

Journal of Materials Chemistry C

Materials for optical, magnetic and electronic devices

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: K. Thapa, J. I. Jones and L. E. Ratcliff, *J. Mater. Chem. C*, 2025, DOI: 10.1039/D5TC01938A.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Cite this: DOI: 00.0000/xxxxxxxxxx

A High Throughput Virtual Screening Approach for Identifying New Thermally Activated Delayed Fluorescence-Based Emitters[†]Kritam Thapa,^{a,‡} Jennifer I. Jones,^{b,‡} and Laura E. Ratcliff^{b,c*}Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Thermally activated delayed fluorescence (TADF) offers the promise of highly efficient organic light emitting diodes (OLEDs), without the heavy metals requirement of the previous generation of OLEDs. However, the design of new TADF emitters is complicated by competing requirements, which require opposing design strategies. High throughput virtual screening (HTVS) approaches, however, offer the possibility of identifying new TADF emitters without necessarily relying on existing design rules. In this work the STONED algorithm [A. Nigam *et al.*, *Chem. Sci.*, 2021, **12**, 7079] is used to impose random structural mutations starting from a set of twenty parent molecules, composed of both traditional donor-acceptor and multiresonant TADF emitters. Following this, successive filters are applied based on features of the atomic structure through to time-dependent density functional theory calculations. Although the randomised approach proves to be ill-suited to rediscovering existing TADF emitters, the resulting workflow leads to the identification of a number of molecules with promising properties for TADF, across a range of emission colours.

1 Introduction

Organic light emitting diodes (OLEDs) are an efficient and sustainable offering to the display and lighting markets, offering advantages over previous technologies like liquid crystal displays and traditional LEDs. OLEDs function via electroluminescence, meaning they do not require a backlight, making them energy efficient, thin and light, as well as allowing for “true black”. Furthermore, they have improved resolution and colour contrast, wider viewing angles, and faster response times^{1,2}, while also offering the potential for flexible and transparent displays^{1,3}. In an OLED, electrons (holes) are injected at the cathode (anode) and recombine to form excitons in the emissive layer, subsequently leading to the emission of a photon as the excited molecules return to their ground states^{1,4,5}. Due to spin statistics, singlet and triplet excitons form in a 1:3 ratio. The first generation of OLED emitters, namely fluorescent emitters, exploit only singlet exci-

tions for light emission, limiting the internal quantum efficiency (IQE) to 25%. To overcome this limitation, second-generation phosphorescent emitters arose, which are able to harvest both singlet and triplet excitons for light emission, leading to a theoretical IQE of 100%. This is enabled by strong spin-orbit coupling due to the incorporation of heavy metal elements, making the triplet state emissive^{5,6}. However, such materials are costly, potentially toxic, and unstable, particularly in the case of blue emitters⁷. The latest generation of emitters can also harvest all excitons radiatively, but are fully organic. These involve reverse intersystem crossing (RISC) from the first triplet (T_1) to the first singlet (S_1) excited state, a process activated by the input of thermal energy when the energy gap between S_1 and T_1 , known as ΔE_{ST} , is sufficiently small, generally on the order of 0.2 eV or smaller⁸. This results in both “prompt” and “delayed” fluorescence, and is hence known as thermally activated delayed fluorescence (TADF)^{5,9}.

Small ΔE_{ST} in TADF emitters is typically achieved via spatial separation of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO)⁴. The most common approach involves twisted combinations of donor (D) and acceptor (A) constituents in which the HOMO (LUMO) is strongly localised on the D (A). However, increased HOMO-LUMO separation also reduces the oscillator strength of the S_1 to S_0 transition (f_{S_1}), thereby reducing the electroluminescent efficiency¹⁰, leading to contradicting requirements of both small ΔE_{ST} and strong f_{S_1} ¹¹. Furthermore, D-A emitters have large excited state struc-

^a Department of Materials, Imperial College London, London SW7 2AZ, United Kingdom.^b Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom.^c Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway.

* Email: laura.ratcliff@bristol.ac.uk

‡ These authors contributed equally to this work.

† Supplementary Information available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.



tural relaxations, and in turn a broad Stokes shift and reduced colour purity¹⁰. In 2016 a new design strategy arose based on the multiple resonance (MR) effect¹². Here the frontier orbitals are instead localised on electron-donating and withdrawing atoms, as well as at their ortho and para positions in a fused polycyclic aromatic structure¹⁰. D-A and MR type emitters can be regarded as the two main classes of TADF emitters, though alternative design mechanisms also exist¹³. There are also various other active avenues for TADF emitter design, including metal-based complexes (see e.g. refs. 14–16), as well as strategies going beyond TADF, like “hot exciton” RISC, in which triplets are harvested *via* RISC from higher lying triplet states, T_n , where internal conversion to the lowest-lying triplet state is suppressed due to a large T_n - T_1 energy gap^{17–19}. TADF emitter design is therefore complex, and while the efficiencies of TADF OLED emitters have begun to compete with those of phosphorescent emitters²⁰, there is still a need for further developments in order to achieve high efficiency, stability and colour purity simultaneously.

One approach which is gaining traction for identifying new TADF emitters is high-throughput virtual screening (HTVS), whereby candidates are filtered out at successive stages based on computational chemistry-based predictions of properties like ΔE_{ST} and f_{S_1} . An example is the work of Gómez-Bombarelli *et al.*⁵, who used machine learning (ML) and time-dependent density functional theory (TDDFT)²¹ to investigate a library of 1.6 million molecules, resulting in a number of experimentally-verified OLED devices. Other examples include the work of Lin *et al.*²², who performed screening based on ionisation energies and electron affinities, which relate to trap-free electron transport in single-layer OLED devices, a recent HTVS study which focused on high-efficiency hosts for blue OLEDs²³, and the work of Zhao *et al.*¹³, who screened a database of known molecular materials for TADF properties. There are also many examples of using ML for property prediction based on structural features^{24–28}, which can further reduce the cost of HTVS workflows.

In this work, we introduce an HTVS approach for identifying potential TADF emitters, beginning with a library of 40,000 molecules generated *via* random structural alterations made to 20 existing “parent” TADF emitters using the STONED algorithm²⁹. We first present the library generation and screening process, as well as benchmarking calculations for calibrating screening parameters. We then present the workflow results, also highlighting molecules with interesting properties. We finish with a summary and suggestions for expanding this work in future.

2 Methods

A generative approach is used to create a library of “child” molecules starting from a known TADF “parent” molecule. As is typical of HTVS workflows, a funnel approach is then employed, with molecules screened out at each stage based on filters. Following library generation, the calculation stages can be broadly separated into initial screening, synthesisability screening, initial geometry optimisations, DFT^{30,31} single point calculations, DFT geometry optimisations and TDDFT calculations. The stages and filters are depicted in Fig. 1, using the parent molecule TXO-TPA as an example, and are explained in detail in the following.

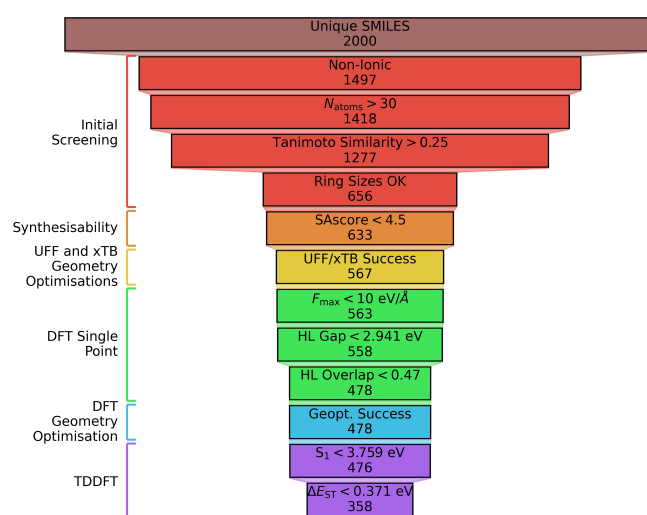


Fig. 1 Funnel depiction of the various calculation steps, screening stages and molecules retained at each point for the parent molecule TXO-TPA. Calculation stages are highlighted in a different colour, with individual filters, threshold values and number of molecules passing each stage given on the funnel. For further details see the main text.

2.1 Library Generation and Initial Filters

Some previous HTVS studies screen a (subset of a) database of known compounds, the advantage of which being that these are known to be both synthesisable and stable³². However, TADF emitters may not lie within the chemical space of such databases²⁸. Therefore, to achieve chemical diversity and increase the likelihood of discovering novel molecules, it can be advantageous to generate a custom library. Previous HTVS studies have done so by combining different D, A and linker moieties^{5,22–27}, though this fails to account for MR structures. Further generations of molecules may be generated from an initial library *via* structural mutations or ML in order to improve diversity^{23,25,27,28,33}. Domain knowledge may also be incorporated into the molecular design process, e.g. *via* morphing operations²⁸. This can hold advantages over randomised approaches in the sense of more relevant molecules being generated, but may lack the novelty of random generation.

In this work a random approach is taken based on mutating known TADF emitters, balancing the advantages of randomness and domain knowledge. This is done using the STONED algorithm²⁹, which produces child molecules through random point mutations to a parent molecule. STONED first reorders an input SMILES string representing the parent, to improve the diversity of generated structures. The resulting SMILES strings are then converted into SELFIES format, an alternative and robust molecular string representation in which all strings represent valid molecules³⁴. In this work a single point mutation is applied, defined as a string substitution corresponding to the addition, deletion or replacement of a single SELFIES character. The mutated SELFIES strings are then converted back into SMILES format and canonicalised. For each parent molecule used in this research, 2000 unique child SMILES strings were generated.



Following molecular generation, a number of rudimentary filters are imposed to refine the search space. First, open-shell molecules are removed, as the focus is on traditional uncharged TADF emitters, and they can also require alternative methods at subsequent workflow steps. Next, a ring size filter is imposed, whereupon candidates without rings or containing rings of sizes outside of 5- or 6- members are removed, since smaller rings are subject to increased ring strain which can negatively affect synthetic accessibility³⁵ and lead to reduced stability. Both smaller (3-4 member) and larger (7+ member) rings are also scarcely observed in the literature, while to the best of our knowledge there are no known TADF structures without rings.

