Article

Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics

Steven A. Lopez, Benjamin Sanchez-Lengeling, Julio de Goes Soares, Alán Aspuru-Guzik

aspuru@chemistry.harvard.edu

HIGHLIGHTS

Statistical study of >51,000 non-fullerene acceptors for organic photovoltaics

Machine learning is used to calibrate theoretical results to experimental data

838 high-performing (PCE >8%) molecules have been identified

Statistical analysis on fragments offers insights to design rules

The power conversion efficiencies of organic photovoltaics (OPVs) have grown tremendously over the last 20 years and represent a low-cost and sustainable solution for harnessing solar energy to power our residences, workplaces, and devices. Fullerene-containing OPVs are relatively expensive and have limited overlap absorbance with the solar spectrum. We used density functional theory calculation and Gaussian processes calibration to predict frontier molecular orbitals and power conversion efficiencies for over 51,000 non-fullerene acceptors with the polymeric electron-donor component, poly[N-9'-heptadecanyl-2,7-carbazole-alt-5,5-(4',7'-di-2-thienyl-2',1',3'-benzothiadiazole)].

CrossMark

Article

Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics

Steven A. Lopez,[1,4] Benjamin Sanchez-Lengeling,[1,4] Julio de Goes Soares,[2] and Alán Aspuru-Guzik[1,3,5,*]

SUMMARY

Non-fullerene acceptors in organic photovoltaics (OPVs) continue to improve upon the shortcomings of many fullerene-based solar cells. We explored the chemical space of over 51,000 non-fullerene acceptors featuring 106 common moieties from the organic electronics literature, including naphthalene diimides, benzathiadiazoles, and fused fluoroanthenediimides. We identify top candidates featuring optimal energy level offsets, based on a well-studied electron-donor: poly[N-9'-heptadecanyl-2,7-carbazole-alt-5,5-(4',7'-di-2-thienyl-2',1',3'-benzothiadiazole)]. The Harvard Clean Energy Project infrastructure was employed through the IBM World Community Grid to carry out quantum mechanical calculations using density functional theory. Gaussian processes regression was utilized to correct the computed frontier molecular orbital energies. Additional time-dependent density functional theory calculations on a subset of the top candidates refined the narrow band-gap chromophores for OPVs. These results on electron-acceptor materials complement the electronic structure calculations of over one million electron-donor materials publicly available through our website.
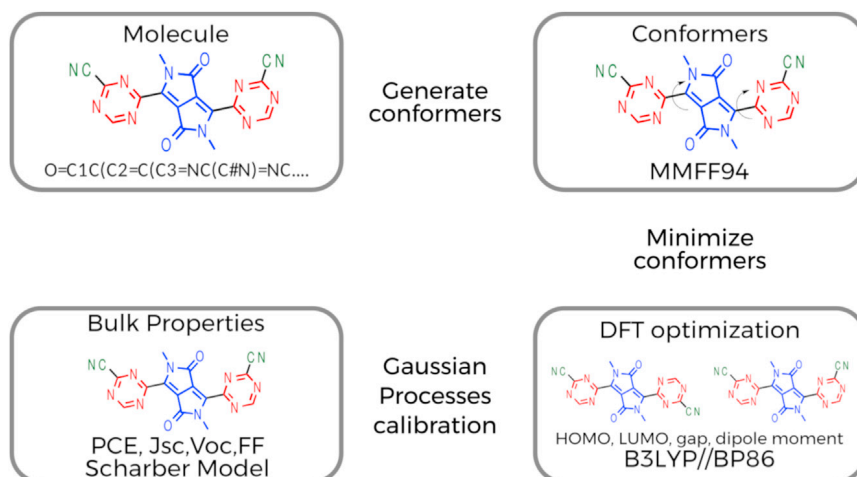
INTRODUCTION

The earth receives 1.52×10^9 TWh of energy annually, which greatly exceeds the 161,000 TWh of power consumed globally, and the 239,000 TWh projected consumption by 2040.[1] The development of affordable photovoltaic cells is therefore one of the most promising long-term solutions for sustainable energy. Low-cost, flexible, organic photovoltaics (OPVs) are easily fabricated and can be produced using inexpensive techniques such as roll-to-roll printing.[2] Despite competition from other photovoltaic devices such as perovskites,[3–7] they have shown long lifetimes and remain an attractive option.[8,9] The number of publications involving OPVs has risen dramatically over the last 15 years. A recent article by Zhao et al.[10] reported a record-high efficiency of 13% with a non-fullerene acceptor indacenodithiophene derivative. Ternary solar cells also have been reported to have power conversion efficiencies (PCEs) greater than 12%.[11] These advances can be attributed to improvements in device architecture and processing, and materials design.[12–18]

Donor materials for OPVs have traditionally received much more attention than the acceptor component because of the continued success of fullerene-based derivatives in OPVs.[19] [6,6]-Phenyl-C_{61}-butyric acid methyl ester (PCBM) is regarded by many as the standard electron acceptor in OPVs.[20] The energy levels of donor materials have been tuned to have optimal overlap with those of PCBM.[21] However, substantial effort has been put forth over the last 10 years to address the high

Context & Scale

Acceptor materials for single-junction organic solar cells (OSCs) are currently dominated by fullerene derivatives, which suffer from high production costs and difficult functionalization. Non-fullerene small-molecules acceptors (NFA) might be able to overcome these shortcomings for OSCs. This work represents a statistical study of around 51,000 NFA molecules utilizing quantum chemistry calculations, the Scharber model to relate molecular frontier orbital energies with photovoltaic conversion efficiency, and machine learning techniques to calibrate theoretical results to experimental data. Within the explored molecular space, 838 high-performing candidate molecules have been identified. Molecular fragments are scored based on performance to identify potential design patterns for future materials. The design rules and top candidates warrant future work on synthesis and device fabrication to characterize their performance. A lost-cost, easy to synthesize molecule might potentially be a driver for wider adoption and higher performance in organic photovoltaics.

