Understanding the Key Factors for Photoinduced Radical Generation in Crystalline Triphenylamines Using Experiment and Machine Learning

Gamage Isuri P. Wijesekera, Fahidat A. Gbadamosi, Muhammad Saddam Hossain, Abhilash Patra, Christopher Sutton*, Linda S. Shimizu*

Department of Chemistry and Biochemistry, University of South Carolina 29208, USA.

KEYWORDS. Electron Paramagnetic Resonance, triphenylamine, photoinduced radicals, machine learning

ABSTRACT: In this study, we combine experiments, calculated properties, and machine learning (ML) to design new triphenylamine-based (TPA) molecules that have a high photoinduced radical (PIR) generation in crystals. A dataset of 34 crystal structures was extracted from the Cambridge Crystallographic Data Centre. Eighteen structures with experimentally reported PIR values from 0 to 0.85% were used to build an ML model trained using Random Forest that achieves an average leave-one-out test set error of 0.173% PIR. The ML model was used to screen the remaining 16 compounds, of which 4 were selected and subsequently compared with the experimentally measured PIR%. The predicted PIR% demonstrated good agreement with the measured values of TPA bis-urea macrocycles host-guest complexes and non-macrocyclic compounds of TPAs. Examining a broad set of molecular architectures/scaffolds allows for investigating the structural and electronic properties that lead to high PIR generation. We found very different trends for macrocycles, linear TPAs, and mono TPAs, where mono TPAs consistently have the lowest PIR generation. Macrocycles tend to have the highest PIR generation, especially for systems with benzene and fluorobenzene guests. Although linear analogs overall perform worse than macrocycles, they display clear trends with increasing excited-state dipole moment, oscillator strength and electron-hole covariance, while decreasing ionization potential and interatomic distance are generally correlated with higher PIRs. What is consistently observed is that higher PIRs are seen for brominated analogs. Our study, therefore, provides guidelines for future design strategies of TPAs for PIR generation.

1. Introduction

Triphenylamine (TPA) derivatives, along with their corresponding radical cation forms, are of interest in organic redox catalysis, organic semiconductors, and magnetic materials.[1–3] The low ionization potential of TPAs makes them susceptible to generating radical cations via chemical and electrochemical oxidation[4] and UV irradiation.[1,5] Photoinduced radicals (PIRs) of TPAs have the advantage of being used as optoelectronic materials such as OLEDs[6] and perovskite solar cells.[7] They have been observed in varying quantities in solution and amorphous and crystalline solids. They are detected by their effects on photophysics, altering absorption/emission spectra, and are sometimes quantified by electron paramagnetic resonance (EPR) spectroscopy versus a standard solution of Magic Blue.[8] Huang *et al.* reported the PIR formation of three TPA derivatives constructed from phenothiazine as the donor and 3-ethylrhodanine or 1,3-indanedione as the acceptor.[9] These π conjugated donor-acceptor systems exhibited radicals in solution and solid-state. The reported percentages of PIRs in solution ranged from 16.54% to 19.36%, while in thin films, these compounds ranged from 0.47% to 6.0%.

The discovery of new TPAs that exhibit PIRs has largely proceeded in a case-by-case manner. This effort is mainly guided by trends found previously in the literature, including intramolecular (e.g., TPA structure, electronic properties) and intermolecular aspects (e.g., intermolecular distances) on the persistence of radicals. Regarding the structural elements, increasing spin delocalization, typically by molecular planarization, is crucial for stabilizing TPA radicals.[10] For example, Hellwinkel's[11] design constrained TPAs through additional ring formation. Since then, many others have observed planarization of propeller-shaped TPA gives radical cations that are considerably more stable due to delocalization over the planarized π-system and heteroatoms.[11,12] Para-substituents on TPAs stabilize TPA radicals primarily by preventing degradation via dimerization reactions.[13,14] Additionally, the incorporation of halides often enhances radical generation in TPAs upon UV irradiation, attributed to spin-orbit coupling induced by the heavy atom effect.[8]

Beyond tuning PIRs in TPAs at the molecular level, there is evidence that the molecule packing, altered through intermolecular interactions, affects radical generation in measurements taken from both solution and solid-state. For example,

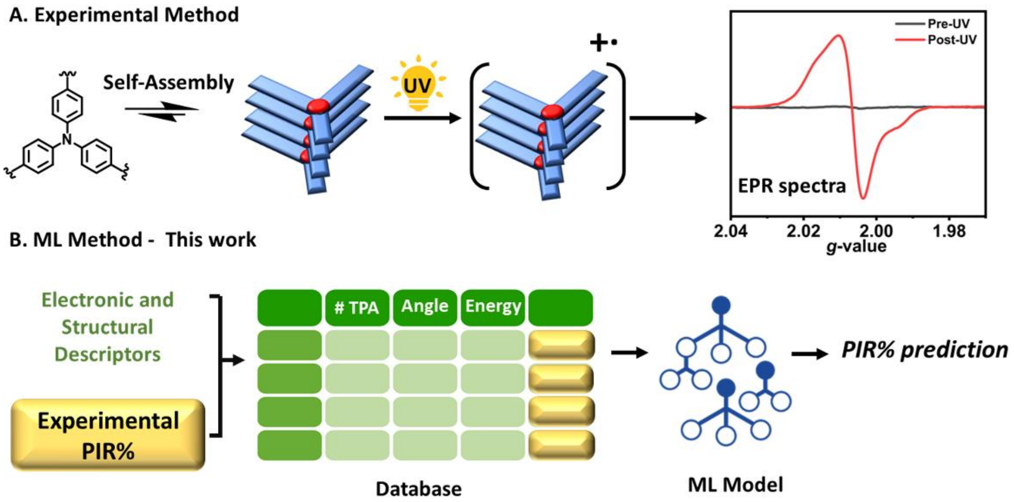


Figure 1. A. Experimental method used to quantify the maximum PIR percentages. Radicals are generated upon UV irradiation on TPAs. Maximum PIR percentage is quantified using EPR. B. ML model trained in this work to predict maximum PIR percentage.

