $\AA^{-1}$), rendering structures to their respective local minima. The genetic operations are then carried out, creating a new generation of most suited samples. This process is repeated until a total number of generations is achieved or a proper convergence of the fitting criteria, which is the total energy in our case. In the end, the best structure goes through a new geometry optimisation with a higher cutoff energy of 600 eV and a k-mesh of $6 \times 6 \times 6$ keeping the same GGA-PBE functional. To remedy the overestimation of the electron delocalization generated by the GGA scheme, a final calculation is performed with the Heyd – Scuseria – Ernzerhof (HSE06) [61] hybrid functional while keeping the same cutoff energy and a new k-mesh of $4 \times 4 \times 4$. The whole refining process described here has been applied for all the molecules presented on Scheme 1 and their respective first two lithiated phases, resulting in a small database of predicted organic crystals.

For all ab-initio calculations, the Grimme-D2 methodology [62] for dispersion interactions were considered, except for lithium atoms as the increasing number of this specie on lithiated phases can lead to an accumulation of errors due to intrinsic limitations of the DFT-D theory. Henceforth, all results related to the solid-state structures presented in this work refer to the framework explained above, with the final electronic properties resulting from the HSE06 step. The lithiation mechanism can be represented by the following chemical reaction, referenced to the lithium metal electrode:

$$Li_{x_0}H + (x_1 - x_0)Li \rightleftharpoons Li_{x_1}H \quad (1)$$

with the voltage profile of such electrochemical process computed through the corresponding reaction free energy by the Nernst equation, such as:

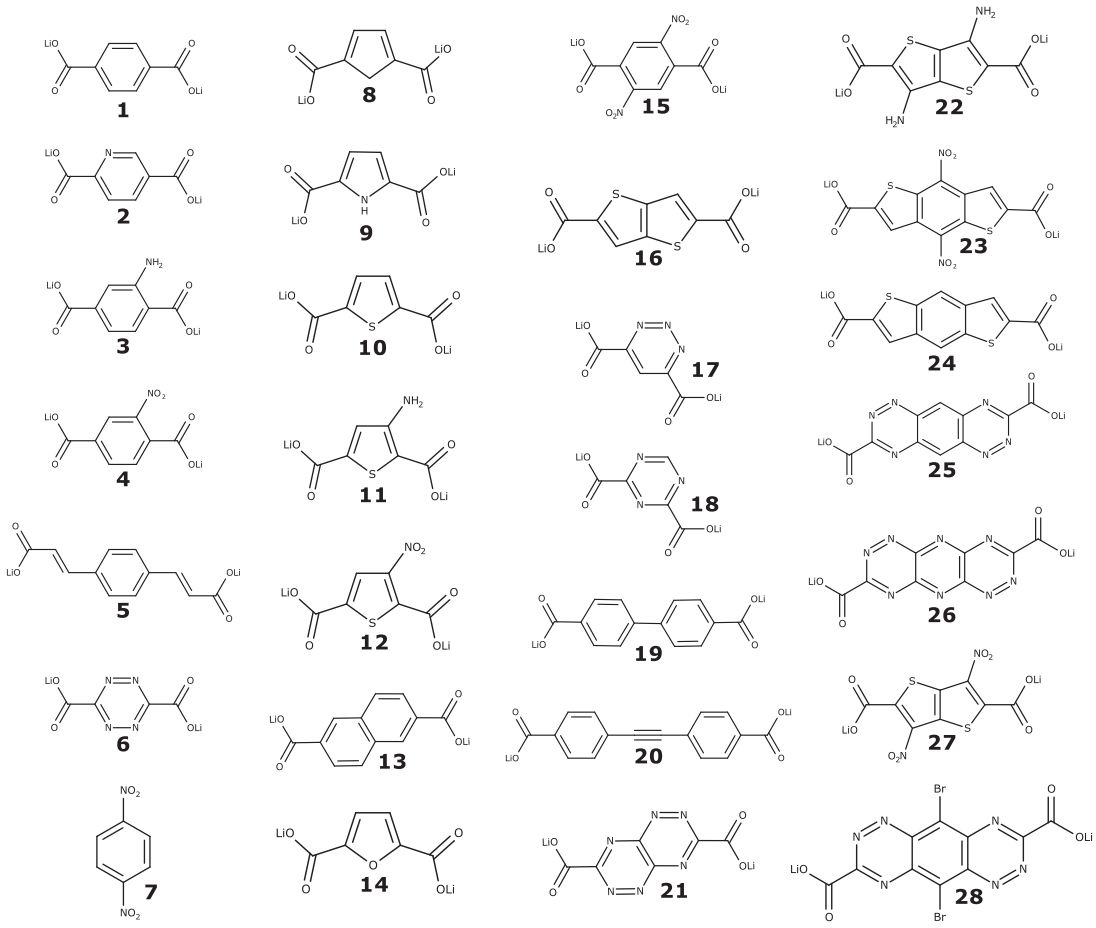
$$V(x) = \frac{-\Delta G_r}{nF} = -\frac{E\left(Li_{x_1}\mathcal{H}\right) - E\left(Li_{x_0}\mathcal{H}\right) - (x_1 - x_0)E(Li)}{x_1 - x_0} \quad (2)$$

Here, the Gibbs free energy is approximated by the electronic total energy (E), as the latter is expected to be the dominant term of the reaction free energy. From this step, the open-circuit voltage (V_{OC}) for the Li-ion insertion reaction as referenced to the Li/Li^+ electrode was obtained for all OEMs. Such methodology has given rise to excellent V_{OC} estimation of several organic and inorganic battery materials [63,64].

The employed combination of an evolutionary algorithm and high-quality DFT calculations render considerably higher-precision estimations of the electrochemical properties as compared to conventionally employed gas-phase calculations. The former approach primarily offers a better representation of the material's chemical environment during the ion insertion reaction. Thereby, it renders improvements in the description of several important aspects of the reaction thermodynamics, such as the inserted-ion coordination shell, changes in the material's crystal structure and electronic structure, etc. In a previous work, for instance, it was demonstrated that the use of a single molecule was not enough to describe the lithiation potential of the dilithium thiophene dicarboxylate, with a discrepancy higher than 50% when comparing with experimental findings [50]. However, the discrepancy drops to 3% when employing this evolutionary algorithm framework [50]. Furthermore, this approach has been successfully applied to predict the crystal structure and Li-ion insertion thermodynamics of the dilithium benzene diacrylate with a good match with the experimentally resolved structure [52].

2.2. Step B – organic materials for energy applications database

We have developed a new database composed of more than 26,000 unique organic molecules together with a list of properties that are relevant for various practical applications, and in particular for EES. These include the HOMO-LUMO energies, dielectric constants, redox potentials, electronic affinities and ionization potentials – all extracted from ab-initio calculations within the DFT framework. This molecular database, here named "The Organic Materials for Energy Applications Database (OMEAD)" (available in the Supplementary Material), is formed by an initial selection of molecules and polymers from a range of energy-related applications – energy harvesting, electrodes for energy storage, electrolytes, light absorbers, to cite a few – and merged with a second set of elements extracted from the database GDB17, composed by small organic molecules with up to 17 non-hydrogen atoms of C, O, N, S, F, Cl and Br. Consequently, this approach enables the development of a richer database in terms of chemical and structural diversity than those possible to cover in Step A. Molecules shown in Scheme 1 are also part of the OMEAD. All DFT calculations in this stage were carried out using the Gaussian 16 [65] software package with the B3LYP [66,67] functional and the 6–31G(d)/6–311G(d,p) Pople basis set [68] theory level while keeping molecules in the gas phase. A detailed description of how the molecular properties have been obtained is provided in the Supplementary Material. The development of this database and all the molecular properties included in it required more than 310 000 DFT calculations.



Scheme 1. Lewis representations of the moieties composing the small dataset of predicted crystal structures (step A). In addition, they are also part of the molecular database (step B).

Although a number of organic molecules databases are available in the literature, none of them presents a systematic calculation of the properties we aimed to cover in this study, and where special attention for energy applications is needed.

2.3. Step C – the AI-kernel

In this step, two distinct statistical-learning models have been developed: a linear and a neural network model. It is the combination of these two that define the AI-kernel. The models exchange information inside the kernel to yield the final prediction. This process is further explained in the following, starting from the isolated functionality of the models.

Linear model: We have statistically tested the following linear model

$$V_{OC} = \alpha X + \beta, \quad (3)$$

where V_{OC} is the open-circuit voltage of the lithiation reaction in the solid-state phase. The independent variable X is a column-vector containing the molecular properties of the corresponding electrode materials' building blocks. Here, we have tested orbital energies, oxidation potentials (P_{Ox}), reduction potentials (P_{Red}), dielectric constants, etc., and their combinations; properties that are included in the OMEAD database. More information can be found in Table S2. The row-vector α and the bias β represent the coefficients obtained through ordinary least square (OLS) regression. In this approach, we are searching for the correlation between the solid-state (crystal) and molecular (gas phase) properties of the organic electrode materials. The properties that display the best correlations are referred to as molecular descriptors, which will facilitate the screening in a much wider materials library as they can be assessed at lower computational costs. It should be pointed that although our database OMEAD contains the relevant properties of 26,218 molecules, the OLS has been carried on the narrower space of 57 compounds for which we have more accurate V_{OC} values. Of these, 29 molecules have been added from experimental results reported in literature (see Supplementary Material for more information) and 28 are those in Scheme 1. As described in step A, the latter require computationally demanding crystal structures predictions.