Candidates are then filtered based on molecular size, with a minimum threshold of 30 atoms (including H atoms), in order to remove small fragment-like child molecules. No upper limit is applied since child molecules are not significantly bigger than their parents. A similarity filter is also included, using the Tanimoto coefficient, which is a measure of the common structural features between two molecules³⁶. ECFP4 fingerprints³⁷ are used for the calculation of Tanimoto scores, and both fingerprints and similarity scores are obtained using the RDKit package³⁸. The minimum similarity threshold is set to 0.25, indicating that the child structure shares at least 25% of structural features with its parent. This ensures that child molecules maintain at least part of the parent molecule structure, and are thus not totally random.

2.2 Synthesisability and 3D Structure Generation

An important yet sometimes overlooked consideration in computational molecular design is synthetic accessibility. Approaches for estimating this may broadly be divided into complexity-based methods which assess the complexity of chemical features, and retrosynthetic methods which take into account the full synthetic process including intermediates³⁹. Full synthetic pathway planning is time-intensive and thus not suitable for an HTVS approach. ML approaches also exist but again require time and expense for training. Cheminformatics tools are therefore the most appropriate choice for a high-throughput workflow; examples include SCScore^{39,40}, RAScore⁴¹ and SYBA⁴². This work employs SAScore, as implemented in RDKit, which takes into account both complex structural features in the form of a complexity penalty, as well as the frequency of molecular fragments among structures in the PubChem database, representing a balance between the aforementioned classes of synthesisability scoring³⁹. Although designed for drug-like molecules⁴³, SAScore has been used in previous HTVS approaches^{5,27}, one of which included experimental validation of candidates⁵. SAScore ranks molecules on a scale from 1 to 10, with lower scores representing increased ease of synthesis and *vice versa*. In this workflow, a threshold of 4.5 was used, at which SAScore has shown equal or better performance to other cheminformatics-based metrics³⁹.

Following SAScore screening, the SMILES strings are converted to 3D structures using RDKit. The universal force field (UFF)⁴⁴, as implemented in RDKit, is then used to optimise up to 10 conformers, following which the lowest energy structure is retained. Other force fields like the Merck Molecular Force Field

(MMFF94)^{45–49} are better suited to organic molecules⁵⁰, but MMFF94 does not support all elements, including B, which is frequently seen in MR emitters. Therefore UFF is used for the initial optimisation, but given its limitations, a further geometry optimisation of the lowest energy UFF-optimised structure is then performed using the GFN2-xTB extended semiempirical tight-binding model⁵¹, which is more expensive than UFF but applicable to a wide range of systems, including organic molecules, and improved the stability of the workflow compared to using UFF alone. A small number of UFF optimisations failed to converge within 10,000 iterations, or led to obviously poor structures, e.g. with overlapping atoms or separated into distinct structures; such molecules are eliminated. A handful of molecules are also eliminated due to failed xTB optimisations.

2.3 DFT and TDDFT Calculations

Having generated a library of 3D molecular structures, excited state calculations can be performed, but given the relatively high cost of TDDFT, particularly for large molecules, it is desirable to have an intermediate screening stage. Based on observed correlations from Penfold for a selection of TADF emitters⁵², the HOMO-LUMO gap and overlap are therefore used to screen out molecules based on predicted values for ΔE_{ST} and S_1 respectively, and so the next workflow stage is single point DFT calculations, which are much cheaper than TDDFT. The corresponding thresholds are based on benchmark calculations described in Section 3.

Following the single point calculations, DFT geometry optimisations are performed to improve the xTB-optimised structures, followed by TDDFT calculations. Despite its limitations, particularly with respect to long range CT excitations, TDDFT is widely used in HTVS workflows due to its balance between accuracy and computational cost. While ultimately adiabatic excitation energies are required for direct comparison with experiment, excited state geometry optimisations are expensive, and thus only vertical excitations are calculated, in line with other workflows, which either focused on vertical excitations only or performed excited state geometry optimisations for select molecules only at a later step^{5,13,25}. As shown in Section 3, this proves to be a reasonable approximation. Explicit filters are imposed on S_1 and ΔE_{ST} , with the thresholds chosen based on benchmark calculations described below. No explicit filter is used for f_{S_1} , this is instead considered qualitatively for molecules which pass all filters.

2.3.1 Computational Details.

Ground state DFT calculations were performed using the BigDFT code⁵³, using HGH-GTH pseudopotentials^{54,55}, with non-linear core corrections where available⁵⁶. Single point calculations employed a wavelet grid spacing of 0.45 bohr and coarse (fine) radius multipliers of 5 (7). A small number of molecules were screened out due to not converging within 80 self-consistent iterations, as defined by a wavefunction gradient of less than 10^{-4} . Molecules were also eliminated if the norm of the DFT forces of the xTB-optimised structure was above $10 \text{ eV}/\text{\AA}$, indicating a possible convergence problem or a poor structure. Geometry optimisations employed the same radius multipliers but with a smaller grid spacing of 0.4 bohr, to allow for more accurate force cal-



culations, and employed the semi-empirical dispersion correction scheme of Grimme (DFT-D3)⁵⁷. Molecules were again screened out if after 500 geometry steps the force norm was not below the target threshold of 0.03 eV/Å. All BigDFT calculations employed the PBE exchange-correlation functional⁵⁸.

TDDFT calculations were performed with the NWChem code⁵⁹ in gas phase, using the Tamm-Dancoff approximation (TDA)⁶⁰, the B3LYP exchange-correlation functional^{61–64} and 6-31G* basis set for all atoms except I, where 6-311G* was used. Initial benchmarks were performed for a subset of parent molecules using cc-pVDZ and cc-pVTZ, where it was found that there was only a small difference in energies at the cost of more expensive calculations. Similarly, initial benchmarks performed using the PBE0 functional⁶⁵ showed little difference with respect to B3LYP. Although not optimal for either D-A emitters, where e.g. optimally tuned range-separated hybrid functionals^{4,52} have been shown to give better agreement with experiment than pure hybrid functionals, or for MR emitters, where TDDFT significantly overestimates ΔE_{ST} , due to the lack of doubly excited determinants^{66,67}, cost was a key consideration. Furthermore, although e.g. absolute ΔE_{ST} values are systematically overestimated for MR emitters, trends between molecules are much less severely affected by this limitation, which is sufficient for this work given the aim is to identify candidate child molecules which show promising properties relative to a specific parent molecule, rather than accurately predict ΔE_{ST} values. A handful of molecules used for benchmarking and one molecule in the final workflow were screened out after TDDFT due to having negative S_1 or T_1 energies, which was assumed to have been caused by convergence issues.

All workflow steps were implemented in the form of Jupyter notebooks, where the DFT and TDDFT stages employed the remotemanager⁶⁸ Python package to submit calculations and retrieve data from a remote supercomputer. PyBigDFT⁶⁸ was used to generate input files and post-process both BigDFT and NWChem results, as well as to wrap xTB calculations.

3 Benchmark Calculations

In order to assess the workflow performance for both D-A and MR molecules, a set of 10 parent TADF emitters was defined for each. These molecules, which are depicted in Fig. 2, were chosen for diversity of both atomic structure and emission wavelength. The employed SMILES strings and a summary of key properties are given in Tables I and II in the SI. Before running the full workflow, benchmark calculations were performed to calibrate the thresholds for the DFT and TDDFT calculations, as discussed below.

3.1 Experimental Benchmarks

Fig. 3a shows a comparison between TDDFT vertical S_1 energies and experimental emission wavelengths for the parent molecules. Despite the lack of excited state geometry optimisation and differences with respect to experimental conditions (e.g. measurements were performed in various solvents or thin films vs. gas phase calculations), as well as the expected differences in TDDFT accuracy between the two types of emitter, a linear fit proves to be surprisingly good at predicting the emission wavelengths, leading to

a smaller mean absolute error in predicted wavelengths than using the expected inverse relationship between energy and wavelength. The linear fit shown in Fig. 3a was therefore used both to predict the emission colour for all subsequent calculations in this work, and for defining a cut-off threshold such that molecules with calculated S_1 values giving a predicted emission wavelength below 380 nm can be neglected due to expected emission outside the visible range. No cut-off was defined at longer wavelengths since none of the child molecules in subsequent benchmarks were predicted to exhibit non-visible emission at this end of the spectrum.

Fig. 3b shows a comparison between calculated and experimental ΔE_{ST} values, separated by emitter type. As expected, absolute ΔE_{ST} values are not well captured, particularly for MR emitters, however the correlation is similar for both types of emitters ($r^2 = 0.61$ (0.59) for D-A (MR)). While not as accurate as the emission wavelength predictions, the trends in ΔE_{ST} are sufficiently well approximated to define a threshold ΔE_{ST} value above which molecules can be discarded. Separate D-A and MR thresholds were found using the linear fits and the maximum parent molecule experimental ΔE_{ST} value. Although the r^2 value remains similar if a linear fit is applied to all emitters collectively, the resulting thresholds would lead to many more DA molecules being retained, reducing workflow efficiency. While it is possible that some promising molecules will be erroneously discarded using this threshold-based approach, these are likely to have ΔE_{ST} on the higher end of the targeted range, while the cost savings with respect to a more expensive method make this a reasonable compromise within the context of an HTVS approach.

3.2 HOMO-LUMO Thresholds

Having determined TDDFT thresholds, further benchmarks were performed to determine thresholds for the HOMO-LUMO gap and spatial overlap. These were done for the parent molecules, and for child molecules generated from a subset of the parent molecules (see Fig. 2). For each parent molecule in this subset, single point DFT calculations, ground state DFT geometry optimisations and single point TDDFT calculations were performed for 200 molecules which passed the initial filters.

