Journal of
Materials Chemistry C

ROYAL SOCIETY
OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d5tc04137f

Received 22nd November 2025,
Accepted 8th March 2026

DOI: 10.1039/d5tc04137f

rsc.li/materials-c

A simplified machine learning workflow for identifying potential singlet fission candidates: benzannulated biphenylenes as a case study

Iqra Sarfraz, Sergei F. Vyboishchikov, Miquel Solà * and Albert Artigas *

Singlet fission (SF) is a physical phenomenon exhibited by some families of organic materials potentially able to boost the power conversion efficiency of solar cells beyond the theoretical Shockley–Queisser limit of 33%. To experience SF, a molecule must fulfil the so-called energy matching conditions (EMCs), which can be evaluated using DFT and TD-DFT calculations. Here, we propose a simple protocol that exploits machine learning workflows to screen large libraries of molecules using a reduced number of quantum chemical calculations. The protocol is based on the AQME and ROBERT platforms and is adapted to users with no experience in data science and basic computational chemistry knowledge. Using this approach, we screened a library of 3835 benzannulated biphenylenes to identify 505 candidates fulfilling the first EMC for SF. Fragment-based statistical analysis was employed to rationalize the structural features associated with SF. The workflow is general and can be applied to other families of compounds for the accelerated discovery of SF materials.

Introduction

Singlet fission (SF) is a spin-allowed photophysical process in which a high-energy singlet exciton transfers its energy to a neighboring ground-state molecule, generating two triplet excitons (Fig. 1a). This mechanism allows the production of two triplets from a single absorbed photon, enabling external quantum efficiencies (EQEs) of up to 200%.[1,2] For this reason, SF has long been proposed as a promising strategy to enhance the power conversion efficiency (PCE) of third-generation photovoltaic (PV) devices.[3,4] If successfully integrated, SF-based solar cells could potentially exceed the Shockley–Queisser limit of 33%, the theoretical maximum efficiency for single-junction solar cells.[5] SF is an exclusive property of certain organic molecules and has only been observed in a limited number of compound families (Fig. 1b).[6–8] Among these, acenes are by far the most extensively studied systems.[9,10] Indeed, SF was first discovered several decades ago in anthracene crystals,[11] and pentacene became the first SF chromophore to demonstrate external quantum efficiencies (EQEs) higher than 100% in organic photovoltaic (OPV) devices produced by the Baldo group.[12,13] The same group fabricated SF-sensitized silicon solar cells achieving EQE > 130%.[14,15] Despite these advances, SF-based PV devices have so far exhibited very low PCEs.[3,7] A key challenge is the limited variety of known SF-active materials, leading to a broad consensus within the SF research community that new chromophore families must be discovered to unlock the full potential of this mechanism in practical applications.

For a molecule to undergo SF, it must satisfy at least two key energy matching conditions (EMCs): (i) the lowest singlet excited state (S_1) must have at least twice the energy of the lowest triplet state (T_1) to ensure exergonic SF, and (ii) the second triplet state (T_2) must also exceed twice the T_1 energy to prevent triplet–triplet annihilation (TTA).[1] However, these two energetic criteria alone do not guarantee efficient SF.[16] The chromophore must also exhibit a high absorption coefficient and an appropriate bandgap to capture high-energy photons. A fast SF rate is essential to outcompete other deactivation pathways, such as charge transfer from the S_1 state to the acceptor, internal conversion to S_0, or excimer formation. Likewise, the material should display a slow triplet–triplet annihilation (TTA) rate. In practical device configurations, additional electronic factors become crucial. For instance, proper alignment between the T_1 energy and the acceptor's bandgap ($\approx$1.1 eV for silicon-based solar cells)[17] or compatible frontier orbital energies (*e.g.*, LUMO matching in organic photovoltaics) to ensure efficient triplet harvesting.[18] Moreover, practical considerations also include chemical stability and synthetic feasibility, which are essential for real-world implementation.

Considering all these aspects, the screening of molecular systems fulfilling the first EMC remains a good starting point for SF material discovery. In this context, the identification of

Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, C/Maria Aurèlia Capmany, 69, 17003 Girona, Catalonia, Spain. E-mail: albert.artigas@udg.edu, miquel.sola@udg.edu

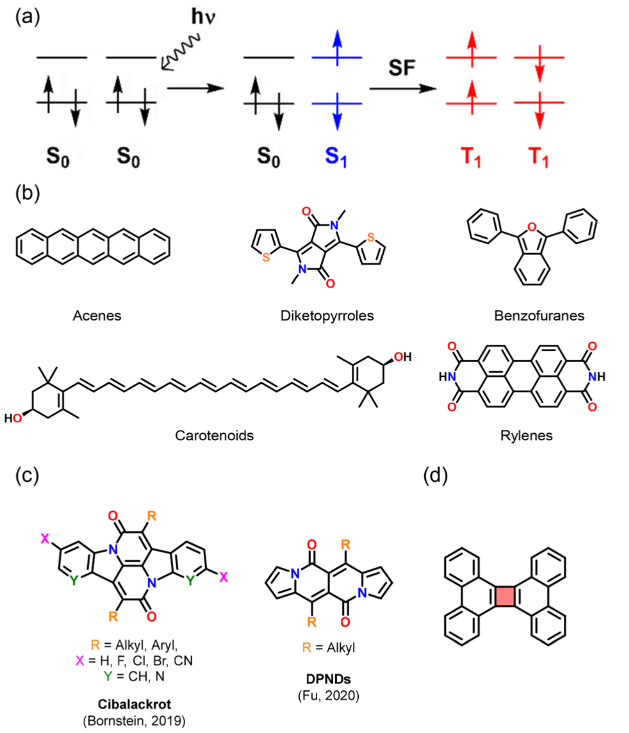

Fig. 1 (a) Simplified SF mechanism. (b) Conventional families of SF materials. (c) Aromatic chameleons (d) Archetypal benzannulated biphenylene with four fused benzene rings. Red color is used to highlight the antiaromatic cyclobutadiene core.

new SF materials has largely relied on *in silico* studies guided by these EMCs.[6] A paradigmatic example is the rational design of 1,3-diphenylisobenzofuran, which is considered the first rationally engineered SF chromophore.[19–21] Subsequent works have demonstrated that SF is operational in other open-shell systems.[22–24]