This exploratory investigation resulted in the reduction potential as the best descriptor, i.e., the property which could best explain V_{OC}. The choice of such potential does not come as a total surprise, especially when considering n-type materials. Alternatively, a principal component analysis (PCA) was also employed to merge the P_{Ox} and P_{Red} potentials in a single principal component (PC). This PC could be further correlated to the lithiation voltage through a new linear regression, achieving similar precision as the OLS. This PCA is shown in the Supplementary Material and is presented as a secondary option for the Linear model. In this study, we have calculated these potentials not only using the reaction enthalpy but also including the full Gibbs free energy.

Neural model: We have developed an Artificial Neural Network based model aiming at predicting the molecular descriptors in a fast and accurate manner, by-passing the computing demanding DFT calculations and only using the molecular structure as input. The training of this model has been carried on top of our database, OMEAD.

Artificial Neural Networks (NN) can be classified as a non-linear learning algorithm inspired by the human brain architecture that is capable of representing any function with a high level of fidelity. In general, it is formed by an arbitrary number of neurons (or nodes) organized in layers, which in turn are interconnected through a set of arithmetic operations between the output of a neuron and a fittable parameter called weight. When each neuron in a given layer is connected to all neurons in the next layer, we have a so-called Fully Connected (or Dense) Neural Network (FCNN or FC-layers), resulting in a set of weights to be fitted. On the other hand, when a set of neurons in a given layer is connected to only one neuron in the next layer, sharing a unique weight parameter, the result is a Convolutional Neural Network (CNN) [69]. Another alternative is the Recurrent Neural Network (RNN) [70,71], in which the input data is seen as an organised sequence of information and the RNN layer tries to understand its sequential meaning. To achieve this, the architecture keeps a hidden internal state vector that, together with the weight parameters, are responsible to uncover the knowledge behind the data ordering. In all cases, a bias term can be added to each layer and an activation function (tanh, sigmoid function, ReLU, SeLU, etc.) [72,73] is applied to the layer's output before it can be forwarded to the next layer to incorporate non-linear characteristics to the model. The general representation of a simple Neural Network is expressed in Eq. (4), where l is any given layer, σ is the activation function, $W^{(l)}$ is a matrix containing the weight parameters linking all neurons of a previous layer $(l-1)$ with the corresponding neurons of the layer l and $b^{(l)}$ are the biases.

$$z^{(l)} = \sigma\left(W^{(l)} z^{(l-1)} + b^{(l)}\right) \tag{4}$$

The 0-layer is referred to as the input layer $z^{(0)}$, where the actual data will feed the network. The input data, in turn, gives rise to a new question to be answered in machine learning methods: how should the information be provided to the machine in a meaningful way? Furthermore, the right choice of input data can be a key factor to correctly describe the desired structure-property relationship. Recently, several works have been trying to answer this question [74–76], presenting a variety of representation models (or fingerprints) for atoms, molecules and crystals.

In this present work, we have developed a methodology based on the Simplified Molecular-Input Line-Entry System (SMILES) as a simple text-based fingerprint and a mixed network architecture containing RNN and FCNN layers. The use of SMILES as a text string in this approach requires a further processing step through the so-called Natural Language Processing (NLP). In this auxiliary encoding system, the SMILES datasets are segmented into smaller pieces to form a unique vocabulary. For example, the string Cc1ccccn1, describing 2-methylpyridine, can be fragmented into [C, c, 1, c, c, c, c, n, 1], which result in a vocabulary of [1, C, c, n] unique entries. The final vocabulary must contain every unique fragment for all the molecules in the dataset. Afterwards, the SMILES string is translated into an index-based language, respecting the proper fragment ordering, where each index is a unique identification of the vocabulary element. The methylpyridine SMILES would be translated, for instance, to [2, 3, 1, 3, 3, 3, 3, 4, 1], by using the limited [1, C, c, n] vocabulary. Finally, an embedding system is employed to transform each vocabulary element into a new unique n-dimensional vector of scores, named embeddings. These scores represent an unknown characteristic that will be adjusted during the training phase in such a way that the neural network is responsible to learn the meaning of each vocabulary entry and how it is connected to the desirable property. In other words, the machine will develop a particular understanding of each element composing the SMILES, such as atoms and bonds, and how they relate to the molecular descriptor (P_{Red}). A carbon ('C') may be represented in this system by the unique vector $[e_1^{(C)}, e_2^{(C)}, \ldots, e_n^{(C)}]$, where e_i are the scores to be fitted during the learning process. This step can also be seen as part of an encoding step in which the Network is creating an internal representation of the molecules. A decoding process then follows in the next neural layers to provide the desired molecular descriptors. This use of the SMILES-representation differs from other recent proposals [77] in the literature where the SMILES has been employed to predict other molecules, i.e., new SMILES strings in a sequence-to-sequence architecture [78] or to obtain topological properties by analysing the SMILES to compose a new fingerprint set [79]. Additionally, the Coulomb Matrix [40] (CM) and the Many-Body Tensor Representation (MBTR) [80] have also been investigated. These three fingerprints have been benchmarked for different combinations of FCNNs, CNNs and RNNs. The details about this investigative exploration are depicted in the Supplementary Material.

It is important to point out that the different representations (CM, MBTR and SMILES) and a collection of processing tools for machine learning have been implemented in a new software package named ANIMA (AdvaNced artificial Intelligence for Materials reseArch). The package is written in the python language and published under an open-source basis (more details in the Supplementary Material). ANIMA and the developed databased compose a novel platform for the design of OEMs and will be freely distributed to the scientific community.

AI-kernel: The AI works by combining the Linear and Neural models in a two-step process where the information is exchanged inside the kernel. First, the SMILES is read and processed by the Neural model, which in its turn predicts the molecular descriptor P_{Red}, thereafter providing them to the Linear model to yield the V_{OC}. All these values can be accessed through the kernel.

High-throughput: After training, the machinery is employed in a high-throughput screening of 20 million molecules following a 2-tier filtering system of voltages → capacities to identify novel organic cathode materials. The voltages (V_{OC}) are evaluated in a fast way by the AI-kernel while the theoretical capacities are estimated by using the Eq. (5) alongside a straightforward analysis of the compounds' SMILES. The maximum amount of inserted Li^+ was asserted with typical redox reactions of specific sites, e.g., reduction of double bonding oxygen or nitrogen and heteroatoms such as sulphur.

$$C = \frac{nF}{3.6M} \tag{5}$$

In Eq. (5), n is the number of redox sites, F the Faraday constant and M the molecule molar mass. As a part of this screening step, a new DFT calculation is performed with the selected candidates, acting as a final filtering process. Moreover, it also serves to evaluate the AI-kernel performance and to feedback the OMEAD.

4. Results and discussion

Following the crystal structure prediction for the molecules presented in Scheme 1 and their respective first two lithiated phases, the average lithiation voltage for a two-step reaction – hereafter simply referred to as lithiation voltage or the open-circuit voltage V_{OC} – can be straightforwardly calculated from Eq. (2). The delithiated structures for some of these compounds are presented in Fig. 3(a). The crystal structure data for all of these compounds and their respective lithiated phases are available in the Supplementary Material. Several molecules in this dataset are based on dicarboxylates due to the fact that they initially form stable crystals, as noticeable from Fig. 3(a). In addition, the dicarboxylate-based building blocks may be further tailored by different mechanisms, thus offering tunable thermodynamic properties. A common signature of these crystals is the formation of a salt layer intercalated by their organic counterpart. In this layer, the Li-ions are often surrounded by four carboxylate oxygens forming tetrahedron coordination. This characteristic contributes significantly to the general stability of these type of organic electrodes, a desirable property for LIBs. Another interesting feature is the presence of cyclic structures, in which different heteroatoms and functional groups may be included, thus allowing a proper engineering of the redox chemistry. In this sense, a quick overview of Scheme 1 exemplifies the different strategies followed in this work with the inclusion of heteroatoms such as N, S and O or functional groups like NO_2 or NH_2. In Fig. 3(b), the V_{OC} (vs Li/Li^+)

