

Systematic Computational Design and Optimization of Light Absorbing Dyes

Jelena Belić,* Bas van Beek, Jan Paul Menzel, Francesco Buda, and Lucas Visscher*

Cite This: *J. Phys. Chem. A* 2020, 124, 6380–6388

Read Online

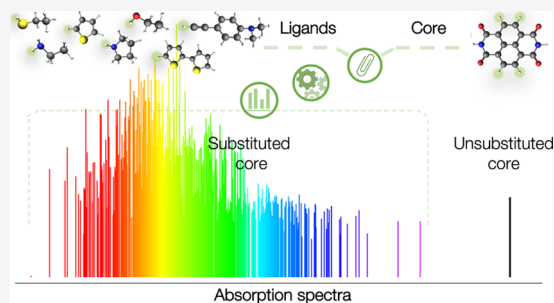
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: We present a workflow to aid the discovery of new dyes for the role of a photosensitive unit in the dye-sensitized photo-electrochemical cells (DS-PECs). New structures are generated in a fully automated way using the Compound Attachment Tool (CAT) introduced in this work. These structures are characterized with efficient approximate density functional theory (DFT) methods, and molecules with favorable optical properties are suggested for possible further use in DS-PECs. As around 2500 structures are generated in this work, and as we aim for still larger volumes of compounds to screen in subsequent applications, we have assessed the reliability of low-cost screening methods and show that simplified time-dependent density functional theory (sTDDFT) provides a satisfying accuracy/cost ratio. From the dyes considered, we propose a set that can be suitable for panchromatic sensitization of the photoelectrode in DS-PECs to further increase DS-PEC efficiency.



INTRODUCTION

Our development into a technologically advanced society has been accompanied by a rapid depletion of fossil fuels that makes the search for a sustainable energy solution of utmost importance. In the long term, one would like to harvest sunlight and use this energy to produce chemicals that can be used as fuels or feedstock. One way to achieve this is to split water inside a photo-electrochemical cell and obtain energy-rich molecular hydrogen.^{1,2} Further processing could then lead to synthesis of methanol, ethanol, heavier hydrocarbons, and synthetic fuels.

Dye-sensitized photo-electrochemical cells (DS-PECs) are promising candidates for this efficient transformation of solar energy. However, the oxidative half-reaction that happens on the anode of the DS-PECs remains a challenging part yet to be fully optimized.^{3,4} The interface between the electrode and photosensitive component is notoriously complex as it involves processes of light absorption, charge separation, and charge transport. It consists of a semiconductor sensitized with a light absorbing molecule (dye) that is coupled to a water oxidation catalyst (WOC), either via a covalent bond or coadsorbed on the semiconductor's surface.^{2,5} After absorbing a photon, the excited dye should promote charge separation and transfer an electron to the semiconductor. Separating these processes into different parts of the device, in analogy to natural photosynthesis,^{6,7} allows individual optimization of the device units—the dye and the semiconductor. In this work, we will focus on optimization of the dyes.

An optimal dye has characteristics such as a wide range of absorption for visible light, affinity toward fast electron injection into the semiconductor, and a low charge

recombination rate. After electron injection, the oxidized dye must go back to the initial state by activating the WOC. In addition, the dye has to be chemically stable and has to show a strong anchoring to the semiconductor surface. The main advantages of employing metal-free organic dyes are their low cost and diversity.⁹ There are many studies on organic dyes in many fields,¹⁰ but we have not been able to find studies that systematically explore the large chemical space of these molecules. We believe that such studies, which are feasible with modern computational techniques, will enable us to quickly identify interesting new targets for synthesis and experimental validation of their properties.

In this paper, we present a tool for automated, systematic generation of molecular structures coupled to a workflow for computational screening of the generated molecules. The step into the area of big data requires automation of all the aspects of scientific research, from preparation to data generation and analysis. As we need to evaluate spectroscopic properties, the use of quantum chemical methods is mandatory, and the computational efficiency of the procedure is of importance. After generating structures by adding one or more substituents from a library to the core molecule, we therefore test the reliability of a number of approximate density functional

Received: May 19, 2020

Revised: July 9, 2020

Published: July 10, 2020



theory (DFT) methods for prediction of the key absorption characteristics. We thereby focus on density functional tight-binding (DFTB),^{11,12} time-dependent DFTB,¹³ and simplified time-dependent density functional theory (sTDDFT),^{14,15} as these methods are all capable of quickly evaluating the position and intensity of absorption peaks in the visible part of the solar spectrum. As a reference method for validation of their accuracy, we use time-dependent density functional theory (TDDFT) as well as comparison with experimentally known spectra.

We first apply our procedure to the 1,4,5,8-naphthalenediimide (NDI) dye as a significant number of its derivatives have been explored experimentally, providing good possibilities for validation of the theoretical predictions. After testing the procedure for this class of dye, we then apply the workflow to the perylene diimide (PDI) and perylene triimides (PTI)^{16,17} as well, leading in total to almost 2500 derivatives that are evaluated.

Spectroscopic Properties of NDI. NDI is a small member of the family of aromatic diimides with a simple structure and relatively easy synthesis.^{18,19} NDI is well-known for its application in artificial photosynthesis, and numerous scientific papers are proof of its extraordinary chemistry.^{10,20–22} In this work, we focus on NDI's light absorbing properties.

Non-substituted NDI absorbs in the UV region, and functionalizing of the naphthalene core (Figure 1) can move

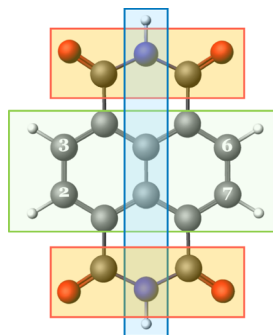


Figure 1. NDI molecule with highlighted possible substitution positions: core substitution (green), shoulder substitution (red), and substitution on the imide (blue).

the absorption to the visible region. Depending on the nature of ligands, electron withdrawing or electron donating, NDI's absorption can be tuned throughout the solar spectrum, resulting in the so-called “rainbow” collection.²¹ Substitution on the imide (Figure 1) side has a negligible effect on the optical properties of NDI,²³ and it is often used to control solubility.

Substitution with different types and numbers of functional groups has been addressed in the literature,^{21,24,25} but most commonly, the effects of mono-substitution, 2,6-di-substitution, and 2,3,6,7-tetra-substitution (Figure 1) are considered. It is shown that electron-donating functional groups like alkoxy,^{26,27} alkylamino,²⁷ alkylthio,²⁶ and thiophenes²⁸ affect optical properties via electronic effects (mainly mesomeric and inductive) on the frontier molecular orbitals, causing a decrease of the HOMO/LUMO gap while increasing the HOMO and LUMO energies.^{29,30} The lowest transition in NDI is susceptible to the substituent effect, while the higher

transition is unaltered and resembles the properties of non-substituted NDI.^{21,26}

As NDIs are planar π systems that tend to aggregate, an additional advantage of introducing the ligands is that the steric hindrance effect can be used to prevent the aggregation, making π – π stacking of the NDIs less favorable.

