UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2177

Organic Electrode Battery Materials

A Journey from Quantum Mechanics to Artificial Intelligence

RODRIGO P. CARVALHO

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

ISSN 1651-6214
ISBN 978-91-513-1571-3
URN urn:nbn:se:uu:diva-481583

Dissertation presented at Uppsala University to be publicly examined in Room 2001, Ångströmlaboratoriet, Lägerhyddsvägen 1, Uppsala, Friday, 30 September 2022 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Prof. Wolfgang Wenzel (Institute of Nanotechnology, Karlsruhe Institute of Technology).

Abstract

Carvalho, R. P. 2022. Organic Electrode Battery Materials. A Journey from Quantum Mechanics to Artificial Intelligence. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2177. 75 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1571-3.

Batteries have become an irreplaceable technology in human life as society becomes progressively more dependent on electricity. The demand for novel battery technologies has increased fast, especially with the popularisation of different portable devices. However, the current battery industry relies heavily on non-renewable resources that are also prone to provoke environmental harm. Among the possible candidates for the next generation of batteries, organic electroactive materials (OEMs) have become attractive due to a series of advantages: vastly accessible from renewable raw materials; highly versatile due to the possible functionalisation mechanisms; possibly lower production costs; reduced environmental impacts; etc. Nevertheless, some drawbacks need to be overcome before OEMs become competitive. Issues with energy density, rate capability and cycling stability hinder their final technological application. This thesis thereby discusses fundamental aspects of OEMs and proposes novel techniques to accelerate the materials discovery process.

The first part of this thesis presents a pathway to systematically investigate organic materials by combining quantum mechanics calculations and crystal structure predictions. An evolutionary algorithm predicts the crystal structure of several OEMs, enabling an initial assessment of the electronic structure and the thermodynamics of the ionic insertion mechanism in these compounds. Furthermore, this first part also suggests an approach to tailor OEMs, identifying their charge storage limits and the possible occurrence of metastable phases during the ion insertion process. However, the presented strategy, while accurate, is seriously limited by its high computational demands, which are unrealistic for high-throughput screening of novel materials.

Since organic materials represent a possibly limitless universe of compounds, alternative techniques are needed. Thus, the second part of this thesis combines quantum mechanics and artificial intelligence (AI), rendering a powerful platform to aid this task. An "AI-\textit{kernel}" was employed to analyse millions of organic compounds, discovering novel possible organic battery materials. Moreover, the AI accurately identified common functional groups associated with higher-voltage electrodes and suggested features that may aid future materials design. Furthermore, the kernel can also identify materials suitable for Na- and K-ion batteries and anticipate their redox stability.

In conclusion, this thesis has focused on investigating fundamental properties of organic electroactive materials, particularly the ionic insertion process in batteries. Furthermore, AI-driven methodologies have also been proposed, accurately evaluating OEMs and enabling fast access to the gigantic organic realm when searching for novel battery electrode materials.

Keywords: Batteries, Artificial Intelligence, Organic electrodes, High-voltage cathode materials, Machine learning, Materials discovery, High throughput screening

Rodrigo P. Carvalho, Department of Physics and Astronomy, Materials Theory, Box 516, Uppsala University, SE-751 20 Uppsala, Sweden. Department of Chemistry - Ångström, Structural Chemistry, Box 538, Uppsala University, SE-751 21 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1571-3
URN urn:nbn:se:uu:diva-481583 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-481583)

"Thus, the task is not so much to see what no one has yet seen, but to think what nobody has yet thought concerning that which everybody sees"
Arthur Schopenhauer

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Carvalho, R. P.**, Marchiori, C. F. N., Brandell, D. & Araujo, C. M. *Understanding the lithiation limits of high-capacity organic battery anodes by atomic charge derivative analysis.* (Submitted to J. Chem. Phys.)

II **Carvalho, R. P.**, Marchiori, C. F. N., Oltean, V.A., Renault, S., Willhammar, T., Gómez, C. P., Araujo, C. M. & Brandell, D. *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(3), 1024-1034 (2021).

III **Carvalho, R. P.**, Marchiori, C. F. N., Brandell, D. & Araujo, C. M. *Tuning the Electrochemical Properties of Organic Battery Cathode Materials: Insights from Evolutionary Algorithm DFT Calculations.* ChemSusChem 13, 2402–2409 (2020).

IV **Carvalho, R. P.**, Alhanash, M., Marchiori, C. F. N., Brandell, D. & Araujo, C. M. *Exploring metastable phases during lithiation of organic battery electrode materials.* ChemSusChem 15, e202200354 (2022).

V **Carvalho, R. P.**, Marchiori, C. F. N., Brandell, D. & Araujo, C. M. *Artificial intelligence driven in-silico discovery of novel organic lithium-ion battery cathodes.* Energy Storage Materials 44, 313-325 (2022).

VI **Carvalho, R. P.**, Brandell, D. & Araujo, C. M. *An AI-kernel discovering redox-stable organic electrode materials for alkali-ion batteries*. (in Manuscript).

Reprints were made with permission from the publishers.

Disclaimer: Certain parts of this thesis are based on my licentiate thesis entitled "In silico Design of Organic Battery Materials: *A Journey from Quantum Mechanics to Artificial Intelligence*" (Uppsala University, 2020).

Comments on my contribution to the papers in this thesis:

In all the Papers, I participated actively in the design of the project and the scientific discussions, performed most of the calculations and data analysis and was the main responsible for writing the manuscript. Particularly for Papers IV, V and VI, I have worked on developing two novel software, later published in an open-source basis in connection with the respective research articles.

List of publications not included in the thesis

I Marchiori, C. F. N., **Carvalho, R. P.**, Ebadi, M., Brandell, D. & Araujo, C. M. Understanding the Electrochemical Stability Window of Polymer Electrolytes in Solid-State Batteries from Atomic-Scale Modeling: The Role of Li-Ion Salts. Chemistry of Materials 32, 7237–7246 (2020).

II **Carvalho, R. P.**, Marchiori, C. F. N., Araujo, C. M. & Brandell, D. Chapter 3 Atomic-scale Modelling of Redox-active Organic Molecules and Polymers for Energy Applications in Redox Polymers for Energy and Nanomedicine 93–136 (The Royal Society of Chemistry, 2021).

III Cavallo, C., Calcagno, G., **Carvalho, R. P.**, Sadd, M., Gonano, B., Araujo, C. M., Palmqvist, A. & Matic, A. Effect of the niobium doping concentration on the charge storage mechanism of mesoporous anatase beads as anode for high-rate Li-ion batteries. ACS Applied Energy Materials 4(1), 215-225 (2020).

IV Liu, H., Zhu, J., Tian, D., **Carvalho, R. P.**, Shi, Z., Cai, Z., Chang, X., Araujo, C.M., Zhou, Y. & Zhu, J. 3D Lattice-Matching Layered Hydroxide Heterostructure with Improved Interfacial Charge Transfer and Ion Diffusion for High Energy Density Supercapacitor. Advanced Materials Interfaces 8(14), 2100429 (2021).

Contents

Abbreviations

AI	Artificial intelligence
AIMD	Ab initio molecular dynamics
ANIMA	AdvaNced artificial Intelligence for Materials reseArch
CNN	Convolutional neural networks
DFT	Density functional theory
DL	Deep learning
DOS	Density of states
EA	Evolutionary algorithm
ECC	Edge-conditioned convolution
FCNN	Fully connected neural networks
GC	Graph convolution
GGA	Generalized gradient approximation
GNN	Graph neural networks
LDA	Local density approximation
LIB	Lithium-ion battery
MAE	Mean absolute error
MAP	Mapion – Potential mapping
ML	Machine learning
NLP	Natural language processing
NN	Neural networks
OCV	Open-circuit voltage
OEM	Organic electroactive material
OLS	Ordinary least squares
OMEAD	Organic materials for energy applications database
PAW	Projector augmented wave
PDF	Probability density function
PES	Potential energy surface

RNN	Recurrent neural networks
SMILES	Simplified molecular-input line-entry system
TMO	Transition metal oxide
XRD	X-ray diffraction

1. Introduction

Energy is one of the most important aspects of human society nowadays and is not only a matter of production but also of storage and distribution. It directly affects the economy, society and the environment. Moreover, the modern human way of life is remarkably dependent on electrical energy, especially with the advent of highly connected devices and electrical vehicles where energy storage and portability have been playing a central role. On top of this, the harsh environmental changes that the planet is facing[1] imposes an urgent demand for alternative technologies that must be aligned with sustainable and environment-oriented philosophies. In this regard, the current ion-based battery industry still has a lot of room for improvement. A battery is composed of several components, from which the positive and negative electrodes are of utmost importance for the overall energy storage performance. During the battery operation, the cation (Li^+, Na^+, K^+, etc.) is transferred between the electrodes in a process controlled by several properties. Among them, the electrode's structure and chemistries.

Figure 1.1 shows a simplified sketch of this mechanism, illustrating only a few components of a battery. When an external power supply (load) is connected to the battery electrodes, electrons and ions are transferred from the cathode (anode) to the anode (cathode). In general, the mainstream state-of-the-art technology for Li-ion batteries (LIBs), for example, is based on graphite and transition metal oxides (TMOs) for the negative and positive electrodes, respectively. These are usually extracted from mining processes, thus a non-renewable resource. Furthermore, the processing steps during the battery life-cycle, i.e., from TMO mining to an uncertain recycling at the battery end-of-life, raise several environmental and economical concerns.[2] Hence, a next-generation of battery technologies is a genuine requirement.

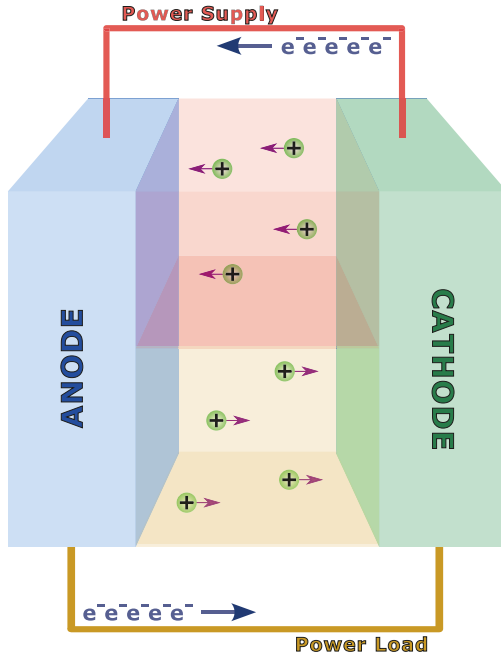


Figure 1.1. A simplified overview of a battery and the charge transfer processes during charge/discharge operations.

In this direction, organic materials have been arising as promising candidates. This class of compounds present numerous advantages:[3,4] a) they can be renewably produced from abundant raw materials like biomass; b) they present versatile synthesis routes by following well-established methodologies of organic chemistry; c) they allow tunable properties due to the numerous tailoring possibilities through functionalisation; to mention a few. In addition, the end-of-life treatment for organic battery materials may follow an easier path than inorganics, for example, by low-temperature combustion into synthesis precursors, which results in a closed life-cycle loop.[5]

When it comes to the battery operation, organics electrode materials generally function in a similar way to their inorganic counterparts, i.e., following the same general principles of redox reactions and ion insertion mechanisms. However, the electrodes' active materials are comprised of small molecules or polymers that often include a redox-active functionalisation. While the former usually offer higher energy densities due to their lower molecular weight,[6] the latter can often present superior rate capabilities due to the faster kinetics of, for example, radical polymers[7–9] or the higher electrical conductivity of conducting polymers.[7,10,11] The general low energy density of conventional electrically conductive polymers can be addressed by different functionalisation techniques or, alternatively, using different forms of organosulfur compounds.[12,13] This latter class of materials, however, usually

relies on reversible bond breaking and recombination processes, often resulting in lower rate capabilities due to poor kinetics.

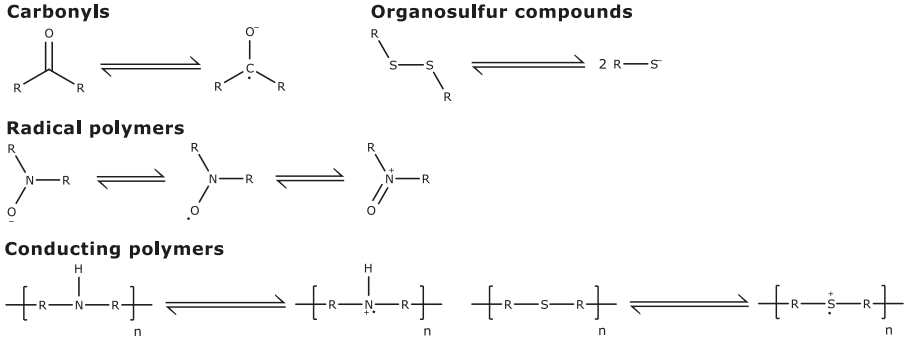


Figure 1.2. A simple example of the redox mechanisms of typical organic electroactive materials for battery electrodes.

Conjugated carbonyl compounds,[4] on the other hand, are a good example of low M_w molecules for battery materials with versatile chemistry. They have been vastly investigated since their electrochemistry was first report in 1969[14] with several examples in recent scientific literature. In general, the reaction mechanism in these compounds relies on the reversible reduction of the double-bonded oxygen of the carbonyl groups, which classify them as *n-type* materials. Figure 1.2 exemplifies the redox mechanisms for some of these classes of compounds. Overall, OEMs can be categorised into three types depending on how their charge state changes during the redox reactions:

1. *p-type*: the OEM is electrochemically active when it is oxidized from the neutral state.
2. *n-type*: the OEM is electrochemically active when it is reduced from the neutral state.
3. *bipolar*: both positively and negatively charged states are electrochemically active.

This is a material-intrinsic characteristic, also connected with its underlying electronic structure. Additionally, this further affects the so-called redox-active center, which is the spatial region where the charge tends to localise after the oxidation/reduction reaction.[13,15] During the last decade, several organic electroactive materials (OEMs) have been proposed as alternatives in LIBs. For example, dilithium terephthalate (Li_2TP) has received special attention after the pioneering work of Tarascon and co-workers.[16] This OEM can be straightly obtained by recycling the polyethylene terephthalate (PET) plastic, displaying a reversible capacity of 300 mAh/g and a lithiation potential of about 0.8 V vs. Li^+/Li.[16] Alternatively, higher-voltage

electrodes have also been suggested, like the quinone derivative dilithium-oxyterephthalate with a lithiation voltage of about 2.5 V vs. Li^+/Li,[17] further increased to about 3.5 V vs. Li^+/Li after engineering the molecule with Mg^{2+} ions. It is also worth to mention the radical polymer poly(2,2,5,5-tetramethyl-3-oxiranyl-3-pyrrolin-1-oxyl ehtylene oxide) (PTMA), which contains the tetramethylpiperidine-1-oxyl (TEMPO) stable nitroxyl radical. This OEM presents a fast electrochemical process due to the radical polymer nature and an open-circuit voltage of about 3.5 V vs. Li^+/Li.[18,19]

Nonetheless, further improvements need to be done until OEM could be economically explored for LIBs. This class of materials still present a few drawbacks, most related to cyclability issues and lower energy densities. In general, the electrochemical performance of these compounds can be engineered by a molecular design approach. Different electron donating/withdrawing groups can be employed to functionalize the material, with the effect of decreasing/increasing the open-circuit voltage.[20] These groups can modify the de/localisation of charges upon the redox reaction, thus tweaking the material's electronic structure and, consequently, the redox-active center[Paper III]. Tuning aromaticity and conjugation in these materials can also affect their electrochemical performance, with significant changes in voltages and conductivity.[21] Polymerization of small molecules, often soluble in typical liquid electrolytes, can also be employed to achieve better cyclability, in exchange for lower capacities.[13] Most of these approaches, however, still suffer from some disadvantages. Therefore, the discovery of novel materials could offer a new paradigm in this scenario and introduce technological breakthroughs.

In this regard, this thesis first presents a systematic approach to deepen the understanding about fundamental aspects of the electrochemical process for ionic insertion into organic compounds. In Section 3.1, an initial investigation on the molecular level is carried out demonstrating the maximum electronic storage of these molecular reservoirs. This initial assessment introduces the novel charge derivative analysis, which sheds light on the "superlithiation" phenomenon reported in the literature.[22] However, a sole molecular analysis may be insufficient to reasonably describe the electrode's chemical environment "seen" by the inserted ion during the battery operation. Thus, a more accurate approach is necessary. This has been achieved by a combination between quantum mechanics calculations, by the means of Density Functional Theory (DFT), and different approaches to assess the crystal structure of the desired compound, such as evolutionary algorithms predictions and metastable phases analysis. Hence, this methodology has enabled the evaluation of key structure-property features, e.g., the electronic structure, thermodynamics of the ion insertion reaction and the structural changes induced by it – thereby systematically improving the overall knowledge of OEMs.

Despite being a powerful method, the evolutionary algorithm strategy still suffers from a severe drawback: the demanded computational efforts. The evolution process (described in Section 3.2) requires an impressive number of calculations to correctly predict the crystal structure for the energy global minimum. On the other hand, faster methods as the metastable phases analysis, proposed in this thesis through a novel software called Mapion, may be inadequate to describe the battery system when global minimum structures are required. This novel method is further discussed in Section 3.2. Therefore, substantial limitations exist when aiming for a high-throughput screening for novel materials. The universe of organic molecules, particularly, offers an almost limitless number of compounds.[23,24] To just give an idea of how gigantic the organic realm is, 166 billion of small organic molecules have recently been enumerated by just considering possible combinations of elements like C, N, O, S and halogens with up to 17 atoms.[25] This is an impressive number that imposes several limitations to be accessed, both experimentally and theoretically through quantum mechanics. As a comparison, it is estimated that there are 100 billion stars in the Milky Way, from which 1.3 billion have recently been mapped by the Gaia mission.[26] Figure 1.3 displays a comparison between these numbers, with a closer zoom at a smaller range in this scale to reveal the Sigma-Aldrich 'in-stock' 14 million compounds from the Aldrich Market Select Chemistry Services.[27]

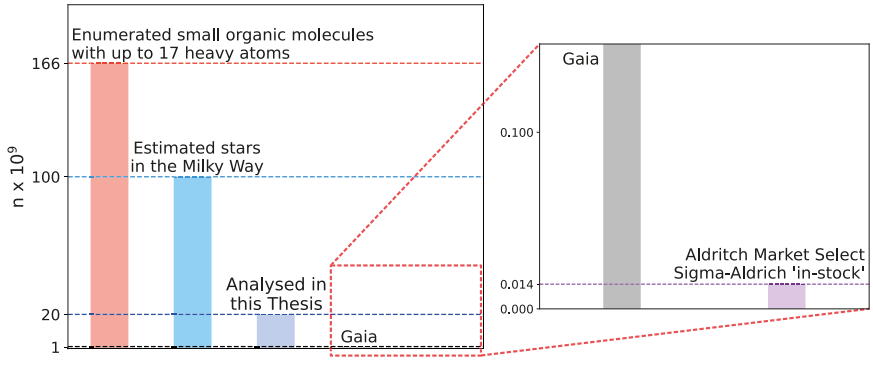


Figure 1.3. Illustration comparing the "size" of representative domains.

When facing this problem, the development of new methodologies is required to explore this vast universe and accelerate the discovery of novel materials with potentially enhanced properties for batteries. Naturally, data-driven approaches may offer interesting strategies for fulfilling this task. Recently, these statistically-based methodologies have been receiving a lot of attention due to modern revolutions in the field of data science, especially in processing and storing data. In particular, machine learning and artificial intelligence have been playing a major role in this scenario due to recent popularisations of these techniques, greatly fuelled by the release of open-source

frameworks from companies like Google and Facebook. Nevertheless, a lot of work is necessary to correctly apply a data-driven approach, like machine learning, in accessing the organic realm before it could be used to boost materials discovery. In this regard, this thesis has delved into this discussion in order to propose an innovative methodology based on machine learning and backed by quantum mechanics. Firstly, a database comprised of unique organic molecules has been developed with more than 41800 moieties. For each entry, a list of electronic properties extracted from quantum mechanics calculations has been included, also considering the oxidized and reduced moiety states.

