Journal of Materials Chemistry C

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: J. Mater. Chem. C, 2021, 9, 13557

High-throughput virtual screening for organic electronics: a comparative study of alternative strategies

Ömer H. Omar, Marcos del Cueto, Tahereh Nematiaram and Alessandro Troisi *

We present a review of the field of high-throughput virtual screening for organic electronics materials focusing on the sequence of methodological choices that determine each virtual screening protocol. These choices are present in all high-throughput virtual screenings and addressing them systematically will lead to optimised workflows and improve their applicability. We consider the range of properties that can be computed and illustrate how their accuracy can be determined depending on the quality and size of the experimental datasets. The approaches to generate candidates for virtual screening are also extremely varied and their relative strengths and weaknesses are discussed. The analysis of high-throughput virtual screening is almost never limited to the identification of top candidates and often new patterns and structure–property relations are the most interesting findings of such searches. The review reveals a very dynamic field constantly adapting to match an evolving landscape of applications, methodologies and datasets.

Received 12th July 2021,
Accepted 13th September 2021

DOI: 10.1039/d1tc03256a

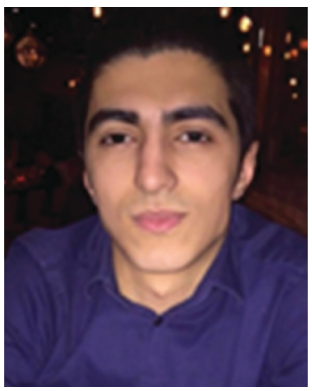
rsc.li/materials-c

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1. Introduction

The vision of employing organic materials as active components of electrical or optical devices, put forward and very actively pursued from the 80s[1] has been remarkably fruitful with a range of products that have reached the mass market, like organic light-emitting diodes (OLEDs),[2] prototype devices that approach their more established competitors, organic photovoltaic (OPV) devices,[3] and components of flexible electronic devices like conductive inks.[4] The field has been able to renew itself and identify new challenges, such as the development of novel emissive materials (dual emission,[5] room temperature phosphorescence,[6] thermally activated delayed fluorescence (TADF)[7]), the exploitation of multiexcitonic states (singlet fission[8] (SF) and up-conversion[9]) and the application into novel domains like organic bioelectronics,[10] neuromorphic[11] and quantum computing.[12] The premise for the successes and the optimism

Department of Chemistry, University of Liverpool, Liverpool, L69 3BX, UK.
E-mail: a.troisi@liverpool.ac.uk



Ömer H. Omar

Ömer H. Omar graduated from the University of Liverpool with an MChem in Chemistry in 2019. He is currently pursuing his PhD at the University of Liverpool under the supervision of Professor Alessandro Troisi. His current research interest lies in using high-throughput virtual screening to find novel materials for organic electronics.

Marcos del Cueto

Marcos del Cueto is a research associate in the Department of Chemistry at the University of Liverpool. He received his PhD in Madrid (Spain) and later did a postdoctoral stay in New Mexico (USA), studying the dynamics and reactive scattering of molecules on surfaces. He currently focuses on the use of data-driven methods to accelerate materials discovery. He specialises in modelling catalysts, solar cells and other energy materials, and the development of predictive models to guide the discovery of novel efficient materials.

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

about the new challenges ahead is that organic materials for electronics can be fine-tuned with exquisite precision to have the desired electronic characteristics and the processing characteristics required for fabrication. While there is a substantial component of chemical fine-tuning in the progress of these fields, the greatest advances coincide with the introduction of novel material classes or model materials (*e.g.* semicrystalline polymers,[13] solution-processable high-mobility crystals,[14] n-type polymeric semiconductors[15]). Such breakthroughs have been historically the result of very extensive labour and an understanding of the physical principles, which only became more consolidated recently.[16–18]

True exploitation of the power of organic synthesis in electronic materials comes with the availability of reliable models that allow realistic prediction of properties, and it is therefore not surprising that computational modelling has accompanied the development of organic electronics throughout.[19] The ambition to guide the discovery of new materials and contribute to innovative breakthroughs is instead more recent[20–22] and has been promoted by (i) the availability of low-cost computational infrastructure, (ii) the facility of access to cheminformatics tools and databases, (iii) the increased robustness of quantum chemical methods and (iv) the penetration of data science methods in chemical and materials discovery.[23] For the purpose of this work we can define high-throughput virtual screening (HTVS) as the computational investigation of a large set of compounds or materials to assess their suitability for a particular function. By "large", here we simply mean sufficiently large to prevent the human inspection of the individual instances and requiring independent statistical validation of the accuracy of the procedure and automatic analysis of the output.

A survey of the literature reveals that works reporting HTVS of organic materials address the same series of questions (what can be computed, how are the results validated against experiment, how can one generate a dataset of candidates to screen and how can the results be interpreted or used). For each specific problem, one encounters multiple bifurcations and, as a result, there are almost limitless ways to perform HTVS for the very same problem (see scheme in Fig. 1). However, most of these choices are independent of each other and it should be possible to define an optimal workflow by analysing each component independently and adopting the best practice of each of them. The importance of developing more standardised workflows can be appreciated from Table 1, which reports a non-exhaustive list of organic electronics technologies and specific properties that require the identification of specialised molecular materials (alongside a representative bibliographic reference with more detail on the technology). The growing number of applications and the variety of requirements from different technologies suggest that HTVS of organic materials will continue to accompany this research field for many years to come and a standardised framework to analyse and improve HTVS would be very beneficial.

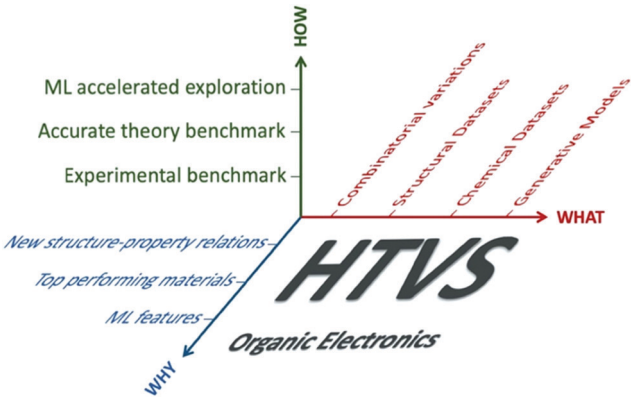


Fig. 1 The dimensions of HTVS in organic electronics explored individually in this review. How to perform and validate HTVS (Section 2), what datasets can be explored (Section 3) and what questions can be answered (Section 4).

Tahereh Nematiaram

Tahereh Nematiaram is a Postdoctoral Researcher at the group of Professor Alessandro Troisi. She holds a double-PhD in Nanophysics and Photonics from the University of Grenoble Alpes and the University of Tabriz. Tahereh is interested in developing quantum models to explain the fundamental physics dominating molecular materials and junctions. Exploiting such theoretical models, currently, her research is focused on discovering materials with targeted optical and electronic properties, determining the realistic physical limit to such properties, and deriving strategies for inverse materials design.

Alessandro Troisi

Alessandro Troisi is a Professor of Chemistry. He obtained his PhD in Bologna (Italy 2002), performed postdoctoral research at Northwestern University (US) and started in Warwick in 2005 as assistant professor, being promoted in 2010 to a full professorship. Since 2017 Prof. Troisi is Chair in the Materials Innovation Factory of the University of Liverpool. Professor Troisi specialises in theoretical and computational chemistry, computer aided materials discovery, electron transport and transfer, organic semiconductors and quantum dynamics. He was awarded the Marlow Medal by the Royal Society of Chemistry in 2007 and the ERC Starting Grant (2009), ERC Consolidator Grant (2014), ERC Advanced Grant (2021).

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 1 Non-exhaustive overview of the technologies in organic electronics requiring the development of novel molecular materials with a sample reference providing an example of research work focused on the identification of the promising molecules for that function

Technology area	Property or function	Sample ref.
Organic/hybrid photovoltaics	Electron acceptors	34
	Electron donors	35
	Singlet fission	36 and 37
	Up-conversion	38
	Hole-transporting materials	39
	Sensitisers	40 and 41
Displays	Light emitters (IR & vis) for OLEDs	42 and 43
	TADF	44 and 45
	Dual emission	46
Lasers/probes	Materials for gain medium	47 and 48
	Deep-tissue optical imaging	49
Transistors	High mobility materials	50
	Biomaterial devices	51
	Phototransistors	52
Detectors	X- and gamma-ray	53
	Artificial vision	54 and 55
	IR detection	56
Sensors	Gas sensors	57 and 58
	Biological sensing	59
Memory	Resistive memories	60 and 61
Energy storage	Nonaqueous redox flow batteries	62
	Organic electrodes	63 and 64
	Pseudo-/super-capacitors	65 and 66
Photocatalytic systems	Organic photocatalysts	72 and 73
	Dye-sensitised photocatalysis	74
Others	Conductive inks	67
	Spin-valves	68
	All-printed RFID tags	69
	Neuromorphic devices	70 and 71

For these reasons, instead of describing, one by one, the main contributions in the area, we organise this review following the questions that should be answered in planning and performing HTVS and, specifically, what properties can be computed and with what accuracy (Section 2), how the datasets to explore can be generated (Section 3) and how the results can be analysed (Section 4). In this work, we will try to highlight areas where a consensus on the methodology is being reached and where contrasting approaches have been proposed. This deconstruction of the field of HTVS in organic materials should be seen as complementary to other reviews of the field.[24–26] While the same questions are posed in search of inorganic materials, this review will not focus on them, and the reader is referred to other works.[27–33]

2. Computable properties: benchmarking and calibration

Although hardware is developing rapidly, the most accurate theoretical methods can be unfeasibly expensive, meaning that approximate methods are vital, especially for HTVS. This implies that benchmarking is often a necessary pre-step for HTVS, whether it is done by the researchers carrying out the screening or provided by other benchmarking studies in the literature. There have been multiple recent works[75–79] with the purpose of accelerating the future computation of a specific property by comparison, modification, and combination of existing theory.

Benchmarking is a way to find a cost-effective methodology to successfully carry out HTVS within the limitations of both project time and the hardware available. To properly benchmark, a small and diverse set of the larger input set can be evaluated by employing different protocols, *e.g.* for density functional theory (DFT),[80] this could include a wide range of exchange–correlation functionals and basis sets. If experimental data for the molecules used are available, each of the methodologies can be assessed in terms of their correlation and error to experiment to figure out the maximum accuracy that can be achieved. As it is likely that the best methods may still be too computationally expensive for HTVS, the most efficient method can be chosen judiciously according to specific project constraints. The method can also be subsequently calibrated by identifying a simple relation (often linear) between computed and experimental data which reduces the

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

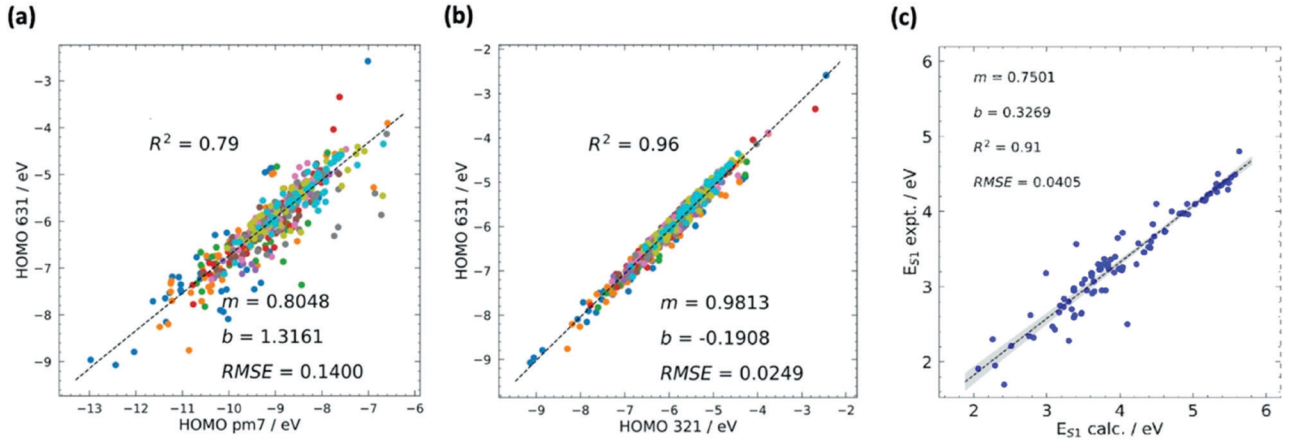


Fig. 2 Calibrations in HTVS has different forms. The most common is the calibration of the lower *versus* the higher level of theory. The comparison between semi-empirical and DFT theory is often quite satisfactory and the comparison of results with different basis sets is typically excellent. (a) Example of calibration of HOMO energy level computed using semi-empirical (PM7) and DFT (B3LYP/6-31G*) methods with squared correlation coefficient R^2, root mean squared error (RMSE), and the m and b values for the linear fit $y = mx + b$. (b) Calibration of HOMO energy level computed using DFT with different basis sets B3LYP/3-21G* and B3LYP/6-31G*. A combination of methods at different accuracies can be used sequentially in multilayer screenings. Comparison against experiment, often carried out with one of the best methods, highlights the systematic errors and the maximum accuracy achievable when this is corrected. (c) Comparison between calculated (M06-2X/def2-TZVP) and experimental values of first excited state energy E_{S1}. Adapted from ref. 37.

systematic error of the calculation (see also Fig. 2). This is only available for screening studies that seek to evaluate directly computable properties such as optical absorption for single molecules where good and homogeneous datasets[81,82] are available.

If there is no available experimental data, comparison with the most accurate (but time-consuming) methods, such as coupled cluster (CC) theory,[83] can be done in the same way. The ultimate goal is to find a method which evaluates the property of interest to a good degree of accuracy, whilst also being computationally feasible for HTVS. In their extensive assessment of functionals to accurately calculate the singlet excited energy using TD-DFT,[84] D. Jacquemin *et al.* suggest that although benchmarking to experiment is a logical philosophy to realise more accurate theoretical results, it is limited by both the uniformity, or lack thereof, of experiment between datasets, and the weakness of not being able to strongly mimic experimental conditions computationally, such as solvation effects. It is especially palpable for excited state computations, whereby only vertical excitations are normally considered; calculations of the more realistic adiabatic energies require further geometry optimisations on the excited state geometries, which is mostly limited to the higher-level refinement stages of a large-scale screening scenario. Benchmarking and calibration to higher levels of theory, *e.g.* state-of-the-art wavefunction methods, remove these conditions and grants uniformity to all reference datasets. In cases such as calibrating classical or semi-empirical methods to DFT, the size of reference datasets can also become substantially large, making the fits potentially quite predictive.[85–88] Although the disadvantage of using theory to calibrate theory is clear, its successful applications have allowed it to readily percolate through the field.

Experimentally accurate post-Hartree–Fock *ab initio* methods, *e.g.* CCSD(T) are often the gold standard for calculating physical properties but, due to their very unfavourable computational scaling with system size,[89] they cannot be used for HTVS of organic electronics. Despite the increased accuracy of intensive wavefunction methods, benchmarking of functionals and basis sets using (TD-)DFT can be enough to reduce the error in the computation of physical properties to acceptable levels for screening.[75,84,90–92] A more available hierarchy is the use of DFT to fit results calculated using semi-empirical methods. This is especially useful when polymeric structures are being considered,[93–97] as the system sizes are too large to evaluate with DFT directly, unless the expensive integrals can be parametrised.

To successfully benchmark and calibrate the output of a computational method, the property in question must be directly computable, *e.g.* ionisation potential (IP), electron affinity (EA), excitation energy *etc.* Properties which are only indirectly computable, *e.g.* SF yield or power conversion efficiency (PCE) and are not based on the outcome of a single calculation, cannot be treated in the way described above. The validation of directly computable properties is discussed in the next two sections.

a. Frontier orbital energy and redox potential

A common focus of HTVS is on the evaluation of ground state properties such as the frontier orbital energies. The HOMO and LUMO energies are especially useful in calculating IP and EA which are often defined in vacuum. For many organic electronics applications (organic semiconductors (OSC), hole transporting materials, *etc.*), the energy level alignment of these values across the device is an essential prerequisite for their function. A range of experimental data have been developed over the years to validate the computation of IP and EA, with inclusion of larger, more relevant molecules for organic electronics taking place more recently. In terms of dedicated

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

databases, the G2-97 test set consists of 88 experimental IPs and 58 EAs of small molecules,[98] and is a collection of measurements with a reported uncertainty of less than 0.05 eV. The NIST Chemistry Webbook has thousands of IP and EA entries for small to medium sized molecules collated from literature sources.[99] The method of measurement and uncertainty (which can be as high as 1 eV) is reported for each entry. In general, photoemission spectroscopy can provide the most reliable results for these properties.[100]

The computation of these quantities has been long assessed and improved with the various quantum chemical methods. For example, evaluating IP and EA using Koopmans' theorem[101] is crude if Kohn–Sham (KS) DFT is employed and can yield large errors, especially for EA.[92] However, the error can be significantly alleviated if long-range corrections are introduced due to better treatment of fractionally occupied orbitals.[102–108] Further investigations[109–118] into the expensive GW[119–122] approach shows improvement against KS DFT, however there is found to be significant improvement when long-range corrected hybrid functionals are used in tandem with GW corrections.[108,123,124] Electron propagator methods[125–129] are an additional way of finding IPs and EAs with good accuracy to highly-correlated wavefunction methods,[130,131] while remaining, however, still too computational expensive for HTVS. This then leaves the various flavours of DFT, or more approximate methods, as the level of approximation used for the initial stage(s) of virtual screening for suitable orbital energy levels.

In order to use DFT for HTVS seeking to calculate IPs and EAs, it should be benchmarked for optimal accuracy. In ref. 92, a diverse set of 11 DFT functionals is assessed by validating against experimental IP and EA values in vacuum. They showed that the HOMO energy can be evaluated relatively accurately *versus* experiment for 27 common small molecules. Despite a systematic upward shift in the calculated energy, linear calibration with the experiment can yield predictive HOMO values, especially when functionals including exact exchange are used, such as KMLYP[132] with R^2 = 0.99 and an average error of 0.73 eV. This is also seen in a larger-scale study by Y. Fu *et al.* where 270 experimental IPs are compiled from the literature to calibrate the B3LYP[133,134]/6-311++G(2df,2p)//B3LYP/6-31+G(d) protocol; a strong linear dependence is found with R^2 = 0.99 and a systematic error of 0.28 eV,[135] indicating the existence of robust measurements for this property. The LUMO eigenvalue calculated in ref. 92 is, unsurprisingly, found to be extremely inaccurate in comparison to experimental values ($R^2 \approx 0$). As the calculation and experimental measurement of the gap (R^2 = 0.91–0.96) and HOMO energy (R^2 = 0.94–0.99) are shown to be more robust, it was proposed that the LUMO accuracy can be increased by subtracting the HOMO energy eigenvalue from the gap energy when using these types of methods.[92]

There are several screening studies,[63,136] including large-scale,[137] which evaluate IP and EA using DFT. For example, M. Korth studied 23 000 known small molecules for IP and EA.[138] Their pre-screening work used the Koopmans approach to assess the accuracy of different quantum chemical methods on IP and EA, with benchmarking against the G21EA and G21IP subsets of the GMTKN24 database.[139,140] Out of their tested DFT methods, PBE[141]/TZVP[142] had the lowest error with mean absolute deviation (MAD) both for IP (MAD = 4.64 kcal mol^{-1}) and for EA (MAD = 7.00 kcal mol^{-1}).