Excited state engineering can also be achieved through aromaticity manipulation. In this sense, polycyclic systems exhibiting ground-state (Hückel) antiaromaticity and excited-state (Baird) aromaticity show simultaneous stabilization of S_1 and T_1 states, enabling tailored SF behavior through aromatic/antiaromatic character modulation.[25] A more sophisticated approach involves the design of the so-called aromatic chameleons (Fig. 1c), which are defined as π-conjugated compounds that can modify their π-electron distributions to satisfy various aromaticity laws in various electronic states, in particular, in the S_0 and T_1 states.[26–31] Compounds that are aromatic in S_0 and T_1 states are candidates for SF applications because, commonly, they have a small singlet–triplet energy gap. While the computational search for new SF has been so far a fruitful strategy, evaluating the EMCs for large compound libraries is often tedious and computationally expensive, as it requires relatively accurate DFT and TD-DFT calculations. To accelerate this discovery phase, many groups have recently turned their attention to genetic algorithms[32] and machine learning (ML) techniques.[33–38] However, ML protocols can be complex to implement and reproduce, particularly for non-expert researchers. To address this challenge, here we propose a streamlined workflow using ROBERT[39] and AQME[40] packages—developed and maintained by the Alegre-Requena group—to efficiently screen large libraries of potential SF chromophores. This workflow is ideally suited for organic and materials chemists seeking to identify promising synthetic targets for SF applications.

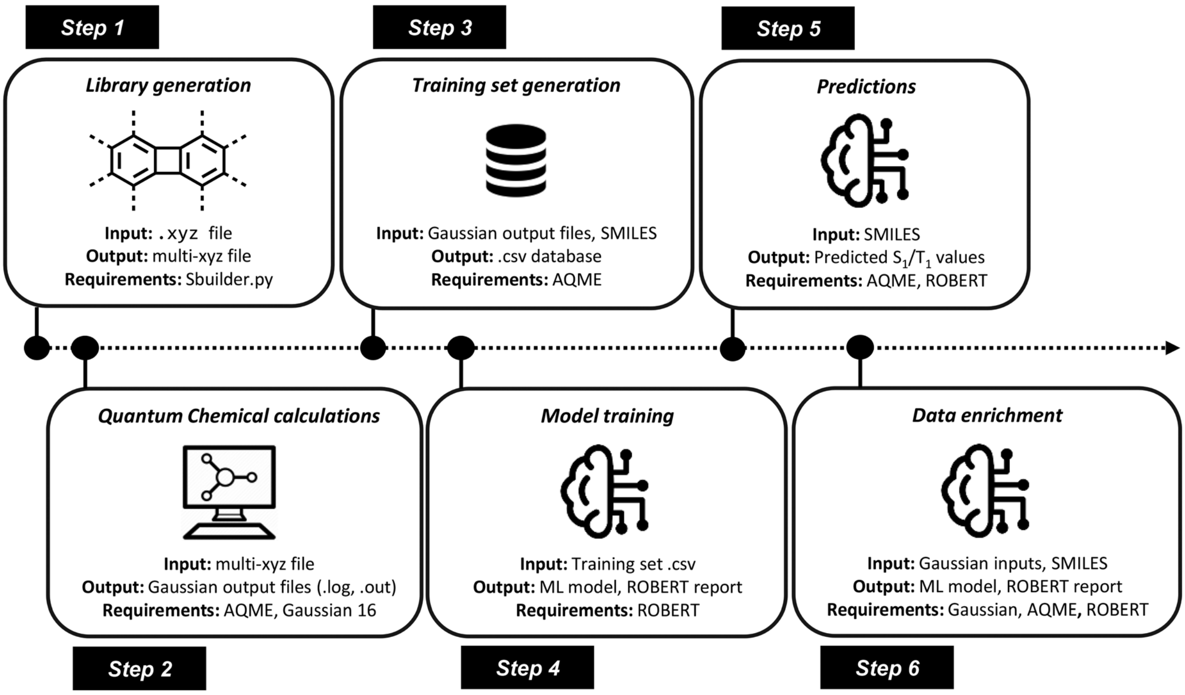
Step 1
Library generation
Input: .xyz file
Output: multi-xyz file
Requirements: Sbuilder.py
Step 2
Quantum Chemical calculations
Input: multi-xyz file
Output: Gaussian output files (.log, .out)
Requirements: AQME, Gaussian 16
Step 3
Training set generation
Input: Gaussian output files, SMILES
Output: .csv database
Requirements: AQME
Step 4
Model training
Input: Training set .csv
Output: ML model, ROBERT report
Requirements: ROBERT
Step 5
Predictions
Input: SMILES
Output: Predicted S_1/T_1 values
Requirements: AQME, ROBERT
Step 6
Data enrichment
Input: Gaussian inputs, SMILES
Output: ML model, ROBERT report
Requirements: Gaussian, AQME, ROBERT

Fig. 2 Overview of the ML protocol employed in this work to predict S_1/T_1 ratio of benzannulated biphenylenes.

Results and discussion

The goal of our study was to computationally screen a large library of benzannulated biphenylenes (Fig. 1d), a synthetically accessible[41,42] family of polycyclic conjugated hydrocarbons (PCH) with a central antiaromatic core. We anticipated these systems to be well suited for SF design due to their dual character: on the one hand, the central cyclobutadiene ring exhibits ground-state (Hückel) antiaromaticity, which destabilizes the S_0 while lowering the T_1 energy through excited-state (Baird) aromaticity. On the other hand, the surrounding polybenzene moieties, which can contain a number of Clar aromatic π-sextets provide thermodynamic stability to both the ground state (S_0) and lowest-lying excited states (S_1 and T_1).[43] This design was intended to identify candidates that meet the first EMC [$(E(S_1)/E(T_1) \geq 2$] to select a group of hit candidates for further development. An overview of the ML protocol employed is presented in Fig. 2 and all necessary input and output.csv files are provided as SI.

