

# Chemical Science

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: J. Shen, L. E. Walker, K. Ma, J. Green, H. Bronstein, K. Butler and T. J. H. Hele, *Chem. Sci.*, 2025, DOI: 10.1039/D5SC04276C.



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.

## Journal Name

## ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Learning Radical Excited States from Sparse Data<sup>†</sup>Jingkun Shen,<sup>a,‡</sup> Lucy E. Walker,<sup>b,c,‡</sup> Kevin Ma,<sup>a,‡</sup> James D. Green,<sup>a,‡</sup> Hugo Bronstein,<sup>b,c</sup> Keith T. Butler<sup>a</sup> and Timothy J. H. Hele<sup>\*a</sup>Received Date  
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Emissive organic radicals are currently of great interest for their potential use in the next generation of highly efficient organic light emitting diode (OLED) devices and as molecular qubits. However, simulating their optoelectronic properties is challenging, largely due to spin-contamination and the multiconfigurational character of their excited states. Here we present a data-driven approach where, for the first time, the excited electronic states of organic radicals are learned directly from experimental excited state data, using a much smaller amount of data than typically required by Machine Learning. We adopt ExROPPP, a fast and spin-pure semiempirical method for the calculation of the excited states of radicals, as a surrogate physical model for which we learn the optimal set of parameters. To achieve this we compile the largest known database of organic radical geometries and their UV-vis data, which we use to train our model. Our trained model gives root mean square and mean absolute errors for excited state energies of 0.24 and 0.16 eV respectively, improving hugely over ExROPPP with literature parameters. Four new organic radicals are synthesised and we test the model on their spectra, finding even lower errors and similar correlation as for the training set. This paves the way for the high throughput discovery of next generation radical-based optoelectronics.

## 1 Introduction

Recent years have shown a great interest in radicals for organic light emitting diodes (OLEDs), which display internal quantum efficiencies (IQE) of near 100% and intense emission in the deep red, near-infrared (NIR) and infrared (IR) spectral regions, features which are unusual and highly desirable.<sup>1–8</sup> These radical OLEDs, based on organic monoradicals, offer an alternative for the next generation of highly efficient lighting. Furthermore, optical readout of the quartet state of some radicals has potential applications in quantum information science and paves the way for next-generation molecular qubits.<sup>9</sup> However, historically many organic radicals have been nonemissive,<sup>10</sup> such that trial-and-error exploration of chemical space is inefficient, and there is therefore a large and unmet need for a method that facilitates the fast, accurate, and spin-pure calculation of the low-lying excited states of a wide variety of radical molecules. Such a computational method would also be invaluable for the high-

throughput screening of radicals for their UV-visible spectra. This work focuses on organic monoradicals, i.e. molecules with only one unpaired electron, however, it should be noted that the excited state properties of organic biradicals and organic radicals with many unpaired electrons are generally different to those discussed here.<sup>11,12</sup>

Calculating the excited electronic states of radicals is challenging due to spin-contamination and their multiconfigurational character. There exist several highly accurate methods for calculation of the excited states of radicals, such as multiconfigurational self-consistent field (MCSCF), complete active space perturbation theory to the 2<sup>nd</sup> order (CASPT2) and Coupled-cluster theory. However, these methods are very computationally expensive, making them unsuitable for high-throughput workflows.<sup>13–15</sup> Moreover, computationally cheaper methods such as conventional Time-Dependent Density Functional Theory (TD-DFT) can lead to spin-contaminated and functional-dependent results for the electronically excited states of radicals.<sup>16,17</sup> Additionally, it has been shown that for the most accurate calculation of excited state energies one must also include nuclear quantum effects.<sup>18</sup> Recently, an alternative, semiempirical method was developed — ExROPPP (Extended Restricted Open-shell Pariser-Parr-Pople theory), which is significantly faster, yet as accurate as higher level methods for calculating excited states of hydrocarbon radicals.<sup>19</sup> ExROPPP is based on the Pariser-Parr-Pople (PPP) Hamiltonian<sup>20–24</sup> with a subsequent Extended Configuration Interaction Singles (XCIS)<sup>25</sup> calculation which ensures spin

<sup>a</sup> Department of Chemistry, Christopher Ingold Building, University College London, WC1H 0AJ, UK.

<sup>\*</sup> E-mail: t.hele@ucl.ac.uk

<sup>b</sup> Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW, UK.

<sup>c</sup> Department of Physics, Cavendish Laboratory, Cambridge University, Cambridge CB3 0HF, UK.

<sup>‡</sup> These authors contributed equally.

<sup>†</sup> Supplementary Information available: details of database collation, the training and validation processes, optimized parameters, parameterization schemes and experimental synthesis and spectra. See DOI: 00.0000/00000000.



purity. PPP theory has gained recent popularity for predicting electronic properties at a significantly reduced computational cost.<sup>19,26–30</sup> Being a semiempirical method, PPP theory and consequently ExROPPP requires parameters which must be specified at the start of a calculation.<sup>20–24</sup> The carbon atom PPP parameters already existing in the literature have been shown to be successful for predicting excited state energies for hydrocarbons in ExROPPP.<sup>19</sup> However, emissive radicals commonly contain nitrogen and chlorine atoms, and we are not aware of any consistent, unified and widely-accepted set of parameters for including heteroatoms such as these in PPP theory or ExROPPP.<sup>24,31,32</sup> The advent of ExROPPP has opened up the possibility of rapid screening of the electronically excited states of radicals, however, extending and generalizing this method requires an optimal set of parameters to be found.

In recent years Machine Learning (ML) has become an indispensable tool for the study of chemical systems.<sup>33</sup> Such models allow for accurate prediction of chemical and physical properties with huge computational savings compared to methods such as DFT, provided sufficient data are available, and are often seen as an alternative to semiempirical and classical force-field methods.<sup>34</sup> ML has seen numerous applications in predicting energies, structures and reactivity patterns of molecules and materials.<sup>33–39</sup> Furthermore it has been applied to calculating the excited states of molecules and simulating excited-state potential energy surfaces.<sup>34,40–46</sup> However, while a wealth of previous studies have been successful for closed-shell species, we find very few examples of ML for the electronically excited states of radicals. In recent work a machine-learned density functional, ML- $\omega$ PBE, previously trained on closed-shell species, was applied to calculate the electronically excited states of radicals in TD-DFT showing good accuracy with a mean absolute error (MAE) for excitation energies of less than 0.3 eV.<sup>47</sup> However, to the best of our knowledge, learning the excited states of organic radicals directly from their experimental excited state data has yet to be attempted. Furthermore, typical ML models, in which no strong priors about the system are assumed at the outset, generally require large amounts of data, e.g. the properties of thousands of molecules or more, in order to be successful<sup>33,34,36</sup> and we are unaware of any such large datasets for the excited states of radicals. In addition, predicting electronically excited states using ML is challenging as they are a largely non-local property which cannot in general be treated using atom-wise descriptors.<sup>43</sup> Moreover, the prediction of primary outputs of quantum chemistry such as the  $N$ -electron wavefunction (and thus the composition of excited electronic states) is a highly desirable feature of an ML model, yet few ML models can predict these quantities for excited states.<sup>43</sup> There has, however, been much recent interest in leveraging wavefunction-based descriptors in quantum ML models allowing them to retain some of the physical intuition of conventional quantum chemistry, and this work follows in a similar spirit to these developments.<sup>48–52</sup>

Due to the lack of sufficient excited state data available in databases for organic radicals and the aforementioned challenges, adopting a trusted physical model such as ExROPPP and learning its optimal parameters may be a viable alternative to conventional ML for the excited states of radicals.<sup>19,26,27,53–59</sup> Us-

ing such a model also allows for the direct prediction of a variety of primary quantum chemical quantities such as molecular orbitals and transition dipole moments. In this paper we will focus on predicting molecular UV-visible linear absorption spectra and leave the computation of emission spectra, which usually requires excited state geometries that are difficult to acquire, for future research.

The linear UV-visible absorption spectra of organic radicals are usually characterised by two main features. These are an intense absorption (or absorptions) in the UV, usually between 300–400nm, and a much weaker absorption in the visible (or near infra-red). The weak visible  $D_1$  state has been investigated using various levels of theory (TD-DFT, PPP, MCSCF).<sup>2,9,19</sup> For alternant hydrocarbons (where the atoms can be divided into two groups such that no two atoms of the same group are directly bonded<sup>60</sup>),  $D_1$  is a minus combination  $|\Psi_{i0}^- \rangle$  of the HOMO-SOMO and SOMO-LUMO excitations (where the SOMO is the singly-occupied molecular orbital) and is essentially dark in the absorption spectrum.<sup>6,19,60–62</sup> Alternant hydrocarbons are usually non-emissive as non-radiative processes outcompete fluorescence.<sup>6,18</sup> Conversely, in non-alternant molecules, the  $D_1$  state may have significantly higher absorption intensity. One widely explored class of non-alternant radicals are those with a donor-acceptor structure, such as TTM-1Cz, in which the  $D_1$  state is bright, charge transfer (CT) in nature and is mostly composed of the HOMO-SOMO excitation.<sup>1,2,9</sup> These radicals are also highly emissive and have been incorporated into high-performing OLEDs.<sup>1,2</sup> Another class of emissive radicals have recently been discovered which lack a donor-acceptor structure and CT characteristics, but instead employ mesityl groups leading to a large increase in the photoluminescence quantum yield. However, mesityl substitution does not significantly affect the absorption characteristics.<sup>5</sup>

In this paper we learn the excited states of organic radicals directly from their experimental data for the first time. We choose to train on experimental excited state data rather than the results of high-level calculations as we are unaware of any databases of high-level calculations (e.g. GMC-QDPT/CASPT2) of organic radicals, and also we aim to directly replicate experimental UV-Visible spectra. To achieve this we use a modest amount of published UV-visible absorption data to learn an optimal set of ExROPPP parameters for organic radicals containing carbon, hydrogen, nitrogen and chlorine. Despite only containing a modest amount of data, we believe our compiled database of UV-Visible absorption data of  $\pi$ -conjugated organic radicals to be the largest of its kind.<sup>47</sup> Four new radicals are synthesised and we test our model on their absorption spectra to demonstrate its predictability and transferability. We find that the trained model has a significantly higher accuracy than the model using parameters taken from the literature and is able to make accurate predictions about the electronic excited states of unseen molecules.