Giuseppone demonstrated the reversible formation of PIRs in soluble supramolecular stacks of amide-tethered TPAs.[15] Similarly, Yang et al. incorporated TPAs within metallocycles and observed that their PIRs stimulated hierarchical assembly, and the TPA radical cations were more stable within the extended structures.[16] Chi et al. observed luminescent PIRs in crystals of tri-*p*-tolueneamine. The radicals quickly decayed upon light removal, leading to sensitive on/off photoswitching. Interestingly, this phenomenon was absent in the amorphous phase.[17]

Crystalline TPA radicals provide an excellent opportunity to explore how structural attributes influence self-assembly and PIR generation. While PIR percentages are lower in crystals, they offer enhanced characterization capabilities for understanding these relationships due to their well-defined structural properties. We reported crystalline self-assembled TPA bis-urea macrocycles and urea-tethered TPA dimers, which showed sizeable differences in the quantities and persistence of their PIRs.[18–22] The variation in observed radical quantities and lifetimes monitored by EPR was attributed to structural and electronic differences in halogen-substituted TPA dimers.[19] Interestingly, the incorporation of guests within structurally similar channels of porous, assembled TPA macrocycles significantly influenced the maximum PIRs.[20]

The trial-and-error discovery of new stable TPAs with high radical generation is time-consuming and resource-intensive. It relies on synthesis, purification, and careful experimental measurements of potential compounds. Alternatively, ML is emerging as a valuable tool to determine material properties that are difficult to measure or calculate.[23–25] Recently, there has been substantial advancement in the accurate prediction of properties of organic crystalline materials using ML, [26,27] such as the band gap for large organic crystal structures[28], thermal properties,[29] and crystalline density[30] among many others. Moreover, the high costs of experimental measurements and the limited availability of quantified PIR% data for crystalline TPA samples make the development of ML methods particularly attractive.

This work explores the structure-property relationship of key crystal structure characteristics, including angles, distances, unit cell data, and electronic properties, such as ionization potential, transition dipole moment, and excitation energy, of crystalline TPA systems with PIR%. We then train an ML model to predict the maximum PIRs in crystalline systems (Figure 1) based on numerical properties derived from the crystalline structure. We applied the ML model to an application dataset of 16 TPA-containing crystal structures from the Cambridge Crystallographic Data Centre (CCDC) to screen for maximum PIR generation. Of these, two were subsequently measured using EPR herein, giving good agreement with the ML predicted values. Additionally, the PIR values of two other structures have recently become available, demonstrating good agreement with the predicted values. Moreover, using correlations among the structural and electronic features with experimental PIRs, we identify that molecular planarization, para substitution, shorter inter-atomic distance of nitrogen of TPAs, low ionization potential, and increased oscillator strength are important for the PIR generation of crystalline TPAs.

2. Methods

2.1 Computational Details

All DFT and TD-DFT calculations were performed using Q-Chem 5.3 at the ωB97XD/6-31G* level. The structures used in these calculations for the single molecules and host-guest complexes were taken directly from the crystallographic information file (cif) and were not optimized with DFT before performing the excited-state and Ip calculations (see Figure S1 and S2 for images of structures). The same level of theory was also used for calculating the electron transfer couplings using the fragment charge difference (FCD) method. [31] In this approach, the coupling values are determined between dimers of either the molecular or host-guest monomers.

2.2 Training set construction

A search of the CCDC identified 34 TPA crystal structures. They were initially exported as cifs with CONQUEST.[32] A subset of 18 crystal structures (refer to Figure 2B, Table S1) was selected for the initial dataset, which comprised data from the literature and our group.[18–22] They were selected based on their

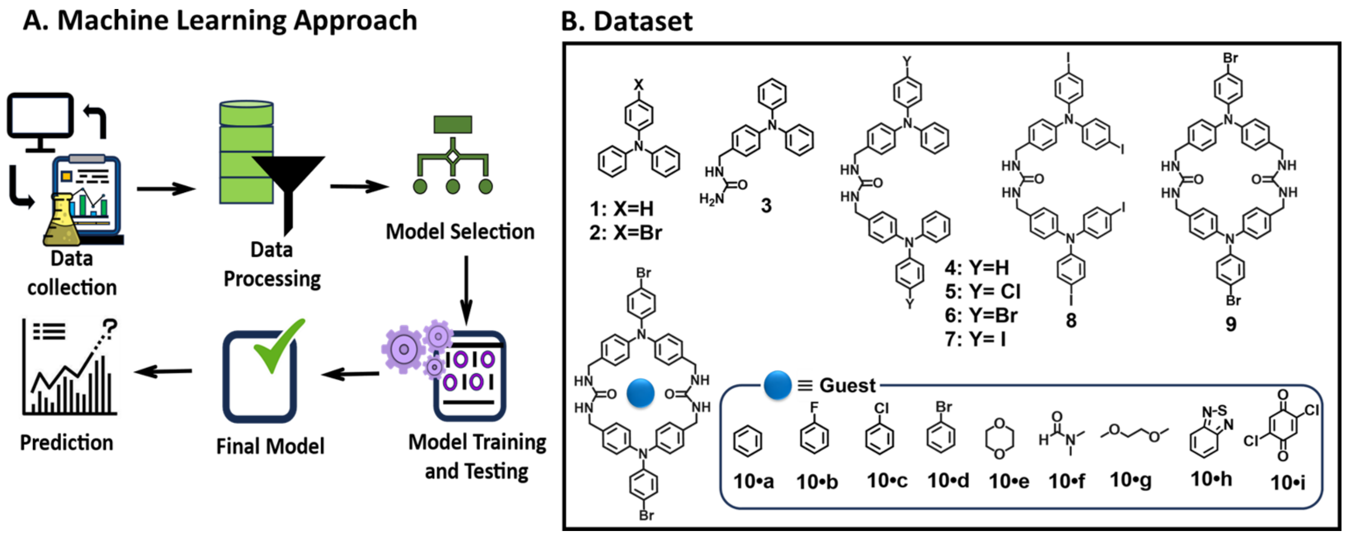


Figure 2. A. Overview of the ML approach used in this study. B. Dataset used to train the ML model to predict PIR%

reported maximum PIR%. The structural data for these compounds and the corresponding maximum PIR%, which serves as the target property for training the model, were extracted. The crystallization of molecules and the measurement of radical percentages have been conducted using consistent methods, ensuring the uniformity and reliability of the training dataset. The quantification of photoinduced radicals (PIRs) during UV irradiation was conducted using a calibration curve established with standard solutions of Magic Blue. Magic Blue, a one-electron oxidant, is a tribromo TPA antimony salt. Comparing the total area of the experimental EPR spectra of the solid sample to the calibration curve established with known concentrations of Magic Blue provided an approximate concentration of PIR in solid TPA[8] (See Supplementary Information). The following equation was used to calculate the reported maximum radical percentages:

$$[\text{PIR\%}] = \frac{[rad]\ of\ calibration\ plot\ \left(\frac{mol}{L}\right) * 0.0001L}{Solid\ TPA\ (mol)} * 100$$

2.3 Feature development and selection

The selection of descriptors for the model was a critical step in building an accurate predictive tool. Descriptors were selected based on their relevance to PIR generation and structure-property relationships. Initially, 24 descriptors were tried (Table S2), but the set was reduced to 12 (see Table 1) to remove features with a high linear correlation between different features, as well as the model's performance and concern for overfitting when too many features are considered given the small dataset size. All the molecules used in the training set are shown in Figure 2B. Tables S3 and S5 contain data of these twelve descriptors for both the training and application datasets, respectively. The training dataset (Table S3) highlights a range of maximum PIR percentages for crystalline TPAs, spanning from 0 to 0.85%.

