

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Learning Radical Excited States from Sparse Data[†]

Jingkun Shen,^{a,‡} Lucy E. Walker,^{b,c,‡} Kevin Ma,^{a,‡} James D. Green,^{a,‡} Hugo Bronstein,^{b,c} Keith T. Butler^a and Timothy J. H. Hele^{*a}

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 multireference 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 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 (RMS) 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 testing set. This model 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, NIR and IR spectral regions, features which are unusual and highly desirable.^{1–8} These radical OLEDs, based on organic monoradicals, offer an alternative for the next generation of highly efficient lighting. Furthermore, the 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.⁹ However, historically many organic radicals have been non-emissive, such that trial-and-error exploration of chemical space is inefficient, and there is therefore a large and unmet need for a method which 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.^{10,11}

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 2nd order (CASPT2) and Coupled-cluster theory, however, these methods are very computationally expensive, making them unsuitable for high-throughput workflows.^{12–14} 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.^{15,16} Additionally, it has been shown that for the most accurate calculation of excited state energies one must also include nuclear quantum effects.¹⁷ 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.¹⁸ ExROPPP is based on the Pariser-Parr-Pople (PPP) Hamiltonian^{19–23} with a subsequent Extended Configuration Interaction Singles (XCIS)²⁴ calculation which ensures spin purity. PPP theory has gained recent popularity for predicting electronic properties at a significantly reduced computational

^a Department of Chemistry, Christopher Ingold Building, University College London, WC1H 0AJ, UK.

*E-mail: t.hele@ucl.ac.uk

^b Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW, UK.

^c Department of Physics, Cavendish Laboratory, Cambridge University, Cambridge CB3 0HF, UK.

[†]These authors contributed equally.

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

cost.^{18,25–29} Being a semiempirical method, PPP theory and consequently ExROPPP requires parameters which must be specified at the start of a calculation.^{19–23} 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.¹⁸ 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.^{23,30,31} 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.³² 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.³³ ML has seen numerous applications in predicting energies, structures and reactivity patterns of molecules and materials.^{32–38} Furthermore it has been applied to calculating the excited states of molecules and simulating excited state potential energy surfaces.^{33,39–45} 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 ML was applied to calculate the electronically excited states of radicals by training on closed-shell molecules, however to our knowledge, learning the excited states directly from excited state data of organic radicals themselves has yet to be attempted.⁴⁶ 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^{32,33,35} and we are unaware of any such large datasets for 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.⁴² 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.⁴² 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.^{47–49}

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.^{18,25,26,50–56} Using 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 300–400 nm, and a much weaker absorption in the visible. The weak visible D_1 state has been investigated using various levels of theory (TD-DFT, PPP, MCSCF).^{2,9,18} In the special case of alternant hydrocarbons, D_1 is a minus combination $|\Psi_{i0}^-\rangle$ of the HOMO-SOMO and SOMO-LUMO excitations and is essentially dark in the absorption spectrum.^{6,18,57–59} Alternant hydrocarbons are usually non-emissive as non-radiative processes outcompete fluorescence.^{6,17} 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.^{1,2,9} These radicals are also highly emissive and have been incorporated into high-performing OLEDs.^{1,2} 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.⁵

In this paper we learn the excited states of organic radicals directly from their experimental data for the first time. 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 π -conjugated organic radicals to be the largest of its kind.⁴⁶ 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.^{2,5–7,9,46,60–68} 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.^{2,5,9,46,61} However, the molecular geometries for some molecules could not be found in the literature so these structures were optimised using unrestricted DFT in GAMESS-US.⁶⁹ 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 first excited doublet states (D_1), the energies E_{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 two absorptions, extracted from linear UV-visible absorption spectra. These target properties are similar to those previously used in ML of molecular spectra³³. 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 typical radicals).⁹ 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

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

where w_{D_1} , w_{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⁻² 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.⁷⁰

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. 3.⁷¹ 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_{brt} set to 1 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.⁵⁷ For heterocycles, all three terms are included with weights of 1 (except for a few molecules whose bright state data

could not be found, 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, lead 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: M₂TTM-4Me, M₂TTM-3PCz, M₂TTM-3TPA and M₂TTM-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_{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 differences (MAD), R² 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).

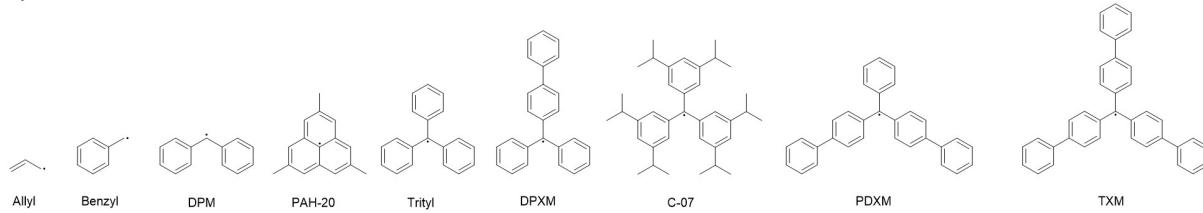
2.5 PPP parameterization

We employ largely the same functional form of PPP theory as in previous work.^{18,23,25,26} The parameters of this model are the one-electron on-site Coulomb ϵ_μ and hopping t parameters, and two-electron Hubbard U and distance scaling r_0 parameters. We use the Mataga-Nishimoto form for the two-electron integrals