## 2 Methodology

### 2.1 Data collection

We obtained spectroscopic data for 81 organic radicals from previously published work whose structures are given in Fig. 1 and



Fig. 2.<sup>2,5–7,9,47,63–71</sup> In order to compile a database of suitable radicals, we considered all radicals we could find in the literature containing only carbon, hydrogen, chlorine and pyrrole, aniline and pyridine type nitrogen atoms. Those radicals whose spectroscopic absorption data could be found were added to the database along with their data. We also obtained DFT optimised molecular geometries for these molecules from previous studies.<sup>2,5,9,47,64</sup> However, the molecular geometries for some molecules could not be found in the literature so most of these structures were optimised using unrestricted DFT with a gradient tolerance of 0.0001 E<sub>H</sub>/a<sub>0</sub> in GAMESS-US.<sup>72</sup> These data constitute the training set of the ML ExROPPP model. Further details pertaining to the collation of the database and geometry calculations can be found in SI Section I.

The target properties used for training are the energies  $E_{D_1}$  of the the first excited doublet states ( $D_1$ ), the energies  $E_{\text{brt}}$  of the brightest absorptions in the UV-visible spectra, and  $I_{D_1}^{\text{rel}} = \epsilon_{D_1}/\epsilon_{\text{brt}}$  which is the ratio of the molar extinction coefficients of these the two absorptions, extracted from linear UV-visible absorption spectra. These target properties are similar to those previously used in ML of molecular spectra<sup>34</sup>. Exceptions are made for TTM-1Cz-An and TTM-1Cz-PhAn whereby due to their unusual electronic structure, their first excited doublet state ( $D_1$ ) is a dark triplet-coupled doublet state and the lowest energy bright doublet state is  $D_2$  which is by majority composed of the carbazole HOMO–TTM SOMO excitation (same orbital parentage as  $D_1$  in a typical donor-acceptor radical such as TTM-1Cz).<sup>9</sup> Therefore, for these exceptional molecules we fit the  $D_2$  (instead of  $D_1$ ) energy and oscillator strength to the corresponding lowest energy  $D_2$  absorption seen in experiment. For these two molecules, we will group this state ( $D_2$ ) in with the  $D_1$  states for all the other molecules when performing the statistical analyses.

## 2.2 Training

We train the ExROPPP model on experimental UV-visible data of known organic radicals, using a fitness function of the computed energies and intensities compared to those obtained from experiment, which quantifies how well the predictions of the ExROPPP model fit with the experimental data. The fitness function takes the form

$$f = w_{D_1}(E_{D_1,\text{calc.}} - E_{D_1,\text{exp.}})^2 + w_{\text{brt}}(E_{\text{brt,calc.}} - E_{\text{brt,exp.}})^2 + w_I(I_{D_1,\text{calc.}}^{\text{rel.}} - I_{D_1,\text{exp.}}^{\text{rel.}})^2, \quad (1)$$

where  $w_{D_1}$ ,  $w_{\text{brt}}$  and  $w_I$  are the weights of the three respective terms in Eq. (1). The weights of the first two terms have units of eV<sup>−2</sup> and  $w_I$  is dimensionless such that  $f$  is a dimensionless quantity. While there are many other fitness functions which we could use, such as those based on the theory of optimal transport (between experimental and calculated spectra), we choose to use the above function as it encapsulates the essential spectral information of organic radicals which we believe is most important to be able to predict and only requires a small amount of raw spectral data for each radical.<sup>73</sup>

Training is achieved by finding a set of ExROPPP parameters which minimises this fitness function utilising the derivative-free

Nelder-Mead optimiser in Python as shown in Fig. 3.<sup>74</sup> The algorithm first reads in the initial parameters, molecular geometries and experimental absorption data for all training molecules and classifies the molecules into hydrocarbons or heterocycles, which are treated slightly differently. For hydrocarbons, the fitness function comprises of only energy terms, with  $w_{D_1}$  and  $w_{\text{brt}}$  set to 1 eV<sup>−2</sup> and  $w_I$  set to zero as the  $D_1$  state for hydrocarbons gives a very weak absorption in experiment and in ExROPPP has zero oscillator strength.<sup>61</sup> For heterocycles,  $w_{D_1}$  and  $w_{\text{brt}}$  are set to 1 eV<sup>−2</sup> and  $w_I$  set to 1 (except for a few molecules whose bright state data could not be found in which case  $w_I$  set to zero, see radicals-spreadsheet.xlsx in Data Availability). Then the parameters are iteratively varied and the fitness calculated on each iteration until convergence.

We found that pre-training separately on subsets of molecules of different heteroatom types to obtain better initial guess parameters before training on all molecules for all parameters, an approach which we call ‘stratified’ training, leads to lower errors than not including these pre-training steps. The results of the trained model presented in the next section are obtained using this stratified approach. Details of the training process are discussed further in SI Section II. These calculations are parallelised for maximum efficiency.

## 2.3 Testing on novel radicals

Four new radicals: M<sub>2</sub>TTM-4Me, M<sub>2</sub>TTM-3PCz, M<sub>2</sub>TTM-3TPA and M<sub>2</sub>TTM-4TPA, shown in Fig. 6 were synthesised, their UV-visible absorption spectra were measured, and their minimum energy geometries were obtained using DFT. The spectroscopic data ( $E_{D_1}$ ,  $E_{\text{brt}}$  and  $I_{D_1}^{\text{rel.}}$ ) and molecular structures of these new molecules form a testing set for the ML ExROPPP model. The molecular geometries and extracted UV-visible absorption data of all molecules as well as initial and optimised sets of parameters are available on the UCL Research Data Repository (see Data Availability).

## 2.4 Statistical analysis

We calculate root mean-squared errors (RMSE), mean absolute errors (MAE), coefficients of determination R<sup>2</sup> and Spearman’s rank correlation coefficients (SRCC) between the experimental data and the simulated data for the training and testing set (see SI Section IV C).

## 2.5 PPP theory

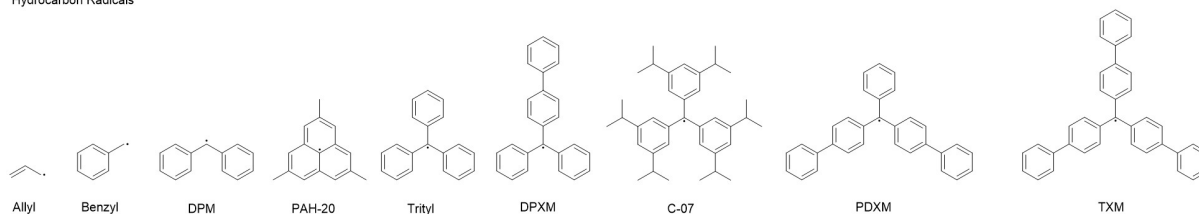
Here we briefly introduce Pariser-Parr-Pople (PPP) theory before describing its extension to the ExROPPP method used as a surrogate model in this article.<sup>20–22,75</sup>

The PPP model extends semi-empirical Hückel theory by explicitly including  $\pi$  electron correlation, thereby reproducing the low-lying excited states of conjugated chromophores with markedly greater fidelity.<sup>20–22,75</sup> The three core approximations that underpin PPP are: (i)  $\sigma$ – $\pi$  separation, which treats the  $\sigma$  framework as an inert, non-polarisable core; (ii) zero differential overlap (ZDO), approximating the overlap matrix **S** as the identity matrix and neglecting three- and four-centre two-electron terms; and