It is, however, important for the discovery of new organic electronics that molecules of relevant sizes are also included within the benchmarking. In a multipart study,[108,123,131,143] the vertical IPs and EAs of acenes, quinones, nitro/nitriles, anhydrides and other electron acceptor structures with experimental measurements from the NIST were evaluated with the best theoretical methods, *i.e.* within ±0.03 eV accuracy of CCSD(T) with complete basis set limit (CBS) extrapolation. The latter, regarded as the gold standard method and unfeasible for larger molecules, is expected to be near the Born–Oppenheimer *ab initio* limit, and major errors *versus* experiment ($\geq$0.1 eV) are thought to be due to nuclear relaxation and environmental effects. This set of 24 known acceptors then forms a homogeneous, high-level reference for calibration of more approximate methods.

An additional step to increase the accuracy of IP and EA calculations, *versus* simply taking the negative of the respective orbital eigenvalue, is to calculate the difference between the total electronic energy of the neutral state and the charged state of the molecule. Furthermore, one can consider computing the adiabatic rather than the vertical transition energy, *i.e.* allowing for the nuclear relaxation of the charge state at the cost of an additional optimisation. It was proposed that this latter effect could be neglected in screening,[138] as it produces larger errors only in few cases.[144] The Electrolyte Genome Project by X. Qu *et al.*[145] included nuclear relaxation effects in their automated and benchmarked high-throughput DFT screening. They sought to ensure the high fidelity of their results by evaluating adiabatic IPs and EAs with the inclusion of the solvation model IEF-PCM.[146] Extensive protocols such as this can be reasonable with smaller scale and lower-level screening efforts.

Additional challenges with respect to the calculation of IP and EA are provided by the accurate evaluation of redox potentials which is essential in, for example, developing new materials for use in organic redox flow batteries and other devices.[137,147] Solvation energy plays a major role in determining the redox potential and must be carefully assessed[147] (this also means that insoluble compounds cannot be measured). Furthermore, some organic compounds suffer irreversible coupling of their monomers and can create additional difficulty in providing reliable values.[148]

An effort by H. Neugebauer *et al.*[149] assessed the performance of the previously developed semi-empirical theory (GFN-xTB),[150,151] PMx,[152,153] and DFT for the calculation of redox potentials for small to medium sized organic structures. Their test set consisted of 193 organic molecules with experimentally determined redox potentials,[154,155] allowing for a relatively robust evaluation of each method. They showed that DFT is bottlenecked by inadequate treatment of solvation effects (MAD = 0.22 V/R^2 = 0.97 for their best method: PWPB95-D4[139]/def2-QZVPP[156]). This finding is supported by a protocol comparison study by M. Isegawa *et al.*, which showed a similar error in the redox potentials for even the highly-correlated

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

CCSD(T) method.[157] Their tight-binding based method (GFN2-xTB) was comparable with DFT (MAD = 0.30 V/R^2 = 0.94), whereas the PMx class of semi-empirical methods were poorer approximations (MAD ≈ 0.60 V/R^2 ≈ 0.88). Ultimately, their novel tight-binding methodologies were approximately 2–3 orders of magnitude faster than the DFT counterparts, and around one order of magnitude faster than the PMx protocols. What is then left for further improvements in the evaluation of redox potentials is related to the improvement of solvation models, and work has been done to validate both explicit solvation,[158] and the effect of considering a diverse range of compounds, rather than picking a homologous set which is known to work well with an empirically parametrised model.[159]

b. Excited state energies

To be able to study interesting photophysical phenomena such as fluorescence,[160,161] phosphorescence,[162,163] SF,[164–167] TADF,[168–171] *etc.*, one must consider the electronically excited states of a structure. There are a range of experimental datasets available to validate a theoretical evaluation of these energies, particularly S_1 and T_1. This is largely due to the ease of the experiments, *e.g.* UV-vis, absorption and luminescence. Most molecules maintain a relatively constant electronic structure when the solvent differs between measurements, unless some of the relevant states have a strong charge transfer character,[172] the transition energies are only moderately affected by solvent effects, making it easier to construct large, reliable experimental datasets to be used for benchmarking and calibration. It should be noted, however, that experimental energies of low triplet states may require less straightforward measurements at low temperature or in polymeric matrices[173] making the availability of data more limited in certain spectral regions.

The extensive TD-DFT benchmark in ref. 84 constructed an experimental training set of 483 molecules from literature sources.[174–182] Using this dataset, they assessed the qualities of 29 different functionals on predicting excited state transition energies. It is found that pure functionals systematically underestimated the energies with mean absolute error (MAE) of approximately 0.35–0.4 eV whereas mixing in some exact exchange, optimally 22% to 25%, reduced the MAE to around 0.26 eV and 0.23 eV, as with the TPSSh[183] and B3LYP[133] functionals respectively. Moving from pure functionals to those with fractions of exact exchange also increased the correlation between the theoretical and experimental values, *i.e.* 0.94 to 0.96, which allowed for a more confident calibration, and indeed, combination of the best results through multiple linear regression[184] pushed the R^2 correlation as high as 0.98.

The Handbook of Photochemistry contains the S_1 and T_1 energies of approximately 500 organic molecules;[81] one such study[37] which used this data to calibrate a previously benchmarked protocol[185] is an HTVS done by our own group to find SF active molecules existing within the Cambridge Structural Database (CSD).[186] The protocol benchmarking study in ref. 185 assessed the quality of local hybrid functionals for the evaluation of S_1, T_1 and T_2 energies, with particular emphasis on the T_1 state due to its potentially poor description by TD-DFT.[187] They used a small test set of 11 SF candidates previously calculated with CASPT2,[188] however updated this reference level to CC2/CBS for better accuracy uniformity since CASPT2 can be sensitive to the selection of active space orbitals.[189] It was found with the best protocol which combines TD-DFT for, S_1 and T_2 and ΔSCF for T_1 (Lh12ct-SsifPW92/def2-QZVPD//BLYP35/def2-TZVP) that the MAE for the vertical S_1 and T_2 energies were around 0.10 eV, and just over 0.20 eV for the T_1 energy. In the following screening effort in ref. 37, the more widely available and similarly accurate M06-2X functional was adopted with the basis set lowered to def2-TZVP for all stages. Using 100 S_1 and T_1 energies of small to medium organic molecules from the reliable experimental dataset,[81] the root-mean-square deviation (RMSD) for the calculated S_1 energy with the experiment was 0.0405 eV with R^2 = 0.91, and RMSD = 0.0537 eV with R^2 = 0.88 for the T_1 energy. This accuracy allowed for the final calibration of results to identify molecules with the desired S_1–T_1 energy difference.

There are other experimental datasets available which are popular for machine learning (ML) techniques but can be used for benchmarking and basic calibration using many datapoints, for example the collection by J. F. Joung *et al.*[190] includes the optical properties for over 7000 unique chromophores, ranging from small molecules to relevant molecules such as pyrene,[191] coumarin,[192] azobenzene[193] *etc.* However, many screening studies opt to use smaller sets of specialised molecules due to the availability of experimental data. This is usually done by searching through literature that includes experimental measurement, either by automatic or manual extraction, to construct a unique training set and benchmark reference for the screening. R. Gómez-Bombarelli *et al.*[42] pieced together a training set of 46 molecules from other works[194–210] to calibrate the results of their TD-DFT calculations in an attempt to find TADF active molecules, which are often large and complex in comparison to molecules existing in experimental datasets. Another study which followed this method is by N. M. O'Boyle *et al.*,[96] where the primary excitation energy of 60 literature-based oligomeric compounds, taken from their previous benchmark study,[97] was compared with experiment, yielding RMSD = 0.28 eV, and R^2 = 0.84 when using the PM6/ZINDO[211] method.

As an alternative to calibration against experimentally determined excited state energies, it is possible to compare with high-reference excited state calculations. If computed data for excited state properties are to be considered for calibration, there are very high-level reference datasets available.[212–217] For example, the excited states, including triplets, of 18 small molecules, *e.g.* water, acetylene *etc.* were evaluated with up to the CCSDTQP level of theory, and almost reached full configuration interaction standard.[218] With the expansion of the molecules considered, now encompassing 27 medium sized molecules such as benzene, thiophene, triazine *etc.*, ref. 219 provides data up to the CCSDTQ level of theory for molecules with 4 heavy atoms, and CCSDT for those with 5 and 6 heavy atoms. It is believed that, for the theoretical geometries, almost 95% of their reported transition energies are chemically accurate

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

to 1 kcal mol^{-1}, and the errors can be attributed to geometry differences and inconsistencies in measuring vertical transitions experimentally.

There are a number of very large datasets which contain theoretical reference data at more feasible levels, *e.g.* TD-DFT and are nominally used for ML-based studies. For example, VERDE Materials DB[220] hosts 1500 organic electronic relevant structures with computed excited state energies at the M06/6-31+G(d,p)[221–223] level. The QM7b[224,225] database provides information on over 7000 structures based on the 7 heavy-atom subset of the GDB-13.[224] The QM8[226,227] includes 20 000 synthetically available small molecules, and provides electronic structure values using both TD-DFT and CC2. The QM-symex[228] has 173 000 compounds with excited state energies calculated at the B3LYP/6-31G level, and provides particular emphasis on molecular symmetry. Such databases could be used for the calibration of lower-level methods, *e.g.* semi-empirical; this can be especially useful for the preliminary stages of HTVS projects where molecules are to be filtered prior to secondary stages. Works based on high-throughput screening of the excited state energies of polymeric structures could benefit most from this philosophy as the main bulk of the calculations will necessarily need to use low-level theory such as semi-empirical methods, with the references computed using TD-DFT, for example.

In contrast to time-dependent methods, the optical gap can be estimated using the frontier orbital energies. Testing on a set of benzofulvene derivatives, S. Tortorella *et al.* showed that the optical gap can be predicted reasonably well, in terms of correlation, by taking the HOMO–LUMO difference when using the ZINDO semi-empirical method on AM 1 optimised geometries (MAD = 3.01 eV/R^2 = 0.83).[91] Naturally, the higher level DFT methods such as B3LYP produced better results when compared with experiment (MAD = 0.26 eV/R^2 = 0.87) especially in terms of the systematic error, though interestingly the same correlation could be achieved if the B3LYP geometry was used with a ZINDO electronic structure calculation (MAD = 2.74 eV/R^2 = 0.87). This finding implies that, if the correlation is good, low-level theories such as semi-empirical methods can be used for rapid pre-screening of even millions of molecules, prior to higher level refinement, which is in line with the typical computational funnel hierarchy of HTVS. However, there should be careful consideration when these very approximate methods are used as they are often parametrised to reproduce experimental values for specific structural motifs, and can fail to generalise.[229–231]

The use of semi-empirical methods, such as ZINDO, AMx,[232–235] PMx and so on[236] to compute optoelectronic properties, *e.g.* band gap, of conjugated polymers has been a focus of screening studies instigated by G. R. Hutchison *et al.* with more recent investigations into soft modelling approaches using genetic algorithms.[95–97] In a similar vein, M. A. Zwijnenburg and co-workers used a semi-empirical high-throughput screening approach to calculate IP, EA, and the optical gap for conjugated polymers.[93] They used the xTB methods developed by S. Grimme *et al.*[150] and linearly calibrated with (TD-)DFT using a set of 40 copolymers; this, to an extent, allowed for (TD-)DFT accuracy for unknown polymeric structures computed with the xTB methods. They followed the same calibration approach to search for novel, diketopyrrolopyrrole based dyes for use in OPVs.[74]

Although there are datasets of organic molecular properties,[81] the field still lacks the availability of a large set of data for crystalline properties except for reviews of selected topics[237,238] from which it is possible to collate information on the optical spectra and luminescence of about ~100 molecular crystals. This set is, however, sufficient to provide validation for the computation of the excitonic coupling[239] which is one of the key elements needed for the study of solid-state optical properties. The other key element is the evaluation of the local exciton phonon coupling[240,241] which is a property of the isolated molecule and is computable with a good accuracy.[242]

A general observation that can be made after the overview of the past two sections is that the calibration against experiments and higher-level theory for both excited and ground state properties is generally less systematic in works focused on organic electronic applications, *i.e. ad hoc* calibrations are often proposed to maximise the accuracy of the predictions within a given technological context. On the contrary, contributions directed toward the community of quantum chemistry have developed more standardised datasets for testing (albeit generally based on smaller molecules). To better track future progress and strike the right balance between accuracy and computational cost it would be ideal to develop common datasets that sample a broad range of compounds used across organic electronics.

c. Charge mobility

Charge carrier mobility is one of the main figures of merit used to determine the suitability of organic semiconducting materials for technological applications. Molecular semiconductors have been considered traditionally more approachable for benchmarking and HTVS studies as, unlike polymers or amorphous materials,[16,243,244] there are no unknown structural or morphological features to be determined. Charge mobility measurements of molecular semiconductors that are robust and reproducible in different research labs have started to appear relatively recently.[245] The measured mobilities, however, depend on the purity of crystals,[246] and the degree of polycrystallinity.[247] Therefore, the intrinsic (defect/trap free) mobility of single crystals is the only reference experimental value to study the relation between crystal structure and mobility. Particularly valuable are therefore those reports of mobility with intrinsic nature of transport verified either by comparison with Hall mobility,[248] or ''band-like'' temperature dependence (*i.e.* mobility decreasing with increasing temperature).[249] Considering these conditions, there are approximately 20 ''reference'' measurements of intrinsic mobilities in thin-film transistors, which provide a limited, but reasonable, set of data to validate the theory.[250]

The success in intrinsic mobility measurements has led to an important advance of technologies based on organic thin-film transistors,[250] but it has also challenged the theory

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

of molecular semiconductors that was thought to be well-understood. It is evident that the measured charge mobilities of high purity crystals (>1 cm^2 V^{-1} s^{-1}) are too high to be rationalised with a simple hopping transport mechanism yet too low to be accurately consistent with the band transport.[251,252] This is more clearly discussed in ref. 253 where through evaluating mobility of $\sim$60 organic crystals using both theories, it is shown that, at room temperature, neither of these transport theories can accurately predict the charge carrier mobility. In molecular semiconductors, the Hamiltonian parameters, including charge transfer integrals, vibrational energies, reorganisation energy, dynamic disorder, and thermal energy at room temperature, differ by not more than an order of magnitude, which makes the evaluation of charge transport a challenging task.[18] Accordingly, over the years, a number of advanced theoretical methods such as small polaron theory,[254,255] mean-field Ehrenfest model,[256,257] trajectory surface hopping method,[258,259] open quantum systems,[260,261] quantum Monte Carlo[262,263] and transient localisation theory (TLT)[264–266] have been developed to evaluate charge transport in this materials class. Most of these methods, despite being promising in predicting mobility in agreement with experimental measurements,[267,268] are slow and, hence, have never been employed in HTVS studies.

Hopping transport theory, despite being not fully applicable to molecular semiconductors, due to its simplicity and low computational cost, remains a frequent method of choice in the majority of HTVS studies, including those using large databases,[269–271] as well as those evaluating smaller libraries of structures.[272–275] Applying this theory to screen the CSD, as the world's largest repository of small organic/metal–organic molecules whose crystalline structures are experimentally known,[186] C. Schober *et al.* made an important contribution demonstrating, for the first time, that large scale screening of materials for transport applications is possible.[269] The transfer integral and the reorganisation energy are the main components of the hopping transport theory. The computation of these two quantities is not very demanding, and their validation against accurate theoretical methods and experimental data is straightforward. There are a number of different methods that can be employed to compute charge transfer integrals.[276] Computed band structures can be validated against more accurate computational methods, *e.g.* larger basis sets in DFT-based methods, as well as experimental data such as those obtained from angle-resolved photoemission experiments (ARPES).[277–279] The calculations of the reorganisation energies rely on the high vibrational frequencies where the routine DFT methods provide reliable results. The results of these studies have shown remarkable agreement with experimental data extracted from infra-red and Raman spectroscopy.[280–282] In some of these hopping theory-based virtual screening studies, experimental verification of the mobility of the targeted molecules is also provided, which shows only a qualitative agreement with theoretical results.[283]

All the advanced theories developed for accurate calculation of mobility of molecular semiconductors, as described above, require the nonlocal electron–phonon couplings elements.[284,285] However, the calculation of this parameter is known to be computationally very demanding. Recent calculations of nonlocal electron–phonon couplings on one[268] or few[286,287] molecules required millions of CPU hours[288] for state-of-the-art methods (*i.e.* not suitable for HTVS) while computationally inexpensive empirical force fields yield inaccurate results as they are not parameterised to reproduce low-frequency phonons.[289] From the experimental side, Raman spectroscopy[280] and terahertz time-domain spectroscopy[290] are the most common methods used to derive information on low-energy phonons. However, these methods provide only gamma phonon energy and, therefore, can only partially verify the computed phonon spectra.[286] High-resolution inelastic neutron scattering measurement is shown to be a more useful method as it allows retrieving information on the low-energy phonons and recently has enabled the validation of low-frequency phonon calculations in the context of evaluating charge mobility of molecular crystals.[288]

The calculations of nonlocal electron–phonon couplings can be substantially accelerated using simplified methods. For example, our group suggested using approximate phonons, assuming that each molecule oscillates independently from the others, to significantly speed up the calculations and make them feasible for HTVS studies. Relying on this strategy and applying the transient localisation theory, which is among the advanced theories that consider all transport parameters on the same footing, the CSD was screened for high mobility materials.[50] The absolute values of calculated mobilities in the framework of TLT are within $\sim$35% of experimental data, and the relative values are well reproduced in families of homogenous compounds.[265,288] Furthermore, as shown in a recent study, this theory is also able to reproduce electromechanical responses.[291] The fragment-orbital based surface hopping method, relying on explicit time propagation of the electron-nuclear dynamics, is another viable approach that is used to capture the impact of thermal fluctuations on charge transport.[259] A recent study, through applying this theory to a set of eight crystalline structures, shows that the excess charge carrier leads to a polaron which is delocalised over 10–20 molecules in highly conductive crystals. Charge mobilities extracted in this study are in remarkable agreement with experimental measurements and correlate strongly with the data obtained using TLT. The comparison between experimental and computed charge mobilities exemplifies a common situation in organic electronics where benchmark experimental data are limited for some properties and the modelling tools are not yet fully optimised. In these cases, the great demand for high performing materials is driving both standardisation of experimental methods[245] and refinement of the theories.[18]

d. Interfacial properties

Many of the functions of organic electronics materials are acquired at the interface with other materials[292] and most organic electronic devices are fabricated as multiple layers of distinct materials[293] or are based on multi-phase components with complex morphologies and often unknown composition.[294]

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

These aspects are at the forefront of current research and, since consensus is yet to be reached on the underlying physical mechanisms, they are not directly tackled by HTVS. The virtual screening in these cases often focuses on easily computable properties that capture only some of the physics. Examples are the search for alternative electron acceptor molecules for OPV that have similar electronic and solubility characteristics to the best performing ones,[295] or the development of statistical models only considering energy levels of the constituent molecules.[34]

While it is and will remain impractical for many years to model many interfaces between soft organic materials with state-of-the-art methods,[296–298] several authors have considered approximate models of the interfaces that allow their exploration in larger numbers. For example, Y Imamura *et al.*[299] presented an approximated model to study the interfacial geometry of 1850 donor–acceptor pairs looking for arrangements more favourable to charge generations. Alternatively, one can build fewer models able to capture more physical details, like the proposal by C. Poelking *et al.* to model the electrostatics of bilayer cells which explain, very well, a range of measurements and could be included into future screening protocols.[300] A type of interface that is amenable to HTVS is that between crystalline phases and chemisorbed molecules found, for example, in dye sensitised solar cell. In these cases it is possible to separate the role of the anchoring group from that of the rest of the dye[301] and screen separately for the best anchoring group[302] and the best dye.[40,74]

e. Crystal structure prediction

For a range of organic electronics applications, the function is determined by molecular arrangement in highly ordered or crystalline domains, *i.e.* the properties are a combination of molecular properties and intermolecular interactions. Very substantial progress has been made in the field of molecular crystal structure prediction (CSP), *i.e.* the prediction of the molecular arrangement in the crystal starting from the chemical topology of the constituent molecules. This research area, more commonly employed in drug discovery and fundamental research, is reviewed often[303–305] and we focus here on its recent ramifications in the field of organic electronics.

