Electronic Supplementary Material (ESI) for Energy & Environmental Science.
This journal is © The Royal Society of Chemistry 2019

SUPPORTING INFORMATION FOR:

Singlet Fission Molecules among Known Compounds:

Finding a Few Needles in a Haystack

Daniele Padula,*,¶ Ömer H. Omar, Tahereh Nematiaram, Alessandro Troisi*

Department of Chemistry, University of Liverpool, Liverpool L69 7ZD, U.K.

¶Current Address: Laboratory for Functional Polymers, Swiss Federal Laboratories for Materials Science and Technology, Empa, Überlandstrasse 129, 8600 Dübendorf, Switzerland.

E-mail: DP dpadula@liverpool.ac.uk, AT atroisi@liverpool.ac.uk

Content

1. Computational Details

2. Generating a dataset of molecular semiconductors from the full CSD

3. Comparison between the results obtained with different density functional

4. Comparison between adiabatic and vertical excitation energies

5. Description of additional files submitted

6. References

1. Computational Details

Quantum Chemical calculations were carried out with the Gaussian16 software.[1] We define various protocols depending on the task to carry out. The protocols are described in the following.

***Molecular Geometry Generation*.** For the validation of the protocol on molecules with available experimental data from the literature, we gathered experimental excited states energies from a few paper and from Table 3a of ref. [2], using data collected in non-polar solvent environment so that gas phase calculations would be appropriate. Data and references are provided in electronic format as ESI in the `expt_calib.csv` file. We obtained the SMILES strings as structural representations from PubChem,[3] or drew the molecules with the MarvinSketch software[4] to obtain SMILES strings.[5–7] Following, the SMILES strings were processed with the RDKit package[8] to obtain 3-D coordinates, to run a preliminary geometry optimisation at MM level (MMFF94s Force Field), refined through DFT at BLYP35/def2-TZVP level (tight convergence criteria), according to a literature protocol.[9]

For the virtual screening of molecules from the CSD, we used experimental geometries as provided in the database, using the CSD Python API to add missing hydrogen atoms and to normalise them (*i.e.* place them at a typical X-H distance). When optimising molecules taken from the CSD, we start from the experimental geometry provided in the CSD.

***Singlet Fission Screening*.** The energies of the electronic states involved are evaluated following a literature protocol applied on the previously generated geometries. The singlet fission screening procedure described in [9] reported results obtained with Coupled-Cluster (CC2) and TDDFT calculations in comparison to previous CASPT2 calculations. As a result of their comparisons, the authors proposed the Lh12ct-SsifPW92/def2-QZVPD density functional/basis set combination. The reason for this functional was the "lower empiricism" with respect to other highly parametrised functionals (*e.g.* M06-2X) that performed similarly. Due to availability and reproducibility reasons, we opted for the broadly accessible M06-2X functional, implemented in Gaussian16.[1] The authors obtained Mean Absolute Errors (MAEs) between the M06-2X and CC2 predictions of ~0.05-0.3 eV, "with expected CC2 errors compared to experiments of around 0.2-0.3 eV"[10]. Concerning the basis set, the authors investigated the effect of the size, allowing for smaller basis sets to be used. Due to the computational effort required for the screening, we checked that we could lower the basis

set down to def2-SVP without significant loss of accuracy. Additionally, for 262 data points in our group A we have performed large basis set calculation that further validated this approach for the data points with the greatest relevance. The initial tests are not shown because more stringent validation was performed as part of this work.

For our preliminary screening, we started from experimental X-ray geometries, and our procedure consisted of:

- Evaluation of E_{S1} at TDDFT level (M062-X/def2-SVP)
- Evaluation of E_{T1} with a ΔSCF procedure, which involves an additional single point calculation (at M062-X/def2-SVP) for the first triplet state (the ground state single point energy is obtained with the previous calculation).

The computational effort required for this preliminary screening can be quantified in 450,000 CPU hours.