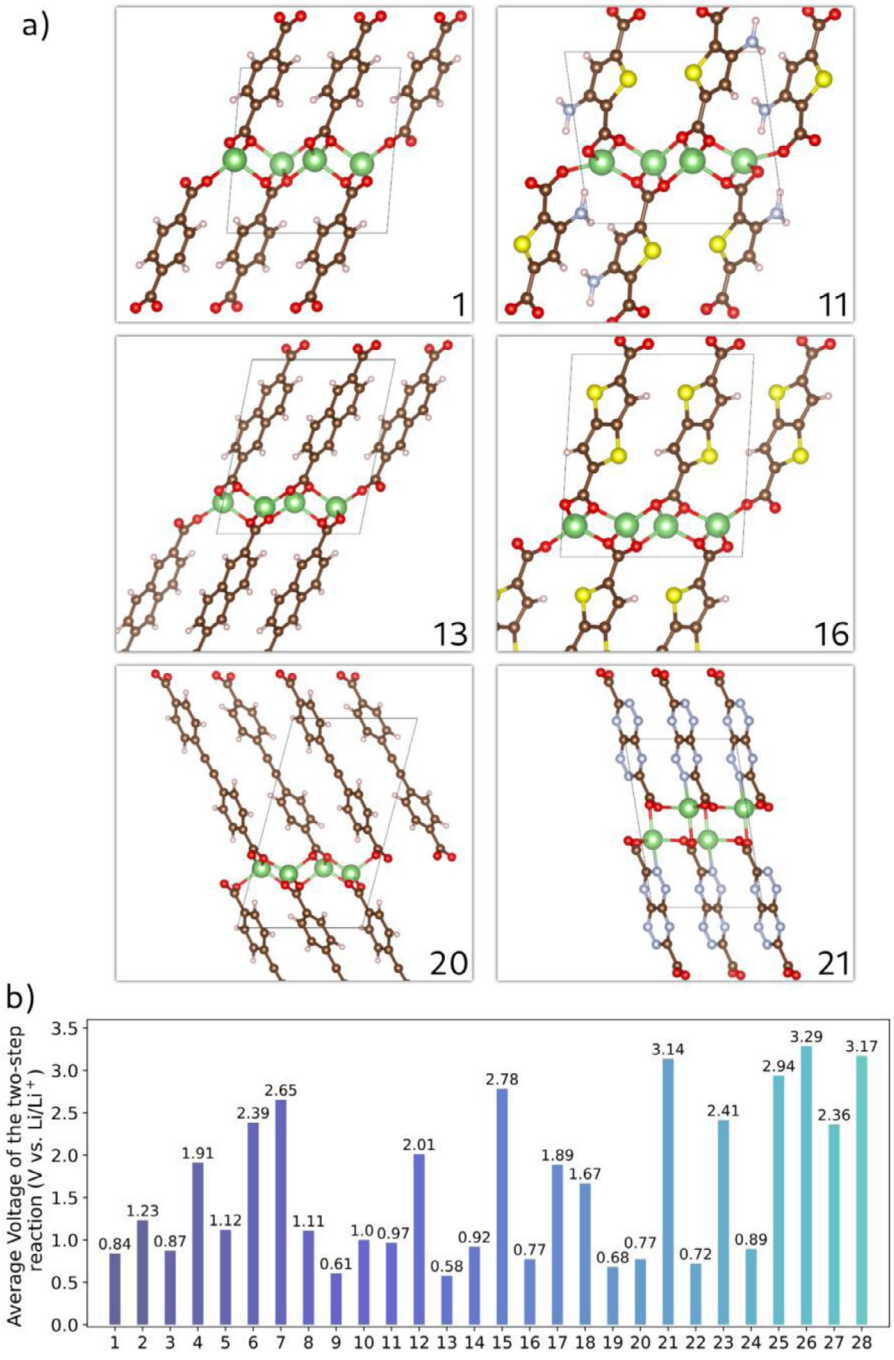


Fig. 3. a) Crystal structure of the delithiated phase for some of the compounds presented in Scheme 1. The colour code for the atoms is as follow: red for oxygen, brown for carbon, white for hydrogen, blue for nitrogen, yellow for sulphur and green for lithium. b) Calculated open-circuit voltage (V_{OC}) for all structures presented in Scheme 1 as obtained by following the evolutionary algorithm depicted in A. The presented values represent the average voltage for a two-step lithiation reaction.

are presented for all these compounds following the same index-base from Scheme 1. With values spanning from 0.58 V to numbers as high as 3.29 V, this group of compounds exhibit both anodic and cathodic potentials with suitable use in LIBs. Moreover, the presence of voltages ranging from anodes to cathodes in this dataset greatly improves its statistical diversity and helps to enhance the linear model performance (discussed below). As expected, molecules based solely on carboxylates or functionalized with NH_2 groups display the lowest potentials (1, 3, 9, 13, 16, 19, 20, 22, 24) while those functionalized with NO_2 or composed with N heteroatoms exhibit the highest ones (4, 6, 7, 12, 15, 21, 25, 26, 28). It is also interesting to note how the presence of Br, a weaker electron-withdrawing group than NO_2, shifts the voltage from 2.94 V in 25 to 3.17 V in 28.

As described in step C, the Linear model was built to assess the open circuit voltages on top of the P_{Red} as molecular descriptor, i.e., input to the model. In addition to the organic crystals predicted in the previous step, the lithiation voltages of several organic electrodes were introduced from experimental results obtained from literature. This fact expands the model diversity and enhances its prediction effectiveness as the added experimental molecules are not exclusively based on carboxylates. Furthermore, the presence of experimental results also validates the methodology. A list of these molecules, lithiation voltages and references is presented in Scheme S1 of the Supplementary Material. Fig. 4 presents the performance of the regression technique and its respective equation. From Fig. 4, it is worth noting the good correlation between predicted and target voltages, especially for the experimental moieties.

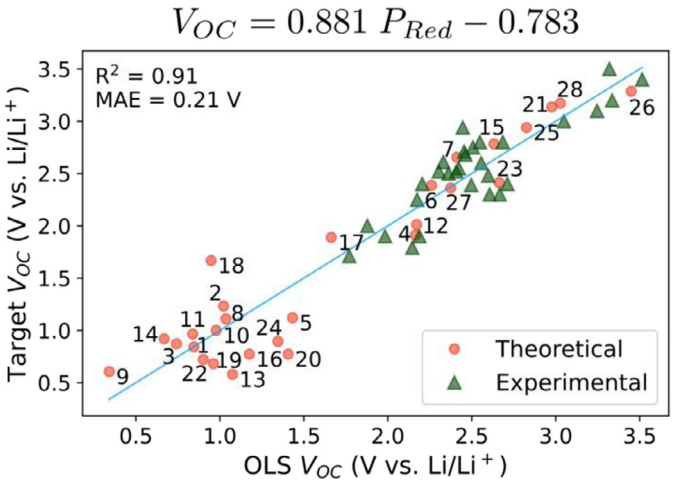


Fig. 4. The OLS model connecting the reduction potential molecular descriptor and the open-circuit voltage. The respective equation for each model is evidenced above the plots. The numerical labels follow the ones in Scheme 1.

To build the Neural model, different combinations of fingerprints and network architectures were benchmarked to generate the most effective model. A detailed explanation of the network architectures can be found in the computational method section and Supplementary Material. The neural networks for all the CM and MBTR combinations were coded on the TensorFlow framework [81,82] while the SMILES were developed on top of PyTorch [83,84]. The mean absolute error (MAE) (Equation S7) was chosen as the training criteria for the networks when analysing the general performance of the different fingerprints and architectures. The training was performed in a portion of the OMEAD molecular database with 18,528 samples, while 2290 were reserved for testing purposes. The scores for all the tested combinations can be seen in Table S2. Somewhat surprisingly, the SMILES representation achieved similar performance as the one for the MBTR, a fingerprint considerably more robust and capable to encode more structural information. Therefore, due to its simplicity and conceptual elegance, the SMILES architecture was the final choice. Furthermore, this approach facilitates the access to a bigger sampling universe as we avoid demanding many-body structural processing required with other fingerprints, e.g., calculating atom pairs distances, angles, etc.

A general schematic of the final Neural model is presented in Fig. 5(a), showing the terephthalic acid SMILES as an example. After the initial indexing of SMILES into the NLP, the data is translated into the embedding vectors. The sequence of these embeddings is fed to four independent RNN (simply named A, B, C and D) and their output is forwarded to four independent FC-layers. This first part may also be seen as an encoding scheme, in which the network is creating an internal representation of each molecular structure while learning the meaning of the SMILES elements (atom species, bonds, etc.) via the embedding process and the knowledge behind their ordering through the recurrent layers. Afterwards, a concatenation operation is responsible for merging the output from the four-level layers before forwarding it to two sequential fully connected layers. Following the encoding scheme, this last step may be understood as a decoding process, where the network translates the desired structure-property relationship. Fig. 5(b) shows the Neural model performance when predicting the reduction potential for the test dataset. More details about the learning process and model performance for training and testing sets can be found in Figure S3 of the Supplementary Material, in which we have also included a second network for the oxidation potential.