The first step of the protocol involved generating a library of benzannulated biphenylenes (step 1 in Fig. 2). To achieve this, we employed an in-house developed Python code[44] that successively attaches benzene rings to a given input geometry (xyz format) in a cata-condensed fashion (see Scheme S1 in the SI for further details). For our study, we considered six generations of cata-benzannulated biphenylenes. Thus, the program generated a family of 3835 distinct molecules, each containing 2–8 benzene rings fused to the Hückel antiaromatic cyclobutadiene core. The molecules are represented as graphs using Python's NetworkX module.[45] Importantly, the program performs a preliminary optimization of each structure with the Merck molecular force field (MMFF)[46] as implemented in RDKit.[47] Redundancy among newly generated molecules is assessed by computing the InChIKey[48] for each structure and checking its presence against the set of previously generated structures within the same generation. The final results are compiled into a single multi-xyz output file. Next step was the obtention of target values (step 2 in Fig. 2). Accordingly, 300 members of the library were randomly selected and transformed into multi-job Gaussian 16[49] input files. The S_0 and T_1 states of this set of molecules were optimized with the (U)M06-2X/def2-SVP method and their energies were obtained at the (U)M06-2X/def2-TZVPP//(U)M06-2X/def2-SVP (gas phase) level of theory.[50,51] This level of theory was selected based on previous results.[6,25] Vertical S_1 energies were obtained from TD-DFT calculations at the S_0 geometries at the same level of theory. Input file generation could be done automatically with a single command line using the QPREP module included in the AQME platform. Successfully terminated calculations with no imaginary frequencies (280 out of 300) were identified and the energies of each state were extracted and assembled in a spreadsheet that contained an identification code for each molecule together with their computed S_1/T_1 ratio.

We are aware that the use of vertical energies for the S_1 state results in somewhat overestimated S_1/T_1 ratios.[52] Additionally, each molecule was annotated with its SMILES code using the Open Babel program,[53] a requirement for the subsequent generation of chemical descriptors to be used during model training.

Training set generation and model training (steps 3 and 4 in Fig. 2) were carried out using the AQME and ROBERT platforms, respectively. Both processes are fully automated and can be executed with a single command line thanks to their streamlined integration. First, AQME's CSEARCH and CMIN modules perform an RDKit based conformational search, followed by xTB[54] geometry optimization starting from SMILES strings. Then, the QDESCP module generates three distinct databases containing a number of Boltzmann weighted xTB and RDKit descriptors, together with the desired target values (in our case, the S_1/T_1 ratio). These three databases—denoted full, denovo, and interpret—contain descriptor sets organized across increasing levels of human interpretability. The interpret database, which contains 22 human-interpretable molecular descriptors, was selected for ML model training due to its lower dimensionality, reduced redundancy, and enhanced interpretability. This database was then treated with the ROBERT platform (step 4 in Fig. 2), which operates through 4 consecutives modules. First, the CURATE module improves data quality by automatically suppressing highly correlated variables.

Next, the GENERATE module screens diverse combinations of built-in scikit-learn[55] algorithms—by default, random forest (RF), gradient boosting (GB), multivariate linear regression (MVL), and dense neural networks (NN) as implemented in sklearn.neural_network. MLPRegressor class—along with different partition sizes. It also employs permutation feature importance (PFI) analysis[56] to quantify the relative contribution of each descriptor to the model's predictive performance. Additionally, SHapley Additive exPlanations (SHAP)[57,58] analysis is used to assess the contribution of each descriptor to the model predictions. This method provides feature importance values based on Shapley theory.[59] In this approach, the values of each descriptor are randomly permuted, and the resulting decrease in model accuracy is used as a measure of its importance. The VERIFY module then carries out a series of tests to assess the performance and reliability of the ML predictors produced by the GENERATE module. Finally, the PREDICT module computes a set of performance metrics for the selected models. All results are compiled into a comprehensive report delivered as a pdf file, which also includes a 1-to-10 score derived from the model's performance metrics. The details regarding the ROBERT platform can be found the original publications[39,40] and the online documentation websites.

The most relevant results obtained from the first round of model training can be found in the corresponding ROBERT report (see SI). Out of the 280 molecules randomly chosen from the family of 3836 compounds, 224 (80%) were allocated to the training set for 10-times repeated 5-fold cross-validation, while the remaining 56 (20%) were set aside by the program as the test set. The best-performing models involved either a NN algorithm using 6 descriptors [model without PFI filtering, denoted as **NN (No PFI)**] or an MVL model using 4 descriptors (model with PFI filtering). In both cases, the overall score was

8/10. The **NN (No PFI)** model yielded particularly strong metrics: R^2 = 0.92, MAE = 0.046, and RMSE = 0.066, with only 3.6% outliers in both the test set. In both models, PFI and SHAP analyses identified the xTB-computed S_0–T_1 gap as the most important descriptor, which is fully consistent with the target variable under study. While the model successfully passed most of the diagnostic tests performed by the VERIFY module, a major issue was identified regarding data distribution (Fig. 3). That is, 86% of the data points were found concentrated in the first two quartiles, which may significantly limit the model's applicability. This bias in data distribution could limit the model's ability to reliably predict molecules with $S_1/T_1 \geq 2$, which are precisely the desired targets.

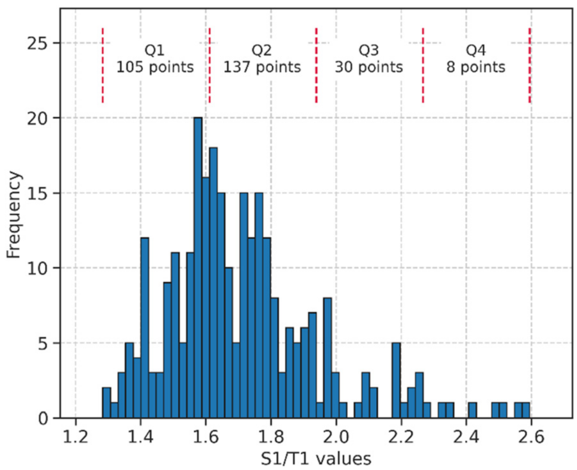

Fig. 3 Data distribution of target values within CV (10 times repeated 5-fold cross-validation) and test sets.

Taking this into account, we used the **NN (No PFI)** model to select additional compounds predicted to belong to the upper quartiles of the S_1/T_1 ratio distribution (step 5 in Fig. 2). While this model was expected to be less reliable in the 2.0–2.6 range, it could still identify molecules for which $S_1/T_1 \geq 2$ with remarkable accuracy. Indeed, all picked systems (100 systems with predicted S_1/T_1 in the 2.25–2.6 range, see SI) were found to satisfy the first EMC. Based on these predictions, a new training set was assembled by repeating steps 2 and 3 of the protocol.