For the accurate calculations on the most promising candidates, we followed the protocol reported in [9] and it consists of:

- Geometry optimisation at DFT level (BLYP35/def2-TZVP)
- Evaluation of E_{S1}, E_{S2}, E_{T2} at TDDFT level (M062-X/def2-TZVP)
- Evaluation of E_{T1} with a ΔSCF procedure, which involves an additional single point calculation (at M062-X/def2-TZVP) for the first triplet state (the ground state single point energy is obtained with the previous calculation).
- Calibration of E_{S1}, E_{T1} to experimental data according to the data in Fig. 1 in the main text.

The computational effort required for this accurate screening can be quantified in 50,000 CPU hours.

2. Generating a dataset of molecular semiconductors from the Cambridge Structural Database (CSD)

Removal of non-organic and polymeric molecules can be carried out thanks to the application programming interface and search criteria provided within the CSD,[11] and allowed us to reduce the molecules to be screened from ~1M to ~250K. We reduced the structures to consider by (i) including only certain elements in the molecules (H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, I), (ii) removing entries with no experimental X-ray structure, (iii) removing entries with more than one molecule type in the unit cell, (iv) limiting the maximum number of atoms in the molecule to 100, (v) removing duplicate entries.

Among these ~250K molecules, we wish to consider only semiconductors, e.g. molecules with a sufficiently low (computed) HOMO-LUMO gap E_{gap}. Due to the high number of molecules to screen, it is impossible to rely on demanding calculations, so we obtained calibration curves to estimate high-level computed gaps (B3LYP/6-31G*, E_{gap}^{H}) from intermediate (B3LYP/3-21G*, E_{gap}^{I}) and low level ones (PM7, E_{gap}^{L}) on a subset of 550 molecules. To span a large interval of orbital energy values, we divided molecules in 10 subsets (molecules with 0-10 atoms, 10-20 atoms, etc) and we picked 55 molecules from each subset. We indicate with $E_{gap}^{L\rightarrow H}$ the expected gap at high-level computation that can be estimated from the low level calculation and the calibration curve in Figure S1 (and an analogous meaning was given to $E_{gap}^{I\rightarrow H}$ using the intermediate level calculations instead). We then carried out calculations at low level on all ~250K molecules to estimate their high level gap $E_{gap}^{L\rightarrow H}$ (see Fig. S5). Considering the distribution of errors from the corresponding calibration curve (σ = 0.5 eV, see Fig. S2), at this stage of the screening we retained all molecules (~175K) showing $2\ \mathrm{eV} \leq E_{gap}^{L\rightarrow H} \leq 5.5\ \mathrm{eV}$ (see Fig. S5). For these ~175K molecules, we repeated electronic calculations at intermediate level to estimate their high level gap $E_{gap}^{I\rightarrow H}$, thanks to the calibration in Fig. S3. 3-21G* calculation show an excellent correlation with 6-31G* results (see Fig. S3), allowing for a reliable estimate. Considering the distribution of errors from the corresponding calibration curve (σ = 0.1 eV, see Fig. S4), at this stage of the screening we retained all molecules (~40K) with $2\ \mathrm{eV} \leq E_{gap}^{I\rightarrow H} \leq 4\ \mathrm{eV}$ (see Fig. S6), that will constitute our set of organic molecular semiconductors for this work. Fig. S5-S6 show that there is substantial computational cost to extend the data set to higher values of the estimated HOMO-LUMO gap.