Consideration was given to several structural descriptors extracted from the crystal structures. Specifically, the maximum angle between phenyl groups in TPA (θ) because planarizing facilitates delocalization over the π-system.[14] Introducing substituents, particularly at the para position of the TPA, can enhance the stability of radical cations by preventing dimerization.[1,3] Therefore, we included the number of para substituents per TPA (N_{para}) as a descriptor. Finally, we calculated the distance between two TPA monomers (*d*) in the crystal structure by hand using Mercury 2022.3.0 because the distance between TPA-containing molecules has been shown to influence through-space coupling interactions.[33,34,35]

Table 1. Structural and calculated descriptors used as inputs into the ML model.

Molecular Descriptor	**Description**
θ	Maximum angle between phenyl groups in TPA
N_{para}	Number of para substituents per TPA
d	Shortest distance between N of TPAs of two monomers
Ip	Ionization potential
μ	Transition dipole moment
Ex	Excitation energy
f	Oscillator Strength
d_{RMS}	Root-mean-square (RMS) electron–hole separation
D	Dipole moment
cov	Electron-hole covariance
V_{01}	Electronic coupling between ground and first singlet excited state
Δq	Charge difference between monomers in ground state

To assess the impact of electronic properties of TPAs on PIR generation, we include several calculated properties such as the ionization potential (Ip), the vertical excitation energy (Ex), oscillator strength (*f*), root-mean-square (RMS) electron–hole separation (d_{RMS}), transition dipole moment (μ), dipole moment

(D), electron-hole covariance (cov), electronic coupling between ground and singlet excited state (V_{01}) and charge difference between monomers (Δq) in a dimer model. The selected descriptors are broadly applicable to both simple TPAs and their complex host-guest systems. Moreover, the motivation for including these properties is that μ plays a crucial role in assessing the interaction strength governing transitions between the ground state and the excited state of TPA PIRs, while d_{RMS} and cov are vital for understanding charge transfer and charge recombination processes within the system.[19,36] The exciton size is quantified by d_{RMS}, while cov effectively includes information about hole size, electron size, and their correlation during the excited-state charge transfer process. V_{01} and Δq were obtained by charge transfer electron coupling calculations between monomers in a dimer model using the FCD method. Also, the impact of the Ip and Ex of TPAs on the generation of PIR is known to be significant.[21,34,35]

Ionization potential (Ip) energies were determined from the energy difference between isolated neutral and radical cation states, using single-point calculations of the closed-shell singlet ground-state geometry. The Ex, f, d_{RMS}, μ, D, cov values for all molecules were calculated for their excited states with the highest oscillator strength and their S1 states (Table S3 and Table S4). We tested whether extracting these properties from the S1 state instead of the state with the highest oscillator strength would affect the model performance and found no difference. Therefore, all the excited-state property analyses were performed using the properties extracted from the state with the highest oscillator strength. Excitation energies, oscillator strengths, and HOMO-LUMO transitions for the first four excited states (S1 to S4) are tabulated in Table S8.

The ML model was trained using a Random Forest (RF) regression implemented in the Python scikit-learn package.[37] RF is an ensemble learning method that combines the predictions of multiple decision trees to provide robust and accurate results.[38–40] It is typically well-suited for small datasets, as it potentially mitigates overfitting by aggregating the outputs of multiple trees.

The RF model was trained with fixed hyperparameters, using 100 estimators, a minimum of 2 samples to split a node and at least 2 samples per leaf. These parameters were determined after experimenting with several combinations to balance model complexity and performance. The model was trained using Leave-One-Out Cross-Validation (LOOCV), where each of the 18 samples was treated as the left-out sample. To ensure robust and unbiased performance evaluation, the 50 intendent models were trained using different random seeds. The mean absolute error (MAE) was calculated for each random seed based on the LOOCV scheme.

The model trained using the random seed = 6 was selected as it produced the median value among the 50 MAEs. The final model has an MAE of 0.17% (see Table S6 and S7 for the prediction from LOOCV for the 18 training set samples). This selected model configuration was then used to predict the PIR values for the application set of 16 samples.

2.4 Experimental Details for PIR Quantification.

Crystals of **10•m** were previously synthesized; **14** was purchased from TCI America and crystallized via vapor diffusion of water to a DMSO solution (10 mg/2 mL). PXRDs were performed for both samples to compare the bulk crystals with the reported structures. Theoretical PXRD patterns of **10•m** and **14** were predicted from the CCDC structures 2206566 and 2090087, respectively. For **10•m** and **14**, the experimental PXRD patterns closely matched the theoretical pattern (Figure S5-S6). This suggests the bulk crystals are in a single phase, and the structure of the bulk powder closely resembles that of the crystals used in the validation dataset.

Freshly triple-recrystallized samples of **14** (5 mg) and **10•m** (6.5 mg) were filtered, dried under an inert atmosphere, Ar (g), and weighed. Samples were loaded into quartz EPR tubes. X-band EPR spectra were recorded before and after UV irradiation using 365 nm LEDs. A cooling fan was employed during the irradiation to prevent any heat buildup. Inspection of EPR spectra of both compounds (Figure 3A and 3C) show no signal before irradiation. However, broad axial powder pattern shapes were observed after UV irradiation with g-values of 2.004 for both compounds. The recorded g-value of 2.004 falls within the typical range for TPA radical cations in solution, which is generally between 2.002 and 2.005.[13] The EPR signals of **14** and **10•m** were measured under increasing irradiation time 1 to 16 h and 1 to 32 h, respectively, to examine whether longer irradiation times would increase the number of radicals.