Simultaneously, a data processing and machine learning software has been developed and released together with the database on an open-source basis. Finally, the framework has shown the capability of predicting the electrochemical performance of the Li^+ insertion reaction in organic materials. Thereafter, a high-throughput screening has been done on 20 million unique molecules, rendering a set of possible cathode candidates for LIBs – a task accomplished in a matter of hours by using a personal computer, but that would require years if attempted by means of quantum mechanics calculations carried out on supercomputers. A detailed discussion on all the construction aspects of this machinery, here named the "AI-*kernel*", is presented in the Section 4.1 of this thesis.

In accordance with the requirements needed to explore the huge space of organic materials, the constructed AI-*kernel* request solely a molecular SMILES string as input in order to evaluate its possible lithiation potential. Interestingly, the AI was capable to correctly associate high-voltage compounds with well-known functional groups used as voltage enhancers – a knowledge acquired independently during the learning process. Moreover, the AI has suggested an interesting combination of electron-withdrawing and electron-donating groups on the same molecule as possible cathode candidates. An effect that could, in principle, lead to a better localisation of the redox-active center on organic electroactive materials.

Furthermore, the AI-*kernel* was expanded to also predict the ion-insertion voltages of Na- and K-ion based batteries and likewise employed in high-throughput screening 40 million molecules in searching for high-voltage cathodes. However, predicting the performance of millions of OEMs by means of lithiation potentials alone, without evaluating their chemical stability, is far from ideal. These molecules can often show different degradation mechanisms during redox reactions taking place within the battery operation. Therefore, redox stability is a straightforward feature to analyse; here defined as the molecule's ability to maintain its overall structure after an oxidation/reduction step, i.e., to not degrade after an electron transfer process. In this direction, an additional AI-powered methodology has been developed to tap into the

complex matter of the stability analysis of organic compounds. With this new functionality, the AI-*kernel* can predict the redox stability of organic electroactive compounds for Li-, Na- and K-ion batteries, presenting a more complete design platform.

In light of these, the AI-*kernel* may be considered a powerful addition to the arsenal of technologies at our disposal to enable straightforward access into the organic realm, thus accelerating the materials discovery process.

2. Theoretical Framework

In the past century, a large set of theories have been developed to solve the Schrödinger equation (Equation 2.1) for molecular and solid-state systems in a combined effort of quantum chemistry and solid-state physics. More specifically, the problem of solving the Schrödinger equation for such systems is a many-body problem by definition and, thus, a challenging task. Several methods have arisen in the past years attempting to fulfil this quest and it may be an understatement that DFT drastically changed the stage in this regard. Based on the Hohenberg and Kohn theorems,[28] DFT arises as a formidable framework to back computational efforts in solving the Schrödinger equation for many-body systems. Furthermore, DFT can be processed with a much lower computational cost than techniques based on wavefunctions, e.g., Hartree - Fock,[29] Møller - Plesset,[30] configuration interaction,[31] etc.

$$H\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{2.1}$$

In this thesis, DFT is the main framework adopted for ab initio calculations in both molecules and solids, thus this section is partially dedicated for a brief explanation of it. The main computational software packages employed for calculations was the Vienna Ab initio Simulation Package (VASP)[32–34] and the Gaussian series of electronic structure programs.[35]

Even presenting an accurate framework with lower computational effort than other ab initio approaches, the use of DFT as a mean to explore and develop novel materials is still handicapped by computational limitations. Thereby, secondary methodologies are necessary. In this sense, the technical revolutions staged by computational science, both by the means of data processing and data storage, has been turning data-driven methods more available and more accurate. Nowadays, the use of supercomputing facilities to process a large number of systems has become quite popular and initiatives like the Materials Project[36] and the Open Quantum Materials Database[37] have alternatively presented a solution to store information regarding many materials. On top of this, there is a recent demand for the development of novel frameworks capable of processing these huge quantities of data and in turn learning structure-property relationships in order to promote breakthrough innovations in Materials Science. In this sense, Machine Learning is emerging as an interesting ally beyond conventional approaches like DFT[38] and molecular dynamics[39] to assess several properties of materials.[40–43] In this thesis,

an Artificial Intelligence alternative has been proposed by the means of Neural Networks and Deep Learning to design novel materials for organic batteries. Therefore, a short description of these methods will also be provided in this chapter.

2.1 Density Functional Theory

As stated before, the problem of solving the Schrödinger equation for molecules and solids constitutes a many-body problem. In general, when dealing with such systems the Hamiltonian in the Equation 2.1 should represent a system of interacting electrons and nuclei, illustrated, in atomic units, as:

$$H = -\frac{1}{2}\sum_i \nabla_i^2 - \frac{1}{2}\sum_A \nabla_A^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{r_{ij}} - \frac{1}{2}\sum_{i\neq A}\frac{Z_A}{r_{iA}} + \frac{1}{2}\sum_{A\neq B}\frac{Z_A Z_B}{R_{AB}}, \tag{2.2}$$

where the lower (upper) case subscripts run over electrons (nuclei) and r_{ij}, r_{iA}, R_{AB} and Z are, respectively, the electron-electron, electron-nucleus and nucleus-nucleus distances and the atomic number. The first two terms account for the kinetic energy of electrons and nuclei while the remaining terms describe the electrostatic interaction between electron-electron, electron-nucleus and nucleus-nucleus, respectively. Hence, a clear challenging many-body system to be addressed. This sort of problem is not exclusive from quantum mechanics. In fact, Poincaré[44] pointed out in the nineteenth century that just a few classical mechanics problems can be solved analytically, mainly due to the famous three (or more) body problem. Therefore, approximations need to be made when dealing with Equation 2.2.

A first approximation is the simple, yet elegant, *adiabatic* (or Born - Oppenheimer) *approximation*[45] in which the nuclear motions are neglected, thus decoupling the electron-nucleus complex dynamical problem. In short, the absence of nuclear motion implies that no heat is exchanged with the outside environment. Therefore, the Hamiltonian now can be written as:

$$H = T_e + V_{ee}(\mathbf{r}) + V_{eN(\mathbf{r},\mathbf{R})}, \tag{2.3}$$

where T and V stand for the kinetic and electrostatic terms, respectively, and the indexes e and N for electron and nucleus. The adiabatic approximation relies on the assumption that the electron dynamics occur in a much faster way than the nuclear one due to the expressive mass difference between them, thus justifying the decoupling. Henceforth, the problem is reduced to the electronic motion in an external potential, exerted by the nuclei. Although the nuclear

positions now enter in Equation 2.3 as parameters, the many-body problem remains. In wavefunction-based methods, the eigenfunctions of Equation 2.1 are represented as a combination of some basis function – regardless of other approximations – and the problem is solved following the Rayleigh - Ritz variational principle.

In 1927, Thomas suggested a simple concept of using the electronic density as a basic variable and expressing the potential as a functional of this density, instead of orbitals, thus introducing the notion of density functional theory.[46] His first efforts in this direction – combined with the work of Fermi[47] – derived *the Thomas - Fermi model*. For being based on the assumption that a uniform electron gas could be the solution of the Equation 2.1, the model failed in representing simple molecular and solid systems for not considering other intrinsic effects. Furthermore, there was no proof of the density functionals existence and the uniqueness of solutions. In 1964 the Hohenberg - Kohn theorems emerge as the basis of the density functional theory, allowing the solution of the electronic problem in terms of the electronic density.[28] Such theorems are stated as:

THEOREM 1: *Given the problem of interacting particles in an external potential, there is a one-to-one correspondence between this potential and the electronic density of the ground state* $\rho(\mathbf{r})$.

This first theorem ensures that the external potential, and thus the Hamiltonian ground state, can be uniquely constructed by the electron density. Therefore, all the system information on the ground state is included in the density and the expectation value of a physical observable $\mathcal{O}$ can be written as a density functional. For the total energy:

$$E = \langle \Psi | E | \Psi \rangle = E[\rho(\mathbf{r})], \tag{2.4}$$

where the electron density should sum to the total number N of electrons on the system, such as:

$$\int \rho(\mathbf{r}) d\mathbf{r} = N \tag{2.5}$$

THEOREM 2: *It is always possible to establish a universal total energy functional of the electronic density for any given external potential acting in an interacting system. The global minimum of this functional and the corresponding electronic density are, respectively, the exact total energy and density of the system ground-state. The total energy functional may be written as:*

$$E[\rho(\mathbf{r})] = F_{HK}[\rho(\mathbf{r})] + \int V_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}, \tag{2.6}$$

with F_{HK} being the Hohenberg - Kohn functional, defined as:

$$F_{HK}[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] \tag{2.7}$$

It is worth to emphasize that F_{HK} is independent of any information about nuclei, thus it is a universal functional for the electronic system. The second theorem offers a variational ansatz in determining the ground-state electronic density of the system since the functional in Equation 2.6 has a global minimum that is the exact ground-state total energy, i.e.,

$$\delta(E - \mu\rho(\mathbf{r})) = 0, \tag{2.8}$$

where μ is a Lagrangian multiplier. This leads to the following relationship:

$$V_{ext}(\mathbf{r}) + \frac{\delta F_{HK}}{\delta\rho(\mathbf{r})} = \mu \tag{2.9}$$

Although the huge advance in solving the many-body problem, the functionals T and V_{ee} appearing in F_{HK} remains unknown. Kohn and Sham[48] proposed a methodology to solve the Equation 2.9 by introducing orbitals in a fashion to compute the kinetic energy, presenting some similarities with the Hartree - Fock approach.[49] By introducing an auxiliary system of non-interacting electrons, in which the density is the same electron density of the interacting system, they proposed an accurate approximation to calculate the kinetic energy as:

$$T_S[\rho] = -\frac{1}{2}\sum_{i}^{N}\langle\phi_i|\nabla_i^2|\phi_i\rangle\,, \tag{2.10}$$

where ϕ are the wavefunctions of the non-interacting auxiliary system. The HK functional can now be written as:

$$F_{HK}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho], \tag{2.11}$$

with $J[\rho]$ representing the classic Coulomb interaction,

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}', \tag{2.12}$$

and $E_{XC}[\rho]$ the so-called exchange-correlation functional, which contains everything that is unknown and may arise from the approximations made so far. Moreover, they are quantum and many-body effects due to the nature of the real interacting electronic motion. The exchange effect, also known as the Fermi correlation, is due to the fermionic nature of electrons, i.e., electrons possessing parallel spin affect the motion of each other. The correlation energy, or the Coulomb correlation, arises from the collective effect of the electronic motion due to the intractable many-body Coulomb interaction, not taken into account by the auxiliary non-interacting system. In practice, the form of $E_{XC}[\rho]$ is unknown and must be approximated. The HK Equation 2.9 can now be written as:

$$V_{eff}(\mathbf{r}) + \frac{\delta T_S}{\delta\rho(\mathbf{r})} = \mu, \tag{2.13}$$

with the instruction of the effective potential:

$$\begin{aligned} V_{eff}(\mathbf{r}) =& V_{ext}(\mathbf{r}) + \frac{\delta J}{\delta\rho(\mathbf{r})} + \frac{\delta E_{XC}}{\delta\rho(\mathbf{r})} \\ =& V_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + V_{XC}(\mathbf{r}), \end{aligned} \tag{2.14}$$

where V_{XC} stands for the exchange-correlation potential that should also be interpreted as a second external potential (apart from the nuclear one) acting on the auxiliary non-interacting system. Following the variation principle in order to minimise the energy functional, one can obtain a Schrödinger-like equation as a condition for the existence of extremes:

$$\left(-\frac{1}{2}\nabla^2 + V_{eff}(\mathbf{r})\right)\phi_i = \varepsilon_i\phi_i, \tag{2.15}$$

which is a set of one-electron Kohn-Sham equations to be solved iteratively in a similar fashion to the Hartree or Hartree-Fock equations, e.g., by following

the self-consistent Roothaan approach.[29] The Lagrangian multiplier μ introduced here brings a curious interpretation, landed by Parr et. al..[50] First, by fixing the external potential, μ can be written as:

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{V_{ext}}, \tag{2.16}$$

which immensely resembles the thermodynamic chemical potential of an element i with n_i moles in a macroscopic system at given temperature T and pressure P:

$$\mu = \left(\frac{\partial G}{\partial n_i}\right)_{P,T,n_j(j\neq i)} \tag{2.17}$$

The relationship presented in Equation 2.16 indicates the escaping tendency of electrons from the system at equilibrium, which is an opposite concept to the electronegativity χ. Kohn, Parr and Becke underlined this analogy in 1996.[51] Furthermore, this definition of electronegativity can be expanded as an average of the ionization potential (I) and the electron affinity (A) by using finite differences:

$$\chi = -\mu = -\left(\frac{\partial E}{\partial N}\right) = \frac{1}{2}(I+A), \tag{2.18}$$

which is a useful interpretation in the batteries context.

Turning our attention back to the exchange and correlation functional $E_{XC}[\rho]$, several approximations have been proposed in the past years. Here, the most widespread Local Density Approximation (LDA) and Generalized Gradient Approximation (GGA) are going to be briefly described. On the former, a simple and clever local approximation to the homogeneous electron gas (HEG) of density $\rho(\mathbf{r})$ is introduced as a way to model the exchange and correlation effects. Thus, the LDA functional can be written, in terms of the HEG exchange-correlation energy density ε_{XC}^{HEG}, as:

$$E_{XC}^{LDA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{XC}^{HEG}[\rho]d\mathbf{r} \tag{2.19}$$

The exchange part of E_{XC}^{LDA} may also be defined as:

$$E_{XC}^{LDA} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int d\mathbf{r}\rho^{4/3}(\mathbf{r}), \tag{2.20}$$

which was proposed by Dirac in 1930 as a way to add the exchange energy in the Thomas-Fermi model, becoming the first exchange functional of the electron density.[52] Nonetheless, the analytical form of the correlation part is still unknown and must be parametrised. For being based on the HEG, the LDA is better suited for slowly varying electron densities. A correction scheme to this functional was first proposed by von Weizsäcker by using the gradient of the electron density, thus recognized as the first *gradient approximation*.[53] The use of gradient-based corrections is a clever attempt to enhance the LDA functional as it could better capture the variations of the electron density. Similarly, the generalized gradient approximation (GGA) adds a term $F_{XC}(\rho, \nabla\rho)$ responsible to capture fluctuations on the electron density as:

$$E_{XC}^{GGA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_X^{HEG}[\rho]F_{XC}(\rho, \nabla\rho)d\mathbf{r} \tag{2.21}$$

GGA is, in fact, a class of functionals with several ways to represent F_{XC}. Throughout this thesis, the approximation presented by Perdew, Burke and Enzerhof (PBE)[54] has been widely adopted. Despite their overall success, LDA and GGA still present the non-physical self-interacting problem.[55] While many approaches trying to address this issue do exist, here the well-known hybrid functionals are going to be used. In this scheme, a portion of the exact Hartree-Fock exchange is mixed with the exchange-correlation from LDA/GGA. The B3LYP functional, which mix the exchange proposed by Becke[56] and the correlation from Lee–Yang-Parr[57] with the HF exchange in a three-parameter equation, has been applied in this thesis for molecular systems when using the Gaussian series of software. For solid-state systems, the Heyd-Scuseria-Ernzerhof (HSE06) is going to be used.[58] In this functional, the short-range HF exact exchange is mixed with the short-range PBE exchange while the long-range exchange and correlation contributions come solely from PBE.

The van der Waals forces are another problematic poorly represented by these functionals. This weak interaction, proportional to $1/r^6$, arises from dipole-induced interactions and are of fundamental importance in molecular systems where dipoles may be permanent, induced or even a momentaneous consequence of random quantum fluctuations. As it is possible to foresee, these interactions are critical for organic molecular crystals and, thus, a necessary addition. Although there are many methodologies to treat the van der Waals interactions, in this thesis the Grimme-D2 scheme will be adopted.[59]

Furthermore, these corrections are switched off for the insertion ion specie when dealing with battery electrodes, i.e., for lithium and sodium ions, as the increasing amount of them in the reduced phases may lead to an error accumulation due to fundamental limitations of the DFT-D approach.

In practice, after defining the exchange-correlation functional, the wavefunctions on the KS one-electron equations (Equation 2.15) can be expanded in a basis set φ_μ with coefficients $c_{i,\mu}$ of the kind

$$\phi_i = \sum_\mu c_{i,\mu}\,\varphi_\mu, \tag{2.22}$$

and, therefore, transforming the KS equation into a matrix equation following the Roothaan method.[29] This Kohn-Sham-Roothaan equation can be written in a matrix form as:

$$\mathbf{F}^{KS}\mathbf{C}_i = \varepsilon_i \mathbf{S}\mathbf{C}_i, \tag{2.23}$$

and thus efficiently calculated self-consistently. Here, the $\mathbf{F}^{KS}$ represents the Kohn-Sham functional, $\mathbf{C}_i$ the set of coefficients to be obtained and $\mathbf{S}$ the overlap matrix. In this thesis, the φ functions were represented by a linear combination of Gaussian functions, namely the contracted Gaussian Functions,[60] employed in the Gaussian software for dealing with molecules. Alternatively, Plane-Waves (PW) was adopted when dealing with periodic systems, i.e., solids in the VASP software.

Although the PW introduces several facilities when dealing with solids, it also presents a new issue: the quick oscillations of the wavefunctions near the nuclei requires an alarming number of PW basis to be properly described, therefore imposing computational hindrances. In this regard, the Projected Augmented Wave (PAW) method[33,34] proposed the ingenious approach of mapping the real wavefunction (ϕ) into an auxiliary pseudo wavefunction $(\tilde{\phi})$ that should reproduce the same results while being less computationally demanding. The scheme starts by dividing the crystal into a non-overlapping augmented region Ω around nuclei (core) and an interstitial residual region elsewhere. In this interstitial space, the pseudo wavefunction should correspond to the real wavefunction, thus PW may be used. However, inside the augmented region the wavefunction is represented as a set of partial waves as the use of PW shall be avoided. These partial waves are chosen to be atom centered localised functions, e.g, spherical Bessel functions. With this basic concept, the PAW method circumvents the high oscillations of the true wavefunction near core regions while presenting a highly efficient way to solve

the KS equations. When referred to calculations regarding crystals by using VASP, this method is implicitly being used through this thesis.

2.2 Artificial Intelligence

Data-driven methods based on statistical theories is not a novelty in the world of science. The use of statistical analysis has been aiding scientists for decades into proving and probing all sort of scientific hypothesis and into predicting new concepts. Nonetheless, the past years witnessed numerous technical revolutions in how efficiently computer sciences store and process information, turning data-driven approaches as attractive as it was never before. As an example, a quick search on the Google Scholar platform for the keyword "artificial neural networks" rendered about 700000 results between the years 2000 and 2020. Correspondingly, Material Science also served as a stage for the uprising use of data-driven techniques, especially by the means of machine learning (ML) applications for both designing and understanding novel materials. ML has appeared in science through several different methodologies, e.g., linear and logistic regressions, support vector machines (SVMs), k-nearest neighbours (kNN), linear and quadratic discriminant analysis (LDA/QDA), decision trees, foresting models and artificial neural networks (NNs). Nevertheless, only recently that the latter experienced a surge in popularity, greatly due to open-source and accessible frameworks like TensorFlow[61,62] and PyTorch.[63,64] Furthermore, neural networks have established a path to access a field named Deep Learning (DL),[65] where the machines can achieve a profound knowledge about some given issue. In Material Sciences, NNs and DL have recently been emerging as a powerful tool to predict several properties of materials and to even by-pass quantum mechanics equations, aiding in the decisive task of designing novel materials. With this in mind, NNs are going to be the main ML approach explored through this thesis to uncover novel organic compounds for batteries and, therefore, this section will be dedicated to briefly describe them. Machine learning is considered a sub-domain of artificial intelligence (AI), in which the idea is to use the first in a fashion way to intelligently execute tasks. Thereby, the term AI and NNs will be constantly interchanged throughout this thesis when referring to the final framework developed.