With the Linear and Neural models trained, the AI-kernel is settled. The next step is to apply the framework in production to explore the organic universe and identify new possible electrodes for LIBs through a high-throughput screening approach. To do so, a total of 20 million molecules were extracted from the GDB17 dataset and processed by the AI-kernel to predict their lithiation voltages (V_{OC}). It is worth emphasizing that the prediction step, apart from the textual pre-processing, required less than 40 min for all the 20 million SMILES on a personal computer. This task would demand more than 10 years of a state-of-the-art supercomputer if the corresponding quantum mechanical calculations were to be employed for all these compounds.

To select possible candidates, a simple voltage filter was applied to identify cathodic compounds with $\boldsymbol{V_{OC}}$ higher than 2.9 V (vs. Li/Li$^+$) and anodic compounds with $\boldsymbol{V_{OC}}$ between 0.0 V and 0.5 V (vs. Li/Li$^+$). Additionally, a second filter based on the theoretical lithiation capacity has been used to narrow down the cathode candidates' list and select compounds with capacities higher than 100 mA h g^{-1}. Out of the 20 million molecules analysed, the filtering selection delivered 1001 possible cathodes and 6014,051 anodes. From the latter category, we randomly selected 1500, hence achieving 2501 samples for further analysis. This contrast in the number of possible candidates further reveals how challenging is to find organic electrodes of higher potentials (> 2.9 V vs. Li/Li$^+$), highlighting the demand for alternative approaches to accelerate materials discovery like the one proposed in this work. In fact, the number of identified cathodes represents only 0.0050% of the total of 20 million moieties analysed.

To get an insight into what molecular features that the AI recommends, Fig. 6 shows the occurrence (in%) and the selection ratio of atomic species (represented by their chemical symbols) and the number of ring structures that the molecules are composed of (represented by integers) for both low and high lithiation voltages. The comparative selection thereby illustrates the likelihood of the selected molecules being composed by specific atom species and the number of rings. Therefore, it reveals how likely it is that any of these elements will affect the lithiation voltage. As can be seen, sulphur, bromine and chlorine show the highest selection ratios when aiming at high-voltage compounds, followed by fluorine. Conversely, these elements show significantly lower selection ratios when it comes to low-voltage electrodes, where the number of ring structures appears to be the most relevant characteristic. In this regard, molecules with a higher number of ring structures seems to be more suited as anode compounds while the opposite is true for cathodes, in which the two rings case displays the highest score. For oxygen and nitrogen, the selection ratio alone does not provide any direct conclusion, but when combined with the occurrence analysis it shows that the AI slightly favours the selection of oxygen/nitrogen-containing molecules as cathodes. This method of counting occurrences, however, fails to identify functional groups. For instance, it is not possible to identify methyl, a group that usually act as an electron-donating and thereby reduces the lithiation voltage.

After the selection process, a second round of DFT calculations was performed for all the 2501 molecules to assess their reduction potentials and $\boldsymbol{V_{OC}}$ as a part of the high throughput workflow. This DFT step act as a third-layer filter, improving the final molecules to be proposed for production by cleaning the statistical noise commonly find in data-driven approaches. Additionally, all these 2501 molecules can now be fed to the OMEAD, creating a continuous feedback loop. Simultaneously, these new results can be used to further analyse the AI-kernel's performance. Fig. 7 presents a collection of benchmarks between DFT and AI data. The probability density function (PDF) is shown in Figs. 7(a) and (b) for reduction potentials and Figs. 7(c) and (d) for the open-circuit voltages of the anodic and cathodic ranges, respectively. The PDF is a Gaussian fit to the data distribution that indicates the likelihood to find a certain outcome ($\boldsymbol{P_{Red}}$ or $\boldsymbol{V_{OC}}$). For both low and high potentials, it is clear that our machinery is offering predictions in the correct range of values. Fig. 7(e) and (f) show the 2D histogram plot for anode and cathode candidates, respectively. The darker areas in the plot represent a higher data correlation between values from DFT and AI. The data

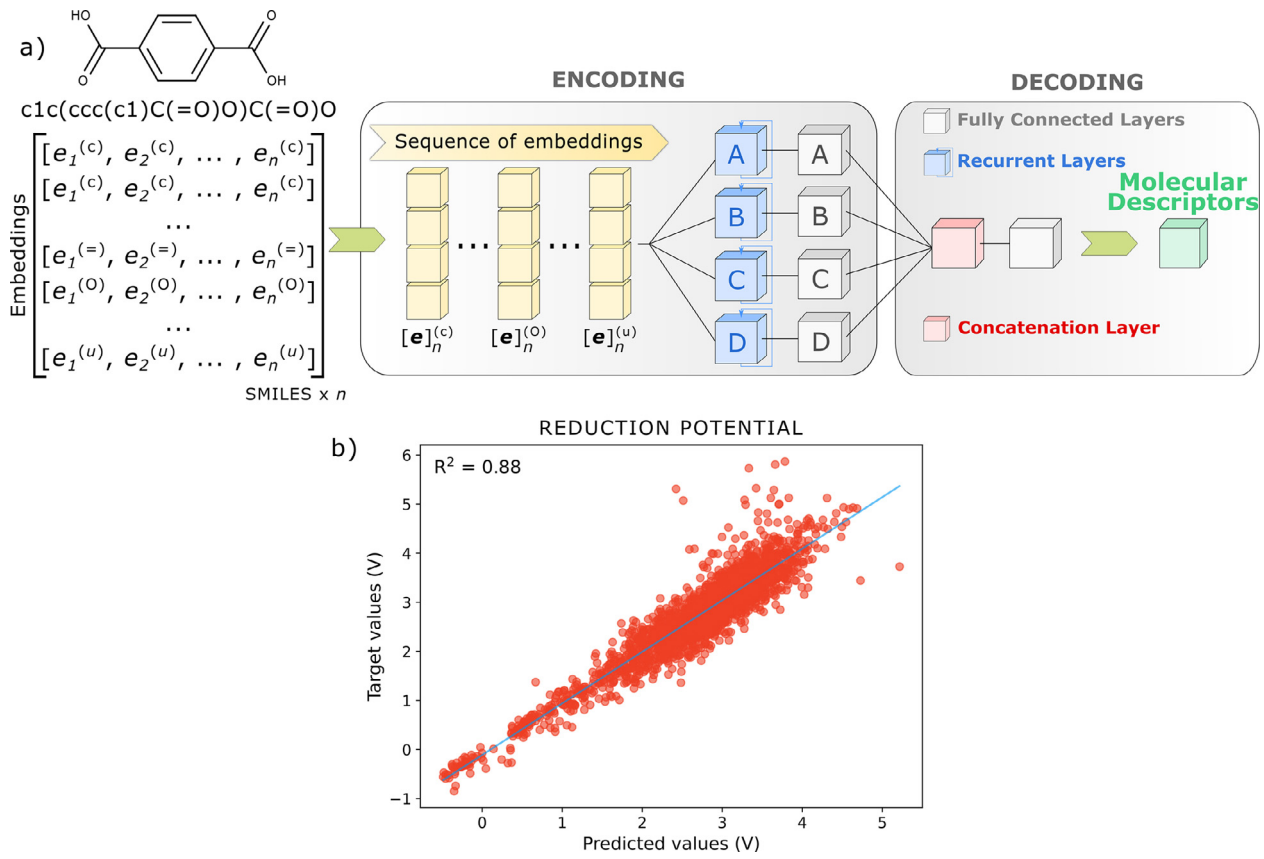


Fig. 5. (a) A simplified representation of the Neural Network model. As an example, the terephthalic acid has its SMILES string converted into a sequence of embedding vectors that will feed the Neural Network layers. (b) Neural model performance in the test dataset of 2290 molecules in predicting reduction potentials.

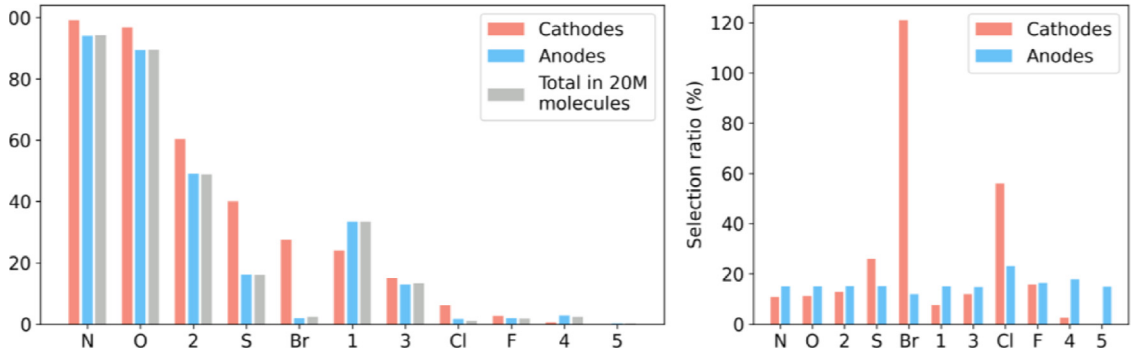


Fig. 6. The occurrences on the left and the selection ratio on the right for the selected molecules in the anodic and cathodic range of potentials. The x-axes list the chemical symbols (excluding carbon) and the number of ring structures the molecules are composed of.

distributions are in the expected region of ~ 0 V and ~ 3 V for both low and high potential ranges.