This new set comprised 369 molecules, including the original 280 systems used in round 1 and 89 extra data points with computed S_1/T_1 in the 2.2–3.0 range (Fig. 4a). The best algorithm identified in a second round of ROBERT model training again the **NN (No PFI)**, which employed 6 AQME-generated

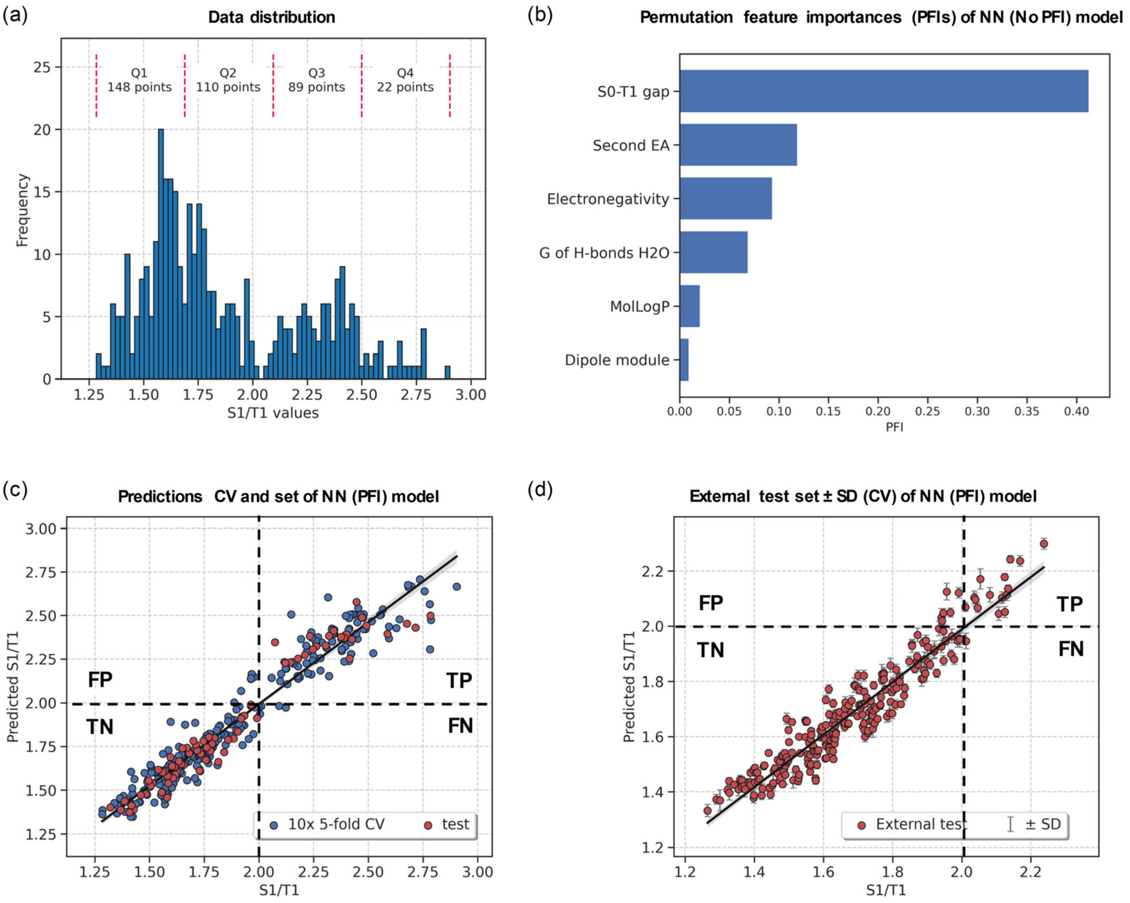

Fig. 4 (a) Data distribution in round 2. (b) PFI analysis of the **NN No PFI** model obtained in round 2. (c) Predictions for the CV (10 times repeated 5-fold cross-validation) and test set of the PFI model in round 2. (d) Predictions for the external set of the PFI model. Abbreviations: FP = False Positive; TP = True Positive: TN = True Negative; FN = False Negative.

descriptors (Fig. 4b): S_0–T_1 gap, second electron affinity, G of H bonds H_2O, electronegativity, MolLogP and dipole moment.† The xTB-computed S_0–T_1 gap once more emerged as the most important contributor based on PFI and SHAP analyses (see ROBERT report in the SI). The overall model score remained at 8/10, and excellent test-set performance metrics were computed using the PREDICT module: R^2 = 0.96, MAE = 0.057, RMSE = 0.076 (Fig. 4c). The proportion of outliers was 3.7% for the cross validation (CV) set and 5.4% for the test set. In terms of binary (qualitative) prediction, an accuracy of 96.2% and an *F*-score of 0.95 was obtained. Moreover, all false negatives yielded predicted S_1/T_1 values above 1.90, very close to the 2.0 singlet-fission threshold. Likewise, all false positives have actual S_1/T_1 that are also above 1.90.

Despite the good predictive metrics obtained, it should be noted that only the three most relevant descriptors identified by PFI and SHAP analyses (Fig. 4b) show a clear connection to the molecular electronic structure and are thus related to the energetic requirements for SF at some extent. In contrast, the remaining descriptors (*i.e.* G of H bonds H_2O, dipole moment, and MolLogP) appear to lack a direct physical relationship with the process. For this reason, we selected the PFI-filtered model [denoted **NN (PFI)** in Fig. 4c and d], which includes only the three most relevant and chemically interpretable descriptors (*i.e.* S_0–T_1 energy gap, electronegativity, and second electron affinity). This simplified model maintained excellent performance (R^2 = 0.94, MAE = 0.066, RMSE = 0.092 for the test set) and achieved an even higher ROBERT score of 9/10. The consistency between descriptor relevance (PFI) and chemical interpretability confirms that the model's predictive power is driven by physically meaningful features. Moreover, this outcome highlights the importance of combining statistical feature selection with chemical interpretability to ensure that the resulting models remain both accurate and insightful.

To further evaluate the workflow's performance, this second round of model training was repeated using three different combinations of two descriptors and S_0–T_1 energy gap alone (see SI for details). While good performance (ROBERT score up to 9) was obtained using an NN and the S_0–T_1 gap alone or in combination with electronegativity, the predictive power in the range of interest was found poorer. Therefore, we retained the **NN (PFI)** model employing all three physically meaningful descriptors as the champion model. To assess the model's robustness, its performance was evaluated on an external random test set of 257 molecules (Fig. 4d). The results remained consistent, with R^2 = 0.90, MAE = 0.053, and RMSE = 0.063. In terms of binary predictions, only 11 of 257 results are misclassified (6 false negatives plus 5 false positives). This corresponds to an accuracy of 95.7%, a precision (true positives divided by all positive predictions) of 86.8%, and an *F*-score of 0.86.