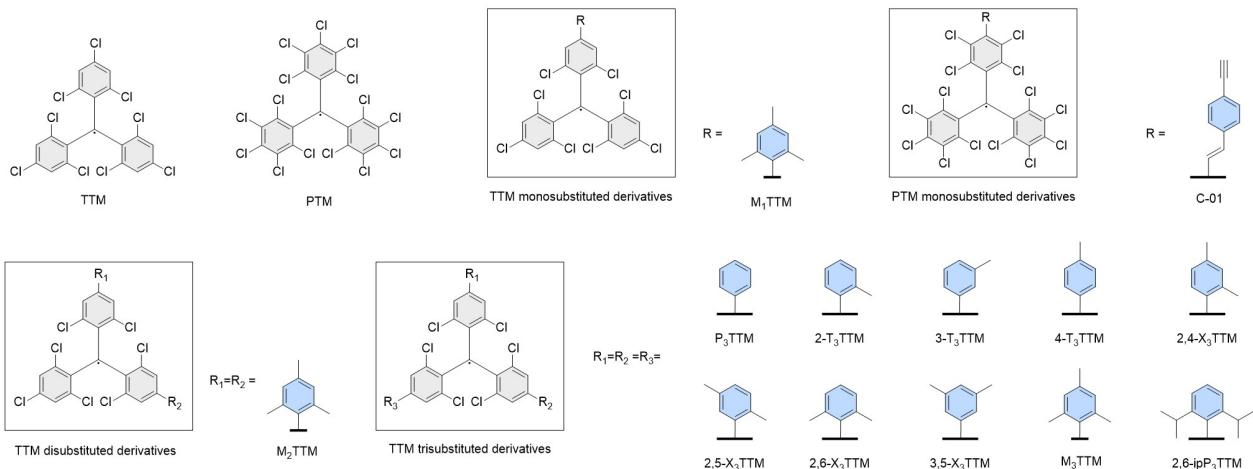
$$\gamma_{\mu\nu} \simeq (\mu\mu|\nu\nu) = \frac{U_{\mu\nu}}{1 + r_{\mu\nu}/r_{0,\mu\nu}}, \quad (2)$$

expressed in terms of atomic orbitals, where $r_{\mu\nu}$ is the scalar distance between atoms μ and ν .²³ However unlike Refs. 18,23,25, 26, but similar to Ref. 72 we elect to use an exponentially decaying function which is scaled by the cosine of the dihedral angle for the hopping term of the form $t_{\mu\nu} = A \exp(-br_{\mu\nu}) \cos \theta$. The PPP parameters are atom specific, with a different ϵ_μ for each atom type and different $t_{\mu\nu}$ for each pair of types of bonded atoms μ and ν . There is only one independent $U_{\mu\mu}$ and $r_{0,\mu\mu}$ for each atom type, and an average of the parameters for different atom types is taken for two-electron interactions between two different types of atoms i.e. $U_{\mu\nu} = \frac{1}{2}(U_{\mu\mu} + U_{\nu\nu})$ and $r_{0,\mu\nu} = \frac{1}{2}(r_{0,\mu\mu} + r_{0,\nu\nu})$. We use different parameters for pyridine and pyrrole/aniline type nitrogen atoms due to their different numbers of π -electrons. Car-

Hydrocarbon Radicals



Heterocyclic Radicals with Chlorine:



Heterocyclic Radicals with Chlorine and Pyrrole-Type Nitrogen:

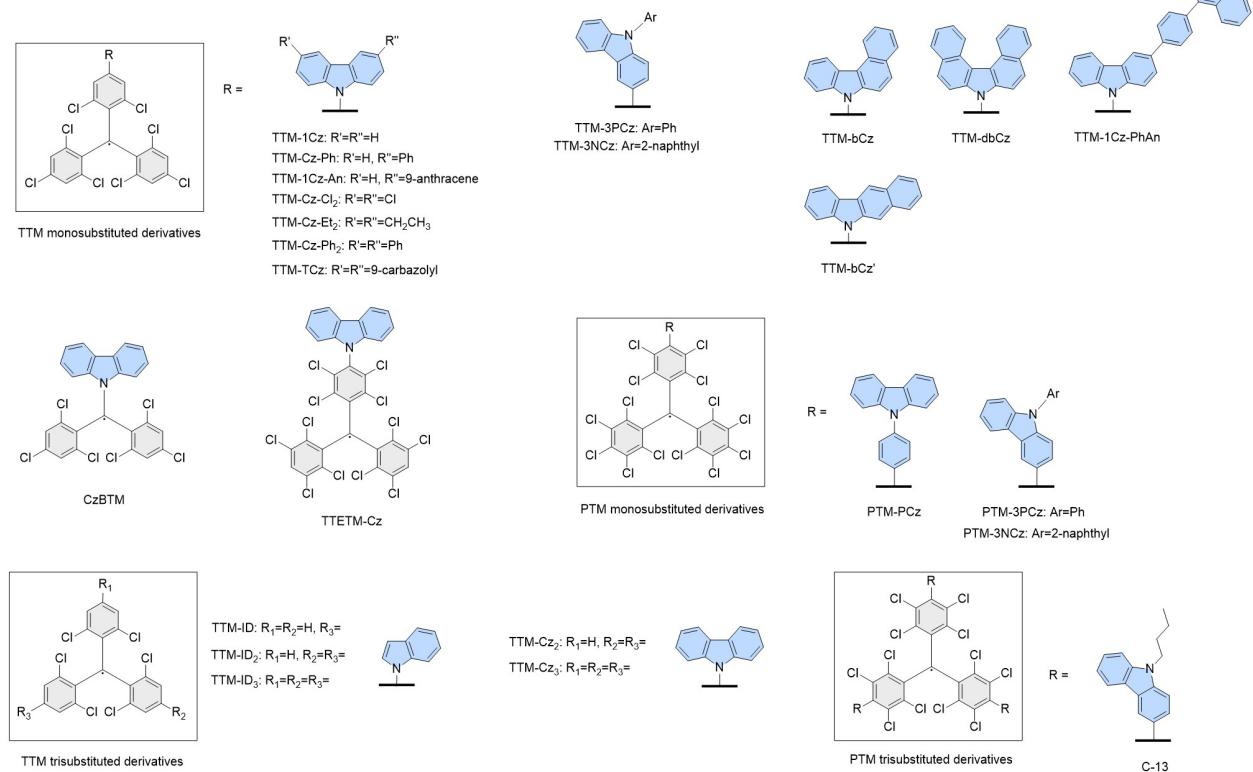
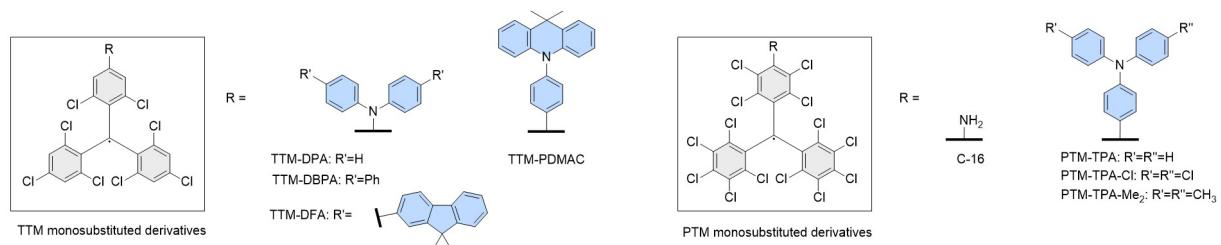
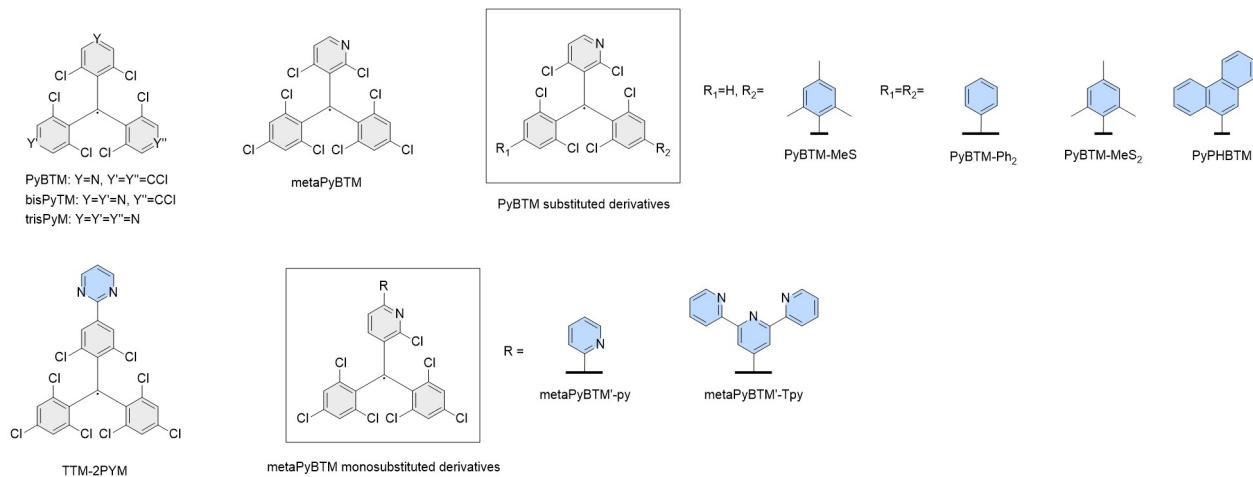


Fig. 1 Structures of the molecules in the training set containing carbon, hydrogen, chlorine and pyrrole-type nitrogen. The central TTM/PTM backbone is coloured in grey, and substituents colored in light blue.

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:

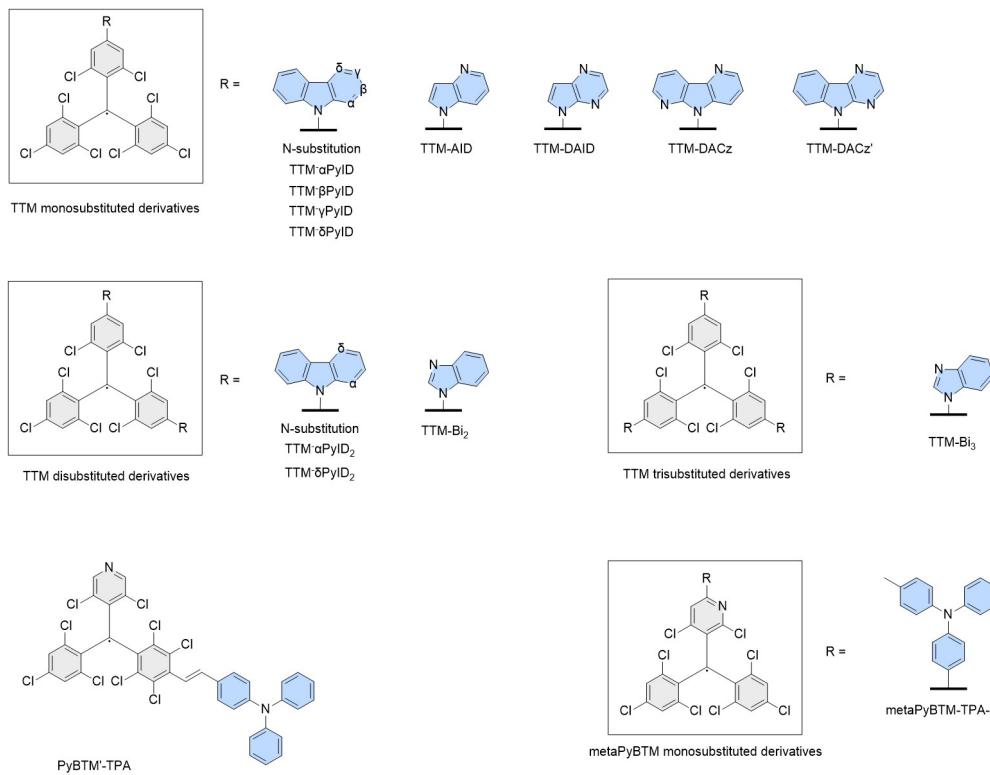


Fig. 2 Structures of the molecules in the training set containing aniline, pyridine and multiple types of nitrogen. The central TTM/PTM backbone is coloured in grey, and substituents colored in light blue.

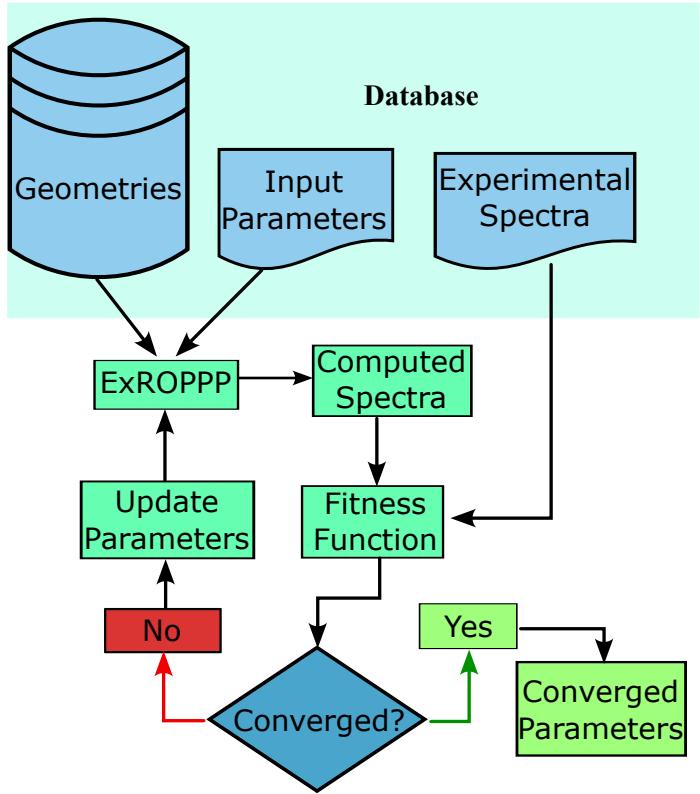


Fig. 3 A flow diagram illustrating our method for training our ExROPPP model on the experimental absorption data of organic radicals.

bon and pyridine type nitrogen atoms contribute one electron to the π -system whereas chlorine and pyrrole/aniline type nitrogen atoms contribute two, from lone pairs. We model the atomic cores formed of the nuclei and core electrons as point charges at the nuclei with effective charge of e for carbon and pyridine type nitrogen, and $2e$ for chlorine and pyrrole/aniline type nitrogen, where e is the electron charge.

To show the improvement of the trained model, we compare the trained parameters with those initially sourced from the literature. Full details of the parameterisation used can be found in SI Section V.

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 MAD 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

Table 1 Total fitness, root mean-squared errors (RMSE), mean absolute differences (MAD), R^2 and Spearman's rank correlation coefficients (SRCC) for the trained ExROPPP model and from K-fold validation 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
MAD (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

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.¹ 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.⁹ 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 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.⁹ Trained ExROPPP does produce an extra absorption around 500nm for both molecules which is not seen in experiment, and could be an artifact of the choice of fitness function. 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. Nevertheless, the ability of the trained ExROPPP model to accurately capture the absorption spectra and excited state features of both typical and anomalous radicals shows its flexibility and robustness.

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 data noise is shown to be small for all trials. Artificially (selected by the user) and randomly distributed folds gave similar errors proving that the distribution of molecules across folds does not significantly affect the model. Also, the K -fold results show that 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

