An evolutionary-driven AI model discovering redox-stable organic electrode materials for alkali-ion batteries

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

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

ABSTRACT

Data-driven approaches have been revolutionizing materials science and materials discovery in the past years. Especially when coupled with other computational physics methods, they can be applied in complex high-throughput schemes to discover novel materials, e.g. for batteries. In this direction, the present work provides a robust AI-driven framework, to accelerate the discovery of novel organic-based materials for Li-, Na- and K-ion batteries. This platform is able to predict the open-circuit voltage of the respective battery and provide an initial assessment of the materials redox stability. The model was employed to screen 45 million small molecules in the search for novel high-potential cathodes, resulting in a proposed shortlist of 3202, 689 and 702 novel compounds for Li-, Na- and K-ion batteries, respectively, considering only the redox stable candidates.

1. Introduction

Data-driven methods, such as machine learning (ML), have been offering impactful innovation in the materials discovery landscape. The trained ML algorithms allow addressing fundamental questions on complex systems that were unfeasible, or very difficult, to tackle with standard methods. ML has been applied, for instance, to by-pass fundamental quantum mechanics calculations in the description of inter-atomic interactions or in the prediction of the materials properties [1–5]. In this context, novel energy-related materials have recently been explored by combinations of experimental-theoretical frameworks and machine learning [6–12]. In fact, such methodologies have been proven useful in identifying novel functional materials and uncovering different structure-property relationships [13–15]. Moreover, and in connection with the actual environmental hurdles the planet is facing, data-driven approaches have the potential to guide the development of greener and more sustainable technologies [16–20]. In general, the accelerated process offered by such methods can speed up the discovery, and subsequent manufacturing, of unexplored materials that could revolutionize the way energy is consumed, distributed and stored.

Recently, several organic-based materials have been proposed as a feasible alternative for electrical energy storage (EES) by means of organic electrodes for lithium-ion batteries (LIBs) [21–29]. These compounds offer key features connecting energy storage and sustainability aspects, e.g. green chemistry synthetic routes, renewable resources and easier end-of-life treatments. However, there are fundamental issues related to volumetric energy density, rate capability and cycling stability that still need to be resolved before organic-based electrodes could be technologically competitive. For instance, the energy density is given by the product between the volumetric charge storage capacity of the electrode materials and cell voltage. The latter is controlled by the electrochemical potential of the redox active moieties, while the former is given by the number of such moieties per unit of volume. Then, anode and cathode materials should display the lowest and highest electrochemical potentials, respectively, to achieve high cell voltage. To achieve such high potentials in n-type (as it will be defined below) organic materials is a challenging task. Therefore, this work aimed at developing efficient methodologies to predict redox potentials in order to facilitate the materials screening in large databases.

In fact, another key feature of the organic electrode materials (OEMs) is the molecular versatility with a large possible variation of chemical compositions and structures. In fact, a materials library containing hundreds of billions of organic molecules can be feasibly obtained by considering a combination of atoms that appear in common organic compounds, such as C, N, O, S, F, Cl and Br [30]. Thereby, novel organic materials could be discovered and engineered to meet the requirements

---

for final technological applications and overcome the aforementioned hurdles. In this context, data-driven approaches play an important role to realistically screen such huge databases. An AI-driven platform was recently proposed and employed to screen a library of 20 million small molecules in the search for high-potential cathodes for Li-ion batteries [16]. This screening identified more than 400 molecules with remarkable gravimetric energy densities. It should be pointed out that OEMs undergo redox reactions during the battery cycling and depending on the corresponding charge states of such reactions they are grouped as *p-*, *n-* or *bipolar-*type materials [31]. More specifically, the n-type materials will change reversibly between the negatively charged and neutral states while the p-type ones change between neutral and positively charged states. The direction of such reactions during the battery charge/discharge process will determine whether they become anode or cathode active materials. The *n*-type materials have been the subject of our previous work [16] and they are the focus of the current study as well.

Many OEMs degrade upon redox reactions (when they are either reduced or oxidized) leading to irreversible chemical bond breaking. We will refer to such molecules as the redox unstable ones. Another stability issue associated with the organic electrodes is the solubility of the organic active materials into the electrolyte, which is in turn the most common mechanism of capacity fading. The molecular stability in electrochemical systems is in fact a complex and multidimensional topic that requires detailed investigations, from both theoretical and experimental viewpoints. Our previously proposed AI framework did not include any aspect of such stability during reversible ion insertion/deinsertion. To start tapping into this issue, the redox stability was investigated here as a first step into assessing the stability of these materials from a data-driven perspective. To accomplish this, a graph neural network model fueled by molecular graphs has been designed to predict the redox stability of a given molecule, reducing the task to a classification problem. With a prediction accuracy of 74.4% in real data, this model can filter-out undesired, i.e., redox unstable, compounds during screening and, together with the previously described AI model, [16] significantly enhance the discovery of suitable materials.

This AI-driven platform has been further developed to predict not only open-circuit voltages of Li-ion batteries, but also of Na- and K-ion batteries. In this context, several organic compounds have been reported to also work as electrodes for these different alkali-ion batteries [32–38]. Thus, by including these new methodologies, the AI model could be a powerful tool in the quest to discover alternative materials for greener batteries. As a first step in this direction, the framework was employed to screen 45 million molecules to identify possible candidates for high-potential cathodes for Li-, Na- and K-ion batteries. As a result, 4047, 860 and 874 cathode candidates were identified for Li-, Na- and K-ion batteries, respectively. It is important to emphasize that these quantities represent less than 0.009% of the 45 million compounds screened, illustrating the enormous challenge of finding high-potential electrodes for these types of batteries. All these compounds have thereafter been further investigated by employing *ab initio* calculations, which also served the purpose of benchmarking the AI model. After a careful filtering process, a shortlist is here being proposed containing 3202, 689 and 702 novel redox stable molecules for Li-, Na- and K-ion batteries, respectively. It should be emphasized that the ultimate validation of the developed computational materials design framework will be the synthesis and the implementation of these proposed materials in the battery devices. Therefore, this work has the potential to inspire and guide future experimental activities in the field of organic batteries.