METHODS

All computational methods used are included in the Amsterdam Modeling Suite (AMS) from Software for Chemistry and Materials (SCM).³¹ Initial geometries for the core and the library of ligands are prepared with the SCM Graphical User Interface (GUI). Generation of new structures is done with the Compound Attachment Tool (CAT).³² CAT is a Python code employing, among others, the Python Library for Automating Molecular Simulation (PLAMS)³³ for generating structures. The spectroscopic characterization is done with a workflow in which a structure optimization is followed by time-dependent (approximate) DFT calculations. The full workflow is depicted in Figure 2 and can be divided into three parts: (i) generating new structures, (ii) calculating the properties, and (iii) processing and analyzing generated data. We will discuss these parts in more detail below.

Structure Generation. The task of CAT is to make multi-substituted molecules by combining the core dye with a library of possible ligands. The user of this program needs to supply the sets of ligands and cores as a list of structure files in a suitable (xyz) format and needs to define the substitution spot(s) by selecting the hydrogen atom(s) to be substituted at the core and at the ligands. The coordinates of the hydrogen atom at the unsubstituted core serve as the initial position of the new substituent. Figure 3 shows how molecule attachment works in CAT. Initially, we treat the ligand as a rigid body that is attached and rotated around the attachment bond axis to determine the position with minimal steric repulsion. After this rotation along the bond axis, the code will run an additional force field optimization in which the ligand is allowed to distort from its free molecule structure to minimize steric repulsion of the ligand with the rest of the molecule.

Multiple substitutions are implemented as series of individual substitutions on the specified substitution spots. As the dye molecules often possess symmetry, it is of importance to remove redundant structures by considering permutation symmetry. As we aim to have a procedure that will work for arbitrary input structures that may not always be generated by applying symmetry constraints in the structure optimization, we leave the specification of symmetry elements to the user. This also easily allows for the use of approximate symmetry, e.g., a larger symmetric dye that is functionalized in a way that breaks the exact symmetry. With this philosophy, the generation of unique derivatives may use techniques for permutational isomers.³⁴ CAT contains a function to eliminate the permutational isomers, taking into account the type of ligand, substitution sites on the core molecule, and symmetry type of the substitution site. For determining the number of unique derivatives, it thereby uses Polya's enumeration theorem.^{35,36} For the cases with up to four substituents that we consider in this work, this is primarily needed for the class of tetra-substituted molecules where the use of Polya's theorem gives a significant reduction of the number of molecules that need to be considered explicitly.

A new single bond (between the core and ligand) increases the number of conformational isomers, in addition to

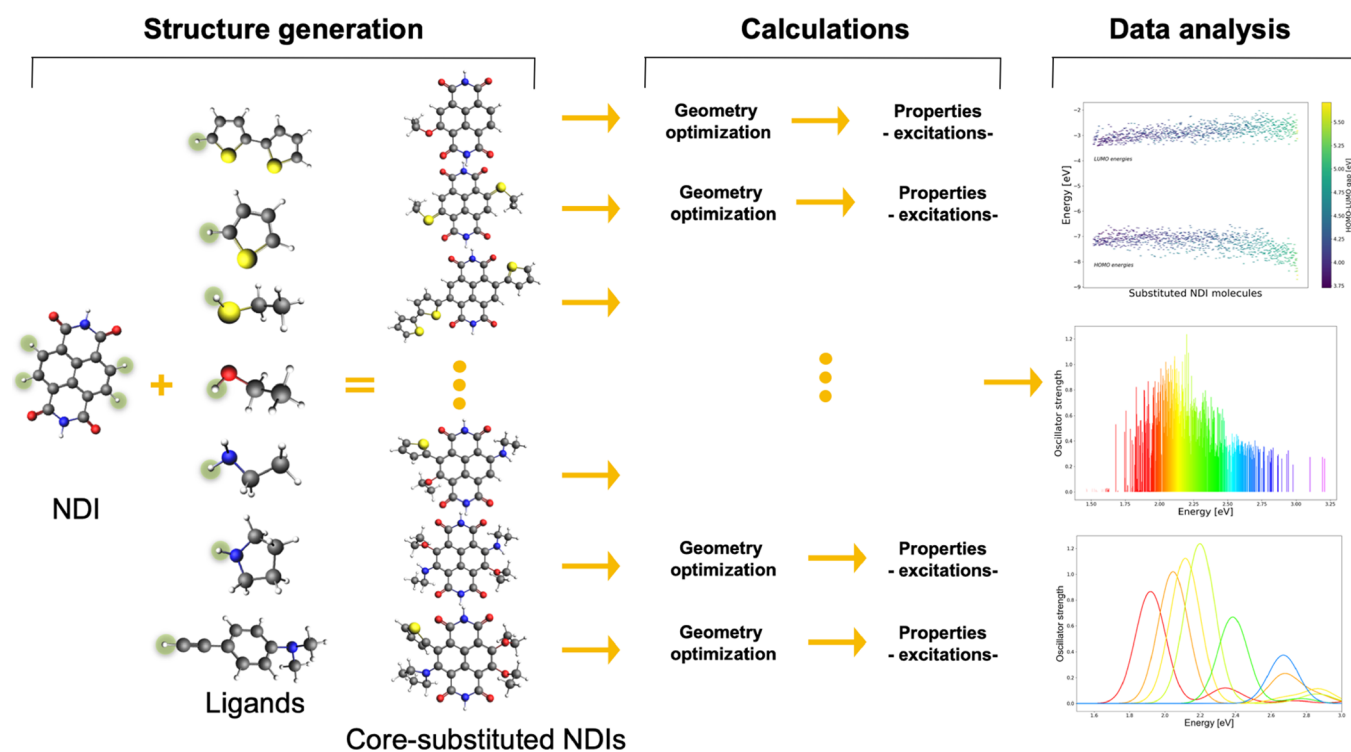


Figure 2. Workflow scheme for the NDI example. The workflow is separated into three sections: (i) generating new structures—attaching the ligands to the substitution spots (highlighted in green) on the NDI in such a manner to avoid steric clashes; (ii) calculating the properties—running the geometry optimization and forwarding the optimized structure to the input for time-dependent calculations; and (iii) processing and analyzing generated data—manipulating data, visualizing data, and applying the selection criteria.