At the moment, it is not possible to perform HTVS where the starting point is a large set of molecules, and the output is their predicted crystal structure and the property of the material. With progress in sampling methods, adopting efficient, accurate atom–atom force fields, and employing parallel, high-performance computing, the situation seems to be changing.[303,306] For instance, a library of 27 structural isomers of pyrido[2,3-*b*]pyrido-[3′,2′:4,5]pyrrolo[3,2-*g*]indole is screened to assess charge mobility in their predicted crystal structures.[307] This study yields two molecules with desirable charge mobility, which are also attractive as synthetic targets. A follow-up study from the same research group, using an evolutionary method, has screened a larger user-specified region of chemical space, containing aza-substituted pentacenes, to identify high-mobility materials.[308] The first step of this analysis realises a set of promising molecules which are eventually evaluated using crystal structure prediction. Results reveal two promising structural motifs: aza-substituted naphtho[1,2-*a*]anthracenes having reorganisation energies comparable to that of pentacene and a series of pyridazine-based molecules possessing both low reorganisation energies and high electron affinities. Similar methods have been applied to classify molecules capable of forming weakly bound polyaromatic hydrocarbon co-crystals.[309] As such, a two-step approach is developed that first utilises all the known molecular combinations forming this class of co-crystals, extracted from the CSD (1722 molecular combinations), to train the model. Then, in the second step, it ranks possible, but unknown, pairs from the ZINC15 database[310] (21 736 possible molecular combinations). The applicability of the employed methodology and discoveries are verified through the experimental realisation of two co-crystals named pyrene-6*H*-benzo[*c*]chromen-6-one and pyrene-9,10-dicyanoanthracene both comprising molecules never considered as co-crystallising in the CSD.

f. Conformational search

Conformational search is an important aspect of computational chemistry established in the field of drug discovery. Many drug molecules possess a large number of rotatable bonds,[311,312] and it is often not enough to consider only a single conformational isomer. Although extending this idea into the field of organic optoelectronics may appear to be an unnecessary step since the chemical space is made up of rigid, conjugated structures and the free rotation of flexible side chains tend to have little effect on the electronic structure of an isolated molecule,[313,314] it has been found that its impact can be far from negligible.[315] It is especially true for functionalised organic electronic materials which contain rigid oligomers connected formally by a single carbon–carbon bond. This architecture allows for large flexibility on the dihedral angle between the monomers which, often modulated by steric effects,[316,317] can generate a range of conformers with different properties.[318] Consideration of a suitable conformational search step prior to HTVS is now considered necessary to ensure optimal accuracy. Works on polymeric structures[93,95] are especially reliant on a good conformational search strategy to achieve higher accuracies. A common method between these studies is to generate a (large) sample of conformers using molecular mechanics (MM) methods at a low cost to find the lowest energy conformer. This conformer is then selected and subject to higher-level computations, rather than a randomly generated or distorted (higher energy) orientation.

Recent considerations by G. R. Hutchison *et al.* showed that finding the lowest energy conformer with MM methods may offer a poor starting geometry when compared to more accurate, quantum chemical methods.[319] In other words, a consequence of using the lowest energy MM conformer prior to screening is that a high-energy geometry will persist with subsequent optimisations and will not reflect the geometry found if quantum chemical methods are used within the conformational search step. The difficulty in incorporating conformational effects in HTVS is particularly severe while modelling charge transport properties in polymers. There is a

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

consensus on the dependence of transport on the local ordering of the polymers[320] and a range of accurate works demonstrating how the local structure requires extensive molecular dynamics simulations,[321] which are inconsistent with high-throughput screening. There is however some preliminary evidence that the conformational study of isolated chain coupled with electronic structure calculations can still provide sufficient guidance for the design of new materials.[322]

g. Acceleration of HTVS by machine learning

Machine Learning (ML) has proved to be a valuable tool to produce results with an accuracy similar to quantum chemistry methods at a fraction of their cost, given that one has enough data to train the model. For an in-depth review of this field, we refer the reader to ref. 323 and 324. Here, we will focus on how the recent developments of ML impact the computation of properties of interest for organic electronics.

M. Rupp *et al.* developed an ML model that calculates the optimisation energies from the nuclear charges and atomic positions of approximately 7000 organic molecules,[325] and obtained an MAE of $\sim$10 kcal mol^{-1}, similar to mean-field electronic structure theory at only a fraction of the computational cost. ML has also been used to correct the PM7 HOMO/LUMO eigenvalues and molecular polarisabilities (as well as other thermochemical properties less relevant for OPV applications) to an accuracy similar to DFT,[326] where the ML correction represents only a fraction of the computational time of the semi-empirical calculation. This approach opens new possibilities and as data availability of quantum chemical calculations increases, the use of ML to improve and bypass new calculations will become more prevalent. ML has also been used to bypass DFT calculations,[327] saddle point searches[328] and force calculations during molecular dynamics.[329] More recently, ML has been used to correct DFT atomisation energies to an accuracy of higher composite methods (G4MP2) with an accuracy of 0.005 eV for approximately 100 000 small molecules, and 0.012 eV for larger molecules with 10–14 heavy atoms.[330] In a relevant application for organic electronics it has been shown how neural networks, trained with 200 000 molecules from the Harvard Clean Energy Project Database,[331] can reproduce HOMO and LUMO results of quantum chemical calculations with a large level of accuracy.[332] F. Jabeen *et al.* used a multilinear regression analysis to predict the refractive index of a set of approximately 100 polymers.[333] ML has also been used to predict the emission wavelength of multiple fluorescent organic molecules, using steric, hydrophobic and electronic properties.[334,335] Other examples of the prediction of excited state properties through machine learning and a more general view on its applicability can be found in ref. 336 and 337. Alternatively, approaches have been proposed in which ML is used to approximate the wavefunction itself,[338,339] and all properties of interest are extracted from it (see the difference with the more common approach of predicting each property separately in Fig. 3). An example of this approach was presented in ref. 340, where ML is used to approximate the atomic charges, instead of calculating the self-consistent charges during density functional tight binding calculations, which are then used to approximate the ground and excited state potential energy surfaces of multiple organic molecules.

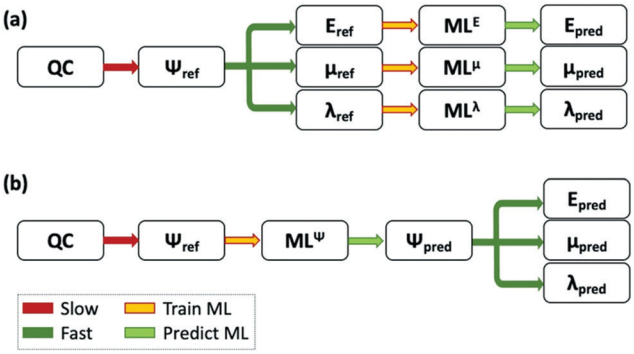


Fig. 3 Scheme of the use of ML to complement QC calculations. (a) A ML model is trained to predict each feature. (b) The ML model is trained to predict the wavefunction of the system, and the model can then be used to predict multiple properties. Adapted from ref. 339, under a Creative Commons Attribution 4.0 International (CC BY) license.

These results show how ML can correct low-level calculations and approximate more expensive quantum chemical calculations for a variety of cases, and such models can be accurately trained with the order of thousands of training samples. However, one should also consider the lack of extrapolation capabilities of ML discussed for example in a recent work by S. Kauwe *et al.*[341] The authors considered a dataset extracted from the Automatic Flow for Materials Discovery (AFLOW) database[342] with a large variety of compounds with evenly distributed DFT properties, and they could consistently identify most of the materials with the 1% highest performing bulk modulus, thermal expansion and other properties, when training the model with the other 99%. Despite the encouraging results, the authors themselves indicate some limitations that can make this identification of highest performing materials not possible, *e.g.* when the high performance of the new material is due to a new physical mechanism not present in the training set, when datasets have imbalanced classes, lack heterogeneity or lack properties of interest, or when rare events cause drastic property changes. To address these problems, new metrics have been proposed to identify molecules outside the models' domain of applicability,[343] increasing the confidence of ML-guided exploration of the chemical space. Explicit quantum chemistry calculations can always be used to correct the erroneous predictions of materials with exceptional properties (false positives), which are typically in very small numbers. The greatest risk of ML methods bypassing direct calculations are the false negatives, *i.e.* exceptional materials not identified as such, which would be missed by the virtual screening.

3. Methods to sample the chemical space

Once the computational methodology for a target property has been determined and validated, the definition of the chemical

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

space to be explored is the next major challenge of HTVS. In an HTVS, the generation of 3D structural information is often not a trivial task. The very first choice is whether the search for a novel property should be done among pre-existing compounds which are inherently synthetically available, or completely original chemical motifs. If pre-existing molecules are to be ignored, the chemical space of organic molecules, which is essentially impossible to sample fully ($\sim 10^{60}$),[344,345] is the vast sea which must be navigated. However, randomly generating molecules which simply follow the rules of organic chemistry is not an efficient way to search for niche phenomena, as this will quickly produce a large list of uninteresting structures and lends no consideration to the ease of synthesis. Synthetic accessibility of hypothetical compounds is arguably the most important aspect of *de novo* library design. If, for example, theory predicts a novel structure to have a higher PCE than any other known OPV, it is not useful if it cannot be synthesised; the only utility for such a case is the potential for accessible derivatives, or the development of new design rules. An alternative method is to use chemical motifs already known to exhibit the desired property as the chromophore bases for the input set.

A variety of approaches have been proposed that differ fundamentally in terms of dimensionality of the space explored, similarity of the molecules with respect to known examples and importance given on the chemical accessibility of the proposed candidates. Each approach is defined and exemplified in the following subsections.

a. Combinatorial modifications

There are several ways to modify, enumerate, substitute, and combine chemical motifs in the pre-screening stages, and there have been multiple efforts developing increasingly intelligent algorithms to do this automatically, *e.g.* genetic algorithms can use physical and electronic data of chemical motifs to find useful combinations based on the screening criteria;[95] they can also be used to find more accurate low-energy conformations for flexible molecules.[346]

Introduction of such algorithms is not always necessary to generate a rich input library. More simplistic approaches where fixed moieties are functionalised at random, or generic chemical formulae are filled in with relevant units have found purchase in the field of HTVS with reasonable success.[74,347–352] The Harvard Clean Energy Project,[331] for example, used 26 common building blocks, *e.g.* thiophene-, furan-, triazine- and benzene-based compounds with predefined connecting points to construct a set of 10 million unique molecules, including oligomeric sequences with up to five units. M. Moral *et al.*,[353] on the other hand, took a more restricted approach and investigated 100 1,4-bis(phenylethynyl)-benzene derivatives with the formula Y–C≡C–X–C≡C–Y. This rigid architecture was used with consideration of the axial rod-like feature, which is prevalent in molecular scale electronics.[354–356] With judicious choice of aromatic motifs in the X and Y position, the screening found structures which independently satisfied ambipolar charge transport characteristics, electron and hole transport, Ohmic contact with common electrodes and so on. Using a more exhaustive strategy like that in ref. 331, N. N. Matsuzawa *et al.* constructed a very large set of rigid acenes with two to eight fused rings and substituted with carbon, oxygen, sulphur and selenium.[270] Since random permutation methods can quickly generate an unfeasible number of molecules, in this case over 7 M structures, a selection of a smaller subset is often necessary for the screening. The selection can be random, guided by some loose logic or by synthetic score. There are multiple ways to score the complexity or synthetic accessibility of molecules, these can range from expert opinion to neural networks, chemical similarity indices and structural descriptors based on graph theory.[357–359] Y. Wen *et al.* considered synthetic accessibility in their DFT and ML combined approach to screen approximately 10 000 novel dyes for DSSC applications.[40] For 500 promising hit candidates, they provide the synthetic accessibility score by Ertl and Schuffenhauer.[360] They pointed out that some of the candidates with high predicted PCE are also predicted to have low synthetic feasibility. This work highlights the importance of considering synthetic feasibility; something that is often overlooked when *de novo* molecular generation is adopted for HTVS.

Creating a collection of novel molecules can be done in a way which better facilitates the ease of synthesis by using stricter design rules or combining the existing moieties in a chemically logical way. There are screening efforts which follow this more intuitive approach to library generation; for example, K. B. Ørnsø *et al.* studied 1029 original, functionalised zinc porphyrin rings for DSSC applications and are generated by substitution of four peripheral positions (Fig. 4a).[361] Their contained set of seven electron donors and three electron acceptors was driven by well-established knowledge of their properties and synthetic availability as they have been investigated experimentally or theoretically prior. A smaller-scale study by J. T. Blaskovits *et al.* looked at finding new SF candidates by pairing known donor and acceptor cores to create 81 novel copolymers.[362] The constituent moieties were chosen to ensure short synthesis with high atom economy.

All of the efforts mentioned above used in-house techniques to modify existing motifs to generate an input set prior to computation of properties, and this may be the most reasonable approach when searching for a certain property, *e.g.* high PCE. It is also possible to generalise this process and consider all chemically plausible molecules that can be created as done in the GDB-13[224] and GDB-17[226] sets. These databases exhaust the chemical space of possible organic molecules with up to 13 and 17 heavy atoms respectively. Naturally, these projects have no prior consideration of synthetic accessibility but can be used to lead investigations into completely new design rules. R. Ramakrishnan *et al.* probed into this diverse, computer-generated chemical universe and provided the ground-state properties of a subset 134 000 structures from the GDB-17 using benchmarked DFT methods.[363] There are other comprehensive works which also consider very large subsets, some with the inclusion of isomerism effects, and are invaluable for ML-based studies. For example, the ANI-1 database reports 20 million

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

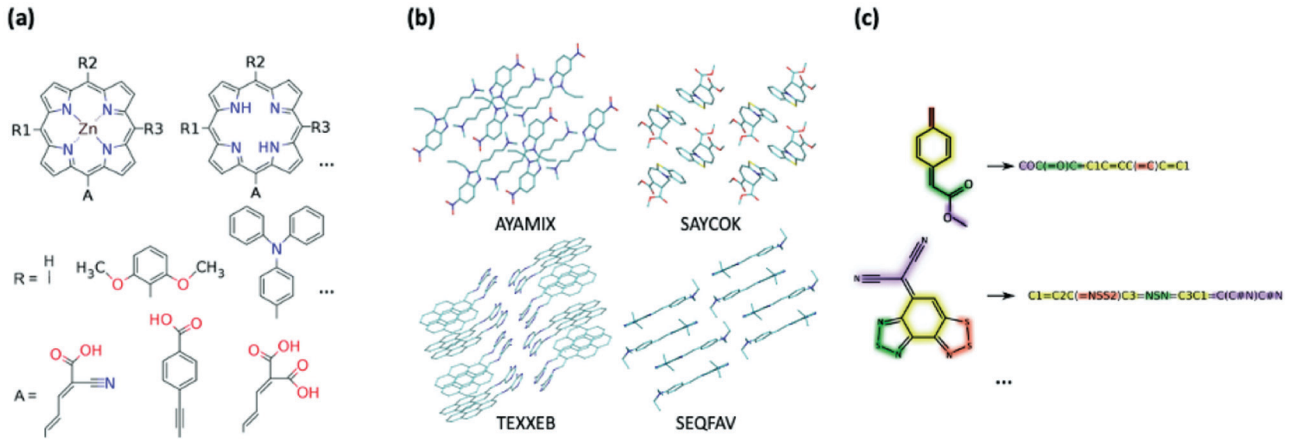


Fig. 4 Example of generation of datasets for HTVP *via* (a) combinatorial modification of chemical motifs (from ref. 361), (b) structural dataset containing geometry and molecular crystal arrangements, (c) chemical dataset represented only through their chemical connectivity, *e.g.* SMILES strings.

structures based on over 57 000 small organic compounds and their conformations.[364] Using benchmarked DFT to provide more chemically accurate physicochemical results, ML algorithms are given a wealth of training data, including the fine structural differences. In a similar vein, the QM7-X set contains physicochemical data on approximately 4.2 million structures,[365] including many isomers for improved application of ML techniques in drug design;[366] however this is limited to seven heavy atoms. As the last few examples indicate, exhaustive combinatorial searches will be unable, for the foreseeable future, to deal with the molecular size relevant for organic electronics.

b. Structural databases

As already alluded to, HTVS efforts which utilise databases of known chemical compounds have the advantage that any interesting compound found through them can be synthesised and is sufficiently stable for its initial characterisation to have taken place. Structural databases give additional access to the molecular geometry (within the crystal, Fig. 4b). For example, the CSD contains over 1 M stable crystalline entries with over 40 000 organic compounds with a relatively small computed HOMO–LUMO gap.[37] In most cases, easy access to the work which outlines the synthesis is also made available. One immediate advantage of using these X-ray geometries is that the conformational search and energy optimisation step can be omitted with a substantial reduction of computational time. In our previous works to find SF[37] and TADF[367] candidates among the CSD, we could directly evaluate the low excited state energies and oscillator strengths on experimental geometries. Geometry optimisation was performed only on the most promising candidates computed as isolated molecules, most of which retained their desirable properties after optimisation. A. Stuke *et al.* also used the CSD for relevance and generated a dataset of 62 000 entries for spectroscopic applications;[368] however, they relaxed the geometries using DFT, prior to computation of orbital energies.

It is particularly advantageous to use structural databases for HTVS when considering properties that depend on the intermolecular arrangement like charge mobility of excitonic properties. C. Schober *et al.* considered crystal arrangements to investigate high-mobility structures within the CSD.[269] By considering a filtered set of approximately 95k crystal structures, they computed the electronic coupling between molecules in contact in the experimental geometry of the crystal and evaluated the mobility in the hopping limit. Another HTVS effort based on the CSD focused on the limit of coherent transport and included the effect of non-local electron phonon coupling.[50] This required the evaluation of the local vibration of molecules embedded in their crystalline environment. Molecular arrangements in the crystal are also needed to study excitonic properties, as illustrated in a recent survey of ~2200 crystals formed by molecules with bright lowest excited states for which the lowest excitonic band was characterised.[239]

c. Chemical databases

In a similar vein to using geometries taken from structural databases, properties taken from chemical databases can also be used for HTVS. Chemical databases, such as PubChem[369] and ZINC,[310] enumerate millions of existing small molecules and have been heavily utilised in the field of drug discovery (a list of chemical databases is provided in ref. 370). The greatest advantage of using these databases is that each molecule is inherently synthesisable and often have chemical vendors linked for immediate purchase. Since experimental geometries are not provided, it is still necessary to construct reliable 3D geometries from the basic structural data encoded in the database, such as the very compact SMILES string (Fig. 4c).[371,372] This means that, for a molecule, the usual geometry construction, conformer search and geometry optimisation should be performed prior to the computation of any specific property.