Scheme 1. An Illustration of the Workflow of the HTVS Process in the Harvard Clean Energy Project

production costs,[22] limited absorbance of visible light, and difficult modulation of the electronic properties of fullerenes.[23,24] Thus, incorporating low-cost, synthesizable non-fullerene acceptors in OPV devices is highly desirable and the subject of substantial efforts in the research community. These efforts have produced a set of non-fullerene acceptor materials with comparable orbital energies with PCBM and its derivatives. Five common classes of these include compounds based on perylene diimides (PDIs),[25–28] tetraazabenzodifluoranthenes (BFIs),[29] diketopyrrolopyrroles (DPPs),[30,31] fluoranthene-fused imides, and benzothiadiazole (BT).[32–36] Recently, Kuzmich et al.[37] showed the importance of a low-lying LUMO + 1 level in top non-fullerene acceptor materials. The advantages of these materials and the greater diversity of molecular packing arrangements suggest that they may overtake fullerenes as the dominant acceptor in low-cost bulk heterojunction OPVs.[38]

Harvard Clean Energy Project

The Harvard Clean Energy Project (CEP) is a distributed computing effort for screening OPV candidates carried out by volunteers connected to the IBM World Community Grid.[39] The first release of CEP screening data for a high-throughput virtual screening (HTVS) featured 1.3 million donor materials from an initial 26 common fragments in the organic electronics literature.[40–42] Other *in silico* screening efforts have focused on small-molecule and polymer electron-donor materials and address charge transport in the solid state.[43,44] Here, we have extended the initial HTVS via the Harvard Clean Energy Project (CEP) to screen a library of more than 51,000 potential non-fullerene acceptor materials. It is a systematic, automated, volunteer-powered tool for HTVS of materials for OPVs; the workflow is illustrated in Scheme 1.

Our automated workflow allows us to efficiently explore the conformational space of each candidate with molecular mechanics. Quantum mechanical calculations are utilized to obtain the electronic structure of the molecule (details given in the computational methods section). Finally, the Gaussian processes (GP) technique is used to calibrate the computed highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) to experimentally determined values. With this workflow, we can screen large libraries of non-fullerene acceptor candidates for top-performing new OPV devices.

[1]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

[2]Instituto de Química, Universidade de São Paulo, São Paulo, 05508-900, Brazil

[3]Canadian Institute for Advanced Research, Bioinspired Solar Energy Program, Toronto, ON M5G 1Z8, Canada

[4]These authors contributed equally

[5]Lead Contact

*Correspondence:
aspuru@chemistry.harvard.edu

These di- and monosubstituted fragments are shown in Scheme 2; each fragment and the substitution pattern are informed by the literature or commercial precedent. They are organized into three categories: cores, spacers, and terminals. Cores and spacers are both disubstituted; terminal groups are monosubstituted. The full screening set contains a total of 107 fragments with 13 cores, 49 spacers, and 45 terminal groups.
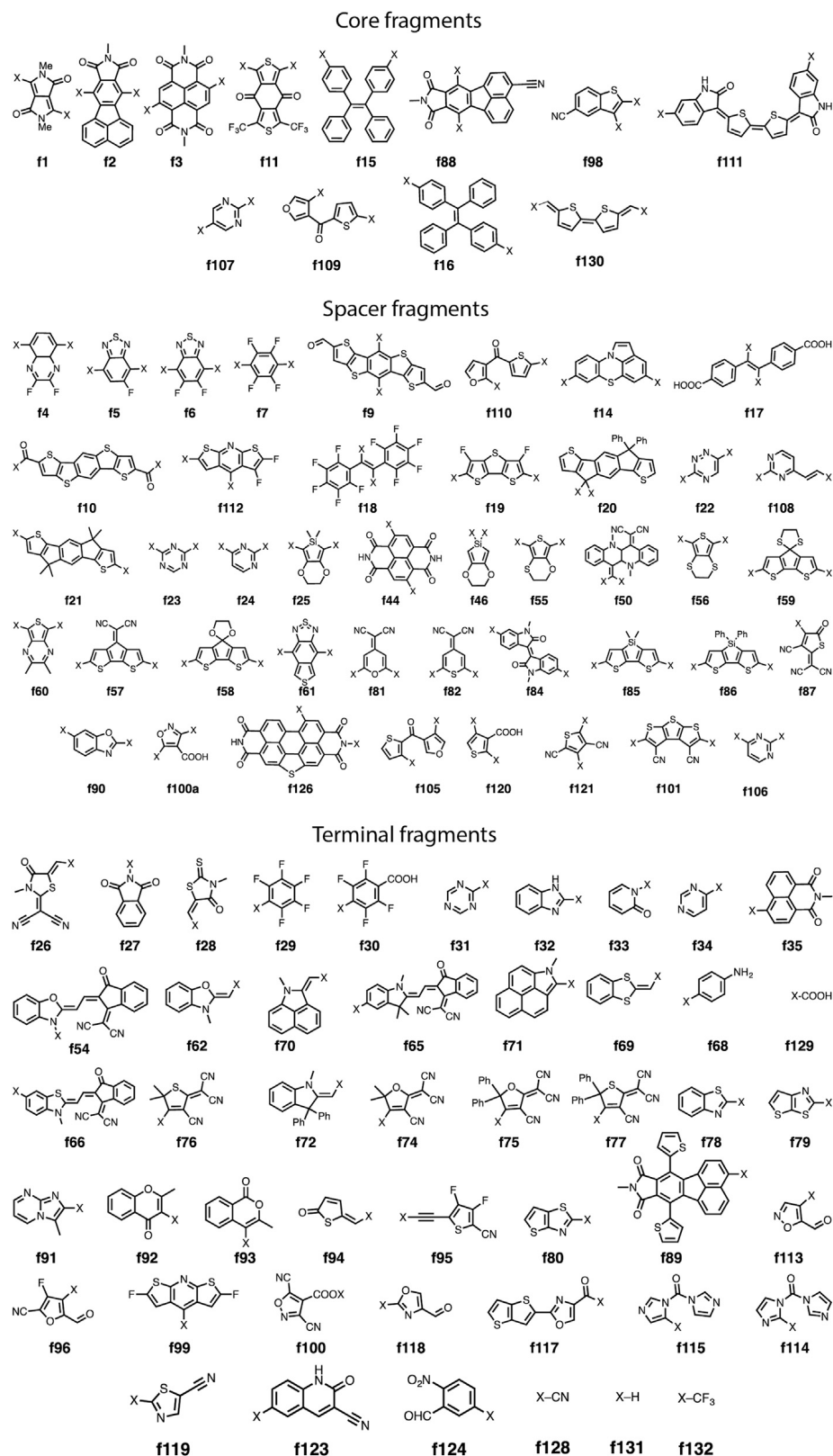
RESULTS