The correlation between HOMO-LUMO gaps and calculated vertical S_1 energies for the benchmark molecules is shown in Fig. 3c. Semi-local functionals such as PBE are not expected to accurately predict band gaps, however the band gaps themselves are not directly relevant, except as a means for predicting the emission colour. Indeed, as shown in Fig. 3c, the correlation between the PBE-calculated HOMO-LUMO gap and the TDDFT/B3LYP S_1 energy is excellent ($r^2 = 0.93$). The linear fit depicted in Fig. 3c was therefore used to determine a threshold for the HOMO-LUMO gap, using the threshold for calculated S_1 values discussed above (see Table III in the SI for values). Also shown in Fig. 3c is the number of molecules which are correctly (wrongly) retained (discarded) using the HOMO-LUMO gap as a predictor. Only 2 out of 2010 molecules are erroneously discarded, with 17 being needlessly calculated. A corresponding breakdown divided by parent molecule is given in Table IV in the



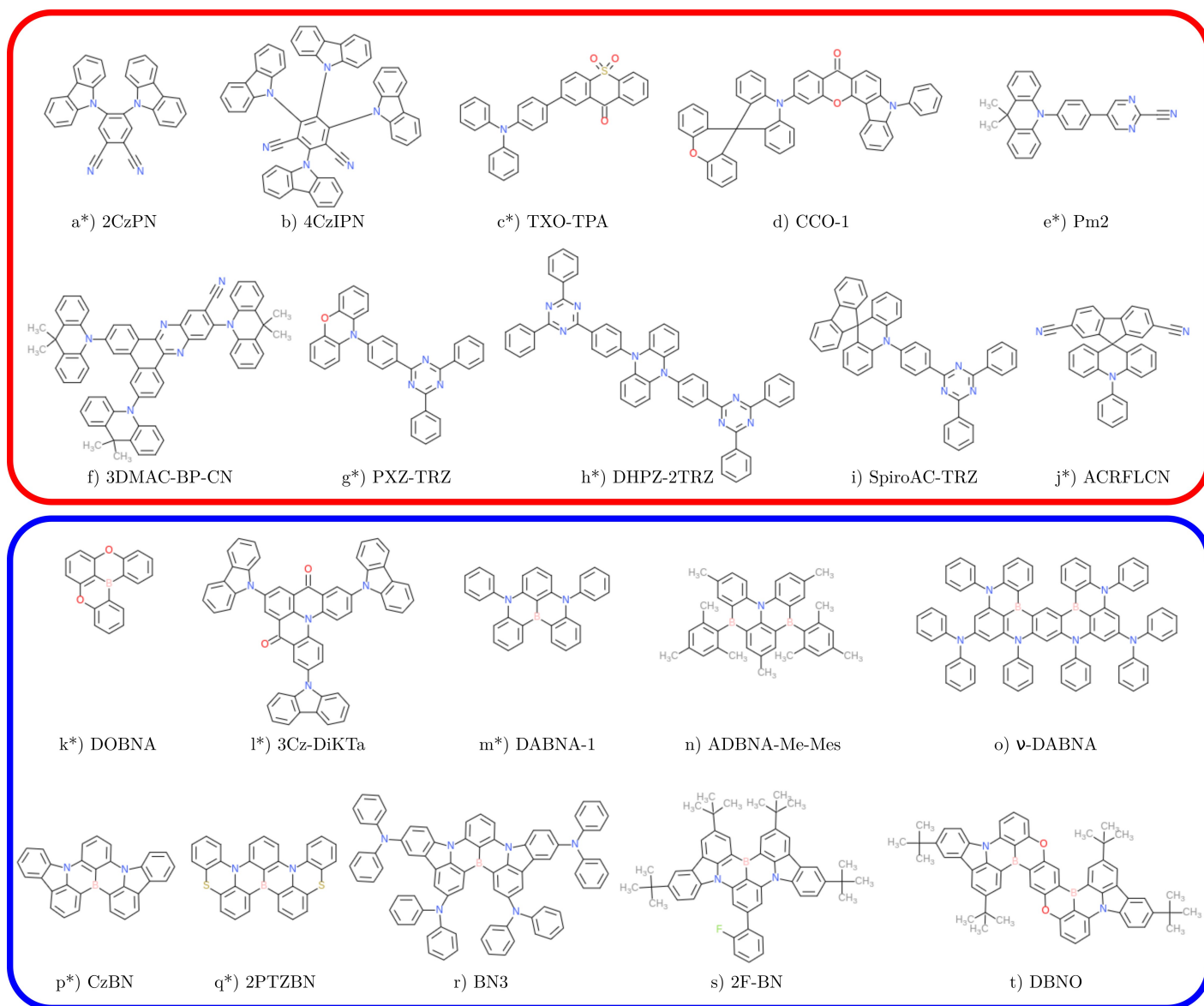


Fig. 2 Depiction of the parent molecules, with D-A (MR) molecules in the red (blue) box. The molecules which were used to generate the set of benchmark molecules are labeled with *. Images were generated using ChemDoodle Web Components⁶⁹.

SI.

Fig. 3d shows a comparison between the HOMO-LUMO spatial overlap and calculated vertical ΔE_{ST} , with D-A and MR molecules plotted separately. The correlation is much weaker than between the HOMO-LUMO gap and S_1 energy, in large part due to the fact that not all S_1 and T_1 states are dominated by a HOMO→LUMO transition. Indeed, the majority of the outliers are molecules which have S_1 and/or T_1 transitions which are not HOMO→LUMO dominated. Despite the weaker correlation, there are only a handful of molecules which have a relatively high HOMO-LUMO spatial overlap and small ΔE_{ST} . Therefore, threshold values for the HOMO-LUMO spatial overlap were chosen such that no more than 1% of molecules with ΔE_{ST} below the threshold would be erroneously discarded, while also ensuring all parent molecules would be retained. As shown in Fig. 3d, this leads to only a handful of molecules being incorrectly lost. Although a

relatively large number of molecules with high ΔE_{ST} are retained, 10% (6%) of D-A (MR) molecules are nonetheless correctly discarded, resulting in significant computational savings due to skipping subsequent workflow steps. Final threshold values and the breakdown of molecules lost/retained divided by parent molecule are given in Tables III and IV in the SI respectively. The use of the HOMO-LUMO gap for predicting ΔE_{ST} was also explored, but this showed no correlation, while the use of the HOMO-LUMO overlap $\times$ the HOMO-LUMO gap led to a higher number of molecules with high ΔE_{ST} being needlessly retained.

4 Workflow Results

Having determined the threshold values, all stages of the workflow were run for each of the 20 parent molecules. In the following the overall performance of the workflow is first discussed, including drawbacks to the workflow. This is followed by a dis-



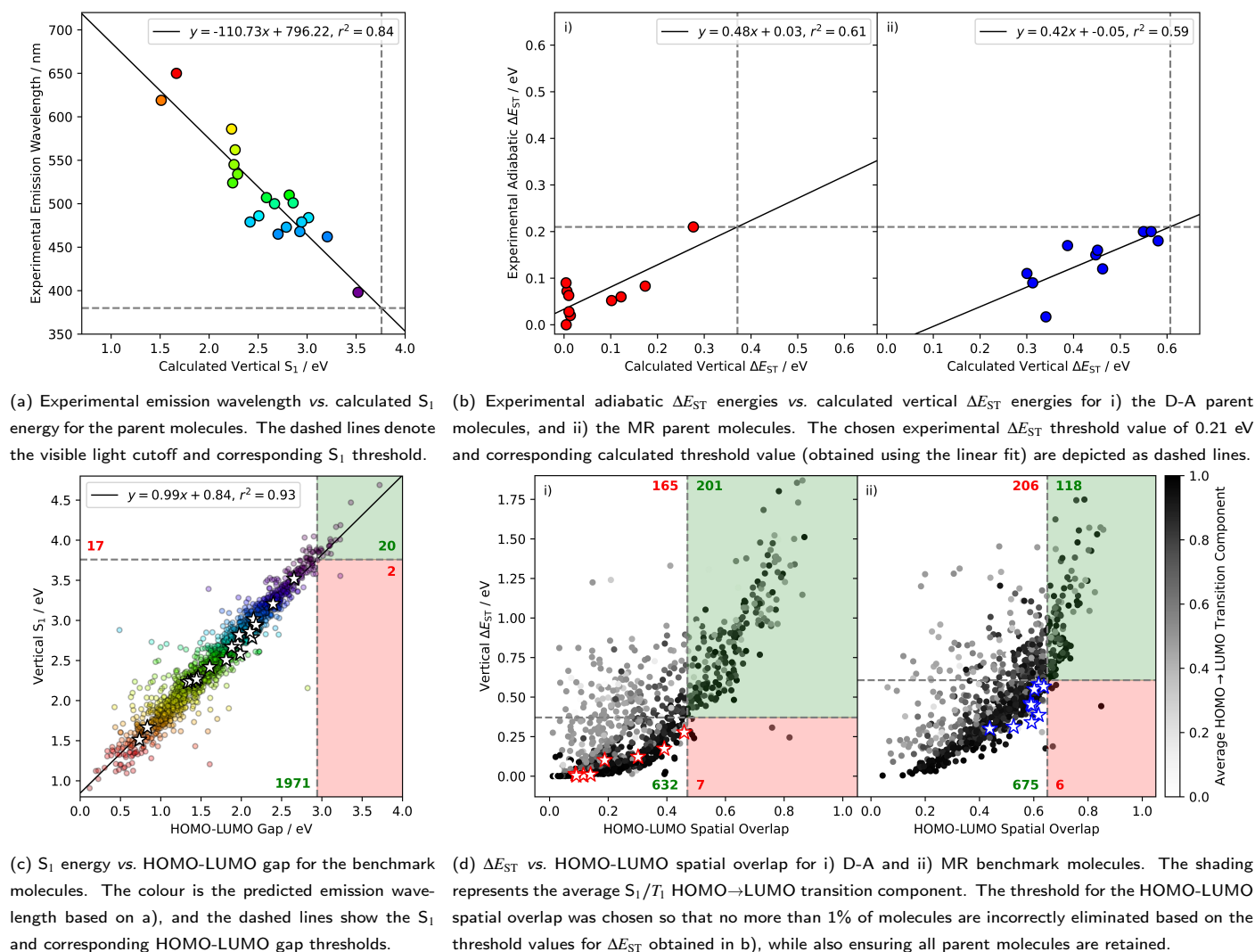


Fig. 3 Benchmarking plots for determining the screening thresholds for HOMO-LUMO gaps and overlaps. The green (red) shaded regions in c) and d) show molecules which are correctly (wrongly) eliminated following the HOMO-LUMO calculation. The corresponding numbers of molecules are written in green (red), as are those which are correctly (needlessly) retained in the unshaded areas. For c) and d) parent molecules are highlighted using the ★ symbol. S_1 , ΔE_{ST} and HOMO→LUMO transition components were calculated using TDDFT at the B3LYP/6-31G* level on the DFT-optimised geometries, while HOMO-LUMO gaps and overlaps were calculated using PBE and a wavelet basis on the xTB-optimised geometries. Experimental values are from Refs^{9,12,70–87}; the ΔE_{ST} value for DHPZ-2TRZ is recorded as “nearly zero”, and so was approximated to zero.

cussion in the trends in properties of the potential candidate child molecules, i.e. those which passed all filters. The section finishes with a discussion highlighting some of the most interesting candidate child molecules.

4.1 Workflow Performance

In the following we discuss the performance of the workflow as a whole, including some of the issues and potential improvements which could be made in future. The number of molecules retained at each filter stage are shown in funnel plots, an example of which is given for TXO-TPA in Fig. 1, with the remainder shown in Figs. 1–4 in the SI.