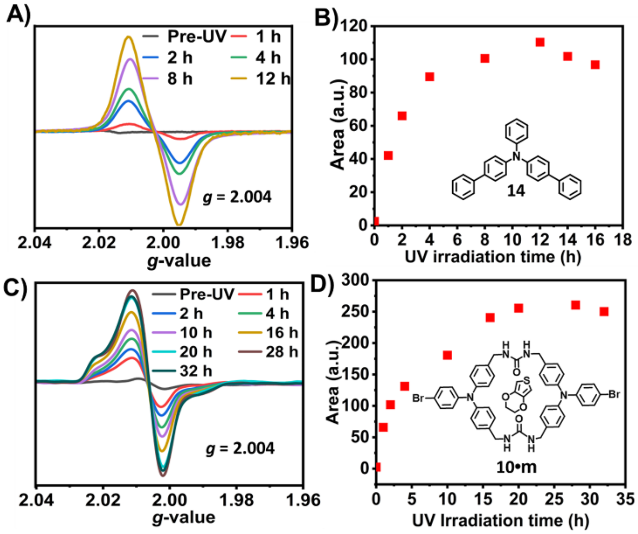


Figure 3. A. EPR with incremental times of UV-irradiation of **14**; B. Double integration over time of UV-irradiation of **14**; C. EPR with incremental times of UV-irradiation of **10•m**; D. Double integration over time of UV-irradiation of **10•m**.

Figure 3A compares the EPR signal of **14** after longer irradiation times and shows the intensity of the signal increases, suggesting the formation of a larger quantity of PIRs. The quantity of radicals was estimated by double integration of the EPR spectra and plotted versus irradiation time (Figure 3B). The concentration of radicals in crystalline **14** increases with irradiation time until it reaches a plateau at ~ 12 h. A calibration curve using Magic Blue standard solutions in dichloromethane (Figure S7) was used to estimate the absolute number of radicals.[20] The maximum number of radicals was estimated by averaging the last four data points of the area vs irradiation time plot (Figure 3B). This means ~1 in 370 molecules (~0.27%) generates a radical. This is equivalent to the number of radicals observed in 100 µL of a 0.34 mM solution of Magic Blue. In the context of **10•m** (as illustrated in Figure 3C and 3D), a maximum PIR concentration of 0.36% was found for 6.5 mg, by averaging the last four data points. Approximately 1 out of 277 molecules produced a radical. This percentage is comparable to the number of radicals detected in a 100 µL sample of a 0.58 mM solution of Magic Blue.

3. Results and Discussion

The ML model was developed on a dataset comprising 18 crystalline TPA structures with PIR% ranging from 0 to 0.85% (Figure 2B and Table 2). Overall, we can divide the TPA molecules into three groups: macrocycles, linear analogs, and mono TPAs. Macrocycles generally exhibit higher PIRs than linear analogs and mono TPA derivatives. All mono TPAs, e.g., with single urea tether (**3**), or missing urea tethers (**1**,**2**) exhibit low PIR generation (< 0.1%).

The low PIR% in mono TPAs (**1**, **2**, and **3**), is partly explained by the lack of para substituents. We observe high PIR% for $N_{para} > 2$ (Figure S3C). All macrocycles have $N_{para} = 3$ and generally exhibit larger PIR% values (although significant differences among these complexes are detailed below). The presence of para substituents hinders TPA dimerization and contributes to persistent PIRs.[14]

Table 2. Training dataset molecules and their reported PIR%.

Molecule	PIR%	Molecule	PIR%
1	0	**10•a**	0.85
2	0	**10•b**	0.45
3	0	**10•c**	0.24
4	0.22	**10•d**	0.23
5	0.13	**10•e**	0.38
6	0.42	**10•f**	0.15
7	0.06	**10•g**	0.28
8	0.12	**10•h**	0.29
9	0.69	**10•i**	0.3

Although macrocycles generally display a higher PIR% than linear analogs, an exception to the trend is the linear analog **6**, which has one of the highest PIR% (> 0.4%) in the dataset. Only three macrocycles (**10•a**, **9**, **10•b**) have a larger PIR% than **6**. Four of these compounds with the highest PIR% (> 0.4%) are observed for bromine-substituted compounds compared with non-brominated analogs in both Br-TPA macrocycles and linear molecules (e.g., **10•a**, **6** vs. **4**), indicating that bromine-substitution on urea-tethered TPAs is particularly favorable for PIR generation. Among the linear analogs, complexes with iodine substitutions (**7** and **8**) tend to generate lower maximum PIR% than those with chlorine and bromine substitutions. This is likely due to the short lifetimes and fast recombination of iodine TPA radicals.[19]

The choice of the substituent affects the ground-state density and ionization potential. Among the linear compounds, **6** has the highest PIR%, and the lowest Ip. Figure S4A shows that Ip exhibits an inverse relationship with PIR% for linear analogs. This is consistent with the literature in which a lower Ip indicates a reduced energy barrier associated with higher radical generations.[35]

Although macrocycles tend to have the highest PIR%, a varying degree of PIR% is observed for macrocycle-guest complexes (**10•a** to **10•i**), suggesting that the radical generation depends on guest inclusion. For example, macrocycles with polar guests have lower PIR%, whereas macrocycles with generally nonpolar guests show higher levels of PIR% (**10•a** vs **10•f**, **10•g**). Moreover, in complexes with halogenated benzene guests, the PIR% tends to decrease as the halogen gets larger (**10•b**>**10•c**>**10•d**). Mono TPAs do not exhibit a clear trend in the Ip and PIR%.

Excited-state energy analysis for these different analogs was conducted using TD-DFT calculations. For the linear analogs and macrocycles, the S1 and S2 states have degenerate excitations localized on the different TPAs and display CT-type excitations (Table S8). [41] In contrast, the S1 state in all mono TPAs is 0.2 eV lower than the S2 state.

In addition to the excited state energies, the oscillator strength, *f,* is correlated with PIR%, which has also been observed for TPAs in solution.[42,43] Most macrocycles exhibit larger *f* values and higher PIR% values (Figure S4B) than linear TPAs. In linear analogs, increasing *f* is also associated with increasing PIR values. For instance, in the case of **6** > **4** > **7,** molecule **6** displays the highest *f* and PIR%, while molecule **7** exhibits the lowest *f* and PIR%.

Increased D is roughly correlated with PIR % for linear analogs (Figure 4, blue circles). For example, compounds **7** and **8** show low D and PIR% compared to compound **6**, which displays the highest D value and PIR%. In contrast to what was observed in linear analogs, a decrease in D in macrocycles leads to an increase in PIR%. Mono TPAs do not exhibit a clear trend in the D and PIR%.