Molecular Generation and DFT Calculations

The molecular generation process builds symmetrical and unsymmetrical candidates for non-fullerene acceptor materials. This involves the combination of a core, two spacers, and two terminal fragments (T-S-C-S-T). The incorporation of non-identical spacer or terminal fragments leads to non-symmetric molecules of this format (i.e., T'-S-C-S-T). Another class of molecules results when a single core or spacer fragment is combined with a terminal fragment (T-C-T) or (T-S-T), which can be symmetric or non-symmetric. Due to restrictions from the World Community Grid computational resource, the molecules that enter this work flow are limited to 306 electrons or less. A simplified molecular-input line-entry system (SMILES)[45] string representation of the molecule is generated for each of the molecules in our library. An initial set of 1,500 3D molecular conformation guesses are generated via the RDKit[46] conformer generator. The proposed 3D structures are minimized using the MMFF force field,[47] and duplicate structures are removed using the *obfit* functionality implemented in the *OpenBabel* software package.[48] Up to 20 lowest energy conformer clusters within 5 kcal mol^{-1} are assumed to represent the most energetically feasible conformations in the solid state for each candidate molecule. The geometries for each of these conformers were optimized with the BP86[49,50]/def2-SVP[51] density functional theory (DFT) model chemistry. For each unique conformation, single-point energies were calculated with B3LYP using the double-ζ def2-SVP level of theory to improve the accuracy of the HOMO and LUMO energies. These frontier molecular orbital energies were then calibrated to correct for the systematic errors in our quantum chemical calculations. The calibration was performed using GP regression[52] based on a set of experimental HOMO and LUMO energies for 94 molecules, including those in the recently published *Harvard Organic Photovoltaic Set*[53](see also Tables S1 and S2).

Machine Learning: Gaussian Processes Regression

A classical approach for improving the predictive power of a theoretical model is to calibrate its biased predictions against experimentally observed data and gradually decrease the gap between experimental and theoretical values. One simple strategy of calibration of frontier orbital energies would be to use linear regression, but this has proven to be inadequate for our data, which do not follow a linear trend. Instead, we utilize GP regression[54] with molecular fingerprints to calibrate molecular orbital energies based on molecular similarity. GP regression is an established probabilistic framework in the field of machine learning to build flexible models, while furnishing uncertainty bounds on predictions. We recently showed that this approach can eliminate the functional dependence of orbital energies[55] and improve upon linear regression for molecular orbital energy prediction.

We generated Morgan fingerprints[56] to measure molecular similarity. Morgan fingerprints are fixed-length binary vectors (bit vectors) indicating the absence (zeros) or presence (ones) of a particular graph-connectivity environment. Each environment captures the local topological information of a molecule by mapping the local vicinity of connected atoms, along with their formal charges, type of chemical bond,

Core fragments

f1 f2 f3 f11 f15 f88 f98 f111

f107 f109 f16 f130

Spacer fragments

f4 f5 f6 f7 f9 f110 f14 f17

f10 f112 f18 f19 f20 f22 f108

f21 f23 f24 f25 f44 f46 f55 f50 f56 f59

f60 f57 f58 f61 f81 f82 f84 f85 f86 f87

f90 f100a f126 f105 f120 f121 f101 f106

Terminal fragments

f26 f27 f28 f29 f30 f31 f32 f33 f34 f35

f54 f62 f70 f65 f71 f69 f68 f129

f66 f76 f72 f74 f75 f77 f78 f79

f91 f92 f93 f94 f95 f80 f89 f113

f96 f99 f100 f118 f117 f115 f114

f119 f123 f124 f128 f131 f132

Scheme 2. Disubstituted Core, Spacer, and Terminal Fragments Used in This Study
The label under each structure indicates the fragment, f, and the number.

and its position relative to a cyclic structure. These environments are indexed and hashed into a position (bit) within the fixed-length vector via a modulus operation. The likelihood of overlapping environments hashed onto the same bit substantially decreases with vector length. Here, the fingerprints feature a length of 1,024 bits and the connectivity radius is four bonds. Molecular similarity between two fingerprints x_1 and x_2 can then be computed via the Tanimoto similarity index (i.e., Jaccard similarity). This approach can be adapted to quantify similarity diversity between molecules in GP regression.

Predictions of GPs can be interpreted as weighted averages of training data, where the weights are probabilistic in nature. GPs are distinct because of their associated covariance functions (i.e., kernel). Equation 1 shows that the kernel we employ is a mixture of a squared exponential kernel and a noise kernel, which increases the robustness of the model.

$$K(x_1, x_2) = A_{\text{kernel}} e^{-0.5 \left(\frac{\text{Tanimoto}(x_1, x_2)}{fp_\sigma} \right)^2} + A_{\text{noise}}^2 I . \qquad \text{(Equation 1)}$$

Equation 1 has three hyperparameters Akernel, fpσ, and Anoise, which control how distances between observations of data are interpolated and smoothed. These parameters are optimized to improve agreement with data and model flexibility. An initial Bayesian optimization is carried out to localize an acceptable starting point,[57] followed by gradient descent to optimize the hyperparameters. The training was done with 94 training points, molecules with experimental and calculated orbital energies. To ensure we are not overfitting, we used leave-one-out cross-validation (LOO-CV), where we train our model on all data points except one point and try to predict its value; this is done for all 94 data points, and so reported predictions are always on untrained data. LOO-CV calibration results are shown in Figure 1. It is also important to note that GPs are inherently robust to overfitting since the training procedure penalizes more complex models (higher-rank kernels).[58]

To account for some of the effects not treated by the Scharber model and to better match our predicted values to experiment, we performed an additional GP regression on the PCE. For this model, we utilized a set of 49 reported experimental values and their predicted Scharber model PCE values. Beside molecular similarity, we found it is crucial to take into account molecular weight and bandgap energies to arrive at a reasonable fit. These additional properties were added to the model via a radial basis function kernel, which substitutes the Tanimoto distance in Equation 1 for Euclidean distance. We measured goodness of fit with two statistics: Pearson's r coefficient of linear correlation and R^2, the coefficient of determination, a more general (non-linear) measure of fitness. R^2 can range from negative infinity to 1, with 0 representing the predictive power of the mean value of the data. Calibration of PCE can improve the extremely poor correlation between experimental and Scharber PCEs (r = 0.30, R^2 = −2.89) to r = 0.43 and R^2 = 0.11 (see Figure S2). This model was also optimized and then validated with LOO-CV.