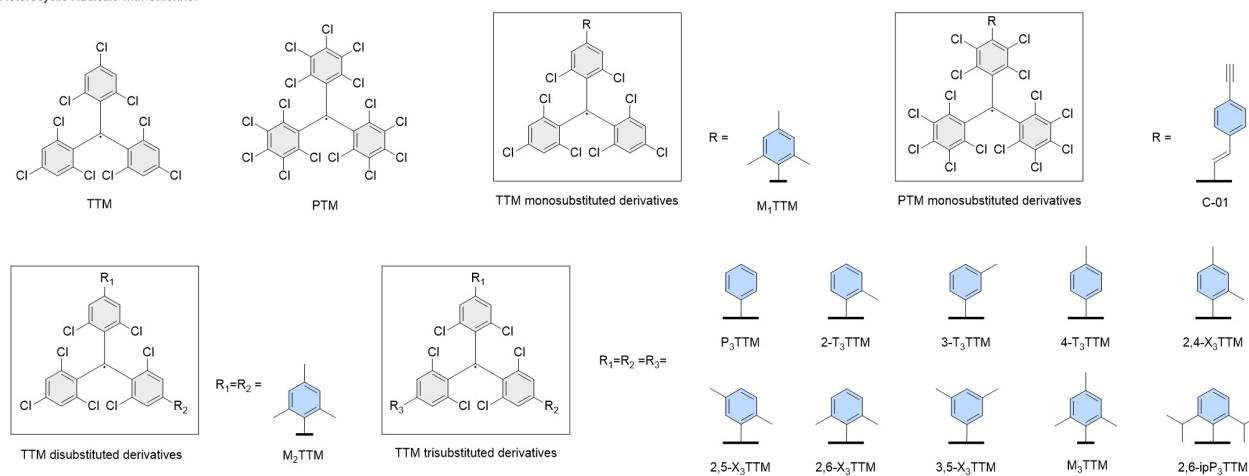




## Hydrocarbon Radicals



## Heterocyclic Radicals with Chlorine:



## Heterocyclic Radicals with Chlorine and Pyrrole-Type Nitrogen:

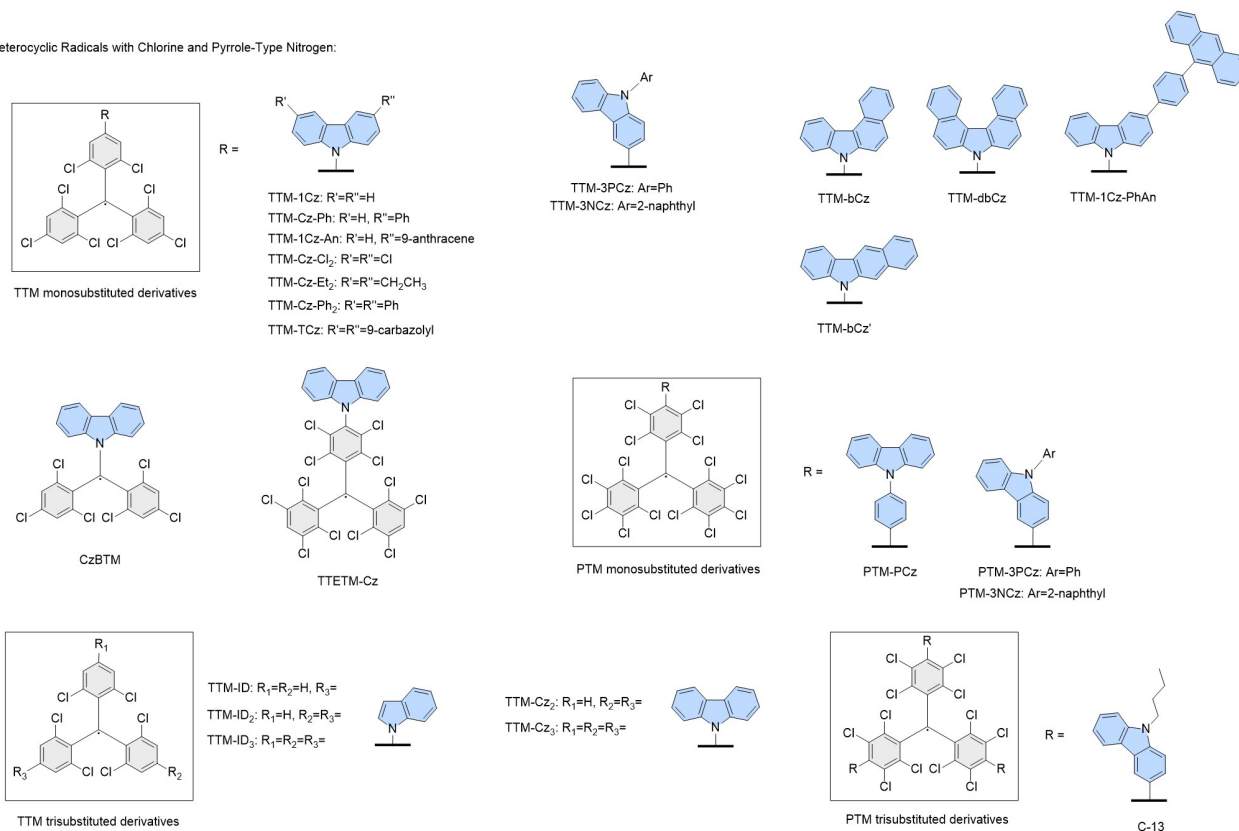
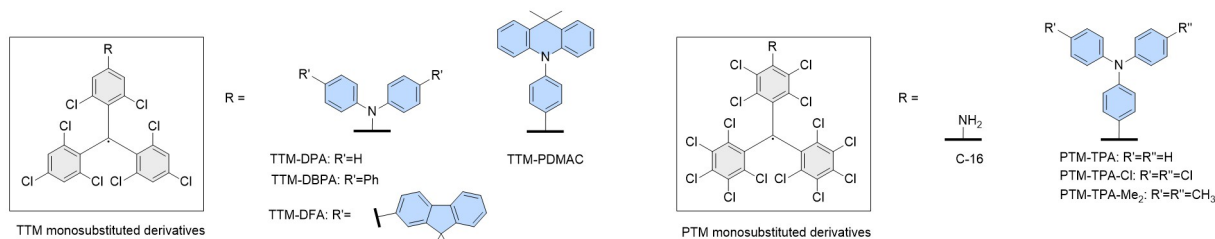


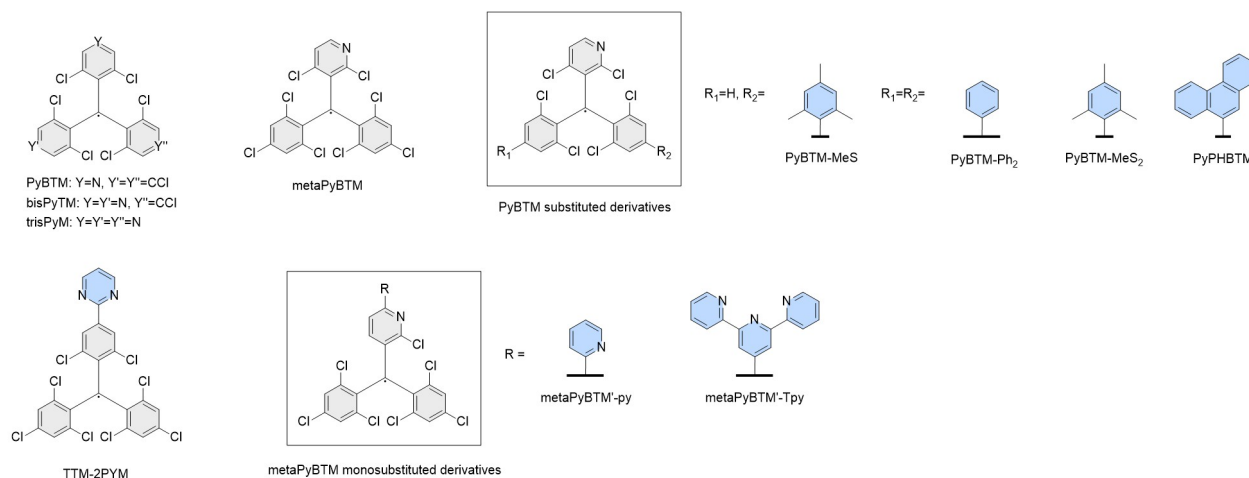
Fig. 1 Structures of the molecules in the training set containing carbon, hydrogen, chlorine and pyrrole-type nitrogen. The central TTM/PTM backbone is coloured in grey, and substituents colored in light blue.



## Heterocyclic Radicals with Chlorine and Aniline-Type Nitrogen:



## Heterocyclic Radicals with Chlorine and Pyridine-Type Nitrogen:



## Heterocyclic Radicals with Chlorine and Mixed Types of Nitrogen:

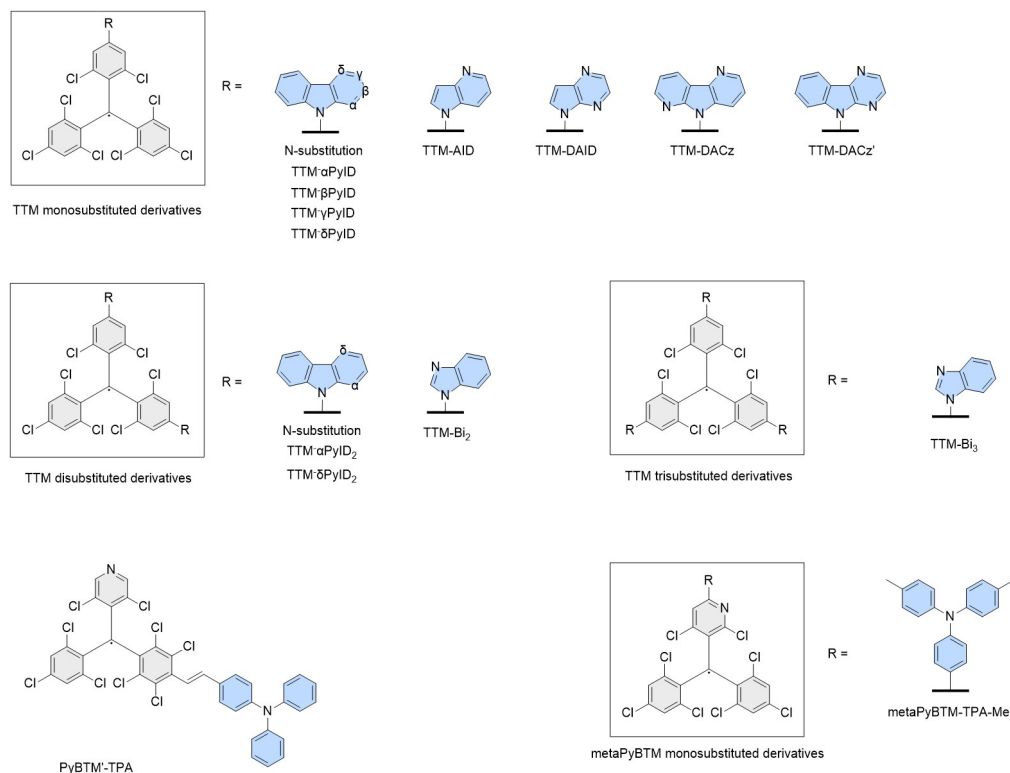


Fig. 2 Structures of the molecules in the training set containing aniline, pyridine and multiple types of nitrogen. The central TTM/PTM backbone is coloured in grey, and substituents colored in light blue.



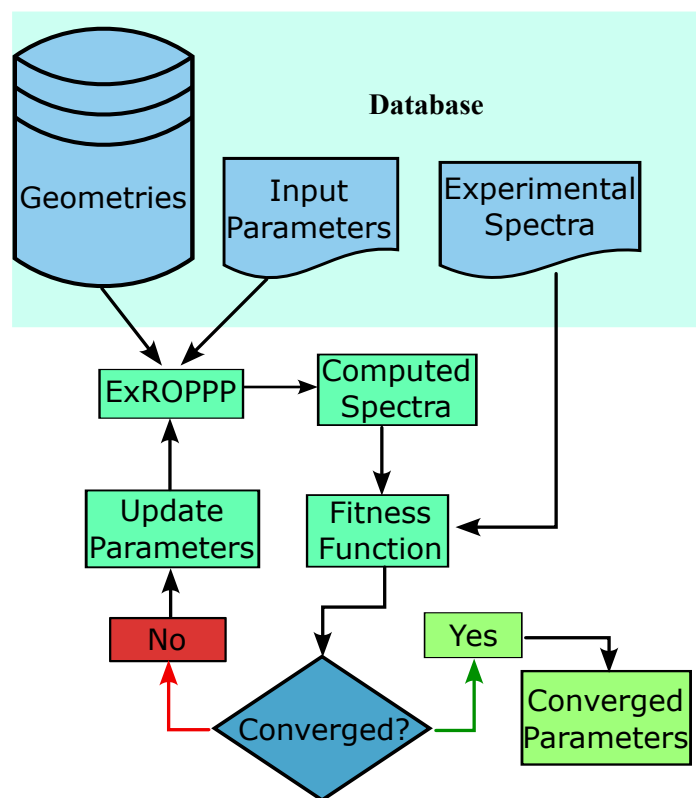


Fig. 3 A flow diagram illustrating our method for training our ExROPPP model on the experimental absorption data of organic radicals.