This model was finally employed to predict once more the S_1/T_1 value of all 3835 compounds in the library. As a result, 505 systems with predicted values $\geq$1.90 were identified as hit compounds for further investigation and development (see Fig. 5a for selected examples and the SI for the full list in multi-xyz format).

While the present study is primarily data-driven, some degree of chemical interpretation of the hit molecules is desirable. To rationalize the identified hits in terms of their benzannulation patterns, we first analyzed the occurrence of all fragments obtained by fusing one to four benzene rings to a single side of the cyclobutadiene core (fragments **F1–F27** in Fig. S1) and compared their relative occurrence within the hit subset to that in the full library of 3835 compounds. Substructure searches were performed using functions available in the RDKit.Chem module.[41] The results indicate that no simple, clear-cut prediction based on the presence of a single substructure is possible. Nevertheless, one fragment, namely the relatively simple cyclobuta[*a*]anthracene motif (**F5** in Fig. 5b), exhibits an increased relative occurrence among the hit molecules. Using the presence of this fragment as a binary classifier yields 398 true positives, 661 false positives, 107 false negatives, and 2669 true negatives. This corresponds to an accuracy of 80%, a precision of 38%, a sensitivity of 79%, and an *F*-score of 0.51. In practical terms, this criterion allows a large fraction of unsuitable candidates to be correctly discarded. However, it is insufficiently selective, as the proportion of false positives (the structures that have the fragment in question but show S_1/T_1 $\leq$ 1.90) remains substantial, limiting its usefulness as a standalone predictor. None of the other investigated substructures display statistically meaningful predictive power. Complete statistics for all fragments are provided in the SI (Table S2).

The preference for the cyclobuta[*a*]anthracene motif indicates that linearly cata-condensed π-extension adjacent to the cyclobutadiene core favors SF propensity within the analyzed family. This trend is chemically consistent with the well-known behavior of linearly fused acenes, which are established SF chromophores.[1,9] It should be noted, however, that isolated anthracene does not satisfy the first energetic matching criterion, whereas longer acenes such as tetracene and pentacene do. Our results suggest that, in benzannulated biphenylenes, a three-ring linear fragment can already be sufficient. This difference is attributed to the synergistic effect of linear conjugation and the presence of the antiaromatic core. From a practical

† S_0–T_1 gap refers to the energy difference between the ground singlet state and the lowest triplet excited state computed with xTB. The values are given in kcal/mol. Second EA refers to the energy difference between the dianion and the monoanion. It is estimated vertically from xTB-computed self-consistent-charge (SCC) energies of the monoanion and dianion plus an xTB correction factor. Values are given in eV. Electronegativity refers to the ability of a molecule to attract electrons. It is estimated as the average of the xTB-computed SCC ionization potential and the xTB-computed SCC electron affinity. Values are given in eV. G of H bonds H_2O refers to the Gibbs energy of interaction through hydrogen bonds between the molecule and an implicit solvation model, the so-called analytical linearized Poisson–Boltzmann (ALPB) that uses water as the solvent. MolLogP refers to the logarithm of the molecule's partition coefficient between octanol and water, which estimates its lipophilicity (hydrophobicity). It is estimated with RDKit using the Wildman–Crippen fragment method,[60] in which predefined atom-environment fragments contribute additively to the final value. This model assigns each atom a fragment constant based on its type and bonding context, and the molecular log*P* is obtained as the sum of these contributions. Dipole moment refers to the xTB computed molecular dipole moment module expressed in Debye (Note that the AQME and ROBERT programs incorrectly refer to this descriptor as dipole module).

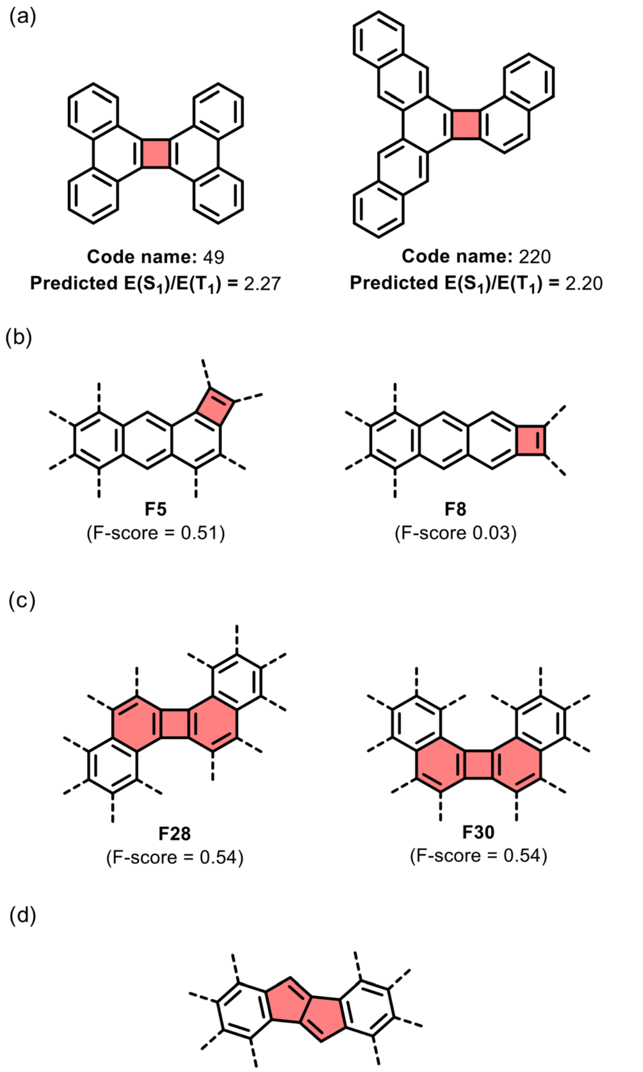

Fig. 5 (a) Two examples of benzannulated biphenylenes predicted as hits. (b) Two isomeric benzocyclobutaanthracene fragments considered in the structural analysis of the hit library. (c) Kinked dibenzobiphenylene fragments associated with singlet fission propensity among the analyzed hits. (d) General structure of benzannulated dibenzopentalenes.