Figure 1 shows that both frontier orbital energies present a linear relationship between the computed and experimental values, in both cases with high linear correlation coefficients (r = 0.91 for HOMO, r = 0.81 for LUMO). As expected, the LUMO energies have a weaker correlation between computed and experimental. Upon calibration, the root-mean-square error decreases by around 40% in both cases. Correlation coefficients r and R^2 increase in both cases; R^2 presents the most substantial gains, which suggests that the GP model more fully captures the overall variance
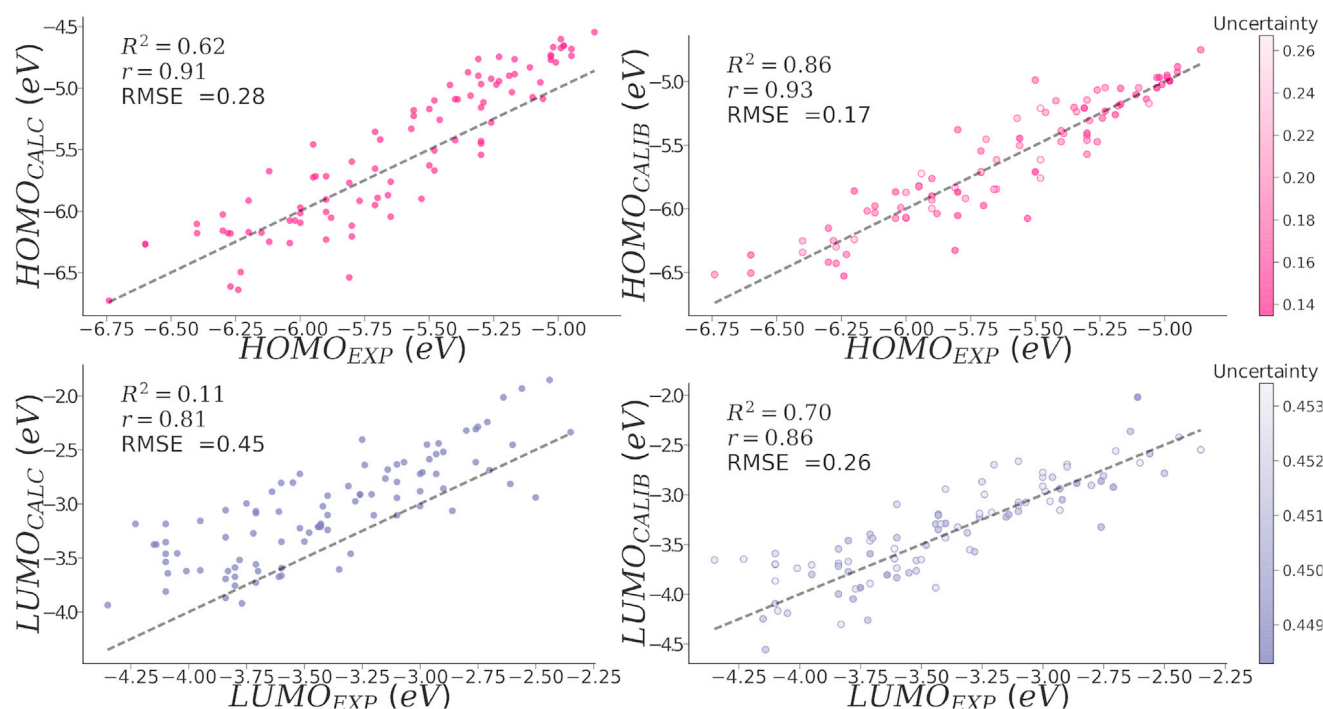
Figure 1. Calibration of HOMOs and LUMOs of the Training Set with Experimental (EXP), Calculated (CALC), and GP-Calibrated LOO-CV Predictions (CALIB)

Calculations were done with B3LYP/def2-SVP//BP86/def2-SVP. Uncertainty increases with lightly shaded data points. Dotted line represents the ideal fit.

and non-linearity of the experimental-theoretical energy difference. To further understand the effects of calibration, we present Figure 2, where we look at how the distribution of the HOMO-LUMO gap shifts before and after calibration and at the contribution of each calibration to the PCE. The gap on average shifts -0.26 eVs with an SD of 0.12 eV. The HOMO-LUMO GP calibration slightly increases the PCE by an average of 0.44%, while the PCE GP calibration presents two modes around zero, one slightly increasing the value and the other decreasing the majority of molecules by around -0.97%. Overall, the PCE is negatively shifted. These calibrated HOMO, LUMO, and gap values enter our modified Scharber model[59,60] (Equation 2) for computing open-circuit voltage for electron-acceptor materials, and further refined with an extra additive term. In essence, it is a modified version of the Shockley-Queisser model calibrated to better match experimental values.[61]

DISCUSSION

Scharber Model: PCE Calculations

Here, we utilize poly[N-9′-heptadecanyl-2,7-carbazole-*alt*-5,5-(4′,7′-di-2-thienyl-2′,1′,3′-benzothiadiazole)] (PCDTBT), a standard electron-donor material. Blouin et al.[62] report estimated HOMO and LUMO energies of PCDTBT based on cyclic voltammetry measurements as -5.50 eV and -3.60 eV, respectively. The open-circuit voltage (V_{OC}) is the ΔE_{DA} (energy difference between the D^{HOMO} and A^{LUMO}), and the short-circuit current density (J_{SC}) assumes that all photons in the solar spectrum with energy greater than the HOMO-LUMO gap of the acceptor material are absorbed. To give a non-zero predicted PCE, the $D^{LUMO} - A^{LUMO}$ offset is required to be larger than 0.3 eV to promote interfacial charge separation.[63] We incorporate an empirical loss parameter of 0.3 eV, which accounts for typical voltage loss in experimental devices. The Scharber model stipulates that the fill factor (FF) and
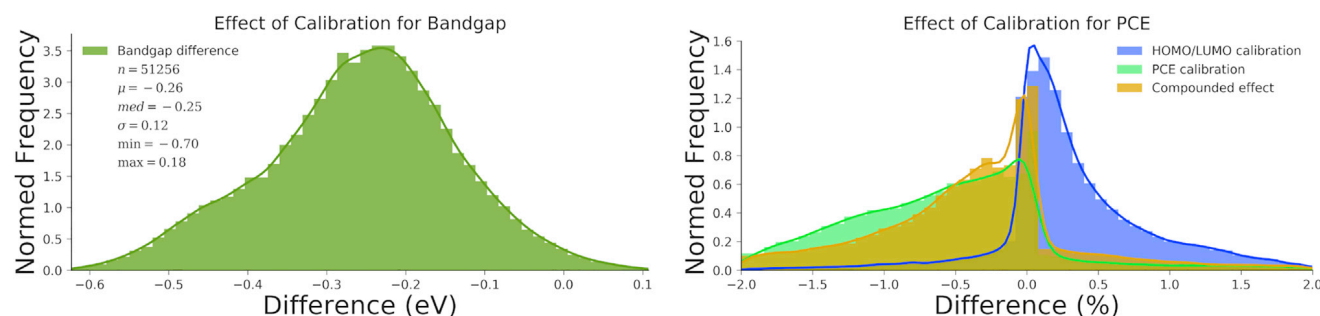
Figure 2. Effects of Calibration on HOMO-LUMO Gaps (Left) and Power Conversion Efficiencies (Right)
Each distribution is presented as a histogram and a kernel density estimate from the difference between a baseline and a calibration. Gap presents the general statistics of the distribution. PCE comparison shows the contribution of each calibration and its total effect. Effects on other photovoltaic properties can be found in Figure S1.