The overall result in Fig. 7 presents a good agreement between DFT and AI, which reasserts the model's performance. Small deviations are mainly due to outliers coming mostly from molecules that went through significant structural changes over the redox process in the DFT calculations. These outliers often result in very large/small voltages that widen the probability densities. Such unrealistic voltages should not be considered toward the final AI-kernel performance. Furthermore, their appearance is connected to physicochemical processes happening during the redox reaction that the AI-kernel is not capable to assess, such as bond cleavage leading to dissociations or irreversible structural changes. The SMILES does not implicitly encode any information about changes in the charged phases and the presence of outliers is therefore anticipated. On the other hand, this fact can help to identify unwanted molecules that are prone to degradation during the lithiation process, which can also present a new pathway for an intelligent design of compounds in the future.

Scheme 2 displays a small set of molecules selected as cathode candidates after the validation process. We present the V_{OC} as obtained from both AI and DFT. Our kernel tends to accurately associate cathodic potentials with molecules containing electron-withdrawing groups, such as nitrile –C ≡ N, carbonyl –C = O and halogen functionalities. In fact, these groups are commonly employed as voltage enhancers in organic electrodes due to their effect over the redox-active centres. At the same time, some electron-donating ligands – like amino -NH_2, hydroxy -OH

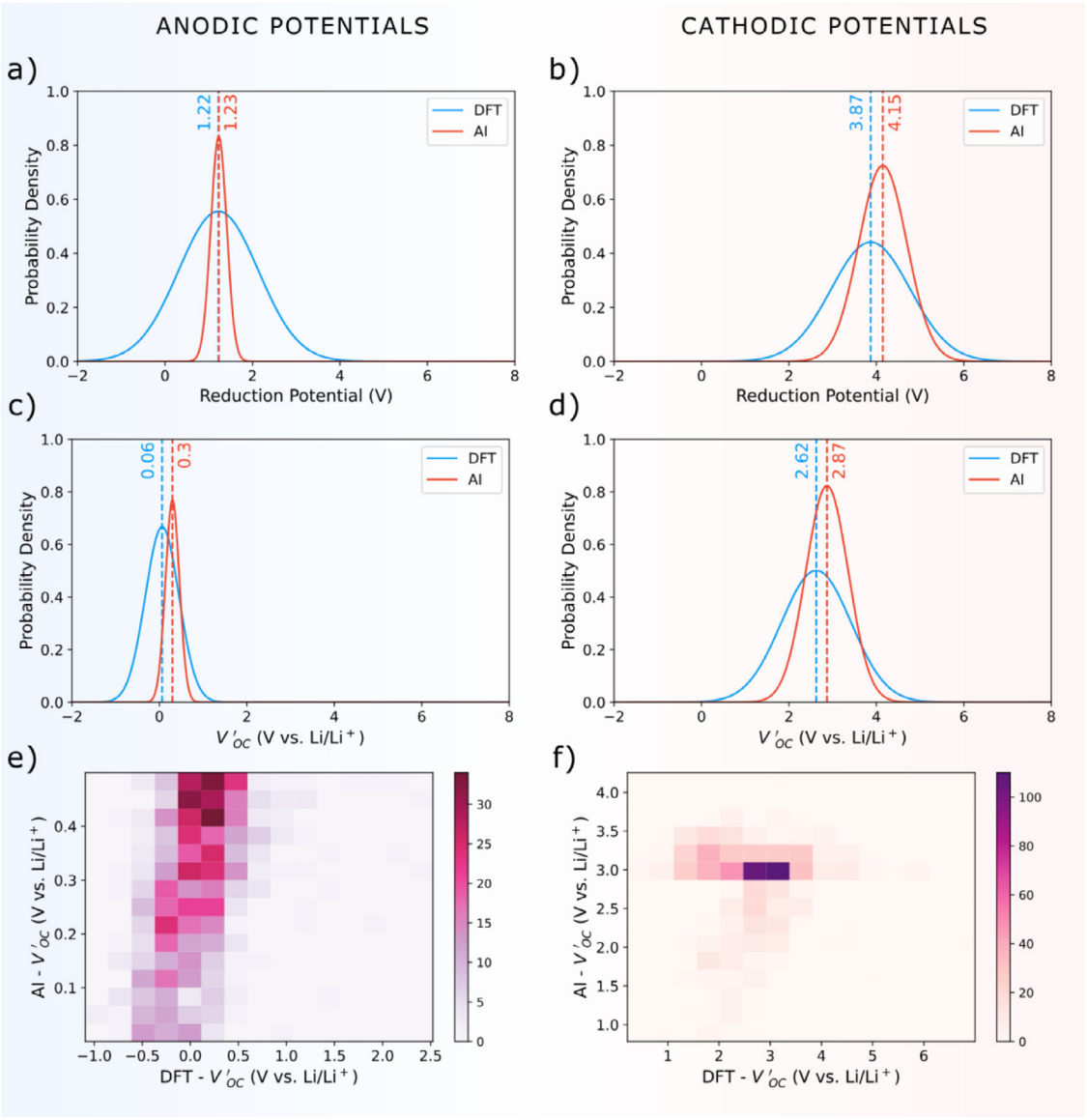
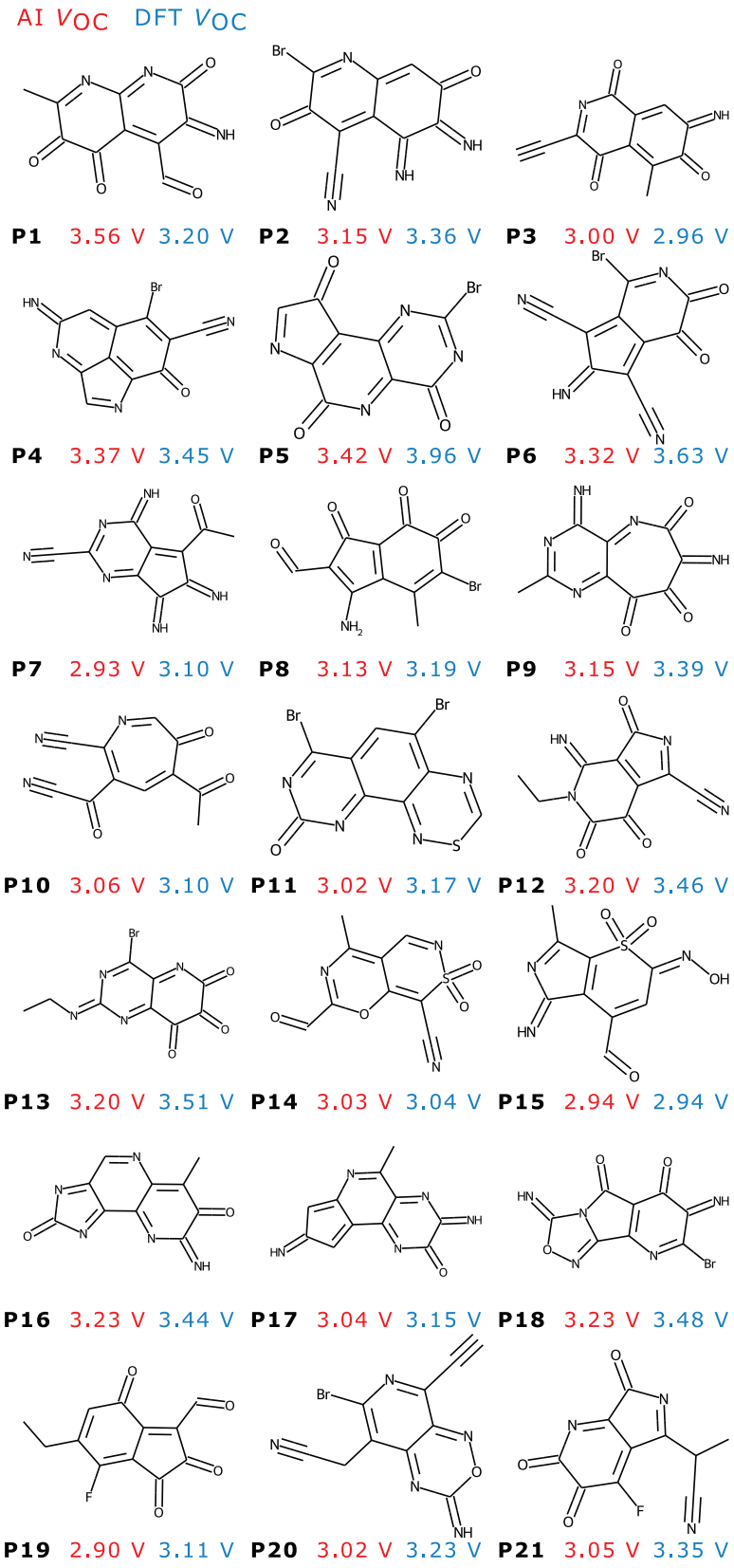


Fig. 7. The reduction potentials and open-circuit voltages (V_{OC}) probability densities for anodic (a, c) and cathodic (b, d) range of potentials, respectively, exhibiting the closeness between the distribution of AI-kernel predictions and DFT-based values. A 2D-histogram plot of AI and DFT V_{OC} for the respective e) anodic and f) cathodic potentials.

and alkyl -C_xH_y groups – are also included for some molecules. This is an interesting and well-known combination in organic optoelectronics and photovoltaics that may lead to a donor-acceptor-like behaviour [85–88]. Although this mechanism is not often explored in the literature of organic batteries, an effective use could help to engineer the electronic structure of the target electroactive compound and, thus, improving the charge localization over the redox-active centre during the Li-ion insertion. In fact, this effect may be one of the reasons for the higher voltage displayed by some of the compounds discovered by the AI.