2. Results & discussions

In our previous work [16], we have reported a model capable of predicting the Li-ion insertion voltages (*vs.* Li/Li$^+$) of organic-based electrodes by relying only on the molecular reduction potentials. The model was able to connect a molecular descriptor with a solid-state property, the battery open-circuit voltage (V_{OC}), with good accuracy.

In the present work, two other models have been designed, targeting Na- and K-ion batteries. Fig. 1 (a) and (b) shows the performance of the Na$^+$ (*vs.* Na/Na$^+$) and K$^+$ (*vs.* K/K$^+$) models, respectively, in evaluating the open-circuit voltages of a few organic electrodes reported in the literature. A complete list of these compounds is included in Table S1 and Table S2 of the Supporting Information. The *N*-Linear Model (LM), with N = [Li, Na, K], being proposed is based on an ordinary least-square regression to estimate the α and β parameters of the linear function $y = \alpha x + \beta$. In this model, x is the molecular reduction potential (P_{Red}) and y the battery open-circuit voltage referred to the respective ion's reference electrode, i.e., vs. Li/Li$^+$, Na/Na$^+$ and K/K$^+$. Therefore, the presented voltages are going to follow this standard, and the reference electrodes will thus be omitted from the description. The resulting equation is also shown in Fig. 1(a) and (b) for the Na-LM and K-LM, respectively. Furthermore, these model parameters also appear to be connected with physical properties of the respective alkali metal. Fig. 1 (c) shows the correlation between α and electron affinities, and Fig. 1(d) the correlation for β and Pauling electronegativities. Interestingly, a higher electron affinity presents a higher slope for the LM, i.e., higher voltages, while a larger electronegativity shows a more negative offset. Although those correlations are far from trivial to understand, they can be connected to different aspects of the insertion ions. For instance, the studied alkali-metal with the highest electronegativity (Li) shows the most negative offset, i.e., acting to reduce the final insertion voltage. This may be representing a balance between the molecule's and the insertion material's electronegativity. Similarly, the slope correlates to the electron affinity and may be balancing the difference between the reference electrode's reaction and the molecule's reduction potential.

The 'Organic Molecules for Energy Application Database' (OMEAD) was here expanded from 26218, used in our previous work [16], to 41800 unique molecules. The database contains several features extracted from density functional theory (DFT) calculations, further explained in the Computational Methods section. Correspondingly, the neural network model predicting the reduction potential, here called Reduction Potential Neural Model (RPNM), was retrained with more molecular samples. Specific details can be found in the Computational Methods section. Once more, the RPNM accesses the molecule's SMILES [16,39] representation to predict its P_{Red} through a complex natural language processing (NLP) and neural network architecture. The predicted reduction potential then feeds the *N*-LM to evaluate the corresponding open-circuit voltage of the battery. These two models, LM and RPNM, work together to form the so-called AI *model*. This digital envelop predicts the compound's V_{OC} without any experimental input or demanding quantum mechanical calculations. However, it does not anticipate the material's stability under battery operational conditions, i.e., molecular stability during redox reactions. The new Redox Stability Neural Model (RSNM) addresses this issue, adding a new functionality layer to the AI model. Fig. 2 (a) shows an overview of the final model. Firstly, any input molecule must pass through the *redox stability layer* (the RSNM) to check if it would be "suitable", i.e. redox stable, for battery applications. Afterwards, the reduction potential of selected molecules is predicted in the *reduction potential layer* (the RPNM). With the P_{Red}, the desired alkali-ion battery open-circuit voltage can be obtained in the *ION layer* (through the LM).

The RSNM is based on a combination of Graph Neural Networks (GNNs) and Feed-Forward Neural Networks (FCNNs). The GNNs are required since in this model molecular graphs have been employed to represent the molecular entries, encoding information within the graph elements. In this model, embeddings were used to represent bonding types through the graph edges and atomic species within graph nodes. Moreover, the network architecture was obtained through a rational evolutionary process based on a differential evolution algorithm [40, 41]. Fig. 2 (b) shows the overall network architecture, in which the evolutionary process was responsible to determine the amount of parallel-stacked structures called *heads* and serial-stacked network layers of different types. In this context, *parallel* means that the stacked
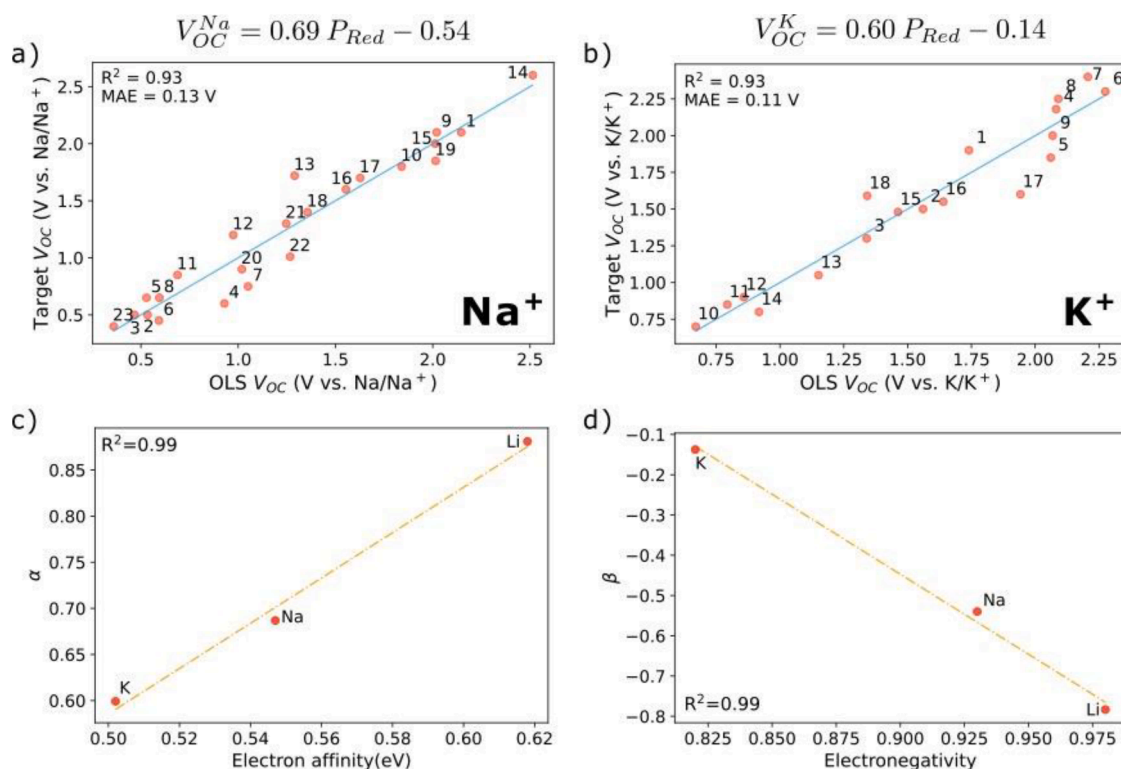
Fig. 1. The Linear Model (LM) performance in predicting the open-circuit voltages for (a) Na- and (b) K-ion batteries. Correlation plots between (c) α and electron affinities and (d) β and electronegativities, where α and β are parameters of the respective LM.