(iii) nearest-neighbour approximations, whereby the one-electron hopping integral  $t_{\mu\nu}$  is non-zero only for bonded pairs. To obtain the orbitals, Roothaan's equations are solved self consistently, which under ZDO reduce to:

$$\mathbf{FC} = \mathbf{CE} \quad (2)$$

with  $\mathbf{F}$  the Fock matrix,  $\mathbf{C}$  the molecular-orbital coefficient matrix, and  $\mathbf{E}$  the diagonal orbital-energy matrix.<sup>76</sup> Pople's formulation of PPP theory uses<sup>22</sup>:

$$F_{\mu\mu} = \epsilon_{\mu} + \frac{1}{2}P_{\mu\mu}\gamma_{\mu\mu} + \sum_{\nu \neq \mu} (P_{\nu\nu} - Z_{\nu})\gamma_{\mu\nu}, \quad (3)$$

$$F_{\mu\nu} = t_{\mu\nu} - \frac{1}{2}P_{\mu\nu}\gamma_{\mu\nu}. \quad (4)$$

where  $\epsilon_{\mu}$  is the one-electron on-site Coulomb term (Hückel  $\alpha$ ) that captures the interactions of the  $\pi$  electron with the frozen  $\sigma$  core,  $t_{\mu\nu}$  is the hopping (resonance) integral (Hückel  $\beta$ ) between neighbouring  $p$  orbitals,  $\gamma_{\mu\mu}$  and  $\gamma_{\mu\nu}$  are the one- and two-centre electron repulsion integrals that depend on the distance between atoms  $\mu$  and  $\nu$ , and  $Z_{\nu}$  is the effective nuclear charge.

## 2.6 PPP Parameterisation

We employ largely the same functional form of PPP theory as in previous work.<sup>19,24,26,27</sup>

**Coulomb term:** In this work, the Pariser-Parr form for  $\epsilon_{\mu}$  is

used:<sup>20</sup>

$$\epsilon_{\mu} = -\text{IP}_{\mu} \quad (5)$$

where  $\text{IP}_{\mu}$  is the valence-state ionization potential of atom  $\mu$ .

**Two electron integral:** To parameterise the two-electron integral, the Mataga-Nishimoto form expressed in atomic orbitals is invoked:<sup>24</sup>

$$\gamma_{\mu\nu} = \frac{U_{\mu\nu}}{1 + r_{\mu\nu}/r_{0,\mu\nu}} \quad (6)$$

with  $U_{\mu\nu} = \frac{1}{2}(U_{\mu\mu} + U_{\nu\nu})$  and  $r_{0,\mu\nu} = \frac{1}{2}(r_{0,\mu\mu} + r_{0,\nu\nu})$ . Here,  $U_{\mu\mu}$  is the Hubbard on-site repulsion and  $r_{\mu\nu}$  is the scalar distance between atoms  $\mu$  and  $\nu$ . This formulation was used previously to describe the two-electron integrals in PPP theory for both closed-shell<sup>26,27</sup> and radical<sup>19</sup> systems.

**Hopping integral:** Here we use a similar form of the hopping integral  $t_{\mu\nu}$  as used in previous studies,<sup>19,26,27</sup> however unlike Refs. 19,26,27, but similar to Ref. 77 we elect to use an exponentially decaying function for the hopping term. This term is also scaled by the cosine of the dihedral angle  $\theta$  similar to Refs. 19,26,27 to accommodate non-planar geometries,

$$t_{\mu\nu} = A_{\mu\nu} \exp(-b_{\mu\nu}r_{\mu\nu}) \cos \theta, \quad (7)$$

where  $A_{\mu\nu}$  and  $b_{\mu\nu}$  are parameters which are different for different pairs of atom types (see SI Section II).

**Effective nuclear charge:** The PPP parameter  $Z_{\mu}$  is the effective nuclear charge (of the nuclei and core electrons) experienced by the  $\pi$  electron(s) on atom  $\mu$ .<sup>20,26</sup> We employ effective nuclear charges  $Z_{\mu}$  of  $e$  for carbon and pyridine type nitrogen as they each contribute one electron to the  $\pi$ -system, and  $2e$  for chlorine and pyrrole/aniline type nitrogen as they each contribute two electrons to the  $\pi$ -system, where  $e$  is the electron charge.

**Atom specificity:** The PPP parameter  $\epsilon_X$  is atom-specific, whereas  $t_{XY}$  is atom pair-specific (where  $X$  and  $Y$  are now used to represent atom types rather than  $\mu$  and  $\nu$  which are atom indices). It would seem like  $U_{XY}$  and  $r_{0,XY}$  are also pair-specific, however, the convention of Ref. 24 used here is that two-electron interactions between dissimilar atoms use arithmetic averages, e.g.  $U_{XY} = \frac{1}{2}(U_{XX} + U_{YY})$  so there is only one unique  $U_{XX}$  and  $r_{0,XX}$  for every atom type  $X$ .

Different parameter sets are adopted for pyridine versus pyrrole/aniline nitrogens to reflect their different  $\pi$ -electron counts. Carbons and pyridine nitrogens donate one  $\pi$  electron, whereas chlorine and pyrrole/aniline nitrogens donate two via lone pairs. For further discussion of historical parameterisation strategies and alternative functional forms, see SI Section V.

## 2.7 ExROPPP method

Conceptually, ExROPPP combines the PPP Hamiltonian<sup>20–22,75</sup> with a restricted open-shell reference<sup>60</sup> and an extended configuration interaction singles (XCIS) scheme<sup>25</sup> to rapidly calculate spin-pure electronically excited states with appreciable accuracy, as demonstrated for hydrocarbon radicals.<sup>19</sup>

**Restricted open-shell reference:** Starting from a single-open-shell determinant  $|\Psi_0\rangle$  (with one Singly Occupied Molecular Or-



bital, SOMO), the molecular orbitals are optimised by minimising the total energy with respect to mixing of singly excited configurations<sup>19,59,60</sup>, yielding the restricted open-shell Fock operator

$$F_{\mu\nu}^{\text{ROHF}} = h_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[ (\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu) \right] \quad (8)$$

where  $h_{\mu\nu}$  is the one-electron integral and  $(\mu\nu|\lambda\sigma)$  a two-electron integral in the atomic orbital basis. The density matrix elements are

$$P_{\lambda\sigma} = \sum_{l=1}^k (2C_{\lambda l}^* C_{\sigma l}) + C_{\lambda 0}^* C_{\sigma 0} \quad (9)$$

with  $C_{\sigma l}$  the coefficient of atomic orbital  $\sigma$  that contributes to molecular orbital  $l$ , where  $l = 0$  corresponds to the SOMO,  $l = 1$  to the HOMO and so on<sup>19,59</sup>. Diagonalising  $F^{\text{ROHF}}$  provides the optimised orbitals  $\{\phi_p\}$  used in the PPP ground-state calculation and subsequent integral evaluations.

**Spin-adapted excitation basis:** Spin-pure doublet and quartet configurations are generated by linear combinations of single  $(|\Psi_i^{j'}\rangle, |\Psi_i^{\bar{j}'}\rangle)$  and double  $(|\Psi_{i0}^{\bar{j}j'}\rangle)$  excitations that are eigenfunctions of  $\hat{S}^2$ . For monoradicals with one unpaired electron in the SOMO, the spin-pure basis comprises<sup>19,59,60</sup>:

$$\{|\Phi_I\rangle\} = \{^2\Psi_0, ^2\Psi_0^{j'}, ^2\Psi_i^0\}, \quad (10)$$

$$^2\Psi_i^{j'}, ^2\Psi_i^{\bar{j}'}, ^4\Psi_i^{j'}\}, \quad (11)$$

where  $^2\Psi_i^{j'}$ ,  $^2\Psi_i^{\bar{j}'}$ ,  $^4\Psi_i^{j'}$  are, respectively, the singlet-coupled, triplet-coupled and quartet states obtained by diagonalizing  $\hat{S}^2$  matrix in the subspace spanned by  $|\Psi_i^{j'}\rangle, |\Psi_i^{\bar{j}'}\rangle$  and  $|\Psi_{i0}^{\bar{j}j'}\rangle$ .

**XCIS Hamiltonian:** Within this spin-pure basis the excited-state Hamiltonian is

$$H_{IJ}^{\text{XCIS}} = \langle \Phi_I | \hat{H}_{\text{PPP}} | \Phi_J \rangle, \quad (12)$$

where  $\hat{H}_{\text{PPP}}$  is built from the same one-electron on site terms  $\varepsilon_\mu$ , hoppings  $t_{\mu\nu}$ , and two-electron Coulomb terms  $\gamma_{\mu\nu}$  as defined in Sec. 2.6. Block-diagonalisation of  $H^{\text{XCIS}}$  yields excitation energies and eigenvectors. See the SI Eqs. 5–8 for calculation of the transition dipole moments and oscillator strengths in the simulated spectra.

**Typical Workflow:** ExROPPP (i) solves the restricted open-shell Roothaan equations with the PPP parameters of Sec. 2.5 to obtain  $\{\varphi_p\}$  and  $E_0$ ; (ii) constructs the spin-adapted XCIS manifold; (iii) Forms and diagonalises  $H^{\text{XCIS}}$ ; and (iv) evaluates transition dipole moments. Absorption spectra are generated by applying Lorentzian broadening (FWHM 20 nm at a reference wavelength of 300 nm) to the discrete  $H^{\text{XCIS}}$  excitation energies.

## 3 Results

### 3.1 Training and Validation

The results of training the 81 molecule model are summarized in Fig. 4 and Table 1. We find that the accuracy of the simulated excited state energies improves significantly on training. The RMSE

reduces from 0.86 eV with literature parameters to 0.24 eV for the trained model and the MAE reduces from 0.80 eV with literature parameters to 0.16 eV for the trained model. In terms of correlation, we find a marked improvement in  $R^2$  from  $-0.71$  to  $0.87$  and a smaller improvement for SRCC going from  $0.79$  to  $0.88$  in the trained model compared with the literature parameters.