In general, neural networks are described as a non-linear algorithm inspired by how the human brain is architectured. It is able to represent any kind of function with high fidelity, and thus capable to learn intrinsic correlation in a given dataset. The basic construction of NNs lies around the neuron (or perceptron), that is, a single unit of the network that represents a linear operation between input and output values. By stacking neurons in a layer-like structure these linear operations can be executed in an efficient matrix form, in which

each neuron i is represented by a w_i weight parameter. Likewise, stacking the layers in a network-like architecture renders the foundation of a general NN. So far, simply combining neurons (linear) will produce a limited linear model, therefore activation functions are introduced in the output of each layer to incorporate non-linearities. These activation functions may be of many different kinds and among the most famous one can find the sigmoid functions, ReLUs and tanh. A further revision on this topic can be seen on the reference.[66]

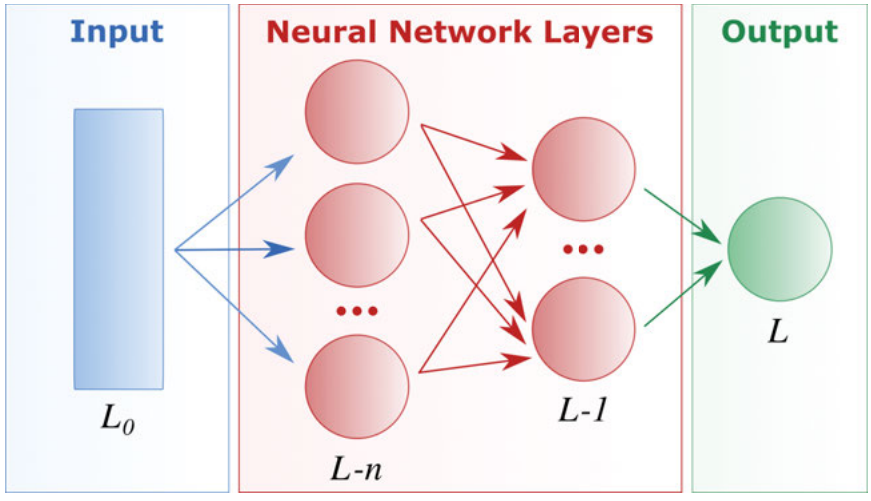


Figure 2.1. General architecture of a basic Neural Network. Circles represent neurons and arrows indicate information being forwarded through the network.

Figure 2.1 brings the general concept of a neural network, where each circle represents a neuron and arrows indicate information being transmitted between layers, i.e., being forwarded among neurons of different layers. For a simple neuron i, this input-output operation is represented, omitting the activation function, by:

$$[\text{output}]_i = w_i[\text{input}] + b_i, \tag{2.24}$$

which is a simple linear equation, also known as the *perceptron model* when considering the activation function. For the complete network, the Equation 2.25 summarizes this forward process in the matrix form:

$$\mathbf{z}^{(L)} = \sigma\left(\mathbf{W}^{(L)}\mathbf{z}^{(L-1)} + \mathbf{b}^{(L)}\right), \tag{2.25}$$

where $\mathbf{z}^{(L)}$, $\mathbf{W}^{(L)}$ and $\mathbf{b}^{(L)}$ are the output, the weight matrix and the bias of the

L-th layer and $\mathbf{z}^{(L-1)}$ the output of the previous layer. The $\mathbf{W}$ matrix contains all the i-th weight parameters of its respective layer – these are the ones to be adjusted during the learning process along with the b-th bias terms. The linear operation inside parentheses, then, passes through the activation function σ, responsible for adding the desired non-linear behaviour before forwarding the information to the next layer.

During the model training, i.e., the learning process, a forward operation is carried out to compose a loss function comparing the NN output $\mathbf{z}^{(L)} = y_{NN}$ with the training examples y. A common loss function in regression problems is the mean absolute error (MAE or L1-loss):

$$\text{Loss}^{MAE} = \frac{1}{n}\sum_{m,n} |y^{NN}_{m,n} - y_{m,n}| \tag{2.26}$$

where m and n stand for the output vector dimension and number of data examples, respectively. The actual learning occurs when the $w_i^{(L)}$ and $b_i^{(L)}$ parameters start to be updated in order to reduce the loss function value. A usual pathway to update the weights is to follow a simple *gradient descent* algorithm:

$$\begin{aligned} \mathbf{W}^{(L)}_{t+1} &= \mathbf{W}^{(L)}_{t} - \lambda \nabla(\text{Loss})_{\mathbf{W}^{(L)}_t} \\ \mathbf{b}^{(L)}_{t+1} &= \mathbf{b}^{(L)}_{t} - \lambda \nabla(\text{Loss})_{\mathbf{b}^{(L)}_t}, \end{aligned} \tag{2.27}$$

being λ a learning rate parameter controlling the step-size. Due to the immense amount of data generally employed in this process, the usual approach involves a stochastic gradient descent technique, dividing the number of examples used for each update in small batches. Likewise, an epoch occurs when all batches are used. Nonetheless, the calculation of the gradients follows a backward chain rule procedure, layer-by-layer, often called *backpropagation*.

Nowadays, neural networks have different forms and operability. The model described so far is known as the Fully Connected (or Dense) Neural Network (FCNN os NN). However, when several neurons are connected through a sharing of w_i, often reducing the number of operations, the outcome is a Convolutional Neural Network (CNN).[65] The forward step now involves a series of convolutions, in which the dimensionality of the input data plays a major role. Alternatively, Recurrent Neural Networks (RNNs) introduces a more sophisticated way of understanding the sequential meaning in temporal-like data. The RNN keeps a hidden internal state vector as a memory of the previous elements on the sequence that, together with w_i, try to capture the

knowledge behind data ordering. For a deeper revision on RNNs the works of Medsker and Jain[67] and Lipton et al.[68] are suggested.

Another issue that needs to be addressed when dealing with data-driven methods is how the input data should be supplied to the model. Several approaches have appeared in the past years, especially around the fields of computational physics, cheminformatics and bioinformatics when regarding material sciences. In general, structures like molecules and solids have to be translated into something more meaningful for the machine, i.e., *representations* (or *fingerprints*) capturing relevant structure-property information. The development of fingerprints is an active research field to date, thereby just a short description of four alternatives are going to be presented here.

COULOMB MATRIX (CM):[69,70] This fingerprint uses a pair matrix to encode the Coulomb interaction between atoms in a molecule, thus, it is better suited for non-periodic systems. The idea is summarized by the following equation:

$$\mathrm{CM}_{ij} = \begin{cases} 0.5Z_i^{2.4} & \text{for} \quad i = j \\ \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}|_j} & \text{for} \quad i \neq j, \end{cases} \tag{2.28}$$

in which Z_i and R_i are the atomic number and position vector of atom i. The diagonal elements are represented by a fitting of atomic energies to their respective atomic number. The CM can be unfolded in three different sub-representation: a common matrix, a one-dimensional reshape of it or its eigenvalues.

MANY-BODY TENSOR REPRESENTATION (MBTR):[71] In this fingerprint, a set of many-body like aspects are encoded in a tensor format, which is suitable for both molecules and solids. The MBTR is defined by the following equation:

$$f_k(x,\mathbf{z}) = \sum_{i=1}^{N_a} w_k(\mathbf{i})\mathcal{D}(x, g_k(\mathbf{i})) \prod_{j=1}^{k} C_{z_j, Z_{i_j}} \tag{2.29}$$

where the index $\mathbf{i} \in \{1,\ldots,N_a\}^k$ runs over the k-group of N_a atoms with $\mathbf{z} \in N_a^k$ atomic numbers that are being represented by the k-body function g_k, w_k a weighting function and $\mathcal{D}$ a smoothing probability distribution. For $k = 1,2,3,4$-body relationships, g_k may be used to describe the respective number of atom species, inverse distances, angles and dihedral angles. The x acts as a real-space axis where the encoding happens.

SMILES: Simplified Molecular-Input Line-Entry System (SMILES) is one of the simplest ways to represent a molecule. Based on a text representation, the molecular characteristics are translated on a string-like sequence of characters encoding atom species, bonds, branches and rings. Atom species are represented by their respective chemical symbols; single bonds are omitted while double and triple bonds are described by "=" and "#"; branches are encapsulated by parentheses "()"; rings are enclosed by numbers, where the first/second number appearance on the sequence means opening/closing the ring. Additionally, special cases can be represented by brackets "[]", e.g., an explicit charged specie [N^+]. As a text-based fingerprint, SMILES often needs to be treated by using Natural Language Processing (NLP).[72]

MOLECULAR GRAPHS: Graph is a mathematical concept that represents pairwise connections between different elements. These elements and connections are the so-called graph *nodes* and *edges*, respectively. As molecular representations, graphs are useful as it offers a systematic way to encode information about atoms, as nodes, and chemical bonds, as edges. In general, a molecular graph is an identical representation of a given molecule, with nodes encoding information about atomic species and edges encoding bonding features.

Overall, graphs represent a special type of structured data, i.e., the way nodes and edges are organized also represents important information. Therefore, a special type of neural network is necessary when working with them: the Graph Neural Networks (GNNs).[73] These are designed to be compatible with graphs as inputs for the network and can also work with other architectures of neural networks. GNNs depend on a message passing (or neighbourhood aggregation) scheme that works by convoluting information from graph nodes and edges. In this way, the neural network's data processing for a given input node also depends on neighbour nodes, i.e., the network follows edge connections. The following Equation summarizes this message passing process:

$$\mathbf{x}_i^{(l)} = \boldsymbol{\phi}^{(l)}\left(\mathbf{x}_i^{(l-1)}, \nabla_{j\in N(i)}\boldsymbol{\psi}^{(l)}\left(\mathbf{x}_i^{(l-1)}, \mathbf{x}_j^{(l-1)}, \mathbf{e}_{j,i}\right)\right), \tag{2.30}$$

where $\mathbf{x}_i^{(l)}$ are features of the node i in the l-layer, $\mathbf{e}_{j,i}$ are features of the edge connecting the nodes j and i, ϕ and ψ are differentiable functions, often a subclass of neural networks like a FCNN, and ∇ represents an aggregation operation, e.g., a sum or a mean considering the $N(i)$ neighbours of the i-th node.

3. Thermodynamics assessment

3.1 The ion insertion process

During the battery cycling, two simultaneous redox reactions are happening at both electrodes while the process of lithium-ion de/insertion is occurring. While discharging, Li-ions leave the anode (lower potential vs. Li^+/Li) and are transferred to the cathode (higher potential vs. Li^+/Li). This reaction mechanism can be summarised, referenced to the lithium metal electrode, as:

$$\mathrm{Li}_{x_0}\mathcal{H} + (x_1 - x_0)\mathrm{Li} \rightleftharpoons \mathrm{Li}_{x_1}\mathcal{H},$$

in which the host material $\mathcal{H}$ is receiving $x_1 - x_0$ Li-ions. The battery voltage $V(x)$ at each equilibrium step of this process can be evaluated by a direct thermodynamic assessment of the reaction Gibbs free-energy ($\Delta_r G$) and the Nernst equation as:

$$V(x) = -\frac{\Delta_r G(x)}{nF} \tag{3.1}$$

with

$$\Delta_r G(x) = -\frac{1}{n_{\mathcal{H}}}\frac{\partial G(x)}{\partial x}, \tag{3.2}$$

where $n_{\mathcal{H}}$ is the number of moles of $\mathcal{H}$, F the Faraday constant, $G(x)$ the Gibbs free-energy of the system at the step x and n the number of electrons transferred in the process. The Gibbs free-energy in this equation is often approximated in solid-state by using the electronic total energy of the system at each step, which is expected to be the dominant term of the reaction free-energy at equilibrium, leading to:

$$V(x) = -\frac{\Delta_r G}{nF} = -\frac{E(\mathrm{Li}_{x_1}\mathcal{H}) - E(\mathrm{Li}_{x_0}\mathcal{H}) - (x_1 - x_0)E(\mathrm{Li})}{(x_1 - x_0)F} \tag{3.3}$$

Additionally, the redox potentials for a given molecular unit $\mathcal{M}$ – electroactive material of interest in this thesis – can be further investigated thermodynamically by breaking down the redox process into the oxidation and reduction reactions:[74]

$$\mathcal{M} \longrightarrow \mathcal{M}^+ + e^-$$
$$\mathcal{M} + e^- \longrightarrow \mathcal{M}^-$$

And following the Nernst equation:

$$\begin{aligned} E^0_{Oxi} &= \frac{-\Delta G_r}{nF} = \frac{-(G_X - G_{X^+})}{nF} \\ E^0_{Red} &= \frac{-\Delta G_r}{nF} = \frac{-(G_{X^-} - G_X)}{nF}. \end{aligned} \tag{3.4}$$

In the gas phase, the Gibbs free-energy is given by:

$$G = H - T(S_{vib} + S_{trans} + S_{rot}), \tag{3.5}$$

where

$$H = E_{elect} + E_{ZPE} + U_{vib} + U_{trans} + U_{rot} + PV, \tag{3.6}$$

and E_{elect} is the electronic energy and E_{ZPE} the zero-point energy. The subscripts *vib*, *trans* and *rot* stand for the vibrational, translational and rotational thermal contributions for the internal energy U and entropy S. The PV is the pressure-volume product. In solution, however, the Gibbs free-energy may be obtained by following the Born-Harber thermodynamic cycle, described in Figure 3.1. This procedure is shown in the following equation:

$$\begin{aligned} \Delta G_{(\text{solvent})}(Red) =& \Delta G_{(\text{gas})}(Red) + \Delta G_{\text{solvation}}(Red) \\ & - \Delta G_{\text{solvation}}(Ox), \end{aligned} \tag{3.7}$$

where the subscripts gas and solvent represent the molecule in gas and solvated phases while $\Delta G_{solvation}$ is the solvation energy of the respective state.

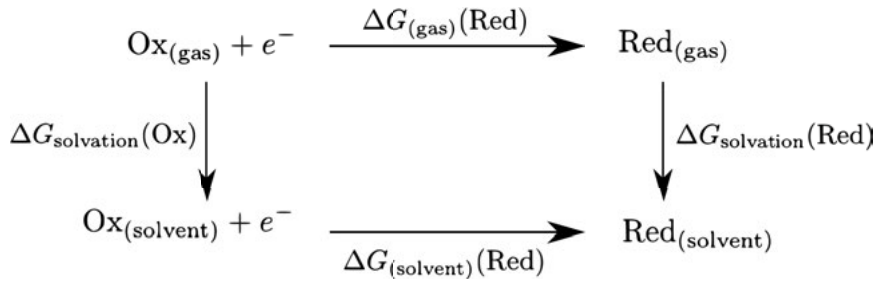


Figure 3.1. The Born-Harber thermodynamic cycle for the redox reaction described as $\mathcal{M} \longrightarrow \mathcal{M}^+ + e^-$. The subscripts gas and solvent describe, respectively, the moiety $\mathcal{M}$ in gas phase and in solution.

Furthermore, the compound's formation energy for a given concentration (x) of the inserted ion can be expressed following the Equation 3.8, below, with respect to the initial (x_i) and final (x_i) electrode compositions. This formation energy is often used to evaluate the stability of alloys,[75] as its dependency on the alloy composition draws a convex hull that could suggest stable phases.

$$E_F = E(\mathrm{Li}_x\mathcal{H}|\mathcal{M}) - \frac{(x-x_i)E(\mathrm{Li}_{x_f}\mathcal{H}|\mathcal{M}) + (x_f-x)E(\mathrm{Li}_{x_i}\mathcal{H}|\mathcal{M})}{x_f - x_i} \tag{3.8}$$

Finally, the Li-ion in the presented methodology can be easily exchanged by Na^+, K^+ or other desired insertion ion.

3.1.1 Case 01: "Superlithiation" – Understanding the energy storage limits of OEMs

In this Case, based on Paper I, a novel analysis is proposed to aid the investigation of the ion insertion reaction thermodynamics. This new tool, driven by the charge derivative of the organic moiety as a function of the inserted ion concentration, provides a new understanding of the charge storage mechanism in OEMs. Furthermore, it offers a simple reasoning for the "superlithiation" phenomenon. In this, some molecules are capable of storing more charge than what is initially anticipated from their possible redox-active centers, which results in a higher charge storage capacity for the OEM. Paper I demonstrates that to assess this phenomenon a voltage analysis alone may not suffice. However, when coupled with the charge derivative analysis, experimental results from the literature are accurately reproduced. This is shown for a set of carbonyl-based molecules, namely, dilithium acetylene-dicarboxylate (Li_2ADC), dilithium tolane-dicarboxylate (Li_2TODC), dilithium thiophene-dicarboxylate (Li_2TDC), dilithium terephthalate (Li_2TP), dilithium benzene-diacrylate (Li_2BDA) and dilithium benzene-dipropiolate (Li_2BDP).

To back this analysis, a "melting-quenching" approach was employed to generate the molecular structures for all the investigated compounds at different lithiation phases. This method allows, during the melting stage, the examination of the molecule's configurational space as it is carried out through a high-temperature Born-Oppenheimer molecular dynamics simulation. Several snapshots are taken from this ab initio molecular dynamics (AIMD), followed by a structural relaxation (within the DFT framework) in order to find the minimum energy geometry. This last process representing the quenching stage. This way, the reaction thermodynamics can be evaluated by employing the minimum energy structure of different lithiation phases. Furthermore, the electronic charge distribution in each phase can also be assessed.

The charge derivative follows the idea that the organic moiety functions as a charge reservoir, accommodating the inserted electron at each reduction step during the ion insertion reaction. Thus, evaluating changes in the moiety's charge state can reveal its electronic storage limits. For example, the charge variation upon lithiation is going to be negative when the moiety has accommodated an inserted electron, and positive otherwise. This can be summarised in the following equation:

$$\frac{\partial Q(n)}{\partial n} < 0, \tag{3.9}$$

where Q is the total molecular charge in units of e, excluding inserted ions, and n the reaction step. Hence, this derivative will be ≥ 0 at the limit where the molecule fails in accommodating additional electrons. Equivalently, the moiety reduction starts to become less favourable than the inserted ion's reduction, which would lead to the electron being transferred to the Li^+ and, thus, the formation of metallic Li.

Figure 3.2 (a) presents the average charge per atom (excluding Li) as a function of the number of Li-ions. The zero-charge level refers to the pristine compound. Equation 3.9 directly correlates with the slope at different lithiation reaction steps in these plots. That is, a downward slope means a negative derivative while an upward slope represents a positive one, i.e., the molecule failing to accommodate extra electrons. The Li_2BDP compound, for instance, shows a steady decrease in charge up to the Li_{14}BDP phase, which agrees with the experimental finding that this moiety can be successfully lithiated up till $Li_{13.5}$BDP.[22] Conversely, the Li_2TP displays an ascending slope after the Li_6TP phase indicating that this would be its limit, which also agrees with reported experimental results.[76] Figure 3.2 (b-g) shows the charge derivative following two possible reaction pathways for all the compounds. These reaction pathways were defined by considering the allowed reaction thermo-

dynamics, i.e.,

$$\frac{\partial V(n)}{\partial n} \leq 0, \tag{3.10}$$

and the formation energies according to Equation 3.8.