In their ML-based screening, P. M. Tagade *et al.* took a randomised subset of the PubChem database with approximately 78k structures to find a way of predicting redox potentials and frontier orbital energies based on chemical structure.[373] They performed geometry optimisations for all of their structures at

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

the B3LYP/6-311+G(d,p) level of theory. Following the same randomised selection logic, P. T. St. John *et al.* chose 40k closed-shell molecules, with 200k corresponding radicals, from the PubChem database for a DFT screening and provided their database for future ML works based on radical chemistry.[374] However, in this case, they imposed criteria such as element selectivity, number of heavy atoms, net charge, bond type, *etc.* to ensure that all entries are relevant. This filtration is an important pre-step when using structural and chemical databases, especially for targeted properties. O. Borodin *et al.*, for example, considered only a specific set of 400 carbonate and phosphate molecules chosen from the PubChem database and screened for electrochemical stability of battery electrolytes.[375]

Considering much larger sets, the PubChemQC Project provides the ground-state electronic structure of 3M molecules from the PubChem database, and low-lying excited-state energies for 2 M of them *via* a web interface.[376] Since the authors have performed calculation in increasing order of molecular weight, lighter molecules of more limited interest to organic electronics are more represented but the project well exemplifies the current capability of HTVS. In choosing the set of molecules to compute, molecules with erroneous representations in their cheminformatics representations (InChI[377] and SMILES) were identified and removed, element selectivity based on the capability of the 6-31G* basis set were chosen and isotopes were ignored. All of the structures, having been generated using the Open Babel code,[378] were subjected to optimisation using the semi-empirical PM3 method,[379,380] Hartree–Fock with the STO-6G basis set, DFT at the B3LYP/6-31G* level and finally TD-DFT at the same level of theory.

d. Generative models

Generative models are trained with a database of molecules and their functionality and are then used to construct new chemical structures that have a desired functionality. This is often referred to as inverse design (see Fig. 5a), as opposed to the conventional approach where one first designs a compound and then predict its properties. Generative models have been used to generate novel information like new human faces[381] or music,[382] and some of these models, like variational auto-encoders[383] and generative adversarial networks[384] have recently been used to suggest prospective organic molecules.[385] These models offer the possibility of aiding HTVS, as they can increase the amount of data analysed and offer an alternative to expand on human chemical intuition, with the potential to reduce human biases and detecting trends that may escape the human eye. Most applications of generative models in chemistry are in the field of drug discovery[386,387] but these techniques have also been gaining traction in organic electronics.

In Fig. 5b, we show an inverse design strategy where a deep neural network (DNN) is used to encode the structural features and predict their target property, and a recurrent neural network (RNN) is used to decode this information and propose molecular structures with a desired target property. This approach has been used to construct an inverse design strategy that analysed 40 000 random chemical structures and was able

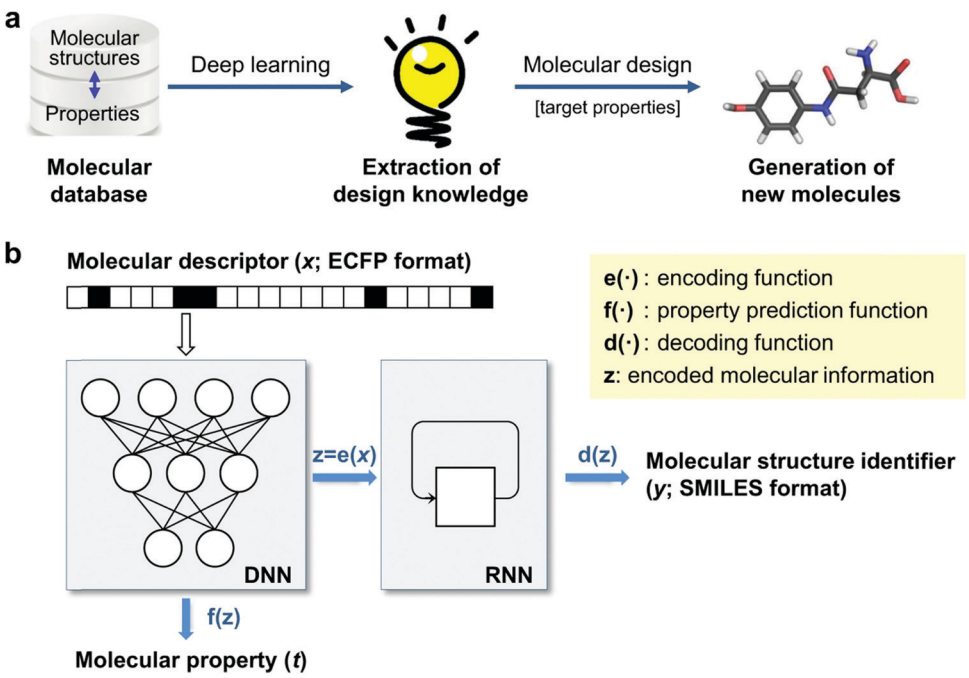


Fig. 5 Data-driven inverse design. (a) Concept of inverse design: hidden knowledge for molecular design is extracted from a given molecular database in a fully data-driven manner using deep-learning, and new molecules with the target properties are generated subsequently. (b) Deep encoder–decoder architecture of inverse design model: the encoding and property prediction functions are obtained by a DNN using the molecular descriptor as an input, and the decoding function is obtained by an RNN using the encoding function as an input to generate the molecular identifier. Reprinted from ref. 388, under a Creative Commons Attribution 4.0 International (CC BY) license.

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

to propose more than 3000 unique chemically plausible structures in the targeted range of $T_1 \geq 3.0$ eV, relevant for phosphorescent organic light-emitting diodes.[388] Using DFT, it was found that 58.7% of the proposed molecules was in the desired range, which is a significant improvement from the 36.2% of molecules in the training set within that range. The correlation between T_1 values predicted by the model and the DFT values was 0.881, and three of the proposed molecules were further validated experimentally, showing the potential for this type of inverse design strategies to target specific properties relevant to organic electronics. A similar approach has been used to construct generative and predictive models that were able to propose new non-fullerene acceptors whose properties were validated through DFT calculations.[389] Other similar generative model has been used to propose new donor–acceptor oligomers with specific electronic properties, like the HOMO–LUMO gap and dipole moment, and it has been showed how the training data can affect the values of the electronic properties of the predicted oligomers.[390]

A critical point of generative models is the selection of the chemical space that is analysed. If one chooses a chemical space very similar to the one already known, one is merely interpolating known data. However, molecules that are too different from any known one can be risky to predict, as they can correspond to structures that are not chemically sensible or cannot be synthesised.[359,391] It has been recently noted that generative models often ignore synthesisability. M. Sumita *et al.*[392] prepared a platform to predict photofunctional organic molecules with excited states at a desired range using Monte Carlo tree search[393] and a recurrent neural network. After just a few days of searching, the platform suggested 86 possible candidates, out of which five were confirmed to be synthesisable and stable. These results are promising, as it shows the potential of computer-aided chemistry to discover new molecules, but it also highlights how only a minority the predicted molecules can be synthesised if synthesisability is not explicitly considered when designing generative models, a problem already encountered in Section 3.a. W. Gao *et al.* used a data-driven computer-aided synthesis planning program to show that state-of-the-art generative models often result in non-synthesisable molecules,[357] although this can be partially addressed by adding synthetic complexity heuristic to the models. In the last years, there have been several advances in computer-assisted synthesis planning (CASP) programs to overcome the lack of synthesisability prediction of generative models, by applying retrosynthetic transformations.[394–397]

4. Analysis

Once an HTVS is completed, the large volume of data can be searched to find materials of a given property. The pool of potential candidates found based on preliminary criteria can be tightened by introducing refinement using higher levels of theory, along with a more rigorous selection of rules. After the number of candidates is reduced to include only the very best predicted by theory, experimentation can verify which, if any, of the small set of candidates can be used for device fabrication. This ''computational funnel approach'' popularised for example by Pyzer-Knapp *et al.*[24] is not the only, or even the most common workflow, of HTVS. The pre-screening stage, where the systems to be studied are selected (Section 2 of this work) is always inexpensive. The first round of electronic structure computations is almost always the most expensive and produce a large set of homogenous data that is particularly precious for validating, disproving or discovering structure–property relations (Fig. 6). The following sections discuss how the results are HTVS can be analysed.

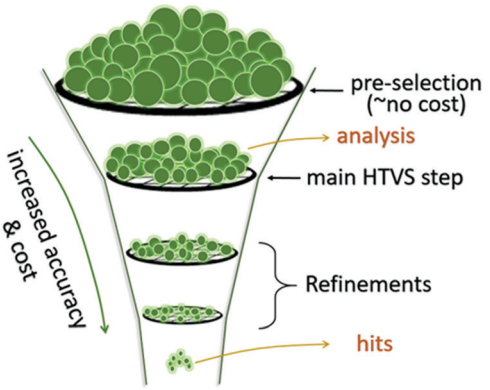


Fig. 6 A revisited schematics of the computational funnel. The first and probably the most important step in is the definition of the systems to study and it normally has a negligible computational cost with respect to the rest. The first level of screening is often dominating the computational costs and it is the one that allows derivation of structure–property relations because of the size and homogeneity of the data produced. Refinements are sometimes needed but are less likely to give important insights.

a. Experimental verification of novel findings

Finding that a property is exhibited experimentally for molecules predicted by HTVS is the most important measure of success for a study seeking to find new optoelectronics. There are many barriers in the pipeline from theory to manufacturing such as chemical stability, environmental conditions and synthetic availability, therefore finding even one new material, among thousands, which can be used in a physical device can be considered as a major advancement. There have been a few attempts in the literature to directly combine HTVS with experimental testing, where some of the best candidates are shown to exhibit the desired property in a practical setting. A. Aspuru-Guzik and co-workers took this approach to seek molecules that exhibit TADF.[42] They successfully found efficient materials, with a reported external quantum efficiency of over 22% in one, for potential use in devices by following their tiered screening approach. In a similar vein, they searched for molecules which exhibit SF by considering over 4000 anthracene derivatives,[36] whereby anthracene is known to exhibit endothermic blue SF.[398,399] Out of this set, they successfully found molecules with the desired energy levels, although with practical limitations.

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Hund's rule has been questioned recently by the existence of singlet–triplet energy inversion in molecules such as heptazine.[400] D. Miyajima *et al.* investigate this extremely novel property by combinatorial generation of a set of approximately 35 000 heptazine derivatives to find those with a negative singlet–triplet gap and with a sufficient oscillator strength, *i.e.* an emissive singlet state.[401] After refinement of the best structures with highly correlated wavefunction methods, it was found that one such candidate exhibited delayed fluorescence from inverted singlet and triplet states experimentally. Seeking more traditional TADF molecules with a low singlet–triplet gap and those for SF with the opposite, T. Chen *et al.* investigated a set of triazines and the effect of their unique derivations on the singlet–triplet gap.[402] With a congruent experimental and theoretical effort, they found that the impact of molecular symmetry and frontier orbital overlap can fine-tune the singlet–triplet splitting to sample a range which can accommodate both TADF and SF among the different triazines. Particularly intriguing is the combination of HTVS with high-throughput experimentation, two fields that are bound to converge.

A distinct but related form of validation is the rediscovery of molecules already known to have the desired property from a database that is not biased toward this property. A similar method was used to validate the screening of molecular semiconductors for SF,[37] TADF[367] and electron acceptors for OPV.[295] In all these cases less than 0.5% of the screened materials had the desirable property but a fraction of them were already known to possess this property providing a good statistical validation of the procedure. There is no consensus among practitioners in the field on whether it is best to pursue experimental validation of HTVS within the same research report. On one hand, it is very useful to demonstrate the potential of the methodology through examples of success. On the other hand, experimental validations may introduce bias in the reporting of HTVS procedure. For example, there are virtually no joint prediction-experiment papers where the experiment disagrees with the prediction, *i.e.* the published literature may depict a too-optimistic view of the prediction accuracy.

It is worth making a connection with the area of autonomous chemical discovery where models and robotics synthesis are integrated.[403–405] This approach is particularly valuable when libraries can be built on the basis of a limited number of precursors and synthetic conditions, *e.g.* donor–acceptor pairs. The virtual screening can be used to narrow down the number of precursors to consider and can become the first step of the autonomous discovery process. The best type of problems to be tackled with autonomous chemistry approach remains the optimisation of experimental conditions or composition of blends.[405] However, the objective of HTVS is often the discovery of completely new lead compounds, requiring novel unpredicted precursors and synthetic routes. In these cases, HTVS and experimental validation cannot be easily integrated and will remain two separate stages of the discovery process.

b. Validation of existing physical hypotheses and approximations

Supporting previous hypotheses and verifying the applicability of widely used approximations is one of the main outcomes of HTVS studies. For example, discovery of new SF materials has been driven by well-established and strict design rules based on biradicaloid structures,[165,166,406,407] which has focused a narrow sight on the well-known acenes and only a few other classes. After expanding the library of predicted SF compounds,[37] a follow-up screening which evaluated the diradical character of the potential candidates showed markedly that the biradicaloid characteristic was well-preserved among the vastly different molecules, but not all of them.[408] Indeed, it is fairly frequent to discover exceptions to the established rules when large datasets are considered and occasionally one finds that the known rules are not valid in a statistical sense. For example, it seems that there is a complete lack of correlation between computed donor–acceptor character in a dye and its efficiency in a dye sensitised solar cell.[409]

In developing theories to describe charge transport in molecular semiconductors, often a number of different assumptions have been considered whose extent of validity can be verified using HTVS analyses. For instance, the analysis of ref. 18 on two structurally different materials extracted from the CSD confirms the viability of considering only a linear nonlocal electron–phonon coupling at least for the considered structures. Furthermore, it shows, in contrast to common assumptions, that the role of high-frequency modes in nonlocal electron–phonon coupling calculations cannot be completely ignored. For instance, the contribution of these modes to the fluctuation of the largest transfer integral of rubrene at room temperature is estimated to be 9%.[285] In spite of strong coupling with high-frequency modes, the percentage is not remarkable because the high-frequency modes are not populated at room temperature. The other common assumption is that the valence band originates only from the HOMO orbital. An extended HTVS study on a set of $\sim$40 000 molecular semiconductors reveals that the median energy separation between HOMO and HOMO$-$1 energy levels is 0.66 eV whereas the charge transfer integral between the neighbouring molecules is never greater than 0.4 eV.[50] This implies the band energies do not overlap effectively and, accordingly, this prevalent approximation is largely valid. The same study confirms the scarcity of molecular semiconductors with bands extending in three dimensions.

c. Discovery of novel structure–property relations

Deriving insights from structure–property relations is an important strategy to develop new functional materials with tailored properties for organic electronics. The HTVS studies, designed to explore a large set of data, provide an excellent framework to derive novel approaches to materials discovery or to verify, in a statistically more rigorous way, the correlations that have been reported on limited samples in the literature. Confirming the lack of correlations between certain parameters is also another advantage of such analysis. The examples below illustrate these ideas in further detail.

A frequent analysis of computed dataset involves the rationalisation of the coupling between electronic and nuclear degrees of freedom. An early example of this is the demonstration that the local electron–phonon coupling and the number

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

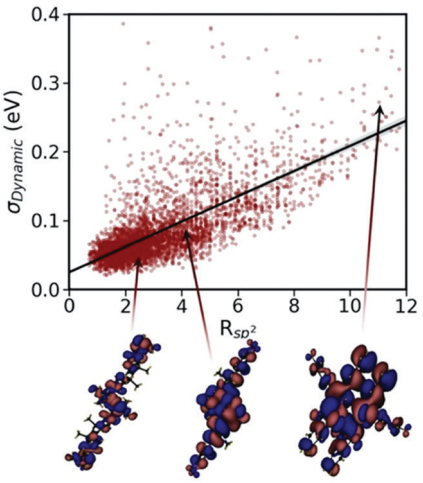


Fig. 7 Correlation between dynamic disorder $\sigma_{Dynamic}$ and fraction of sp^2 atoms in contact (expressed in percentage) in molecular crystals: head to tail arrangement is surprisingly favourable for charge transport.

of π bonded atoms are inversely proportional.[410] In a more recent example, a large set of ~5000 molecular semiconductors is screened to derive strategies for designing molecular crystals with a small level of dynamic disorder $\sigma_{Dynamic}$.[411] According to this analysis, the dynamic disorder between two molecules increases with the fraction of sp^2-hybridised atoms in vdW contact, a property explainable with the nature of the overlap between molecular orbitals with many nodal planes (see Fig. 7). The very counterintuitive implication is that one should not try to design crystals with the maximal superposition of the molecular π-orbitals, as done for many years, because dynamic disorder is minimal when molecules are arranged in a "head-to-tail" configuration and their conjugated cores do not lie on the same plane. When the computed mobility itself is compared with the key system parameters,[50] one finds a strong and expected correlation with the area per molecule, isotropy of the material and parameters controlling nonlocal and local electron phonon couplings and a relatively weak correlation with the transfer integral, in line with recent theories[265] but not anticipated in the early days of organic electronics. Another interesting correlation is the recently reported link, in OPV cells, between electron–phonon coupling and non-radiative voltage losses for a large set of published and new material combinations data implying that the latter are, to some extent, unavoidable.[412]

Several works have attempted to identify the relation between luminescence and chemical structure. For instance, an extensive research incorporating cyclic voltammetry, thermogravimetric analysis, spectroscopic, and theoretical studies of spirobifluorene-derivatives reports that the molecular structure, conformational twisting, structural rigidity, and supramolecular packing play key roles in the photoluminescence of these molecules.[413] Through generating a large set of molecular structures using ChemTS, a python library for *de novo* molecular generation,[414] combined with a proper score function, the lack of correlation between the molecular shape and the high luminescence/absorption dissymmetry factor is reported.[415] Structure–property relations have been sought in a database of 80 non-fullerene electron acceptors,[416] finding that all high performing materials are characterised by a small gap between LUMO and LUMO+1 orbitals as proposed earlier on the basis of elementary models.[417]

The realisation of potential correlations between molecular structure and intra-/inter-molecular interactions has also been of interest. As such, a combined CSP and HTVS study on 27 structural isomers of pyrido[2,3-*b*]pyrido[3′,2′:4,5]pyrrolo[3,2-*g*]indole highlights the lack of correlation between molecular symmetry and preferred crystal packing and consequently intermolecular interactions.[307] A recent work considering a set of over 2000 conjugated polyelectrolytes confirms known relations, *e.g.* that structures composed of alternating electron-rich and electron-poor heterocycle have lower bandgaps, and suggests several additional insights, *e.g.* that a more negatively charged anionic group and a shorter side-chain length are both correlated with higher HOMO level.[418]

d. Determination of the physical limits of certain properties

Screening a large set of data can, not only, potentially identify a number of materials with targeted properties but also can be used to define the plausible physical limit to such properties. The latter is of particular technological interest as it determines how well one can expect to optimise each property. In performing such analysis in a recent project, through screening the CSD, our group identified the maximum achievable excitonic bandwidth (~1.16 eV) and spectral red-shift (~0.6 eV) in excitonic materials.[239] In a similar analysis, a diverse set of conjugated molecules are evaluated to establish the relation between the exciton size and the size of the π-system.[419] This analysis suggests that the exciton binding energy of ~0.3 eV can be considered as the lower limit of this property which is in accordance with the typical theoretical data reported in the literature.[420]

The largest plausible physical limit to charge carrier mobility in molecular semiconductors is evaluated through screening CSD as well.[50] As such, confirming that the main parameters of high mobility materials are either uncorrelated or constructively correlated, a mobility of ~70 $cm^2\ V^{-1}\ s^{-1}$ (Fig. 8) is obtained for a hypothetical material where all its important parameters are set simultaneously to the best one percentile of the distribution found for real materials.