elements receive the same input data while *serial* stands for a sequential processing of data, i.e., a layer receives the output of its previous layer. The *head* structure itself is also a serial stacking of different network layers defined by the evolution mechanism. The final RSNM architecture was obtained after 397 network evaluations from the evolutionary scheme. This process is further detailed in the Computational Methods section and the Supporting Information.

The final RSNM was trained on 15424 molecules and tested on 3856 molecules, with a 0.5/0.5 ratio of redox stable/unstable samples. The final model accuracy on classifying molecules in redox stable/unstable was 80% and 72% for the training and testing dataset, respectively. It is important to note that, among other techniques, early stopping [41,42] was employed to avoid overfitting. Otherwise, the training performance would reach approximately 99% while the testing accuracy would drop to 68%. These accuracies prove the model's overall performance, given the limitation imposed by such a small dataset.

With the AI model settled, a high-throughput screening was performed on 45 million small molecules extracted from the GDB17 [30] dataset searching for high-potential cathodes for Li-, Na- and K-ion batteries. The GDB17 dataset is comprised of moieties with up to 17 non-hydrogen atoms of C, O, N, S, F, Cl and Br. The reduction potential, open-circuit voltages and redox stability for all the 45 million molecules were assessed through the AI model, followed by a two-step filter process of voltages → capacities to select potential candidates. The storage capacities were determined by the following equation

$$C = \frac{nF}{3.6M},\tag{1}$$

where F is the Faraday constant, M the molecule's molar mass and n the number of redox sites, anticipated by analyzing the possible reduction sites from the molecular units. The threshold of 100 mAh/g was considered for capacities while the voltage cutoff was of 3.0 V for Li$^+$ and 2.6 V for Na$^+$ and K$^+$. The screening resulted in 4047, 860 and 874 cathode candidates for the Li-, Na- and K-ions, respectively.

For the molecules classified by the neural network model as redox stable, Fig. 3 (a) and (b) display, respectively, the occurrence (in %) and the selection ratio of atomic species (represented by their chemical symbols) together with the number of ring structures composed in the molecules (represented by integers). The occurrence is defined as the percentage of molecules containing the specific element or ring structure in the respective sampling group (group1 containing the total 45M molecules and group 2 containing the molecules classified as redox stable). The selection ratio, in turn, is defined as the percentage of molecules classified as redox-stable which contains the specific element or ring structure. This analysis provides insights on which type of molecular features that the AI recommended for generating higher redox stability. As can be seen, the selection ratio for the ring structures follows the order 1(47.6%) > 2(39.1%) > 3(31.0%), indicating that the increase in the number of rings leads to an increase in the likelihood of the molecule to degrade upon the redox reaction. On the other hand, the halogens tend to favor redox stability with the selection ratio following the trend Br (51.4%) < Cl (52.9%) < F (78.9%). This trend correlates well with the variation of the electronegative, first ionization potential and atomic radius down the halogen group in the periodic table. In the case of the heteroatoms S, N and O, the following trend was found: S (45,5%) > N (41.9%) > O (39,6%). Thus, they tend to favor the redox instability since they display selection rates below 50%.