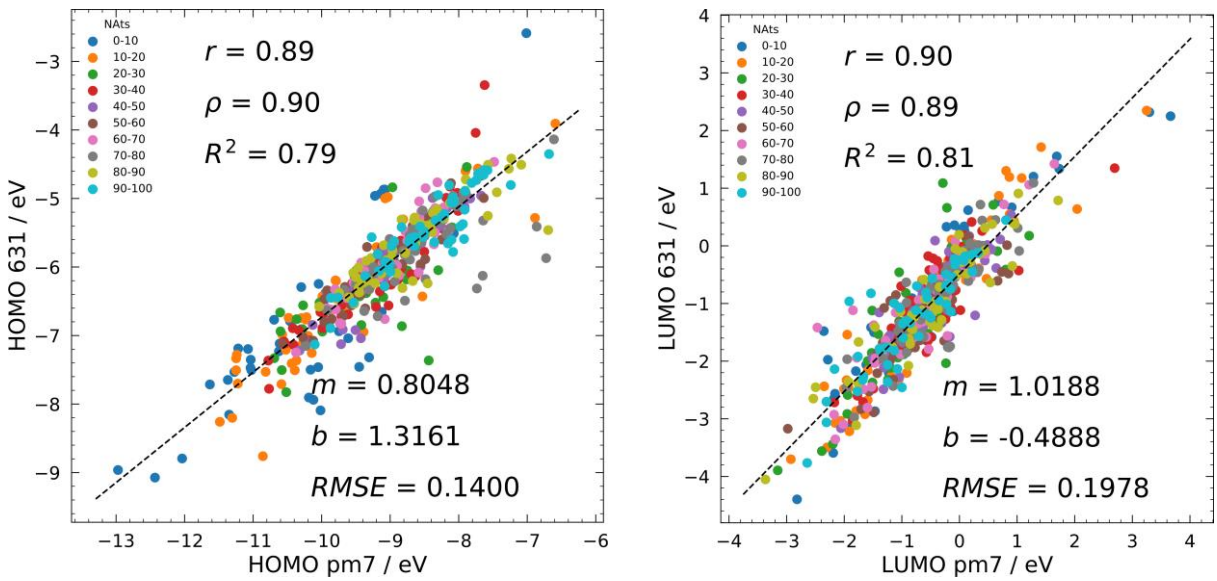

Fig. S1: calibration of $E_{HOMO}^{L\to H}$ (left) and $E_{LUMO}^{L\to H}$ (right), with Pearson (*r*) and Spearman (*ρ*) Correlation coefficients, R^2 metric, root mean squared error (RMSE), and the *m* and *b* values for the linear fit $y = mx + b$.

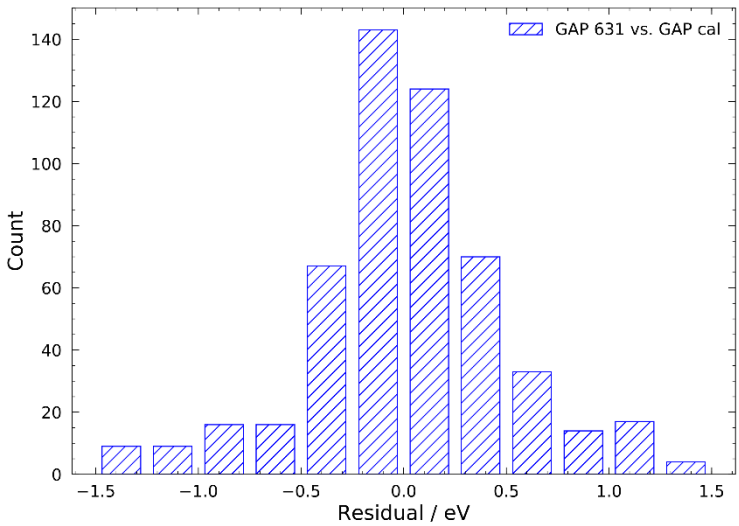

Fig. S2: distribution of residuals $E_{gap}^{H} - E_{gap}^{L\to H}$, with a standard deviation σ = 0.5 eV.