external quantum efficiency are set to 65%. PCEs are computed using Equation 3. We are cognizant that these predicted PCEs represent PCE values based on the highly simplified, but widely utilized Scharber model given PCDTBT as the donor material. In particular, the Scharber model does not take account of morphological considerations of the bulk or interfacial effects. Our predicted PCE values are point estimates with uncertainty bounds, and as such should be interpreted as estimates that permit the screening of many candidates with unfavorable intrinsic electronic structures. When comparing with experimental values, it should be noted that device optimization can play a huge role that is not accounted for in the model. For example, $P3HT/PC_{61}BM$ reported experimental values range from 0% to 6.5% from a survey of 579 publications.[64] Our PCE calibration model includes 49 unique donor-acceptor combinations. As a result, the variance of experimental values for each device cannot be included.[65] This is reflected in the weak correlation between Scharber PCEs and experimental PCEs.

$$V'_{OC} = \left(\frac{1}{e}\right)\left(|E^{donor}HOMO| - |E^{acceptor}LUMO|\right) - 0.3 \quad eV,$$ (Equation 2)

$$PCE = 100\frac{V_{OC} \cdot FF \cdot J_{SC}}{P_{in}}.$$ (Equation 3)

Figure 2 shows that calibration results in an overall decrease of the HOMO-LUMO gap. This is mostly due to an overall lowering of the acceptor LUMO energies from their uncalibrated values. This results in improved computed J_{SC} values and an overall improvement of predicted PCEs. The PCE shows that calibration increases the number of promising candidates, but not evenly, since there are cases where the PCE contribution is negative. Our approach has identified a significant population of molecules with appropriate electronic structures for OPVs (see Figure S3), while eliminating the clear majority of molecules that do not meet the minimum electronic requirements for photon absorption or charge separation.

We then performed a statistical analysis of the initial fragments to determine which fragments are overrepresented or underrepresented relative to their population in the initial fragment library. For these purposes, we computed Z scores, The hypergeometric distribution gives the probability of finding k instances in a subpopulation n, given K observations in the entire population N.[66] In our context, for a given fragment F, k is the frequency of molecules containing F, within the population of top molecules (size n) compared with the frequency K of molecules also containing F within the population of non-top molecules (size N). From this probability distribution, we can compute hypergeometric Z scores, which are similar to Gaussian Z scores. The Z score
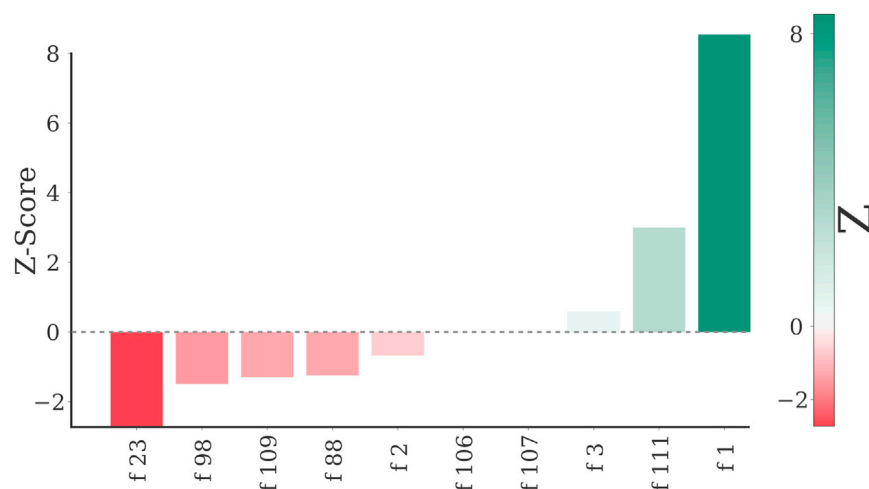
Figure 3. Barplots of Z Scores of the Core Fragments Pertaining to the Top Candidates with PCEs > 8.0%

The overrepresented (positive) Z scores are shown in green and the underrepresented (negative) Z scores are shown in red. Fragments not shown are not contained in top candidate molecules.

quantifies the probability of finding a fragment within potential target molecules relative to the mean (Z = 0) in units of SD according to Equation 4.

$$z = \frac{(k - \langle k \rangle)}{\sigma(k)} \,. \tag{Equation 4}$$

Here, k is the frequency of a fragment in the top set (PCE > 8.0%); $\langle k \rangle$ is the expectation value given by $\langle k \rangle = nK/N$. Z scoring leverages average PCE and variance into a single measure for comparison. Positive Z scores indicate that an individual fragment is overrepresented in the top candidates; negative Z scores indicate that a particular fragment is underrepresented among the top candidates. Z scores were computed for cores, spacers, and terminals to elucidate design principles for top non-fullerene acceptor materials. Figure 3 shows the Z score distribution for the core fragments. Molecules that contain higher Z scored fragments also have on average, higher PCE values. This relationship is further explored in Figures S4 and S5, which contain PCE boxplots and population counts of fragments for top and non-top molecules. Fragments not contained in top candidate molecules cannot be Z scored and are not included in such analysis.

Figure 3 shows that the DPP, NDI, and quinoidal oligothiophene core fragments (f1, f3, and f111, respectively) are statistically overrepresented in top candidates. These results are congruent with the literature. NDI and DPP derivatives frequently appear in the OPV literature involving non-fullerene acceptor components. Reports involving quinoidal thiophenes are considerably less common, but this class of molecules appears to be quite promising because of their improved absorbance in the visible range of the solar spectrum, relatively straightforward synthetic routes, and PCEs reaching 5.2% in TiO_2 dye-sensitized solar cells.[67] We then extended this analysis to "spacer" fragments, which is shown in Figure 4. Diazines and aryl ketones that are particularly electron deficient (f107, f109) are underrepresented in the top set, as demonstrated by negative Z scores.