To find promising novel TADF emitters across the visible spectrum, in ref. 42, a large set of 400 000 molecules are screened using TD-DFT. The promising molecules are selected among those possessing high oscillator strength and low singlet–triplet gap and are synthesised and assessed experimentally to evaluate the predictive power of the employed screening protocol. The experimental measurements of the study, in agreement with theoretical predictions, record 22% as the highest electroluminescence quantum efficiency in the targeted molecules of the considered database.

Determining the limits of optical properties is also of great technological relevance. For instance, the transparency of millions

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

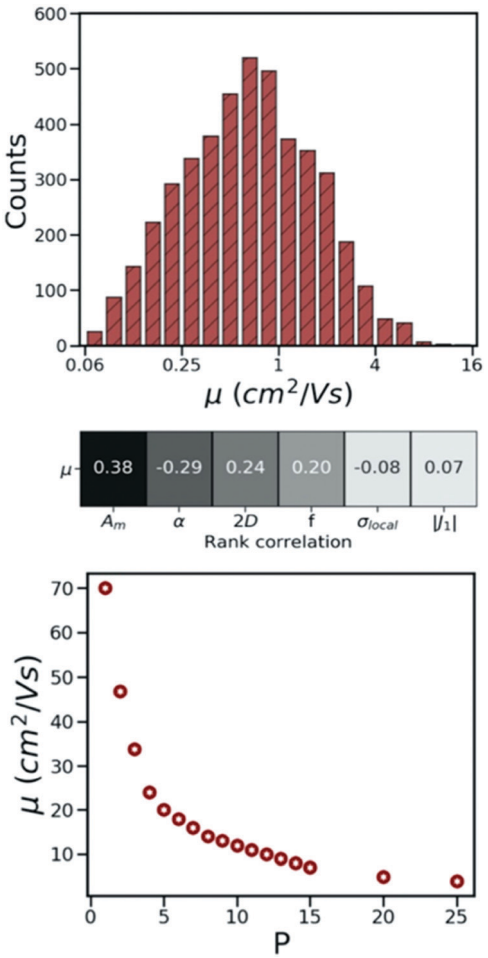


Fig. 8 (top) Distribution of the computed mobility for a set of 5000 molecular semiconductors extracted from CSD. (middle) The rank correlation array between the mobility and important characteristics (molecular area A_m, effective dynamic disorder α, transport two-dimensionality 2D, band renormalisation factor f, and the largest transfer integral J_1). (bottom) Expected mobility for a hypothetical material with the important parameters being set simultaneously in the best P percentile of the distribution of real materials.

of device configurations has been evaluated in an HTVS to identify the best semi-transparent photovoltaics device for power-generating windows to be used in buildings or automobiles.[421] Accordingly, 11% PCE and 30% visible light transmittance are estimated as the upper limit of these structures' characteristics. In another study, through applying the Lorenz–Lorentz equation to a set of over 60 polymers, the theoretical lower limit of the refractive index of organic polymers is evaluated to be 1.29.[422]

e. Machine learning techniques to predict complex properties

Some complex properties, like the PCE of solar cells, cannot be computed from the knowledge of the chemical composition alone because of intrinsic limitations in the modelling capability, including an incomplete physical understanding of the underlying physics. Some simple models offer an approximation to these properties, for example, Scharber's model[423] links the frontier molecular orbitals to the PCE, although it has been shown to have some limitations in reproducing recent experimental data.[424,425] In these situations, ML models are especially useful and they can learn the correlations between the property of interest and specific features. For example, several studies have recently reported ML models able to predict PCE of OPVs with a large accuracy when using molecular fingerprints and a variety of physical features.[424,426–429] These developments will not be reviewed in this work and we refer the reader to recent reviews on how ML has been applied to OPVs[430] and energy materials as a whole.[431] In typical situations, the ML model is a function that predicts experimental properties using, as the input, features of the systems that can be easily accessed. In organic electronic applications one often wishes to predict the properties of devices made with novel compounds for which there is no experimental information available. For this reason, many features used in ML methods are computed electronic structure parameters and a typical work identifies a functional relation between computed features and experimental figure of merit, sometimes followed by the generation of novel structures to be screened with the ML model (see, *e.g.* Fig. 9). Examples of ML guiding HTVS can be found in several fields like OPVs with fullerene acceptors[425] and non-fullerene acceptors,[432] polymer-fullerene solar cells,[433,434] dye-sensitised solar cells,[40] and organic light-emitting diodes.[42]

One should note that available datasets, especially in organic electronics, will likely have clusters of data points around known well-performing materials, while other areas of the chemical space will be sparsely populated also because of the reduced reporting of negative results. ML models, generally built for very large data sets, perform very poorly when trained on biased datasets. J. Sieg *et al.* have tried to tackle this problem and they have shown possible ways to correct for these biases,[436] such as improved baseline assessments and scoring functions, or use higher quality/more abundant data. Some possible solutions to this data biases include the inclusion of failed experiments in the datasets,[437] improved design of experiments[438] and alternative metrics to judge the performance of ML models.[439,440] Additionally, one should be careful when using ML to extrapolate the properties of new materials, *i.e.* calculating the properties of families of materials not present in the dataset, as their capacity to do so is contentious.[441,442] Other recent studies have focused on how improving the ML-guided exploration of materials space, and for example active machine learning has proven able to identify OSCs with larger charge conduction properties than competing exploration strategies.[443]

Because of the way ML methods are conceived, they are unlikely to provide guidance on the maximum achievable performance on each technological domain as in the examples presented in Section 4.d but they can be very powerful in discovering novel correlations and ranking the importance of the features included in any model in particular with the advances of explainable ML.[439] For example, recent studies have showed how ML can be a fast and efficient way to select effective molecular features correlated with the PCE of polymer materials,[444] and can be used to rationalise the effects that the

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

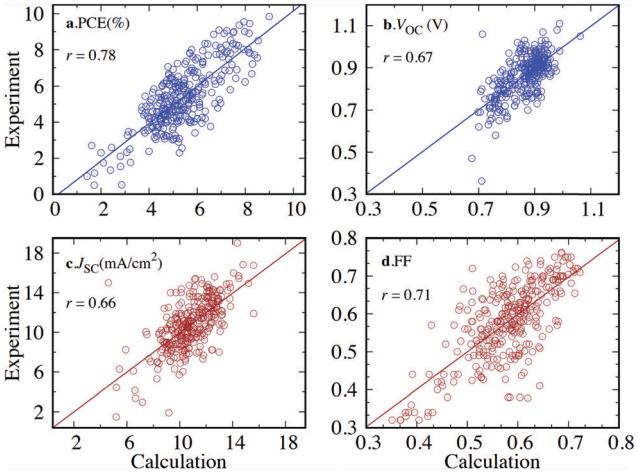


Fig. 9 ML-predicted *versus* experimental data for PCE, V_{OC}, J_{SC}, and FF for all data points (300 OPVs) using the leave-one-out cross validation approach. Reprinted with permission from ref. 435. Copyright 2021 American Chemical Society.

feature's trends have on the open-circuit voltage (V_{OC}), short-circuit current (J_{SC}) and fill factor (FF) of OPVs.[435]

5. Conclusions

As this review of the literature has shown, the application of the high-throughput screening method in organic electronics is constantly evolving and is now solidly a part of the digital discovery tools of molecular sciences. This evolution is driven by the ever-increasing number of applications where novel combinations of properties are required (see, *e.g.* Table 1), the improved understanding of the underlying physics (see *e.g.* Section 2.c), the acceleration in computational methodologies (see, *e.g.* Section 2.b), the collection of more robust datasets of materials and their properties (see, *e.g.* Section 2.a) and the ability to generate numerous and more efficient hypothetical materials to be explored (see, *e.g.* Section 3.d). The dynamicity of the field implies that it is very unlikely that any well-designed virtual screening protocol will provide the definitive answer for any given technological problem, not least because the scope of the screening can vary, *e.g.* from exploring all small variations of a given compound to considering a totally unexplored chemical space for breakthroughs. It is conceivable that each of the many quests for organic electronic materials will be accompanied for many years by a range of virtual screening protocols, just as it is the case for drug discovery. In this sense, being able to classify the different protocols in terms of a sequence of choices to be made and validated is helpful not only for navigating the literature but, more importantly, to introduce systematic improvements in the field.

Specific directions for improvement can be identified for all components of the protocol. Experimental datasets for validation should be expanded to cover a broader set of compounds of relevance for organic electronics using standardised characterization methodologies and, possibly, automation. Standardised computational datasets for the calibration of more approximate methods should be developed to accelerate the direct comparison of different approximate methodologies. The outcome of HTVS should be measured against novelty (with respect to common chemical knowledge) and feasibility (or cost effectiveness) of the discovered compounds. These are somewhat less quantitative aspects that nevertheless can be built in the construction of the virtual dataset to be explored and determine the impact of such explorations on technology.

Finally, high-throughput virtual screening will increasingly play a crucial role in the development of molecular science itself. It is, in many ways, the culmination of physical (or statistical) understanding where the degree of accuracy of the predictions is sufficient to determine the direction of future investigations and, in this sense, it can be seen as the ultimate objective of theoretical modelling. It also sets a new iteration in the process of scientific discovery with technologists exploiting the predictions validated by experiments and scientists focusing on what went wrong in the models and how to improve them. The large size of the datasets introduces, however, an important change to the traditional theory-experiment cycle. A well-executed high-throughput virtual screening does not allow the exclusion of predicted or experimental data points because, by definition, the data are too dense for individual inspection. The quantity of data prevents bias in the assessment of the quality of the model and this will unavoidably yield better models followed by better materials.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We are grateful to Zhi-Wen Zhao, Daniele Padula and Ke Zhao for helpful discussions. We acknowledge the financial support of EPSRC, ERC (Grant No. 862102).

References

1 A. J. Heeger, S. Kivelson, J. R. Schrieffer and W. P. Su, *Rev. Mod. Phys.*, 1988, **60**, 781–850.
2 S. J. Zou, Y. Shen, F. M. Xie, J. De Chen, Y. Q. Li and J. X. Tang, *Mater. Chem. Front.*, 2020, **4**, 788–820.
3 M. Riede, D. Spoltore and K. Leo, *Adv. Energy Mater.*, 2021, **11**, 2002653.
4 C. Y. Yang, M. A. Stoeckel, T. P. Ruoko, H. Y. Wu, X. Liu, N. B. Kolhe, Z. Wu, Y. Puttisong, C. Musumeci, M. Massetti, H. Sun, K. Xu, D. Tu, W. M. Chen, H. Y. Woo, M. Fahlman, S. A. Jenekhe, M. Berggren and S. Fabiano, *Nat. Commun.*, 2021, **12**, 2354.
5 J. Gierschner, S. K. Behera and S. Y. Park, *Angew. Chem., Int. Ed.*, 2021, **60**, 2–17.

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

6 W. Zhao, Z. He and B. Z. Tang, *Nat. Rev. Mater.*, 2020, **5**, 869–885.
7 X. Yin, Y. He, X. Wang, Z. Wu, E. Pang, J. Xu and J. Wang, *Front. Chem.*, 2020, **8**, 725.
8 A. Rao and R. H. Friend, *Nat. Rev. Mater.*, 2017, **2**, 17063.
9 N. Li, N. Eedugurala, D. S. Leem, J. D. Azoulay and T. N. Ng, *Adv. Funct. Mater.*, 2021, **31**, 2100565.
10 S. G. Higgins, A. Lo Fiego, I. Patrick, A. Creamer and M. M. Stevens, *Adv. Mater. Technol.*, 2020, **5**, 2000384.
11 A. Melianas, T. J. Quill, G. LeCroy, Y. Tuchman, H. V. Loo, S. T. Keene, A. Giovannitti, H. R. Lee, I. P. Maria, I. McCulloch and A. Salleo, *Sci. Adv.*, 2020, **6**, 27.
12 M. A. Castellanos, A. Dodin and A. P. Willard, *Phys. Chem. Chem. Phys.*, 2020, **22**, 3048–3057.
13 H. Sirringhaus, P. J. Brown, R. H. Friend, M. M. Nielsen, K. Bechgaard, B. M. W. Langeveld-Voss, A. J. H. Spiering, R. A. J. Janssen, E. W. Meijer, P. Herwig and D. M. De Leeuw, *Nature*, 1999, **401**, 685–688.
14 J. E. Anthony, J. S. Brooks, D. L. Eaton and S. R. Parkin, *J. Am. Chem. Soc.*, 2001, **123**, 9482–9483.
15 H. Yan, Z. Chen, Y. Zheng, C. Newman, J. R. Quinn, F. Dötz, M. Kastler and A. Facchetti, *Nature*, 2009, **457**, 679–686.
16 S. Fratini, M. Nikolka, A. Salleo, G. Schweicher and H. Sirringhaus, *Nat. Mater.*, 2020, **19**, 491–502.
17 S. Few, J. M. Frost and J. Nelson, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2311–2325.
18 T. Nematiaram and A. Troisi, *J. Chem. Phys.*, 2020, **152**, 190902.
19 J. L. Brédas, D. Beljonne, V. Coropceanu and J. Cornil, *Chem. Rev.*, 2004, **104**, 4971–5003.
20 Y. Cui, P. Zhu, X. Liao and Y. Chen, *J. Mater. Chem. C*, 2020, **8**, 15920–15939.
21 P. Friederich, A. Fediai, S. Kaiser, M. Konrad, N. Jung and W. Wenzel, *Adv. Mater.*, 2019, **31**, 1808256.
22 R. Pollice, G. Dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao and A. Aspuru-Guzik, *Acc. Chem. Res.*, 2021, **54**, 849–860.
23 L. Himanen, A. Geurts, A. S. Foster and P. Rinke, *Adv. Sci.*, 2019, **6**, 1900808.
24 E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre and A. Aspuru-Guzik, *Annu. Rev. Mater. Res.*, 2015, **45**, 195–216.
25 S. Hirata and K. Shizu, *Nat. Mater.*, 2016, **15**, 1056–1057.
26 A. Saeki and K. Kranthiraja, *Jpn. J. Appl. Phys.*, 2020, **59**, SD0801.
27 S. Luo, T. Li, X. Wang, M. Faizan and L. Zhang, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, 1489.
28 G. A. McCarver, T. Rajeshkumar and K. D. Vogiatzis, *Coord. Chem. Rev.*, 2021, **436**, 213777.
29 X. Zhang, A. Chen and Z. Zhou, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2019, **9**, 1385.
30 H. Daglar and S. Keskin, *Coord. Chem. Rev.*, 2020, **422**, 213470.
31 G. Fang, L. Xu, Y. Cao and A. Li, *Coord. Chem. Rev.*, 2016, **322**, 94–103.
32 M. A. Makeev and N. N. Rajput, *Curr. Opin. Chem. Eng.*, 2019, **23**, 58–69.
33 A. Jain, Y. Shin and K. A. Persson, *Nat. Rev. Mater.*, 2016, **1**, 15004.
34 S. A. Lopez, B. Sanchez-Lengeling, J. de Goes Soares and A. Aspuru-Guzik, *Joule*, 2017, **1**, 857–870.
35 J. Hachmann, R. Olivares-Amaya, A. Jinich, A. L. Appleton, M. A. Blood-Forsythe, L. R. Seress, C. Román-Salgado, K. Trepte, S. Atahan-Evrenk, S. Er, S. Shrestha, R. Mondal, A. Sokolov, Z. Bao and A. Aspuru-Guzik, *Energy Environ. Sci.*, 2014, **7**, 698–704.
36 C. F. Perkinson, D. P. Tabor, M. Einzinger, D. Sheberla, H. Utzat, T. A. Lin, D. N. Congreve, M. G. Bawendi, A. Aspuru-Guzik and M. A. Baldo, *J. Chem. Phys.*, 2019, **151**, 121102.
37 D. Padula, Ö. H. Omar, T. Nematiaram and A. Troisi, *Energy Environ. Sci.*, 2019, **12**, 2412–2416.
38 F. Hakim and M. K. Alam, *Sol. Energy*, 2019, **191**, 300–310.
39 C. H. Teh, R. Daik, E. L. Lim, C. C. Yap, M. A. Ibrahim, N. A. Ludin, K. Sopian and M. A. Mat Teridi, *J. Mater. Chem. A*, 2016, **4**, 15788–15822.
40 Y. Wen, L. Fu, G. Li, J. Ma and H. Ma, *Sol. RRL*, 2020, **6**, 2000110.
41 Z. Yang, C. Liu, C. Shao, C. Lin and Y. Liu, *J. Phys. Chem. C*, 2015, **119**, 21852–21859.
42 R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams and A. Aspuru-Guzik, *Nat. Mater.*, 2016, **15**, 1120–1127.
43 A. Zampetti, A. Minotto and F. Cacialli, *Adv. Funct. Mater.*, 2019, **29**, 1807623.
44 X. K. Chen, D. Kim and J. L. Brédas, *Acc. Chem. Res.*, 2018, **51**, 2215–2224.
45 J. Lee, N. Aizawa and T. Yasuda, *J. Mater. Chem. C*, 2018, **6**, 3578–3583.
46 I. Bhattacharjee, N. Acharya, H. Bhatia and D. Ray, *J. Phys. Chem. Lett.*, 2018, **9**, 2733–2738.
47 Y. Jiang, Y. Y. Liu, X. Liu, H. Lin, K. Gao, W. Y. Lai and W. Huang, *Chem. Soc. Rev.*, 2020, **49**, 5885–5944.
48 M. Kaur and D. H. Choi, *Chem. Soc. Rev.*, 2015, **44**, 58–77.
49 J. Li and K. Pu, *Chem. Soc. Rev.*, 2019, **48**, 38–71.
50 T. Nematiaram, D. Padula, A. Landi and A. Troisi, *Adv. Funct. Mater.*, 2020, **30**, 2001906.
51 B. Yang, Y. Lu, D. Jiang, Z. Li, Y. Zeng, S. Zhang, Y. Ye, Z. Liu, Q. Ou, Y. Wang, S. Dai, Y. Yi and J. Huang, *Adv. Mater.*, 2020, **32**, 2001227.
52 J. Tao, D. Liu, Z. Qin, B. Shao, J. Jing, H. Li, H. Dong, B. Xu and W. Tian, *Adv. Mater.*, 2020, **32**, 1907791.
53 M. Chen, C. Wang and W. Hu, *J. Mater. Chem. C*, 2021, **9**, 4709–4729.
54 N. Martino, D. Ghezzi, F. Benfenati, G. Lanzani and M. R. Antognazza, *J. Mater. Chem. B*, 2013, **1**, 3768–3780.
55 W. Deng, X. Zhang, R. Jia, L. Huang, X. Zhang and J. Jie, *NPG Asia Mater.*, 2019, **11**, 77.