4.1.1 Molecular Generation and Synthesisability

The initial filters are responsible for eliminating a large number of child molecules: on average roughly a quarter are eliminated due

to being ionic, while nearly three quarters have been eliminated by the time synthesisability is taken into account. A large number of these molecules are eliminated due to undesirable ring sizes, in the majority of cases due to the existence of large rings (7+ members), which can arise when changes are made involving branch- or ring-related SELFIE characters. While this demonstrates the high level of chemical diversity that can be achieved via STONED even with just a single mutation, the random nature of STONED is not very efficient at generating relevant molecules, indeed in the worst cases the generated molecules are small fragments or unrealistic structures. Nonetheless, the initial filters are very cheap, and are able to filter out the majority of the obviously unsuitable molecules, so that this is not a significant drawback.

The most noticeable variations between parent molecules are due to the SAScore filter. This is partly due to the fact that larger molecules tend to have higher SAScores, e.g. DBNO has



a high SAScore of 4.26, close to the threshold of 4.5, so that only 86 molecules passed this stage. At the other extreme, 2CzPN, being much smaller and having an SAScore of 2.29, had 653 molecules pass this stage. In future, it might therefore be advisable to vary the threshold between parent molecules, or investigate other methods for predicting synthesizability, although as discussed above, the cost of approaches such as full synthetic route planning tools which might be more reliable is too high for a HTVS approach, and thus some trade-offs must be made. However, it could be worth exploring in future whether such tools could be used later in the workflow as part of a more detailed investigation into promising candidate molecules. In general, MR molecules have fewer child molecules passing all filters, influenced by the larger average molecule size and correspondingly higher SAScores. By the end of the workflow the gap between the number of remaining molecules narrows, with e.g. 77, 180 and 358 molecules passing the final filters for DBNO, 2CzPN and TXO-TPA respectively, where the latter is the parent molecule for which the most child molecules passed the final filters.

4.1.2 Duplicate Molecules

The workflow was run independently for each parent molecule. Despite the relatively high similarity of some parent molecules, e.g. DBNO and 2F-BN have a Tanimoto similarity of 0.70 (see Fig. 5 in the SI for full parent molecule similarity matrix), there were only six pairs of duplicate child molecules making it through to the first DFT calculation stage; only two of these duplicate pairs made it through the final filters. These all came from D-A parent molecules, with any duplicates generated between MR parent molecules or between D-A and MR parent molecules (primarily small fragments) eliminated at early stages. In future a database approach could be used to eliminate molecules which have already been generated. Another minor limitation of the current implementation is that STONED occasionally generates molecules which are identical to the parent molecule. When such molecules had identical SMILES strings they were filtered out, but there were a few instances of molecules with different SMILES strings which nonetheless had a Tanimoto similarity of 1. In some cases the calculated properties were slightly different from those of the parent molecule due to small differences in the geometry, highlighting the dependency on the initial SMILES string.

4.1.3 Rediscovery of Existing Emitters

One test for evaluating the effectiveness of an HTVS workflow is if existing emitters are rediscovered. To this end, the final child molecules were compared to a set of around 40 known TADF emitters, chosen due to structural similarities with the parent molecules. Only one such molecule was rediscovered: Py2 (Pm2 child molecule 114), which is somewhat surprising, particularly in cases where these known emitters have high similarity with respect to the related parent molecule, e.g. 4CzIPN has a Tanimoto similarity of 0.98 with respect to 4CzTPN, but was nonetheless not rediscovered. One possible explanation is that a single mutation is insufficient to explore a diverse enough chemical space, however repeating the library generation step for increasing numbers of mutations, with 10,000 molecules gener-

ated each for 1 through to 10 mutations (i.e. generating 100,000 unique SMILES per parent molecule) did not result in any further rediscoveries. Furthermore, the average Tanimoto similarity with respect to the parent molecule drops off rapidly with the number of mutations, so that in most cases by 4-5 mutations the average Tanimoto similarity with respect to the parent molecule is less than 0.25. Thus while increasing the number of mutations increases diversity, this rapidly tends towards a region of chemical space which is far from the parent molecule. At the same time, while the STONED algorithm applies the addition, deletion and replacement of characters with equal probability, the nature of these mutations is such that e.g. inserting a character can result in a smaller molecule. Indeed, for the parent molecules considered here, we observed that the average size of the resulting child molecules tends to decrease with increasing number of mutations. Since some of the existing emitters are much larger than the most similar parent molecule, they are thus seemingly out of reach of the STONED algorithm. Similarly, many of the existing emitters have high symmetry, which a random approach tends to break. It might therefore be interesting in future to combine a random approach with e.g. a genetic algorithm or ML-based approach to target the mutations more effectively. However, as shown below, the workflow resulted in a number of child molecules which have promising properties, thus it was nonetheless successful in exploring a relevant and useful region of chemical space around the parent molecules.

4.2 Trends in Candidate Child Molecule Properties

The distribution of ΔE_{ST} values for child molecules which passed all filters is shown in Fig. 4. Most of the parent molecules have peaks around the ΔE_{ST} of the parent molecule, with e.g. 2F-BN and DBNO having very sharp peaks, while other molecules like 2CzPN have a wider spread of values. The majority of D-A parent molecules have a peak in ΔE_{ST} values close to 0 eV. There are two exceptions: 2CzPN and 4CzIPN have notably fewer molecules with very small ΔE_{ST} . This is in line with the higher parent ΔE_{ST} values, where 2CzPN also has a high experimental ΔE_{ST} value, while that of 4CzIPN is on the higher end and overestimated with respect to experiment. Nonetheless both molecules still have a number of molecules with small ΔE_{ST} . Unsurprisingly the MR child molecules typically have larger calculated ΔE_{ST} values, due to both higher experimental values and TDDFT overestimating. Nonetheless, for each MR parent molecule the child molecules show a range of ΔE_{ST} values, including very small values. There is also a spread in HOMO-LUMO spatial overlap values (Fig. 6 in the SI). Although only applicable for HOMO-LUMO dominated excitations, it is interesting to note that there are MR child molecules with very small overlap, indicating possible D-A character. Thus while the majority of child molecules are of the same character as their parent, the random nature of STONED allows for the possibility of generating D-A molecules from MR parents and *vice versa*, as well as molecules which may have aspects of both, which could be explored further in future work.

There were also a wide range of S_1 oscillator strengths across the final child molecules (see Figs. 7-10 in the SI). Slightly differ-



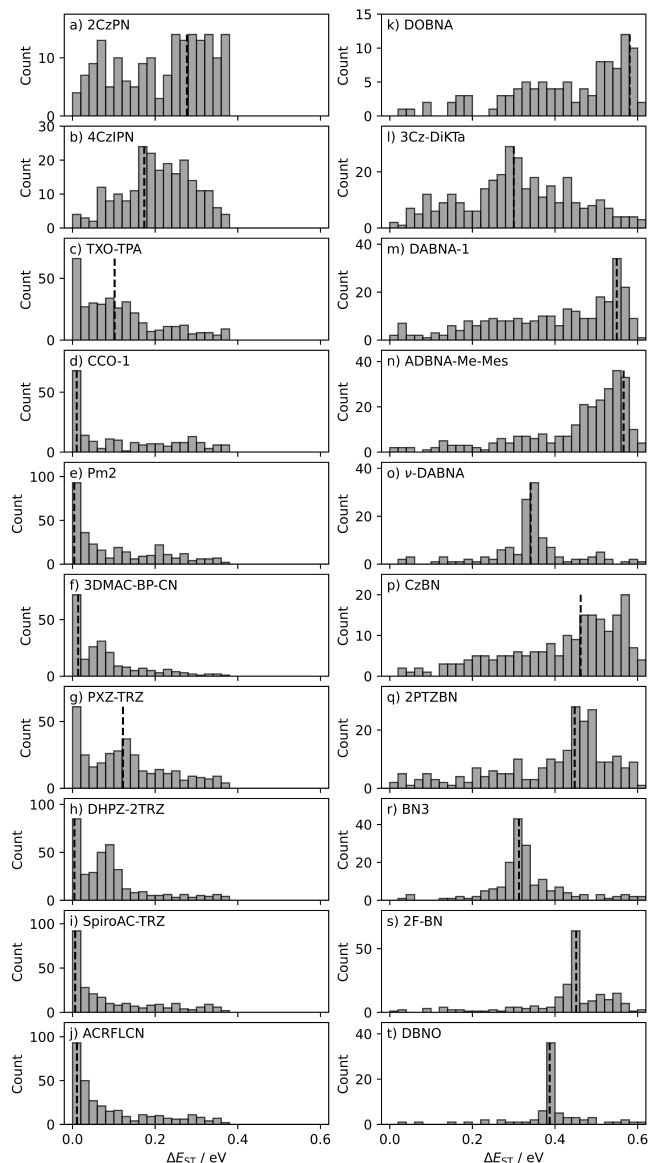


Fig. 4 ΔE_{ST} values for molecules which passed all workflow filters. Values for parent molecules are depicted with a dashed vertical line.

ent trends were seen when considering ΔE_{ST} vs. f_{S_1} for the D-A molecules compared to the MR molecules, with a greater number of D-A child molecules having very small f_{S_1} , in line with a number of D-A parent molecules also having very small f_{S_1} (Fig. 11 in the SI). However, in both cases, as is to be expected ΔE_{ST} and f_{S_1} are in competition, with the molecules having the smallest ΔE_{ST} also having smaller f_{S_1} values. For this reason, and because of the very small f_{S_1} values of some parent molecules, it was decided to not filter molecules based on a particular f_{S_1} threshold, and instead balance both factors when considering candidate child molecules. Given the differences in behaviour for different parent molecules, this was easier than e.g. using a combined function of merit, as has been done in previous work²⁵.

The final distribution of (vertical) S_1 energies is shown in Fig. 5, including the predicted emission colours. In most cases there is a peak around the parent molecule value, although some

have very sharp peaks, e.g. BN3 and DBNO, while others like 2CzPN have more even distributions of S_1 energies. Although there are limitations to this approach to predicting emission colour, it is nonetheless clear that some parent molecules more easily lend themselves to colour tuning across the full spectrum. Indeed, most of the MR parent molecules tend to have a narrower spread of values, with relatively few red emitters, although this is also highly influenced by the distribution of emission colours of the MR parent molecules themselves, as well as a number of the predicted emission colours for the MR parent molecules being blue-shifted compared to experiment.