After the high-throughput screening, DFT calculations were performed for all the selected compounds following the same procedure described for the OMEAD database (see the Supporting Information). The DFT-evaluated reduction potentials were employed to obtain the open-circuit voltages through the LM. Fig. 4 shows a collection of benchmarks evaluating the AI model performance, comparing AI- and DFT-based results through a probability density function (PDF). This PDF is a Gaussian fit to the respective data distribution, indicating the likelihood of observing a certain outcome. Fig. 4 (a), (b) and (c) shows the PDF of reduction potentials for the Na, Li and K cases, respectively. Similarly, Fig. 4 (d), (e) and (f) shows the open-circuit voltages' PDFs for
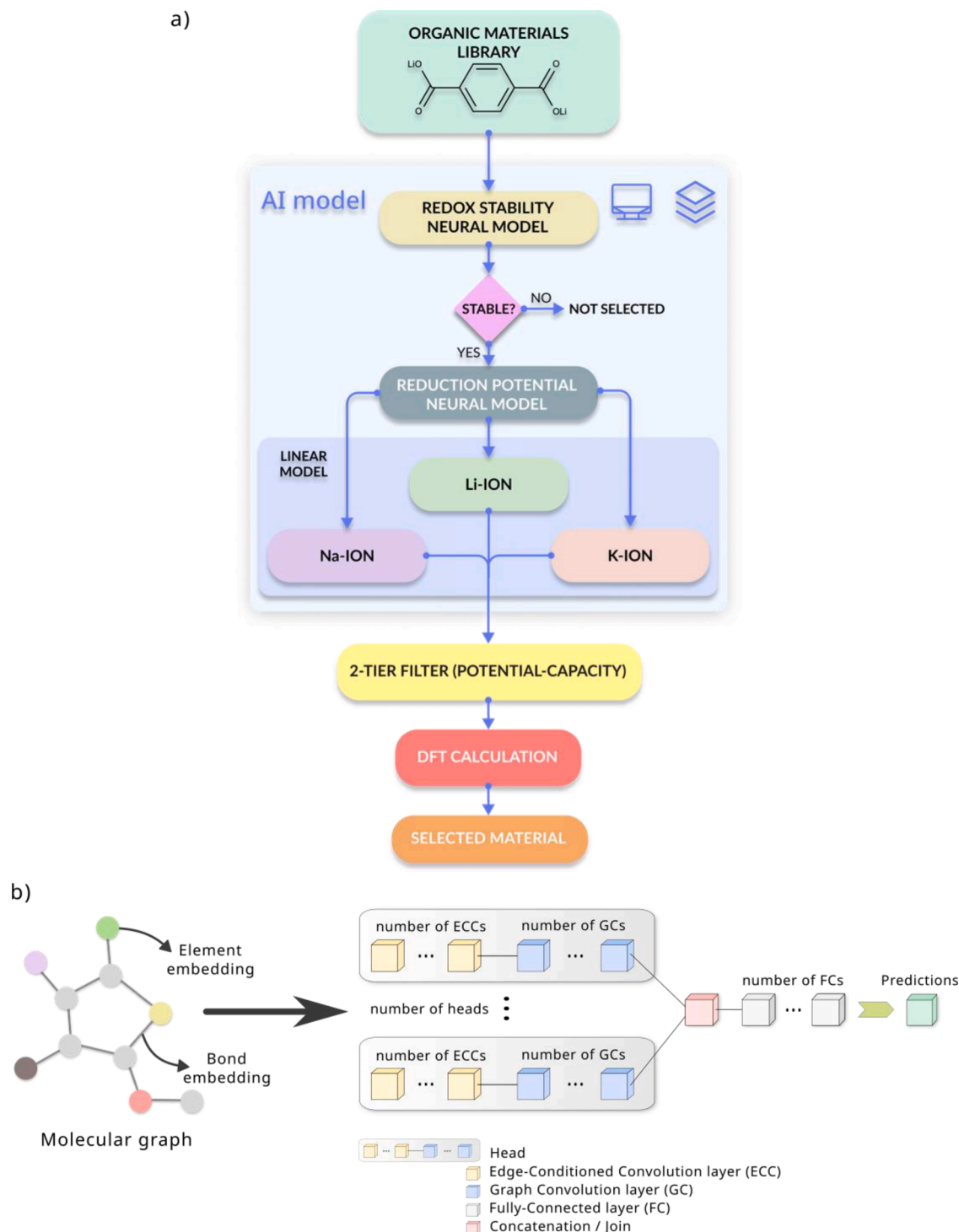
Fig. 2. (a) The AI model summarized in a flowchart, showing the different sub-layers and functionalities. (b) Schematics showing the overall Redox Stability Neural Model.

the same ion-type batteries. Overall, the DFT and AI values are in close agreement, with broader distributions and averages shifting to lower potentials for DFT results. This is due to the fact that the molecular geometry can change significantly after the relaxation step, especially for reduced phases, which largely contributes to the broader PDF distributions and lower potentials. The AI, on the other hand, has no mechanisms to anticipate these molecular changes, which indeed requires a more sophisticated quantum mechanics-based approach. Nevertheless, the model was able to correctly identify high-potential electrodes for all three scenarios/ions. Finally, the RSNM accuracy in

predicting the redox stability of the selected molecules was 74.4% when compared with the DFT findings, i.e., molecules that degraded during the DFT reduction reaction simulation. Fig. 4 (h) shows the confusion matrix for the Li^+ case to further evaluate the model's performance, from which it is possible to derive an F1-score of 83%, precision of 88.3% and recall of 78% when classifying redox stable molecules. This metric represents a harmonic mean of the model's precision and recall and is a traditional way to benchmark a classification model's performance. In other words, it combines the likelihood of the model in correctly identifying true positives (precision) and in finding all the
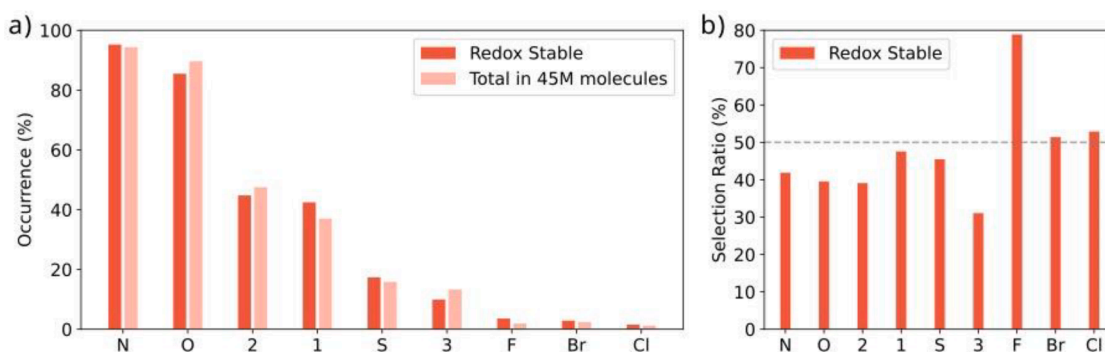
Fig. 3. Occurrences (%) on the left representing the amount of molecules containing the referred element (or number of ring structures; 1, 2 or 3). On the right, selection ratios (%) representing how often those elements are found, normalized by their occurrences. The x-axes shows the chemical symbols and the number of ring structures within the molecule.

possible positives (recall). Therefore, the AI is shown to be useful for filtering most of the unsuited molecules for redox active applications, such as batteries, when screening large materials libraries.

It should be highlighted that such redox stability neural model could be further optimized, to increase its accuracy, by using a significantly larger materials library (with a wider chemical diversity) in the training process. This type of models includes a large number of parameters to be optimized, and are therefore *data hungry*. However, the outcome obtained here is sufficient to prove the suitability of the developed model (including the criteria of classifying redox stability, the molecular graphs representation, the graph-based Neural Network and the evolutionary approach) to avoid redox unstable molecules during the high-throughput screening.