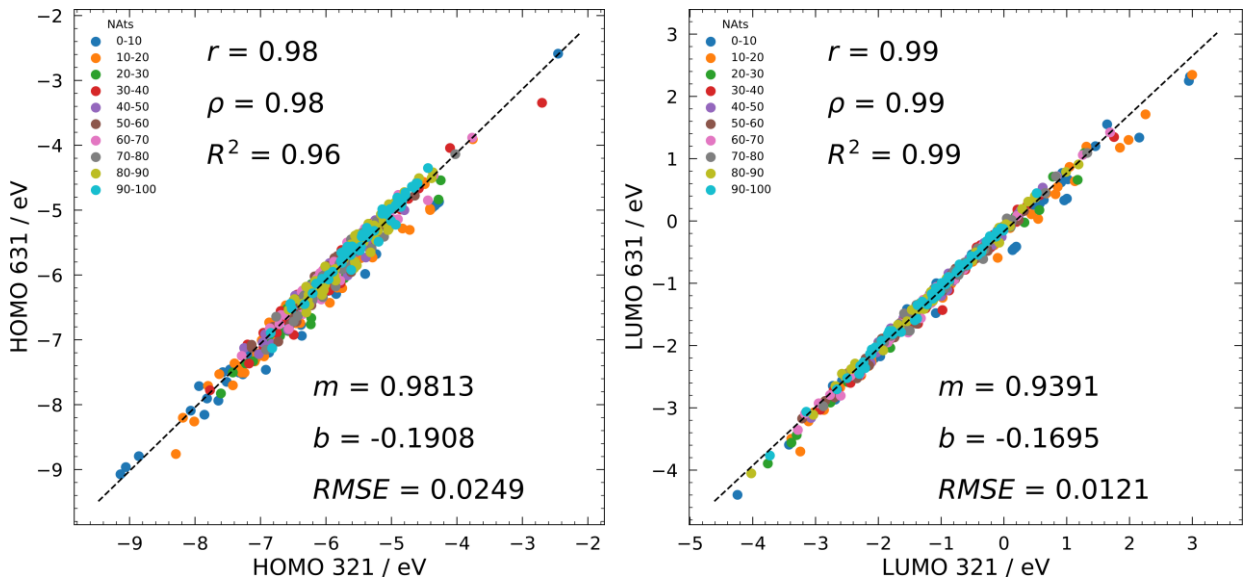

Fig. S3: calibration of $E_{HOMO}^{I\to H}$ (left) and $E_{LUMO}^{I\to H}$ (right) , with Pearson (*r*) and Spearman (*ρ*) Correlation coefficients, R^2 metric, root mean squared error (RMSE), and the *m* and *b* values for the linear fit $y = mx + b$.

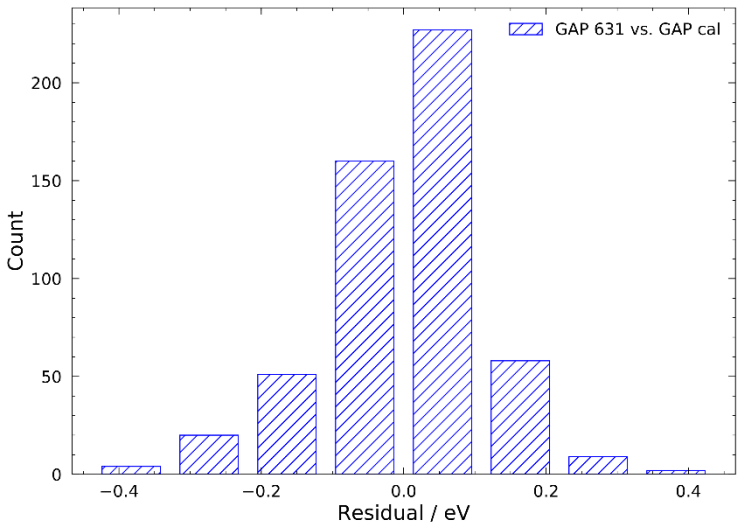

Fig. S4: distribution of residuals $E_{gap}^{H} - E_{gap}^{I\to H}$, with a standard deviation σ = 0.1 eV