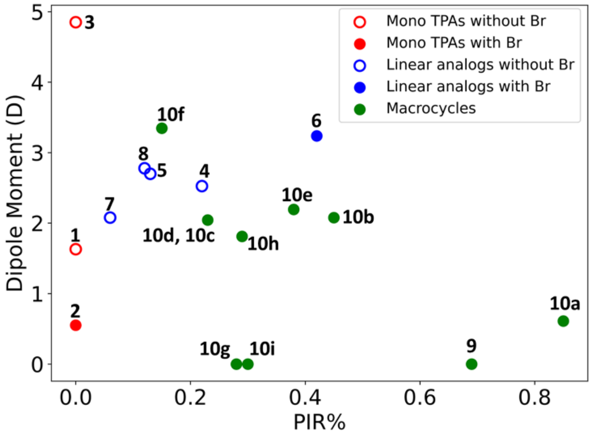


Figure 4. Variation of Dipole Moment versus PIR% for mono TPAs, linear analogs, and macrocycles.

Linear analogs exhibit an increased PIR% with cov (Figure S4C). For instance, within the set of linear molecules **8**,**4** and **6**, molecule **6** has the highest cov and PIR%, while molecule **8** has the lowest cov and PIR%. In contrast, macrocycles do not display a clear correlation between cov and PIR%. However, most macrocycles generally have higher cov values compared to most of the linear analogs while mono TPAs tend to have both low cov values and PIR%. No clear trends were observed for Ex, d_{RMS}, μ, V_{01} and Δq (Figures S4D-S4H).

Beyond electronic properties, structural features extracted from the crystal describe the variation of the higher PIRs in linear analogs. Molecules with higher PIR% tend to possess shorter *d*-values (Figure S3A). In the linear analog series, as *d* decreases, PIR% increases, following the order of **6** > **4** > **8** > **7**. A lower interatomic distance (smaller *d* value) may promote enhanced PIR via through-space electronic coupling that facilitates electron transfer.[33] However, we note that macrocycles do not follow a clear trend with *d* due to the rigid scaffold.

Another structural feature that shows a strong trend is that TPAs with higher maximum PIR percentages tend to have smaller θ values, the maximum angle between phenyl groups in TPAs. Overall, macrocycles exhibit low θ values compared to both linear and mono TPAs (Figure S3B). The low θ values in macrocycles are attributed to their assembly's additional aryl stacking interactions, which can further constrain out-of-plane tilting. In contrast, linear analogs and mono TPAs (**8** and **1** vs **10•a**) have more rotational freedom, generally resulting in larger θ values. The increased θ values reduce PIR% due to the absence of radical cation delocalization over the planarized π-system, consistent with the literature.[11]

The ML model trained to our experimental dataset (described in section 2.3) was applied to an application set of 16 molecules from the CCDC that lacked reported PIR values, shown in Figure 5. Among these structures are molecules featuring TPA bis-urea macrocycles with and without guests, bis-TPA derivatives linked with urea, TPA derivatives with amide linkages, and planarized TPA molecules with oxygen as a bridging atom. The predicted values and standard deviations are shown in Table S7. We aim to use simple ML models to predict the PIR% accurately for totally unseen compounds.

.

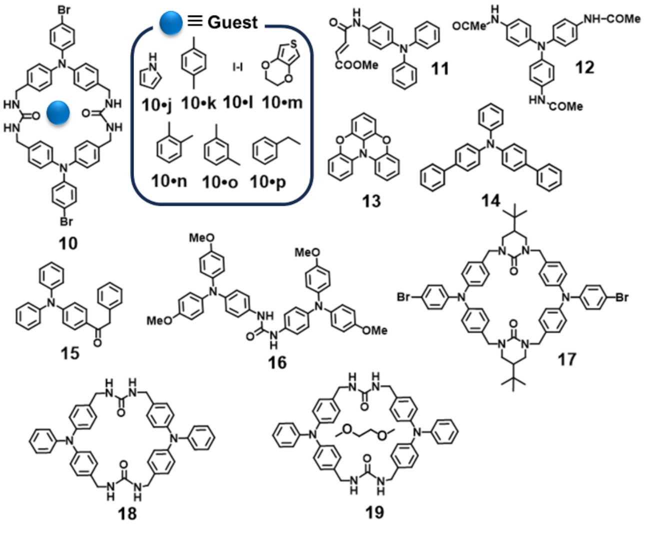


Figure 5. Molecules obtained from CCDC to apply the model.

Interestingly, the ML model predicts four compounds (**10•j**, **10•k**, **10•l** and **18**) to have high PIRs of > 0.35% with a maximum of 0.41% (Table S7). The predictions were tested on an independent validation dataset to evaluate the ML model's performance. Given their commercial or prior synthesis in our labs, four compounds were chosen: **10•m**, **14**, **18**, **19**. Compound **18**, is predicted to have high radical production, while other three molecules are expected to exhibit moderate PIR% values. The PIRs for triply recrystallized samples of **10•m** and **14** were experimentally quantified by EPR spectroscopy (See the Experimental Section), while values for **18** and **19** have recently become available.[44] These measured and predicted values are tabulated in Table 3. The model achieves an accurate ranking of the compounds based on their PIR% (**18** > **10•m** > **14** > **19**), where molecule **18** exhibits the highest PIR% and **19** the lowest.

Next, we analyze how the model performed versus the experimental measurements (Table 3). The host-guest complexes of **10** are well represented in the model training set. Not surprisingly, the model did quite well with **10•m**, which had a predicted value of 0.35% (standard deviation 0.031%) compared with the experimental value of 0.36%. The related TPA macrocycle **18** lacks exterior bromide and a guest, which differs from the macrocycles used in the training set. It exhibited a surprising experimental PIR of 1.4%.[44] The predicted value is 0.36% with a standard deviation of 0.026% and is an outlier. The deviation between the experimental and predicted PIR% arises because the model was initially trained on PIR values with a maximum of 0.85%. Consequently, it struggles to predict values as high as 1.4%. Comparatively, its' host: guest complex with dimethoxyethane, **19**, is well predicted by the model, with a predicted PIR% of 0.16% compared to its experimental value of 0.24%. Compound **14** is a mono TPA with two para substituents (N_{para}) and has a predicted value of 0.17% (with a standard deviation of 0.033%) versus a measured PIR of 0.27%. Overall, although the deviation between the predicted and experimental PIR values for molecule **18** is greater than that for molecules **14**, **19,** and **10•m**, the model accurately ranks the compounds by their PIR% (**18** > **10•m** > **14** > **19**).

Table 3. Actual and predicted PIR% of **10•m**, **14**, **18**, **19**.