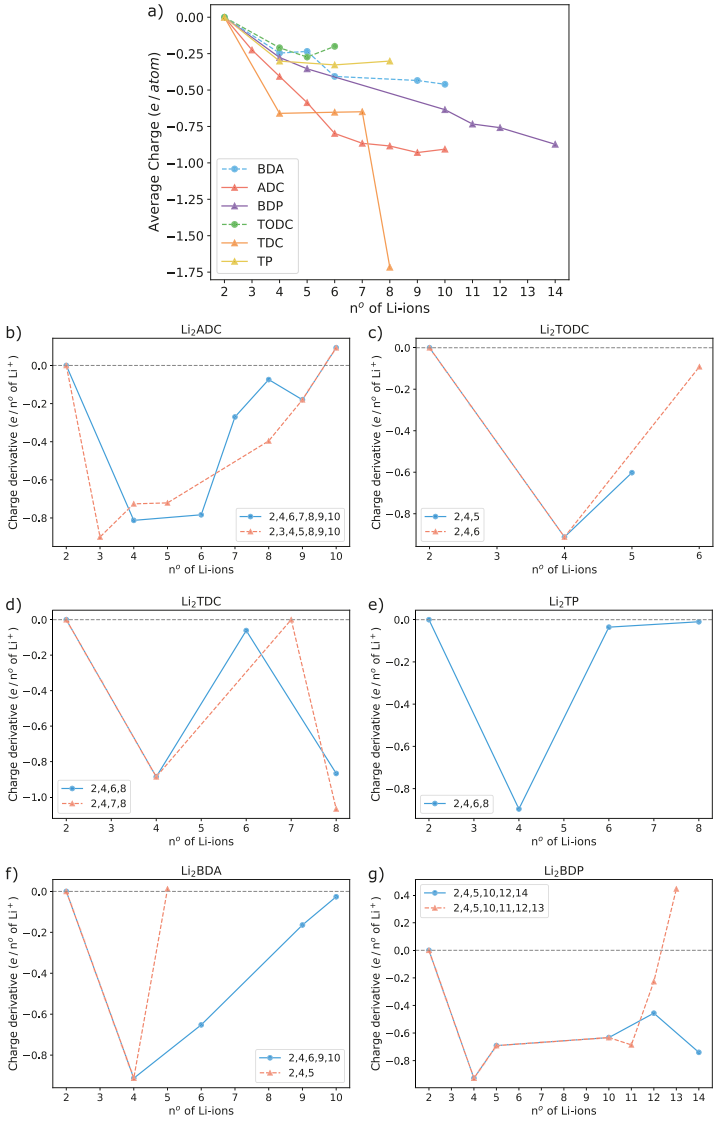


Figure 3.2. a) Average charge per atom (excluding Li) as a function of the number of Li-ions and referred to the pristine phase. The charge derivatives for the b) Li_2ADC, c) Li_2TODC, d) Li_2TDC, e) Li_2TP, f) Li_2BDA and g) Li_2BDP compounds. Reprinted from Paper I.

3.2 Evolutionary algorithm, Potential Mapping and crystal structure predictions

In order to evaluate the thermodynamics in the solid-state level, the crystal structure of the organic molecular crystal and its respective lithiated phases are required. These structures may be obtained experimentally by, for example, X-Ray Diffraction (XRD) techniques. This is, in general, very difficult for the lithiated phases as operando spectroscopy techniques might be necessary. Therefore, the first-principles structure predictions are very important contribution to unveil the ion insertion reaction mechanism. In the past years, several different methodologies have been proposed as a way to predict the crystal structure of a solid, e.g., metadynamics,[77] simulated annealing[78] and basin hopping.[79] In this thesis, a particularly successful approach based on the interplay between DFT and an evolutionary algorithm (EA) has been used. This methodology is implemented on the software USPEX[80–82] and work in conjunction with VASP. The evolutionary algorithm works by evolving a set of possible solutions, i.e., different crystals of a given composition, by applying a series of genetic operations. First, an initial population of structure-candidates for the selected compound is randomly created from a list of possible group symmetries. In the case of molecular crystals, this process can be immensely optimised by supplying the molecule as a whole unit. For each candidate, a full geometry optimisation is carried out within DFT/VASP and the final energies collected. Afterwards, the population goes under a set of evolutionary operations, such as heredity, mutation and permutation, in order to compose a new generation. Using energies as the "survival of the fittest" criteria, the *heredity*, as in nature, creates new structures by mixing two parenting candidates. This ensures that structures with the smallest energy minima are going to pass their characteristics for the next generation. *Mutation* creates a new child by mutating some aspect of a single parenting structure, e.g., by applying a strain on the lattice vectors or a random perturbation on atom positions. *Permutation* swaps two atoms in a selected structure to create a new child.

Following the evolution operations, the new generation is created and the optimisation process is then repeated, composing others generations in sequence. This evolution process stops when a certain convergence criterion is achieved with the, hopefully, global minimum energy structure for the given compound. Thereafter, the reduced phases of the compound after the added insertion ion can also be predicted through this method. Assuming each of the predicted structure phases as the respective equilibrium step, the reaction voltage can be calculated by using the Equation 3.3.

In general, the EA approach is more likely to find a global energy minimum structure in the configurational space as it is essentially an algorithm designed for this purpose. However, metastable phases may happen in different stages of the ion insertion process, affecting the overall battery operation.

Therefore, a different strategy is necessary to evaluate metastability in these organic compounds. A novel computational methodology has been developed in this thesis to address this issue. This new approach, hereafter named Potential Mapping (or MAP), relies on a successive mapping of the potential energy surface (PES) within the compound's crystal structure to introduce "educated guesses" on the inserted ion's position. As a drawback, MAP requires an initial crystal structure (commonly the pristine phase) to start the process. This initial phase can be obtained either through the EA or different experimental methods. Overall, the PES can be assessed by employing DFT, for example, and the ion position can be guessed after considering the ion's charge state. For cations, like Li^+ or Na^+, the logical guess would be the PES maximum. Nonetheless, this initial educated guess must be further improved by considering the interactions between the inserted ion and the host system, e.g., other atoms. This can be amended by carrying out a DFT geometry optimization after the insertion, relaxing the structure to a local energy minimum. Optionally, the PES can be re-evaluated for this new relaxed structure containing the inserted ion and used in sequence to guess the position of a new ion being inserted. The process can be repeated as many times as necessary to represent the desired insertion reaction steps. Figure 3.3 shows a workflow summarising the entire mechanism. This novel methodology was also implemented in a new software package coded in the python language and named Mapion; it is freely distributed on an open-source basis at GitLab and further discussed in Paper IV.

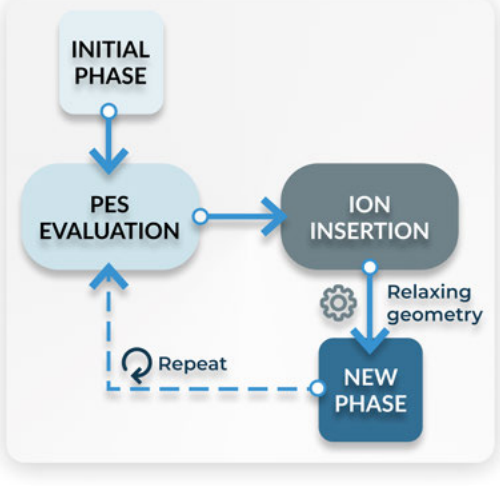


Figure 3.3. Workflow illustrating the potential mapping mechanism. PES = potential energy surface. Reprinted from Paper IV. Copyright 2022, John Wiley & Sons, Inc.

3.2.1 Case 02: Predicting the crystal structure of organic electrodes

This section is based on Paper II, in which a successful use of the evolutionary approach aforementioned rendered the prediction of the crystal structure for the Na_2BDA compound with an excellent match with the experimentally resolved structure. Figure 3.4 (a) shows two different views of the experimental and theoretically predicted structures. For the former, the structure was obtained by using the 3D Electron Diffraction (3DED) technique combined with a powder X-ray diffraction (PXRD). As for the latter, the structure was predicted by following the evolutionary process of two benzene diacrylate units and four Li atoms. Both experimental and predicted structures possess the same $P2_1/c$ space group, with the former exhibiting the lattice parameters $a = 14.49$ Å, $b = 5.50$ Å, $c = 7.21$ Å, 90.7^o, 101.1^o and 89.6^o while the latter $a = 13.77$ Å, $b = 5.34$ Å, $c = 6.89$ Å, 90.01^o, 97.31^o and 89.95^o.

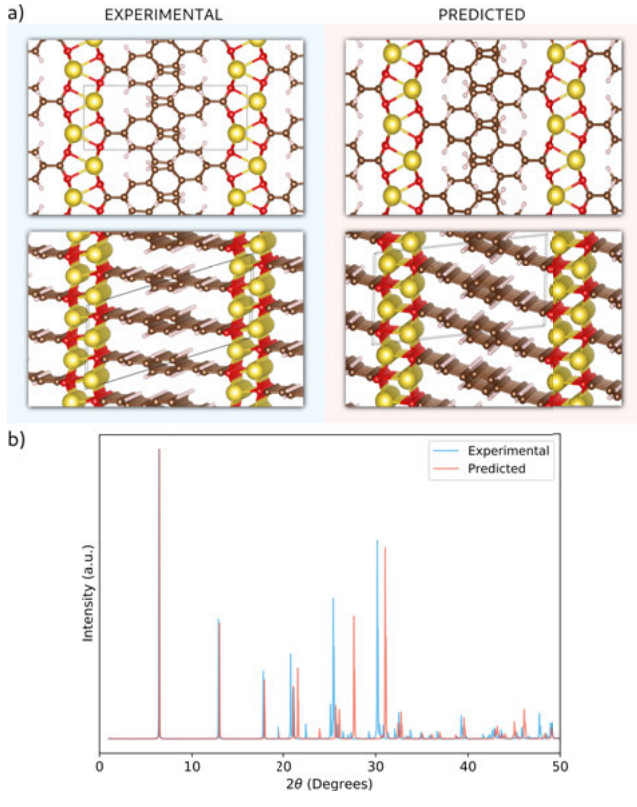


Figure 3.4. a) Comparative split of the experimental (blue shade) and theoretically predicted crystal structures of the Na_2BDA. The colour code for the atoms is: red for oxygen, brown for carbon, white for hydrogen and yellow for sodium. b) Calculated powder-XRD diffractogram for both experimental and predicted structures. Reprinted from Paper II. Copyright 2021, Royal Society of Chemistry.

In Figure 3.4 (b), the calculated XRD diffractogram further illustrate the similarities between the crystals. The evolutionary step was carried out as explained in the Paper II, having VASP as the software handling the underlying DFT calculations by the means of GGA-PBE functional, including the D2 correction scheme as presented in the Theoretical Framework section, and the PAW method. Furthermore, the outstanding agreement between theory and experiment demonstrates how powerful this theoretical methodology is when predicting molecular crystals and corroborates with its use to access the sodiated phases of the Na_2BDA compound. Experimentally resolving these reduced phases is a challenging task to achieve, requiring the implementation of sophisticated *in-operando* techniques. Therefore, the evolutionary approach emerges as a promising alternative. Figure 3.5 shows the crystal structure for the sodiated Na_3BDA and Na_4BDA phases. By accessing such structures, distortions in the lattice may be evaluated in more details. The first sodiation leads to a loss of symmetry in the crystal structure, with the new crystal space group being P1. This is reverted when the second sodium is inserted, turning the crystal back to the $P2_1/c$ group. Moreover, while the first added Na^+ appears to interact more with the carboxylic units the second Na^+ insertion leads to a stronger interaction with the acrylate arms.

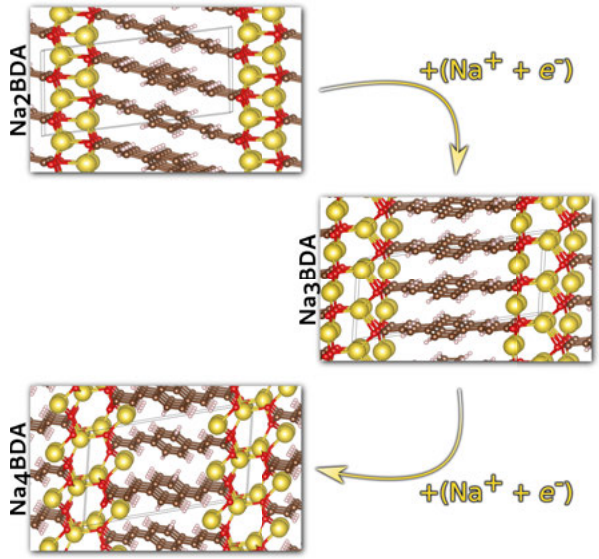


Figure 3.5. Predicted crystal structures for the de/sodiated phases of the Na_2BDA. Atom colours as Figure 3.4. Reprinted from Paper II. Copyright 2021, Royal Society of Chemistry.

The thermodynamics of the sodiation reaction can, now, be evaluated by following the Equation 3.3. Prior to it, an additional single-point calculation has been carried out for each system with the HSE06 functional in order

to improve the electronic structure description. Figure 3.6 present the sodiation voltage for each reaction step by using PBE (a) and HSE06 (b) functionals. Although the average voltage for the two-step reaction (dashed line) of both methods are relatively close, 0.69 V vs. Na/Na^+ (GGA) and 0.63 V vs. Na/Na^+ (HSE06), the latter shows a better agreement with the experimentally reported value of 0.6 V vs. Na/Na^+.[83] Figure 3.6 (c) present the formation energy of the first sodiated phase with respect to Na_2BDA and Na_4BDA as calculated by following the Equation 3.8. The HSE06 functional indicates that the Na_3BDA phase is energetically unfavourable, which would reveal a disproportionation during the sodiation that could directly stabilize Na_4BDA phase in a two-step process. As a matter of fact, the experimental result reported in the reference further support this conclusion when showing a single plateau in the battery charge/discharge curve.[83]

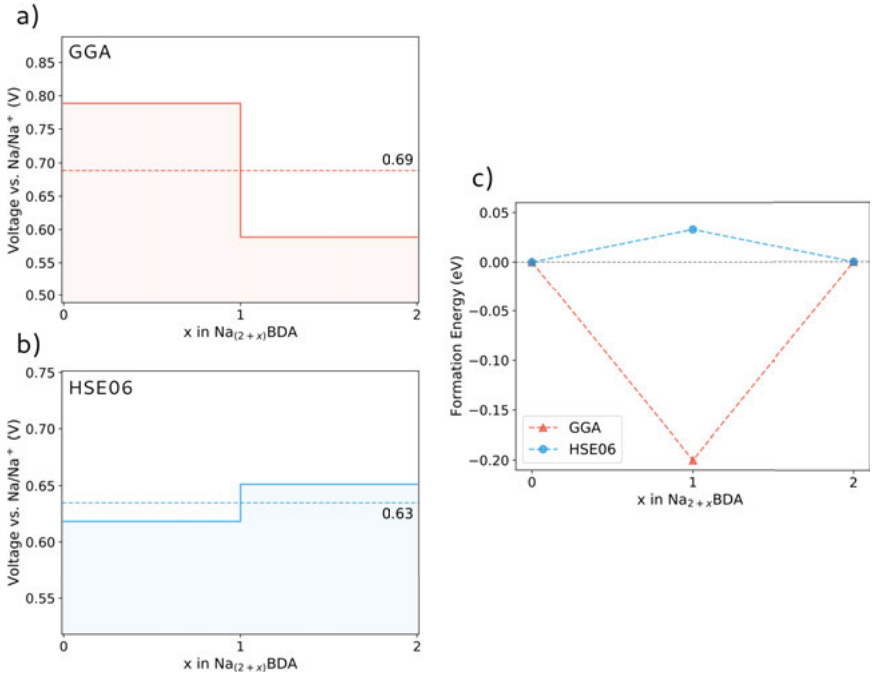


Figure 3.6. Potential profile of the Na-ion insertion process for the Na_2BDA as obtained with the GGA-PBE a) and HSE06 b) functionals. c) Formation energy of the Na_3BDA as referred to the Na_2BDA and Na_4BDA. Reprinted from Paper II. Copyright 2021, Royal Society of Chemistry.

The crystal structure for the Li_2BDA and its first two lithiated phases have also been predicted by following the same methodology. Figure 3.7 briefly summarises the collected data. In Figure 3.7 (a), the total and the segment-projected density of states (DOS) unveil the electronic structure of Li_2BDA. A particularly interesting feature to analyse is the first unoccupied

band. Among other characteristics, the band composition could reveal the redox-active center as this first unoccupied band will be populated by the additional electrons during the electrochemical reaction. In the presented case, the band shows a reasonably fair composition by the molecular fragments, which suggests an electronic delocalisation upon reduction. This fact is further illustrated by Figure 3.7 (b), where the charge distribution for the two extra electrons in Li_4BDA is presented. The formation energy for the Li_3BDA with respect to the Li_2BDA and Li_4BDA is displayed in Figure 3.7 (c). Once more, the results obtained with the HSE06 functional suggests the two-electron process with the Li_3BDA being energetically unfavourable, in stark contrast with the PBE functional. The lithiation voltage has been evaluated by using the Equation 3.3, presenting a two-electron reaction of 1.12 V vs. Li^+/Li in good agreement with the 1.2 V vs. Li^+/Li reported experimentally.[84]

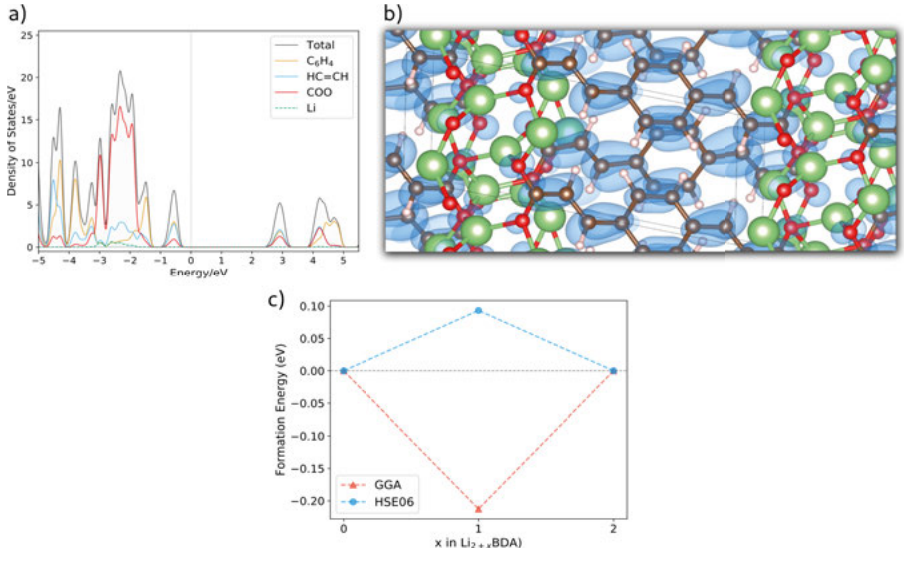
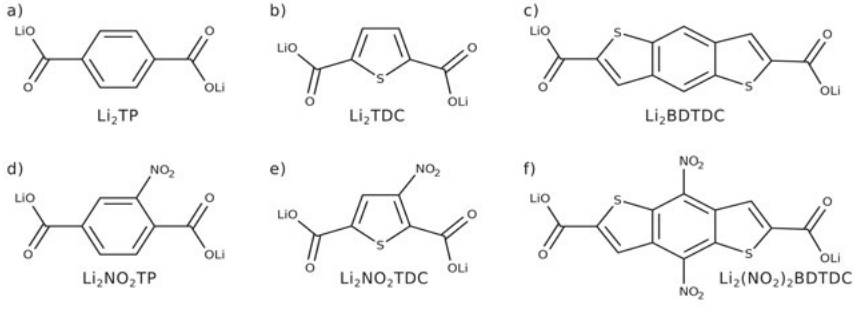


Figure 3.7. a) The fragment-projected density of states (DOS) for Li_2BDA and b) the charge density for the extra two electrons received upon two consecutive lithiation. c) Formation energy of the Li_3BDA referred to Li_2BDA and Li_4BDA. Colour code for the atoms: red for oxygen, brown for carbon, white for hydrogen and green for lithium. Reprinted from Paper II. Copyright 2021, Royal Society of Chemistry.

3.2.2 Case 03: Tailoring organic cathodes

This Case is based on the Paper III. With the evolutionary algorithm approach accomplishing an excellent agreement with the experimental findings, as described in the previous Case, here a step further has been taken. The EA framework has been employed to tailor three basic molecular compounds with anodic characteristics: the dilithium terephthalate (Li_2TP), the dilithium thiophene dicarboxylate (Li_2TDC) and the dilithium benzodithiophene dicarboxy-

late (Li_2BDTDC). By being functionalised with the nitro (NO_2) group, these molecules displayed a significant change in their lithiation potentials, thus shifting to a cathodic character. The Scheme 3.1 present the Lewis structures of these molecules.