The simulated spectra of the training set are more accurately reproduced by the trained model than by the model with literature parameters. To illustrate this we have included the spectra of two emissive radicals in the training set which are relevant to optoelectronics: TTM-1Cz and TTM-1Cz-An (see Fig. 5). TTM-1Cz is a prototypical and widely-studied emissive radical which has been implemented in functioning OLEDs, and is a good reference point for the trained ExROPPP model.<sup>1</sup> On the other hand, TTM-1Cz-An is an atypical organic radical which has a complex and unusual electronic structure owing to its first excited state being a quartet and should be a challenging test case for ExROPPP. TTM-1Cz-An has been investigated for potential applications in quantum information technology.<sup>9</sup>

We find that the trained model reproduces the  $D_1$  ( $D_2$  for Cz-An) energies of these molecules significantly more accurately than does the literature parameters. The accuracy for the bright states also improves with the trained parameters. Furthermore, ExROPPP predicts that the quartet state  $Q_1$  of TTM-1Cz-An is lower in energy than the lowest energy bright state  $D_2$  in both sets of parameters, in line with experimental data and higher-level calculations in the literature. However, the energy of the  $Q_1$  state of TTM-1Cz-An in trained ExROPPP (1.10eV) is quite far from the value obtained from higher-level calculations (2.26eV in NEVPT2) which is not surprising given that ExROPPP was never trained on quartet state energies.<sup>9</sup>

Trained ExROPPP does produce an extra absorption around 500nm for both molecules which is not seen in experiment, and this absorption arises due to ground state mixing of excited configurations (which is allowed for radicals as the form of ROHF we use does not exactly satisfy Brillouin's theorem).<sup>19,25,59,60</sup> When this mixing is set to zero in the calculation, the states responsible for these extra absorptions have similar energies but now have almost zero intensity in the spectrum (see SI Figures 3 and 4). Ground state mixing was not set to zero for the training process as the original implementation of ExROPPP included all Hamiltonian mixing terms and gave accurate energies and intensities for the excited states of hydrocarbon radicals.<sup>19</sup> This being said, it is common practice in XCIS to set these elements to zero and training could be repeated without ground state mixing for more accurate intensities.<sup>25</sup> It is also possible that the choice of fitness function allows a minimum point in the parameter space to be reached where the ground state mixing is too large and leads to inaccuracies in the predicted intensities of certain states in the UV-visible spectra. The fitness only depends on the energies and relative intensities of the  $D_1$  ( $D_2$  for Cz-An) state and the intense bright state and thus does not take into account other states in the spectrum. We leave further investigation of these spectral features for future research.

Despite the apparent inaccuracy for intensities, the ability of the trained ExROPPP model to accurately capture the excitation





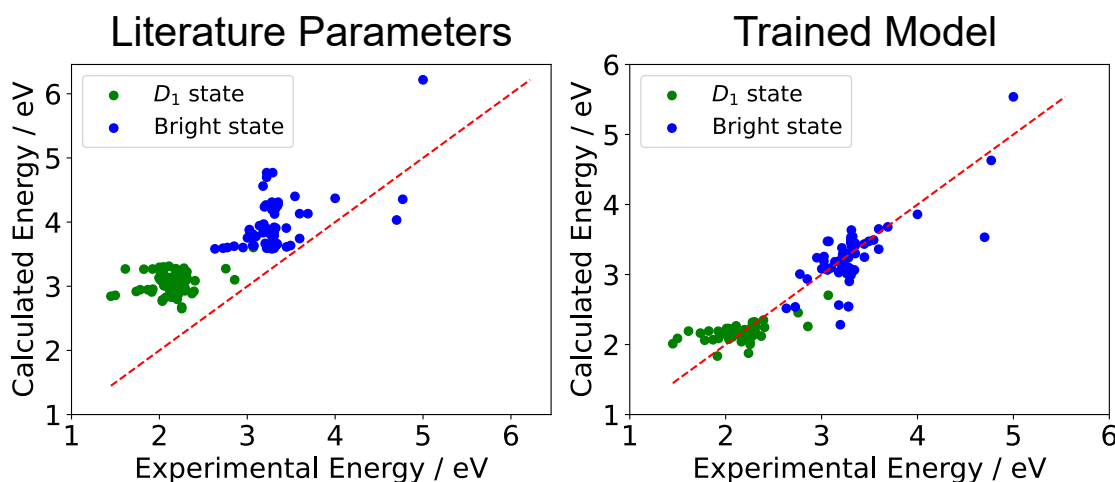


Fig. 4 Regression plots of the excited state energies of the 81 molecule training set calculated by ExROPPP compared with experimentally determined energies, using parameters obtained from the literature (left) and those of the trained ExROPPP model (right). The trained model predicts the energies of UV/Visible absorptions much closer to experiment (red line) than do the literature parameters.

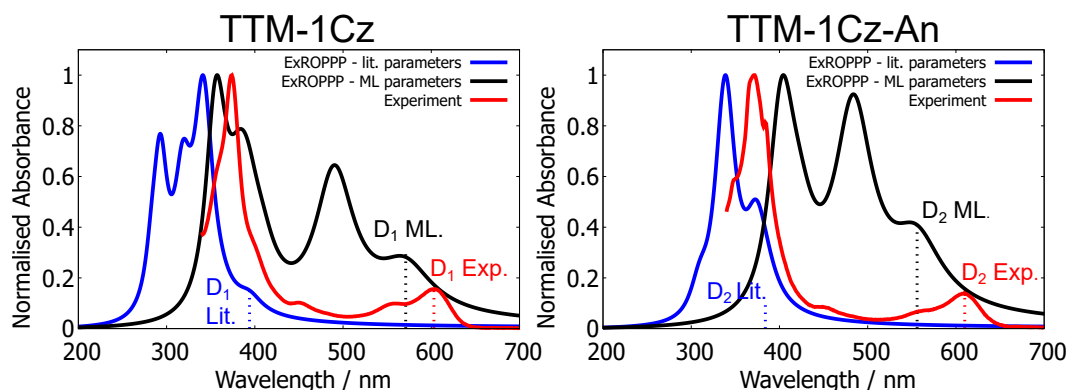


Fig. 5 UV-visible absorption spectra of TTM-1Cz and TTM-1Cz-An measured in 200  $\mu$ M toluene solution at room temperature (red), simulated using ExROPPP with literature parameters (blue) and with the trained 81 molecule model (black).<sup>9</sup> The trained model substantially improves on the literature parameters in both cases.

energies and excited state features of both typical and anomalous radicals shows its flexibility and robustness. In the SI Table XIV, we compute the fitness, MAE,  $R^2$  and SRCC including and excluding the two anomalous molecules (using the converged parameters from ML which includes the anomalous molecules) and find them to be extremely similar, indicating that their inclusion does not significantly affect the computed figures of merit.

In addition, we employed  $K$ -fold cross-validation on the 81-radical training set to validate the robustness of the model against data noise, using several different choices of folds. The average errors of the  $K$ -fold validation are similar to those for the whole training set in the trained model. In going from the trained model to the average  $K$ -fold model, the RMSE only slightly increases from 0.24 eV to 0.27 eV, the MAE goes from 0.16 eV to 0.18 eV, the  $R^2$  decreases slightly from 0.87 to 0.84 and the SRCC also decreases slightly from 0.88 to 0.86 (see Table 1). The data noise is also shown to be small for all trials. Artificially (selected by the user) and randomly distributed folds gave similar errors demonstrating that the distribution of molecules across folds does not significantly affect the model. Also, the  $K$ -fold results show that

Table 1 Total fitness, RMSE, MAE,  $R^2$  and SRCC for the trained ExROPPP model and for the  $K$ -fold validation (total fitness for all folds, errors averaged across all folds) are compared to ExROPPP with parameters obtained from the literature, calculated for all states in the training set of 81 organic radicals. For further details on  $K$ -fold validation see SI section III B.

	Literature parameters	Trained Model	$K$ -fold Model	Target
Total Fitness	117.44	10.00	15.21	
RMSE (all states)/eV	0.86	0.24	0.27	< 0.3
MAE (all states)/eV	0.80	0.16	0.18	< 0.3
$R^2$ (all states)	-0.71	0.87	0.84	close to 1
SRCC (all states)	0.79	0.88	0.86	close to 1



the stratified model has better robustness than the model without pre-training steps. Further details can be found in the SI Section III.

### 3.2 Transferability

Here we briefly consider the extent to which the trained ExROPPP model can reproduce the qualitative orbital structure of radicals. This is a particularly challenging test given that ExROPPP was not trained on orbital data.

A qualitative comparison of the singly occupied (SOMO) and highest occupied (HOMO) molecular orbitals from ExROPPP and Restricted Open-Shell DFT (B3LYP/6-31G(d,p)) demonstrates good overall agreement in both shape and localization, highlighting the transferability of the ExROPPP-based parametrisation. In TTM-1Cz, TTM-3NCz, and M<sub>3</sub>TTM, the SOMOs remain largely centered on the triarylmethyl carbon, with moderate extension onto the phenyl rings, whereas the HOMOs for TTM-1Cz and TTM-3NCz localize more strongly on the carbazole substituent. Notably, the local/global symmetry (e.g., C<sub>2</sub> for TTM-1Cz, C<sub>3</sub> for M<sub>3</sub>TTM) and corresponding irreps are preserved. These observations underscore the physical interpretability and transferability of parameters derived via an optimizer-assisted physics-informed model, which can extrapolate untrained properties. Minor discrepancies occur for the M<sub>3</sub>TTM HOMO, where ExROPPP underestimates the small orbital coefficients in the outer ring. Further details and figures illustrating these points are provided in SI Section IV E.

### 3.3 Novel organic radicals

To test our model, we synthesized four novel trityl radicals, specifically designed to probe various state-of-the-art concepts previously identified in mono-radical systems (see Fig. 6). Each radical was based on a mesitylated TTM framework, which has been shown to enhance photoluminescence quantum efficiency (PLQE) by augmenting the radiative decay rate.<sup>5</sup> To evaluate the ExROPPP model with an asymmetric structure and the absence of charge transfer (CT), toluene was appended to the unsubstituted site of the mesitylated trityl radical core through its 4-position. The three other radicals incorporated CT groups, namely 9-phenylcarbazole (PCz) and triphenylamine (TPA), which contain non-bonding nitrogen lone pairs. These non-bonding electrons have been shown to enhance photoluminescent efficiency through a reduction in excitonic coupling to high-frequency vibrational modes.<sup>78</sup> Through the inclusion of PCz and TPA moieties, we aimed to test the model across electron-donating groups of varying strengths, with TPA being the stronger donor due to the hybridization of its nitrogen heteroatom which influences lone pair availability. Additionally, TPA units were linked to the trityl radical core through both the 3- and 4-positions to assess accuracy in predicting spectroscopic outcomes for different stereoisomers. The combination of mesitylation and non-bonding CT groups provides a promising strategy for developing highly efficient radical emitters and it is crucial that the ExROPPP method can predict outcomes for these cutting-edge radical designs.