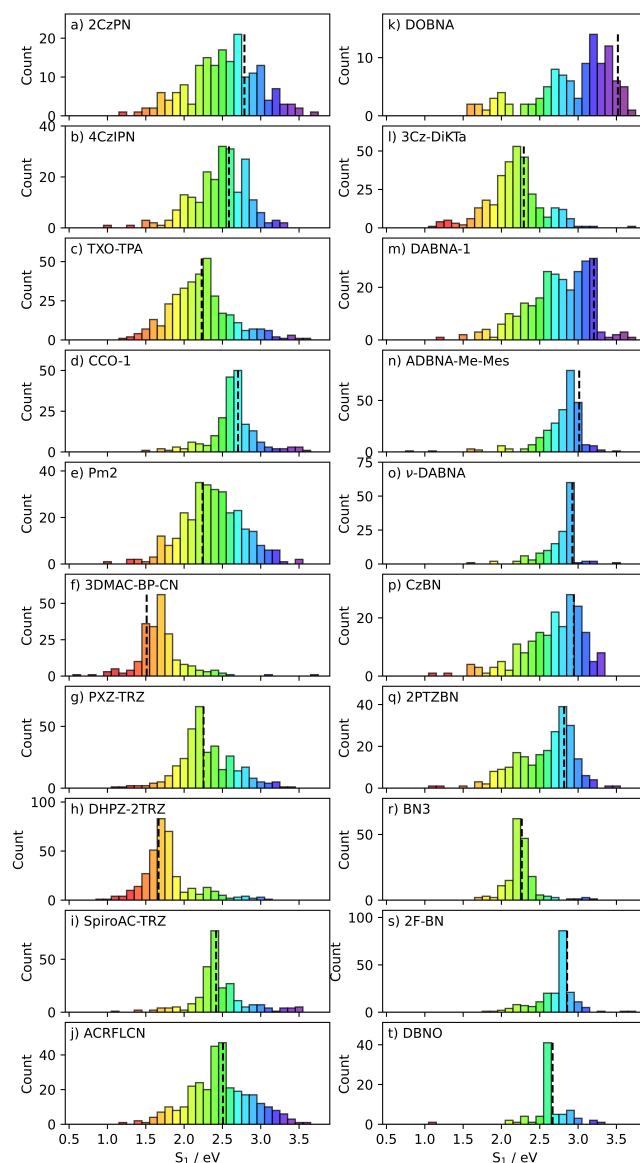


Fig. 5 Vertical S_1 values for molecules which passed all workflow filters, with corresponding predicted emission colour. Values for parent molecules are depicted with a dashed vertical line.

4.3 Candidate Child Molecules

We conclude the results by discussing interesting candidate child molecules for each parent. This includes extremes of emission



colour, small ΔE_{ST} , and balancing small ΔE_{ST} with reasonable f_{S1} . Where applicable, we also discuss any notable structural features. Child molecules are indexed separately for each parent molecule. The number of candidate molecules meeting various ΔE_{ST} and f_{S1} criteria for each parent molecule are also summarised in Tables V and VI in the SI, while key information for the molecules with both smaller ΔE_{ST} and larger f_{S1} than their respective parent are also given in Tables VII–XVII in the SI.

4.3.1 D-A Molecules.

4.3.1.1 2CzPN. There is a broad spread of predicted emission colours for the candidate child molecules, although most are blue or green, in line with 2CzPN's blue emission. At the extremes, there are four red/orange-red candidates; of these only **1392** has a small ΔE_{ST} (0.09 eV), below that of 2CzPN. There are several violet emitters, with three having $\Delta E_{ST} < 0.2$ eV (**1583**, **1657** and **1723**), although all have $f_{S1} \leq 0.001$. A total of 38 molecules have $\Delta E_{ST} < 0.1$ eV, of which 14 are below 0.05 eV, covering a broad range of colours. Balancing ΔE_{ST} and f_{S1} , there are no molecules with $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.1$, although 2CzPN itself has $f_{S1} = 0.05$. There are however six candidates with $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.05$, which are all blue or green emitters. Twelve molecules have smaller ΔE_{ST} and larger f_{S1} than 2CzPN, with a wide range of Tanimoto similarities with respect to the parent molecule and blue through to yellow emission. Both S and I heteroatoms feature in these molecules.

4.3.1.2 4CzIPN. Many child molecules are green emitters, like 4CzIPN itself. There are two red emitters (**1277** and **1690**), where **1277** has $\Delta E_{ST} < 0.1$ eV but also small f_{S1} (0.003), as well as a few orange emitters. There are various violet and deep blue emitters, but many of these have larger ΔE_{ST} than 4CzIPN, with **772** being the shortest wavelength emitter (459 nm) that has smaller ΔE_{ST} ; it also has $f_{S1} = 0.06$, just below the parent molecule. There are 29 molecules with $\Delta E_{ST} < 0.1$ eV, of which seven have $\Delta E_{ST} < 0.05$ eV, spanning blue to red emission. Two molecules have $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.1$ (**193** and **1574**), which are both blue. Seven molecules have smaller ΔE_{ST} and larger f_{S1} than the parent molecule, with blue to green emission, and Tanimoto similarities with respect to 4CzIPN ranging from 0.63 to 0.91. Some of these have lost functional groups compared to 4CzIPN. P and Cl heteroatoms feature, with S, Br and O heteroatoms also appearing among molecules with small ΔE_{ST} .

4.3.1.3 TXO-TPA. There is a broad peak of predicted emission colours around green, in line with the predicted emission colour of TXO-TPA, whose emission is predicted to be green instead of yellow. There are various red and dark orange emitters, including three with predicted wavelengths above 640 nm (**438**, **569** and **974**), all with small ΔE_{ST} but also very small f_{S1} (< 0.001). There are also various violet and deep blue emitters, including five with predicted wavelengths under 420 nm. Of these, both **248** and **529** have ΔE_{ST} values smaller than TXO-TPA, but they also have smaller f_{S1} than the parent. There are 186 molecules with $\Delta E_{ST} < 0.1$ eV, of which 109 have $\Delta E_{ST} < 0.05$ eV, ranging from blue to red. Three molecules have $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.1$ (**748**, **1423** and **1849**), which are various shades of green. A number of

heteroatoms feature, in particular P, which is also present in two of the six molecules which have both ΔE_{ST} and larger f_{S1} than TXO-TPA. These molecules are also either green or yellow, and have a range of Tanimoto similarities with respect to TXO-TPA.

4.3.1.4 CCO-1. There is a sharp peak of emitters with green or blue-green emission, in line with the emission colour of CCO-1, which is predicted to be blue-green, somewhat redshifted compared to the blue experimental emission. There are no predicted red emitters and few orange, including two with emission around 600 nm (**1040** and **1054**), and one at 619 nm (**193**). The latter has small ΔE_{ST} (0.001 eV) but also small f_{S1} (4×10^{-4}), although this is larger than the very small f_{S1} of the parent molecule (10^{-5}). There are various violet and deep blue emitters, including seven with emission wavelengths below 420 nm, but none have $\Delta E_{ST} < 0.1$ eV, and those below 0.2 eV have small f_{S1} . There are 105 molecules with $\Delta E_{ST} < 0.1$ eV, of which 85 have $\Delta E_{ST} < 0.05$ eV, spanning blue to orange emission. One molecule has $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.1$ (**108**), with predicted green emission. Seventeen molecules have both smaller ΔE_{ST} and larger f_{S1} than the parent molecule, spanning blue to orange. These molecules mostly have Tanimoto similarities above 0.7 compared to CCO-1, while P, Cl and S heteroatoms all feature.

4.3.1.5 Pm2. There is a broad peak in emission energies around green, in line with Pm2's predicted green emission, which is red-shifted compared to experiment. At the red end of the spectrum, five molecules have predicted wavelengths above 630 nm, all with $\Delta E_{ST} < 0.1$ eV, with **1746** having the highest f_{S1} (0.04), although all are higher than Pm2's very small f_{S1} (2×10^{-5}). There are two violet outliers (**266** and **844**), where **844** has very small ΔE_{ST} (< 0.01 eV), but also small f_{S1} (2×10^{-4}). There are 141 molecules with $\Delta E_{ST} < 0.05$ eV, spanning violet to red, and 34 with $0.05 \leq \Delta E_{ST} < 0.1$ eV. There are 15 molecules with $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.1$, and 15 which have both smaller ΔE_{ST} and larger f_{S1} than Pm2 again spanning a range of emission colours, and with a range of Tanimoto similarities with respect to Pm2. P heteroatoms are present in five of the fifteen, while a range of other heteroatoms feature among the molecules with small ΔE_{ST} .

4.3.1.6 3DMAC-BP-CN. Most child molecules are red or yellow, in line with the orange-red emission of the parent. This includes two molecules with predicted emission wavelengths above 700 nm (**36** and **382**), where **36** has the smaller ΔE_{ST} of the two but also smaller f_{S1} . There is a single blue (**1966**) and a single violet molecule (**268**), both of which are much smaller molecules than 3DMAC-BP-CN, while also having larger ΔE_{ST} and small f_{S1} . There are 165 molecules with $\Delta E_{ST} < 0.1$ eV, of which 98 have $\Delta E_{ST} < 0.05$ eV, spanning green to red emission. There are 11 molecules with $\Delta E_{ST} < 0.2$ eV and $f_{S1} > 0.1$, and 20 which have both smaller ΔE_{ST} and larger f_{S1} than 3DMAC-BP-CN, where most of the latter have Tanimoto similarities above 0.7 with respect to the parent molecule, while a range of heteroatoms feature.

4.3.1.7 PXZ-TRZ. In line with PXZ-TRZ's predicted green emission, which is in excellent agreement with experiment, there is a spread of emission colours around green. There are various



red and orange emitters, where seven have predicted emission around 630 nm or above. Three of these have smaller ΔE_{ST} than PXZ-TRZ, but with $f_{S_1} < 10^{-4}$. There are various blues and violets, including six with predicted emission below 440 nm, of which only **1617** has ΔE_{ST} smaller than PXZ-TRZ, with $f_{S_1} = 0.02$. There are 93 molecules with $\Delta E_{ST} < 0.05$ eV, with blue through to red emission, and 54 with $0.05 \leq \Delta E_{ST} < 0.1$ eV. There are 73 molecules with $\Delta E_{ST} < 0.2$ eV and $f_{S_1} > 0.1$, and two with both smaller ΔE_{ST} and larger f_{S_1} than PXZ-TRZ (**664** and **1831**), which are green and yellow, and have high similarities with respect to the parent molecule (0.7-0.8). Like many of the emitters with small ΔE_{ST} , they both contain heteroatoms, in this case B and S.