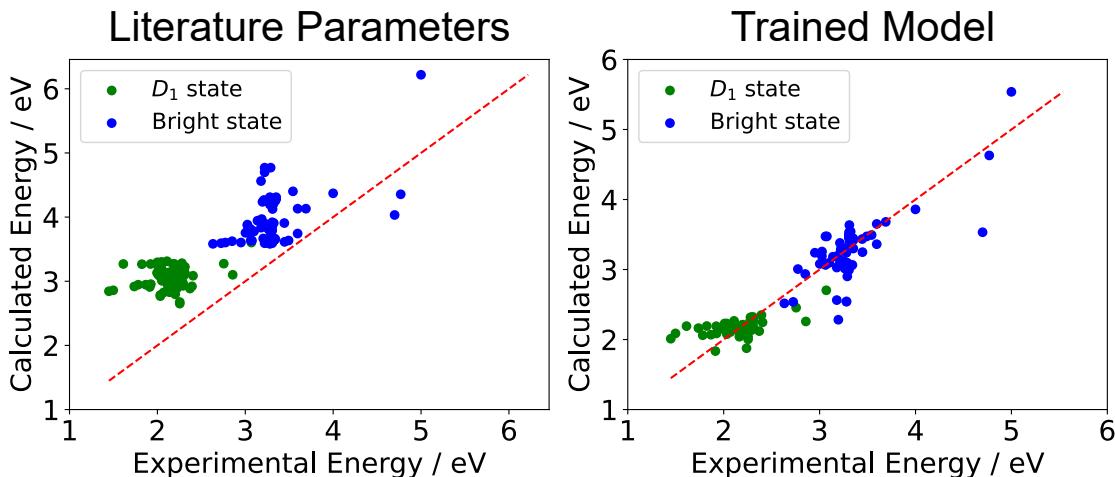


Fig. 4 Regression plots of the excited state energies of the 81 molecule training set calculated by ExROPPP compared with experimentally determined energies, using parameters obtained from the literature (left) and those of the trained ExROPPP model (right). The trained model predicts the energies of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

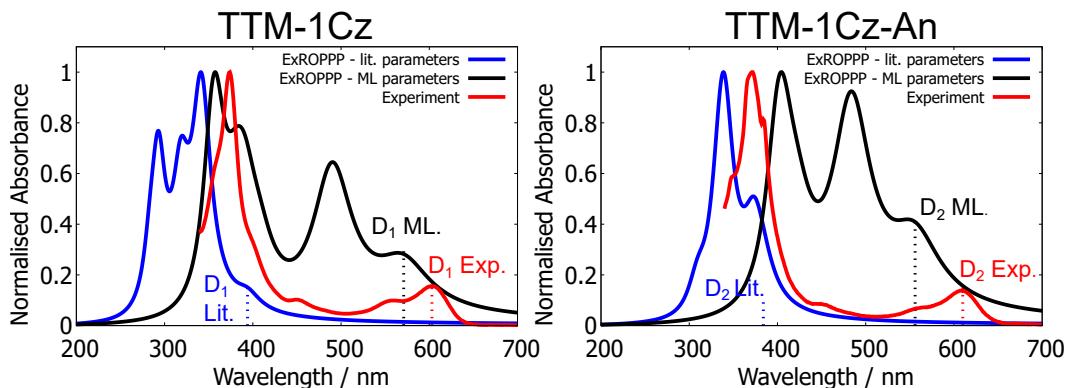


Fig. 5 UV-visible absorption spectra of TTM-1Cz and TTM-1Cz-An measured in $200 \mu\text{M}$ toluene solution at room temperature (red), simulated using ExROPPP with literature parameters (blue) and with the trained 81 molecule model (black).⁹ The trained model substantially improves on the literature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

trained on orbital data.

A qualitative comparison of the singly occupied (SOMO) and highest occupied (HOMO) molecular orbitals from ExROPPP and ROHF (B3LYP/6-31G(d,p)) demonstrates good overall agreement in both shape and localization, highlighting the transferability of the ExROPPP-based parametrization. In TTM-1Cz, TTM-3NCz, and M₃TTM, the SOMOs remain largely centered on the triaryl-methyl 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., C₂ for TTM-1Cz, C₃ for M₃TTM) 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 M₃TTM 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.⁵ 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.⁷³ 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

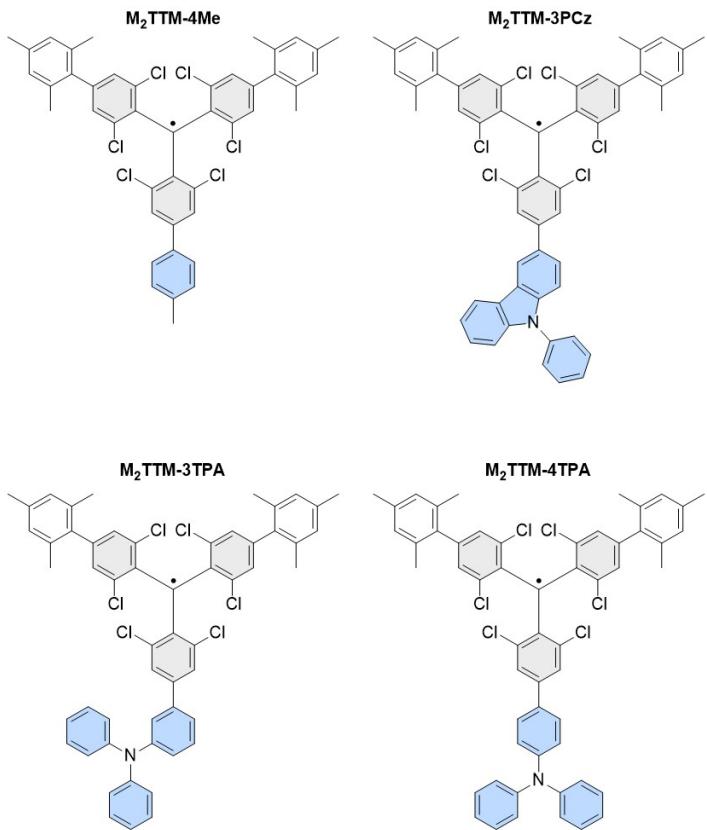


Fig. 6 Structures of the four newly synthesised radicals reported in this work: M₂TTM-4Me, M₂TTM-3PCz, M₂TTM-3TPA and M₂TTM-4TPA, which constitute the testing set.