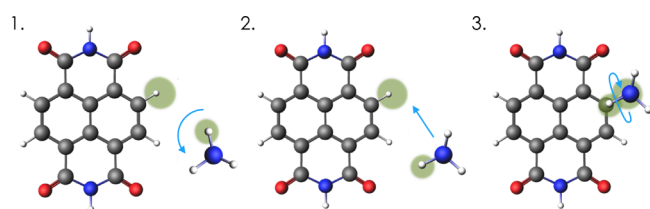


Figure 3. Visualization of CAT's mono-substitution process; (1) rotation of the ligand in such a way that bonds of specified atoms are parallel; (2) translation of the ligand to the position of a specified atom at the core; (3) rotation around the bond axes in order to avoid steric clashes.

conformational isomers of ligands and cores alone. The current version of CAT does not search for the lowest energy conformation. However, for a future version, we plan to include an interface to one of the already existing codes for conformer search.^{37,38}

Prediction of Spectroscopic Properties. The initial molecular structures are optimized with the DFTB3³⁹ method using the 3-ob parameter set.^{40–43} Convergence to a stable minimum is checked by calculating the lowest frequency normal modes.⁴⁴ We allow for some numerical noise and discard only structures with imaginary frequencies above 20 cm^{-1} in this step. We proceed with the remaining structures to the evaluation of molecular properties and compute the 10 lowest-lying singlet excited states with TDDFTB¹³ with the same 3-ob parameter set, as well as sTDDFT¹⁴ employing the range-separated⁴⁵ Coulomb-Attenuated B3LYP (CAM-B3LYP)⁴⁶ functional and the DZP basis set. The parameters for sTDDFT in combination with CAM-B3LYP were set as recommended in the paper by Risthaus, Hansen, and

Grimme.¹⁵ All the calculations are performed for isolated molecules in the gas phase.

For a subset of the final structures, properties are also computed with a reference TDDFT method, for which we also took the CAM-B3LYP⁴⁶ functional and the DZP basis set. CAM-B3LYP was reported to be reliable in predicting the low-lying excited states for the π -conjugated systems in general^{47–49} and was recently favorably evaluated for NDIs in particular.⁵⁰

Data Analysis. The results obtained from the calculations were processed using Python packages for data analysis and visualization, used in conjunction with SCM-GUI for visualization of molecular orbitals and production of simulated spectra.

Selection criteria in this analysis are the dye's optical properties—excitation energies and oscillator strengths. Ideally, the lowest excited state should be above the threshold for water oxidation reaction that is about 950 nm (1.35 eV),⁵¹ while the upper limit could be set at 3.2 eV due to the low intensity of UV light on the Earth's surface. However, as we work with approximate DFT methods and apply additional approximations in our model (most importantly, the neglect of solvent and other environmental effects), we should also account for errors in the computed values when analyzing the results. For this reason, we discard molecules that have the lowest excitation with oscillator strengths below 0.001 as these are considered likely to be “spurious”.^{52,53}

Of the remaining dyes, we consider the ones that have an intense, lowest transition above the energy threshold as the most promising. As not all promising molecules will be easy to synthesize, we also consider the use of a synthetic accessibility

(SA)⁵⁴ score. This is a value that indicates how easy (1) or difficult (10) the synthesis will be.

RESULTS

To generate the NDI derivatives, seven ligands with electron-donating characteristics were chosen from the literature.^{21,26–28,55} The ligands contain sulfur, oxygen, or nitrogen and are numbered as follows: 1 is ethanethiol, 2 is thiophene, and 3 is bithiophene; 4 is ethanol; 5 is ethanamine, 6 is pyrrole, and 7 is 4-ethynyl-*N,N*-dimethylaniline.

The hydrogen atoms on the NDI core (Figure 1) are substituted in the following order, the first position is 2 then 6, 3, and 7. If represented as a list, the first element of the list corresponds to substitution on position 2, the second element corresponds to substitution at position 6, etc.; if we use numbers to represent ligands as in Figure 4, we can represent

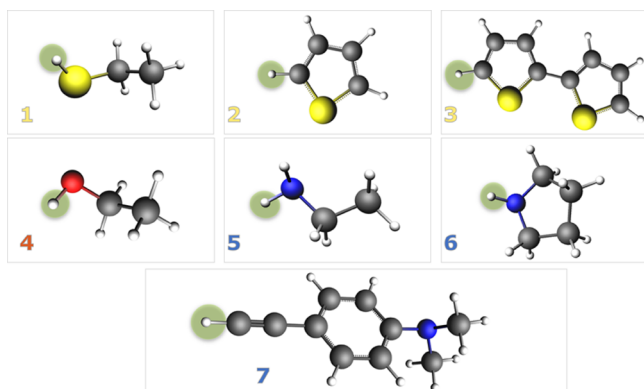


Figure 4. Ligands used to functionalize the NDI core. The atom highlighted in green is the connection point to the NDI core. 1, 2, and 3 are ligands containing sulfur (yellow); 4 contains oxygen (red); and 5, 6, and 7 include nitrogen atoms (blue). Carbon and hydrogen atoms are in gray and white, respectively.

NDI derivatives in a short notation as NDI(1,1) for 2,6-di-alkylthio-substituted NDI or even shorter as NDI-11. This way of substitution is found to be most common in the literature.^{21,26,29,55} Using CAT, a total of 1015 NDI derivatives are generated: 7 mono-substituted, 28 di-substituted, 343 tri-substituted, and 637 tetra-substituted derivatives. The molecule sizes vary from 33 to 106 atoms. Experimental characterization is available for 11 molecules from this set. In order to assess the reliability of the predicted absorption spectra, we compared computed vertical excitation energies, E_{VE} , with the absorption maxima, λ_{max} measured in solution (Figure 5).

The excitation energy values computed with the reference TDDFT(CAM-B3LYP) method show in general good agreement with trends in the experimentally obtained spectra. The method is able to predict the shifts observed in the experiment, e.g., the shift of the absorption band from two alkoxy substituents (NDI-44) to one alkoxy and one alkylamino substituent (NDI-54) and finally to two alkylamino substituents (NDI-55). Also, it predicts the correct electronic transitions; e.g., for the NDI-55, NDI-44, and NDI-54, the lowest energy peak is attributed to the HOMO→LUMO electronic transition, while the second peak is attributed to the HOMO-1→LUMO electronic transition. The second peak remains at the same position for these three molecules as it is a characteristic of the NDI core absorption.^{26,27} An equivalent

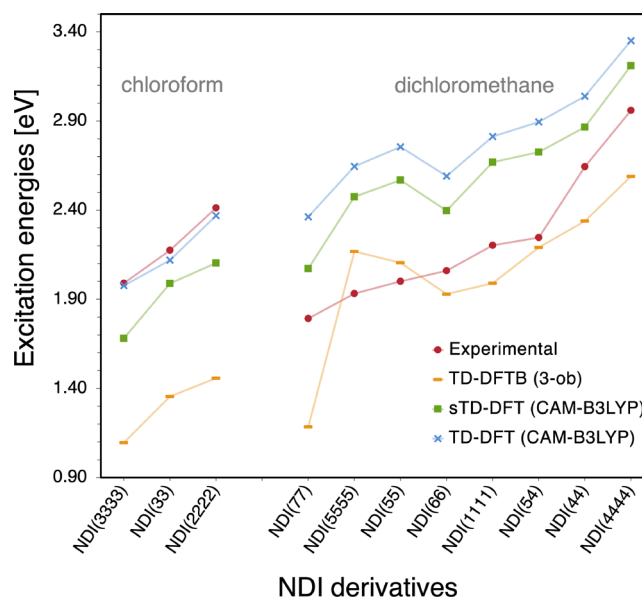


Figure 5. Excitation energies of the isolated NDI derivatives calculated with TDDFTB, sTDDFT, and TDDFT compared to the experimentally obtained absorption maxima. Data points are sorted by the experimental values, increasing from left to right for each solution, chloroform (left) and dichloromethane (right).