Molecule	Experimental PIR%	Predicted PIR%
10•m	0.36	0.35
14	0.27	0.17
18	1.4	0.36
19	0.24	0.16

4. Conclusion

This study highlights the relationship between crystalline TPA derivatives' structural and electronic properties and their corresponding PIR percentages. By analyzing 18 TPA derivatives sourced from the CCDC, we trained a ML model to predict the experimentally generated PIR generation using structural and calculated properties. The model utilized Random Forest with leave-one-out cross-validation and displayed an MAE = 0.17%.

Analysis of trends within the TPAs used for training the model illustrates several design rules for a high PIR generation. Specifically, macrocycles consistently display higher PIR**%** compared to linear and mono TPA analogs, which is attributed to their rigid scaffold and molecular planarization. Furthermore, our analysis reveals the favorable impact of bromine substitutions on PIR generation, contrasting with the low PIR**%** associated with iodine substitutions. This emphasizes the role of bromine substitution in modulating the photophysical properties of TPA derivatives. Moreover, analysis of these molecules unveils intriguing correlations between their structural and electronic features with PIR percentages. In linear analogs, molecules with higher PIR**%** tend to exhibit shorter *d*, implying that reduced *d*-values may facilitate enhanced through-space coupling or electron transfer processes, potentially leading to increased PIR generation. Conversely, while macrocycles also demonstrate reduced *d*-values compared to linear analogs, their PIR**%** do not significantly vary with these values, likely due to their rigid scaffold structures. Additionally, TPA derivatives with higher PIR**%** tend to have smaller θ values, indicative of increased planarization and radical cation delocalization. The presence of para substituents on TPAs contributes to PIRs, likely by impeding degradation processes. Regarding electronic features decreased ionization potential (Ip) and increased excited state oscillator strength (*f*) correlate with increased PIR generation. Electron-hole covariance (cov) increases with increasing PIR% in linear analogs while macrocycles showed a complex variation with PIR and cov. Excited state dipole moment (D) shows a complex trend in PIR generation in macrocycles and linear analogs, whereas root-mean-square (RMS) electron-hole separation (d_{RMS}), transition dipole moment (μ), excitation energy (Ex), electronic coupling between ground and first singlet excited state (V_{01}) and charge difference between monomers in ground state (Δq) do not show clear trends.

Notably, the study extends beyond elucidating these relationships to demonstrate the practical utility of the ML model in predicting PIR**%** in new TPA-containing crystal structures. By applying our model to 16 molecules obtained from the CCDC, we illustrate its potential for screening and predicting PIR**%** in diverse molecular contexts. The model gave good predictions of the hierarchy of PIR% compared with the experimental values of an independent validation dataset of TPA molecules. Expanding our dataset to a broader range of crystalline TPA derivatives holds promise for further enhancing the accuracy and applicability of our ML model. Additionally, we plan to implement this model as a pre-synthesis screening tool to identify more TPA macrocycle-guest systems that generate higher PIR%.

ASSOCIATED CONTENT

Supporting Information. Computational data used for the training and validation datasets, Training data structures, Validation data structures, Model Training and Validation, ^{1}H NMR of **14** and **1•D**, PXRD and EPR data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Linda S. Shimizu - Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, United States; Orcidhttps://orcid.org/0000-0001-5599-4960; Email: shimizls@mailbox.sc.edu.

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported in part by National Science Foundation (NSF) CHE-2203830, CHE-1904386, OIA-1655740. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NSF.

ACKNOWLEDGMENT

We thank Dr. Faizul Islam for providing samples of brominated TPA bis-urea macrocycle.

ABBREVIATIONS

AFM, Atomic force microscopy; DLS, Dynamic Light Scattering; DMSO, Dimethyl sulfoxide; EPR, Electron Paramagnetic Resonance; LED, Light Emitting Diode; MAE, Mean Absolute Error; ML, Machine Learning; TPA, Triphenylamine; PIRs, Photo Induced Radicals; LOOCV, leave-one-out cross validation; FCD, Fragment Charge Difference

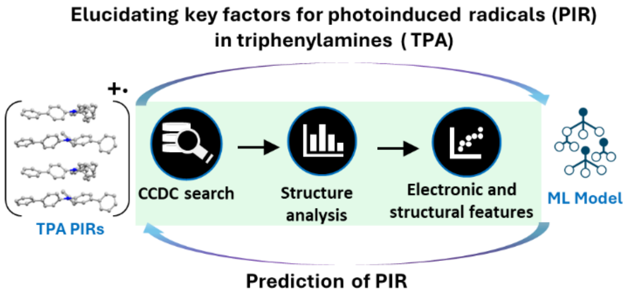
REFERENCES

1. Moulin, E.; Armao, J. J.; Giuseppone, N. Triarylamine-Based Supramolecular Polymers: Structures, Dynamics, and Functions. *Acc. Chem. Res.* **2019**, *52* (4), 975–983. https://doi.org/10.1021/acs.accounts.8b00536.
2. Wang, J.; Liu, K.; Ma, L.; Zhan, X. Triarylamine: Versatile Platform for Organic, Dye-Sensitized, and Perovskite Solar Cells. *Chem. Rev.* **2016**, *116* (23), 14675–14725. https://doi.org/10.1021/acs.chemrev.6b00432.
3. Tan, G.; Wang, X. Isolable Bis (Triarylamine) Dications: Analogues of Thiele's, Chichibabin's, and Müller's Hydrocarbons. *Acc. Chem. Res.* **2017**, *50* (8), 1997–2006. https://doi.org/10.1021/acs.accounts.7b00229.
4. Quinton, C.; Alain-Rizzo, V.; Dumas-Verdes, C.; Miomandre, F.; Clavier, G.; Audebert, P. Redox-Controlled Fluorescence Modulation (Electrofluorochromism) in Triphenylamine Derivatives. *RSC Adv.* **2014**, *4* (65), 34332–34342. https://doi.org/10.1039/C4RA02675F.
5. Busseron, E.; Cid, J.-J.; Wolf, A.; Du, G.; Moulin, E.; Fuks, G.; Maaloum, M.; Polavarapu, P.; Ruff, A.; Saur, A.-K., et al. Light-Controlled Morphologies of Self-Assembled Triarylamine–Fullerene Conjugates. *ACS Nano* **2015**, *9* (3), 2760–2772. https://doi.org/10.1021/nn506646m.
6. Shih, P.-I.; Chien, C.-H.; Wu, F.-I.; Shu, C.-F. A Novel Fluorene-Triphenylamine Hybrid That Is a Highly Efficient Host Material for Blue-, Green-, and Red-Light-Emitting Electrophosphorescent Devices. *Adv. Funct. Mater.* **2007**, *17* (17), 3514–3520. https://doi.org/10.1002/adfm.200700191.
7. Farokhi, A.; Shahroosvand, H.; Monache, G. D.; Pilkington, M.; Nazeeruddin, M. K. The Evolution of Triphenylamine Hole Transport Materials for Efficient Perovskite Solar Cells. *Chem. Soc. Rev.* **2022**, *51* (14), 5974–6064. https://doi.org/10.1039/D1CS01157J.