Fragments with negative Z scores are typically more electrophilic (isoxazole [f100a] and perfluorinated phenyl [f7] spacers). These cores interact with the spacers through to a push-pull effect that decreases the HOMO-LUMO gap and improves
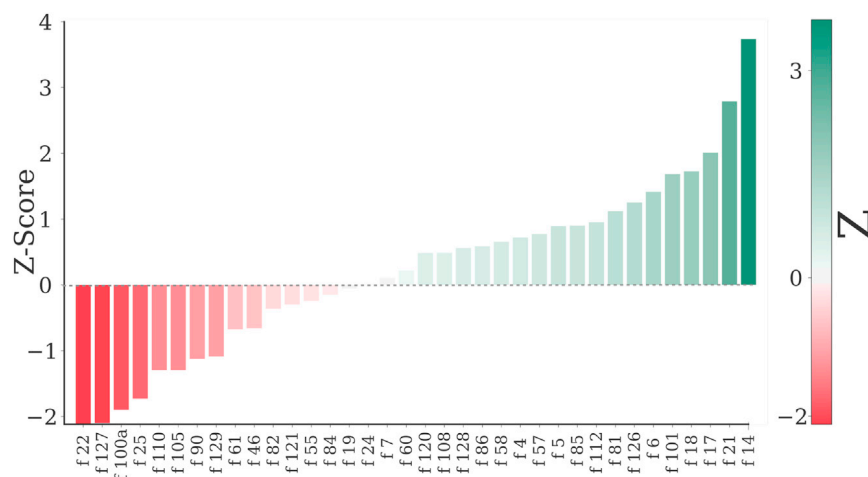
Figure 4. Barplots of Z Scores of the Spacer Fragments Pertaining to the Top Candidates with PCEs > 8.0%

The overrepresented (positive) Z scores are shown in green and the underrepresented (negative) Z scores are shown in red. Fragments not shown are not contained in the top candidate molecules.

J_{SC} and in turn, PCE. The electron-poor nature of these fragments decreases both HOMO and LUMO energies. As a result, the HOMO-LUMO gap remains approximately constant or increases with two adjacent electrophilic fragments.[68,69] The p-thiophene-phenylene-thiophene[70] spacer group appears as overrepresented in the top candidates, congruent with its recent use in organic electronics.[71] Finally, we moved onto the terminal fragments to determine how the molecules should be capped. Figure 5 shows the distribution of Z scores for the terminal fragments.

As with the spacers, the best terminal fragments are weakly electron-donating aryl groups. This extended π-system decreases HOMO-LUMO gaps, which results in higher J_{SC} and PCE values. Carbonyl-based and electron-poor thiophene terminal groups strongly deactivate the activities of terminal substituents. Bulky terminal groups (i.e., indoles) cause substantial twisting along the backbone, which decreases π conjugation. The main design principles that can be inferred from this analysis are as follows: use an electron-deficient central π-conjugated molecular fragment and build outward converging toward an ambiphilic or electron-donating fragment. This will result in materials with small HOMO-LUMO gaps and substantial V_{OC}.

Fingerprint Statistics

Alternatively, 2D structural insights can be gained by looking at statistical patterns found in the binary vectors computed from the Morgan fingerprints. Each position within the vector (bit) encodes a particular graph-connectivity environment; activation (1) of a bit indicates a particular pattern is present in a molecule. We used 8,192 bit vectors to avoid hashing collisions; on average 94 bits are activated in each molecule, the SD is of 20 bits, giving confidence that this representation captures a wide array of patterns that are not focused on the initial library fragments, which might provide insights for construction of novel fragments for future libraries and can be used in predictive modeling. Using a 4-bond radius fingerprint includes the connectivity information for a distance of four bonds from any particular atom. For example, bit number 5476 encodes the presence of a carbon atom, a bit that is activated in all molecules. While the least common bit, a 4-bond radius bit with number 1174 is only activated in 8 molecules. Utilizing hypergeometric Z scoring
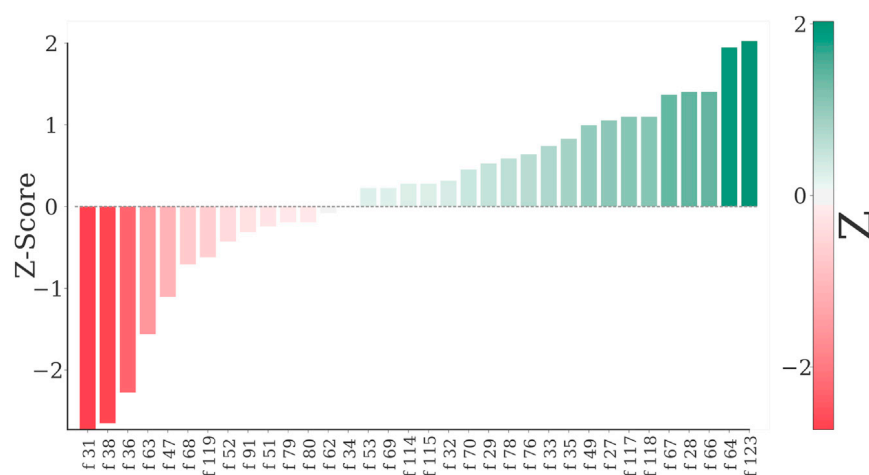
Figure 5. Barplots of Z Scores of the Terminal Fragments Pertaining to the Top Candidates with PCEs > 8.0%

The overrepresented (positive) Z scores are shown in green and the underrepresented (negative) Z scores are shown in red. Fragments not shown are not contained in top candidate molecules.

with bit vectors, we can quantify if certain molecular environments are more likely to be present in molecules with higher predicted PCE. In the Supplemental Information (see Figures S6–S9), we show plots for activated bits for a representative sample of all bits, along with Z scores and boxplots of each bit. A complete analysis of these features is left for future work.

Selected Top Candidates

In this section, we focused on symmetric molecules with PCEs >9.0% for four categories of non-fullerene acceptor materials: quinoidal thiophenes, dicyano-cyclopentyldienyl derivatives, and DPPs. Figure 6 shows these non-fullerene acceptors and computed photophysical properties, including adiabatic excitation energy and oscillator strength.

The high maximum PCE values reflect the favorable electronic properties of materials for PCDTBT-containing organic photovoltaic devices. These materials are predicted to indeed have substantially improved light absorbance in the solar spectrum, while maintaining similar or lower LUMO energies as PCBM. These molecules feature calibrated HOMO-LUMO gaps and computed $S_0 \rightarrow S_1$ transitions in the range of visible light (1.4–2.2 eV). The oscillator strength indicates the likelihood of electronic transition, and this value varies considerably based on the core fragment.