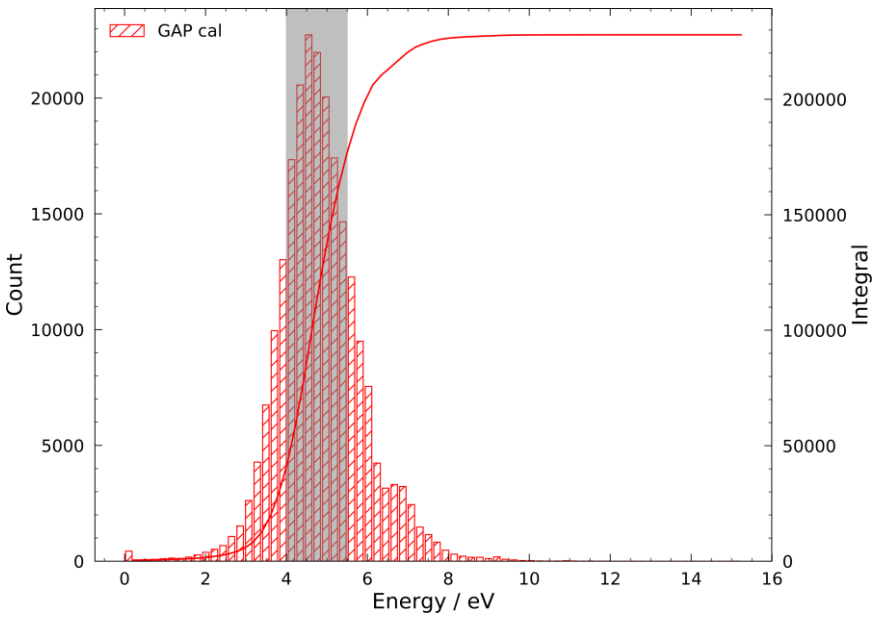

Fig. S5: distribution of $E_{gap}^{L\rightarrow H}$, with cumulative integral (red line) and shaded area between 4.0-5.5 eV.

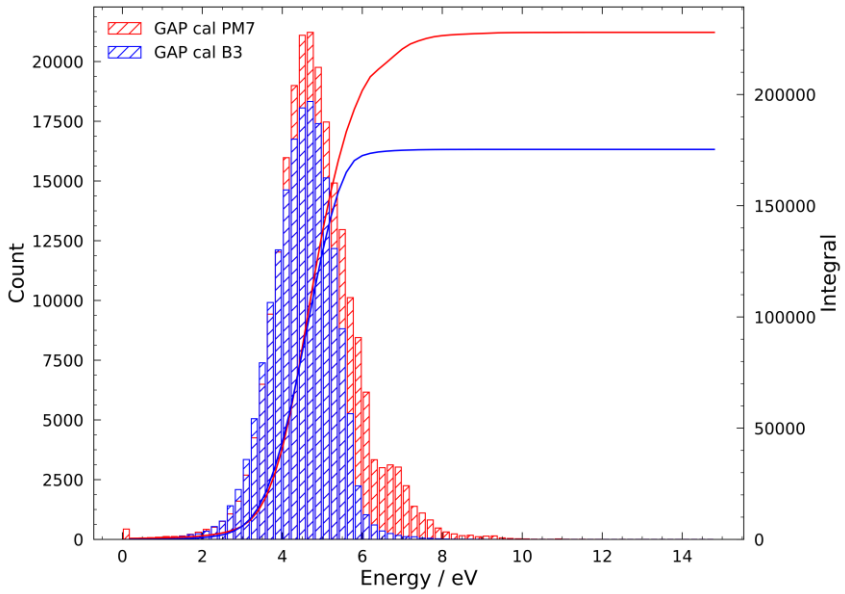

Fig. S6: distribution of $E_{gap}^{L\rightarrow H}$ (red) and $E_{gap}^{I\rightarrow H}$ (blue).

3. Comparison between the results obtained with different density functionals

For a subset of ~70 molecules, we computed excited state energies with the protocol described using two more functionals, starting from experimental X-ray geometries.