behavior occurs for tetra-substituted NDI-5555 compared to di-substituted NDI-55, with a shift toward shorter wavelengths.²⁶ On the other side of the energy scale, the NDI-4444 spectrum arises from three very close intense excitations, with the lowest on HOMO→LUMO transition followed by HOMO-1→LUMO and other higher transitions. The presence of the higher excitations has been suggested by Röger and Würthner,²⁶ as the mirror-image fluorescence band is absent. The molecule giving the lowest excitation energy in this figure (Figure 5) is NDI-77, the NDI core di-substituted with the 4-ethynyl-*N,N*-dimethylaniline group. This large shift to lower energies is due to the extended π -conjugated system that is created by directly connecting the triple C–C bond to the NDI core and due to the electron-rich amino group at the other end of the ligand.⁵⁵

Neglecting solvent effects on the geometry as well as on the spectra is partly responsible for the discrepancy between experimentally measured and calculated values. In the case of NDI-55 and NDI-5555, outliers in Figure 5, the automatically generated geometry differs from experimentally observed—the hydrogen attached to the nitrogen is not forming a hydrogen bond with the oxygen from the imide—and reduces the redshift.^{27,29} For the non-polar chloroform solvent, we find an excellent quantitative prediction of the trend with TDDFT(CAM-B3LYP).

Turning now the attention to the performance of the approximate methods, we consider two possible screening methods, sTDDFT(CAM-B3LYP) and TDDFTB(3-ob), as well as the reference method, TDDFT(CAM-B3LYP). The statistical analysis is shown in Table 1 for the same data points from Figure 5 for values measured in dichloromethane. Looking at Table 1, we see the quantification of the consistent overestimation of excitation energies by TDDFT(CAM-B3LYP) and sTDDFT(CAM-B3LYP) that is already visible in Figure 5.

Table 1. Statistical Analysis^a of the Computed Excitation Energies (eV) for the Isolated Molecule Compared with Experimental Results for the Molecule in Dichloromethane and Average Time (*t*) per Calculation^b

method	MD	MAD	RMSD	R ²	<i>t</i>
TDDFT(CAM-B3LYP)	0.58	0.58	0.59	0.92	27 min
sTDDFT(CAM-B3LYP)	0.39	0.41	0.41	0.88	4.5 min
TDDFTB(3-ob)	-0.17	0.25	0.30	0.60	3.5 s

^aMD stands for the mean deviation, MAD for the mean absolute deviation, RMSD is for the root-mean-square deviation, and R² is the squared correlation. ^bBased on all molecules shown in Figure 5 on two Intel Xeon nodes (48 cores) of the Dutch national super-computer Cartesius.

The values for the mean deviation (MD) and mean absolute deviation (MAD) are higher than for TDDFTB(3-ob), which for some molecules underestimate and for others overestimate the excitation energy (MD = -0.17 and MAD = 0.25). However, the R² value is more important as this quantifies whether shifts due to substitution are predicted correctly. Here, we see that, as expected, the correlation between theory and experiment is highest for the TDDFT(CAM-B3LYP), 0.92. sTDDFT(CAM-B3LYP) scores with 0.88, clearly better than TDDFTB(3-ob) that gives a value of only 0.60. The final column in Table 1 is the average time per calculation, which is indicative of the computational cost of the workflow. Looking at its reliability and efficiency, sTDDFT(CAM-B3LYP) was found to be the optimal method for the purpose of massive screening of dyes. We also note that this method should be easier to combine with inclusion of solvent effects as it derives directly from a DFT ground state calculation. This analysis suggests that the range separation is indeed necessary, and instead of the local functional, TDDFTB(3-ob), its long-range corrected TDDFTB (lc-TDDFTB)⁵⁶ variant would be more appropriate for comparison. Unfortunately, this parameter set is not yet implemented in the SCM software suite.

Apart from solvent effects, the difference between computed values and experimental data can originate from various other factors.⁵⁷ A trivial empirical correction can be applied by shifting the excitation energy in dichloromethane down by the MAD value shown in Table 1. For sTDDFT(CAM-B3LYP), this would yield a modest 0.4 eV correction factor. To summarize the results of this first workflow application, we find that out of 1015 structures, 1013 are characterized as stable (no significant imaginary frequencies). Analysis of excitation energies gives 886 molecules with the lowest (non-spurious) excitations that fit the selection criteria. These excitations are shown in Figure 6 and demonstrate that their energies span the energy range from 1.7 to 3.2 eV. By partitioning this range into smaller intervals and selecting the most intense dye from each interval (Table 2), it is indeed possible to achieve panchromatic sensitization of the photoelectrode.

In Table 2, we notice that all compounds that have the highest absorption intensity contain ligand 7. This ligand gives rise to an extended π conjugation of the NDI core, which reduces the HOMO-LUMO gap of the NDI and thus lowers the excitation energy. The molecule with the highest oscillator strength has 4-ethynyl-*N,N*-dimethylaniline (7) at positions 2 and 6; with a third substituent, ethanamine (5), at position 3; the fourth substituent is ethanol (4) at position 7. Due to its low excitation energy and high oscillator strength, it can be considered as a promising molecule for further study. The

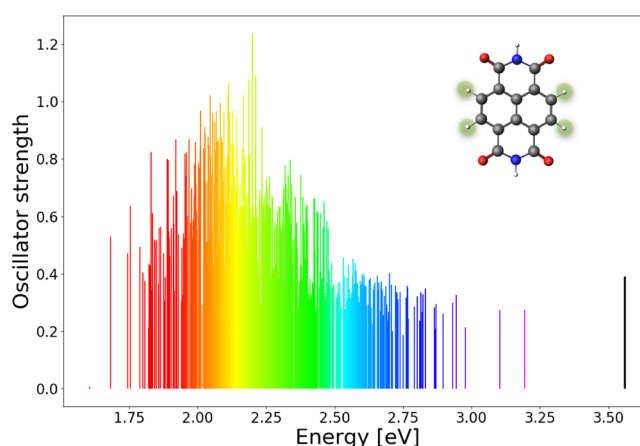


Figure 6. Absorption spectra based on the lowest excitation peaks from each promising compound predicted with sTDDFT. The color of the peaks resembles the color corresponding to the excitation. The absorption peak in black is from the non-substituted NDI ($E_{VE}^{sTDDFT} = 3.56$ eV; $f = 0.39$).