perspective, this finding is also insightful, since shorter linear acene segments are generally more stable and synthetically more accessible than longer acenes. The latter typically suffer from reduced stability and handling difficulties. In contrast, kinked annulation patterns tend to shift the excited state energies outside the SF window. Intriguingly, the closely related and more linear cyclobuta[*b*]anthracene (**F8** in Fig. 5b) exhibits a very low *F*-score of 0.03, suggesting that a strictly collinear alignment of the substituents across the two sides of the cyclobutadiene core is disfavored. To further examine the linear *versus* kinked annulation, we analyzed the occurrence of four fragments that correspond to second-generation structures (**F28–F31** in Fig. S1). Among these, the two kinked isomeric motifs (**F28** and **F30**) are significantly enriched among the hit compounds, with *F*-scores of 0.54. In contrast, the linear topology (**F29** and **F31**) is strongly depleted, with an *F*-score of 0.03 and 0.01, respectively. Overall, the later results confirm that although locally linear benzannulation on each side of the cyclobutadiene core correlates with SF propensity, the most favorable architectures are not globally linear. Instead, an overall kinked topology—*i.e.*, when considering the relative positioning of the fused benzene rings on the left and right sides of the cyclobutadiene core—is preferred.

Finally, in order to assess the robustness and transferability of the ML workflow, we applied the same descriptor set and training protocol to an independent dataset of 302 benzannulated dibenzopentalenes, a structurally distinct yet chemically related family of ground-state antiaromatic polycyclic molecules (Fig. 5d, see SI). The resulting model displays predictive performance comparable to that obtained for benzannulated biphenylenes (test set R^2 = 0.85, MAE = 0.039, RMSE = 0.047; ROBERT score = 8/10), indicating that the learned structure–property relationships are not specific to a single molecular family. When both datasets are combined into a single, more chemically diverse training set, the predictive accuracy is further improved (R^2 = 0.94, MAE = 0.052, RMSE = 0.073), demonstrating that the dominant descriptors remain meaningful across families and that increased chemical diversity enhances the overall generalization of the model.

Conclusions

To conclude, by leveraging the capabilities of the AQME and ROBERT platforms, we have accurately predicted the S_1/T_1 ratio of over 3800 benzannulated biphenylenes, a particular class of polycyclic conjugated hydrocarbons with an antiaromatic core and whose S_0 and T_1 states are stabilized by a number of Clar π-aromatic sextets. The use of machine learning (ML) techniques as implemented in the ROBERT package proved highly advantageous, as quantum chemical calculations were required for fewer than 10% of the molecules (369 compounds). The protocol is well suited to synthetic organic and materials chemists seeking predictive tools to guide the design and selection of target compounds. Although the protocol is constrained to compounds fulfilling only the first EMC, subsequent filtering based on T_2 energies and additional electronic and structural criteria can further refine the selection. In this sense, the proposed workflow serves as a valuable initial step in the identification of promising SF compounds. Notably, the hit compounds identified in this work are currently undergoing further computational refinement, and the synthesis of two optimized lead biphenylenes is underway. Results will be reported in due course. We showed this protocol is generalizable and can be adapted to other families of molecules, by tailoring the library generation algorithm accordingly.

Author contributions

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

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting the results of this study, including full input and output datasets, a multi-xyz file and ROBERT reports are openly available in the CORA repository at: https://doi.org/10.34810/data2473.

Supplementary information (SI): computational details, tables with metrics of model trainings, chemical structure of the fragments considered, compound library (multi-xyz file), input and output datasets (.csv files), and ROBERT reports (pdf files). See DOI: https://doi.org/10.1039/d5tc04137f.

Acknowledgements

We are grateful for financial support from the Agencia Española de Investigación (MCIN/AEI/10.13039/501100011033) for projects PID2023-147424NB-I00, PID2023-146849NB-I00, and RED2024-154178-T, and from the Generalitat de Catalunya for Project 2021SGR623 and ICREA Academia prize 2024 to M.S. The authors thankfully acknowledge RES resources provided by BSC-CNS in the Consorci de Serveis Universitaris de Catalunya (CSUC) to perform the QH-2025-1-0002 project. I. S. is grateful to the Universitat de Girona for the IFUdG2023/64 PhD fellowship and A. A. thanks the European Union's Framework Programme for Research and Innovation Horizon Europe under the Marie Skłodowska-Curie Grant Agreement No. 101106492 (project title: Full-Fission). The authors are indebted to Dr Juan V. Alegre-Requena for insightful recommendations on the use of ROBERT and AQME.

Notes and references

1 M. B. Smith and J. Michl, Singlet Fission, *Chem. Rev.*, 2010, **110**, 6891–6936.

2 D. Casanova, Theoretical Modeling of Singlet Fission, *Chem. Rev.*, 2018, **118**, 7164–7207.

3 A. J. Baldacchino, M. I. Collins, M. P. Nielsen, T. W. Schmidt, D. R. McCamey and M. J. Y. Tayebjee, Singlet fission photovoltaics: Progress and promising pathways, Chem, *Phys. Rev.*, 2022, **3**, 021304.

4 M. C. Hanna and A. J. Nozik, Solar conversion efficiency of photovoltaic and photoelectrolysis cells with carrier multiplication absorbers, *J. Appl. Phys.*, 2006, **100**, 74510.

5 W. Shockley and H. J. Queisser, Detailed Balance Limit of Efficiency of p-n Junction Solar Cells, *J. Appl. Phys.*, 1961, **32**, 510–519.

6 D. Padula, Ö. H. Omar, T. Nematiaram and A. Troisi, Singlet fission molecules among known compounds: finding a few needles in a haystack, *Energy Environ. Sci.*, 2019, **12**, 2412–2416.

7 R. Casillas, I. Papadopoulos, T. Ullrich, D. Thiel, A. Kunzmann and D. M. Guldi, Molecular insights and concepts to engineer singlet fission energy conversion devices, *Energy Environ. Sci.*, 2020, **13**, 2741–2804.

8 T. Ullrich, D. Munz and D. M. Guldi, Unconventional singlet fission materials, *Chem. Soc. Rev.*, 2021, **50**, 3485–3518.

9 J. Li, H. Cao, Z. Zhang, S. Liu and Y. Xia, Research Progress on Singlet Fission in Acenes and Their Derivatives, *Photonics*, 2022, **9**, 689.