Scheme 3.1. The Lewis structures of the investigated compounds a) dilithium terephthalate (Li_2TP), b) dilithium thiophene-dicarboxylate (Li_2TDC), c) dilithium benzodithiophene-dicarboxylate (Li_2BDTDC), d) dilithium nitro-terephethalate (Li_2NO_2TP), e) dilithium nitro-thiophene-dicarboxylate (Li_2NO_2TDC) and f) dilithium dinitro-benzodithione-dicarboxylate ($Li_2(NO_2)_2$BDTDC). Reprinted from Paper III. Copyright 2020, John Wiley & Sons, Inc.

Figure 3.8 (a) exhibit the average bond length deviation (AD) for atoms composing the organic ring in the respective molecular unit. The AD has been calculated by averaging the absolute differences between bond lengths d_i and the mean bond length $\bar{d}$ of the considered ring. An AD close to zero means very similar bonds around the ring, which is a signature of the benzene. Therefore, the variation of AD upon lithiation quantifies geometric changes in the molecular ring. In this sense, it can be seen from Figure 3.8 (a) that the NO_2 presence improves the structural stability of the rings upon lithiation, except for the TP. This effect is connected with the charge localisation and redox centers discussed ahead.

$$\overline{\Delta} = \frac{\sum_i^n |d_i - \bar{d}|}{n} \tag{3.11}$$

The average voltage for the two-step lithiation reaction is shown in Figure 3.8 (b). The two-step $Li_x\mathcal{H} + 2Li \longrightarrow Li_{x+2}\mathcal{H}$ reaction is suitable to compare the overall performance of each compound and the electrochemical effects induced by the NO_2 group. In general, the functional group has shifted the lithiation potential of all compounds to a higher value, with the most expressive change occurring for the $Li_2(NO_2)_2$BDTDC with an increase from 0.92 V vs. Li^+/Li to 2.66 V vs. Li^+/Li. This difference represents a shift of

1.8 V, enough to change the electrode character from the one of an anode to a cathode.

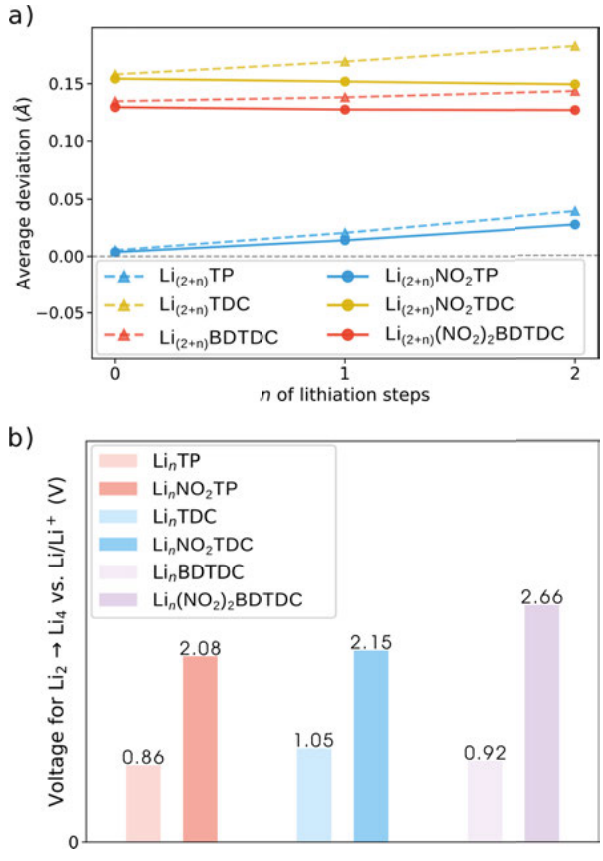


Figure 3.8. a) The average bond length deviation for the molecular ring of the compounds as a function of the lithiation step. b) Calculated average voltage for the two-electron process for each investigated compound. Reprinted from Paper III. Copyright 2020, John Wiley & Sons, Inc.

The changes on the voltage could be better understood on the light of the electronic structure. In this regard, Figure 3.9 presents the fragment-projected density of states (DOS) for the Li_2NO_2TP (a), Li_2NO_2TDC (c) and $Li_2(NO_2)_2BDTDC$ (e). One more time, the first unoccupied band indicates the species that would be reduced upon lithiation, and, therefore, occupying the referred band. For all the compounds, this band is dominated by the NO_2, which seems to be aligned with the fact that this group is known to be a strong electron-withdrawing one. To further illustrate this, the Figure 3.9 also displays the charge density of the two extra electrons receive upon lithiation for the Li_4NO_2TP (a), Li_4NO_2TDC (c) and $Li_4(NO_2)_2BDTDC$ (e). It is possible to note that the charge is mainly localised around the NO_2, especially for

the benzodithiophene. This charge localisation is a common feature of cathodic materials, in which the redox-active center may be better defined.

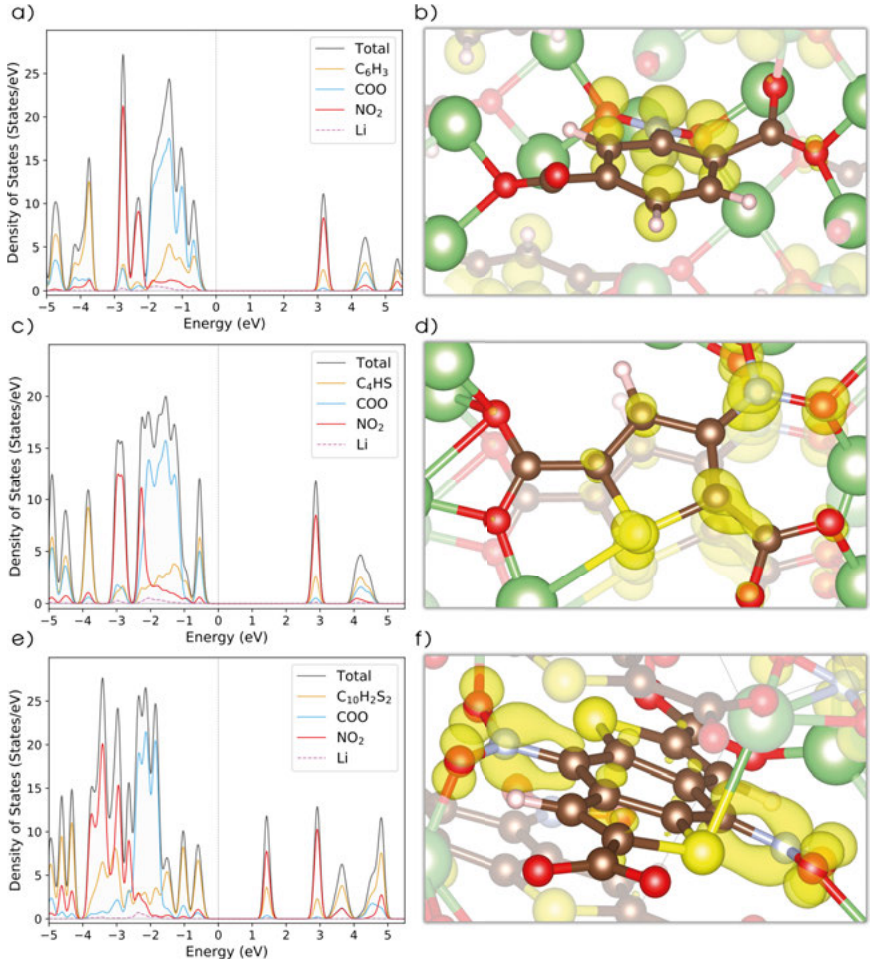


Figure 3.9. HSE06 Density of states (DOS) and charge densities of the extra electron received upon Li^+ uptake for: a) Li_2NO_2TP and b) Li_4NO_2TP, c) Li_2NO_2TDC and d) Li_4NO_2TDC, e) $Li_2(NO_2)_2BDTDC$ and f) $Li_4(NO_2)_2BDTDC$. The colour code for the atoms: red for oxygen, brown for carbon, yellow for sulfur, white for hydrogen, blue nitrogen and green for lithium. Reprinted from Paper III. Copyright 2020, John Wiley & Sons, Inc.

Finally, the Bader charge analysis can be a useful tool in unveiling the charge transfer upon lithiation, pinpointing the charge localisation in specific elements of the compound. The total electronic charge for every fragment per lithiation phase can be compared with their respective delithiated charge. This indicates the localisation of the extra electrons received in conjunction

with the Li^+ uptake. Figure 3.10 present this charge analysis for the pristine (dashed-lines) and the functionalised (solid-lines) compounds.

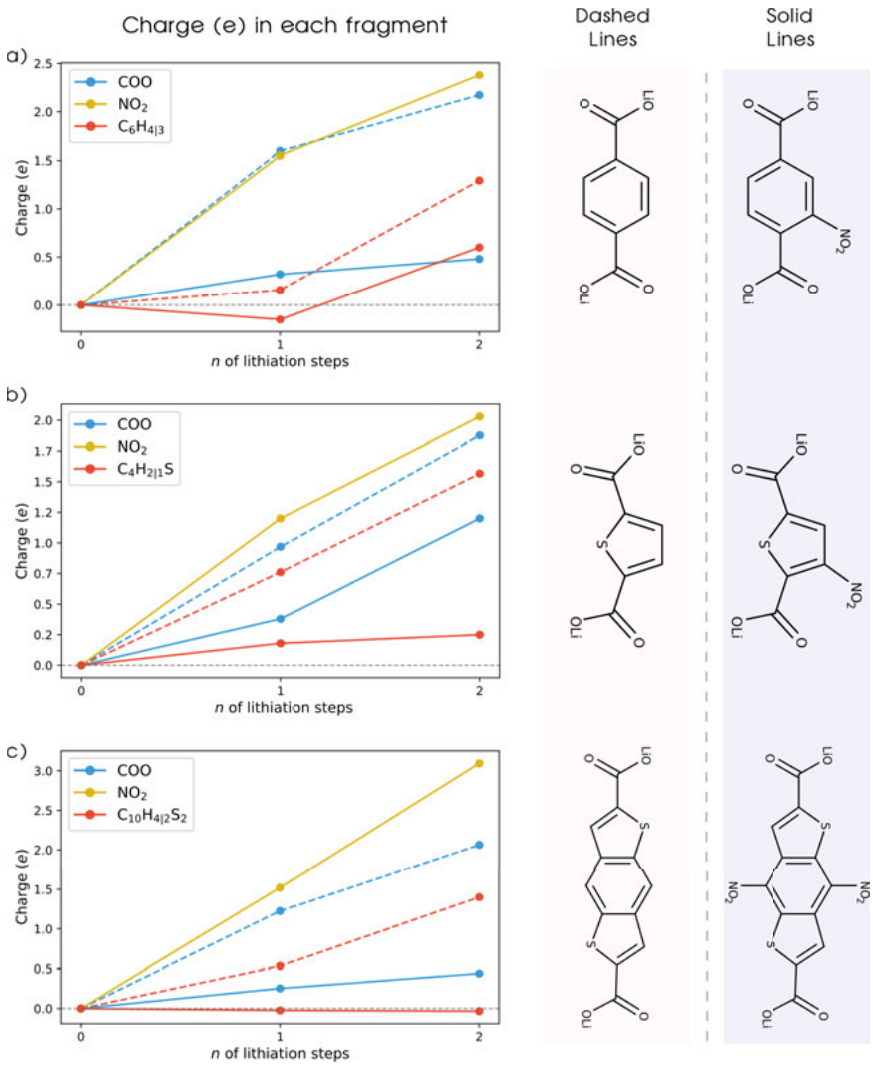


Figure 3.10. Charge variation in each molecular fragment as a function of the lithiation step provided by the Bader analysis. The solid (dashed) lines refer to substituted (non-substituted) compounds. Reprinted from Paper III. Copyright 2020, John Wiley & Sons, Inc.

For the non-substituted compounds, the charge is distributed over the carboxylic units and the molecular ring. Conversely, the charge is sensibly localised in the NO_2 when it is present, thus better defining the redox-active center. Thereby, these results reinforce that the NO_2 group modifies the low-potential anodic character of these compounds to the one of a high-potential

cathode. Furthermore, the ring on the NO_2TP seems to have a share of charge, in contrast with NO_2TDC and $(NO_2)_2$BDTDC where the charge is negligible. This fact confirms the structural stabilisation of the ring proposed in Figure 3.10 (a), in which the TP still suffers from geometrical changes regardless of the functional group presence.

3.2.3 Case 04: Exploring metastable phases during Li-ion insertion processes

In the Paper IV, the lithiation mechanism in OEMs was further investigated by considering not only the evolutionary algorithm scheme but also the novel potential mapping (MAP) algorithm discussed in Section 3.2. These methods were employed in obtaining the lithiated phases for the Li_2DHT compound.[17,85] EA and MAP proposed structurally different phases due to how these prediction schemes are designed. The former is more suited to find global minimum structures while the former is more prone to identify metastable phases. Figure 3.11 (a) shows the Lewis structure of the Li_2DHT compound. Figure 3.11 (b) presents the average insertion voltage for the two-step Li-insertion reaction from Li_2DHT to Li_4DHT considering three distinct theory levels: PBE, HSE06-NO and HSE06-O. In PBE, GGA-PBE functional is employed to relax the predicted structures and for final electronic calculations. In HSE06-NO, the structure is obtained as in PBE, but a final electronic calculation is carried with the HSE06 hybrid functional. Finally, in the HSE06-O both structure relaxation and electronic calculation steps are performed with the HSE06 functional. In this way, Figure 3.11 (b) compares not only the EA and MAP methods but also the fundamental differences when obtaining the final relaxed structure by employing two different functionals. In fact, when comparing these voltages with the experimentally reported value of approximately 2.6 V vs. Li^+/Li (dashed line), we see that MAP offers a better agreement with experimental findings. Moreover, the hybrid functional further improves the result. Overall, the exchange-correlation interactions play an important role in molecular bonding and the improved description of this interaction offered by the hybrid function may justify this outcome. Figure 3.11 (c) shows the electronic energy as a function of the lithiation step and referred to the respective delithiated structure energy from different methods. As expected, the EA presents the final phase (Li_4DHT) with a lower energy than MAP. However, as the latter offers a better agreement with experimental results, it is feasible to conclude that the lithiation is happening through metastable phases for this compound.

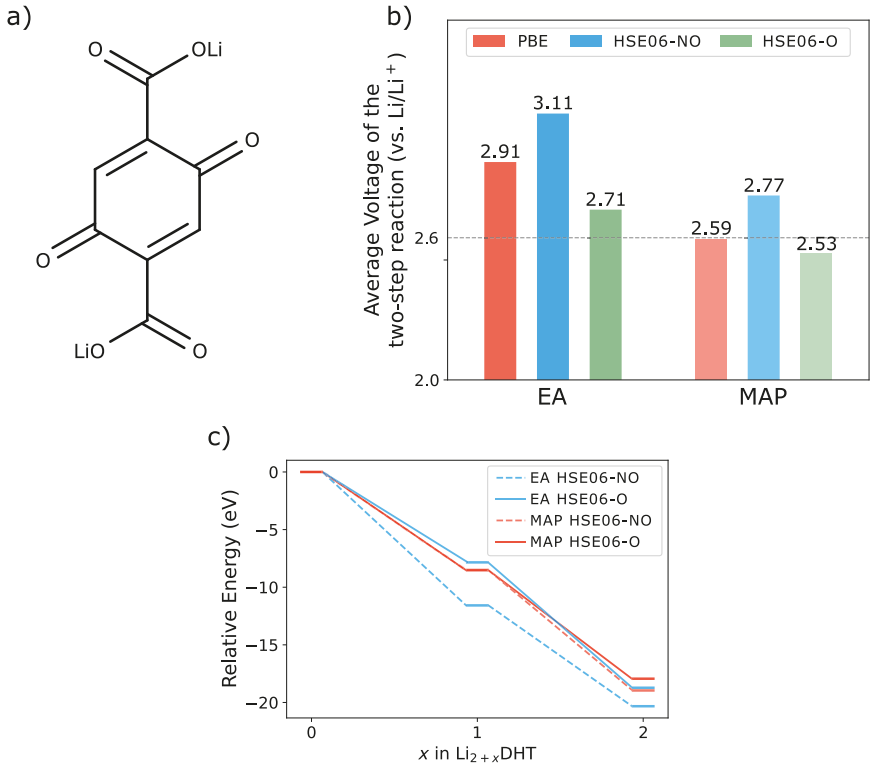


Figure 3.11. a) Lewis representation of the Li_2DHT molecule. b) Average voltage of the two-step reaction (vs. Li^+/Li) for all the investigated prediction schemes and theory levels. The dashed line represents the experimentally reported value.[17,85] c) The electronic energy of each lithiated phase relative to the pristine Li_2DHT. Reprinted from Paper IV. Copyright 2022, John Wiley & Sons, Inc.

In fact, the cyclic voltammogram (CV) reported by Chem and colleagues[85] shows a single oxidation peak at 2.7 V for the first cycle, which disappears in the next cycle giving place to two peaks near 2.5 V and 2.6 V. In their study, the Li_4DHT phase was chemically synthesised to assemble the battery cathode, which means that it is more likely to be close to the global minimum structure and, therefore, the crystal structure obtained from the EA. This is further supported by the fact that the EA insertion voltage is 2.71 V (HSE06-O), in agreement with the CV first cycle 2.7 V peak. However, in the next cycle the CV peaks of 2.5 V / 2.6 V agree with the 2.53 V from MAP. Thereafter, it seems that the Li_4DHT "as-synthesised" is not fully recovered after the first cycle and the lithiation mechanism happens metastable phases.

4. The data-driven approach

4.1 Case 05: AI-*kernel* – An AI-driven in-silico platform to discover novel lithium-ion battery cathodes

Up to this point, the evolutionary algorithm (EA) has proven to be an accurate and useful tool in investigating and engineering organic electrodes. However, this approach has a fateful drawback: the significant computational effort involved in the prediction scheme. The crystal structure evolution process commonly requires hundreds, if not thousands, DFT calculations to correctly predict the global minimum structure of the system. Therefore, applying this methodology in the high-throughput screening of novel organic materials is a challenging task. Moreover, the practical synthesis of billions of organic compounds is an unreasonable deed to be accomplished, both humanly and economically.

Nonetheless, the demand for novel energy storage technologies have been rapidly increasing due to the energy consumption profile of our society. To solve this dilemma, an alternative route to accelerate materials discovery has been arising in the means of machine learning (ML) techniques. This approach has already shown success in several tasks, e.g., accelerating ab-initio molecular dynamics,[39] proposing new perovskites,[86] predicting several properties of materials[40,41,43,87] and even helping to solve quantum mechanics equations.[38] This section, based on the Paper V, is going to present an innovative framework based on Neural Networks (NNs) and backed by DFT calculations in designing novel organic battery materials.

The idea behind this framework, summarised by Figure 4.1, is divided into three development steps: (A) creating a small dataset of organic crystals for Li-ion batteries by using the evolutionary algorithm; (B) generating a database of organic molecules using DFT calculations; (C) the artificial intelligence machinery development combining data from A and B. This AI-*kernel* relies only on the molecular structure as input to predict the battery open-circuit voltages, completely by-passing the time-demanding DFT calculations. Therefore, it enables a fast assessment of novel battery materials. Additionally, Figure 4.1 also shows a possible feedback loop, responsible to further expand the database.