Table 2. Computed Lowest Excitation Energies with the Highest Oscillator Strength in Intervals of 0.2 eV between 1.7 and 3.2 eV^a

energy range [eV]	NDI derivative	E_{VE}^{sTDDFT} [eV]	<i>f</i>	SA score
1.6–1.7	NDI-3333	1.68	0.53	7.91
1.7–1.8	NDI-3337	1.75	0.64	5.84
1.8–1.9	NDI-3377	1.83	0.82	6.87
1.9–2.0	NDI-3677	1.97	0.87	6.71
2.0–2.1	NDI-5677	2.05	1.02	7.95
2.1–2.2	NDI-5477	2.20	1.24	7.78
2.2–2.3	NDI-775	2.21	1.09	7.58
2.4–2.5	NDI-5447	2.43	0.66	7.38
2.5–2.6	NDI-126	2.53	0.46	7.13
2.6–2.8	NDI-5462	2.70	0.40	7.32
2.8–3.0	NDI-5454	2.83	0.35	7.05
3.0–3.2	NDI-444	3.10	0.27	6.80

^aResults are obtained with the sTDDFT method.

derivatives NDI-3333 and NDI-77 have indeed already been recognized in the literature as promising candidates,^{28,54} which confirms the workflow reliability. The last column in Table 2 shows the SA score, which for all molecules falls between 6 and 8. This value is close to the value estimated for 11 experimentally known structures (Figure 5) that are rated from 6.7 to 7.9. We also note that the SA score of non-substituted NDI is 6.44, which makes this value not very selective. This probably also has to do with the limited diversity of our ligands, and we believe that this value may become more useful for larger sets for which further reduction of the number of candidate structures is needed.

To proceed further into uncharted chemical space, we take three less well-known members of the perylene family and combine these with the same set of ligands. We hereby consider perylene diimide (PDI) and two core-extended PDIs, perylene triimides PTI1 and PTI2 (Figure 7).

The PDI core can be substituted at 4 positions (Figure 7a), giving rise to the same number (1015) of mono-, di-, tri-, and tetra-substituted derivatives as the NDIs discussed previously. Calculated with sTDDFT, the lowest excitation energy of non-substituted PDI is 2.43 eV with an oscillator strength of 0.77. Compared to non-substituted NDI, its absorption is shifted

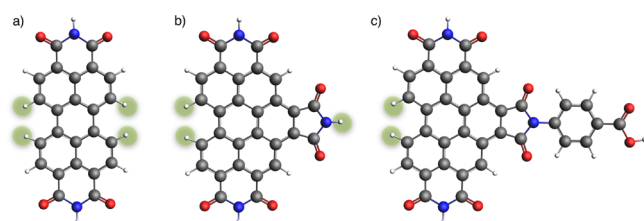


Figure 7. Three perylene-based chromophores (a) PDI, (b) PTI1, and (c) PTI2. Highlighted in green are substitution spots.

dramatically toward the lower energies due to the vertical extension of the naphthalene core. Electron-donating substituents additionally shift the absorption to lower energies, but the effect of substitution is less diverse compared to the NDI core. PDI derivatives have the lowest excitations from 1.6 to 2.3 eV. Nine hundred eighty-six molecules had a stable structure from which 488 fit to the criteria (Figure 8), as

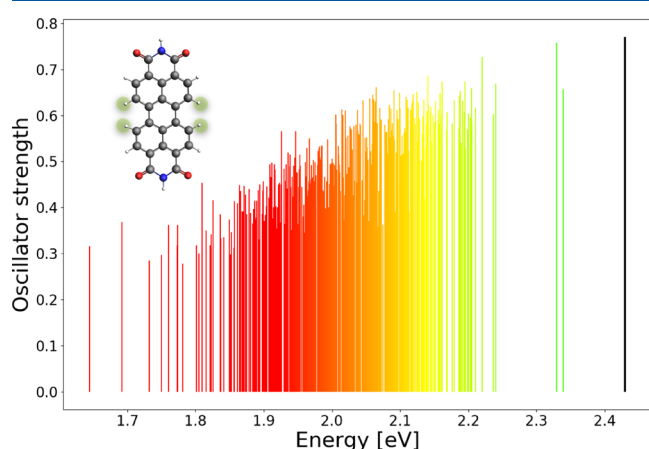


Figure 8. Absorption spectra based on the lowest excitation peaks from each promising PDI derivative calculated with sTDDFT. The color of the peaks resembles the color corresponding to the excitation. The absorption peak in black is from non-substituted PDI.

shown together with non-substituted PDI (black) in Figure 8. We see a systematic decrease in oscillator strength at lower energies and list in Table 3 the most promising PDI derivatives for each energy range.

The PTI1 core has three substitution spots, one at the core-extending imide group and two equivalent positions on the other side of the core (Figure 7b). For this molecule, we first attach ligands to the extended side of the core leading to seven possible mono-substituted molecules (or new cores). Ligands

Table 3. Computed Lowest Excitation Energies with the Highest Oscillator Strength in Small eV Intervals in the Energy Range of PDI Derivatives^a

energy range [eV]	PDI derivative	E_{VE}^{sTDDFT} [eV]	f	SA score
1.6–1.8	PDI-7133	1.77	0.39	6.96
1.8–1.9	PDI-5512	1.90	0.46	7.61
1.9–2.0	PDI-6744	1.93	0.57	7.92
2.0–2.1	PDI-4444	2.06	0.66	7.34
2.1–2.2	PDI-444	2.14	0.69	7.16
2.2–2.3	PDI-44	2.22	0.73	6.99
2.3–2.4	PDI-4	2.33	0.76	6.83

^aResults are obtained with sTDDFT(CAM-B3LYP)

are then attached to the two equivalent spots on the left side of the core, leading to 252 molecules in total, with 35 derivatives for each new core. The PTI2 core has only two positions for substitution (Figure 7c), which therefore leads to 35 PTI2 derivatives, seven mono-substituted and 28 di-substituted molecules. None of these 35 molecules have imaginary frequencies. The non-substituted PTI1 and PTI2 have almost the same lowest excitation energy, 2.88 and 2.89 eV, respectively, with also the same oscillator strength of 0.67. This indicates that the ligand attached on the core-extending imide group has no effect on the absorption properties.