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

56 C. Wang, X. Zhang and W. Hu, *Chem. Soc. Rev.*, 2020, **49**, 653–670.
57 H. J. Cheon, X. Li, Y. J. Jeong, M. J. Sung, Z. Li, I. Jeon, X. Tang, H. G. Girma, H. Kong, S. K. Kwon, T. K. An, S. H. Kim and Y. H. Kim, *J. Mater. Chem. C*, 2020, **8**, 8410–8419.
58 R. Song, Z. Wang, X. Zhou, L. Huang and L. Chi, *ChemPlusChem*, 2019, **84**, 1222–1234.
59 J. Borges-González, C. J. Kousseff and C. B. Nielsen, *J. Mater. Chem. C*, 2019, **7**, 1111–1130.
60 Y. Li, Q. Qian, X. Zhu, Y. Li, M. Zhang, J. Li, C. Ma, H. Li, J. Lu and Q. Zhang, *InfoMat*, 2020, **2**, 995–1033.
61 M. Jia, Y. Li, M. Huang, K. J. Kim, Y. Liu and S. Cao, *Synth. Met.*, 2020, **266**, 116431.
62 K. M. Pelzer, L. Cheng and L. A. Curtiss, *J. Phys. Chem. C*, 2017, **121**, 237–245.
63 S. Sun, Y. Chen and J. Yu, *J. Phys. Chem. C*, 2015, **119**, 25770–25777.
64 O. Allam, B. W. Cho, K. C. Kim and S. S. Jang, *RSC Adv.*, 2018, **8**, 39414–39420.
65 J. Banerjee, K. Dutta, M. A. Kader and S. K. Nayak, *Polym. Adv. Technol.*, 2019, **30**, 1902–1921.
66 H. Wang, H. Yang, Y. Diao, Y. Lu, K. Chrulski and J. M. D'Arcy, *ACS Nano*, 2021, **15**, 7799–7810.
67 M. J. Griffith, S. Cottam, J. Stamenkovic, J. A. Posar and M. Petasecca, *Front. Phys.*, 2020, **8**, 1–21.
68 J. Devkota, R. Geng, R. C. Subedi and T. D. Nguyen, *Adv. Funct. Mater.*, 2016, **26**, 3881–3898.
69 V. Fiore, P. Battiato, S. Abdinia, S. Jacobs, I. Chartier, R. Coppard, G. Klink, E. Cantatore, E. Ragonese and G. Palmisano, *IEEE Trans. Circuits Syst. I Regul. Pap.*, 2015, **62**, 1668–1677.
70 P. A. Shaposhnik, S. A. Zapunidi, M. V. Shestakov, E. V. Agina and S. A. Ponomarenko, *Russ. Chem. Rev.*, 2020, **89**, 1483–1506.
71 S. Pecqueur, D. Vuillaume and F. Alibart, *J. Appl. Phys.*, 2018, **124**, 151902.
72 M. K. Wecławski, M. Jakešová, M. Charyton, N. Demitri, B. Koszarna, K. Oppelt, S. Sariciftci, D. T. Gryko and E. D. Głowacki, *J. Mater. Chem. A*, 2017, **5**, 20780–20788.
73 Y. Z. Chen, W. H. Li, L. Li and L. N. Wang, *Rare Met.*, 2018, **37**, 1–12.
74 I. Heath-Apostolopoulos, D. Vargas-Ortiz, L. Wilbraham, K. E. Jelfs and M. A. Zwijnenburg, *Sustainable Energy Fuels*, 2021, **5**, 704–719.
75 R. Sarkar, M. Boggio-Pasqua, P. F. Loos and D. Jacquemin, *J. Chem. Theory Comput.*, 2021, **17**, 1117–1132.
76 Q. Yang, M. Mendolicchio, V. Barone and J. Bloino, *Front. Astron. Space Sci.*, 2021, **8**, 77.
77 É. Brémond, A. Ottochian, Á. J. Pérez-Jiménez, I. Ciofini, G. Scalmani, M. J. Frisch, J. C. Sancho-García and C. Adamo, *J. Comput. Chem.*, 2021, **42**, 970–981.
78 A. G. Donchev, A. G. Taube, E. Decolvenaere, C. Hargus, R. T. McGibbon, K. H. Law, B. A. Gregersen, J. L. Li, K. Palmo, K. Siva, M. Bergdorf, J. L. Klepeis and D. E. Shaw, *Sci. Data*, 2010, **8**, 55.
79 Y. Xu, R. Friedman, W. Wu and P. Su, *J. Chem. Phys.*, 2021, **154**, 194106.
80 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, 3B.
81 M. Montalti, A. Credi, L. Prodi and M. T. Gandolfi, *Handbook of Photochemistry*, 2006.
82 E. J. Beard, G. Sivaraman, Á. Vázquez-Mayagoitia, V. Vishwanath and J. M. Cole, *Sci. Data*, 2019, **6**, 307.
83 R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.*, 2007, **79**, 291–352.
84 D. Jacquemin, V. Wathelet, E. A. Perpète and C. Adamo, *J. Chem. Theory Comput.*, 2009, **5**, 2420–2435.
85 M. Nakata, T. Shimazaki, M. Hashimoto and T. Maeda, *J. Chem. Inf. Model.*, 2020, **60**, 5891–5899.
86 D. Balcells and B. B. Skjelstad, *J. Chem. Inf. Model.*, 2020, **60**, 6135–6146.
87 K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller and A. Tkatchenko, *Nat. Commun.*, 2017, **8**, 13890.
88 S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt and K. R. Müller, *Sci. Adv.*, 2017, **3**, 5.
89 F. Neese, A. Hansen, F. Wennmohs and S. Grimme, *Acc. Chem. Res.*, 2009, **42**, 641–648.
90 D. Jacquemin, I. Duchemin and X. Blase, *J. Chem. Theory Comput.*, 2015, **11**, 5340–5359.
91 S. Tortorella, M. M. Talamo, A. Cardone, M. Pastore and F. De Angelis, *J. Phys.: Condens. Matter*, 2016, **28**, 7.
92 G. Zhang and C. B. Musgrave, *J. Phys. Chem. A*, 2007, **111**, 1554–1561.
93 L. Wilbraham, E. Berardo, L. Turcani, K. E. Jelfs and M. A. Zwijnenburg, *J. Chem. Inf. Model.*, 2018, **58**, 2450–2459.
94 I. Heath-Apostolopoulos, L. Wilbraham and M. A. Zwijnenburg, *Faraday Discuss.*, 2019, **215**, 98–110.
95 I. Y. Kanal, S. G. Owens, J. S. Bechtel and G. R. Hutchison, *J. Phys. Chem. Lett.*, 2013, **4**, 1613–1623.
96 N. M. Oboyle, C. M. Campbell and G. R. Hutchison, *J. Phys. Chem. C*, 2011, **115**, 16200–16210.
97 G. R. Hutchison, M. A. Ratner and T. J. Marks, *J. Phys. Chem. A*, 2002, **106**, 10596–10605.
98 L. A. Curtiss, P. C. Redfern, K. Raghavachari and J. A. Pople, *J. Chem. Phys.*, 1998, **109**, 42–55.
99 S. G. Lias, H. M. Rosenstock, K. Draxl, B. W. Steiner, J. T. Herron, J. L. Holmes, R. D. Levin, J. F. Liebman and S. A. Kafafi, *Natl. Inst. Stand. Technol.*, 2011, **20899**, 69.
100 J. C. Rienstra-Kiracofe, G. S. Tschumper, H. F. Schaefer, S. Nandi and G. B. Ellison, *Chem. Rev.*, 2002, **102**, 231–282.
101 T. Koopmans, *Physica*, 1934, **1**, 104–113.
102 S. Refaely-Abramson, R. Baer and L. Kronik, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 075144.
103 T. Stein, H. Eisenberg, L. Kronik and R. Baer, *Phys. Rev. Lett.*, 2010, **105**, 266802.
104 T. Körzdörfer, J. S. Sears, C. Sutton and J. L. Brédas, *J. Chem. Phys.*, 2011, **135**, 204107.
105 R. Baer, E. Livshits and U. Salzner, *Annu. Rev. Phys. Chem.*, 2010, **61**, 85–109.
106 H. Sun and J. Autschbach, *J. Chem. Theory Comput.*, 2014, **10**, 1035–1047.
107 T. Zeng, N. Ananth and R. Hoffmann, *J. Am. Chem. Soc.*, 2014, **136**, 12638–12647.

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

108 L. Gallandi, N. Marom, P. Rinke and T. Körzdörfer, *J. Chem. Theory Comput.*, 2016, **12**, 605–614.
109 F. Bruneval and M. A. L. Marques, *J. Chem. Theory Comput.*, 2013, **9**, 324–329.
110 N. Marom, F. Caruso, X. Ren, O. T. Hofmann, T. Körzdörfer, J. R. Chelikowsky, A. Rubio, M. Scheffler and P. Rinke, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 245127.
111 X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter and M. Scheffler, *New J. Phys.*, 2012, **14**, 053020.
112 X. Blase, C. Attaccalite and V. Olevano, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 115103.
113 A. Stan, N. E. Dahlen and R. Van Leeuwen, *Europhys. Lett.*, 2006, **76**, 298–304.
114 M. J. Van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers and P. Rinke, *J. Chem. Theory Comput.*, 2015, **11**, 5665–5687.
115 F. Caruso, P. Rinke, X. Ren, M. Scheffler and A. Rubio, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 081102.
116 T. A. Pham, H. V. Nguyen, D. Rocca and G. Galli, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 155148.
117 K. Krause, M. E. Harding and W. Klopper, *Mol. Phys.*, 2015, **113**, 1952–1960.
118 S. H. Ke, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 205415.
119 F. Aryasetiawan and O. Gunnarsson, *Rep. Prog. Phys.*, 1998, **61**, 237–312.
120 M. S. Hybertsen and S. G. Louie, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **34**, 5390–5413.
121 L. Hedin, *Phys. Rev.*, 1965, **139**, 3A.
122 G. Onida, L. Reining and A. Rubio, *Rev. Mod. Phys.*, 2002, **74**, 601–659.
123 J. W. Knight, X. Wang, L. Gallandi, O. Dolgounitcheva, X. Ren, J. V. Ortiz, P. Rinke, T. Körzdörfer and N. Marom, *J. Chem. Theory Comput.*, 2016, **12**, 615–626.
124 T. Rangel, S. M. Hamed, F. Bruneval and J. B. Neaton, *J. Chem. Theory Comput.*, 2016, **12**, 2834–2842.
125 J. V. Ortiz, *Computational Chemistry: Reviews of Current Trends*, 1997, vol. 2, pp. 1–61.
126 R. Flores-Moreno and J. V. Ortiz, *Practical Aspects of Computational Chemistry: Methods, Concepts and Applications*, 2010, pp. 1–17.
127 J. V. Ortiz, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 123–142.
128 J. V. Ortiz, *Annual Reports in Computational Chemistry*, 2017, vol. 13, pp. 139–182.
129 H. H. Corzo and J. V. Ortiz, *Advances in Quantum Chemistry*, 2017, vol. 74, pp. 267–298.
130 M. Díaz-Tinoco, H. H. Corzo and J. V. Ortiz, *J. Chem. Theory Comput.*, 2018, **14**, 5881–5895.
131 O. Dolgounitcheva, M. Díaz-Tinoco, V. G. Zakrzewski, R. M. Richard, N. Marom, C. D. Sherrill and J. V. Ortiz, *J. Chem. Theory Comput.*, 2016, **12**, 627–637.
132 J. K. Kang and C. B. Musgrave, *J. Chem. Phys.*, 2001, **115**, 11040–11051.
133 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
134 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
135 Y. Fu, L. Liu, H. Z. Yu, Y. M. Wang and Q. X. Guo, *J. Am. Chem. Soc.*, 2005, **127**, 7227–7234.
136 H. Phillips, Z. Zheng, E. Geva and B. D. Dunietz, *Org. Electron.*, 2014, **15**, 1509–1520.
137 S. Er, C. Suh, M. P. Marshak and A. Aspuru-Guzik, *Chem. Sci.*, 2015, **6**, 885–893.
138 M. Korth, *Phys. Chem. Chem. Phys.*, 2014, **16**, 7919–7926.
139 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.
140 M. Korth and W. Thiel, *J. Chem. Theory Comput.*, 2011, **7**, 2929–2936.
141 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
142 A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829–5835.
143 R. M. Richard, M. S. Marshall, O. Dolgounitcheva, J. V. Ortiz, J. L. Brédas, N. Marom and C. D. Sherrill, *J. Chem. Theory Comput.*, 2016, **12**, 595–604.
144 T. J. Van Huis, J. M. Galbraith and H. F. Schaefer, *Mol. Phys.*, 1996, **89**, 607–631.
145 X. Qu, A. Jain, N. N. Rajput, L. Cheng, Y. Zhang, S. P. Ong, M. Brafman, E. Maginn, L. A. Curtiss and K. A. Persson, *Comput. Mater. Sci.*, 2015, **103**, 56–67.
146 J. Tomasi, B. Mennucci and E. Cancès, *THEOCHEM*, 1999, **464**, 211–226.
147 K. Wedege, E. Dražević, D. Konya and A. Bentien, *Sci. Rep.*, 2016, **6**, 39101.
148 J. B. Henry and A. R. Mount, *J. Phys. Chem. A*, 2009, **113**, 13023–13028.
149 H. Neugebauer, F. Bohle, M. Bursch, A. Hansen and S. Grimme, *J. Phys. Chem. A*, 2020, **124**, 7166–7176.
150 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.
151 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
152 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.
153 J. J. P. Stewart, *J. Mol. Model.*, 2013, **19**, 1–32.
154 H. G. Roth, N. A. Romero and D. A. Nicewicz, *Synlett*, 2016, 714–723.
155 M. H. Baik and R. A. Friesner, *J. Phys. Chem. A*, 2002, **106**, 7407–7412.
156 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
157 M. Isegawa, F. Neese and D. A. Pantazis, *J. Chem. Theory Comput.*, 2016, **12**, 2272–2284.
158 C. M. Sterling and R. Bjornsson, *J. Chem. Theory Comput.*, 2019, **15**, 52–67.
159 A. R. McNeill, S. E. Bodman, A. M. Burney, C. D. Hughes and D. L. Crittenden, *J. Phys. Chem. C*, 2020, **124**, 24105–24114.
160 C. F. Wang, W. Y. Hung, M. H. Cheng, J. S. Hwang, M. K. Leung and K. T. Wong, *Org. Electron.*, 2013, **14**, 1958–1965.

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

161 S. Sohn, M. J. Kim, S. Jung, T. J. Shin, H. K. Lee and Y. H. Kim, *Org. Electron.*, 2015, **24**, 234–240.
162 M. A. Baldo, D. F. O'Brien, Y. You, A. Shoustikov, S. Sibley, M. E. Thompson and S. R. Forrest, *Nature*, 1998, **395**, 151–154.
163 S. Reineke, F. Lindner, G. Schwartz, N. Seidler, K. Walzer, B. Lüssem and K. Leo, *Nature*, 2009, **459**, 234–238.
164 M. Pope and C. E. Swenberg, *Electronic processes in organic crystals and polymers*, 1999.
165 M. B. Smith and J. Michl, *Chem. Rev.*, 2010, **110**, 6891–6936.
166 M. B. Smith and J. Michl, *Annu. Rev. Phys. Chem.*, 2013, **64**, 361–386.
167 D. Casanova, *Chem. Rev.*, 2018, **118**, 7164–7207.
168 F. Perrin, *Ann. Phys.*, 1929, **10**, 169–275.
169 G. N. Lewis and D. Lipkin, *J. Am. Chem. Soc.*, 1941, **63**, 3005–3018.
170 F. B. Dias, T. J. Penfold and A. P. Monkman, *Methods Appl. Fluoresc.*, 2017, **5**, 012001.
171 Z. Yang, Z. Mao, Z. Xie, Y. Zhang, S. Liu, J. Zhao, J. Xu, Z. Chi and M. P. Aldred, *Chem. Soc. Rev.*, 2017, **46**, 915–1016.
172 N. J. Turro, V. Ramamurthy and J. C. Scaiano, *Photochem. Photobiol.*, 2012, **88**, 1033.
173 S. Reineke and M. A. Baldo, *Sci. Rep.*, 2014, **4**, 3797.
174 M. Adachi and S. Nakamura, *Dyes Pigm.*, 1991, **17**, 287–296.
175 A. Matsuura, H. Sato, W. Sotoyama, A. Takahashi and M. Sakurai, *THEOCHEM*, 2008, **860**, 119–127.
176 D. Guillaumont and S. Nakamura, *Dyes Pigm.*, 2000, **46**, 85–92.
177 J. Fabian, L. A. Diaz, G. Seifert and T. Niehaus, *THEOCHEM*, 2002, **594**, 41–53.
178 D. Jacquemin, E. A. Perpète, G. E. Scuseria, I. Ciofini and C. Adamo, *J. Chem. Theory Comput.*, 2008, **4**, 123–135.
179 D. Jacquemin, E. A. Perpte, O. A. Vydrov, G. E. Scuseria and C. Adamo, *J. Chem. Phys.*, 2007, **127**, 094102.
180 D. Jacquemin, E. A. Perpète, G. E. Scuseria, I. Ciofini and C. Adamo, *Chem. Phys. Lett.*, 2008, **465**, 226–229.
181 M. J. G. Peach, P. Benfield, T. Helgaker and D. J. Tozer, *J. Chem. Phys.*, 2008, **128**, 044118.
182 M. Dierksen and S. Grimme, *J. Chem. Phys.*, 2004, **120**, 3544–3554.
183 V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129–12137.
184 D. A. Freedman, *Statistical models: Theory and practice*, 2009.
185 R. Grotjahn, T. M. Maier, J. Michl and M. Kaupp, *J. Chem. Theory Comput.*, 2017, **13**, 4984–4996.
186 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.
187 M. J. G. Peach, M. J. Williamson and D. J. Tozer, *J. Chem. Theory Comput.*, 2011, **7**, 3578–3585.
188 J. Wen, Z. Havlas and J. Michl, *J. Am. Chem. Soc.*, 2015, **137**, 165–172.
189 V. Veryazov, P. Å. Malmqvist and B. O. Roos, *Int. J. Quantum Chem.*, 2011, **111**, 3329–3338.
190 J. F. Joung, M. Han, M. Jeong and S. Park, *Sci. Data*, 2020, **7**, 295.
191 T. M. Figueira-Duarte and K. Müllen, *Chem. Rev.*, 2011, **111**, 7260–7314.
192 R. L. Vekariya, J. V. Vaghasiya and A. Dhar, *Org. Electron.*, 2017, **48**, 291–297.
193 J. Wakabayashi, M. Gon, K. Tanaka and Y. Chujo, *Macromolecules*, 2020, **53**, 4524–4532.
194 M. H. Tsai, T. H. Ke, H. W. Lin, C. C. Wu, S. F. Chiu, F. C. Fang, Y. L. Liao, K. T. Wong, Y. H. Chen and C. I. Wu, *ACS Appl. Mater. Interfaces*, 2009, **1**, 567–574.
195 H. Tanaka, K. Shizu, H. Nakanotani and C. Adachi, *Chem. Mater.*, 2013, **25**, 3766–3771.
196 S. Huang, Q. Zhang, Y. Shiota, T. Nakagawa, K. Kuwabara, K. Yoshizawa and C. Adachi, *J. Chem. Theory Comput.*, 2013, **9**, 3872–3877.
197 Q. Zhang, J. Li, K. Shizu, S. Huang, S. Hirata, H. Miyazaki and C. Adachi, *J. Am. Chem. Soc.*, 2012, **134**, 14706–14709.
198 H. Tanaka, K. Shizu, H. Nakanotani and C. Adachi, *J. Phys. Chem. C*, 2014, **118**, 15985–15994.
199 Y. Sagara, K. Shizu, H. Tanaka, H. Miyazaki, K. Goushi, H. Kaji and C. Adachi, *Chem. Lett.*, 2015, **44**, 360–362.
200 T. Takahashi, K. Shizu, T. Yasuda, K. Togashi and C. Adachi, *Sci. Technol. Adv. Mater.*, 2014, **15**, 034202.
201 J. Li, Q. Zhang, H. Nomura, H. Miyazaki and C. Adachi, *Appl. Phys. Lett.*, 2014, **105**, 013301.
202 S. Youn Lee, T. Yasuda, H. Nomura and C. Adachi, *Appl. Phys. Lett.*, 2012, **101**, 093306.
203 T. Serevičius, T. Nakagawa, M. C. Kuo, S. H. Cheng, K. T. Wong, C. H. Chang, R. C. Kwong, S. Xia and C. Adachi, *Phys. Chem. Chem. Phys.*, 2013, **15**, 15850–15855.
204 M. Numata, T. Yasuda and C. Adachi, *Chem. Commun.*, 2015, **51**, 9443–9446.
205 J. Lee, K. Shizu, H. Tanaka, H. Nakanotani, T. Yasuda, H. Kaji and C. Adachi, *J. Mater. Chem. C*, 2015, **3**, 2175–2181.
206 S. Wu, M. Aonuma, Q. Zhang, S. Huang, T. Nakagawa, K. Kuwabara and C. Adachi, *J. Mater. Chem. C*, 2014, **2**, 421–424.
207 J. Lee, K. Shizu, H. Tanaka, H. Nomura, T. Yasuda and C. Adachi, *J. Mater. Chem. C*, 2013, **1**, 4599–4604.
208 K. Nasu, T. Nakagawa, H. Nomura, C. J. Lin, C. H. Cheng, M. R. Tseng, T. Yasuda and C. Adachi, *Chem. Commun.*, 2013, **49**, 10385–10387.
209 Q. Zhang, B. Li, S. Huang, H. Nomura, H. Tanaka and C. Adachi, *Nat. Photonics*, 2014, **8**, 326–332.
210 R. Ishimatsu, S. Matsunami, T. Kasahara, J. Mizuno, T. Edura, C. Adachi, K. Nakano and T. Imato, *Angew. Chem., Int. Ed.*, 2014, **126**, 7113–7116.
211 J. Ridley and M. Zerner, *Theor. Chim. Acta*, 1973, **32**, 111–134.
212 M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer and W. Thiel, *J. Chem. Phys.*, 2008, **128**, 134110.
213 M. Véril, A. Scemama, M. Caffarel, F. Lipparini, M. Boggio-Pasqua, D. Jacquemin and P. F. Loos, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, e1517.