10 E. Busby, T. C. Berkelbach, B. Kumar, A. Chernikov, Y. Zhong, H. Hlaing, X. Y. Zhu, T. F. Heinz, M. S. Hybertsen, M. Y. Sfeir, D. R. Reichman, C. Nuckolls and O. Yaffe, Multiphonon relaxation slows singlet fission in crystalline hexacene, *J. Am. Chem. Soc.*, 2014, **136**, 10654–10660.

11 S. Singh, W. J. Jones, W. Siebrand, B. P. Stoicheff and W. G. Schneider, Laser Generation of Excitons and Fluorescence in Anthracene Crystals, *J. Chem. Phys.*, 1965, **42**, 330–342.

12 D. N. Congreve, J. Lee, N. J. Thompson, E. Hontz, S. R. Yost, P. D. Reusswig, M. E. Bahlke, S. Reineke, T. Van Voorhis and M. A. Baldo, External quantum efficiency above 100% in a singlet-exciton-fission-based organic photovoltaic cell, *Science*, 2013, **340**, 334–337.

13 J. Lee, P. Jadhav, P. D. Reusswig, S. R. Yost, N. J. Thompson, D. N. Congreve, E. Hontz, T. Van Voorhis and M. A. Baldo, Singlet Exciton Fission Photovoltaics, *Acc. Chem. Res.*, 2013, **46**, 1300–1311.

14 M. Einzinger, T. Wu, J. F. Kompalla, H. L. Smith, C. F. Perkinson, L. Nienhaus, S. Wieghold, D. N. Congreve, A. Kahn, M. G. Bawendi and M. A. Baldo, Sensitization of silicon by singlet exciton fission in tetracene, *Nature*, 2019, **571**, 90–94.

15 N. Nagaya, K. Lee, C. F. Perkinson, A. Li, Y. Lee, X. Zhong, S. Lee, L. P. Weisburn, J. Z. Wang, T. K. Baikie, M. G. Bawendi, T. Van Voorhis, W. A. Tisdale, A. Kahn, K. Seo and M. A. Baldo, Exciton fission enhanced silicon solar cell, *Joule*, 2025, 101965.

16 J. C. Johnson, A. J. Nozik and J. Michl, The Role of Chromophore Coupling in Singlet Fission, *Acc. Chem. Res.*, 2013, **46**, 1290–1299.

17 C. J. Lee, A. Sharma, N. A. Panjwani, I. M. Etchells, E. M. Gholizadeh, J. M. White, P. E. Shaw, P. L. Burn, J. Behrends, A. Rao and D. Jones, Toward Silicon-Matched Singlet Fission: Energy-Level Modifications Through Steric Twisting of Organic Semiconductors, *Adv. Opt. Mater.*, 2024, **12**, 2301539.

18 C. J. Priya Jadhav, P. R. Brown, N. Thompson, B. Wunsch, A. Mohanty, S. R. Yost, E. Hontz, T. Van Voorhis, M. G. Bawendi, V. Bulovic, M. A. Baldo, P. J. Jadhav, A. Mohanty, V. Bulovic, M. A. Baldo, P. R. Brown, N. Thompson, B. Wunsch, M. G. Bawendi, E. Hontz, T. Van Voorhis and S. R. Yost, Triplet Exciton Dissociation in Singlet Exciton Fission Photovoltaics, *Adv. Mater.*, 2012, **24**, 6169–6174.

19 I. Paci, J. C. Johnson, X. Chen, G. Rana, D. Popović, D. E. David, A. J. Nozik, M. A. Ratner and J. Michl, Singlet fission for dye-sensitized solar cells: Can a suitable sensitizer be found?, *J. Am. Chem. Soc.*, 2006, **128**, 16546–16553.

20 J. C. Johnson and J. Michl, 1,3-Diphenylisobenzofuran: a Model Chromophore for Singlet Fission, *Top. Curr. Chem.*, 2017, **375**, 1–29.

21 J. L. Ryerson, J. N. Schrauben, A. J. Ferguson, S. C. Sahoo, P. Naumov, Z. Havlas, J. Michl, A. J. Nozik and J. C. Johnson, Two Thin Film Polymorphs of the Singlet Fission Compound 1,3-Diphenylisobenzofuran, *J. Phys. Chem. C*, 2014, **118**, 12121–12132.

22 T. Minami and M. Nakano, Diradical character view of singlet fission, *J. Phys. Chem. Lett.*, 2012, **3**, 145–150.

23 O. Varnavski, N. Abeyasinghe, J. Aragó, J. J. Serrano-Pérez, E. Ortí, J. T. López Navarrete, K. Takimiya, D. Casanova, J. Casado and T. Goodson, High yield ultrafast intramolecular singlet exciton fission in a quinoidal bithiophene, *J. Phys. Chem. Lett.*, 2015, **6**, 1375–1384.

24 S. Lukman, J. M. Richter, L. Yang, P. Hu, J. Wu, N. C. Greenham and A. J. Musser, Efficient Singlet Fission and Triplet-Pair Emission in a Family of Zethrene Diradicaloids, *J. Am. Chem. Soc.*, 2017, **139**, 18376–18385.

25 O. El Bakouri, J. R. Smith and H. Ottosson, Strategies for Design of Potential Singlet Fission Chromophores Utilizing a Combination of Ground-State and Excited-State Aromaticity Rules, *J. Am. Chem. Soc.*, 2020, **142**, 5602–5617.

26 Y. Wu, Y. Wang, J. Chen, G. Zhang, J. Yao, D. Zhang and H. Fu, Intramolecular Singlet Fission in an Antiaromatic Polycyclic Hydrocarbon, *Angew. Chem., Int. Ed.*, 2017, **56**, 9400–9404.

27 K. J. Fallon, P. Budden, E. Salvadori, A. M. Ganose, C. N. Savory, L. Eyre, S. Dowland, Q. Ai, S. Goodlett, C. Risko, D. O. Scanlon, C. W. M. Kay, A. Rao, R. H. Friend, A. J. Musser and H. Bronstein, Exploiting Excited-State Aromaticity to Design Highly Stable Singlet Fission Materials, *J. Am. Chem. Soc.*, 2019, **141**, 13867–13876.

28 S. Yadav, O. El Bakouri, K. Jorner, H. Tong, C. Dahlstrand, M. Solà and H. Ottosson, Exploiting the Aromatic Chameleon Character of Fulvenes for Computational Design of Baird-Aromatic Triplet Ground State Compounds, *Chem. – Asian J.*, 2019, **14**, 1870–1878.