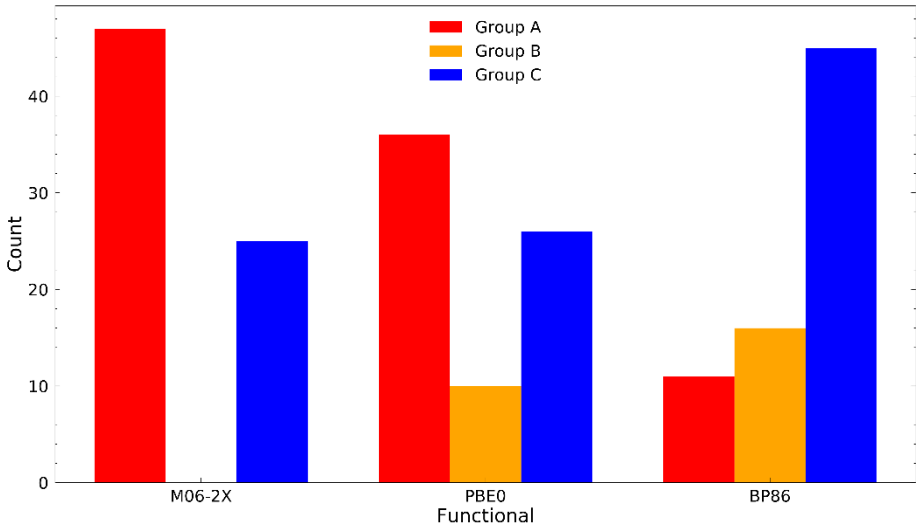

Fig. S7: comparison between M06-2X, PBE0, and BP86 results.

The results show little variability using a global hybrid functional (PBE0), while they vary a lot when adopting a GGA functional: (i) when changing from M06-2X to PBE0, 10 molecules pass from group A to group B, and 1 molecule from group A to group C (ii) when passing from M06-2X to BP86, 16 molecules pass from group A to group B, and 20 molecules from group A to group C.

These results are in line with what already shown in ref. [9], where the poor performance of many density functionals was assigned to an insufficient description of the T1 state.

For a subset of ~20 singlet fission candidates (one for each of the identified families), we repeated geometry optimisations and excited state energies calculations with the protocol described using two more functionals.