Nonetheless, a list of 3202, 689 and 702 novel molecules for Li-, Na- and K-ion batteries, respectively, is here being proposed, considering the redox stable molecules that remained after the DFT filtering process. Fig. 4 shows some of these molecules, followed by their respective open-circuit voltages for K-, Na- and Li-ion batteries. Some of these compounds exhibit interesting working potentials with different redox sites, like carbonyl groups, for example. Moreover, molecules like G or J in Fig. 4 can offer high gravimetric energy densities if all the available redox sites would be electroactive. Furthermore, the AI model identified compounds containing electron-withdrawing groups such as -Cl, -Br, -F and $\equiv$N, commonly employed as voltage enhancers, combined with electron-donating groups such as amines and methyl groups. This combination was pointed out in our previous work as a possible factor contributing to the higher insertion voltages displayed by these molecules. Interestingly, most of the compounds classified as redox stable by this approach contains one or more molecular rings. In fact, chain-based molecules have a higher rotational freedom than ring structures, which can facilitate the molecule to bend into a more favorable geometry for a given reaction to occur. In addition, p orbitals interacting in ring structures may also help to improve the overall stability of the molecule [41–46].

3. Conclusion

In this work, an AI-driven framework was developed for discovery of organic battery cathode materials, greatly expanding from previously proposed methodology. This novel "AI model" has here proven to be a powerful tool for finding such novel organic-based materials not only for Li-ion, but also for Na- and K-ion batteries. The workflow consists in three steps: (i) evaluating the redox stability of candidate molecules through a complex graph neural network architecture (the RSNM); (ii) predicting the reduction potential of redox stable molecules through a combination of natural language processing and neural networks (the RPNM); (iii) using the predicted reduction potential to assess the open-circuit voltage of the respective ion insertion reaction.

The framework was employed to screen 45 million small molecules in order to find novel cathode materials for Li-, Na- and K-ion batteries. From this step, 4047, 860 and 874 cathodes candidates, respectively, were identified by following voltage and capacity filters. All these molecules have been further analyzed through DFT calculations to better evaluate their properties and to benchmark the AI-driven method. From this step, it is possible to see the AI model overall performance. Additionally, the model has shown an accuracy of 74.4% and a F1-score of 83% when identifying redox stable compounds, showing its potential application in finding suitable molecules for redox-driven applications. Finally, a shortlist of 3202, 689 and 702 novel molecules for Li-, Na- and K-ion batteries is being proposed after considering the redox stable candidates (Fig. 5).

4. Computational Methods

Following the same methodology presented in our previous work, [16] the Organic Materials for Energy Applications Database (OMEAD) was expanded and now contains 41800 unique molecules and a list of high-level properties extracted from DFT simulations. Among these features, the redox potentials were obtained by evaluating the calculations to obtain the Gibbs free energy and optimized molecular geometries of the neutral, reduced and oxidized phases. Specific details about this step can be found in the Supporting Information. Similarly, the Reduction Potential Neural Model (RPNM) follows the same architecture first introduced in our previous work [16,47], but was here retrained using 27866 unique molecules and tested on 3800. In Fig. S2, the training and testing results show that the model has similar performance as previously reported.

The Linear Model (LM) for the Li-, Na- and K-ion open-circuit voltages were obtained following the same steps as presented in our previous works, i.e., statistically testing the model:

$$V_{OC}^{Ion} = \alpha x + \beta , \quad \text{(X)}$$

where V_{OC}^{Ion} is the open-circuit voltage of the respective ion insertion reaction happening in the electrode and x the independent variable representing the molecular reduction potential as obtained within the DFT calculation scheme. The parameters α and β were obtained through an ordinary least squares (OLS) regression.

The Redox Stability Neural Model (RSNM) was developed to predict the so-called *redox stability* of the molecular compounds. This characteristic is here presented as a yes/no classification, and the ground truth to train the prediction model was obtained by following changes in the bond length of neighboring atoms upon redox reactions simulations (from OMEAD) within the DFT framework. If the bond length of any neighbors in a cutoff radius of 2.1 Å varies more than 30% upon oxidation or reduction reaction, the compound is classified as redox unstable (class no) and redox stable otherwise (class yes). The 30% limit
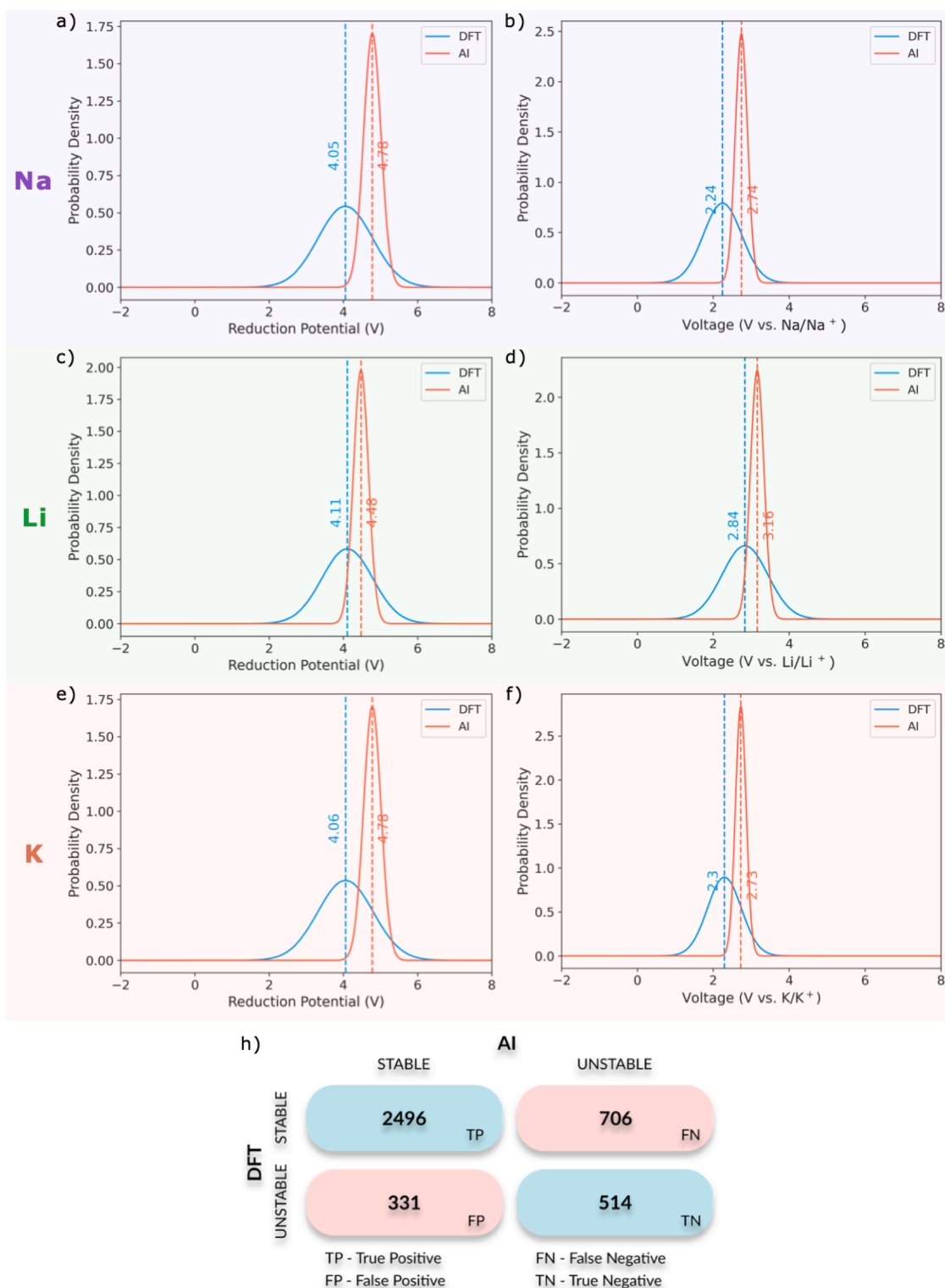
Fig. 4. Probability density function (PDF) plots of reduction potentials for the (a) Na-, (b) Li- and (c) K-ion cases comparing AI and DFT results. Similarly, PDF plots of open-circuit voltages for the (d) Na, (e) Li and (f) K ion insertion reaction. (h) Confusion matrix illustrating the classification model's performance.