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

214 Š. Budzák, G. Scalmani and D. Jacquemin, *J. Chem. Theory Comput.*, 2017, **13**, 6237–6252.
215 N. O. C. Winter, N. K. Graf, S. Leutwyler and C. Hättig, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6623–6630.
216 S. Grimme and E. I. Izgorodina, *Chem. Phys.*, 2004, **305**, 223–230.
217 P. F. Loos, A. Scemama, M. Boggio-Pasqua and D. Jacquemin, *J. Chem. Theory Comput.*, 2020, **16**, 3720–3736.
218 P. F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel and D. Jacquemin, *J. Chem. Theory Comput.*, 2018, **14**, 4360–4379.
219 P. F. Loos, F. Lipparini, M. Boggio-Pasqua, A. Scemama and D. Jacquemin, *J. Chem. Theory Comput.*, 2020, **16**, 1711–1741.
220 B. G. Abreha, S. Agarwal, I. Foster, B. Blaiszik and S. A. Lopez, *J. Phys. Chem. Lett.*, 2019, **10**, 6835–6841.
221 M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265–3269.
222 W. J. Hehre, K. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
223 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
224 L. C. Blum and J. L. Reymond, *J. Am. Chem. Soc.*, 2009, **131**, 8732–8733.
225 G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K. R. Müller and O. A. Von Lilienfeld, *New J. Phys.*, 2013, **15**, 095003.
226 L. Ruddigkeit, R. Van Deursen, L. C. Blum and J. L. Reymond, *J. Chem. Inf. Model.*, 2012, **52**, 2864–2875.
227 R. Ramakrishnan, M. Hartmann, E. Tapavicza and O. A. Von Lilienfeld, *J. Chem. Phys.*, 2015, **143**, 08411.
228 J. Liang, S. Ye, T. Dai, Z. Zha, Y. Gao and X. Zhu, *Sci. Data*, 2020, **7**, 400.
229 M. Caricato, B. Mennucci and J. Tomasi, *J. Phys. Chem. A*, 2004, **108**, 6248–6256.
230 J. Fabian, *Theor. Chem. Acc.*, 2001, **106**, 199–217.
231 D. Jacquemin and E. A. Perpète, *Chem. Phys. Lett.*, 2006, **429**, 147–152.
232 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.
233 P. Winget, A. H. C. Horn, C. Selçuki, B. Martin and T. Clark, *J. Mol. Model.*, 2003, **9**, 408–414.
234 P. Winget and T. Clark, *J. Mol. Model.*, 2005, **11**, 439–456.
235 H. Kayi and T. Clark, *J. Mol. Model.*, 2007, **13**, 965–979.
236 W. Thiel, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 145–157.
237 J. Gierschner, S. Varghese and S. Y. Park, *Adv. Opt. Mater.*, 2016, **4**, 348–364.
238 J. Gierschner, J. Shi, B. Milián-Medina, D. Roca-Sanjuán, S. Varghese and S. Y. Park, *Adv. Opt. Mater.*, 2021, **9**, 2002251.
239 T. Nematiaram, D. Padula and A. Troisi, *Chem. Mater.*, 2021, **33**, 3368–3378.
240 F. C. Spano, L. Silvestri, P. Spearman, L. Raimondo and S. Tavazzi, *J. Chem. Phys.*, 2007, **127**, 184703.
241 F. C. Spano and H. Yamagata, *J. Phys. Chem. B*, 2011, **115**, 5133–5143.
242 L. Silvestri, S. Tavazzi, P. Spearman, L. Raimondo and F. C. Spano, *J. Chem. Phys.*, 2009, **130**, 234701.
243 F. Suzuki, S. Kubo, T. Fukushima and H. Kaji, *Sci. Rep.*, 2018, **8**, 5203.
244 D. Alberga, A. Perrier, I. Ciofini, G. F. Mangiatordi, G. Lattanzi and C. Adamo, *Phys. Chem. Chem. Phys.*, 2015, **17**, 18742–18750.
245 H. H. Choi, K. Cho, C. D. Frisbie, H. Sirringhaus and V. Podzorov, *Nat. Mater.*, 2017, **17**, 2–7.
246 O. D. Jurchescu, J. Baas and T. T. M. Palstra, *Appl. Phys. Lett.*, 2004, **84**, 3061–3063.
247 G. Horowitz, M. E. Hajlaoui and R. Hajlaoui, *J. Appl. Phys.*, 2000, **87**, 4456–4463.
248 V. Podzorov, E. Menard, J. A. Rogers and M. E. Gershenson, *Phys. Rev. Lett.*, 2005, **95**, 226601.
249 N. A. Minder, S. Ono, Z. Chen, A. Facchetti and A. F. Morpurgo, *Adv. Mater.*, 2012, **24**, 503–508.
250 G. Schweicher, G. Garbay, R. Jouclas, F. Vibert, F. Devaux and Y. H. Geerts, *Adv. Mater.*, 2020, **32**, 1905909.
251 H. Oberhofer, K. Reuter and J. Blumberger, *Chem. Rev.*, 2017, **117**, 10319–10357.
252 S. Fratini, D. Mayou and S. Ciuchi, *Adv. Funct. Mater.*, 2016, **26**, 2292–2315.
253 I. Yavuz, *Phys. Chem. Chem. Phys.*, 2017, **19**, 25819–25828.
254 T. Holstein, *Ann. Phys.*, 1959, **8**, 343–389.
255 T. Holstein, *Ann. Phys.*, 2000, **281**, 725–773.
256 J. B. Delos and W. R. Thorson, *J. Chem. Phys.*, 1979, **70**, 1774–1790.
257 G. D. Billing, *Int. Rev. Phys. Chem.*, 1994, **13**, 309–335.
258 O. V. Prezhdo and P. J. Rossky, *J. Chem. Phys.*, 1997, **107**, 5863–5878.
259 J. Spencer, F. Gajdos and J. Blumberger, *J. Chem. Phys.*, 2016, **145**, 064102.
260 T. Nematiaram, A. Asgari and D. Mayou, *J. Chem. Phys.*, 2020, **152**, 044109.
261 M. Lian, Y. C. Wang, Y. Ke and Y. Zhao, *J. Chem. Phys.*, 2019, **151**, 044115.
262 A. S. Mishchenko, N. Nagaosa, G. De Filippis, A. De Candia and V. Cataudella, *Phys. Rev. Lett.*, 2015, **114**, 146401.
263 G. De Filippis, V. Cataudella, A. S. Mishchenko, N. Nagaosa, A. Fierro and A. De Candia, *Phys. Rev. Lett.*, 2015, **114**, 086601.
264 S. Ciuchi, S. Fratini and D. Mayou, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 081202.
265 S. Fratini, S. Ciuchi, D. Mayou, G. T. De Laissardière and A. Troisi, *Nat. Mater.*, 2017, **16**, 998–1002.
266 T. Nematiaram, S. Ciuchi, X. Xie, S. Fratini and A. Troisi, *J. Phys. Chem. C*, 2019, **123**, 6989–6997.
267 S. Giannini, A. Carof, M. Ellis, H. Yang, O. G. Ziogos, S. Ghosh and J. Blumberger, *Nat. Commun.*, 2019, **10**, 3843.
268 N. E. Lee, J. J. Zhou, L. A. Agapito and M. Bernardi, *Phys. Rev. B*, 2018, **97**, 115203.
269 C. Schober, K. Reuter and H. Oberhofer, *J. Phys. Chem. Lett.*, 2016, **7**, 3973–3977.
270 N. N. Matsuzawa, H. Arai, M. Sasago, E. Fujii, A. Goldberg, T. J. Mustard, H. S. Kwak, D. J. Giesen, F. Ranalli and M. D. Halls, *J. Phys. Chem. A*, 2020, **124**, 1981–1992.

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

271 C. Kunkel, C. Schober, J. T. Margraf, K. Reuter and H. Oberhofer, *Chem. Mater.*, 2019, **31**, 969–978.
272 A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C. B. Mannsfeld, A. P. Zoombelt, Z. Bao and A. Aspuru-Guzik, *Nat. Commun.*, 2011, **2**, 437.
273 B. Özen, F. Fadaei Tirani, K. Schenk, K. H. Lin, R. Scopelliti, C. Corminboeuf and H. Frauenrath, *Chem. – Eur. J.*, 2021, **27**, 3348–3360.
274 P. Hu, J. Ye and H. Jiang, *J. Mater. Chem. C*, 2019, **7**, 5858–5873.
275 Y. Zhu, X. Zeng, T. Fu, Y. Cao and H. Meng, *Phys. Chem. Chem. Phys.*, 2019, **21**, 4641–4649.
276 C. P. Hsu, *Acc. Chem. Res.*, 2009, **42**, 509–518.
277 H. Ding, C. Reese, A. J. Mäkinen, Z. Bao and Y. Gao, *Appl. Phys. Lett.*, 2010, **96**, 222106.
278 Y. Nakayama, Y. Uragami, S. Machida, K. R. Koswattage, D. Yoshimura, H. Setoyama, T. Okajima, K. Mase and H. Ishii, *Appl. Phys. Express*, 2012, **5**, 111601.
279 J. Nitta, K. Miwa, N. Komiya, E. Annese, J. Fujii, S. Ono and K. Sakamoto, *Sci. Rep.*, 2019, **9**, 9645.
280 A. Brillante, I. Bilotti, R. G. Della Valle, E. Venuti and A. Girlando, *CrystEngComm*, 2008, **10**, 937–946.
281 M. Abdulla, K. Refson, R. H. Friend and P. D. Haynes, *J. Phys.: Condens. Matter*, 2015, **27**, 375402.
282 D. Di Nuzzo, C. Fontanesi, R. Jones, S. Allard, I. Dumsch, U. Scherf, E. Von Hauff, S. Schumacher and E. Da Como, *Nat. Commun.*, 2015, **6**, 6460.
283 P. Friederich, V. Meded, A. Poschlad, T. Neumann, V. Rodin, V. Stehr, F. Symalla, D. Danilov, G. Lüdemann, R. F. Fink, I. Kondov, F. von Wrochem and W. Wenzel, *Adv. Funct. Mater.*, 2016, **26**, 5757–5763.
284 Y. Yi, V. Coropceanu and J. L. Brédas, *J. Chem. Phys.*, 2012, **137**, 164303.
285 X. Xie, A. Santana-Bonilla and A. Troisi, *J. Chem. Theory Comput.*, 2018, **14**, 3752–3762.
286 N. Bedoya-Martínez, A. Giunchi, T. Salzillo, E. Venuti, R. G. Della Valle and E. Zojer, *J. Chem. Theory Comput.*, 2018, **14**, 4380–4390.
287 V. Dantanarayana, T. Nematiaram, D. Vong, J. E. Anthony, A. Troisi, K. Nguyen Cong, N. Goldman, R. Faller and A. J. Moulé, *J. Chem. Theory Comput.*, 2020, **16**, 3494–3503.
288 T. F. Harrelson, V. Dantanarayana, X. Xie, C. Koshnick, D. Nai, R. Fair, S. A. Nuñez, A. K. Thomas, T. L. Murrey, M. A. Hickner, J. K. Grey, J. E. Anthony, E. D. Gomez, A. Troisi, R. Faller and A. J. Moulé, *Mater. Horiz.*, 2019, **6**, 182–191.
289 J. Nyman, O. S. Pundyke and G. M. Day, *Phys. Chem. Chem. Phys.*, 2016, **18**, 15828–15837.
290 G. Schweicher, G. D'Avino, M. T. Ruggiero, D. J. Harkin, K. Broch, D. Venkateshvaran, G. Liu, A. Richard, C. Ruzié, J. Armstrong, A. R. Kennedy, K. Shankland, K. Takimiya, Y. H. Geerts, J. A. Zeitler, S. Fratini and H. Sirringhaus, *Adv. Mater.*, 2019, **31**, 1902407.
291 A. Landi, A. Peluso and A. Troisi, *Adv. Mater.*, 2021, **33**, 2008049.
292 M. Fahlman, S. Fabiano, V. Gueskine, D. Simon, M. Berggren and X. Crispin, *Nat. Rev. Mater.*, 2019, **4**, 627–650.
293 D. K. Dubey, D. Thakur, R. A. K. Yadav, M. Ram Nagar, T. W. Liang, S. Ghosh and J. H. Jou, *ACS Appl. Mater. Interfaces*, 2021, **13**, 26204–26217.
294 S. Marina, N. P. Kaufmann, A. Karki, E. Gutiérrez-Meza, E. Gutiérrez-Fernández, J. Vollbrecht, E. Solano, B. Walker, J. H. Bannock, J. de Mello, C. Silva, T. Q. Nguyen, D. Cangialosi, N. Stingelin and J. Martín, *Adv. Mater.*, 2020, **32**, 2005241.
295 Z. W. Zhao, Ö. H. Omar, D. Padula, Y. Geng and A. Troisi, *J. Phys. Chem. Lett.*, 2021, **12**, 5009–5015.
296 D. P. McMahon, D. L. Cheung and A. Troisi, *J. Phys. Chem. Lett.*, 2011, **2**, 2737–2741.
297 G. D'Avino, S. Mothy, L. Muccioli, C. Zannoni, L. Wang, J. Cornil, D. Beljonne and F. Castet, *J. Phys. Chem. C*, 2013, **117**, 12981–12990.
298 T. Koch, J. Bachmann, T. Lettmann and N. L. Doltsinis, *Phys. Chem. Chem. Phys.*, 2021, **23**, 12233–12250.
299 Y. Imamura, M. Suganuma and M. Hada, *J. Phys. Chem. C*, 2019, **123**, 17678–17685.
300 C. Poelking, M. Tietze, C. Elschner, S. Olthof, D. Hertel, B. Baumeier, F. Würthner, K. Meerholz, K. Leo and D. Andrienko, *Nat. Mater.*, 2015, **14**, 434–439.
301 N. Martsinovich and A. Troisi, *J. Phys. Chem. C*, 2011, **115**, 11781–11792.
302 F. Ambrosio, N. Martsinovich and A. Troisi, *J. Phys. Chem. Lett.*, 2012, **3**, 1531–1535.
303 A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, *Nat. Rev. Mater.*, 2019, **4**, 331–348.
304 D. H. Bowskill, I. J. Sugden, S. Konstantinopoulos, C. S. Adjiman and C. C. Pantelides, *Annu. Rev. Chem. Biomol. Eng.*, 2021, **12**, 593–623.
305 G. M. Day, *Crystallogr. Rev.*, 2011, **17**, 3–52.
306 J. Nyman and S. M. Reutzel-Edens, *Faraday Discuss.*, 2018, **211**, 459–476.
307 J. Yang, S. De, J. E. Campbell, S. Li, M. Ceriotti and G. M. Day, *Chem. Mater.*, 2018, **30**, 4361–4371.
308 C. Y. Cheng, J. E. Campbell and G. M. Day, *Chem. Sci.*, 2020, **11**, 4922–4933.
309 A. Vriza, A. B. Canaj, R. Vismara, L. J. Kershaw Cook, T. D. Manning, M. W. Gaultois, P. A. Wood, V. Kurlin, N. Berry, M. S. Dyer and M. J. Rosseinsky, *Chem. Sci.*, 2021, **12**, 1702–1719.
310 T. Sterling and J. J. Irwin, *J. Chem. Inf. Model.*, 2015, **55**, 2324–2337.
311 O. Gürsoy and M. Smieško, *J. Cheminf.*, 2017, **9**, 29.
312 G. Poli, T. Seidel and T. Langer, *Front. Chem.*, 2018, **6**, 229.
313 H. Wen, X. Gong, P. Han, B. Lin, L. Zhang, S. Ye, Y. Sun, X. Zhang and H. Yang, *RSC Adv.*, 2016, **6**, 69277–69281.
314 F. Franco, *Mol. Simul.*, 2017, **43**, 222–227.
315 J. T. Blaskovits, K.-H. Lin, R. Fabregat, I. Swiderska, H. Wu and C. Corminboeuf, *ChemRxiv*, 2021, This content is a preprint and has not been peer-reviewed.
316 C. J. Ou, C. Zhu, X. H. Ding, L. Yang, J. Y. Lin, L. H. Xie, Y. Qian, C. X. Xu, J. F. Zhao and W. Huang, *J. Mater. Chem. C*, 2017, **5**, 5345–5355.