Nonetheless, the redox sites of the selected molecules appear to be dominated by double bonding oxygens in carbonyls and sulfonyls and double/triple bonding nitrogen. These redox sites are commonly associated with cathodes in organic batteries [7,17,18,89], which further supports the performance of the AI-kernel in identifying molecules with a cathodic character. The complete list of proposed cathode candidates is available in the Supplementary Material, in which we have removed molecules that presented any sort of degradations during the DFT redox process calculation.

A closer look into the molecules shown in Scheme 2 may further reveal their possible performance as active materials for cathodes. For instance, the reduction sites analysis (Eq. (5)) for the compounds P1, P2, P5, P9, P12, P14, P18 and P21 reveals that they could offer a maximum theoretical lithiation capacity (or if limited for two inserted Li-ions) of about 564 (188), 394 (158), 524 (150), 718 (179), 568 (189), 527 (176), 524 (150) and 477 (238) mA h g^{-1}, respectively. These compounds could offer energy densities higher than 600 W h kg^{-1} when considering the two-lithium insertion process. Alternatively, theoretical energy densities superior to 1000 W h kg^{-1} could possibly be achieved if the maximum capacities were to be accessed. This represents an appealing alternative to some of the state-of-the-art technologies available to date, which offer energy densities close to 800 W h kg^{-1}. The realisation of such enhanced materials could place organic-based batteries in a favourable position as a next generation technology for energy-demanding applications where the combination of high gravimetric energy density and battery sustainability is necessary.



Scheme 2. Lew is representations of a few selected molecules for cathode candidates and their respective lithiation voltages (V_{OC} vs. Li/Li^+) as predicted by the AI-kernel and by using the DFT redox potentials.

5. Conclusions

A framework based on Artificial Intelligence and backed by density functional theory calculations has been developed to accelerate the discovery of novel organic battery materials. The workflow consists of four main steps: (i) first-principles crystal structure prediction of OEMs using an evolutionary algorithm and the corresponding structural modifications upon insertion of ions and electrons during battery cycling; (ii) the development of a database of organic energy materials including key molecular properties assessed through high-level DFT calculations; (iii) the design of an AI-kernel by combining two machine learning (Neural and Linear) models to allow a fast access of a bigger chemical space of organic compounds; (iv) high-throughput screening of 20 million molecules when searching for new organic cathodes candidates.

A new software written in the python language (ANIMA) has been developed to work in conjunction with the neural network architectures that have been investigated. The NNs have been trained on DFT calculated properties of 18,528 molecules while the properties of another 2290 molecules were used to test the quality of the model. The SMILE fingerprints combined with a robust network architecture composed of RNN and FCNN multilayers turned out to be the most efficient algorithm to explore the potential OEMs. Finally, a Linear model was developed to connect the molecular and solid-state electrode properties.

The kernel was then employed to screen 20 million molecules extracted from the GDB17 database, assessing the redox potentials and lithiation voltages, a task that required less than an hour (excluding the textual pre-processing of SMILES). From this set, 1001 cathodes and 1500 anodes candidates were selected by applying voltage and capacity filters. A detailed analysis of these molecules revealed how relevant a few key characteristics are for the AI to obtain cathodes and anodes, such as atomic species and the number of ring structures composing the molecule. Thereafter, DFT calculations were carried out on the 2501 selected molecules to improve the cathodes selection and to benchmark the AI-kernel, which in turn presented its capability of correctly locating the voltages in the respective anodic/cathodic range. Moreover, the kernel accurately identified common functional groups, like nitro and nitrile, that lead to higher voltage electrodes and pointed out an interesting donor-acceptor-like effect that may drive the design of novel cathode materials. Herein, a final list of 459 promising molecules selected after the DFT-filtering process is being provided with the Supplementary Material.

In light of these results, one may see how the use of AI-based methodologies could speed up the discovering process of new electroactive materials for Li-ion batteries and be a revolutionising tool into tapping the huge universe of organic materials. Finally, a list of novel high-voltage cathodes is proposed as promising candidates for the next generations of organic batteries, with some compounds offering the potential to exhibit energy densities superior to 1000 W h kg^{-1}.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Rodrigo P. Carvalho: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – review & editing. **Cleber F.N. Marchiori:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – review & editing. **Daniel Brandell:** Conceptualization, Validation, Resources, Writing – review & editing, Supervision, Funding acquisition. **C. Moyses Araujo:** Conceptualization, Validation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

We acknowledge support from the Swedish Research Council (Grant numbers 2018–04506 and 2020–05223), the Swedish Energy Agency (Grant number: 45420–1) and STandUP for Energy. The computational infrastructure has been provided by the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre (NSC) at Linköping University.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ensm.2021.10.029.

References

[1] N.S. Lewis, D.G. Nocera, Powering the planet: chemical challenges in solar energy utilization, in: Proceedings of the National Academy of Sciences of the United States of America, 103, 2006, pp. 15729–15735, doi:10.1073/pnas.0603395103.
[2] D.M. Schultz, T.P. Yoon, Solar synthesis: prospects in visible light photocatalysis, Science (2014) 343, doi:10.1126/science.1239176.
[3] S. Renault, V.A. Oltean, C.M. Araujo, A. Grigoriev, K. Edström, D. Brandell, Superlithiation of Organic Electrode Materials: the Case of Dilithium Benzenedipropiolate, Chem. Mater. 28 (2016) 1920–1926, doi:10.1021/acs.chemmater.6b00267.
[4] S.E. Burkhardt, J. Bois, J.M. Tarascon, R.G. Hennig, H.D. Abruña, Li-carboxylate anode structure-property relationships from molecular modeling, Chem. Mater. 25 (2013) 132–141, doi:10.1021/cm302839z.
[5] P. Poizot, F. Dolhem, Clean energy new deal for a sustainable world: from non-CO2 generating energy sources to greener electrochemical storage devices, Energy Environ. Sci. 4 (2011) 2003–2019, doi:10.1039/c0ee00731e.
[6] D. Larcher, J.M. Tarascon, Towards greener and more sustainable batteries for electrical energy storage, Nat. Chem. 7 (2015) 19–29, doi:10.1038/nchem.2085.
[7] S. Muench, A. Wild, C. Friebe, B. Häupler, T. Janoschka, U.S. Schubert, Polymer-Based Organic Batteries, Chem. Rev. 116 (2016) 9438–9484, doi:10.1021/acs.chemrev.6b00070.
[8] F. Li, Z. Wei, A. Manthiram, Y. Feng, J. Ma, L. Mai, Sodium-based batteries: from critical materials to battery systems, J. Mater. Chem. A 7 (2019) 9406–9431, doi:10.1039/c8ta11999f.
[9] H. Banda, D. Damien, K. Nagarajan, A. Raj, M. Hariharan, M.M. Shaijumon, Twisted Perylene Diimides with Tunable Redox Properties for Organic Sodium-Ion Batteries, Adv. Energy Mater. 7 (2017) 1701316, doi:10.1002/aenm.201701316.
[10] Z. Luo, L. Liu, Q. Zhao, F. Li, J. Chen, An Insoluble Benzoquinone-Based Organic Cathode for Use in Rechargeable Lithium-Ion Batteries, Angew. Chem. 129 (2017) 12735–12739, doi:10.1002/ange.201706604.
[11] J.K. Kim, Y. Kim, S. Park, H. Ko, Y. Kim, Encapsulation of organic active materials in carbon nanotubes for application to high-electrochemical-performance sodium batteries, Energy Environ. Sci. 9 (2016) 1264–1269, doi:10.1039/c5ee02806j.
[12] H.W. Kim, H.J. Kim, H. Byeon, J. Kim, J.W. Yang, Y. Kim, J.K. Kim, Binder-free organic cathode based on nitroxide radical polymer-functionalized carbon nanotubes and gel polymer electrolyte for high-performance sodium organic polymer batteries, J. Mater. Chem. A 8 (2020) 17980–17986, doi:10.1039/d0ta04526h.
[13] B. Esser, F. Dolhem, M. Becuwe, P. Poizot, A. Vlad, D. Brandell, A perspective on organic electrode materials and technologies for next generation batteries, J. Power Sources 482 (2021) 228814, doi:10.1016/j.jpowsour.2020.228814.
[14] J.R. Smith, P.A. Cox, S.A. Campbell, N.M. Ratcliffe, Application of density functional theory in the synthesis of electroactive polymers, J. Chem. Soc. Faraday Trans. 91 (1995) 2331–2338, doi:10.1039/FT9959102331.
[15] T.W. Kemper, R.E. Larsen, T. Gennett, Relationship between molecular structure and electron transfer in a polymeric Nitroxyl-radical energy storage material, J. Phys. Chem. C 118 (2014) 17213–17220, doi:10.1021/jp501628z.
[16] P. Acker, L. Rzesny, C.F.N. Marchiori, C.M. Araujo, B. Esser, π-Conjugation Enables Ultra-High Rate Capabilities and Cycling Stabilities in Phenothiazine Copolymers as Cathode-Active Battery Materials, Adv. Funct. Mater. 29 (2019) 1906436, doi:10.1002/adfm.201906436.
[17] M.E. Bhosale, S. Chae, J.M. Kim, J.Y. Choi, Organic small molecules and polymers as an electrode material for rechargeable lithium ion batteries, J. Mater. Chem. A 6 (2018) 19885–19911, doi:10.1039/c8ta04906h.
[18] P. Poizot, F. Dolhem, J. Gaubicher, Progress in all-organic rechargeable batteries using cationic and anionic configurations: toward low-cost and greener storage solutions? Curr. Opin. Electrochem. 9 (2018) 70–80, doi:10.1016/j.coelec.2018.04.003.
[19] B. Haeupler, A. Wild, U.S. Schubert, Carbonyls: powerful organic materials for secondary batteries, Adv. Energy Mater. 5 (2015) 1402034.
[20] T.B. Schon, B.T. McAllister, P.-F. Li, D.S. Seferos, The rise of organic electrode materials for energy storage, Chem. Soc. Rev. 45 (2016) 6345–6404, doi:10.1039/C6CS00173D.
[21] Z. Song, H. Zhou, Towards sustainable and versatile energy storage devices: an overview of organic electrode materials, Energy Environ. Sci. 6 (2013) 2280–2301, doi:10.1039/c3ee40709h.
[22] Y. Liang, Z. Tao, J. Chen, Organic Electrode Materials for Rechargeable Lithium Batteries, Adv. Energy Mater. 2 (2012) 742–769, doi:10.1002/aenm.201100795.