The synthesis of the radical species commenced with the

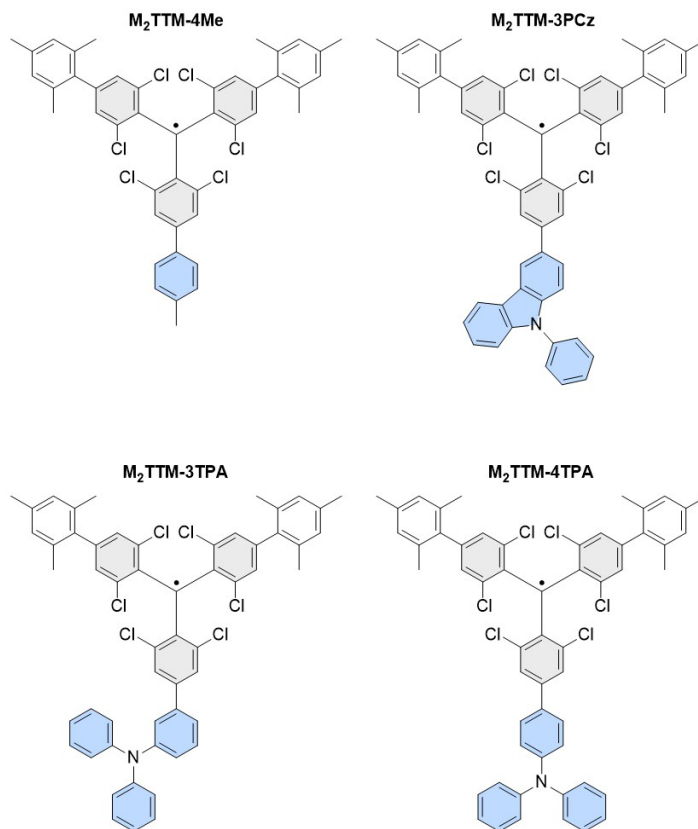


Fig. 6 Structures of the four newly synthesised radicals reported in this work: M<sub>2</sub>TTM-4Me, M<sub>2</sub>TTM-3PCz, M<sub>2</sub>TTM-3TPA and M<sub>2</sub>TTM-4TPA, which constitute the testing set.

formation of  $\alpha$ HM<sub>2</sub>TTM as previously reported by Murto *et al.*<sup>5</sup> Following this,  $\alpha$ HM<sub>2</sub>TTM was reacted with the respective 3- and 4-linked boronic acids of PCz and TPA to form  $\alpha$ HM<sub>2</sub>TTM-3PCz and  $\alpha$ HM<sub>2</sub>TTM-4TPA. To create the  $\alpha$ H precursors for the other two radical species, the remaining para-chlorine of  $\alpha$ HM<sub>2</sub>TTM was converted to a boronic ester through a Miyaura borylation before being coupled with 4-iodotoluene or 3-bromotriphenylamine. To convert into their respective radicals, all four  $\alpha$ H species were subjected to tetrabutylammonium hydroxide, to form the monoanion, before being oxidised to the radical using para-chloranil. M<sub>2</sub>TTM-4TPA, M<sub>2</sub>TTM-3TPA, M<sub>2</sub>TTM-3PCz and M<sub>2</sub>TTM-4T were formed in a 13%, 56%, 86% and 37% yield respectively. UV-vis absorption measurements were carried out for the radicals in a 0.1 mM toluene solution. All four radicals display an intense absorption feature around 370-400 nm, which is characteristic of a local excitation within the TTM radical core. For M<sub>2</sub>TTM-3PCz and M<sub>2</sub>TTM-4TPA, additional absorption peaks can be seen at 590 and 630 nm respectively. These are attributed to CT transitions between the electron donating group and the electron-accepting TTM core.

We find that the four new molecules confirm the structure-property predictions made in 2020<sup>6</sup> that, in order for a significant D<sub>1</sub> absorption the molecule should not be an alternant hydrocarbon and that the HOMO on the donor (4Me, 3PCz or TPA in this case) has orbital amplitude on the atom through which it is joined to the acceptor (TTM). M<sub>2</sub>TTM-4Me is predicted to have



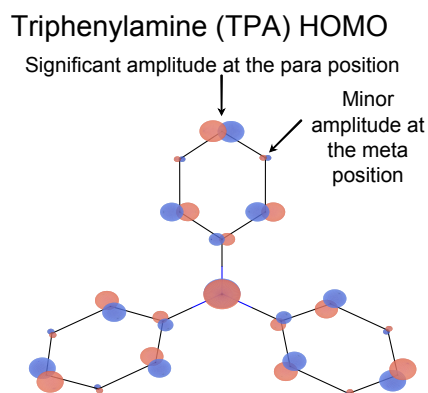


Fig. 7 HOMO of TPA calculated by closed-shell PPP (with the optimised parameters obtained from training on 81 radicals). There is significant HOMO amplitude at the para (4) position but minimal amplitude at the meta (3) position, such that the design rules correctly predict  $M_2$ TTM-4TPA to have a significant low-energy visible absorption and  $M_2$ TTM-3TPA not to have one.

minimal  $D_1$  oscillator strength as it is a *de facto* alternant hydrocarbon, as is observed experimentally.  $M_2$ TTM-3PCz contains a five-membered ring and a nitrogen, both of which break alternacy symmetry leading to a bright  $D_1$  state, as is experimentally observed. A simple PPP calculation on TPA alone finds that the HOMO has significant amplitude at the para (4) position but minimal amplitude at the meta (3) position, as shown in Fig. 7. This therefore predicts that the TPA to TTM charge transfer excitation will be dark in  $M_2$ TTM-3TPA but bright in  $M_2$ TTM-4TPA, as is observed experimentally in Fig. 8. We believe this is the first direct experimental confirmation of the design rule concerning the HOMO amplitude.

### 3.4 Testing

We tested the trained 81-molecule model on our four new organic radicals:  $M_2$ TTM-4Me,  $M_2$ TTM-3PCz,  $M_2$ TTM-3TPA and  $M_2$ TTM-4TPA shown in Fig. 6, which make up the testing set. The same parameters optimised for the 81 radicals in the training set in Section 3.1 (presented in SI Table XII) are applied to the four testing molecules here, and are not re-trained or altered in any way for the testing set. We find that the trained model performs well on the testing set, predicting both  $D_1$  and bright state energies with a significantly higher accuracy than the literature parameters, as can be seen in Fig. 9. We also calculated the RMSE, MAE,  $R^2$  and SRCC for the testing data, presented in Table 2. We find similar values for the errors and correlation metrics for the testing set as seen previously for the training set, again with RMSE and MAE less than 0.3 eV and  $R^2$  and SRCC of 0.93 and 0.76 respectively. The fact that the errors (RMSE and MAE) for the testing set of the four newly-synthesised molecules in Table 2 (which the model was never trained on) are actually slightly lower than for the training set (both for the learning and cross-validation in Table 1) further indicates that overfitting did not occur.

We also compare the ExROPPP simulated UV-visible absorption spectra, with both literature and trained parameters, with the ex-

Table 2 RMSE, MAE,  $R^2$  and SRCC for the testing set of four newly synthesised molecules:  $M_2$ TTM-4Me,  $M_2$ TTM-3PCz,  $M_2$ TTM-3TPA and  $M_2$ TTM-4TPA, using (i) ExROPPP with literature parameters (ii) ExROPPP with parameters trained on the training set of 81 molecules. For RMSE, MAE and  $R^2$  the trained model vastly improves on the literature parameters, and the SRCC is unchanged.

	Literature parameters	Trained Model	Target
RMSE (all states) / eV	0.73	0.15	< 0.3
MAE (all states) / eV	0.71	0.13	< 0.3
$R^2$ (all states)	-0.83	0.93	close to 1
SRCC (all states)	0.76	0.76	close to 1

perimental spectra for these four molecules as shown in Fig. 8. The simulated spectra of all four unseen molecules are significantly improved after training and overall the trained model accurately reproduces their absorption spectra. However, the spectrum of  $M_2$ TTM-3TPA is slightly less well captured by ExROPPP than for the other three molecules, where it predicts a larger  $D_1$  intensity and lower  $D_1$  energy than seen in experiment. Similar extra absorptions are also seen around 500 nm for  $M_2$ TTM-3PCz,  $M_2$ TTM-3TPA and  $M_2$ TTM-4TPA, as seen for TTM-1Cz and TTM-1Cz-An, but are less pronounced. This again highlights that while very accurate at reproducing excitation energies, the trained model is less accurate at predicting intensities of the UV-visible absorptions.

## 4 Conclusions

In this article we have presented the first demonstration of learning the electronically excited states of radicals from experimental data. We achieve this by using the spin-pure ExROPPP method as a surrogate model, both to avoid the spin-contamination problem, and to address the limited experimental data in the literature. We find that the trained ExROPPP model performs far better at computing spectral features of organic radicals than the literature parameters. Four new radicals are synthesised and we test our model by comparing computed spectra against experimental data, finding good agreement and demonstrating its wider applicability as a predictive model. While the trained model accurately reproduces excitation energies, it seems to be less accurate at predicting intensities when compared to experimental UV-visible spectra. More generally, this research shows that ML can be effective with sparse data when using a careful choice of surrogate model.