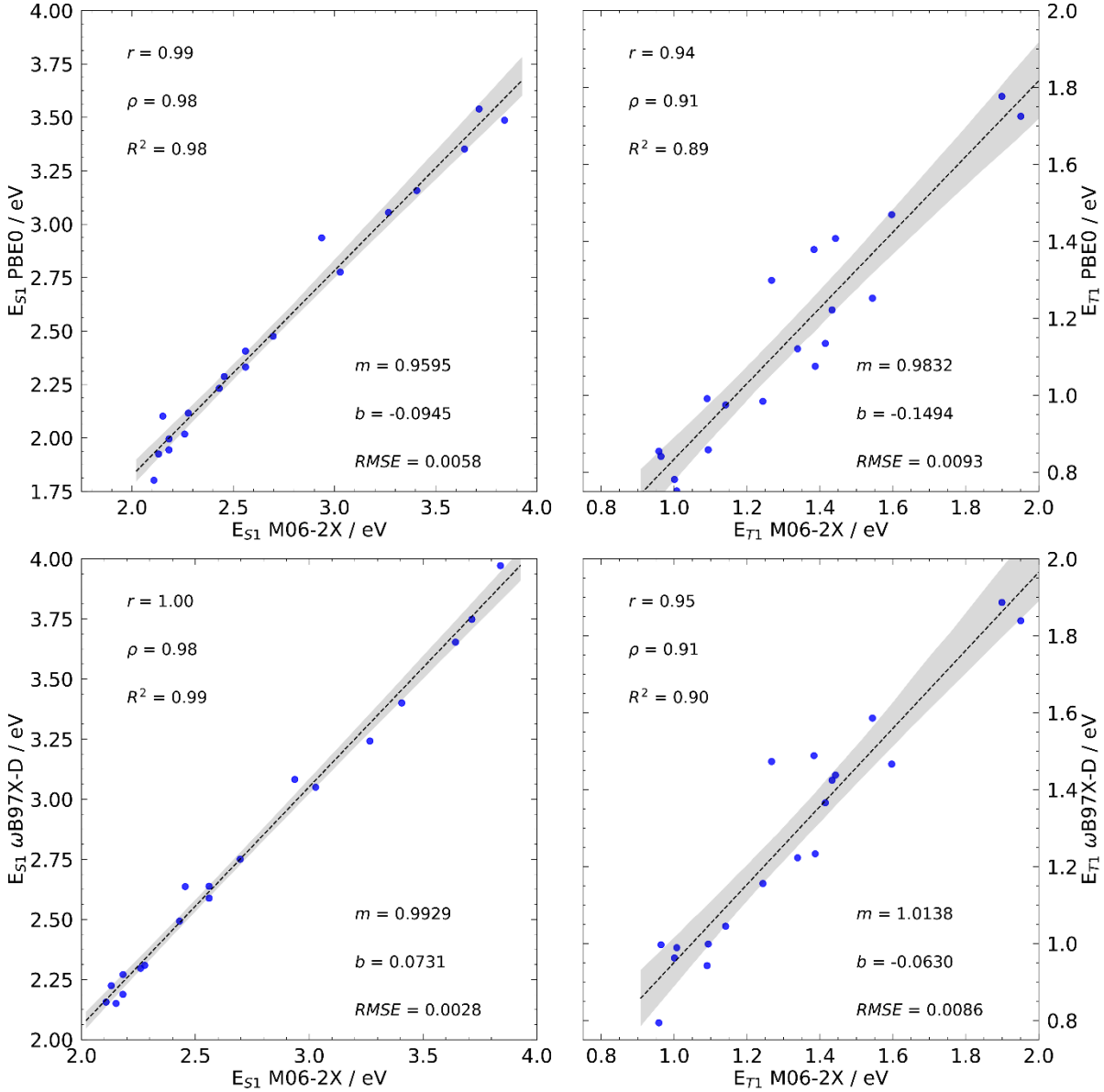

Fig. S8: comparison between M06-2X, PBE0, and ωB97X-D results for ~20 materials from the database.

The results show a very close relationship between M06-2X and ωB97X-D, while comparisons are slightly worse for PBE0. In general, the agreement between singlet energies is excellent, while triplet energies are more sensitive to the functional, in agreement with what observed in ref. [9].

4. Comparison between adiabatic and vertical excitation energies

The main energy condition for singlet fission is more easily verified if the energy in the optimized geometry of the respective state is considered (e.g. adiabatic energy rather than vertical energy differences). This is because the Stokes shift for the triplet state is approximately similar to that of the singlet state but two triplets are generated in the singlet fission process. The table illustrates the case of few molecules that miss the singlet fission energy criteria when vertical energy is considered and meet the criteria when adiabatic energy differences are considered.

Table S1: Reversal of the evaluated Singlet Fission attitude for a subset of molecules close to the boundary in passing from Vertical (V) to Adiabatic (A) energy differences. Excitation energies were not calibrated to experimental data.

CSD ID	$E_{S1}^{Opt,V}$ / eV	$E_{T1}^{Opt,V}$ / eV	$E_{S1}^{Opt,A}$ / eV	$E_{T1}^{Opt,A}$ / eV
HANTAT	2.14	1.08	1.93	0.78
LENHER	2.43	1.23	2.17	0.78
YISFIP	2.64	1.34	2.44	0.92
AJEPOW	2.26	1.14	1.99	0.76
URANIM	2.47	1.26	2.28	0.87
GEGRUI	2.21	1.11	2.06	0.93
HETKUP	2.64	1.33	2.52	1.19

5. Description of additional files submitted

The files and explanations described below are provided in electronic format in a .tar.gz archive available from the publisher, or they can be downloaded from a public BitBucket repository available at https://bitbucket.org/dpadula85/singlet-fission-screening. They will enable redrawing all figures in the manuscript, repeating the calculations and performing further analyses.