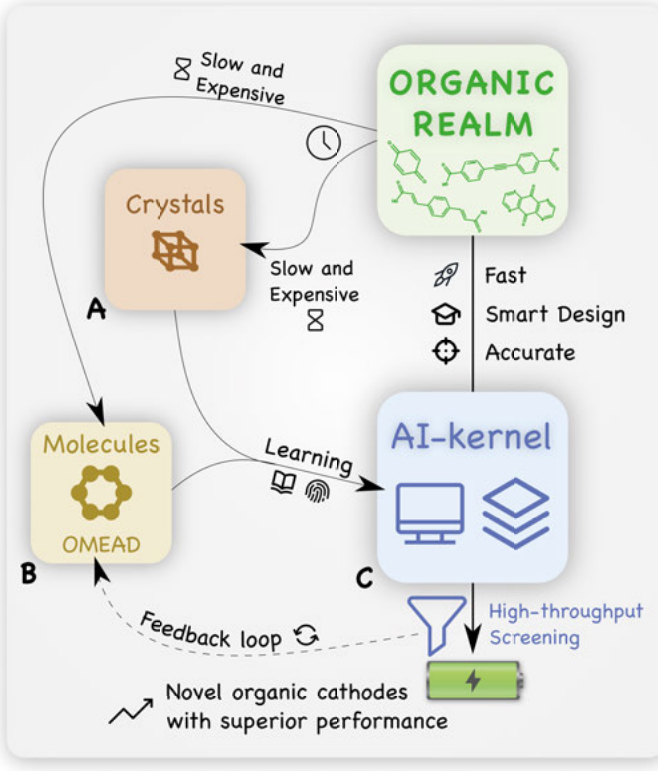
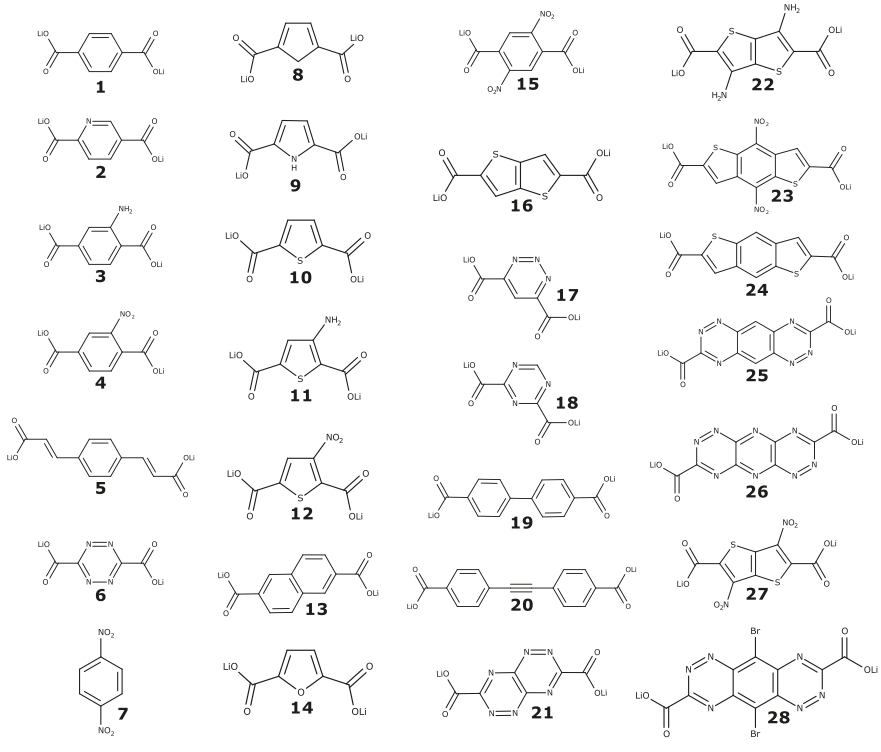


Figure 4.1. Flowchart illustrating the workflow of the proposed framework and how the AI-*kernel* operates, enabling a fast access to the world of organics. OMEAD stands for "Organic Materials for Energy Applications Database". Reprinted from Paper V. Copyright 2022, Elsevier.

Initially, the crystal structure for a set of carboxylate-based electrode candidates and their respective lithiated phases was predicted by using the EA. The lithiation voltage of these systems was calculated by following the Equation 3.3. Carboxylated structures tend to form stable crystals, usually presenting the organic part intercalated by a salt layer where Li-ions are surrounded by the carboxylate oxygens. This characteristic contributes to the overall stability of the crystal structure in this type of compounds. Scheme 4.1 exhibit the Lewis structures of the molecular units of this dataset (step A). Following the step B development, more than 26000 unique small molecules with up to 17 heavy atoms like C, N, O, S and halogens have been extracted from the organic universe and systematically being calculated under DFT. These moieties have been fully optimised without any structural constraints in their neutral, oxidized and reduced phases. The database, here named Organic Molecules for Energy Applications Database (OMEAD), is comprised of the respective molecular entry and a list of properties extracted from the quantum mechanics

calculations, including dielectric constants, HOMO-LUMO energies, redox potentials, electronic affinities and ionization potentials. In order to complete this step, more than 310000 DFT calculations had to be performed. Although a number of organic molecules database exist in the literature, none of them present a systematic calculation of the properties aimed here and with focus on energy applications.



Scheme 4.1. The Lewis representation of the moieties composing the small dataset of predicted crystal structures. Reprinted from Paper V. Copyright 2022, Elsevier.

As for step C, two distinct statistical learning models was developed: a linear and a neural network models. The Linear Model (LM) follows a simple ordinary least squares (OLS) regression to estimate the α and β parameters of the linear function $y = \alpha x + \beta$. In this, y is the open-circuit voltage (V_{OC}) and x the molecular reduction potential (P_{Red}), here chosen as the molecular descriptor after a careful statistical investigation of several molecular properties, e.g., oxidation potential, HOMO and LUMO energies, etc. Figure 4.2 shows the trained model's equation and performance. In addition, new molecular examples extracted from the literature, represented by the green triangles, were introduced to expand the Linear Model diversity and to enhance its prediction effectiveness as these new entries are not based in carboxylates.

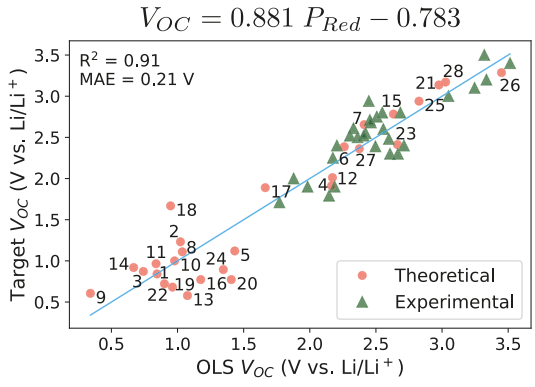


Figure 4.2. The Linear Model connecting reduction potential and open-circuit voltages. Above the plot, the equation obtained through the ordinary least squares (OLS) regression. The numerical labels follow the ones in Scheme 4.1. Reprinted from Paper V. Copyright 2022, Elsevier.

Now the task was to design the Neural Model to predict the molecular descriptors, thus suppressing any need for further DFT calculations. The molecular descriptors predicted by the Neural Model, i.e., the reduction potential, serve as input for the Linear Model in a joint artificial intelligence machinery, the AI-*kernel*. In the development process, several combinations of molecular fingerprints and NN architectures have been benchmarked to render the most efficient model. Figure 4.3 (a) present a simple schematic of the final Neural Model. A molecule, represented by a SMILES string, is initially converted into a sequence of embedding vectors where each string element has a unique vector representation. Embedding is a powerful technique in natural language processing (NLP) and speech recognition that, combined with NNs, enables the model to learn the underlying meaning of each embedded element.[88] Thereafter, this sequence is fed to RNN layers, responsible to understand the knowledge behind the SMILES sequence, and finally to FCNN layers, in charge to process the RNN output and translate it into the desired descriptor. The database was divided into two sets: training, with 18528 samples used on the learning process; testing, with 2290 samples employed to evaluate the model performance. Figure 4.3 (b) exhibit the model performance on the testing dataset for the reduction potential, with the mean absolute error (MAE) of 0.24 V.

With Linear and Neural Models ready, the organic universe may now be explored to uncover new possible high-potential cathodes for Li-ion batteries. In this regard, a total of 20 million new molecules have been analysed and their respective reduction potentials predicted by the Neural Model, serving in sequence as input for the Linear model in obtaining their voltages, referred here as AI-V_{OC} for simplification. Out of curiosity, the AI prediction step required

about 40 minutes for all the 20 million molecules on a personal computer, apart from the SMILES textual pre-processing. This task would require more than 10 years when considering state-of-the-art supercomputing facilities if quantum mechanics calculations were to be employed.

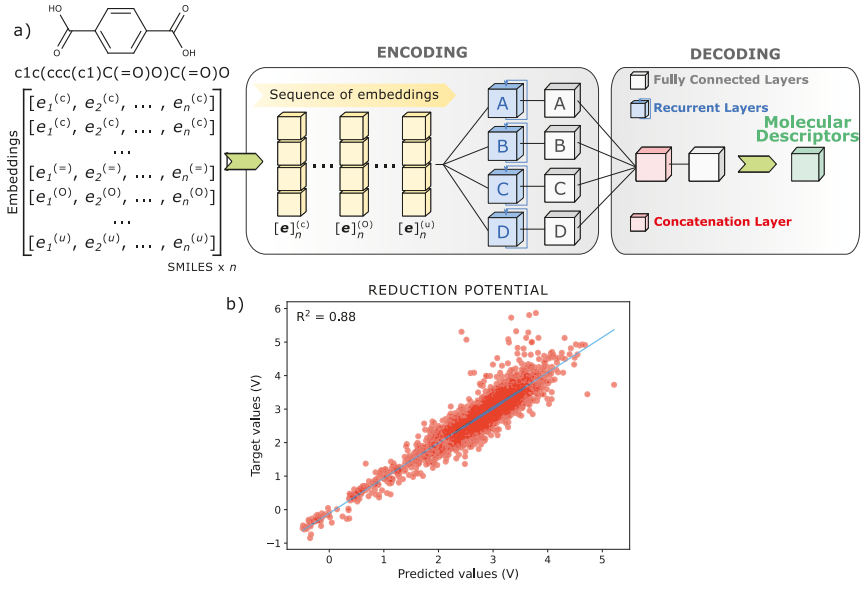


Figure 4.3. a) A simplified schematic of the Neural Model. As an example, the terephthalic acid has its SMILES converted into a sequence of embedding vectors before feeding the network layers. b) Neural Model performance in predicting reduction potentials of the 2290 molecules in the test dataset. Reprinted from Paper V. Copyright 2022, Elsevier.

This huge library has passed through a simple voltage filters in order to identify possible cathode candidates, with V_{OC} higher than 2.9 V (vs. Li^+/Li), and anodes, with V_{OC} between 0.0 V and 0.5 V (vs. Li^+/Li). Furthermore, the cathodes subset has been narrowed down by applying a theoretical capacity filter, selecting compounds with capacities higher than 100 mAh/g. The theoretical capacity may be calculated by following the Equation 4.1, in which n is the number of possible redox sites, F the Faraday constant and M the molecule molar mass.

$$C = \frac{nF}{3.6M} \tag{4.1}$$

The filtering process rendered 1001 cathodes and 6,014,051 anodes candidates, from which 1500 were randomly selected. Out of the 20 million moieties, only about 0.0050% were identified as possible cathodic materials.

This further highlight the challenge to find novel high-potential organic electrodes, thus the need for alternative approaches to accelerate materials discovery. Following the selection step, new DFT calculations has been carried out for all the 2501 molecules to evaluate their P_{Red} and V_{OC}. This step acts as a third-layer filter, part of the high throughput workflow, to improve the list of proposed molecules for production by cleaning the statistical noise offered by data-driven methods. At the same time, these calculations can be used to analyse the AI-*kernel* performance.

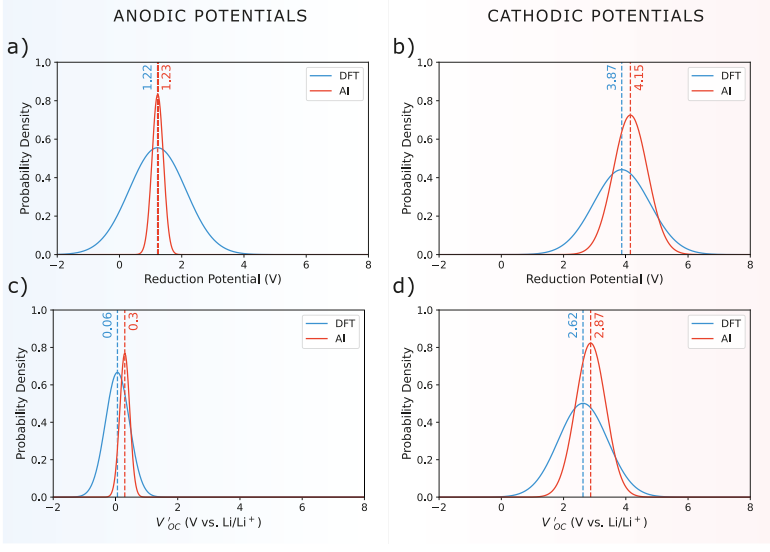
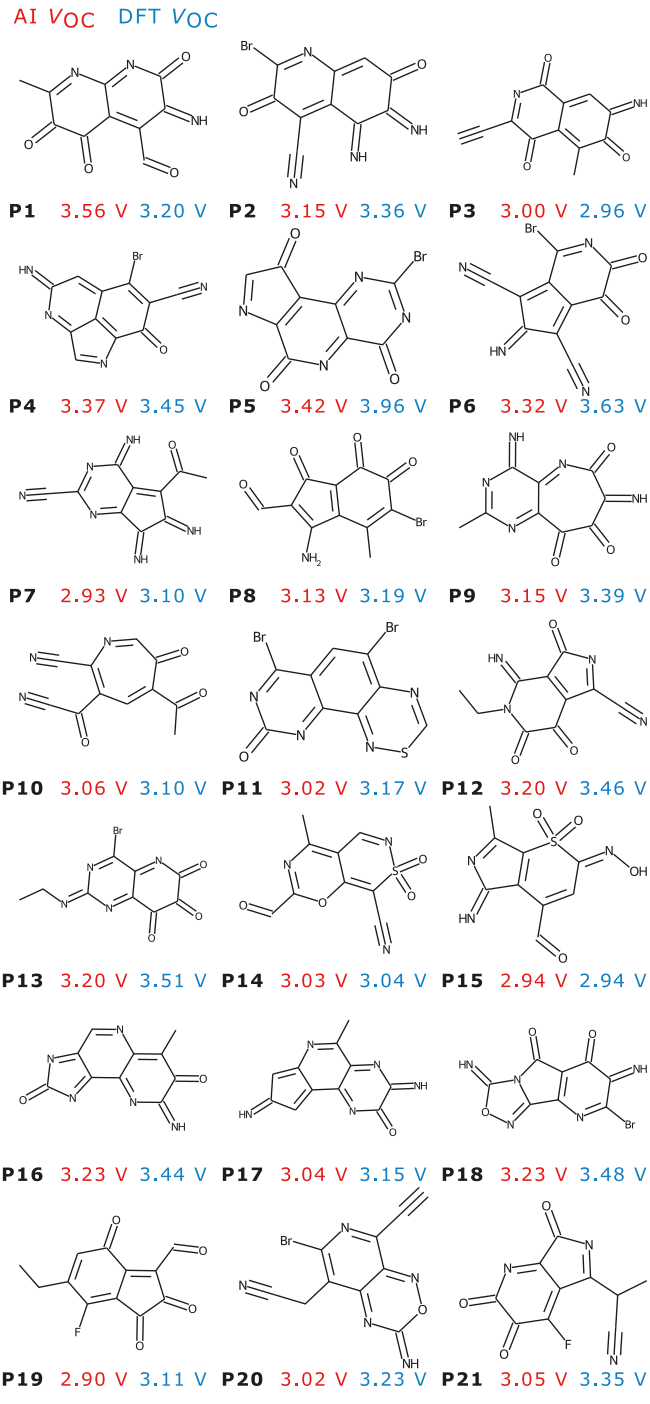


Figure 4.4. Reduction potentials and open-circuit voltages (V_{OC}) probability density functions for the selected anodes (a, c) and cathodes (b, d), respectively. Reprinted from Paper V. Copyright 2022, Elsevier.

Figure 4.4 shows a collection of probability density functions (PDFs) benchmarking AI and DFT results. These PDFs are a Gaussian fit to the data distribution indicating the likelihood to find a certain outcome, i.e., a certain value of P_{Red} and V_{OC}. In Figure 4.4 (a) and (b), the PDFs for the reduction potentials of anodic and cathodic compounds, respectively. Similarly, Figure 4.4 (c) and (d), the PDFs for the V_{OC} of anodes and cathodes, respectively. DFT V_{OC} was obtained by using the reduction potential from the DFT calculations together with the Linear Model to evaluated open-circuit voltages. From these results, it is clear that the machinery is correctly predicting low and high potential compounds, i.e., it can dissociate the two range of potentials. Small deviations caused by outliers are mainly a result of the DFT calculation step, in which these molecule-outliers went through significant structural changes upon the redox process. These changes often result in very large/small unrealistic voltages that widen the PDFs.



Scheme 4.2. Lewis representation for a small group of molecules extracted from the cathode candidates and their respective lithiation voltages (V_{OC} vs. Li^+/Li) as predicted by the AI-*kernel* and evaluated by using the DFT reduction potential. Reprinted from Paper V. Copyright 2022, Elsevier.

The Scheme 4.2 displays a few molecules from the cathode list with their respective AI/DFT voltages referred vs. Li^+/Li. It is interesting to note how the AI tends to accurately identify cathode candidates for molecules containing electron-withdrawing groups, like nitrile $-C{\equiv}N$, carbonyl $-C{=}O$ and halogens. Due to their effect over the redox-active centers, these groups have usually been applied as voltage elevators in organic electrodes for Li-ion batteries.[13] Conversely, electron-donating groups, like the amino ($-NH_2$), hydroxy ($-OH$) and alkyl, are also present for a few molecules. This combination of electron-withdrawing and electron-donating effects has been used in organic optoelectronics and photovoltaics as it may lead to donor-acceptor-like behaviour.[89–91] For Li-ion batteries, this effect may act as a mechanism to improve the charge localisation over the redox-active center by engineering the electronic structure of the compound, therefore enhancing the voltage upon lithium insertion. In addition, the redox sites of the selected cathodes appear to be dominated by carbonyls, nitro groups and double/triple bonding nitrogen, which are commonly associated with high potential electrodes in organic batteries. The combination of these results further corroborates the performance of the AI-framework in predicting the lithiation voltage of organic electroactive materials.

4.2 Case 06: Predicting the redox stability of OEMs with the AI-*kernel* and the performance of [Li, Na, K]-ion batteries

The AI-*kernel* was demonstrated in the previous section to be a powerful ally in discovering novel organic materials for Li-ion batteries. However, the *kernel* did not address an important issue regarding the screened materials: stability. Overall, stability is a complex topic that can lead to several discussions. To start tapping into this issue, the redox stability was investigated in the Paper VI, i.e., the stability of the molecule during redox reactions. Firstly, the OMEAD database was expanded from approximately 26000 to 41800 unique molecules following the same methodology described in the previous section and Paper V. For all these molecules, the *redox stability* was evaluated by analysing changes in the bond length of neighbouring atoms upon oxidation/reduction reactions within the DFT scope. Considering a cutoff radius of 2.1 Å to identify neighbouring atoms, the compounds were classified as redox

unstable (class no) if changes in the bond length were higher than 30%, and redox stable otherwise (class yes).

With this classification method, a small dataset of redox stable/unstable molecules was developed and used as training data for a new prediction layer within the AI-*kernel*. This new layer, here named Redox Stability Neural Model (RSNM) was based on molecular graphs as inputs and a combination of different Graph Neural Networks to predict the redox stability of a given molecule, i.e., to classify the molecule as redox stable/unstable. In other words, the RSNM answers with a "yes" or "no" the question: "is this molecule redox stable?"

Additionally, the Linear Model (LM), presented in the previous section and responsible for predicting the open-circuit voltage of Li-ion batteries from the molecular reduction potential, was further developed to include Na- and K-ion batteries. Now, the LM is subdivided in three models, [Li, Na, K]-LM, that can predict the open-circuit voltage of the respective ion insertion reaction, referred to the Li, Na and K reference electrode, respectively. Figure 4.5 (a) shows the new AI-*kernel* with its different layers and functionalities. In summary:

- **Redox Stability Layer**: *This layer runs the RSNM and is responsible to identify, and filter-out, redox unstable molecules.*
- **Reduction Potential Layer**: *In this layer, the neural model discussed in the previous section, predicts the reduction potential of the input molecule.*
- **Ion Layer**: *With the redox stable molecules and their respective reduction potentials, this layer predicts the battery open-circuit voltage for the desired ion insertion reaction.*

The reduction potential is once more predicted by the neural model, discussed in the previous section, here named Reduction Potential Neural Model (or RPNM) to avoid confusion with the redox stability model. Figure 4.5 (b) shows a general schematic of the RSNM, in which GC and ECC means Graph Convolution and Edge-Conditioned Convolutions, respectively. The final network architecture was obtained through an evolutionary approach, following a differential evolution mechanism.[92] Details about the different types of graph neural networks and the RSNM development is further discussed in Paper VI. After training, the model showed a prediction accuracy of 80% and 72% for the training and testing data, respectively. To avoid overfitting during the learning process, *early stopping* was employed together with other techniques. Otherwise, the training prediction accuracy would reach 99%, but the test accuracy would drop to 68%. With all layers of the AI-*kernel* settled, a high-throughput screening was performed on 45 million molecules to identify high-potential cathodes for Li-, Na- and K-ion batteries.