8. Sindt, A. J.; DeHaven, B. A.; McEachern, D. F.; Dissanayake, D. M. M. M.; Smith, M. D.; Vannucci, A. K.; Shimizu, L. S. UV-Irradiation of Self-Assembled Triphenylamines Affords Persistent and Regenerable Radicals. *Chem. Sci.* **2019**, *10* (9), 2670–2677. https://doi.org/10.1039/C8SC04607G.
9. Rong, X.; Liu, J.; Wu, J.; Li, C.; Wang, K.; Lu, Z.; Liu, Y.; Gu, M.; Huang, Y. A Feasible Strategy to Obtain Air-Stable Triphenylamine Radicals in the Solid State by Introducing Conjugated Donor–Acceptor Modules. *J. Mater. Chem. C* **2023**, *11* (28), 9640–9648. https://doi.org/10.1039/D3TC01187A.
10. Schaub, T. A.; Mekelburg, T.; Dral, P. O.; Miehlich, M.; Hampel, F.; Meyer, K.; Kivala, M. A Spherically Shielded Triphenylamine and Its Persistent Radical Cation. *Chem. Eur. J* **2020**, *26* (15), 3264–3269. https://doi.org/10.1002/chem.202000355.
11. Hellwinkel, D.; Melan, M. Heteropolycyclen Vom Triangulen-Typ, I. 8.12-Dihydro-4H-Benzo[1.9]Chinolizino[3.4.5.6.7-Defg]Acridin-Trion-(4.8.12) Und 5.9-Dihydro-Chino[3.2.1-de]Acridin-Dion-(5.9). *Chem. Ber.* **1971**, *104* (4), 1001–1016. https://doi.org/10.1002/cber.19711040406.
12. Kuratsu, M.; Kozaki, M.; Okada, K. 2,2′:6′,2″:6″,6-Trioxytriphenylamine: Synthesis and Properties of the Radical Cation and Neutral Species. *Angew. Chem. Int. Ed.* **2005**, *44* (26), 4056–4058. https://doi.org/10.1002/anie.200500397.
13. Forrester, A. R.; Hay, J. M.; Thompson, R. H. Organic Chemistry of Stable Free Radicals; Academic Press: New York, **1968**.
14. Mao, L.; Zhou, M.; Shi, X.; Yang, H. B. Triphenylamine (TPA) Radical Cations and Related Macrocycles. *Chin. Chem. Lett.* **2021**, *32* (11), 3331–3341. https://doi.org/10.1016/j.cclet.2021.05.004.
15. Moulin, E.; Niess, F.; Maaloum, M.; Buhler, E.; Nyrkova, I.; Giuseppone, N. The Hierarchical Self-Assembly of Charge Nanocarriers: A Highly Cooperative Process Promoted by Visible Light. *Angew. Chem. Int. Ed.* **2010**, *49* (39), 6974–6978. https://doi.org/10.1002/anie.201001833.
16. Huo, G.-F.; Shi, X.; Tu, Q.; Hu, Y.-X.; Wu, G.-Y.; Yin, G.-Q.; Li, X.; Xu, L.; Ding, H.-M.; Yang, H.-B. Radical-Induced Hierarchical Self-Assembly Involving Supramolecular Coordination Complexes in Both Solution and Solid States. *J. Am. Chem. Soc.* **2019**, *141* (40), 16014–16023. https://doi.org/10.1021/jacs.9b08149.
17. Mu, Y.; Liu, Y.; Tian, H.; Ou, D.; Gong, L.; Zhao, J.; Zhang, Y.; Huo, Y.; Yang, Z.; Chi, Z. Sensitive and Repeatable Photoinduced Luminescent Radicals from A Simple Organic Crystal. *Angew. Chem. Int. Ed.* **2021**, *60* (12), 6367–6371. https://doi.org/10.1002/anie.202014720.
18. Sindt, A. J.; Smith, M. D.; Berens, S.; Vasenkov, S.; Bowers, C. R.; Shimizu, L. S. Single-Crystal-to-Single-Crystal Guest Exchange in Columnar Assembled Brominated Triphenylamine Bis-Urea Macrocycles. *Chem. Commun.* **2019**, *55* (39), 5619–5622. https://doi.org/10.1039/c9cc01725a.
19. Hossain, M. S.; Sindt, A. J.; Goodlett, R. L.; Shields, D. J.; O'Connor, C. J.; Antevska, A.; Karakalos, S. G.; Smith, M. D.; Garashchuk, S.; Do, T. D.; Gudmundsdottir, A. D.; Shimizu, L. S. Effects of Self-Assembly on the Photogeneration of Radical Cations in Halogenated Triphenylamines. *J. Phys. Chem. C* **2021**, *125* (36), 19991–20002. https://doi.org/10.1021/acs.jpcc.1c04933.
20. Sindt, A. J.; Dehaven, B. A.; Goodlett, D. W.; Hartel, J. O.; Ayare, P. J.; Du, Y.; Smith, M. D.; Mehta, A. K.; Brugh, A. M.; Forbes, M. D. E., et al. Guest Inclusion Modulates Concentration and Persistence of Photogenerated Radicals in Assembled Triphenylamine Macrocycles. *J. Am. Chem. Soc.* **2020**, *142* (1), 502–511. https://doi.org/10.1021/jacs.9b11518.
21. Hossain, M. S.; Ahmed, F.; Karakalos, S. G.; Smith, M. D.; Pant, N.; Garashchuk, S.; Greytak, A. B.; Docampo, P.; Shimizu, L. S. Structure–Property Investigations in Urea Tethered Iodinated Triphenylamines. *Phys. Chem. Chem. Phys.* **2022**, *24* (31), 18729–18737. https://doi.org/10.1039/D2CP01856J.
22. Islam, M. F.; Sindt, A. J.; Hossain, M. S.; Ayare, P. J.; Smith, M. D.; Vannucci, A. K.; Garashchuk, S.; Shimizu, L. S. Assembled Triphenylamine Bis-Urea Macrocycles: Exploring Photodriven Electron Transfer from Host to Guests. *Phys. Chem. Chem. Phys.* **2021**, *23* (41), 23953–23960. https://doi.org/10.1039/D1CP03000K.
23. Shi, Y.-F.; Yang, Z.-X.; Ma, S.; Kang, P.-L.; Shang, C.; Hu, P.; Liu, Z.-P. Machine Learning for Chemistry: Basics and Applications. *Engineering* **2023,** *27*, 70-83. https://doi.org/10.1016/j.eng.2023.04.013.
24. Zhang, S.-Q.; Xu, L.-C.; Li, S.-W.; Oliveira, J. C. A.; Li, X.; Ackermann, L.; Hong, X. Bridging Chemical Knowledge and Machine Learning for Performance Prediction of Organic Synthesis. *Chem. Eur. J.* **2023**, *29* (6), e202202834. https://doi.org/10.1002/chem.202202834
25. Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **2021**, *121* (16), 9816–9872. https://doi.org/10.1021/acs.chemrev.1c00107.
26. Jiang, Y.; Yang, Z.; Guo, J.; Li, H.; Liu, Y.; Guo, Y.; Li, M.; Pu, X. Coupling Complementary Strategy to Flexible Graph Neural Network for Quick Discovery of Coformer in Diverse Co-Crystal Materials. *Nat. Commun.* **2021**, *12* (1), 5950. https://doi.org/10.1038/s41467-021-26226-7.
27. Clements, R. J.; Dickman, J.; Johal, J.; Martin, J.; Glover, J.; Day, G. M. Roles and Opportunities for Machine Learning in Organic Molecular Crystal Structure Prediction and Its Applications. *MRS Bull.* **2022**, *47* (10), 1054–1062. https://doi.org/10.1557/s43577-022-00434-y.
28. Olsthoorn, B.; Geilhufe, R. M.; Borysov, S. S.; Balatsky, A. V. Band Gap Prediction for Large Organic Crystal Structures with Machine Learning. *Adv. Quantum Technol.* **2019**, *2* (7–8), 1900023. https://doi.org/10.1002/qute.201900023.
29. Tawfik, S. A.; Isayev, O.; Spencer, M. J. S.; Winkler, D. A. Predicting Thermal Properties of Crystals Using Machine Learning. *Adv. Theory Simul.* **2020**, *3* (2), 1900208. https://doi.org/10.1002/adts.201900208.
30. Nguyen, P.; Loveland, D.; Kim, J. T.; Karande, P.; Hiszpanski, A. M.; Han, T. Y.-J. Predicting Energetics Materials' Crystalline Density from Chemical Structure by Machine Learning. *J. Chem. Inf. Model.* **2021**, *61* (5), 2147–2158. https://doi.org/10.1021/acs.jcim.0c01318.
31. Voityuk, A. A.; Rösch, N. Fragment Charge Difference Method for Estimating Donor–Acceptor Electronic Coupling: Application to DNA π-Stacks. *J. Chem. Phys.* **2002**, *117* (12), 5607–5616. https://doi.org/10.1063/1.1502255
32. Macrae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.; Platings, M.; Shields, G. P.; Stevens, J. S.; Towler, M., et al. Mercury 4.0: From Visualization to Analysis, Design and Prediction. *J. Appl.*