was considered a good choice as it covers variations on the average covalent radii of atoms present in our dataset for multiple bonds [48–51]. Following this tight threshold, 9640 molecules were considered redox unstable. Furthermore, 9640 random redox stable molecules were extracted from OMEAD to form a well-balanced classification dataset for model training.

The RSNM was developed based on molecular graphs as inputs to a

Graph-based Neural Network. Graphs are mathematical structures encapsuling pairwise connections between different elements. These connections are often referred as *edges* and the elements as *nodes*. They are useful representation of molecules as there is a systematic way to encode information about the atoms as node elements and about the chemical bonds as edge elements. In this work, the information encoded in edges and nodes are embeddings [47] relative to the atom elements
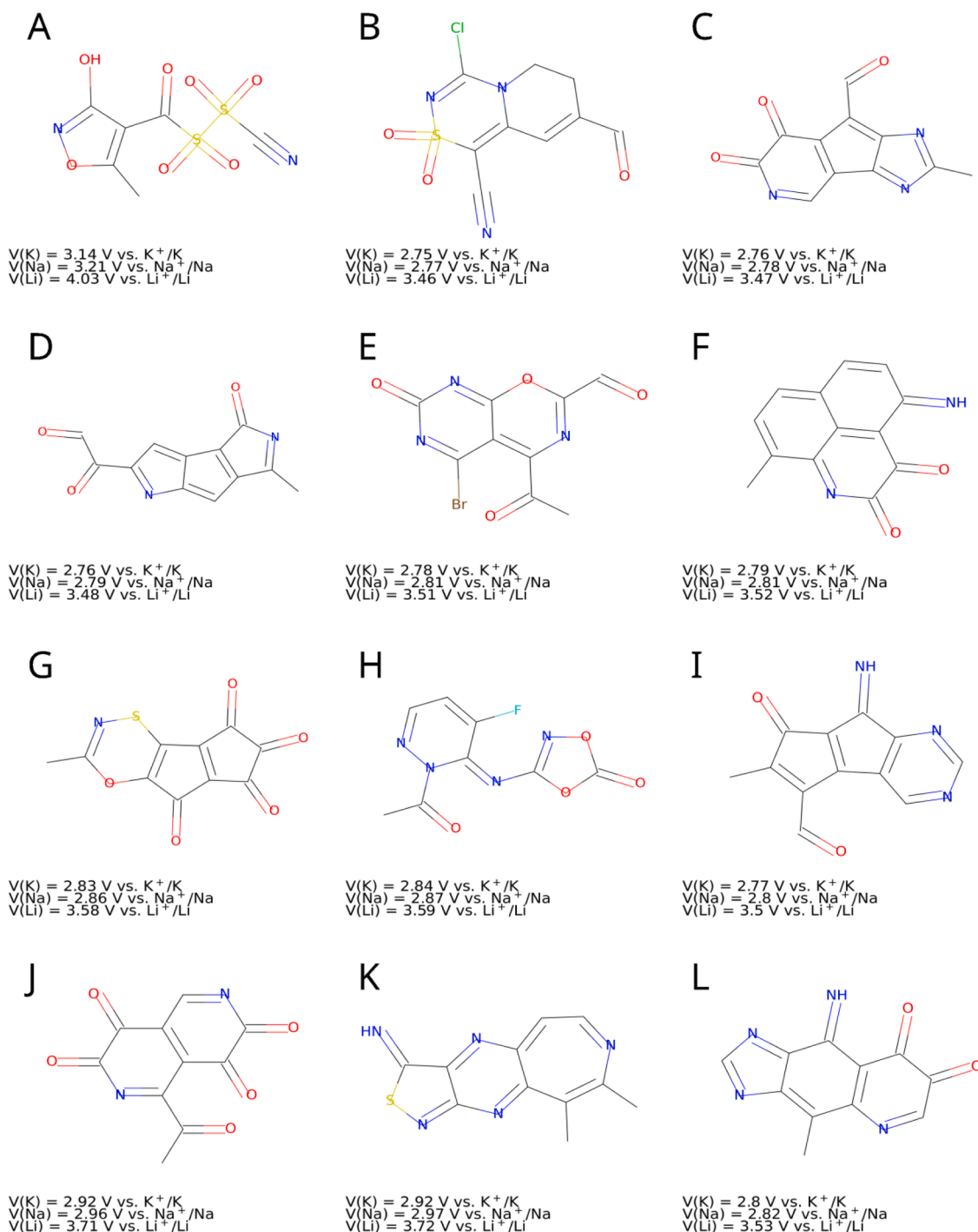
Fig. 5. Lewis representations of a few predicted molecules and theirs respective open-circuit voltages for K-, Na- and Li-ion batteries as predicted by the AI model.