4.3.1.8 DHPZ-2TRZ. The majority of child molecules are orange or yellow, in line with the predicted orange emission of DHPZ-2TRZ, which is blue-shifted compared to the red of experiment. There are many red emitters, including e.g. five with predicted emission wavelengths above 670 nm; of these **681** has the smallest ΔE_{ST} (0.01 eV), but also with very small f_{S_1} ($< 10^{-4}$), which is nonetheless still higher than the very small f_{S_1} of DHPZ-2TRZ (10^{-8}). There are no violet but a few blue emitters, including four at 460 nm or above, although all have $\Delta E_{ST} \geq 0.28$ eV. There are 124 molecules with $\Delta E_{ST} < 0.05$ eV, spanning green to red, and 125 with $0.05 \leq \Delta E_{ST} < 0.1$ eV. There are 85 molecules with $\Delta E_{ST} < 0.2$ eV and $f_{S_1} > 0.1$, and 51 which have both smaller ΔE_{ST} and larger f_{S_1} than DHPZ-2TRZ, which are green-yellow to red, and mostly have high Tanimoto similarities to the parent molecule. Some of these feature heteroatoms, though the majority have other structural changes, e.g. modified or broken rings.

4.3.1.9 SpiroAC-TRZ. Most child molecules are green, in line with the predicted green emission of the parent molecule, which differs from the blue experimental emission. There is one red outlier (**284**), which has high f_{S_1} (0.13), and reasonably small ΔE_{ST} (0.16 eV), albeit higher than that of SpiroAC-TRZ. There are also a number of orange emitters e.g. two with predicted emission wavelengths of 630 nm or above (**701** and **1265**). Of these, **701** is more promising, having both smaller ΔE_{ST} and larger f_{S_1} than the parent. There are many blue and violet emitters, e.g. thirteen with predicted emission wavelengths below 430 nm, which have a range of ΔE_{ST} values, but all larger than SpiroAC-TRZ. There are 168 molecules with $\Delta E_{ST} < 0.1$ eV, of which 132 have $\Delta E_{ST} < 0.05$ eV. Four molecules have $\Delta E_{ST} < 0.2$ eV and $f_{S_1} > 0.1$, spanning green to red, and 26 molecules have both smaller ΔE_{ST} and larger f_{S_1} than SpiroAC-TRZ, with a range of Tanimoto similarities with respect to the parent molecule. A range of heteroatoms feature among the more promising child molecules.

4.3.1.10 ACRFLCN. There is a broad spread of predicted emission colours, with green being the most common, in line with the predicted green emission of ACRFLCN, which differs from the blue of experiment. There is one red (**420**) and three dark orange (**269**, **486** and **1561**) emitters, all with larger ΔE_{ST} than the parent. There are many blue and violet, e.g. ten with predicted emission wavelengths below 430 nm, two of which have $\Delta E_{ST} < 0.1$ eV (**990** and **1884**). There are 156 molecules with $\Delta E_{ST} < 0.05$ eV, spanning a range of colours, and 50 with $0.05 \leq \Delta E_{ST} < 0.1$ eV. No molecules have $\Delta E_{ST} < 0.2$ eV and $f_{S_1} > 0.1$, or even $\Delta E_{ST} < 0.2$ eV

and $f_{S_1} > 0.05$, with most child molecules having small f_{S_1} , in common with the small f_{S_1} of ACRFLCN (4×10^{-4}). There are 15 molecules with smaller ΔE_{ST} and larger f_{S_1} than ACRFLCN, which are various shades of green, and have a range of Tanimoto similarities with respect to ACRFLCN. A range of heteroatoms feature in the molecules with small ΔE_{ST} .

4.3.2 MR Molecules.

4.3.2.1 DOBNA. Most child molecules are towards the violet end of the emission spectrum, like DOBNA, but there is still a spread of predicted emission wavelengths, including four molecules above 600 nm. Of these only **1152** has $\Delta E_{ST} < 0.1$ eV, it also has reasonable $f_{S_1} = 0.04$, albeit smaller than that of DOBNA. A further three molecules have $\Delta E_{ST} < 0.1$ eV (**248**, **835** and **1303**), while eight more have $0.1 \leq \Delta E_{ST} < 0.2$ eV. These have a range of f_{S_1} values, with **1152** having the largest, and range from violet to orange. There are no molecules with $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$, and only one with $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.05$: **1374**, which is predicted to be violet. There are five molecules with both smaller ΔE_{ST} and larger f_{S_1} than DOBNA, all with blue/violet emission, and most of which have high similarity to the parent molecule, with e.g. the only change being the addition/substitution of a single heteroatom.

4.3.2.2 3Cz-DiKTa. While there is a peak around the green emission of 3Cz-DiKTa, there is a range of emission colours. There are a number of red and orange emitters, with e.g. 13 molecules with predicted emission wavelengths above 630 nm, several of which have small ΔE_{ST} . There is one violet outlier (**289**), which is much smaller than the parent molecule, plus one other (**874**) with emission below 450 nm. Both have larger ΔE_{ST} and smaller f_{S_1} than the parent molecule. There are 27 molecules with $\Delta E_{ST} < 0.1$ eV, of which **323** has the largest f_{S_1} (0.02) and 42 molecules with $0.1 \leq \Delta E_{ST} < 0.2$ eV. Three molecules have both $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$ (**371**, **1904** and **1974**), which lie in the green to yellow emission range, where **371** is interesting in that the only difference is the replacement of the N in the DiKTa core with a B atom. There are 13 molecules which have both smaller ΔE_{ST} and larger f_{S_1} than 3Cz-DiKTa, which have green to orange emission, and Tanimoto similarities with respect to the parent molecule of between 0.5 to 0.75.

4.3.2.3 DABNA-1. Most child molecules are blue or green, in line with the deep blue emission of DABNA-1. There is one red molecule (**1577**) and a further five with predicted emission above 600 nm, all of which have ΔE_{ST} smaller than the parent. Of these, **312** and **855** have an S heteroatom bonded to one of the N atoms, and have f_{S_1} around 0.05. Four molecules have predicted emission below 400 nm, all with smaller ΔE_{ST} than DABNA-1, and two of which have $f_{S_1} > 0.05$ (**1450** and **1777**). There are 37 molecules with $\Delta E_{ST} < 0.2$ eV of which 14 have $\Delta E_{ST} < 0.1$ eV. A number of these contain either P or B atoms. Five molecules have $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$, which are all blue or green. Sixteen molecules have smaller ΔE_{ST} and higher f_{S_1} than DABNA-1, all with predicted blue emission and all but two with a Tanimoto similarity below 0.6 with respect to DABNA-1.



4.3.2.4 ADBNA-Me-Mes. The majority of child molecules of ADBNA-Me-Mes are blue or green, in line with the blue emission of the parent molecule. There are two predicted red emitters (**845** and **1895**), but **1895** has a larger ΔE_{ST} than ADBNA-Me-Mes and **845** has very small f_{S_1} . There are a further four molecules with predicted emission above 600 nm, all with smaller ΔE_{ST} than ADBNA-Me-Mes, with **1874** being the most promising with $f_{S_1} = 0.07$. At the violet end there is one outlier with predicted emission of 401 nm (**1585**), which has a smaller ΔE_{ST} than the parent and $f_{S_1} = 0.02$, and various other molecules with predicted violet/deep blue emission. Seven molecules have $\Delta E_{ST} < 0.1$ eV, with blue to orange emission, and 16 have $0.1 \leq \Delta E_{ST} < 0.2$ eV. Three molecules have both $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$ (**489**, **1332** and **1553**), which are blue or green. A total of 38 molecules have smaller ΔE_{ST} and higher f_{S_1} than ADBNA-Me-Mes, with blue or green emission and a wide range of Tanimoto similarities with respect to the parent molecule. A range of heteroatoms are present in the molecules with small ΔE_{ST} .

4.3.2.5 v-DABNA. Most child molecules are predicted to have blue emission like v-DABNA, with many also being green. There is one orange outlier (**1820**), with a similar ΔE_{ST} to v-DABNA, but a much smaller f_{S_1} , and two yellow emitters (**619** and **920**), with small ΔE_{ST} (0.12/0.21 eV) and moderate f_{S_1} (0.02/0.07). There is also a violet outlier (**1499**), which is much smaller than the parent molecule and has a larger ΔE_{ST} and smaller f_{S_1} . Five molecules have $\Delta E_{ST} < 0.1$ eV, all of which are predicted to be green, with a range of f_{S_1} . Three of these contain a P heteroatom, although other molecules containing P have ΔE_{ST} values similar to or worse than v-DABNA, depending on the substitution position. Eight have $0.1 \leq \Delta E_{ST} < 0.2$ eV, with f_{S_1} values up to 0.06, and a range of emission colours, from cyan to yellow. There are 13 molecules with $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$, mostly blue/cyan. Ten molecules have smaller ΔE_{ST} and higher f_{S_1} than the parent molecule, all with blue emission and all but one with Tanimoto similarity > 0.77 with respect to v-DABNA.

4.3.2.6 CzBN. Most child molecules are blue or green, in line with the blue emission of CzBN. There are various orange emitters, and two predicted red emitters (**17** and **497**), both with smaller E_{ST} but also smaller f_{S_1} than CzBN. There are eight violet emitters with predicted emission wavelengths below 430 nm, of which only **503** and **1197** have smaller ΔE_{ST} than the parent, with **1197** having the larger f_{S_1} of the two (0.1 vs. 0.03). Nineteen molecules have $\Delta E_{ST} < 0.2$ eV, of which six $\Delta E_{ST} < 0.1$ eV, which collectively span blue to red. A number of these contain either P or S heteroatoms. Only two molecules have both $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$: **629** and **1686**, which are blue and green respectively. A single molecule has both smaller ΔE_{ST} and higher f_{S_1} than CzBN: **273**, which has almost identical predicted emission colour to CzBN, with the only difference in structure being the substitution of one of the H atoms with Cl.