radical core through both the 3- and 4-positions to assess accuracy in predicting spectroscopic outcomes for different stereoisomers. The combination of mesylation 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 formation of α HM₂TTM as previously reported by Murto *et al.*⁵ Following this, α HM₂TTM was reacted with the respective 3- and 4-linked boronic acids of PCz and TPA to form α HM₂TTM-3PCz and α HM₂TTM-4TPA. To create the α H precursors for the other two radical species, the remaining para-chlorine of α HM₂TTM 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 α H species were subjected to tetrabutylammonium hydroxide, to form the monoanion, before being oxidised to the radical using para-chloranil. M₂TTM-4TPA, M₂TTM-3TPA, M₂TTM-3PCz and M₂TTM-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 M₂TTM-3PCz and M₂TTM-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

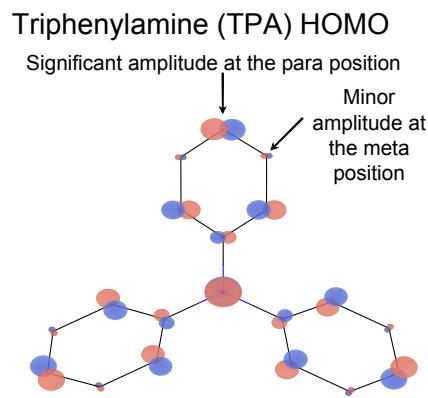


Fig. 7 HOMO of TPA calculated by closed-shell PPP (with the optimised parameters obtained from training on 81 radicals). There is significant HOMO amplitude at the para (4) position but minimal amplitude at the meta (3) position, such that the design rules correctly predict M₂TTM-4TPA to have a significant low-energy visible absorption and M₂TTM-3TPA not to have one.

electron-accepting TTM core.

We find that the four new molecules confirm the structure-property predictions made in 2020⁶ that, in order for a significant D₁ 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). M₂TTM-4Me is predicted to have minimal D₁ oscillator strength as it is a *de facto* alternant hydrocarbon, as is observed experimentally. M₂TTM-3PCz contains a five-membered ring and a nitrogen, both of which break alternacy symmetry leading to a bright D₁ 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 M₂TTM-3TPA but bright in M₂TTM-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: M₂TTM-4Me, M₂TTM-3PCz, M₂TTM-3TPA and M₂TTM-4TPA shown in Fig. 6, which make up the testing set. We find that the trained model performs well on the testing set, predicting both D₁ 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, MAD, R² 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 MAD less than 0.3 eV and R² and SRCC of 0.93 and 0.76 respectively. The fact that the errors (RMSE and MAD) for the testing set are actually slightly lower than for the training set further indicates that overfitting did not occur.