[23] J. Wang, A.E. Lakraychi, X. Liu, L. Sieuw, C. Morari, P. Poizot, A. Vlad, Conjugated sulfonamides as a class of organic lithium-ion positive electrodes, Nat. Mater. (2020) 1–9, doi:10.1038/s41563-020-00869-1.
[24] M. Yao, H. Senoh, S.I. Yamazaki, Z. Siroma, T. Sakai, K. Yasuda, High-capacity organic positive-electrode material based on a benzoquinone derivative for use in rechargeable lithium batteries, J. Power Sources 195 (2010) 8336–8340, doi:10.1016/j.jpowsour.2010.06.069.
[25] J.E. Bachman, L.A. Curtiss, R.S. Assary, Investigation of the redox chemistry of anthraquinone derivatives using density functional theory, J. Phys. Chem. A 118 (2014) 8852–8860, doi:10.1021/jp5060777.
[26] J.H. Park, T. Liu, K.C. Kim, S.W. Lee, S.S. Jang, Systematic Molecular Design of Ketone Derivatives of Aromatic Molecules for Lithium-Ion Batteries: first-Principles DFT Modeling, ChemSusChem 10 (2017) 1584–1591, doi:10.1002/cssc.201601730.
[27] L. Ruddigkeit, R. van Deursen, L.C. Blum, J.L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17, J. Chem. Inf. Model. 52 (2012) 2864–2875, doi:10.1021/ci300415d.
[28] K. Hatakeyama-Sato, T. Tezuka, Y. Nishikitani, H. Nishide, K. Oyaizu, Synthesis of lithium-ion Conducting Polymers Designed by Machine Learning-based Prediction and Screening, Chem. Lett. 48 (2019) 130–132, doi:10.1246/cl.180847.
[29] O. Allam, B.W. Cho, K.C. Kim, S.S. Jang, Application of DFT-based machine learning for developing molecular electrode materials in Li-ion batteries, RSC Adv. 8 (2018) 39414–39420, doi:10.1039/c8ra07112h.
[30] T. Parthiban, R. Ravi, N. Kalaiselvi, Exploration of artificial neural network [ANN] to predict the electrochemical characteristics of lithium-ion cells, Electrochim. Acta 53 (2007) 1877–1882, doi:10.1016/j.electacta.2007.08.049.
[31] Y. Liu, B. Guo, X. Zou, Y. Li, S. Shi, Machine learning assisted materials design and discovery for rechargeable batteries, Energy Storage Materials 31 (2020) 434–450, doi:10.1016/j.ensm.2020.06.033.
[32] R.P. Joshi, J. Eickholt, L. Li, M. Fornari, V. Barone, J.E. Peralta, Machine Learning the Voltage of Electrode Materials in Metal-Ion Batteries, ACS Appl. Mater. Interfaces 11 (2019) 18494–18503, doi:10.1021/acsami.9b04933.
[33] B. Olsthoorn, R.M. Geilhufe, S.S. Borysov, A.v. Balatsky, Band Gap Prediction for Large Organic Crystal Structures with Machine Learning, Advanced Quantum Technologies 2 (2019) 1900023, doi:10.1002/qute.201900023.
[34] J. Lee, A. Seko, K. Shitara, K. Nakayama, I. Tanaka, Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques, Phys. Rev. B 93 (2016), doi:10.1103/PhysRevB.93.115104.
[35] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, Sci. Rep. 3 (2013), doi:10.1038/srep02810.
[36] G. Pilania, J.E. Gubernatis, T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, Comput. Mater. Sci. 129 (2017) 156–163, doi:10.1016/j.commatsci.2016.12.004.
[37] L. Wilbraham, R.S. Sprick, K.E. Jelfs, M.A. Zwijnenburg, Mapping binary copolymer property space with neural networks, Chem. Sci. 10 (2019) 4973–4984, doi:10.1039/c8sc05710a.
[38] O. Çaylak, A. Yaman, B. Baumeier, Evolutionary Approach to Constructing a Deep Feedforward Neural Network for Prediction of Electronic Coupling Elements in Molecular Materials, J. Chem. Theory Comput. 15 (2019) 1777–1784, doi:10.1021/acs.jctc.8b01285.
[39] G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, A.v. Lilienfeld, K. Müller, Learning Invariant Representations of Molecules For Atomization Energy Prediction, Advances in Neural Information Processing Systems(NIPS'12), 2012.
[40] M. Rupp, A. Tkatchenko, K.R. Müller, O.A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, Phys. Rev. Lett. (2012) 108, doi:10.1103/PhysRevLett.108.058301.
[41] E.v. Podryabinkin, E.v. Tikhonov, A.v. Shapeev, A.R. Oganov, Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning, Phys. Rev. B 99 (2019), doi:10.1103/PhysRevB.99.064114.
[42] V. Botu, R. Ramprasad, Adaptive machine learning framework to accelerate ab initio molecular dynamics, Int. J. Quantum Chem. 115 (2015) 1074–1083, doi:10.1002/qua.24836.
[43] M. Gastegger, J. Behler, P. Marquetand, Machine learning molecular dynamics for the simulation of infrared spectra, Chem. Sci. 8 (2017) 6924–6935, doi:10.1039/c7sc02267k.
[44] J. Behler, Perspective: machine learning potentials for atomistic simulations, J. Chem. Phys. (2016) 145, doi:10.1063/1.4966192.
[45] P.v. Balachandran, A.A. Emery, J.E. Gubernatis, T. Lookman, C. Wolverton, A. Zunger, Predictions of new AB O3 perovskite compounds by combining machine learning and density functional theory, Phys. Rev. Materials 2 (2018), doi:10.1103/PhysRevMaterials.2.043802.
[46] F. Brockherde, L. Vogt, L. Li, M.E. Tuckerman, K. Burke, K.R. Müller, Bypassing the Kohn-Sham equations with machine learning, Nat. Commun. 8 (2017) 1–10, doi:10.1038/s41467-017-00839-3.
[47] P.M. Attia, A. Grover, N. Jin, K.A. Severson, T.M. Markov, Y.-.H. Liao, M.H. Chen, B. Cheong, N. Perkins, Z. Yang, P.K. Herring, M. Aykol, S.J. Harris, R.D. Braatz, S. Ermon, W.C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature 578 (2020) 397–402, doi:10.1038/s41586-020-1994-5.
[48] X. Chen, X. Liu, X. Shen, Q. Zhang, Applying Machine Learning to Rechargeable Batteries: From the Microscale to the Macroscale, Angewandte Chemie International Edition, 2021, doi:10.1002/ANIE.202107369.
[49] Y. Okamoto, Y. Kubo, Ab Initio Calculations of the Redox Potentials of Additives for Lithium-Ion Batteries and Their Prediction through Machine Learning, ACS Omega 3 (2018) 7868–7874, doi:10.1021/ACSOMEGA.8B00576.
[50] C.F.N. Marchiori, D. Brandell, C.M. Araujo, Predicting Structure and Electrochemistry of Dilithium Thiophene-2,5-Dicarboxylate Electrodes by Density Functional Theory and Evolutionary Algorithms, J. Phys. Chem. C 123 (2019) 4691–4700, doi:10.1021/acs.jpcc.8b11341.
[51] R.P. Carvalho, C.F.N. Marchiori, D. Brandell, C.M. Araujo, Tuning the Electrochemical Properties of Organic Battery Cathode Materials: insights from Evolutionary Algorithm DFT Calculations, ChemSusChem 13 (2020) 2402–2409, doi:10.1002/cssc.201903450.
[52] R.P. Carvalho, C.F.N. Marchiori, V.-.A. Oltean, S. Renault, T. Willhammar, C. Pay Gómez, C.M. Araujo, D. Brandell, Structure–property relationships in organic battery anode materials: exploring redox reactions in crystalline Na- and Li-benzene diacrylate using combined crystallography and density functional theory calculations, Materials Advances 2 (2021) 1024–1034, doi:10.1039/d0ma00900h.
[53] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Physical Review B - Condensed Matter and Materials Physics 59 (1999) 1758–1775, doi:10.1103/PhysRevB.59.1758.
[54] G. Kresse, J. Hafner, Ab initio molecular dynamics for liquid metals, Phys. Rev. B 47 (1993) 558–561, doi:10.1103/PhysRevB.47.558.
[55] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Physical Review B - Condensed Matter and Materials Physics 54 (1996) 11169, doi:10.1103/PhysRevB.54.11169.
[56] A.R. Oganov, C.W. Glass, Crystal structure prediction using ab initio evolutionary techniques: principles and applications, J. Chem. Phys. 124 (2006) 244704, doi:10.1063/1.2210932.
[57] A.O. Lyakhov, A.R. Oganov, H.T. Stokes, Q. Zhu, New developments in evolutionary structure prediction algorithm USPEX, Comput. Phys. Commun. 184 (2013) 1172–1182, doi:10.1016/j.cpc.2012.12.009.
[58] A.R. Oganov, A.O. Lyakhov, M. Valle, How evolutionary crystal structure prediction works-and why, Acc. Chem. Res. 44 (2011) 227–237, doi:10.1021/ar1001318.
[59] C.W. Glass, A.R. Oganov, N. Hansen, USPEX-Evolutionary crystal structure prediction, Comput. Phys. Commun. 175 (2006) 713–720, doi:10.1016/j.cpc.2006.07.020.
[60] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865.
[61] J. Heyd, G.E. Scuseria, M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, J. Chem. Phys. 118 (2003) 8207–8215, doi:10.1063/1.1564060.
[62] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, J. Comput. Chem. 27 (2006) 1787–1799, doi:10.1002/jcc.20495.
[63] Y.S. Meng, M.E. Arroyo-De Dompablo, First principles computational materials design for energy storage materials in lithium ion batteries, Energy Environ. Sci. 2 (2009) 589–609, doi:10.1039/b901825e.
[64] R.B. Araujo, A. Banerjee, P. Panigrahi, L. Yang, M. Strømme, M. Sjödin, C.M. Araujo, R. Ahuja, Designing strategies to tune reduction potential of organic molecules for sustainable high capacity battery application, J. Mater. Chem. A 5 (2017) 4430–4454, doi:10.1039/c6ta09760j.
[65] G.A. Frisch, M.J. Trucks, G.W. Schlegel, H.B. Scuseria, G.E. Robb, M.A. Cheeseman, J.R. Scalmani, G. Barone, V. Petersson, Gaussian 16, Rev. C.01, Gaussian, Inc, 2016.
[66] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A 38 (1988) 3098–3100, doi:10.1103/PhysRevA.38.3098.
[67] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785–789, doi:10.1103/PhysRevB.37.785.
[68] R. Ditchfield, W.J. Hehre, J.A. Pople, Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules, J. Chem. Phys. 54 (1971) 720–723, doi:10.1063/1.1674902.
[69] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, doi:10.1038/nature14539.
[70] Z.C. Lipton, J. Berkowitz, C. Elkan, A Critical Review of Recurrent Neural Networks for Sequence Learning, ArXiv Preprint ArXiv, 2015 1506.00019 http://arxiv.org/abs/1506.00019 .
[71] L.R. Medsker, L.C. Jain, Recurrent neural networks: design and applications International Series on Computational Intelligence, Book. (2000). http://books.google.co.uk/books?id=ME1SAkN0PyMC.
[72] S. Sharma, S. Sharma, A. Athaiya, Activation Functions in Neural Networks, Int. J. Eng. Appl. Sci. Technol. 04 (2020) 310–316, doi:10.33564/ijeast.2020.v04i12.054.
[73] B. Karlik, A.V. Olgac, Performance analysis of various activation functions in generalized MLP architectures of neural networks, International Journal of Artificial Intelligence and Expert Systems 1 (2011) 111–122.
[74] K.T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K.R. Müller, E.K.U. Gross, How to represent crystal structures for machine learning: towards fast prediction of electronic properties, Phys. Rev. B (2014) 89, doi:10.1103/PhysRevB.89.205118.
[75] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, in: Machine Learning in Materials informatics: Recent applications and Prospects, Npj Computational Materials, 2017, p. 3, doi:10.1038/s41524-017-0056-5.
[76] B. Parsaeifard, D.S. De, A.S. Christensen, F.A. Faber, E. Kocer, S. De, J. Behler, A. von Lilienfeld, S. Goedecker, An assessment of the structural resolution of various fingerprints commonly used in machine learning, Mach. Learn. (2020), doi:10.1088/2632-2153/abb212.
[77] G.B. Goh, N. Hodas, C. Siegel, A. Vishnu, SMILES2VEC: Predicting chemical Properties from Text Representations, ICLR, 2018 2018.