For these molecules also, other ligands do not affect the optical properties significantly if attached to the core-extending imide group. In Figure 9, we noticed clusters of excitation

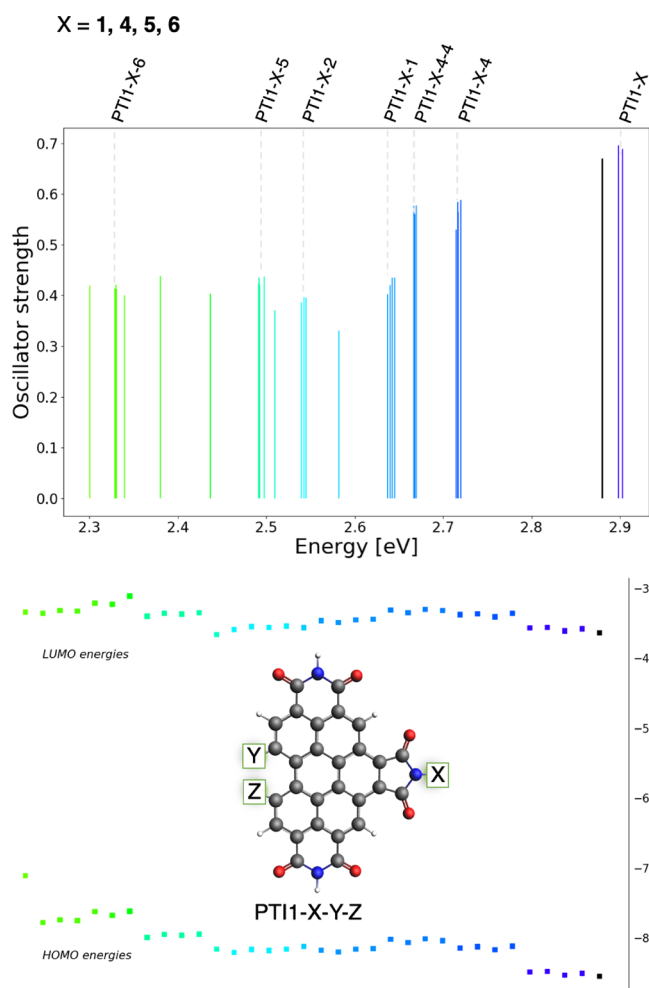


Figure 9. (top) sTDDFT(CAM-B3LYP) prediction of absorption spectra based on the lowest excitation peaks from each promising PTI1. The color of the peaks resembles the color corresponding to the excitation. The absorption peak in black is from the non-substituted PTI1 core; (bottom) HOMO and LUMO energies in eV for the promising PTI1 derivatives.

peaks in the set of promising derivatives. These are four derivatives where X is a ligand (1, 4, 5, and 6). Y and Z are ligands that determine the position of the peak, and the core-extending imide group is acting as a node. Other three ligands at X are showing the same peaks, but the criterion that the lowest excitation is the most intense one is not fulfilled. Figure 9 (bottom) shows HOMO and LUMO energies for the

promising derivatives. The black squares show the HOMO and LUMO energies for the non-substituted PTI1. Compared to non-substituted PTI1, the LUMO energies are slightly higher, up to 0.5 eV. The effect on the HOMO level is a shift of up to 1.5 eV. The lowest energy peak at 2.3 eV (Figure 9) belongs to PTI1-7-6. Table 4 shows the most promising PTI1 derivatives for each energy range. PTI2 provides six promising structures (Table 5), with the effects of its ligands being quite similar to PTI1.

Table 4. Lowest Excitation, Oscillator Strength, and SA Score for the Promising PTI1 Derivatives

energy range [eV]	PTI1 derivative	E_{VE}^{sTDDFT} [eV]	f	SA score
2.3–2.4	PTI1-66	2.38	0.44	7.39
2.4–2.5	PTI1-65	2.50	0.44	7.27
2.5–2.6	PTI1-12	2.54	0.40	7.24
2.6–2.7	PTI1-644	2.67	0.58	7.42
2.7–2.8	PTI1-644	2.72	0.59	7.23
2.8–2.9	PTI1-1	2.90	0.70	7.06

Table 5. Lowest Excitation, Oscillator Strength, and SA Score for All the Promising PTI2 Derivatives

PTI2 derivative	E_{VE}^{sTDDFT} [eV]	f	SA score
PTI2-6	2.36	0.44	7.37
PTI2-5	2.47	0.42	7.25
PTI2-2	2.51	0.38	7.36
PTI2-1	2.62	0.42	7.37
PTI2-44	2.64	0.54	7.40
PTI2-4	2.70	0.55	7.20

For the application in DS-PEC dyes, besides the strong light harvesting, the presence of a potential gradient is also important. For an example of a triad system, as in work of Monti, De Groot, and Buda,⁵⁸ the light absorbing dye also needs to fit the molecular redox properties, ground state oxidation potential (GSOP), and excited state oxidation potential (ESOP) of the two other components of the triad. Calculating these quantities is better done with a dedicated procedure,⁵⁹ the performance of which we plan to investigate in a separate paper. We note that, on the basis of the data calculated with the current workflow, one may give a simple estimate of the GSOP/ESOP using Koopman's theorem as shown in Table 6. The MAD of the LUMO estimate is high, but results are considerably more consistent compared to the HOMO energies and possibly useful to provide insights into the relative ESOP of one derivative as compared to the others. The better performance of the LUMO estimate of ESOP can also be used to obtain a slightly better estimation for GSOP by

Table 6. Statistical Analysis^a of the Simple DFT Estimates with the Experimentally Obtained GSOP and ESOP

method	MD	MAD	RMSD	R^2
ESOP(LUMO)	0.79	0.79	0.80	0.76
GSOP(HOMO) ^b	−1.95	1.95	1.97	0.57
GSOP(HOMO) ^{est, c}	0.15	0.24	0.28	0.76

^aMD stands for the mean deviation, MAD for the mean absolute deviation, RMSD is for the root-mean-square deviation, and R^2 is the squared correlation. ^bHOMO energy from DFT; ^cHOMO^{est} = LUMO − E_{VE}^{sTDDFT} .

subtracting the computed excitation energy from the LUMO, GSOP^{est} = $E(\text{LUMO}) - E_{VE}$.

CONCLUSIONS

We presented a tool for generating derivatives by systematic functionalization of a molecule to aid in exploring the chemical space of molecular derivatives. This generation tool can be combined with workflows for the screening of large numbers of molecules on the basis of computed molecular properties that fulfill certain criteria. We applied such a workflow to find molecules with potential to be used as photosensitizers in DS-PECs, using desirable optical properties as selection criteria. Results are shown for NDI's, PDI's, PTI1's, and PTI2's derivatives. The derivatives create a very diverse set of chromophores allowing panchromatic sensitization of the photoelectrode if adsorbed together on a semiconductor surface. The outcome is around 1400 derivatives that fulfill the criteria to be used as photosensitizers in DS-PECs. The workflow applied to NDI selects already known sensitizers out of 1015 structures, which shows its reliability. We also considered estimating the GSOP and ESOP by HOMO/LUMO energy levels but observed that this simple procedure based on DFT(CAM-B3LYP) values is not predictive enough for this purpose. The CAT code is available as an open source and is thereby extendable to other applications, where systematic screening of compound derivatives is needed, like in the discovery of new photoswitches or in the discovery of new water oxidation catalysts.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.0c04506>.