In future work the accuracy for UV-visible intensities predicted by ExROPPP could possibly be improved by removing ground state mixing in the ExROPPP Hamiltonian and employing an alternative fitness function. This model could also be further extended to predicting the emission spectra of radicals, and also to other atoms and groups common in organic radicals such as O, S and F, nitrile, nitro, aminoxyl and trifluoromethyl.<sup>63</sup> Moreover, while ExROPPP is limited to  $\pi$ -conjugated organic radicals, by incorporating the equations of XCIS into an extended Hückel method such as ZINDO, an analogous but more general approach could be formulated which would allow, for example, for the spin-pure excited states of radicals containing heavy atoms



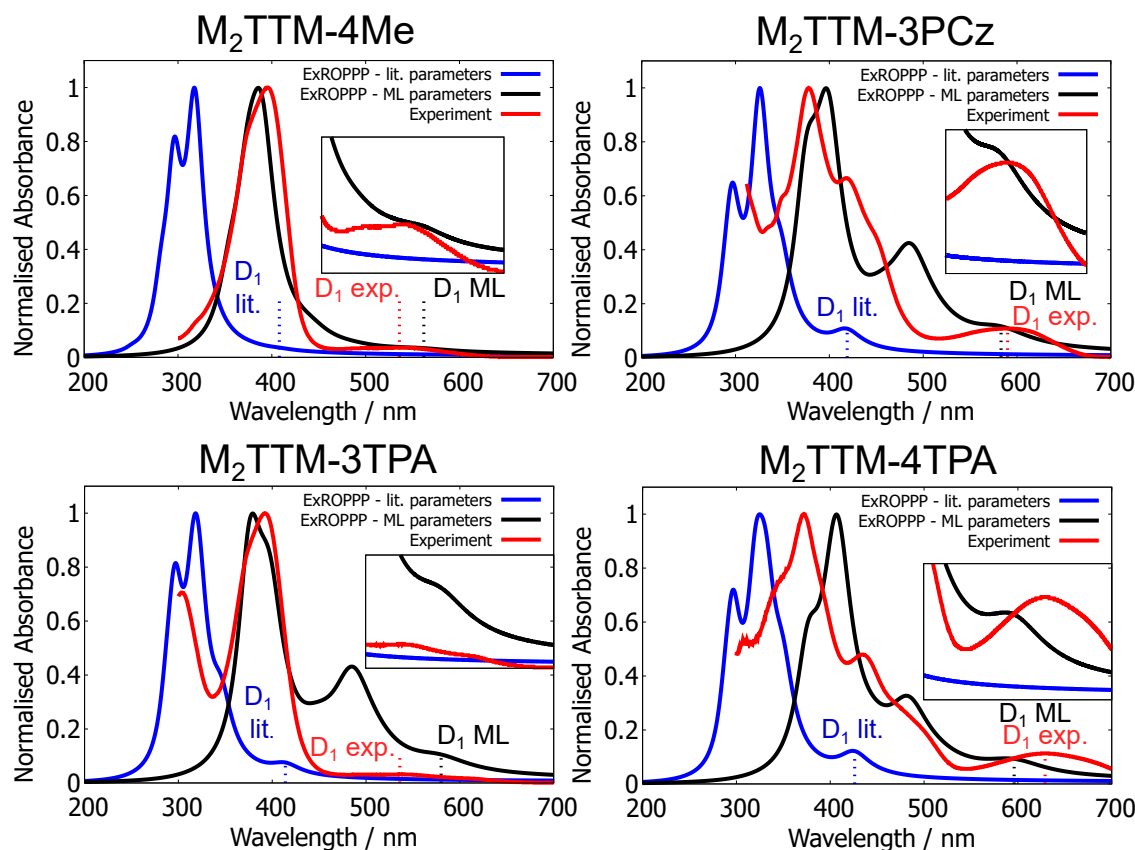


Fig. 8 UV-visible absorption spectra of the newly-synthesised  $M_2$ TTM-4Me (top left),  $M_2$ TTM-3PCz (top right),  $M_2$ TTM-3TPA (bottom left) and  $M_2$ TTM-4TPA (bottom right) measured in 0.1 mM toluene solution (red), simulated using ExROPPP with literature parameters (blue) and of the trained 81 molecule ExROPPP model (black). Trained ExROPPP (black) reproduces the experimental spectra (red) more accurately than untrained ExROPPP with literature parameters (blue).

such as transition metal complexes to be rapidly simulated.<sup>79</sup> Taken together, this work serves as a major step forward for high-throughput screening and inverse molecular design of radicals with applications in OLEDs and qubits.

## 5 Experimental Methods

### 5.1 Characterization techniques of organic radicals

NMR spectra were acquired using a 400 MHz Bruker Avance III HD spectrometer ( $^1\text{H}$ , 400 MHz;  $^{13}\text{C}$ , 100 MHz). Chemical shifts are reported in  $\delta$  (ppm) relative to the solvent peak: chloroform- $d$  ( $\text{CDCl}_3$ :  $^1\text{H}$ , 7.26 ppm;  $^{13}\text{C}$ , 77.16 ppm) and dichloromethane- $d_2$  ( $\text{CD}_2\text{Cl}_2$ :  $^1\text{H}$ , 5.32 ppm;  $^{13}\text{C}$ , 53.84 ppm). Mass spectra were obtained on a Waters Xevo G2-S benchtop QTOF mass spectrometer equipped with a electrospray ionization (ESI) or an atmospheric solid analysis probe (ASAP). Flash chromatography was carried out using Biotage Isolera Four System and Biotage SNAP/Sfär Silica flash cartridges.

### 5.2 Steady-state UV-visible spectroscopy

UV-visible spectra were measured with a commercially available Shimadzu UV-1800 spectrophotometer.

## Author Contributions

TJHH conceived the research idea with assistance from JDG and JS. Computational research for training the ExROPPP model was undertaken by JDG, JS and KM, supervised by TJHH. Organic synthesis, characterization and measurement of UV-Visible absorption spectra was performed by LEW, supervised by HB. JS, KM, LEW and JDG all contributed to DFT calculations. The orbital visualiser was written by KM. KB advised on machine learning. The initial draft of the manuscript was written by JDG with assistance from TJHH, JS and KM and the organic synthetic parts written by LEW. All authors contributed to the revising and reviewing of the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability statement

The Extended Data including the molecular geometries and literature UV-visible absorption data for all molecules, and initial and optimised sets of parameters is available at the UCL Research Data Repository at [https://rdr.ucl.ac.uk/articles/dataset/Research\\_data\\_for\\_Learning\\_Radical\\_Excited\\_States\\_from\\_Sparse\\_Data\\_/29457698/1](https://rdr.ucl.ac.uk/articles/dataset/Research_data_for_Learning_Radical_Excited_States_from_Sparse_Data_/29457698/1).





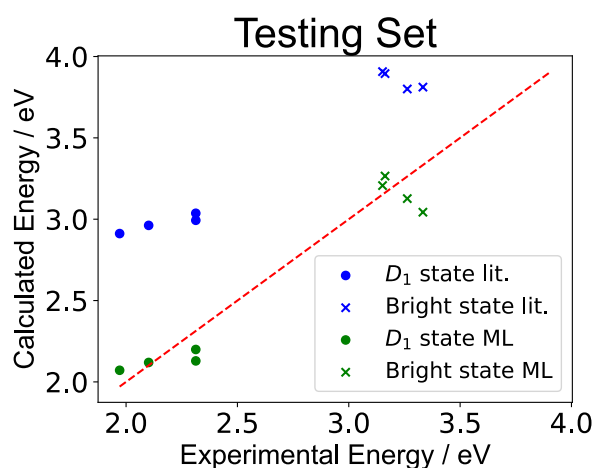


Fig. 9 Regression plot of the excited state energies of the radicals in the testing set calculated by ExROPPP and compared with experimentally determined energies. 'lit.' refers to the parameters sourced from the literature and 'ML' refers to the parameters of the trained 81 molecule model. It can be clearly seen that for this testing set the trained ExROPPP model more accurately reproduces the experimental values than ExROPPP with literature parameters.

## Acknowledgments

We thank Richard H. Friend for helpful discussions and Emrys W. Evans for providing geometries of TTM-3NCz, TTM-3PCz, TTM- $\alpha$ PyID, TTM- $\beta$ PyID, TTM- $\gamma$ PyID and TTM- $\delta$ PyID. TJHH acknowledges a Royal Society University Research Fellowship URF\R1\201502 and a startup grant from University College London. HB acknowledges the grant EP/W017091/1. L.E.W. acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation program grant agreement No. 101020167. KTB acknowledges funding from UKRI grants EP/Y000552/1 and EP/Y014405/1. We acknowledge use of AI tools for data formatting but this was not used to draft the manuscript. We thank Feng Li for permission to reuse the structure of CzBTM.

## Notes and references

- Q. Peng, A. Obolda, M. Zhang and F. Li, *Angew. Chem., Int. Ed.* **54** (2015), 7091.
- X. Ai, E. W. Evans, S. Dong, A. J. Gillett, H. Guo, Y. Chen, T. J. H. Hele, R. H. Friend and F. Li, *Nature* **563** (2018), 536.
- Z. Guo, S. Park, J. Yoon and I. Shin, *Chemical Society Reviews* **43** (2014), 16.
- M. Blanco and I. Villarroja, *TrAC Trends in Analytical Chemistry* **21** (2002), 240.
- P. Murto, R. Chowdhury, S. Gorgon, E. Guo, W. Zeng, B. Li, Y. Sun, H. Francis, R. H. Friend and H. Bronstein, *Nature Communications* **14** (2023), 4147.
- A. Abdurahman, T. J. Hele, Q. Gu, J. Zhang, Q. Peng, M. Zhang, R. H. Friend, F. Li and E. W. Evans, *Nature materials* **19** (2020), 1224.
- F. Li, A. J. Gillett, Q. Gu, J. Ding, Z. Chen, T. J. Hele, W. K. Myers, R. H. Friend and E. W. Evans, *Nature Communications*