Experimental Calibration:

The molecules used for calibration to experimental data are provided in the file `expt_calib.csv`, containing:

- DOI
- SMILES representation
- experimental S1 energy (eV)
- experimental T1 energy (eV)
- computed S1 energy (eV, BLYP35/def2-TZVP geometry, M06-2X/def2-TZVP)
- computed T1 energy (eV, BLYP35/def2-TZVP geometry, M06-2X/def2-TZVP)

Groups A1-A3:

Molecules belonging to groups A1-A3 are provided in the file `group_A_opt.csv` with the results of our calculations, and some data from the CSD, containing:

- CSD identifier
- DOI
- SMILES representation
- computed S1 energy (eV, X-ray geometry, M06-2X/def2-SVP)
- computed S1 oscillator strength (X-ray geometry, M06-2X/def2-SVP)
- computed T1 energy (eV, X-ray geometry, M06-2X/def2-SVP)
- computed S1 energy (eV, BLYP35/def2-TZVP geometry, M06-2X/def2-TZVP)
- computed S1 oscillator strength (BLYP35/def2-TZVP geometry, M06-2X/def2-TZVP)
- computed T1 energy (eV, BLYP35/def2-TZVP geometry, M06-2X/def2-TZVP)
- computed T2 energy (eV, BLYP35/def2-TZVP geometry, M06-2X/def2-TZVP)

The 262 optimised molecules are gathered in the file `opt_structs.xyz`. Geometries have been optimised at BLYP35/def2-TZVP level with the Gaussian16 software,[1] starting from experimental X-ray geometries provided within the CSD. To split the geometries in single files, install OpenBabel[12] (https://openbabel.org) and run the shell command

```
obabel opt_structs.xyz -OCCDC_.xyz --separate -m
```

This will create 262 files named `CCDC_N.xyz` containing the geometries. The CSD identifier is provided as a comment in each .xyz entry.

The 262 molecules have been classified in families, and files with the SMILES and CSD identifiers for easy visualisation are provided in the families folder.

Group B:

Molecules belonging to group B are provided in the file `group_B.csv` with the results of our calculations, and some data from the CSD, containing:

- CSD identifier
- DOI
- SMILES representation
- computed S1 energy (eV, X-ray geometry, M06-2X/def2-SVP)
- computed S1 oscillator strength (X-ray geometry, M06-2X/def2-SVP)
- computed T1 energy (eV, X-ray geometry, M06-2X/def2-SVP)

Molecules are listed according to increasing distance from the decision boundary, so ideally they should be studied in the order provided, checking the oscillator strength of S1.

6. References

(1) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. Gaussian16 Revision B.01. 2016.

(2) Montalti, M.; Credi, A.; Prodi, L.; Gandolfi, M. T. *Handbook of Photochemistry*, 3rd Editio.; CRC Press: Boca Raton, 2006.

(3) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res.* **2019**, *47* (D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033.

(4) ChemAxon. MarvinSketch. 2019.

(5) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36. https://doi.org/10.1021/ci00057a005.

(6) Weininger, D. Smiles. 3. Depict. Graphical Depiction of Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1990**, *30* (3), 237–243. https://doi.org/10.1021/ci00067a005.

(7) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (2), 97–101. https://doi.org/10.1021/ci00062a008.

(8) Landrum, G. RDKit: Open-Source Cheminformatics.

(9) Grotjahn, R.; Maier, T. M.; Michl, J.; Kaupp, M. Development of a TDDFT-Based Protocol with Local Hybrid Functionals for the Screening of Potential Singlet Fission Chromophores. *J. Chem. Theory Comput.* **2017**, *13* (10), 4984–4996. https://doi.org/10.1021/acs.jctc.7b00699.

(10) Christiansen, O.; Koch, H.; Jørgensen, P.; Helgaker, T. Integral Direct Calculation of CC2 Excitation Energies: Singlet Excited States of Benzene. *Chem. Phys. Lett.* **1996**, *263* (3–4), 530–539. https://doi.org/10.1016/S0009-2614(96)01245-6.

(11) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72* (2), 171–179. https://doi.org/10.1107/S2052520616003954.
(12) Boyle, N. M. O.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33. https://doi.org/10.1186/1758-2946-3-33.