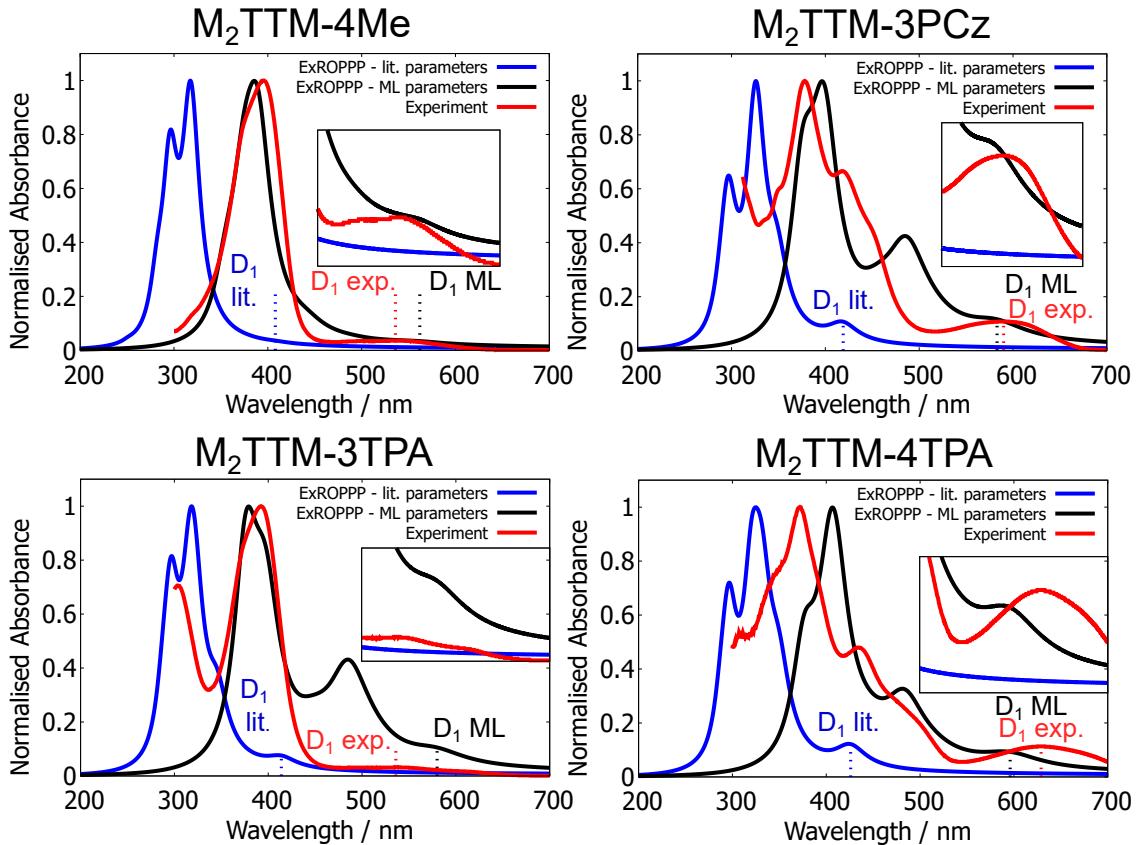


Fig. 8 UV-visible absorption spectra of the newly-synthesised M₂TTM-4Me (top left), M₂TTM-3PCz (top right), M₂TTM-3TPA (bottom left) and M₂TTM-4TPA (bottom right) measured in 0.1 mM toluene solution (red), simulated using ExROPPP with literature parameters (blue) and of the trained 81 molecule ExROPPP model (black). Trained ExROPPP (black) reproduces the experimental spectra (red) more accurately than untrained ExROPPP with literature parameters (blue).

Table 2 Root mean-squared errors (RMSE), mean absolute differences (MAD), R² and Spearman's rank correlation coefficients (SRCC) for the testing set of four newly synthesised molecules: M₂TTM-4Me, M₂TTM-3PCz, M₂TTM-3TPA and M₂TTM-4TPA, using i) the trained 81 molecule model and ii) ExROPPP with parameters obtained from the literature.

	Literature parameters	Trained Model	Target
RMSE (all states) / eV	0.73	0.15	< 0.3
MAD (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

We also compare the ExROPPP simulated UV-visible absorption spectra, with both literature and trained parameters, with the experimental spectra for these four molecules as shown in Fig. 8. The simulated spectra of all four molecules are significantly improved after training. As well as a significant improvement in accuracy of the D₁ and bright state energies, the shape of the spectra are overall also better captured by the trained ExROPPP model. The only slight outlier is M₂TTM-3TPA, for which ExROPPP predicts a larger D₁ intensity and lower D₁ energy than seen in experiment. Overall, however, the trained model accurately reproduces the absorption spectra of these four unseen molecules.

4 Conclusions

In this article we have presented the first demonstration of learning the 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. In future work this model could 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, aminoxy and trifluoromethyl.⁶⁰ In summary, 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 (¹H, 400 MHz; ¹³C, 100 MHz). Chemical shifts are reported in δ (ppm) relative to the solvent peak: chloroform-d

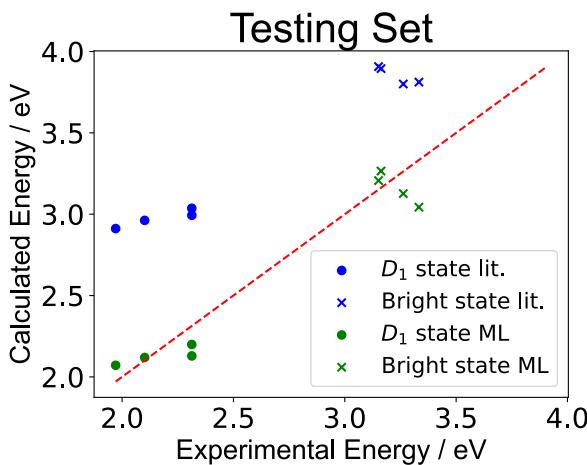


Fig. 9 Regression plot of the excited state energies of the radicals in the testing set calculated by ExROPPP and compared with experimentally determined energies. 'lit.' refers to the parameters sourced from the literature and 'ML' refers to the parameters of the trained 81 molecule model. It can be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

(CDCl₃: ¹H, 7.26 ppm; ¹³C, 77.16 ppm) and dichloromethane-d₂ (CD₂Cl₂: ¹H, 5.32 ppm; ¹³C, 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 <link to be provided>

Acknowledgments

We thank Richard H. Friend for helpful discussions and Emrys W. Evans for providing geometries of TTM-3NCz, TTM-3PCz, TTM- α PyID, TTM- β PyID, TTM- γ PyID and TTM- δ 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 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. Villarroya, *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.
- 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.