4.3.2.7 2PTZBN. Most child molecules are blue or green, in line with the predicted emission of 2PTZBN, which is slightly blue-shifted compared to the green of experiment. There are two red molecules (**977** and **1425**), as well as one orange (**1659**) and various other yellow/yellow-orange molecules. Of these both **977**

and **1659** have a smaller ΔE_{ST} than PTZBN, but **977** has a very small f_{S_1} : 0.006 compared to 0.02 for **1659**. There are two violet outliers with predicted emission wavelengths below 420 nm (**447** and **697**), where **447** has a smaller ΔE_{ST} than PTZBN, but also a smaller f_{S_1} (0.02). Sixteen molecules have $\Delta E_{ST} < 0.1$ eV, and 12 have $0.1 \leq \Delta E_{ST} < 0.2$ eV, collectively spanning red to blue emission. Two molecules have both $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$: **1307** and **1852**. There are five molecules with both ΔE_{ST} and higher f_{S_1} than 2PTZBN, all of which have high Tanimoto similarities with respect to 2PTZBN (> 0.7) and very similar predicted emission wavelengths. Both I and P heteroatoms feature among these molecules, with other heteroatoms such as Cl appearing among the molecules with small ΔE_{ST} .

4.3.2.8 BN3. Most child molecules are yellow or green, in line with BN3, whose predicted emission is a yellowish green, with a wavelength that is slightly underestimated compared to experiment. Four molecules have predicted orange emission, of which all three have $\Delta E_{ST} < 0.3$ eV (**648**, **654** and **1496**), while f_{S_1} values range from 0.005 to 0.08. There are seven blue/violet emitters, ranging from cyan to violet, of which **1379** and **1738** are the only ones with $\Delta E_{ST} < 0.3$ eV, with f_{S_1} values of 0.003 and 0.04 respectively. Four of these, including **1379**, are significantly smaller than the parent molecule. Four molecules have $\Delta E_{ST} < 0.1$ eV, with f_{S_1} ranging from 4×10^{-4} to 0.02; all with yellow predicted emission. Five have $0.1 \leq \Delta E_{ST} < 0.2$ eV, with f_{S_1} ranging from 0.01 to 0.09, and again mostly yellow emission. There are 32 molecules with $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$, ranging from green to yellow emission. There are seven molecules with smaller ΔE_{ST} and larger f_{S_1} than BN3, all with a Tanimoto similarity above 0.8 with respect to BN3, and predicted green emission.

4.3.2.9 2F-BN. The predicted emission wavelength for 2F-BN is slightly underestimated, being a bright blue rather than green. Most child molecules are therefore also predicted to be blue, with many also being green. There is one orange outlier (**234**) and one clear yellow outlier (**1971**), with **1971** having smaller ΔE_{ST} than 2F-BN (0.31 eV), and $f_{S_1} = 0.08$. There are also some green-yellow emitters. There are two violet outliers (**743** and **1521**), which both have larger ΔE_{ST} than 2F-BN. Six molecules have $\Delta E_{ST} < 0.1$ eV, with f_{S_1} ranging from 6×10^{-4} to 0.02, and blue or green emission. Nine have $0.1 \leq \Delta E_{ST} < 0.2$ eV, with f_{S_1} ranging from 0.001 to 0.13, all of which are predicted to be green, and two of which feature S heteroatoms. Overall, five molecules have $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$, all of which are blue or green. There are eleven molecules with smaller ΔE_{ST} and larger f_{S_1} than 2F-BN, all with predicted blue emission and a high Tanimoto similarity with respect to the parent molecule (> 0.74).

4.3.2.10 DBNO. Most child molecules are predicted to have green emission like DBNO, but there is a noticeable outlier in the red emission of **1290**. However it has a very small f_{S_1} (0.007), and slightly higher ΔE_{ST} than DBNO. The next most red-shifted child molecule is **578**, which is predicted to emit yellow/yellow-green light, and has a higher f_{S_1} (0.33) and smaller ΔE_{ST} compared to DBNO (0.27 eV). Three molecules are violet (**68**, **947** and **1902**); all have reasonable f_{S_1} (> 0.2), but higher ΔE_{ST} than DBNO. In common with **1290**, they are much smaller molecules



than DBNO. Two molecules (**216** and **1168**) have $\Delta E_{ST} < 0.1$ eV, but have correspondingly small f_{S_1} (0.01 and 0.002). Two molecules (**492** and **1663**) have $0.1 \leq \Delta E_{ST} < 0.2$ eV, with larger f_{S_1} (0.07 and 0.11); all four of these molecules are predicted to have green emission. Overall, four molecules have $\Delta E_{ST} < 0.3$ eV and $f_{S_1} > 0.1$; two are blue and two green. There are five emitters with smaller ΔE_{ST} and larger f_{S_1} than DBNO, all green and with high Tanimoto similarity with respect to the parent (> 0.88).

5 Conclusions

In this work we presented an HTVS workflow for identifying potential new TADF emitters, using the STONED algorithm to generate an initial library of child molecules based on random mutations to a set of parent molecules. Subsequent steps included initial structural and synthesizability filters, as well as quantum chemistry calculations. Benchmark calculations were performed to justify the use of intermediate filters based on the HOMO-LUMO gap and spatial overlap, leading to significant computational savings due to eliminating unsuitable molecules before the more expensive DFT geometry optimisations and TDDFT calculations. The STONED algorithm allows for a balance between exploring relevant chemical space and achieving diversity, but it proved to be ineffective at rediscovering existing TADF emitters. Nonetheless, the workflow was able to identify a number of molecules worthy of further investigation, without any bias towards existing design rules. In future, it might be interesting to combine random mutations with a more targeted approach, e.g. guiding mutations toward molecular frameworks with established TADF characteristics or integrating domain-specific design principles into the workflow.

The workflow was applied to twenty known TADF parent emitters, split between D-A and MR emitters. While the number of molecules passing all filters varied significantly across parent molecules, in each case a number of potentially promising emitters were identified, covering a range of predicted emission colours. In total 171 D-A and 111 MR molecules were predicted to have both smaller ΔE_{ST} and larger f_{S_1} than their respective parent molecules. The data produced in this work also offers the possibility of being used to explore relationships between atomic structure and target properties which go beyond known design rules, potentially also supplemented by further runs with additional parent molecules, to generate a larger dataset.

The current work focused on using robust approaches with relatively low computational cost, in order to screen large numbers of molecules. In order to predict ΔE_{ST} , f_{S_1} and emission wavelengths more accurately and thus determine which might be worth exploring experimentally, it would be desirable to perform further calculations in future using a higher level of theory for the most promising candidate molecules. This would include for example using tuned range-separated hybrid functionals for the TDDFT calculations of D-A emitters, alternative methods beyond TDDFT for the MR emitters, and excited state geometry optimisations for both classes.

Author contributions

Kritam Thapa: Conceptualisation, Methodology, Software, Investigation, Writing – review and editing. Jennifer Jones: Methodology, Software, Investigation, Data curation, Writing – original draft. Laura Ratcliff: Conceptualisation, Methodology, Software, Investigation, Resources, Data curation, Writing – original draft, Visualisation, Supervision.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including relaxed molecular geometries and calculated properties, are available at Gitlab at <https://gitlab.com/lratcliff/tadf-htvs>.

Acknowledgements

LER acknowledges support from an EPSRC Early Career Research Fellowship (EP/P033253/1). KT acknowledges support from the Engineering and Physical Sciences Research Council (EP/S515085/1). We are also grateful for computational support from the UK national high performance computing service, ARCHER2, for which access was obtained via the UKCP consortium and funded by EPSRC grant ref EP/X035891/1. Parts of this work were also carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol – <http://www.bristol.ac.uk/acrc/>. The authors would like to thank Louis Beal, William Dawson and Luigi Genovese, for their technical input, including advice on the remote submission workflows.

Notes and references

- 1 G. Hong, X. Gan, C. Leonhardt, Z. Zhang, J. Seibert, J. M. Busch and S. Bräse, *Adv. Mater.*, 2021, **33**, 2005630.
- 2 D. Nayak and R. B. Choudhary, *Microelectron. Reliab.*, 2023, **144**, 114959.
- 3 J. Eng and T. J. Penfold, *Chem. Rec.*, 2020, **20**, 831–856.
- 4 H. Sun, C. Zhong and J.-L. Brédas, *J. Chem. Theory Comput.*, 2015, **11**, 3851–3858.
- 5 R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. K. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. Chae, M. Einzinger, D.-G. Ha, T. C. Wu, G. Markopoulos, S. O. 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**, 1120–7.
- 6 Y. Gao, S. Wu, G. Shan and G. Cheng, *Micromachines*, 2022, **13**.
- 7 T. J. P. Fernando B Dias and A. P. Monkman, *Methods Appl. Fluoresc.*, 2017, **5**.
- 8 T. J. Penfold, F. B. Dias and A. P. Monkman, *Chem. Commun.*, 2018, **54**, 3926–3935.
- 9 H. Uoyama, K. Goushi, K. Shizu, H. Nomura and C. Adachi, *Nature*, 2012, **492**, 234–238.
- 10 H. J. Kim and T. Yasuda, *Adv. Opt. Mater.*, 2022, **10**, 2201714.
- 11 H. Nakanotani, Y. Tsuchiya and C. Adachi, *Chem. Lett.*, 2021, **50**, 938–948.