The high oscillator strengths shown for the two DPP derivatives in Figure 6 are consistent with literature reports of visible-light absorption of DPP derivatives.[72] Quinoidal thiophene derivatives have relatively high oscillator strengths (*f* is generally greater than 2.0) and experimental reports involving quinoidal thiophenes are not as common as NDI or DPP non-fullerene acceptors. However, this class of molecules appears to be quite promising because of their improved absorbance in the visible range of the solar spectrum, relatively straightforward synthetic routes, and PCEs reaching 5.2% in TiO_2 dye-sensitized solar cells.[67] These results support our earlier analysis that quinoidal thiophene core fragments should be explored further for hybrid organic-inorganic photovoltaics and OPVs because of their predicted strong electronic absorbance properties. Dicyanocylopentadienyl derivatives
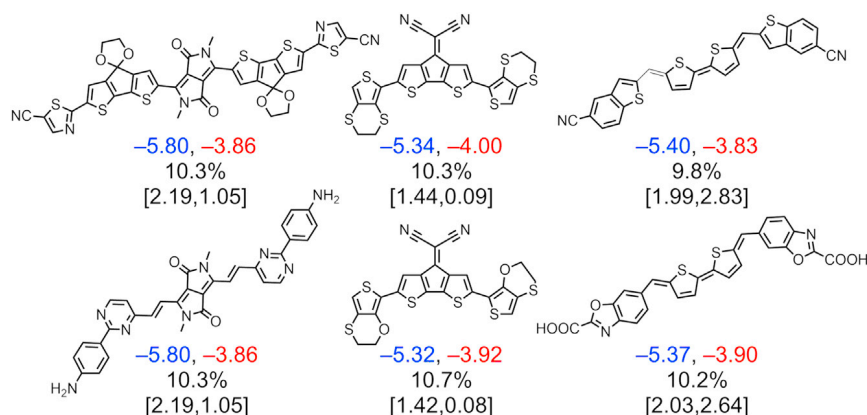
Figure 6. A Selection of Top Molecules from the Quinoidal Thiophene, Dicyanocylopentadienyl, and DPP Fragments with Predicted PCEs > 9%

Under each molecule, we report three lines of calibrated and computed values. The calibrated HOMO (blue) and LUMO (red) energies are given in the first row and are reported in eV. The second row shows the PCE based on calibrated HOMO and LUMO energies. The third row gives the computed $S_0 \rightarrow S_1$ transition energy (reported in eV) and oscillator strength computed using TD-CAM-B3LYP[74]-(D3BJ)[75,76]-IEFPCM(ε = 4)/6-31G(d).

show *f* less than 0.10 and will likely not be utilized as an efficient chromophore in fullerene-free OPVs.

Conclusion

This work represents an important step toward navigating the chemical space of non-fullerene acceptor materials for organic photovoltaics. The Harvard Clean Energy Project in conjunction with the IBM World Community Grid has been used to efficiently screen over 51,000 molecules to identify new classes and fragment ligation patterns. In general, the central fragment should be electron deficient and feature outward-building fragments with increasing electron-donating ability. Additional TD-DFT calculations on a selected subset of top non-fullerene acceptor candidates has increased our understanding of their absorption properties. DPP and quinoidal thiophene derivatives should be explored further because of their direct improvement of absorption properties over fullerenes. Future efforts should focus on improving correlation via the incorporation of more molecular features[73] that relate better to the microscopic and mesoscopic features of OPV devices. We are actively working with experimental collaborators to choose targets for synthesis and device fabrication. We welcome inquiries for more lead candidates from groups around the world.

EXPERIMENTAL PROCEDURES

Experimental Procedures are detailed in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes nine figures, two tables, and five data files and can be found with this article online at https://doi.org/10.1016/j.joule.2017.10.006.

AUTHOR CONTRIBUTION

S.A.L. determined the molecular fragments used in the screening set and performed the quantum mechanical calculations on the test set. B.S.L. performed the calibration and statistical analysis. Both S.A.L. and B.S.L. contributed equally to the writing

REFERENCES

1. U.S. Energy Information Administration. (2016). International Energy Outlook. pp. 1–276. https://www.eia.gov/outlooks/ieo/pdf/0484(2016).pdf.

2. Krebs, F.C., Espinosa, N., Hösel, M., Søndergaard, R.R., and Jørgensen, M. (2014). 25th anniversary article: rise to power — OPV-based solar parks. Adv. Mater. 26, 29–39.

3. Kojima, A., Teshima, K., Shirai, Y., and Miyasaka, T. (2009). Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. J. Am. Chem. Soc. 131, 6050–6051.

4. Lee, M.M., Teuscher, J., Miyasaka, T., Murakami, T.N., and Snaith, H.J. (2012). Efficient hybridge solar cells based on meso-superstructured organometal halide perovskites. Science 38, 643–647.

5. Saparov, B., and Mitzi, D.B. (2016). Organic-inorganic perovskites: structural versatility for functional materials design. Chem. Rev. 116, 4558–4596.

6. Sutherland, B.R., and Sargent, E.H. (2016). Perovskite photonic sources. Nat. Photon. 10, 295–302.

7. Ergen, O., Gilbert, S.M., Pham, T., Turner, S.J., Tan, M.T.Z., Worsley, M.A., and Zetti, A. (2016). Graded bandgap perovskite solar cells. Nat. Mater. 16, 522–525.

8. Mazzio, K.A., and Luscombe, C.K. (2015). The future of organic photovoltaics. Chem. Soc. Rev. 44, 78–90.

9. Lu, L., Zheng, T., Wu, Q., Schneider, A.M., Zhao, D., and Yu, L. (2015). Recent advances in bulk heterojunction polymer solar cells. Chem. Rev. 115, 12666–12731.

10. Zhao, W., Li, S., Yao, H., Zhang, S., Zhang, Y., Yang, B., and Hou, J. (2017). Molecular optimization enables over 13% efficiency in organic solar cells. J. Am. Chem. Soc. 139, 7148–7151.

11. Huang, H., Yang, L., and Sharma, B. (2017). Recent advances in organic ternary solar cells. J. Mater. Chem. A 139, 7148–7151.

12. Brabec, C.J., and Durrant, J.R. (2008). Solution-processed organic solar cells. MRS Bull. 33, 670–675.

13. Helgesen, M., Sondergaard, R., and Krebs, F.C. (2010). Advanced materials and processes for polymer solar cell devices. J. Mater. Chem. 20, 36–60.

14. Krebs, F.C. (2009). Fabrication and processing of polymer solar cells: a review of printing and coating techniques. Sol. Energ. Mat. Sol. Cells 93, 394–412.

15. Yang, X., Loos, J., Veenstra, S.C., Verhees, W.J.H., Wienk, M.M., Kroon, J.M., Michels, M.A.J., and Janssen, R.A. (2005). Nanoscale morphology of high-performance polymer solar cells. J. Nano Lett. 5, 579–583.

16. Li, G., Yao, Y., Yang, H., Shrotriya, V., Yang, G., and Yang, Y. (2007). Solvent-annealing effect in polymer solar cells based on poly(3-hexylthiophene) and methanofullerenes. Adv. Funct. Mater. 17, 1636–1644.

17. Peet, J., Kim, J.Y., Coates, N.E., Ma, W.L., Moses, D., Heeger, A.J., and Bazan, G.C. (2007). Efficiency enhancement in low-bandgap polymer solar cells by processing with alkane diols. Nat. Mater. 6, 497–500.