**13** (2022), 2744.

- H.-H. Cho, S. Gorgon, G. Londi, S. Giannini, C. Cho, P. Ghosh, C. Tonnelé, D. Casanova, Y. Olivier, T. K. Baikie *et al.*, *Nature Photonics* (2024), 1.
- S. Gorgon, K. Lv, J. Grüne, B. H. Drummond, W. K. Myers, G. Londi, G. Ricci, D. Valverde, C. Tonnelé, P. Murto, A. S. Romanov, D. Casanova, V. Dyakonov, A. Sperlich, D. Beljonne, Y. Olivier, F. Li, R. H. Friend and E. W. Evans, *Nature* **620** (2023), 538.
- H. Hiratsuka, S. Rajadurai, P. K. Das, G. L. Hug and R. W. Fessenden, *Chemical Physics Letters* **137** (1987), 255.
- C. P. Yu, R. Chowdhury, Y. Fu, P. Ghosh, W. Zeng, T. B. Mustafa, J. Grüne, L. E. Walker, D. G. Congrave, X. W. Chua *et al.*, *Science Advances* **10** (2024), eado3476.
- A. Mizuno, R. Matsuoka, S. Kimura, K. Ochiai and T. Kusamoto, *Journal of the American Chemical Society* **146** (2024), 18470, PMID: 38921686.
- K. Andersson, P.-Å. Malmqvist and B. O. Roos, *The Journal of chemical physics* **96** (1992), 1218.
- H. Nakano, *The Journal of Chemical Physics* **99** (1993), 7983.
- G. D. Purvis III and R. J. Bartlett, *The Journal of chemical physics* **76** (1982), 1910.
- C. He, Z. Li, Y. Lei, W. Zou and B. Suo, *The Journal of Physical Chemistry Letters* **10** (2019), 574.
- Z. Li and W. Liu, *Journal of Chemical Theory and Computation* **12** (2016), 238, PMID: 26672389.
- T. J. H. Hele, B. Monserrat and A. M. Alvertis, *The Journal of Chemical Physics* **154** (2021), 244109.
- J. D. Green and T. J. H. Hele, *The Journal of Chemical Physics* **160** (2024), 164110.
- R. Pariser and R. G. Parr, *The Journal of Chemical Physics* **21** (1953), 466.
- R. Pariser, *The Journal of Chemical Physics* **24** (1956), 250.
- J. A. Pople, *Trans. Faraday Soc.* **49** (1953), 1375.
- J. A. Pople and R. K. Nesbet, *The Journal of Chemical Physics* **22** (1954), 571.
- N. Mataga and K. Nishimoto, *Zeitschrift für Physikalische Chemie* **13** (1957), 140.
- D. Maurice and M. Head-Gordon, *The Journal of Physical Chemistry* **100** (1996), 6131.
- T. J. H. Hele, E. G. Fuemmeler, S. N. Sanders, E. Kumarasamy, M. Y. Sfeir, L. M. Campos and N. Ananth, *The Journal of Physical Chemistry A* **123** (2019), 2527, PMID: 30802051.
- J. D. Green, E. G. Fuemmeler and T. J. H. Hele, *The Journal of Chemical Physics* **156** (2022), 180901.
- M. Bedogni, D. Giavazzi, F. Di Maiolo and A. Painelli, *Journal of Chemical Theory and Computation* **20** (2023), 902.
- M. Dubbini, F. Bonvini, L. Savi and F. Di Maiolo, *The Journal of Physical Chemistry C* **128** (2024), 18158.
- K. Jorner, R. Pollice, C. Lavigne and A. Aspuru-Guzik, *The Journal of Physical Chemistry A* **128** (2024), 2445, PMID: 38485448.
- J. Hinze and D. L. Beveridge, *Journal of the American Chemical Society* **93** (1971), 3107.



- 32 F. Van-Catledge, *The Journal of Organic Chemistry* **45** (1980), 4801.
- 33 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature* **559** (2018), 547.
- 34 G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller and O. A. Von Lilienfeld, *New Journal of Physics* **15** (2013), 095003.
- 35 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science* **361** (2018), 360.
- 36 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Science Advances* **3** (2017), e1701816.
- 37 V. L. Deringer, M. A. Caro and G. Csányi, *Advanced Materials* **31** (2019), 1902765.
- 38 Š. Sršen, O. A. von Lilienfeld and P. Slaviček, *Phys. Chem. Chem. Phys.* **26** (2024), 4306.
- 39 W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral and G. Cui, *The Journal of Physical Chemistry Letters* **9** (2018), 6702.
- 40 R. Ramakrishnan, M. Hartmann, E. Tapavicza and O. A. von Lilienfeld, *The Journal of Chemical Physics* **143** (2015), 084111.
- 41 L. M. Roch, S. K. Saikin, F. Häse, P. Friederich, R. H. Goldsmith, S. León and A. Aspuru-Guzik, *ACS Nano* **14** (2020), 6589, PMID: 32338888.
- 42 Z. Chen, F. C. Bononi, C. A. Sievers, W.-Y. Kong and D. Donadio, *Journal of Chemical Theory and Computation* **18** (2022), 4891, PMID: 35913220.
- 43 J. Westermayr and P. Marquetand, *Chemical Reviews* **121** (2021), 9873, PMID: 33211478.
- 44 P. O. Dral and M. Barbatti, *Nature Reviews Chemistry* **5** (2021), 388.
- 45 P. O. Dral, M. Barbatti and W. Thiel, *The journal of physical chemistry letters* **9** (2018), 5660.
- 46 B.-X. Xue, M. Barbatti and P. O. Dral, *The Journal of Physical Chemistry A* **124** (2020), 7199, PMID: 32786977.
- 47 C.-W. Ju, Y. Shen, E. J. French, J. Yi, H. Bi, A. Tian and Z. Lin, *The Journal of Physical Chemistry A* **128** (2024), 2457, PMID: 38382058.
- 48 J. Westermayr, M. Gastegger and P. Marquetand, *The Journal of Physical Chemistry Letters* **11** (2020), 3828, PMID: 32311258.
- 49 A. Fabrizio, K. R. Briling and C. Corminboeuf, *Digital Discovery* **1** (2022), 286.
- 50 K. R. Briling, Y. Calvino Alonso, A. Fabrizio and C. Corminboeuf, *Journal of Chemical Theory and Computation* **20** (2024), 1108, PMID: 38227222.
- 51 E. Cignoni, D. Suman, J. Nigam, L. Cupellini, B. Mennucci and M. Ceriotti, *ACS Central Science* **10** (2024), 637.
- 52 D. Suman, J. Nigam, S. Saade, P. Pegolo, H. Türk, X. Zhang, G. K.-L. Chan and M. Ceriotti, *Journal of Chemical Theory and Computation* **21** (2025), 6505, PMID: 40556546.
- 53 R. Hoffmann, *The Journal of Chemical Physics* **39** (1963), 1397.
- 54 M. J. Dewar and W. Thiel, *Journal of the American Chemical Society* **99** (1977), 4899.
- 55 J. J. Stewart, *Journal of Molecular modeling* **13** (2007), 1173.
- 56 M. P. Repasky, J. Chandrasekhar and W. L. Jorgensen, *Journal of computational chemistry* **23** (2002), 1601.
- 57 A. S. Christensen, T. Kubař, Q. Cui and M. Elstner, *Chemical Reviews* **116** (2016), 5301, PMID: 27074247.
- 58 E. Fabiano, T. Keal and W. Thiel, *Chemical Physics* **349** (2008), 334, electron Correlation and Molecular Dynamics for Excited States and Photochemistry.
- 59 T. J. H. Hele, *Physical Chemistry of Semiconductor Materials and Interfaces XX*, International Society for Optics and Photonics, SPIE (2021), volume 11799, 117991A.
- 60 H. C. Longuet-Higgins and J. A. Pople, *Proceedings of the Physical Society. Section A* **68** (1955), 591.
- 61 T. Li Chu and S. I. Weissman, *The Journal of Chemical Physics* **22** (1954), 21.
- 62 J. M. Hudson, T. J. H. Hele and E. W. Evans, *Journal of Applied Physics* **129** (2021), 180901.
- 63 A. Mizuno, R. Matsuoaka, T. Mibu and T. Kusamoto, *Chemical Reviews* **124** (2024), 1034, PMID: 38230673.
- 64 X. Ai, Y. Chen, Y. Feng and F. Li, *Angewandte Chemie International Edition* **57** (2018), 2869.
- 65 S. R. Ruberu and M. A. Fox, *The Journal of Physical Chemistry* **97** (1993), 143.
- 66 K. Matsuda, R. Xiaotian, K. Nakamura, M. Furukori, T. Hosokai, K. Anraku, K. Nakao and K. Albrecht, *Chem. Commun.* **58** (2022), 13443.
- 67 Y. Gao, A. Obolda, M. Zhang and F. Li, *Dyes and Pigments* **139** (2017), 644.
- 68 S. Dong, A. Obolda, Q. Peng, Y. Zhang, S. Marder and F. Li, *Mater. Chem. Front.* **1** (2017), 2132.
- 69 H. Guo, Q. Peng, X.-K. Chen, Q. Gu, S. Dong, E. W. Evans, A. J. Gillett, X. Ai, M. Zhang, D. Credgington *et al.*, *Nature materials* **18** (2019), 977.
- 70 S. Dong, W. Xu, H. Guo, W. Yan, M. Zhang and F. Li, *Phys. Chem. Chem. Phys.* **20** (2018), 18657.
- 71 C. Yan, D. An, W. Chen, N. Zhang, Y. Qiao, J. Fang, X. Lu, G. Zhou and Y. Liu, *CCS Chemistry* **4** (2022), 3190.
- 72 G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Włoch, P. Xu, F. Zahariev and M. S. Gordon, *The Journal of Chemical Physics* **152** (2020), 154102.
- 73 N. A. Seifert, K. Prozument and M. J. Davis, *The Journal of Chemical Physics* **155** (2021).
- 74 J. A. Nelder and R. Mead, *The Computer Journal* **7** (1965), 308.
- 75 J. A. Pople, *Proceedings of the Physical Society. Section A* **68** (1955), 81.



- 76 C. C. J. Roothaan, *Rev. Mod. Phys.* **32** (1960), 179.
- 77 S. de Bruijn, *Chemical Physics Letters* **8** (1971), 19.
- 78 P. Ghosh, A. M. Alvertis, R. Chowdhury, P. Murto, A. J. Gillett, S. Dong, A. J. Sneyd, H.-H. Cho, E. W. Evans, B. Monserrat *et al.*, *Nature* **629** (2024), 355.
- 79 W. P. Anderson, W. D. Edwards and M. C. Zerner, *Inorganic Chemistry* **25** (1986), 2728.



The Extended Data including the molecular geometries and literature UV-visible absorption data for all molecules, and initial and optimised sets of parameters are available at the UCL Research Data Repository at [https://rdr.ucl.ac.uk/articles/dataset/Research\\_data\\_for\\_Learning\\_Radical\\_Excited\\_States\\_from\\_Sparse\\_Data\\_/29457698/1](https://rdr.ucl.ac.uk/articles/dataset/Research_data_for_Learning_Radical_Excited_States_from_Sparse_Data_/29457698/1)

Article Online  
DOI: 10.1039/D5SC04276C