Crystallogr. **2020**, *53* (1), 226–235. https://doi.org/10.1107/S1600576719014092.

33. Sun, D.-L.; Rosokha, S. V; Lindeman, S. V; Kochi, J. K. Intervalence (Charge-Resonance) Transitions in Organic Mixed-Valence Systems. Through-Space versus Through-Bond Electron Transfer between Bridged Aromatic (Redox) Centers. *J. Am. Chem. Soc.* **2003**, *125* (51), 15950–15963. https://doi.org/10.1021/ja037867s.
34. Preat, J. Photoinduced Energy-Transfer and Electron-Transfer Processes in Dye-Sensitized Solar Cells: TDDFT Insights for Triphenylamine Dyes. *J. Phys. Chem. C* **2010**, *114* (39), 16716–16725. https://doi.org/10.1021/jp1050035.
35. Pan, J.-H.; Chiu, H.-L.; Chen, L.; Wang, B.-C. Theoretical Investigations of Triphenylamine Derivatives as Hole Transporting Materials in OLEDs: Correlation of the Hammett Parameter of the Substituent to Ionization Potential, and Reorganization Energy Level. *Comput. Mater. Sci.* **2006**, *38* (1), 105–112. https://doi.org/10.1016/j.commatsci.2006.01.011
36. Patra, A.; Pipim, G. B.; Krylov, A. I.; Mallikarjun Sharada, S. Performance of Density Functionals for Excited-State Properties of Isolated Chromophores and Exciplexes: Emission Spectra, Solvatochromic Shifts, and Charge-Transfer Character. *J. Chem. Theory Comput.* **2024,** *20* (6), 2520–2537. https://pubs.acs.org/doi/10.1021/acs.jctc.4c00005
37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
38. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947–1958. https://doi.org/10.1021/ci034160g.
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.
40. Genuer, R.; Poggi, J.-M. Random Forests. In *Random Forests with R*; Genuer, R., Poggi, J.-M., Eds.; Springer International Publishing: Cham, **2020**; pp 33–55. https://doi.org/10.1007/978-3-030-56485-8_3.
41. Huang, C.-Y.; Hsu, C.-Y.; Yang, L.-Y.; Lee, C.-J.; Yang, T.-F.; Hsu, C.-C.; Ke, C.-H.; Su, Y. O. A Systematic Study of Electrochemical and Spectral Properties for the Electronic Interactions in Porphyrin–Triphenylamine Conjugates. *Eur. J. Inorg. Chem.* **2012**, *2012* (7), 1038–1047. https://doi.org/10.1002/ejic.201101033
42. Kanno, M.; Ono, Y.; Kono, H.; Fujimura, Y. Laser-Polarization Effects on Coherent Vibronic Excitation of Molecules with Quasi-Degenerate Electronic States. *J. Phys. Chem. A* **2012**, *116* (46), 11260–11272. https://doi.org/10.1021/jp305284w.
43. Saeki, A.; Seki, S.; Satoh, N.; Yamamoto, K.; Tagawa, S. Long-Lived Hole Stabilized at a Triphenylamine Core and Shielded by Rigid Phenylazomethine Dendrons: A Pulse Radiolysis Study. *J. Phys. Chem. B* **2008**, *112* (49), 15540–15545. https://doi.org/10.1021/jp805266v.
44. Islam, M. F.; Wijesekera, G. I. P.; Sindt, A. J.; Smith, M. D.; Shimizu, L. S. Examination of Photoinduced Radicals in Two Crystal Forms of Triphenylamine Bis-Urea Macrocycles. *CrystEngComm*. **Submitted.** CE-ART-08-2024-000839.



TOC Graphic