29 L. Wang, L. Lin, J. Yang, Y. Wu, H. Wang, J. Zhu, J. Yao and H. Fu, Singlet Fission in a Pyrrole-Fused Cross-Conjugated Skeleton with Adaptive Aromaticity, *J. Am. Chem. Soc.*, 2020, **142**, 10235–10239.

30 W. Zeng, O. El Bakouri, D. W. Szczepanik, H. Bronstein and H. Ottosson, Excited state character of Cibalackrot-type compounds interpreted in terms of Hückel-aromaticity: a rationale for singlet fission chromophore design, *Chem. Sci.*, 2021, **12**, 6159–6171.

31 M. Purdy, J. R. Walton, K. J. Fallon, D. T. W. Toolan, P. Budden, W. Zeng, M. K. Corpinot, D. K. Bučar, L. van Turnhout, R. Friend, A. Rao and H. Bronstein, Aza-Cibalackrot: Turning on Singlet Fission Through Crystal Engineering, *J. Am. Chem. Soc.*, 2023, **145**, 10712–10720.

32 L. Schaufelberger, J. T. Blaskovits, R. Laplaza, K. Jorner and C. Corminboeuf, Inverse Design of Singlet-Fission Materials with Uncertainty-Controlled Genetic Optimization, *Angew. Chem., Int. Ed.*, 2025, **64**, e202415056.

33 X. Wang, S. Gao, Y. Luo, X. Liu, R. Tom, K. Zhao, V. Chang and N. Marom, Computational Discovery of Intermolecular Singlet Fission Materials Using Many-Body Perturbation Theory, *J. Phys. Chem. C*, 2024, **128**, 7841–7864.

34 C. H. Li and D. P. Tabor, Generative organic electronic molecular design informed by quantum chemistry, *Chem. Sci.*, 2023, **14**, 11045–11055.

35 L. Borislavov, M. Nedyalkova, A. Tadjer, O. Aydemir and J. Romanova, Machine Learning-Based Screening for Potential Singlet Fission Chromophores: The Challenge of Imbalanced Data Sets, *J. Phys. Chem. Lett.*, 2023, **14**, 10103–10112.

36 X. Liu, X. Wang, S. Gao, V. Chang, R. Tom, M. Yu, L. M. Ghiringhelli and N. Marom, Finding predictive models for singlet fission by machine learning, *npj Comput. Mater.*, 2022, **8**, 70.

37 F. Weber and H. Mori, Machine-learning assisted design principle search for singlet fission: an example study of cibalackrot, *npj Comput. Mater.*, 2022, **8**, 176.

38 S. Ye, J. Liang and X. Zhu, Catalyst deep neural networks (Cat-DNNs) in singlet fission property prediction, *Phys. Chem. Chem. Phys.*, 2021, **23**, 20835–20840.

39 D. Dalmau and J. V. Alegre-Requena, ROBERT: Bridging the Gap Between Machine Learning and Chemistry, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2024, **14**, e1733.

40 J. V. Alegre-Requena, S. S. Sowndarya V, R. Pérez-Soto, T. M. Alturaifi and R. S. Paton, AQME: Automated quantum mechanical environments for researchers and educators, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2023, **13**, e1663.

41 H. Takano, T. Ito, K. S. Kanyiva and T. Shibata, Recent Advances of Biphenylene: Synthesis, Reactions and Uses, *Eur. J. Org. Chem.*, 2019, 2871–2883.

42 C. Hong, J. Baltazar and J. D. Tovar, Manifestations of Antiaromaticity in Organic Materials: Case Studies of Cyclobutadiene, Borole, and Pentalene, *Eur. J. Org. Chem.*, 2022, e202101343.

43 R. Ayub, O. El Bakouri, K. Jorner, M. Solà and H. Ottosson, Can Baird's and Clar's Rules Combined Explain Triplet State Energies of Polycyclic Conjugated Hydrocarbons with Fused 4nπ- and (4n + 2)π-Rings?, *J. Org. Chem.*, 2017, **82**, 6327–6340.

44 S. F. Vyboishchikov. Sbuilder, GitHub repository, 2024, https://github.com/vyboishchikov/Sbuilder (accessed 17 July 2025).

45 A. Hagberg, D. A. Schult and P. J. Swart, Exploring Network Structure, Dynamics, and Function Using NetworkX, *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, ed Varoquaux, G., Vaught, T., Millman, J., Pasadena, CA, USA, 2008, 11–15.

46 T. A. Halgreen, MMFF VI. MMFF94s Option for Energy Minimization Studies, *J. Comput. Chem.*, 1999, **20**, 720–729.

47 G. Landrum. *RDKit: Open-source cheminformatics software*. https://www.rdkit.org (accessed 17 July 2025).

48 S. R. Heller, A. McNaught, I. Pletnev, S. Stein and D. Tchekhovskoi, InChI, the IUPAC International Chemical Identifier, *J. Cheminf.*, 2015, **7**, 23.

49 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg,

D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.02*, Gaussian, Inc., Wallingford CT, 2016.

50 Y. Zhao and D. G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

51 F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

52 E. Säbb, P. J. Mayer and H. Ottosson, Can heteroatom and heteroarene annelations make pentalenes suitable as singlet fission chromophores?, *Phys. Chem. Chem. Phys.*, 2026, DOI: 10.1039/D5CP04492H.

53 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, Open Babel: An Open chemical toolbox, *J. Cheminform.*, 2011, **3**, 33.

54 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, Extended tight-binding quantum chemistry methods, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1493.

55 F. Pedregosa, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

56 L. Breiman, Random Forests, *Mach. Learn.*, 2001, **45**, 5–32.

57 S. M. Lundberg, P. G. Allen and S.-I. Lee, A Unified Approach to Interpreting Model Predictions, *Adv. Neural Inf. Process Syst.*, 2017, **30**, 4765–4774.

58 S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.-I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.*, 2020, **2**, 56–67.

59 L. S. Shapley, A Value for n-person Games, in *Contributions to the Theory of Games*, ed H. W. Kuhn and A. W. Tucker, Princeton University Press, Princeton, 1953, vol. **28**, pp. 307–317.

60 S. A. Wildman and G. M. Crippen, Prediction of Physicochemical Parameters by Atomic Contributions, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 868–873.