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

317 M. Urban, P. Górka, K. Nawara, K. Woźniak, K. Durka and S. Luliński, *Dalton Trans.*, 2018, **47**, 15670–15684.
318 S. Monti, F. Cicogna, E. Passaglia, G. Prampolini and V. Barone, *Phys. Chem. Chem. Phys.*, 2011, **13**, 21471–21478.
319 I. Y. Kanal, J. A. Keith and G. R. Hutchison, *Int. J. Quantum Chem.*, 2018, **118**, 25512.
320 R. Noriega, J. Rivnay, K. Vandewal, F. P. V. Koch, N. Stingelin, P. Smith, M. F. Toney and A. Salleo, *Nat. Mater.*, 2013, **12**, 1038–1044.
321 N. E. Jackson, B. M. Savoie, K. L. Kohlstedt, M. Olvera De La Cruz, G. C. Schatz, L. X. Chen and M. A. Ratner, *J. Am. Chem. Soc.*, 2013, **135**, 10475–10483.
322 R. Manurung, P. Li and A. Troisi, *J. Phys. Chem. B*, 2021, **125**, 6338–6348.
323 P. O. Dral, *J. Phys. Chem. Lett.*, 2020, **11**, 2336–2347.
324 O. A. von Lilienfeld, *Angew. Chem., Int. Ed.*, 2018, **57**, 4164–4169.
325 M. Rupp, A. Tkatchenko, K. R. Müller and O. A. Von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.
326 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.
327 F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke and K. R. Müller, *Nat. Commun.*, 2017, **8**, 872.
328 Z. D. Pozun, K. Hansen, D. Sheppard, M. Rupp, K. R. Müller and G. Henkelman, *J. Chem. Phys.*, 2012, **136**, 174101.
329 Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 096405.
330 L. Ward, B. Blaiszik, I. Foster, R. S. Assary, B. Narayanan and L. Curtiss, *MRS Commun.*, 2019, **9**, 891–899.
331 J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway and A. Aspuru-Guzik, *J. Phys. Chem. Lett.*, 2011, **2**, 2241–2251.
332 E. O. Pyzer-Knapp, K. Li and A. Aspuru-Guzik, *Adv. Funct. Mater.*, 2015, **25**, 6495–6502.
333 F. Jabeen, M. Chen, B. Rasulev, M. Ossowski and P. Boudjouk, *Comput. Mater. Sci.*, 2017, **137**, 215–224.
334 C. Nantasenamat, K. Srungboonmee, S. Jamsak, N. Tansila, C. Isarankura-Na-Ayudhya and V. Prachayasittikul, *Chemom. Intell. Lab. Syst.*, 2013, **120**, 42–52.
335 Z. R. Ye, I. S. Huang, Y. Te Chan, Z. J. Li, C. C. Liao, H. R. Tsai, M. C. Hsieh, C. C. Chang and M. K. Tsai, *RSC Adv.*, 2020, **10**, 23834–23841.
336 J. Westermayr and P. Marquetand, *Chem. Rev.*, 2021, **121**, 9873–9926.
337 P. O. Dral and M. Barbatti, *Nat. Rev. Chem.*, 2021, **5**, 388–405.
338 M. Sugawara, *Comput. Phys. Commun.*, 2001, **140**, 366–380.
339 K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller and R. J. Maurer, *Nat. Commun.*, 2019, **10**, 5024.
340 M. Babaei, Y. T. Azar and A. Sadeghi, *Phys. Rev. B*, 2020, **101**, 115132.
341 S. K. Kauwe, J. Graser, R. Murdock and T. D. Sparks, *Comput. Mater. Sci.*, 2019, **174**, 109498.
342 S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko and D. Morgan, *Comput. Mater. Sci.*, 2012, **58**, 218–226.
343 J. P. Janet, C. Duan, T. Yang, A. Nandy and H. J. Kulik, *Chem. Sci.*, 2019, **10**, 7913–7922.
344 T. Fink, H. Bruggesser and J.-L. Reymond, *Angew. Chem., Int. Ed.*, 2005, **117**, 1528–1532.
345 J. L. Reymond, R. Van Deursen, L. C. Blum and L. Ruddigkeit, *MedChemComm*, 2010, **1**, 30–38.
346 A. Supady, V. Blum and C. Baldauf, *J. Chem. Inf. Model.*, 2015, **55**, 2338–2348.
347 Y. Wang, W. Hao, W. Huang, H. Zhao, J. Zhu and W. Fang, *J. Phys. Chem. Lett.*, 2020, **11**, 3928–3933.
348 L. Wilbraham, D. Smajli, I. Heath-Apostolopoulos and M. A. Zwijnenburg, *Commun. Chem.*, 2020, **3**, 14.
349 M. A. F. Afzal, M. Haghighatlari, S. P. Ganesh, C. Cheng and J. Hachmann, *J. Phys. Chem. C*, 2019, **123**, 14610–14618.
350 Y. Imamura, M. Tashiro, M. Katouda and M. Hada, *J. Phys. Chem. C*, 2017, **121**, 28275–28286.
351 C. Zanlorenzi and L. Akcelrud, *J. Polym. Sci., Part B: Polym. Phys.*, 2017, **55**, 919–927.
352 P. Li, H. Zhang and A. Troisi, *J. Phys. Chem. C*, 2018, **122**, 23890–23898.
353 M. Moral, A. Garzón-Ruiz, M. Castro, J. Canales-Vázquez and J. C. Sancho-García, *J. Phys. Chem. C*, 2017, **121**, 28249–28261.
354 T. A. Gschneidtner, Y. A. Diaz Fernandez and K. Moth-Poulsen, *J. Mater. Chem. C*, 2013, **1**, 7127–7133.
355 N. Xin, J. Guan, C. Zhou, X. Chen, C. Gu, Y. Li, M. A. Ratner, A. Nitzan, J. F. Stoddart and X. Guo, *Nat. Rev. Phys.*, 2019, **1**, 211–230.
356 S. Ghasemi and K. Moth-Poulsen, *Nanoscale*, 2021, **13**, 659–671.
357 W. Gao and C. W. Coley, *J. Chem. Inf. Model.*, 2020, **60**, 5714–5723.
358 C. W. Coley, L. Rogers, W. H. Green and K. F. Jensen, *J. Chem. Inf. Model.*, 2018, **58**, 252–261.
359 F. T. Szczypiński, S. Bennett and K. E. Jelfs, *Chem. Sci.*, 2021, **12**, 830–840.
360 P. Ertl and A. Schuffenhauer, *J. Cheminf.*, 2009, **1**, 8.
361 K. B. Ørnsø, J. M. Garcia-Lastra and K. S. Thygesen, *Phys. Chem. Chem. Phys.*, 2013, **15**, 19478–19486.
362 J. T. Blaskovits, M. Fumanal, S. Vela and C. Corminboeuf, *Chem. Mater.*, 2020, **32**, 6515–6524.
363 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *Sci. Data*, 2014, **1**, 140022.
364 J. S. Smith, O. Isayev and A. E. Roitberg, *Sci. Data*, 2017, **4**, 170193.
365 J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio and A. Tkatchenko, *Sci. Data*, 2021, **8**, 43.
366 J. Mc Conathy and M. J. Owens, *Prim. Care Companion J. Clin. Psychiatry*, 2003, **05**, 70–73.
367 K. Zhao, Ö. H. Omar, T. Nematiaram, D. Padula and A. Troisi, *J. Mater. Chem. C*, 2021, **9**, 3324–3333.
368 A. Stuke, C. Kunkel, D. Golze, M. Todorović, J. T. Margraf, K. Reuter, P. Rinke and H. Oberhofer, *Sci. Data*, 2020, **7**, 58.

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

369 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nucleic Acids Res.*, 2021, **49**, D1388–D1395.
370 University of Cambridge, Chemistry Library, https://www-library.ch.cam.ac.uk/list-useful-databases, (accessed July 2021).
371 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
372 D. Weininger, A. Weininger and J. L. Weininger, *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 97–101.
373 P. M. Tagade, S. P. Adiga, M. S. Park, S. Pandian, K. S. Hariharan and S. M. Kolake, *J. Phys. Chem. C*, 2018, **122**, 11322–11333.
374 P. C. St. John, Y. Guan, Y. Kim, B. D. Etz, S. Kim and R. S. Paton, *Sci. Data*, 2020, **7**, 244.
375 O. Borodin, M. Olguin, C. E. Spear, K. W. Leiter and J. Knap, *Nanotechnology*, 2015, **26**, 35.
376 M. Nakata and T. Shimazaki, *J. Chem. Inf. Model.*, 2017, **57**, 1300–1308.
377 S. R. Heller, A. McNaught, I. Pletnev, S. Stein and D. Tchekhovskoi, *J. Cheminf.*, 2015, **7**, 23.
378 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.
379 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209–220.
380 J. J. P. Stewart, *J. Comput. Chem.*, 1991, **12**, 320–341.
381 M. Ranzato, J. Susskind, V. Mnih and G. Hinton, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011, pp. 2857–2864.
382 L. C. Yang and A. Lerch, *Neural Comput. Appl.*, 2020, **32**, 4773–4784.
383 D. P. Kingma and M. Welling, in *2nd International Conference on Learning Representations, ICLR 2014 – Conference Track Proceedings*, 2014.
384 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, *Commun. ACM*, 2020, **63**, 139–144.
385 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360–365.
386 S. D'Souza, K. V. Prema and S. Balaji, *Drug Discovery Today*, 2020, **25**, 748–756.
387 L. David, A. Thakkar, R. Mercado and O. Engkvist, *J. Cheminf.*, 2020, **12**, 56.
388 K. Kim, S. Kang, J. Yoo, Y. Kwon, Y. Nam, D. Lee, I. Kim, Y. S. Choi, Y. Jung, S. Kim, W. J. Son, J. Son, H. S. Lee, S. Kim, J. Shin and S. Hwang, *npj Comput. Mater.*, 2018, **4**, 67.
389 S. P. Peng and Y. Zhao, *J. Chem. Inf. Model.*, 2019, **59**, 4993–5001.
390 Q. Yuan, A. Santana-Bonilla, M. A. Zwijnenburg and K. E. Jelfs, *Nanoscale*, 2020, **12**, 6744–6758.
391 D. Schwalbe-Koda and R. Gómez-Bombarelli, *Lecture Notes in Physics*, 2020, vol. 968, pp. 445–467.
392 M. Sumita, X. Yang, S. Ishihara, R. Tamura and K. Tsuda, *ACS Cent. Sci.*, 2018, **4**, 1126–1133.
393 C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis and S. Colton, *IEEE Trans. Comput. Intell. AI Games*, 2012, **4**, 1–43.
394 M. H. S. Segler, M. Preuss and M. P. Waller, *Nature*, 2018, **555**, 604–610.
395 I. A. Watson, J. Wang and C. A. Nicolaou, *J. Cheminf.*, 2019, **11**, 1.
396 C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. John Hart, T. F. Jamison and K. F. Jensen, *Science*, 2019, **365**, 557.
397 R. Shibukawa, S. Ishida, K. Yoshizoe, K. Wasa, K. Takasu, Y. Okuno, K. Terayama and K. Tsuda, *J. Cheminf.*, 2020, **12**, 52.
398 G. Klein, R. Voltz and M. Schott, *Chem. Phys. Lett.*, 1973, **19**, 391–394.
399 G. Klein, R. Voltz and M. Schott, *Chem. Phys. Lett.*, 1972, **16**, 340–344.
400 J. Ehrmaier, E. J. Rabe, S. R. Pristash, K. L. Corp, C. W. Schlenker, A. L. Sobolewski and W. Domcke, *J. Phys. Chem. A*, 2019, **123**, 8099–8108.
401 D. Miyajima, N. Aizawa, Y.-J. Pu, A. Nihonyanagi, R. Ibuka, H. Inuzuka, B. Dhara, Y. Koyama and F. Araoka, Research Square, 2021. This content is a preprint and has not been peer-reviewed.
402 T. Chen, L. Zheng, J. Yuan, Z. An, R. Chen, Y. Tao, H. Li, X. Xie and W. Huang, *Sci. Rep.*, 2015, **5**, 10923.
403 C. W. Coley, N. S. Eyke and K. F. Jensen, *Angew. Chem.*, 2020, **59**, 22858–22893.
404 T. Dimitrov, C. Kreisbeck, J. S. Becker, A. Aspuru-Guzik and S. K. Saikin, *ACS Appl. Mater. Interfaces*, 2019, **11**, 24825–24836.
405 B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick and A. I. Cooper, *Nature*, 2020, **583**, 237–241.
406 M. Nakano, *Chem. Rec.*, 2017, **17**, 27–62.
407 T. Minami and M. Nakano, *J. Phys. Chem. Lett.*, 2012, **3**, 145–150.
408 Ö. H. Omar, D. Padula and A. Troisi, *ChemPhotoChem*, 2020, **4**, 5223–5229.
409 C. M. Ip and A. Troisi, *J. Phys. Chem. Lett.*, 2016, **7**, 2989–2993.
410 A. Devos and M. Lannoo, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 8236–8239.
411 T. Nematiaram and A. Troisi, *Mater. Horiz.*, 2020, **7**, 2922–2928.
412 J. Benduhn, K. Tvingstedt, F. Piersimoni, S. Ullbrich, Y. Fan, M. Tropiano, K. A. McGarry, O. Zeika, M. K. Riede, C. J. Douglas, S. Barlow, S. R. Marder, D. Neher, D. Spoltore and K. Vandewal, *Nat. Energy*, 2017, **2**, 17053.
413 H. Bilgili, G. Yakall, N. Kaya, M. Karaman and Ş. Demiç, *J. Phys. Chem. C*, 2021, **125**, 4893–4908.
414 X. Yang, J. Zhang, K. Yoshizoe, K. Terayama and K. Tsuda, *Sci. Technol. Adv. Mater.*, 2017, **18**, 972–976.
415 M. Tashiro, Y. Imamura and M. Katouda, *J. Comput. Chem.*, 2021, **42**, 136–143.

 This journal is © The Royal Society of Chemistry 2021

Open Access Article. Published on 16 Setembar 2021. Downloaded on 20/08/2025 00:36:37.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

416 A. Kuzmich, D. Padula, H. Ma and A. Troisi, *Energy Environ. Sci.*, 2017, **10**, 395–401.
417 T. Liu and A. Troisi, *Adv. Funct. Mater.*, 2014, **24**, 925–933.
418 Y. Wan, F. Ramirez, X. Zhang, T. Q. Nguyen, G. C. Bazan and G. Lu, *npj Comput. Mater.*, 2021, **7**, 69.
419 S. Kraner, R. Scholz, F. Plasser, C. Koerner and K. Leo, *J. Chem. Phys.*, 2015, **143**, 244905.
420 T. N. Aram, A. Asgari, M. Ernzerhof, P. Quémerais and D. Mayou, *EPJ Photovoltaics*, 2017, **8**, 85503.
421 R. Xia, C. J. Brabec, H. L. Yip and Y. Cao, *Joule*, 2019, **3**, 2241–2254.
422 W. Groh and A. Zimmermann, *Macromolecules*, 1991, **24**, 6660–6663.
423 M. C. Scharber, D. Mühlbacher, M. Koppe, P. Denk, C. Waldauf, A. J. Heeger and C. J. Brabec, *Adv. Mater.*, 2006, **18**, 789–794.
424 D. Padula, J. D. Simpson and A. Troisi, *Mater. Horiz.*, 2019, **6**, 343–349.
425 H. Sahu, F. Yang, X. Ye, J. Ma, W. Fang and H. Ma, *J. Mater. Chem. A*, 2019, **7**, 17480–17488.
426 H. Sahu, W. Rao, A. Troisi and H. Ma, *Adv. Energy Mater.*, 2018, **8**, 1801032.
427 Z. W. Zhao, M. Del Cueto, Y. Geng and A. Troisi, *Chem. Mater.*, 2020, **32**, 7777–7787.
428 M. H. Lee, *Adv. Energy Mater.*, 2019, **9**, 1900891.
429 W. Sun, Y. Zheng, K. Yang, Q. Zhang, A. A. Shah, Z. Wu, Y. Sun, L. Feng, D. Chen, Z. Xiao, S. Lu, Y. Li and K. Sun, *Sci. Adv.*, 2019, **5**, 11.
430 J. Feng, H. Wang, Y. Ji and Y. Li, *Nano Sel.*, 2021, 1–13.
431 C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng and S. P. Ong, *Adv. Energy Mater.*, 2020, **10**, 1903242.
432 Y. Wu, J. Guo, R. Sun and J. Min, *npj Comput. Mater.*, 2020, **6**, 120.
433 P. B. Jørgensen, M. Mesta, S. Shil, J. M. García Lastra, K. W. Jacobsen, K. S. Thygesen and M. N. Schmidt, *J. Chem. Phys.*, 2018, **148**, 241735.
434 S. Nagasawa, E. Al-Naamani and A. Saeki, *J. Phys. Chem. Lett.*, 2018, **9**, 2639–2646.
435 H. Sahu and H. Ma, *J. Phys. Chem. Lett.*, 2019, **10**, 7277–7284.
436 J. Sieg, F. Flachsenberg and M. Rarey, *J. Chem. Inf. Model.*, 2019, **59**, 947–961.
437 P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
438 B. Cao, L. A. Adutwum, A. O. Oliynyk, E. J. Luber, B. C. Olsen, A. Mar and J. M. Buriak, *ACS Nano*, 2018, **12**, 7434–7444.
439 B. Kailkhura, B. Gallagher, S. Kim, A. Hiszpanski and T. Y. J. Han, *npj Comput. Mater.*, 2019, **5**, 108.
440 M. del Cueto and A. Troisi, *Phys. Chem. Chem. Phys.*, 2021, **23**, 14156–14163.
441 B. Meredig, E. Antono, C. Church, M. Hutchinson, J. Ling, S. Paradiso, B. Blaiszik, I. Foster, B. Gibbons, J. Hattrick-Simpers, A. Mehta and L. Ward, *Mol. Syst. Des. Eng.*, 2018, **3**, 819–825.
442 Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu and J. Hu, *Comput. Mater. Sci.*, 2020, **171**, 109203.
443 C. Kunkel, J. T. Margraf, K. Chen, H. Oberhofer and K. Reuter, *Nat. Commun.*, 2021, **12**, 2422.
444 Y. Huang, J. Zhang, E. S. Jiang, Y. Oya, A. Saeki, G. Kikugawa, T. Okabe, T. Okabe, F. S. Ohuchi and F. S. Ohuchi, *J. Phys. Chem. C*, 2020, **124**, 12871–12882.