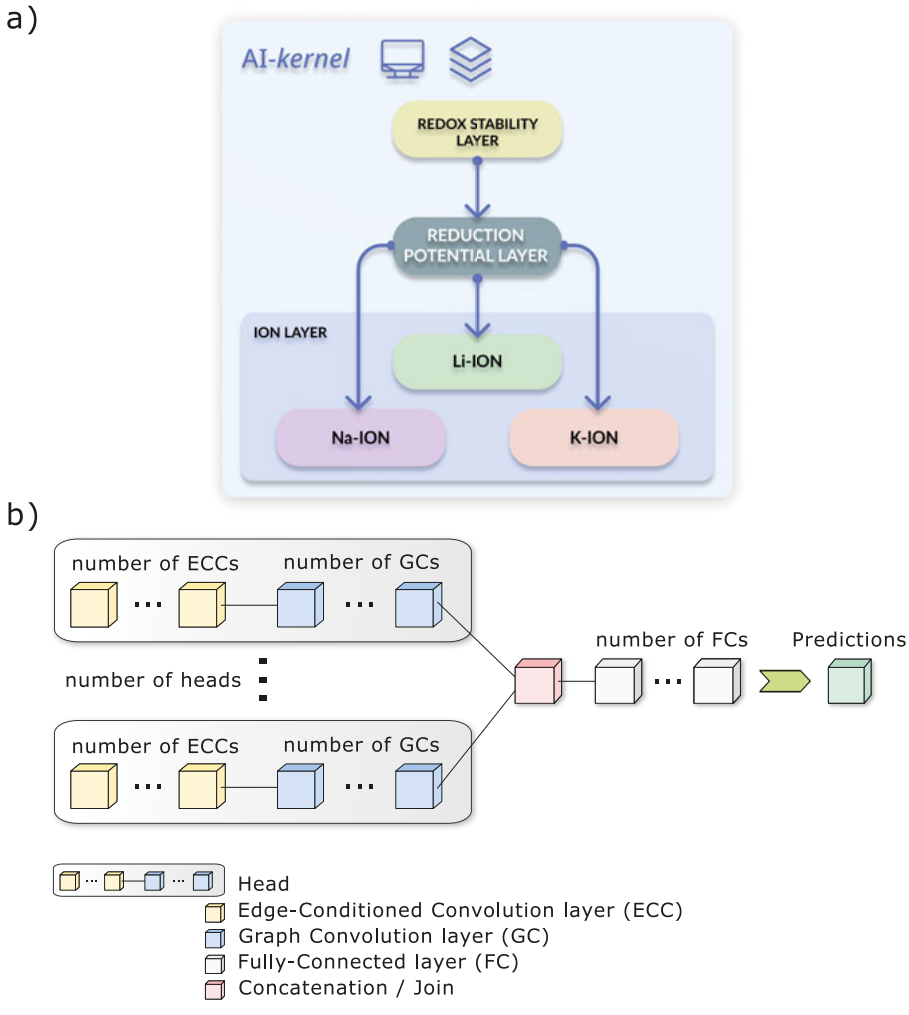


Figure 4.5. a) The AI-kernel summarized in a flowchart, showing the different sub-layers and functionalities. b) Schematics showing the overall Redox Stability Neural Model. Figure from Paper VI.

Similarly with the previous Case, a two-step selection process of voltages → capacities was performed to filter potential candidates. The threshold of 100 mAh/g was considered for capacities while for open-circuit voltages a cutoff of 3.0 V for Li and 2.6 V for Na and K. The screening resulted in 4047, 860 and 874 cathodes candidates for the Li-, Na- and K-ions, respectively. Thereafter, a DFT calculations round was carried out for all the selected candidates to better investigate molecular properties and to benchmark the *kernel*. Once again, probability density functions were employed in the task of comparing AI and DFT outcomes.

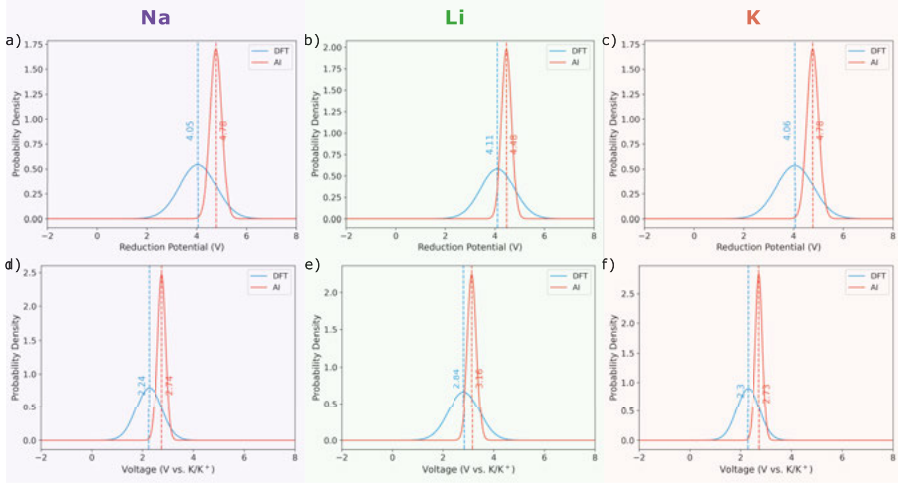


Figure 4.6. Probability density function (PDF) plots of reduction potentials for the a) Na-, b) Li- and c) K-ion cases comparing AI and DFT results. Similarly, PDF plots of open-circuit voltages for the d) Na, e) Li and f) K ion insertion reaction. Figure from Paper VI.

Figure 4.6 (a), (b) and (c) shows a collection of these evaluations for reduction potentials while Figure 4.6 (d), (e) and (f) for open-circuit voltages. From these plots, it is possible to see that DFT and AI values are in close agreement. However, DFT results show broader distributions and averages shifting to lower potentials. This happens as molecular geometries can significantly change after the DFT relaxation step, especially for reduced phases, largely contributing to the broadening of PDF distributions and lowering 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 kernel was able to correctly identify high-potential electrodes for all three ions. Finally, the RSNM accuracy in predicting the redox stability of the selected molecules was 74.4% when comparing with the DFT findings, i.e., molecules that did degraded during the DFT reduction reaction simulation. This shows the *kernel* performance and, once more, how it can be a powerful platform to accelerate materials discovery for battery applications.

5. Concluding remarks

As organic materials have the potential to positively impact the global Energy scenario, both in production and storage, a deeper understanding of them is necessary. In this sense, this thesis was divided into two interconnected parts: (i) an initial investigation of organic-based materials suitable as battery electrodes, uncovering structure-property relationships controlling the overall electrochemical performance of such compounds; (ii) the development of a novel data-driven framework, backed by artificial intelligence and as accurate as quantum mechanics calculations, to accelerate the discovery of novel OEMs.

Initially, it has been demonstrated here how to evaluate the thermodynamics of the ion insertion process, providing a way to understand and anticipate the electron transfer and storage mechanisms in molecules. Furthermore, a simple charge analysis has been proposed to identify the charge storage limits of OEMs, tapping into the "superlithiation" phenomenon, which is not entirely understood by the scientific community so far. Thereafter, this thesis shows a systematic way to predict the crystal structure of organic compounds by combining an evolutionary algorithm (EA) and density functional theory calculations. This methodology has correctly predicted the crystal structure of the disodium benzenediacrylate (Na_2BDA) with a good agreement with the experimentally resolved structure. Both experiment and prediction agree with the space group of the crystal structure, Na-ion coordination in the desodiated phase and lattice parameters. Validated by this agreement, the evolutionary approach has been employed to assess the crystal structure of the sodiated phases for Na_2BDA and the de/lithiated phases for the analogous compound Li_2BDA. Thereafter, the thermodynamics of the Na^+/Li^+ insertion reactions have been evaluated, displaying a good agreement with the experimentally reported electrochemistry. More precisely, 0.63 (0.6) V vs. Na/Na^+ for the Na-ion insertion and 1.12 (1.2) V vs. Li^+/Li for the Li-ion insertion as theoretically (experimentally) determined. In addition, this investigation has shown that the use of the correct theoretical formalism is essential to describe the thermodynamics of these reactions. For instance, the GGA-PBE functional has suggested a stepwise one-electron reaction for the ion insertion while the more robust hybrid functional HSE06 has indicated at two-step reaction of these compounds. This also appears to be on par with the reported experimental findings.

An efficient use of the evolutionary algorithm into engineering organic materials has also been demonstrated, with the tailoring of three anodic materials: dilithium terephthalate (Li_2TP), dilithium thiophene (Li_2TDC) and

dilithium benzodithiophene (Li_2BDTDC). By functionalisation with the nitro (-NO_2) group, these compounds have changed from a lower-potential anode to a high-potential cathode, with the Li_2BDTDC exhibiting the highest voltage shift from 0.92 to 2.66 V vs. Li^+/Li for the average lithiation voltage of the two-step reaction. This effect has further been investigated and correlated with the charge localisation induced by the NO_2 unit, which contributes to a better definition of the redox-active center (RAC) in these molecules. A well-defined RAC is usually present in high-potential electrodes for Li-ion batteries, in contrast to low-potential counterparts where the charge tends to be delocalized over the entire compound. This was also the case for the molecules without NO_2 present. An apparent structural stabilization of the molecular ring has been suggested as another effect of the charge localisation. Accordingly, the results here have shown that less charge is being shared within the ring as new electrons of the reduced phases are being localised onto the nitro group.

In general, the evolutionary algorithm was designed to be a global minimum search approach. However, metastable phases can control the lithiation mechanism in some battery electrodes. To tap into this, a new methodology was developed to identify metastable phases during the ion insertion process. Based on this novel method, here named MAP, the Li_2DHT compound was investigated and its lithiation mechanism uncovered, showing the contrasting results between EA and MAP. For this material, the lithiation process seems to be controlled by metastable phases. Therefore, this technique can be useful to investigate the battery operation and, combined with the EA, uncover which phases control the lithiation reaction. Overall, a valuable addition to battery materials manufacturing. Additionally, the Mapion software, derived from the development of this technique, was distributed on an open-source basis.

In the latter part of this thesis, the discussion about the possible upscaling of this methodology for a larger number of materials has been introduced. As the evolutionary algorithm requires a significant computational effort, alternatives are required for an extensive screening of the organic universe. In this sense, another pathway has been proposed by means of machine learning techniques. Backed by a substantial number of quantum mechanics calculations, this novel approach rendered a powerful framework, the AI-*kernel* capable of predicting the lithiation voltage of electrodes based solely on the SMILES string of the molecule desired as the active unit. Firstly, for a small set of predicted organic crystals, this investigation has shown that the reduction potential of the molecular units correlates with the lithiation voltages in the solid-state, resulting in a linear model to assess the latter by only supplying the former. Henceforth, a database, named Organic Materials for Energy Applications Database (OMEAD), comprised of organic molecules has been developed using DFT calculations, having several electronic and thermodynamic properties extracted from the calculations.

The AI-*kernel* was designed by combining two distinct models: (1) a neural network architecture, trained on top of the OMEAD, capable of predicting the reduction potentials from molecular SMILES; (2) a linear model responsible to predict battery open-circuit voltages (vs. Li^+/Li) from molecular reduction potentials. In this manner, a molecule/SMILES has its reduction potential evaluated in (1), which is used by (2) to predict the corresponding open-circuit voltage. Thereafter, the AI has been employed to assess the lithiation voltages of 20 million new molecules extracted from the organic universe in order to identify possible cathode candidates for organic batteries. It is important to stress that this task would require several years to be accomplished by DFT calculations even if state-of-the-art supercomputing facilities available to date were to be used, not to mention the challenging logistic to store the data generated by the quantum mechanics calculations. Finally, the AI predictions have been verified and endorsed by new DFT calculations carried out for the suggested molecules, which has shown that the model was capable of correctly locating anodic and cathodic range of potentials. Moreover, the AI has appeared to accurately understand the connection between some functional groups and high voltage compounds. Interestingly, the proposed molecules present an intriguing combination of electron-withdrawing and electron-donating groups, which may lead to a donor-acceptor-like behaviour that could enhance the redox-active center, being one of the reasons for the cathodic characteristics. Finally, the AI-*kernel* has been expanded to include Na- and K-ion batteries. In this way, the kernel now can be employed to screen the library of organic materials for different alkali batteries. Furthermore, a new model was included to predict the redox stability of the desired molecular compound, i.e., to anticipate if the molecule would degrade under redox reactions. This third model is based on graph neural networks and relies on molecular graphs as inputs. The AI-*kernel* design demanded the development of a novel machine learning platform, named ANIMA and coded in python. ANIMA is distributed under an open-source license.

Aside from the advances introduced so far in understanding and designing organic electroactive materials, additional efforts are still needed. In this sense, it is a natural process that future works shall promote investigations of different phenomena regarding OEMs, such as interfacial effects, improvements of electronic and ionic conductivity, etc. Moreover, and albeit the presented AI performance, an upscaling of this methodology shall also be promoted to go even further into the organic realm.

In light of these outcomes, this thesis helps to move towards the realisation of a next generation of battery materials by means of organic compounds – an alternative, greener and more sustainable solution to current technologies. Moreover, it provides an understanding of fundamental aspects of organics when working as battery materials by, for example, shedding light on the lithiation mechanism and the underlying electronic changes upon the redox

reactions for these materials. This knowledge is useful to overcome different limitations of the OEMs known to date regarding energy storage performance and cyclability. In addition, the AI-driven platform has addressed the bottleneck of discovering novel suitable materials, providing an accelerated and streamlined workflow capable of breaking through current state-of-the-art limitations. Nevertheless, organics still have a long and challenging road ahead before becoming a competitive technology and, in short, this thesis contributes to solving some of these challenges.

Popular Science Summary

It is undeniable that batteries are of utmost importance to our society. They have become a cornerstone for different technologies, especially with the advent of portable devices, such as smartphones, wearables, computers, etc., and electric vehicles. Furthermore, batteries must have desirable performances to meet their application's requirements. After all, no one wants to have a "dead" smartphone by the end of the day. Overall, these facts have been driving a great deal of research from the materials science and industry in order to design better batteries, with much attention given to the Li-ion battery technologies. However, this fundamentally depends on the underlying material employed in battery manufacturing. Nowadays, when it comes to battery electrodes, for example, state-of-the-art technology relies on inorganic materials. In general, the negative electrode (or anode) is usually based on graphite while the positive electrode (or cathode) uses transition metal oxides (TMOs), based on cobalt, iron, aluminium, manganese, among others. These materials are heavily dependent on mining processes, therefore a non-renewable resource. On top of that, the mining and processing of these compounds result in severe environmental hazards, with possible catastrophic impacts. In addition, the actual low efficiency in recycling these batteries usually result in a linear value chain.

Moving towards more sustainable alternatives, with less to no environmental damage, this Thesis dives into an exploratory quest for organic-based alternatives for these battery materials. Organics can offer a set of interesting features, such as clean synthesis routes from renewable sources, tunable properties through molecular engineering and a closed life cycle, i.e., a closed loop value chain, illustrated in Figure 5.1. In fact, several organic electrodes have been proposed in the past with reasonable performance. However, these materials still presented some drawbacks that need to be addressed before an "Organic Battery" could be successfully proposed for production. For example, some of these materials offer low energy densities, which means more times a day with that smartphone plugged in. Discovering novel organic battery materials with enhanced properties is one path to address this issue. However, this task can be exceptionally challenging if we consider how vast the universe of organic compounds is. Possibly, a limitless number of molecules and polymers are yet to be discovered, and synthesising all of them in a laboratory is exceptionally impossible. Therefore, computational approaches are necessary.

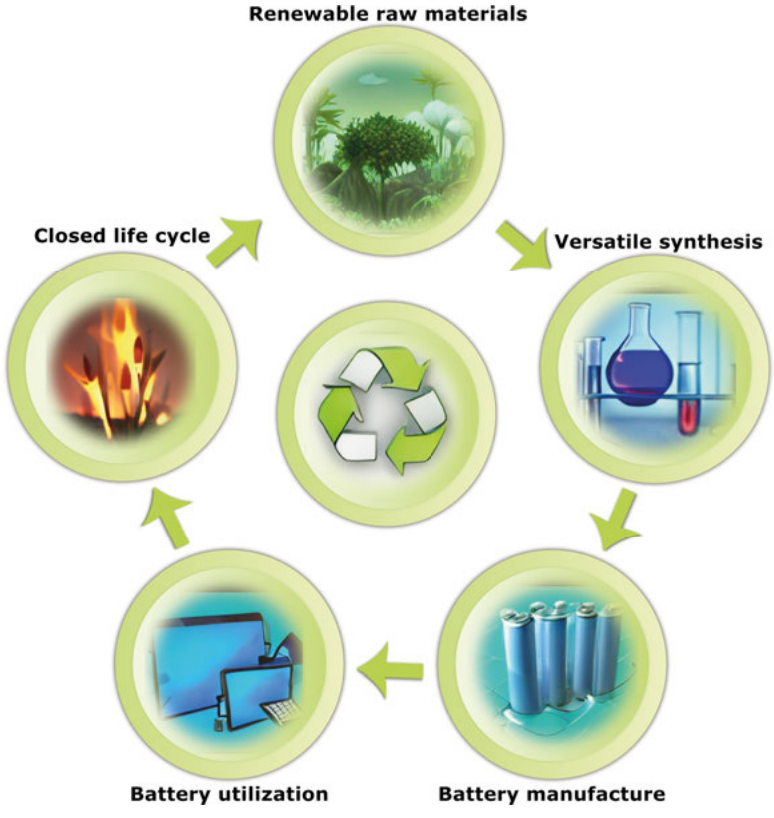


Figure 5.1. The closed loop value chain of organic-based batteries brings sustainability into the battery manufacturing value chain. The small illustrations in this Figure were generated by an artificial intelligence algorithm found at craiyon.com.

In the first part of this Thesis, a computational quantum mechanics framework is employed to investigate several organic electrode candidates. From predicting the structure of these compounds to assessing their electronic properties, the Thesis uncovers structure-property relationships controlling the battery's performance. These results are consistently backed by experimental findings whenever possible, ensuring the overall accuracy of the method. From this part, it is possible to grasp what molecular properties affect desirable battery features, like the open-circuit voltage and maximum energy storage capacity. Furthermore, it is shown alternative pathways to engineering organic molecules to improve a given aspect of the battery, such as voltage and capacity. However, this quantum mechanics-based approach, although accurate, is substantially demanding when aiming to screen the limitless library of organic materials. Thus, an alternative route backed by data-driven techniques is discussed in the second part of the Thesis.

This data-driven method, based on artificial intelligence (AI), is developed on top of the aforementioned information. The AI learned from the quan-

tum mechanics framework, deriving in a complex platform, here named AI-*kernel*, capable of predicting battery properties such as open-circuit voltages and assessing stability aspects of a given molecule or electrode candidate. This framework completely bypasses the demanding computational quantum mechanics step, immensely accelerating the process of materials discovery. For instance, the AI-*kernel* was employed in screening 45 million novel molecules, searching for high-performance cathodes for Li-, Na and K-ion batteries. This search was performed in a matter of hours, which would demand more than two decades if following traditional computational quantum mechanics methods. This screening resulted in a shortlist of proposed organic electrode candidates with the potential to be future battery materials with not only desirable performance but also offering the aimed sustainability for this class of energy storage technology.