Excitation energies, oscillator strengths, and HOMO-LUMO gaps for all evaluated dyes (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jelena Belić – *Vrije Universiteit Amsterdam, Amsterdam 1081 HV, The Netherlands*; Email: j.belic@vu.nl

Lucas Visscher – *Vrije Universiteit Amsterdam, Amsterdam 1081 HV, The Netherlands*; orcid.org/0000-0002-7748-6243; Email: lvisscher@vu.nl

Authors

Bas van Beek – *Vrije Universiteit Amsterdam, Amsterdam 1081 HV, The Netherlands*

Jan Paul Menzel – *Leiden Institute of Chemistry, Leiden University, Leiden 2300 RA, The Netherlands*; orcid.org/0000-0002-1312-5000

Francesco Buda – *Leiden Institute of Chemistry, Leiden University, Leiden 2300 RA, The Netherlands*

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpca.0c04506>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research has been financially supported by the NWO Solar to Products program (project number 733.000.007). We acknowledge the use of supercomputer facilities at SURFsara

sponsored by NWO Physical Sciences, with financial support from The Netherlands Organization for Scientific Research (NWO).

REFERENCES

- (1) Fujishima, A.; Honda, K. Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature* **1972**, *238*, 37–38.
- (2) Youngblood, W. J.; Lee, S. H. A.; Kobayashi, Y.; Hernandez-Pagan, E. A.; Hoertz, P. G.; Moore, T. A.; Moore, A. L.; Gust, D.; Mallouk, T. E. Photoassisted Overall Water Splitting in a Visible Light-Absorbing Dye-Sensitized Photoelectrochemical Cell. *J. Am. Chem. Soc.* **2009**, *131*, 926–927.
- (3) Wu, L. Z.; Chen, B.; Li, Z. J.; Tung, C. H. Enhancement of the Efficiency of Photocatalytic Reduction of Protons to Hydrogen via Molecular Assembly. *Acc. Chem. Res.* **2014**, *47*, 2177–2185.
- (4) Yu, Z.; Li, F.; Sun, L. Recent Advances in Dye-Sensitized Photoelectrochemical Cells for Solar Hydrogen Production Based on Molecular Components. *Energy Environ. Sci.* **2015**, *8*, 760–775.
- (5) Moore, G. F.; Blakemore, J. D.; Milot, R. L.; Hull, J. F.; Song, H. E.; Cai, L.; Schmuttenmaer, C. A.; Crabtree, R. H.; Brudvig, G. W. A Visible Light Water-Splitting Cell with a Photoanode Formed by Codeposition of a High-Potential Porphyrin and an Iridium Water-Oxidation Catalyst. *Energy Environ. Sci.* **2011**, *4*, 2389–2392.
- (6) Dau, H.; Zaharieva, I. Principles, Efficiency, and Blueprint Character of Solar-Energy Conversion in Photosynthetic. *Acc. Chem. Res.* **2009**, *42*, 1861–1870.
- (7) Dau, H.; Limberg, C.; Reier, T.; Risch, M.; Roggan, S.; Strasser, P. The Mechanism of Water Oxidation: From Electrolysis via Homogeneous to Biological Catalysis. *ChemCatChem* **2010**, *2*, 724–761.
- (8) Gust, D.; Moore, T. A.; Moore, A. L. Solar Fuels via Artificial Photosynthesis. *Acc. Chem. Res.* **2009**, *42*, 1890–1898.
- (9) Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. Dye-Sensitized Solar Cells. *Chem. Rev.* **2010**, *110*, 6595–6663.
- (10) Kobaisi, M. A.; Bhosale, S. V.; Latham, K.; Raynor, A. M.; Bhosale, S. V. Functional Naphthalene Diimides: Synthesis, Properties, and Applications. *Chem. Rev.* **2016**, *116*, 1168511796.
- (11) Rüger, R.; Yakovlev, A.; Philippen, P.; Borini, S.; Melix, P.; Oliveira, A. F.; Franchini, M.; van Vuren, T.; Soini, T.; de Reus, M.; et al. SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands. *AMS DFTB version 2019*, 3 <http://www.scm.com>.
- (12) Oliveira, A. F.; Seifert, G.; Heine, T.; Duarte, H. A. Density-Functional Based Tight Binding: An Approximate DFT Method. *J. Braz. Chem. Soc.* **2009**, *20*, 1193–1205.
- (13) Rüger, R.; Van Lenthe, E.; Lu, Y.; Frenzel, J.; Heine, T.; Visscher, L. Efficient Calculation of Electronic Absorption Spectra by Means of Intensity-Selected Time-Dependent Density Functional Tight Binding. *J. Chem. Theory Comput.* **2015**, *11*, 157–167.
- (14) Bannwarth, C.; Grimme, S. A Simplified Time-Dependent Density Functional Theory Approach for Electronic Ultraviolet and Circular Dichroism Spectra of Very Large Molecules. *Comput. Theor. Chem.* **2014**, *1040-1041*, 45–53.
- (15) Risthaus, T.; Hansen, A.; Grimme, S. Excited States Using the Simplified TammDancoff-Approach for Range-Separated Hybrid Density Functionals: Development and Application. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14408–14419.
- (16) Chen, H.-C.; Hsu, C.-P.; Reek, J. N.; Williams, R. M.; Brouwer, A. M. Highly Soluble Benzo[ghi]perylene triimide Derivatives: Stable and Air-Insensitive Electron Acceptors for Artificial Photosynthesis. *ChemSusChem* **2015**, *8*, 3639–3650.
- (17) Chen, H.-C.; Williams, R. M.; Reek, J. N.; Brouwer, A. M. Robust Benzo[ghi]perylene triimide Dye-Sensitized Electrodes in Air-Saturated Aqueous Buffer Solution. *Chem. Eur. J.* **2016**, *22*, 5489–5493.
- (18) Suraru, S. L.; Würthner, F. Strategies for the Synthesis of Functional Naphthalene Diimides. *Angew. Chem., Int. Ed.* **2014**, *53*, 7428–7448.
- (19) Chopin, S.; Chaignon, F.; Blart, E.; Odobel, F. Syntheses and Properties of Core-Substituted Naphthalene Bisimides with Aryl Ethynyl or Cyano Groups. *J. Mater. Chem.* **2007**, *17*, 4139–4146.
- (20) Bhosale, S. V.; Jani, C. H.; Langford, S. J. Chemistry of Naphthalene Diimides. *Chem. Soc. Rev.* **2008**, *37*, 331–342.
- (21) Sakai, N.; Mareda, J.; Vauthey, E.; Matile, S. Core-Substituted Naphthalenediimides. *Chem. Commun.* **2010**, *46*, 4225–4237.
- (22) Zhou, N.; Facchetti, A. Naphthalenediimide (NDI) Polymers for All-Polymer Photovoltaics. *Mater. Today* **2018**, *21*, 377–390.
- (23) Thalacker, C.; Röger, C.; Würthner, F. Synthesis and Optical and Redox Properties of Core-Substituted Naphthalene Diimide Dyes. *J. Org. Chem.* **2006**, *71*, 8098–8105.
- (24) Jones, B. A.; Facchetti, A.; Wasielewski, M. R.; Marks, T. J. Tuning Orbital Energetics in Arylene Diimide Semiconductors. Materials Design for Ambient Stability of n-Type Charge Transport. *J. Am. Chem. Soc.* **2007**, *129*, 15259–15278.
- (25) Gsänger, M.; Bialas, D.; Huang, L.; Stolte, M.; Würthner, F. Organic Semiconductors Based on Dyes and Color Pigments. *Adv. Mater.* **2016**, *28*, 3615–3645.
- (26) Röger, C.; Würthner, F. Core-Tetrasubstituted Naphthalene Diimides: Synthesis, Optical Properties, and Redox Characteristics. *J. Org. Chem.* **2007**, *72*, 8070–8075.
- (27) Würthner, F.; Ahmed, S.; Thalacker, C.; Debaerdemaeker, T. Core-Substituted Naphthalene Bisimides: New Fluorophors with Tunable Emission Wavelength for FRET Studies. *Chem. – Eur. J.* **2002**, *8*, 4742–4750.
- (28) Krüger, H.; Janietz, S.; Sainova, D.; Dobrova, D.; Koch, N.; Vollmer, A. Hybrid Supramolecular Naphthalene Diimide-thiophene Structures and their Application in Polymer Electronics. *Adv. Funct. Mater.* **2007**, *17*, 3715–3723.
- (29) Mulder, J. R.; Guerra, C. F.; Slootweg, J. C.; Lammertsma, K.; Bickelhaupt, F. M. Substituent Effects on the Optical Properties of Naphthalenediimides: A Frontier Orbital Analysis Across the Periodic Table. *J. Comput. Chem.* **2015**, *37*, 304–313.
- (30) Kishore, R. S. K.; Kel, O.; Banerji, N.; Emery, D.; Bollot, G.; Mareda, J.; GomezCasado, A.; Jonkheijm, P.; Huskens, J.; Maroni, P.; et al. Ordered and Oriented Supramolecular n/p-Heterojunction Surface Architectures: Completion of the Primary Color Collection. *J. Am. Chem. Soc.* **2009**, *131*, 11106–11116.
- (31) SCM *Theoretical Chemistry*; Vrije Universiteit: Amsterdam, The Netherlands, <http://www.scm.com>.
- (32) van Beek, B.; Belić, J. CAT version 0.8.7; GitHub, Inc. <https://github.com/nlesc-nano/CAT>. <https://doi.org/10.5281/zenodo.3832787>, 2020.
- (33) PLAMS, SCM *Theoretical Chemistry*; Vrije Universiteit, Amsterdam, The Netherlands. <https://github.com/SCM-NV/PLAMS>.
- (34) Hässelbarth, W. Substitution Symmetry. *Theor. Chim. Acta* **1985**, *67*, 339–367.
- (35) Read, R. C. Pólya's Theorem and Its Progeny. *Math. Mag.* **2018**, *60*, 275–282.
- (36) Read, R. C.; Polya, G. *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*; Springer-Verlag New York Inc., 1987.
- (37) Landrum, G. *RDKit: Open-source cheminformatics*; GitHub, 2006, <http://www.rdkit.org>.
- (38) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.
- (39) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- (40) Kubillus, M.; Kubar, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.
- (41) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *J. Phys. Chem. B* **2015**, *119*, 1062–1082.