- 12 T. Hatakeyama, K. Shiren, K. Nakajima, S. Nomura, S. Nakatsuka, K. Kinoshita, J. Ni, Y. Ono and T. Ikuta, *Adv. Mater.*, 2016, **28**, 2777–2781.
- 13 K. Zhao, Ö. H. Omar, T. Nematiram, D. Padula and A. Troisi, *J. Mater. Chem. C*, 2021, **9**, 3324–3333.
- 14 W.-P. To, G. Cheng, G. S. M. Tong, D. Zhou and C.-M. Che, *Front. Chem.*, 2020, **Volume 8 - 2020**.
- 15 A. Farokhi, S. Lipinski, L. M. Cavinato, H. Shahroosvand, B. Pashaei, S. Karimi, S. Bellani, F. Bonaccorso and R. D. Costa, *Chem. Soc. Rev.*, 2025, **54**, 266–340.
- 16 V. Ferraro, C. Bizzarri and S. Bräse, *Adv. Sci.*, 2024, **11**, 2404866.
- 17 L. Yao, B. Yang and Y. Ma, *Sci. China Chem.*, 2014, **57**, 335–345.
- 18 Y. Xu, X. Liang, X. Zhou, P. Yuan, J. Zhou, C. Wang, B. Li, D. Hu, X. Qiao, X. Jiang, L. Liu, S.-J. Su, D. Ma and Y. Ma, *Adv. Mater.*, 2019, **31**, 1807388.
- 19 Y. Xu, P. Xu, D. Hu and Y. Ma, *Chem. Soc. Rev.*, 2021, **50**, 1030–1069.
- 20 Y. Olivier, J.-C. Sancho-Garcia, L. Muccioli, G. D'Avino and D. Beljonne, *J. Phys. Chem. Lett.*, 2018, **9**, 6149–6163.
- 21 E. Runge and E. K. U. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.
- 22 K.-H. Lin, G.-J. A. H. Wetzelaer, P. W. M. Blom and D. Andrienko, *Front. Chem.*, 2021, **9**, 800027.
- 23 S. An, Y. H. Jung, G. Nam, E. Jeon, J. H. Ham, S. C. Cha, M. Y. Chae, J. H. Kwon and Y. Jung, *Chem. Eng. J.*, 2025, **505**, 159697.
- 24 C. Tu, W. Huang, S. Liang, K. Wang, Q. Tian and W. Yan, *RSC Adv.*, 2022, **12**, 30962–30975.
- 25 Y. Shu and B. G. Levine, *J. Chem. Phys.*, 2015, **142**, 104104.
- 26 Y. Bu and Q. Peng, *J. Phys. Chem. C*, 2023, **127**, 23845–23851.
- 27 Z. Tan, Y. Li, Z. Zhang, X. Wu, T. Penfold, W. Shi and S. Yang, *ACS Omega*, 2022, **7**, 18179–18188.
- 28 M. Ge, L. Xu, W. Shao, H. Jin, M. Liu and Q. Wang, *ChemRxiv*, 2024.
- 29 A. Nigam, R. Pollice, M. Krenn, G. d. P. Gomes and A. Aspuru-Guzik, *Chem. Sci.*, 2021, **12**, 7079–7090.
- 30 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.
- 31 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
- 32 Ö. H. Omar, M. del Cueto, T. Nematiram and A. Troisi, *J. Mater. Chem. C*, 2021, **9**, 13557–13583.
- 33 C. Rupakheti, R. Al-Saadon, Y. Zhang, A. M. Virshup, P. Zhang, W. Yang and D. N. Beratan, *J. Chem. Theory Comput.*, 2016, **12**, 1942–1952.
- 34 M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N. C. Frey, P. Friederich, T. Gaudin, A. A. Gayle, K. M. Jablonka, R. F. Lameiro, D. Lemm, A. Lo, S. M. Moosavi, J. M. Nápoles-Duarte, A. Nigam, R. Pollice, K. Rajan, U. Schatzschneider, P. Schwaller, M. Skreta, B. Smit, F. Strieth-Kalthoff, C. Sun, G. Tom, G. Falk von Rudorff, A. Wang, A. D. White, A. Young, R. Yu and A. Aspuru-Guzik, *Patterns*, 2022, **3**, 100588.
- 35 E. R. Sauvé, D. M. Mayder, S. Kamal, M. S. Oderinde and Z. M. Hudson, *Chem. Sci.*, 2022, **13**, 2296–2302.
- 36 J. W. Godden, L. Xue and J. Bajorath, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 163–166.
- 37 D. Rogers and M. Hahn, *J. Chem. Inf. and Model.*, 2010, **50**, 742–754.
- 38 RDKit: Open-source Cheminformatics, <https://www.rdkit.org>, <https://www.rdkit.org>.
- 39 G. Skoraczynski, M. Kitlas, B. Miasojedow and A. Gambin, *J. Cheminform.*, 2023, **15**, 6.
- 40 C. W. Coley, L. Rogers, W. H. Green and K. F. Jensen, *J. Chem. Inf. Model.*, 2018, **58**, 252–261.
- 41 A. Thakkar, V. Chadimová, E. J. Bjerrum, O. Engkvist and J.-L. Reymond, *Chem. Sci.*, 2021, **12**, 3339–3349.
- 42 M. Voršilák, M. Kolář, I. Čmelo and D. Svozil, *J. Cheminform.*, 2020, **12**, 35.
- 43 P. Ertl and A. Schuffenhauer, *J. Cheminform.*, 2009, **1**, 8.
- 44 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. I. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
- 45 T. A. Halgren, *J. Comp. Chem.*, 1996, **17**, 490–519.
- 46 T. A. Halgren, *J. Comp. Chem.*, 1996, **17**, 520–552.
- 47 T. A. Halgren, *J. Comp. Chem.*, 1996, **17**, 553–586.
- 48 T. A. Halgren and R. B. Nachbar, *J. Comp. Chem.*, 1996, **17**, 587–615.
- 49 T. A. Halgren, *J. Comp. Chem.*, 1996, **17**, 616–641.
- 50 T. Lewis-Atwell, P. A. Townsend and M. N. Grayson, *Tetrahedron*, 2021, **79**, 131865.
- 51 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- 52 T. J. Penfold, *J. Phys. Chem. C*, 2015, **119**, 13535–13544.
- 53 L. E. Ratcliff, W. Dawson, G. Fiscaro, D. Caliste, S. Mohr, A. Degomme, B. Videau, V. Cristiglio, M. Stella, M. D'Alessandro, S. Goedecker, T. Nakajima, T. Deutsch and L. Genovese, *J. Chem. Phys.*, 2020, **152**, 194110.
- 54 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710.
- 55 C. Hartwigsen, S. Goedecker and J. Hutter, *Phys. Rev. B*, 1998, **58**, 3641–3662.
- 56 A. Willand, Y. O. Kvashnin, L. Genovese, Á. Vázquez-Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch and S. Goedecker, *J. Chem. Phys.*, 2013, **138**, 104109.
- 57 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 58 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 59 M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus and W. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477–1489.
- 60 S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.*, 1999, **314**, 291–299.
- 61 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 62 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.



- 63 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
- 64 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 65 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 66 D. Hall, J. C. Sancho-García, A. Pershin, G. Ricci, D. Beljonne, E. Zysman-Colman and Y. Olivier, *J. Chem. Theory Comput.*, 2022, **18**, 4903–4918.
- 67 L. Kunze, T. Froitzheim, A. Hansen, S. Grimme and J.-M. Mewes, *J. Phys. Chem. Lett.*, 2024, **15**, 8065–8077.
- 68 W. Dawson, L. Beal, L. E. Ratcliff, M. Stella, T. Nakajima and L. Genovese, *Electron. Struct.*, 2024, **6**, 027003.
- 69 M. C. Burger, *J. Cheminform.*, 2015, **7**, 35.
- 70 G. H. Kim, R. Lampande, J. B. Im, J. M. Lee, J. Y. Lee and J. H. Kwon, *Mater. Horiz.*, 2017, **4**, 619–624.
- 71 H. Wang, L. Xie, Q. Peng, L. Meng, Y. Wang, Y. Yi and P. Wang, *Adv. Mater.*, 2014, **26**, 5198–5204.
- 72 Y. Fu, H. Liu, B. Z. Tang and Z. Zhao, *Nat. Commun.*, 2023, **14**, 2019.
- 73 K.-C. Pan, S.-W. Li, Y.-Y. Ho, Y.-J. Shiu, W.-L. Tsai, M. Jiao, W.-K. Lee, C.-C. Wu, C.-L. Chung, T. Chatterjee, Y.-S. Li, K.-T. Wong, H.-C. Hu, C.-C. Chen and M.-T. Lee, *Adv. Funct. Mater.*, 2016, **26**, 7560–7571.
- 74 F.-X. Huang, H.-Z. Li, F.-M. Xie, X.-Y. Zeng, Y.-Q. Li, Y.-Y. Hu, J.-X. Tang and X. Zhao, *Dyes Pigm.*, 2021, **195**, 109731.
- 75 H. Tanaka, K. Shizu, H. Miyazaki and C. Adachi, *Chem. Commun.*, 2012, **48**, 11392–11394.
- 76 S. Huang, Q. Zhang, Y. Shiota, T. Nakagawa, K. Kuwabara, K. Yoshizawa and C. Adachi, *J. Chem. Theory Comput.*, 2013, **9**, 3872–3877.
- 77 J. Lee, K. Shizu, H. Tanaka, H. Nakanotani, T. Yasuda, H. Kaji and C. Adachi, *J. Mater. Chem. C*, 2015, **3**, 2175–2181.
- 78 T.-A. Lin, T. Chatterjee, W.-L. Tsai, W.-K. Lee, M.-J. Wu, M. Jiao, K.-C. Pan, C.-L. Yi, C.-L. Chung, K.-T. Wong and C.-C. Wu, *Adv. Mater.*, 2016, **28**, 6976–6983.
- 79 G. Méhes, H. Nomura, Q. Zhang, T. Nakagawa and C. Adachi, *Ange. Chem.*, 2012, **51**, 11311–11315.
- 80 N. Ikeda, S. Oda, R. Matsumoto, M. Yoshioka, D. Fukushima, K. Yoshiura, N. Yasuda and T. Hatakeyama, *Adv. Mater.*, 2020, **32**, 2004072.
- 81 S. Wu, W. Li, K. Yoshida, D. Hall, S. Madayanad Suresh, T. Sayner, J. Gong, D. Beljonne, Y. Olivier, I. D. W. Samuel and E. Zysman-Colman, *ACS Appl. Mater. Interfaces*, 2022, **14**, 22341–22352.
- 82 S. Oda, B. Kawakami, R. Kawasumi, R. Okita and T. Hatakeyama, *Org. Lett.*, 2019, **21**, 9311–9314.
- 83 Y. Kondo, K. Yoshiura, S. Kitera, H. Nishi, S. Oda, H. Gotoh, Y. Sasada, M. Yanai and T. Hatakeyama, *Nat. Photonics*, 2019, **13**, 678–682.
- 84 S. Xu, Q. Yang, Y. Zhang, H. Li, Q. Xue, G. Xie, M. Gu, J. Jin, L. Huang and R. Chen, *Chin. Chem. Lett.*, 2021, **32**, 1372–1376.
- 85 T. Hua, L. Zhan, N. Li, Z. Huang, X. Cao, Z. Xiao, S. Gong, C. Zhou, C. Zhong and C. Yang, *Chem. Eng. J.*, 2021, **426**, 131169.
- 86 Y. Zhang, D. Zhang, J. Wei, Z. Liu, Y. Lu and L. Duan, *Angew. Chem. Int. Ed.*, 2019, **58**, 16912–16917.
- 87 X. Cai, J. Xue, C. Li, B. Liang, A. Ying, Y. Tan, S. Gong and Y. Wang, *Angew. Chem. Int. Ed.*, 2022, **61**, e202200337.



Data for this article, including relaxed molecular geometries and calculated properties, are available at Gitlab at <https://gitlab.com/lratchliff/tadf-htvs>