[78] B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, S. Ho, J. Sloane, P. Wender, V. Pande, Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models, ACS Cent Sci 3 (2017) 1103–1113, doi:10.1021/acscentsci.7b00303.

[79] G.A. Pinheiro, J. Mucelini, M.D. Soares, R.C. Prati, J.L.F. da Silva, M.G. Quiles, Machine Learning Prediction of Nine Molecular Properties Based on the SMILES Representation of the QM9 Quantum-Chemistry Dataset, J. Phys. Chem. A 124 (2020) 9854–9866, doi:10.1021/acs.jpca.0c05969.

[80] H. Huo, M. Rupp, in: Unified Representation of Molecules and Crystals for Machine Learning, ArXiv Preprint ArXiv, 2017, p. 13754. http://arxiv.org/abs/1704.06439. 1704.06439.

[81] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: a system for large-scale machine learning, in: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, 2016, OSDI, 2016, pp. 265–283.

[82] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, ArXiv Preprint ArXiv, 2016 1603.04467 http://arxiv.org/abs/1603.04467 .

[83] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Automatic differentiation in pytorch, NIPS 2017 Autodiff Workshop: The Future of Gradient-Based Machine Learning Software and Techniques, 2017 https://openreview.net/forum?id=BJJsrmfCZ.

[84] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, high-performance deep learning library, Advances in Neural Information Processing Systems, 2019.

[85] G.B. Damas, F. Von Kieseritzky, J. Hellberg, C.F.N. Marchiori, C.M. Araujo, Symmetric Small-Molecules with Acceptor-Donor-Acceptor Architecture for Efficient Visible-Light Driven Hydrogen Production: optical and Thermodynamic Aspects, J. Phys. Chem. C 123 (2019) 30799–30808, doi:10.1021/acs.jpcc.9b07721.

[86] G. Damas, C.F.N. Marchiori, C.M. Araujo, On the Design of Donor-Acceptor Conjugated Polymers for Photocatalytic Hydrogen Evolution Reaction: first-Principles Theory-Based Assessment, J. Phys. Chem. C 122 (2018) 26876–26888, doi:10.1021/acs.jpcc.8b09408.

[87] X. He, B. Cao, T.C. Hauger, M. Kang, S. Gusarov, E.J. Luber, J.M. Buriak, Donor-acceptor small molecules for organic photovoltaics: single-atom substitution (se or s), ACS Appl. Mater. Interfaces 7 (2015) 8188–8199, doi:10.1021/acsami.5b01063.

[88] U. Salzner, Effect of donor-acceptor substitution on optoelectronic properties of conducting organic polymers, J. Chem. Theory Comput. 10 (2014) 4921–4937, doi:10.1021/ct500816c.

[89] S. Lee, G. Kwon, K. Ku, K. Yoon, S.K. Jung, H.D. Lim, K. Kang, Recent Progress in Organic Electrodes for Li and Na Rechargeable Batteries, Adv. Mater. 30 (2018), doi:10.1002/adma.201704682.