and bonding types, respectively. For example, the fragment C=C is represented by two different graph nodes storing the *carbon embedding* with a graph edge connecting them and storing the *double bond embedding*. However, when working with mathematical graphs a special type of neural networks is required, the Graph Neural Networks (GNNs) [52, 53]. These are constructed to be compatible with graphs as inputs for the network and can work in conjunction with other types of neural networks, like Fully-Connected, Recurrent, Convolutional, etc. In general, GNNs relies on a message passing (or neighborhood aggregation) scheme to convolute information from the graph nodes and edges. Eq. 2 summarizes this message passing operation:

$$x_i^{(l)} = \phi^{(l)}\left(x_i^{(l-1)}, \nabla_{j \in N(i)} \psi^{(l)}\left(x_i^{(l-1)}, x_j^{(l-1)}, e_{j,i}\right)\right), \quad (2)$$

in which $x_i^{(l)}$ is the information (or features) of the node i in the $l-$ layer, $e_{j,i}$ the edge features between nodes j and i, ϕ and ψ are differentiable functions, usually a subclass of neural networks, and ∇ represents an aggregation operation like a sum or a mean considering the $N(i)$ neighbors of the i-th node. In this work, we have employed the Graph Convolutional [53] (GC) and the Edge-Conditioned Convolutional [54, 55] (ECC) layers, as shown in the Fig. 2, combined with a set of Fully-Connected (FC) layers. For the GC, Eq. 2 assumes the following form:

$$x_i^{(l)} = \frac{1}{|N(i)|} \sum_{j \ \in N(i)} \left(\Theta^{(l)} \cdot x_j^{(l-1)} \right) + b^{(l)}, \tag{3}$$

where the fraction is a normalization factor that depends on the *i*-th node's neighborhood, Θ and b the weight and bias parameters, respectively, similarly to a FC layer. The ECC, on the other hand, is given by:

$$x_i^{(l)} = \Theta^{(l)} \cdot x_i^{(l-1)} + \sum_{j \ \in N(i)} x_j^{(l-1)} \cdot h(e_{ij}) + b^{(l)}, \tag{4}$$

in which h is a FC layer and e_{ij} are features of the edge connecting i and j nodes. The network architecture, however, was obtained through an evolutionary approach by employing a differential evolution-based (DE) [40,41] code to find the optimal set of parameters for this network. The DE works by randomly initializing a group of different solution candidates, named population, and then evolving it through the *mutation* and *recombination* evolution operations. During the evolution, new candidates are created, potentially rendering better results. The resulting network prediction accuracy on a test dataset was chosen as the metric to evaluate the solution candidate fitness. A total of 397 network evaluations were performed by this scheme to obtain the final RSNM, which was retrained one more time considering a dropout of 20% on all layers to avoid overfitting. The DE approach was chosen due to the challenging task of defining a suitable model architecture for this specific problem. In fact, other attempts resulted in poor accuracy models that were excluded from the present work. More details of this scheme can be found in the Supporting Information. It should be mentioned that methods like k-fold validation could further support the evaluation of our model. However, we here had to consider the additional computational demands associated with the implementation of the DE approach. For instance, if a common 10-fold validation were to be employed, a total of 3970 trainings would be required. Although the data-driven approach is significantly faster than the quantum mechanics-based counterpart, it still relies on heavy computational resources. Therefore, having a training/testing split with the same examples made the comparison between the evolution scheme more consistent. Moreover, we have also carried out DFT-calculations for all molecules predicted by the AI model. This additional tier in the materials screening reduce our dependence on the accuracy of the NN-model.

CRediT authorship contribution statement

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

Declaration of Competing Interest

There are no conflicts to declare.

Data availability

The data is shared through an external library, further detailed in the electronic supplementary material.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ensm.2023.102865.