- (42) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2012**, *9*, 338–354.
- (43) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014**, *10*, 1518–1537.
- (44) Deglmann, P.; Furche, F. Efficient Characterization of Stationary Points on Potential Energy Surfaces. *J. Chem. Phys.* **2002**, *117*, 9535–9538.
- (45) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A Long-Range-Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2004**, *120*, 8425–8433.
- (46) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (47) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128*, No. 044118.
- (48) Pastore, M.; Mosconi, E.; De Angelis, F.; Grätzel, M. A Computational Investigation of Organic Dyes for Dye-Sensitized Solar Cells: Benchmark, Strategies, and Open Issues. *J. Phys. Chem. C* **2010**, *114*, 7205–7212.
- (49) Jacquemin, D.; Planchat, A.; Adamo, C.; Mennucci, B. TD-DFT Assessment of Functionals for Optical 0-0 Transitions in Solvated Dyes. *J. Chem. Theory Comput.* **2012**, *8*, 2359–2372.
- (50) Narsaria, A. K.; Ruijter, J. D.; Hamlin, T. A.; Ehlers, A. W.; Guerra, C. F.; Lammertsma, K.; Bickelhaupt, F. M. Performance of TDDFT Vertical Excitation Energies of Core-Substituted Naphthalene Diimides. *J. Comput. Chem.* **2020**, 1448–1455.
- (51) Yun, S.; Vlachopoulos, N.; Qurashi, A.; Ahmad, S.; Hagfeldt, A. Dye Sensitized Photoelectrolysis Cells. *Chem. Soc. Rev.* **2019**, *48*, 3705–3722.
- (52) Heßelmann, A.; Görling, A. Blindness of the Exact Density Response Function to Certain Types of Electronic Excitations: Implications for Time-Dependent Density Functional Theory. *Phys. Rev. Lett.* **2009**, *102*, 233003.
- (53) Casida, M. E. Time-Dependent Density-Functional Theory for Molecules and Molecular Solids. *J. Mol. Struct. Theochem.* **2009**, *914*, 3–18.
- (54) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-Like Molecules Based on Molecular Complexity and Fragment Contributions. *Aust. J. Chem.* **2009**, *1*, 8.
- (55) Chopin, S.; Chaignon, F.; Blart, E.; Odobel, F. Syntheses and Properties of Core-Substituted Naphthalene Bisimides with Aryl Ethynyl or Cyano Groups. *J. Mater. Chem.* **2007**, *17*, 4139–4146.
- (56) Humeniuk, A.; Mitrić, R. Long-Range Correction for Tight-Binding TD-DFT. *J. Chem. Phys.* **2015**, *143*, 134120.
- (57) Ferrer, F. J. A.; Cerezo, J.; Stendardo, E.; Impropa, R.; Santoro, F. Insights for an Accurate Comparison of Computational Data to Experimental Absorption and Emission Spectra: Beyond the Vertical Transition Approximation. *J. Chem. Theory Comput.* **2013**, *9*, 2072–2082.
- (58) Monti, A.; De Groot, H. J. M.; Buda, F. In-Silico Design of a Donor-Antenna-Acceptor Supramolecular Complex for Photo-induced Charge Separation. *J. Phys. Chem. C* **2014**, *118*, 15600–15609.
- (59) Pastore, M.; Fantacci, S.; De Angelis, F. Ab Initio Determination of Ground and Excited State Oxidation Potentials of Organic Chromophores for Dye-Sensitized Solar Cells. *J. Phys. Chem. C* **2010**, *114*, 22742–22750.