Svensk Sammanfattning

Batterier är oförnekligen en nyckelkomponent i dagens samhälle. De har blivit en förutsättning för dagens teknologi, särskilt med tillkomsten av bärbara och portabla enheter, såsom smartphones, smartklockor, laptops, etc. samt inte minst elfordon. Därmed ställs det höga krav på batteriers prestanda att uppfylla dessa olika applikationers alla krav. När allt kommer omkring vill ingen att ens smartphone ska ”dö” innan dagen är slut, och många vittnar om laddningsångest när det gäller sina elfordon. Att utveckla batterier har därför lett till mycket forskning inom akademisk materialvetenskap och tillhörande industri, särskilt med avseende på litium-jonbatterier. Prestandan hos dessa beror främst på de underliggande materialen som används i batterierna. Dagens batterielektroder är baserade på främst oorganiska material. Exempelvis är den negativa elektroden (eller anoden) vanligtvis baserad på grafit medan den positiva elektroden (eller katoden) använder övergångsmetalloxider (TMO) baserade på bland annat kobolt, järn, aluminium och/eller mangan. Dessa resurser är icke-förnybara eftersom utvinningen sker ur gruvor, och brytningen medför ofta allvarliga miljörisker. Dessutom är effektiviteten vid återvinningen av dessa material låg, vilket leder till en linjär – och inte cirkulär – värdekedja.

Denna avhandling har ett fokus på hållbara materialalternativ, där mindre eller inga miljöskador förekommer. Därmed strävar forskningen till att utforska organiskt baserade alternativ som ersättning för dagens batterimaterial. Organiska material uppvisar en löftesrik uppsättning egenskaper såsom rena syntesvägar från förnybara källor, kontrollerbara egenskaper genom molekylär ingenjörskonst, och en sluten livscykel. Med andra ord, en cirkulär materialekonomi blir möjlig, vilket illustreras i Figur 5.2. Faktum är också att flera organiska elektroder har uppvisat en rimlig prestanda avseende en eller flera egenskaper som krävs av ett batteri. Dock uppvisar dessa material ett antal problem som måste åtgärdas innan ett ”organiskt batteri” framgångsrikt tas i produktion. Exempelvis innehar dessa material en låg energidensitet, vilket kan översättas till att en telefon behöver laddas oftare eller är större.

Att upptäcka nya organiska material med förbättrade egenskaper är därmed önskvärt för att lösa dessa problem. Denna uppgift kan dock vara exceptionellt utmanande, om vi tar hänsyn till hur stort universumet av organiska föreningar är. Det existerar ett nästan obegränsat antal molekyler och polymerer som inte ens är upptäckta än, och det är osäkert om de kan syntetiseras i ett laboratorium. Detta är också exceptionellt tidskrävande. Därmed är materialvetenskapliga beräkningsmetoder nödvändiga för att lösa denna uppgift.

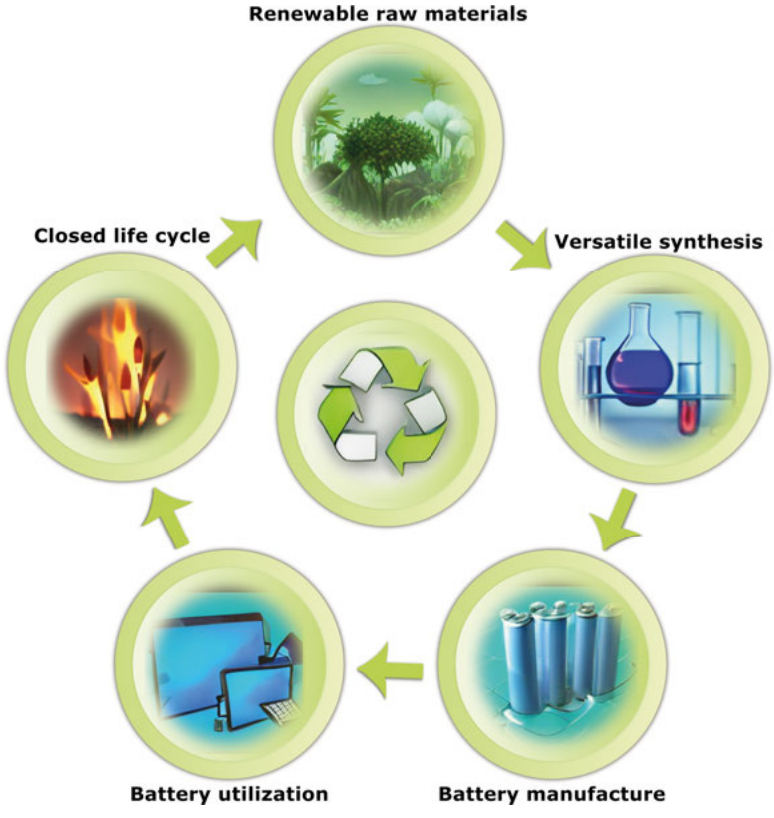


Figure 5.2. En cirkulär värdekedja för ekologiskt baserade batterier vilket möjliggör en hållbar batteritillverkning. De små illustrationerna i denna figur genererades av en artificiell intelligensalgoritm som finns på craiyon.com.

I den första delen av denna avhandling används kvantmekaniska metoder för att undersöka vissa specifika organiska elektrodkandidater. Genom att förutse strukturen hos dessa föreningar kan deras elektroniska egenskaper avgöras, vilket avslöjar struktur-egenskapsförhållanden, vilket i sin tur styr batteriets prestanda. Dessa beräkningsvetenskapliga resultat stöds av experiment när detta är möjligt, vilket stärker metodernas övergripande noggrannhet. Detta gör det möjligt att förstå vilka molekylära egenskaper som påverkar önskvärda batteriegenskaper, som spänning och maximal laddningskapacitet. Dessutom påvisas alternativa vägar till att konstruera organiska molekyler för att förbättra en given aspekt av batteriet, såsom spänning och kapacitet.

Emellertid, även om detta beräkningsmässiga ramverk i princip är en korrekt metod att använda, så kräver det en avsevärd mängd undersökningar av den näst intill gränslösa floran av organiska material. Således utvecklades en alternativ väg baserad på datadrivna metoder i den andra delen av avhandlingen. Denna datadrivna metod, baserad på artificiell intelligens (AI), är utvecklad genom användandet av den tidigare nämnda informationen. AI:n utveck-

lades därmed inom det kvantmekaniska ramverket, härlett från en sammansatt plattform, vilket här kallas för en AI-*kernel*. Denna kan användas till att kunna förutsäga batteriegenskaper såsom batterispänning eller stabilitetsaspekter hos en given molekyl eller elektrodkandidat. Fördelen med detta ramverk är att det förbigår det krävande kvantmekaniska steget i beräkningarna, vilket påskyndar processen för att upptäcka nya material. Exempelvis användes denna AI-*kernel* för att gå igenom hela 45 miljoner nya molekyler, för att bland dessa söka efter högpresterande katoder för Li-, Na- och K-jonbatterier. Denna undersökning utfördes av datorn på några timmar, men skulle kräva mer än två decennier om traditionella kvantmekaniska beräkningsmetoder användes. Detta resulterade i en kortare lista över föreslagna organiska elektrodkandidater vilka har potential att bli framtida batterimaterial: inte bara med önskvärd prestanda, utan också eftersträvandes hållbarhet.

Acknowledgements

First of all, I would like to thank you for reading my thesis and being interested in my work. I sincerely hope that you have found a pleasant read and an interesting scientific discussion that may help you on your journey.

Naturally, I am grateful to my advisors Moyses Araujo and Daniel Brandell for all the support and guidance so far. I know how "cliché" this will sound, but I could not have asked for better advisors. Thanks for the challenges, for the encouragements and the understanding (even on my "stubborn" moments). I also would like to thank Susanne Mirbt for all the kind moments of guidance and motivation. Mentions to Cleber and Muhammad for all the scientific discussions about organic batteries that may, or may not, have had a relation with the main goal of my work. Both of you helped a lot in my understanding of Organic Chemistry.

I am deeply thankful to my family for all the support, love and understanding that they always have offered me. The most difficult part of these four years of doctoral studies was being far from all of you! Thanks, Dad! I will forever miss you!

I would also like to thank all my friends from all over the world for every moment we have had together during this time – you are a really important share of my life and, even not realising, somehow contributed to this work. Thanks to my colleagues at Uppsala University for all the time together, the fikas, pub nights and the relaxing chats. **Obrigado**!

References

1. P. T. Brown, K. Caldeira, *Nature* **2017**, *552*, 45–50.
2. M. Romare, L. Dahllöf, The Life Cycle Energy Consumption and Greenhouse Gas Emissions from Lithium-Ion Batteries, tech. rep. C 243, **2017**, p. 58.
3. T. B. Schon, B. T. McAllister, P.-F. Li, D. S. Seferos, *Chem. Soc. Rev.* **2016**, *45*, 6345–6404.
4. B. Haeupler, A. Wild, U. S. Schubert, *Advanced Energy Materials* **2015**, *5*, 1402034.
5. S. Renault, D. Brandell, K. Edström, *ChemSusChem* **2014**, *7*, 2859–2867.
6. M. E. Bhosale, S. Chae, J. M. Kim, J.-Y. Choi, *J. Mater. Chem. A* **2018**, *6*, 19885–19911.
7. S. Muench, A. Wild, C. Friebe, B. Haupler, T. Janoschka, U. S. Schubert, *Chemical reviews* **2016**, *116*, 9438–9484.
8. T. Janoschka, M. D. Hager, U. S. Schubert, *Advanced Materials* **2012**, *24*, 6397–6409.
9. K. Nakahara, K. Oyaizu, H. Nishide, *Chemistry Letters* **2011**, *40*, 222–227.
10. D. Mecerreyes, L. Porcarelli, N. Casado, *Macromolecular Chemistry and Physics* **2020**, *221*, 1900490.
11. X. Jia, Y. Ge, L. Shao, C. Wang, G. G. Wallace, *ACS Sustainable Chemistry & Engineering* **2019**, *7*, 14321–14340.
12. Y. Liang, Z. Tao, J. Chen, *Advanced Energy Materials* **2012**, *2*, 742–769.
13. 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, **2018**.
14. D. L. Williams, J. J. Byrne, J. S. Driscoll, *Journal of The Electrochemical Society* **1969**, *116*, 2.
15. M. Saubanère, J.-S. Filhol, M.-L. Doublet in *Physical Multiscale Modeling and Numerical Simulation of Electrochemical Devices for Energy Conversion and Storage: From Theory to Engineering to Practice*, (Eds.: A. A. Franco, M. L. Doublet, W. G. Bessler), Springer London, London, **2016**, pp. 1–36.

16. M. Armand, S. Grugeon, H. Vezin, S. Laruelle, P. Ribière, P. Poizot, J. M. Tarascon, *Nature Materials* **2009**, *8*, 120–125.

17. A. Jouhara, N. Dupré, A.-C. Gaillot, D. Guyomard, F. Dolhem, P. Poizot, *Nature Communications* **2018**, *9*, 4401.

18. K. Oyaizu, H. Nishide, *Advanced Materials* **2009**, *21*, 2339–2344.

19. T. Suga, Y. J. Pu, S. Kasatori, H. Nishide, *Macromolecules* **2007**, *40*, 3167–3173.

20. Y. Lu, Q. Zhang, L. Li, Z. Niu, J. Chen, *Chem* **2018**, *4*, 2786–2813.

21. M. Tang, S. Zhu, Z. Liu, C. Jiang, Y. Wu, H. Li, B. Wang, E. Wang, J. Ma, C. Wang, *Chem* **2018**, *4*, 2600–2614.

22. S. Renault, V. A. Oltean, C. M. Araujo, A. Grigoriev, K. Edström, D. Brandell, *Chemistry of Materials* **2016**, *28*, 1920–1926.

23. P. Kirkpatrick, C. Ellis, Chemical space, **2004**.

24. T. Hoffmann, M. Gastreich, The next level in chemical space navigation: going far beyond enumerable compound libraries, **2019**.

25. L. Ruddigkeit, R. Van Deursen, L. C. Blum, J. L. Reymond, *Journal of Chemical Information and Modeling* **2012**, *52*, 2864–2875.

26. European Space Agency, ESA: Second Data Release from ESA's GAIA Mission, **2018**.

27. Merck, Aldrich Market Select Chemistry Services, **2018**.

28. P. Hohenberg, W. Kohn, *Physical Review* **1964**, *136*, B864–B871.

29. C. C. Roothaan, *Reviews of Modern Physics* **1951**, *23*, 69–89.

30. C. Møller, M. S. Plesset, *Physical Review* **1934**, *46*, 618–622.

31. I. Shavitt in *Methods of Electronic Structure Theory*, **1977**, pp. 189–275.

32. G. Kresse, J. Hafner, *Physical Review B* **1993**, *47*, 558–561.

33. G. Kresse, J. Furthmüller, *Physical Review B - Condensed Matter and Materials Physics* **1996**, *54*, 11169–11186.

34. G. Kresse, D. Joubert, *Physical Review B - Condensed Matter and Materials Physics* **1999**, *59*, 1758–1775.

35. M. J. Frisch, G. W. Trucks, H. E. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. Petersson, F. O., J. B. Foresman, J. D. Fox, Gaussian 16, **2016**.

36. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. a. Persson, *APL Materials* **2013**, *1*, 11002.

37. S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, *npj Computational Materials* **2015**, *1*, 15010.

38. F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, K. R. Müller, *Nature Communications* **2017**, *8*, 1–10.

39. V. Botu, R. Ramprasad, *International Journal of Quantum Chemistry* **2015**, *115*, 1074–1083.

40. G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, *Scientific Reports* **2013**, *3*, 1–6.

41. J. Lee, A. Seko, K. Shitara, K. Nakayama, I. Tanaka, *Physical Review B* **2016**, *93*, DOI 10.1103/PhysRevB.93.115104.

42. B. Olsthoorn, R. M. Geilhufe, S. S. Borysov, A. V. Balatsky, *Advanced Quantum Technologies* **2019**, *2*, 1900023.

43. O. Allam, B. W. Cho, K. C. Kim, S. S. Jang, *RSC Advances* **2018**, *8*, 39414–39420.

44. T. Tsuneda, *Density Functional Theory in Quantum Chemistry*, **2014**.

45. M. Born, R. Oppenheimer, *Annalen der Physik* **1927**, *389*, 457–484.

46. L. H. Thomas, *Mathematical Proceedings of the Cambridge Philosophical Society* **1927**, *23*, 542–548.

47. E. Fermi, *Zeitschrift für Physik* **1928**, *48*, 73–79.

48. W. Kohn, L. J. Sham, *Physical Review* **1965**, *140*, A1133.

49. P. C. Hohenberg, W. Kohn, L. J. Sham, *Advances in Quantum Chemistry* **1990**, *21*, 7–26.

50. R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, *The Journal of Chemical Physics* **1977**, *68*, 3801–3807.

51. W. Kohn, A. D. Becke, R. G. Parr, *Journal of Physical Chemistry* **1996**, *100*, 12974–12980.

52. P. A. M. Dirac, *Mathematical Proceedings of the Cambridge Philosophical Society* **1930**, *26*, 376–385.

53. C. F. Weizsäcker, *Zeitschrift für Physik* **1935**, *96*, 431–458.

54. J. P. Perdew, K. Burke, M. Ernzerhof, *Physical review letters* **1996**, *77*, 3865.

55. J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048–5079.

56. A. D. Becke, *Physical Review A* **1988**, *38*, 3098–3100.

57. C. Lee, W. Yang, R. G. Parr, *Physical Review B* **1988**, *37*, 785–789.

58. J. Heyd, G. E. Scuseria, M. Ernzerhof, *Journal of Chemical Physics* **2003**, *118*, 8207–8215.

59. S. Grimme, *Journal of Computational Chemistry* **2006**, *27*, 1787–1799.

60. R. Ditchfield, W. J. Hehre, J. A. Pople, *The Journal of Chemical Physics* **1971**, *54*, 720–723.

61. 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 in Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, **2016**, pp. 265–283.

62. 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, *arXiv preprint arXiv:1603.04467* **2016**.

63. A Paszke, S Gross, S Chintala, G Chanan, E Yang, ... in NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques, **2017**.

64. 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 in Advances in Neural Information Processing Systems, *Vol. 32*, **2019**.

65. Y. Lecun, Y. Bengio, G. Hinton, Deep learning, **2015**.

66. B. Karlik, A. V. Olgac, *International Journal of Artificial Intelligence and Expert Systems* **2011**, *1*, 111–122.

67. L. R. Medsker, L. C. Jain, *Book* **2000**.

68. Z. C. Lipton, J. Berkowitz, C. Elkan, *arXiv preprint arXiv:1506.00019* **2015**.

69. M. Rupp, A. Tkatchenko, K. R. Müller, O. A. Von Lilienfeld, *Physical Review Letters* **2012**, *108*, DOI 10.1103/PhysRevLett.108.058301.

70. G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, O. A. Von Lilienfeld, K. R. Müller in Advances in Neural Information Processing Systems, *Vol. 1*, **2012**, pp. 440–448.

71. H. Huo, M. Rupp, *arXiv preprint arXiv:1704.06439* **2017**, *13754*.

72. C. D. Manning, H. Schütze, G. Weikurn, *SIGMOD Record* **2002**, *31*, 37–38.

73. J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, *AI Open* **2020**, *1*, 57–81.

74. R. P. Carvalho, C. F. N. Marchiori, C. M. Araujo, D. Brandell in *Redox Polymers for Energy and Nanomedicine*, The Royal Society of Chemistry, **2021**, pp. 93–136.

75. Y. S. Meng, M. E. Arroyo-De Dompablo, First principles computational materials design for energy storage materials in lithium ion batteries, **2009**.

76. H. H. Lee, Y. Park, K.-H. Shin, K. T. Lee, S. Y. Hong, *ACS Applied Materials & Interfaces* **2014**, *6*, PMID: 25285535, 19118–19126.

77. R. Martoňák, A. Laio, M. Parrinello, *Physical Review Letters* **2003**, *90*, 4.

78. J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, V. Caignaert, *Nature* **1990**, *346*, 343–345.

79. S. Goedecker, *Journal of Chemical Physics* **2004**, *120*, 9911–9917.

80. A. R. Oganov, C. W. Glass, *Journal of Chemical Physics* **2006**, *124*, 244704.

81. C. W. Glass, A. R. Oganov, N. Hansen, *Computer Physics Communications* **2006**, *175*, 713–720.

82. A. O. Lyakhov, A. R. Oganov, H. T. Stokes, Q. Zhu, *Computer Physics Communications* **2013**, *184*, 1172–1182.

83. V. A. Mihali, S. Renault, L. Nyholm, D. Brandell, *RSC Advances* **2014**, *4*, 38004–38011.

84. S. Renault, D. Brandell, T. Gustafsson, K. Edström, *Chemical Communications* **2013**, *49*, 1945–1947.

85. S. Wang, L. Wang, K. Zhang, Z. Zhu, Z. Tao, J. Chen, *Nano Letters* **2013**, *13*, PMID: 23978244, 4404–4409.

86. P. V. Balachandran, A. A. Emery, J. E. Gubernatis, T. Lookman, C. Wolverton, A. Zunger, *Physical Review Materials* **2018**, *2*, DOI `10.1103/PhysRevMaterials.2.043802`.

87. K. Hatakeyama-Sato, T. Tezuka, Y. Nishikitani, H. Nishide, K. Oyaizu, *Chemistry Letters* **2019**, *48*, 130–132.

88. S. Bengio, G. Heigold in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, **2014**, pp. 1053–1057.

89. G. Damas, C. F. Marchiori, C. M. Araujo, *Journal of Physical Chemistry C* **2018**, *122*, 26876–26888.

90. G. B. Damas, F. Von Kieseritzky, J. Hellberg, C. F. Marchiori, C. M. Araujo, *Journal of Physical Chemistry C* **2019**, *123*, 30799–30808.

91. X. He, B. Cao, T. C. Hauger, M. Kang, S. Gusarov, E. J. Luber, J. M. Buriak, *ACS Applied Materials and Interfaces* **2015**, *7*, 8188–8199.

92. R. Storn, K. Price, *Journal of global optimization* **1997**, *11*, 341–359.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2177

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-481583

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022