- 17 T. J. H. Hele, B. Monserrat and A. M. Alvertis, *The Journal of Chemical Physics* **154** (2021), 244109.
- 18 J. D. Green and T. J. H. Hele, *The Journal of Chemical Physics* **160** (2024), 164110.
- 19 R. Pariser and R. G. Parr, *The Journal of Chemical Physics* **21** (1953), 466.
- 20 R. Pariser, *The Journal of Chemical Physics* **24** (1956), 250.
- 21 J. A. Pople, *Trans. Faraday Soc.* **49** (1953), 1375.
- 22 J. A. Pople and R. K. Nesbet, *The Journal of Chemical Physics* **22** (1954), 571.
- 23 N. Mataga and K. Nishimoto, *Zeitschrift für Physikalische Chemie* **13** (1957), 140.
- 24 D. Maurice and M. Head-Gordon, *The Journal of Physical Chemistry* **100** (1996), 6131.
- 25 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.
- 26 J. D. Green, E. G. Fuemmeler and T. J. H. Hele, *The Journal of Chemical Physics* **156** (2022), 180901.
- 27 M. Bedogni, D. Giavazzi, F. Di Maiolo and A. Painelli, *Journal of Chemical Theory and Computation* **20** (2023), 902.
- 28 M. Dubbini, F. Bonvini, L. Savi and F. Di Maiolo, *The Journal of Physical Chemistry C* **128** (2024), 18158.
- 29 K. Jorner, R. Pollice, C. Lavigne and A. Aspuru-Guzik, *The Journal of Physical Chemistry A* **128** (2024), 2445, pMID: 38485448.
- 30 J. Hinze and D. L. Beveridge, *Journal of the American Chemical Society* **93** (1971), 3107.
- 31 F. Van-Catledge, *The Journal of Organic Chemistry* **45** (1980), 4801.
- 32 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature* **559** (2018), 547.
- 33 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.
- 34 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science* **361** (2018), 360.
- 35 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Science Advances* **3** (2017), e1701816.
- 36 V. L. Deringer, M. A. Caro and G. Csányi, *Advanced Materials* **31** (2019), 1902765.
- 37 Š. Sršení, O. A. von Lilienfeld and P. Slavíček, *Phys. Chem. Chem. Phys.* **26** (2024), 4306.
- 38 W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral and G. Cui, *The Journal of Physical Chemistry Letters* **9** (2018), 6702.
- 39 R. Ramakrishnan, M. Hartmann, E. Tapavicza and O. A. von Lilienfeld, *The Journal of Chemical Physics* **143** (2015), 084111.
- 40 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.
- 41 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.
- 42 J. Westermayr and P. Marquetand, *Chemical Reviews* **121** (2021), 9873, pMID: 33211478.
- 43 P. O. Dral and M. Barbatti, *Nature Reviews Chemistry* **5** (2021), 388.
- 44 P. O. Dral, M. Barbatti and W. Thiel, *The journal of physical chemistry letters* **9** (2018), 5660.
- 45 B.-X. Xue, M. Barbatti and P. O. Dral, *The Journal of Physical Chemistry A* **124** (2020), 7199, pMID: 32786977.
- 46 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.
- 47 A. Fabrizio, K. R. Briling and C. Corminboeuf, *Digital Discovery* **1** (2022), 286.
- 48 K. R. Briling, Y. Calvino Alonso, A. Fabrizio and C. Corminboeuf, *Journal of Chemical Theory and Computation* **20** (2024), 1108, pMID: 38227222.
- 49 J. Westermayr, M. Gastegger and P. Marquetand, *The Journal of Physical Chemistry Letters* **11** (2020), 3828, pMID: 32311258.
- 50 R. Hoffmann, *The Journal of Chemical Physics* **39** (1963), 1397.
- 51 M. J. Dewar and W. Thiel, *Journal of the American Chemical Society* **99** (1977), 4899.
- 52 J. J. Stewart, *Journal of Molecular modeling* **13** (2007), 1173.
- 53 M. P. Repasky, J. Chandrasekhar and W. L. Jorgensen, *Journal of computational chemistry* **23** (2002), 1601.
- 54 A. S. Christensen, T. Kubař, Q. Cui and M. Elstner, *Chemical Reviews* **116** (2016), 5301, pMID: 27074247.
- 55 E. Fabiano, T. Keal and W. Thiel, *Chemical Physics* **349** (2008), 334, electron Correlation and Molecular Dynamics for Excited States and Photochemistry.
- 56 T. J. H. Hele, *Physical Chemistry of Semiconductor Materials and Interfaces XX*, International Society for Optics and Photonics, SPIE (2021), volume 11799, 117991A.
- 57 T. Li Chu and S. I. Weissman, *The Journal of Chemical Physics* **22** (1954), 21.
- 58 H. C. Longuet-Higgins and J. A. Pople, *Proceedings of the Physical Society. Section A* **68** (1955), 591.
- 59 J. M. Hudson, T. J. H. Hele and E. W. Evans, *Journal of Applied Physics* **129** (2021), 180901.
- 60 A. Mizuno, R. Matsuoka, T. Mibu and T. Kusamoto, *Chemical Reviews* **124** (2024), 1034, pMID: 38230673.
- 61 X. Ai, Y. Chen, Y. Feng and F. Li, *Angewandte Chemie International Edition* **57** (2018), 2869.
- 62 S. R. Ruberu and M. A. Fox, *The Journal of Physical Chemistry* **97** (1993), 143.
- 63 K. Matsuda, R. Xiaotian, K. Nakamura, M. Furukori, T. Hosokai, K. Anraku, K. Nakao and K. Albrecht, *Chem. Commun.* **58** (2022), 13443.
- 64 Y. Gao, A. Obolda, M. Zhang and F. Li, *Dyes and Pigments* **139** (2017), 644.
- 65 S. Dong, A. Obolda, Q. Peng, Y. Zhang, S. Marder and F. Li, *Mater. Chem. Front.* **1** (2017), 2132.

- 66 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.
- 67 S. Dong, W. Xu, H. Guo, W. Yan, M. Zhang and F. Li, *Phys. Chem. Chem. Phys.* **20** (2018), 18657.
- 68 C. Yan, D. An, W. Chen, N. Zhang, Y. Qiao, J. Fang, X. Lu, G. Zhou and Y. Liu, *CCS Chemistry* **4** (2022), 3190.
- 69 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.
- 70 N. A. Seifert, K. Prozument and M. J. Davis, *The Journal of Chemical Physics* **155** (2021).
- 71 J. A. Nelder and R. Mead, *The Computer Journal* **7** (1965), 308.
- 72 S. de Bruijn, *Chemical Physics Letters* **8** (1971), 19.
- 73 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.