References

[1] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Nature 559 (2018) 547–555, 2018 559:7715.
[2] S. Axelrod, D. Schwalbe-Koda, S. Mohapatra, J. Damewood, K.P. Greenman, R. Gómez-Bombarelli, Acc. Mater. Res. 3 (2022) 343–357.
[3] J. Kirkpatrick, B. McMorrow, D.H.P. Turban, A.L. Gaunt, J.S. Spencer, A.G.D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L.R. Castellanos, S. Petersen, A.W.R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, A.J. Cohen, Science 374 (2021) 1385–1389, 1979.
[4] J. Behler, Angew. Chem. Int. Ed. 56 (2017) 12828–12840.
[5] F. Brockherde, L. Vogt, L. Li, M.E. Tuckerman, K. Burke, K.R. Müller, Nat. Commun. 8 (2017) 1–10, 2017 8:1.
[6] R. Shi, S. Jiao, Q. Yue, G. Gu, K. Zhang, Y. Zhao, C. Kai Zhang, Exploration (2022), 0220066.
[7] D.C. Elton, Z. Boukouvalas, M.S. Butrico, M.D. Fuge, P.W. Chung, Sci. Rep. 8 (2018) 1–12, 2018 8:1.
[8] F. Häse, L.M. Roch, P. Friederich, A. Aspuru-Guzik, Nat. Commun. 11 (2020) 1–11, 2020 11:1.
[9] Y. Liu, B. Guo, X. Zou, Y. Li, S. Shi, Energy Storage Mater. 31 (2020) 434–450.
[10] A. Chen, X. Zhang, Z. Zhou, InfoMat 2 (2020) 553–576.
[11] Y. Liu, T. Zhao, W. Ju, S. Shi, J. Materi. 3 (2017) 159–177.
[12] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S. Ping Ong, C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S.P. Ong, Adv. Energy Mater. 10 (2020), 1903242.
[13] Z. Wang, Q. Wang, Y. Han, Y. Ma, H. Zhao, A. Nowak, J. Li, Energy Storage Mater. 39 (2021) 45–53.
[14] A. Wang, Z. Zou, D. Wang, Y. Liu, Y. Li, J. Wu, M. Avdeev, S. Shi, Energy Storage Mater. 35 (2021) 595–601.
[15] Z. Wang, Y. Han, J. Cai, S. Wu, J. Li, Energy Storage Mater. 45 (2022) 1201–1211.
[16] R.P. Carvalho, C.F.N. Marchiori, D. Brandell, C.M. Araujo, Energy Storage Mater. 44 (2022) 313–325.
[17] G.H. Gu, J. Noh, I. Kim, Y. Jung, J. Mater. Chem. A Mater. 7 (2019) 17096–17117.
[18] X. Zhu, D.C.W. Tsang, L. Wang, Z. Su, D. Hou, L. Li, J. Shang, J. Clean. Prod. 273 (2020), 122915.
[19] I. Tsamardinos, G.S. Fanourgakis, E. Greasidou, E. Klontzas, K. Gkagkas, G. E. Froudakis, Microporous Mesoporous Mater. 300 (2020), 110160.
[20] Y. Wang, B. Seo, B. Wang, N. Zamel, K. Jiao, X.C. Adroher, Energy AI 1 (2020), 100014.
[21] P. Poizot, F. Dolhem, Energy Environ. Sci. 4 (2011) 2003–2019.
[22] P. Poizot, F. Dolhem, J. Gaubicher, Curr. Opin. Electrochem. 9 (2018) 70–80.
[23] B. Esser, F. Dolhem, M. Becuwe, P. Poizot, A. Vlad, D. Brandell, J. Power Sources 482 (2021), 228814.
[24] H. Chen, M. Armand, G. Demailly, F. Dolhem, P. Poizot, J.M. Tarascon, ChemSusChem 1 (2008) 348–355.
[25] C.P. Grey, J.M. Tarascon, Nat. Mater. 16 (2017) 45–56.
[26] S. Renault, D. Brandell, K. Edström, ChemSusChem 7 (2014) 2859–2867.
[27] D. Larcher, J.M. Tarascon, Nat. Chem. 7 (2015) 19–29.
[28] H. Yang, J. Lee, J.Y. Cheong, Y. Wang, G. Duan, H. Hou, S. Jiang, I.D. Kim, Energy Environ. Sci. 14 (2021) 4228–4267.
[29] D. Wilkinson, M. Bhosale, M. Amores, G. Naresh, S.A. Cussen, G. Cooke, ACS Appl. Energy Mater. 4 (2021) 12084–12090.
[30] L. Ruddigkeit, R. van Deursen, L.C. Blum, J.L. Reymond, J. Chem. Inf. Model. 52 (2012) 2864–2875.
[31] S. Lee, G. Kwon, K. Ku, K. Yoon, S.K. Jung, H.D. Lim, K. Kang, Adv. Mater. (2018) 30.
[32] R.R. Kapaev, P.A. Troshin, J. Mater. Chem. A Mater. 8 (2020) 17296–17325.
[33] W. Zhang, W. Huang, Q. Zhang, Chem. A Eur. J. 27 (2021) 6131–6144.
[34] S. Xu, Y. Chen, C. Wang, J. Mater. Chem. A Mater. 8 (2020) 15547–15574.
[35] Y. Chen, W. Luo, M. Carter, L. Zhou, J. Dai, K. Fu, S. Lacey, T. Li, J. Wan, X. Han, Y. Bao, L. Hu, Nano Energy 18 (2015) 205–211.
[36] Y.F. Shen, D.D. Yuan, X.P. Ai, H.X. Yang, M. Zhou, Electrochem. Commun. 49 (2014) 5–8.
[37] M. Lee, J. Hong, J. Lopez, Y. Sun, D. Feng, K. Lim, W.C. Chueh, M.F. Toney, Y. Cui, Z. Bao, Nature Energy 2 (2017) 861–868, 2017 2:11.
[38] Y. Xu, M. Zhou, Y. Lei, Mater. Today 21 (2018) 60–78.
[39] G.A. Pinheiro, J. Mucelini, M.D. Soares, R.C. Prati, J.L.F. da Silva, M.G. Quiles, J. Phys. Chem. A 124 (2020) 9854–9866.
[40] R. Storn, K. Price, J. Glob. Optim. 11 (1997) 341–359, 1997 11:4.
[41] K.v. Price, Intell. Syst. Ref. Libr. 38 (2013) 187–214.
[42] Y. Shao, G.N. Taff, S.J. Walsh, IEEE Geosci. Remote Sens. Lett. 8 (2011) 113–117.
[43] L. Prechelt, 1998, 55–69.
[44] C. Heinemann, T. Müller, Y. Apeloig, H. Schwarz, J. Am. Chem. Soc. 118 (1996) 2023–2038.
[45] R.F.W. Bader, H.J.T. Preston, Int. J. Quantum Chem. 3 (1969) 327–347.
[46] W. England, K. Ruedenberg, Theor. Chim. Acta 22 (1971) 196–213.

[47] S. Bengio, G. Heigold, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2014, pp. 1053–1057.

[48] B. Cordero, V. Gómez, A.E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, S. Alvarez, Dalton Trans. (2008) 2832–2838.

[49] P. Pyykkö, M. Atsumi, Chem. Eur. J. 15 (2009) 186–197.

[50] P. Pyykkö, M. Atsumi, Chem. Eur. J. 15 (2009) 12770–12779.

[51] P. Pyykkö, S. Riedel, M. Patzschke, Chem. Eur. J. 11 (2005) 3511–3520.

[52] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, IEEE Trans. Neural Netw. 20 (2009) 61–80.

[53] T.N. Kipf, M. Welling, in: Proceedings of the 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2023, https://doi.org/10.48550/arxiv.1609.02907.

[54] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, in: Proceedings of the 34th International Conference on Machine Learning 3, ICML 2017, 2017, pp. 2053–2070.

[55] M. Simonovsky, N. Komodakis, in: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 29–38, 2017-January.