18. Rogers, J.T., Schmidt, K., Toney, M.F., Kramer, E.J., and Bazan, G.C. (2011). Structural order in bulk heterojunction films prepared with solvent additives. Adv. Mater. 23, 2284–2288.

19. Holliday, S., Li, Y., and Luscombe, C.K. (2017). Recent advances in high performance donor-acceptor polymers for organic photovoltaics. Prog. Polym. Sci. 70, 34–51.

20. Hummelen, J.C., Knight, B.W., Lepeq, F., Wudl, F., Yao, J., and Wilkins, C.L. (1995). Preparation and characterization of fulleroid and methanofullerene derivatives. J. Org. Chem. 60, 532–538.

21. Mishra, A., and Bauerle, P. (2012). Small molecule organic semiconductors on the move: promises for future solar energy technology. Angew. Chem. Int. Ed. 51, 2020–2067.

22. He, Y.J., and Li, Y.F. (2011). Fullerene derivative acceptors for high performance polymer solar cells. Phys. Chem. Chem. Phys. 13, 1970–1983.

23. Wang, H., He, Y., Li, Y., and Su, H. (2012). Photophysical and electronic properties of five PCBM-like C60 derivatives: spectral and quantum chemical view. J. Phys. Chem. A 116, 255–262.

24. Nielsen, C.B., Holliday, S., Chen, H.-Y., Cryer, S.J., and McCulloch, I. (2015). Non-fullerene electron acceptors for use in organic solar cells. Acc. Chem. Res. 48, 2803–2812.

25. Ling, M.-M., Erk, P., Gomez, M., Koenemann, M., Locklin, J., and Bao, Z. (2007). Air-stable n-channel organic semiconductors based on perylene diimide derivatives without strong electron withdrawing groups. Adv. Mater. 19, 1123–1127.

26. Hendsbee, A.D., Sun, J.-P., Law, W.K., Yan, H., Hill, I.G., Spasyuk, D.M., and Welch, G.C. (2016). synthesis, self-assembly, and solar cell performance of n-annulated perylene diimide non-fullerene acceptors. Chem. Mater. 28, 7098–7109.

27. McAfee, S.M., Topple, J.M., Hill, I.G., and Welch, G.C. (2015). Key components to the recent performance increases of solution

processed non-fullerene small molecule acceptors. J. Mater. Chem. A 3, 16393–16408.

28. Chen, W., Yang, X., Long, G., Wan, X., Chen, Y., and Zhang, Q. (2015). A perylene diimide (PDI)-based small molecule with tetrahedral configuration as a non-fullerene acceptor for organic solar cells. J. Mater. Chem. C 3, 4698–4705.

29. Li, H., Kim, F.S., Ren, G., Hollenbeck, E.C., Subramaniyan, S., and Jenekhe, S.A. (2013). Tetraazabenzodifluoranthene diimides: building blocks for solution-processable n-type organic semiconductors. Angew. Chem. Int. Ed. 52, 5513–5517.

30. Li, W., Hendriks, K.H., Wienk, M.M., and Janssen, R.A. (2016). Diketopyrrolopyrrole polymers for organic solar cells. Acc. Chem. Res. 49, 78–86.

31. Falzon, M.-F., Zoombelt, A.P., Wienk, M.M., and Janssen, R.A.J. (2011). Diketopyrrolopyrrole-based acceptor polymers for photovoltaic applications. Phys. Chem. Chem. Phys. 13, 8931–8939.

32. Shin, R.Y.C., Kietzke, T., Sudhakar, S., Dodabalapur, A., Chen, Z.-K., and Sellinger, A. (2007). N-type conjugated materials based on 2-vinyl-4,5-dicyanoimidazoles and their use in solar cells. Chem. Mater. 19, 1892–1894.

33. Kietzke, T., Shin, R.Y.C., Egbe, D.A.M., Chen, Z.-K., and Sellinger, A. (2007). Effect of annealing on the characteristics of organic solar cells: polymer blends with 2-vinyl-4,5-dicyanoimidazole derivative. Macromolecules 40, 4424–4428.

34. Ooi, Z.E., Tam, T.L., Shin, R.Y.C., Kietzke, T., Sellinger, A., Baumgarten, M., Mullen, K., and deMello, J.C. (2008). Solution processable bulk-heterojunction solar cells using a small molecule acceptor. J. Mater. Chem. 18, 4619–4622.

35. Schubert, M., Yin, C.H., Castellani, M., Bange, S., Tam, T.L., Sellinger, A., Horhold, H.H., Kietzke, T., and Neher, D. (2009). Heterojunction topology vs. fill factor correlations in novel hybrid small-molecular/polymeric solar cells. J. Chem. Phys. 130, 094703.

36. Shin, R.Y.C., Sonar, P., Siew, P.S., Chen, Z.-K., and Sellinger, A. (2009). Electron-accepting conjugated materials based on 2-vinyl-4,5-dicyanoimidazoles for application in organic electronics. J. Org. Chem. 74, 3293–3298.

37. Kuzmich, A., Padula, D., Ma, H., and Triosi, A. (2017). Trends in the electronic and geometric structure of non-fullerene based acceptors for organic solar cells. Energy Environ. Sci. https://doi.org/10.1039/c6ee03654f.

38. Lin, Y., and Zhan, X. (2014). Non-fullerene acceptors for organic photovoltaics. Mater. Horiz. 1, 470–488.

39. World Community Grid. http://www.worldcommunitygrid.org/.

40. Hachmann, J., Olivares-Amaya, R., Jinich, A., Appleton, A.L., Blood-Forsythe, M.A., Seress, L.R., Román-Salgado, C., Trepte, K., Atahan-Evrenk, S., Er, S., et al. (2014). Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the

harvard clean energy project. Energy Environ. Sci. 7, 698–704.

41. Hachmann, J., Olivares-Amaya, R., Atahan-Evernk, S., Amador-Bedolla, C., Sanchez-Carrera, R.S., Gold-Parker, A., Vogt, L., Brockway, A.M., and Aspuru-Guzik, A. (2011). The Harvard Clean Energy Project: large-scale computational screening and design of organic photovoltaics on the world community grid. J. Phys. Chem. Lett. 2, 2241–2251.

42. Olivares–Amaya, R., Amador-Bedolla, C., Hachmann, J., Atahan-Evernk, S., Sanchez-Carrera, R.S., Vogt, L., and Aspuru-Guzik, A. (2011). Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics. Energy Environ. Sci. 3, 4849–4861.

43. Kanal, I.Y., Owens, S.G., Bechtel, J.S., and Hutchinson, G.R. (2014). Efficient computational screening of organic polymer photovoltaics. J. Phys. Chem. Lett. 4, 1613–1623.

44. Gagorik, A.G., Mohin, J.W., Kowalewski, T., and Hutchinson, G.R. (2015). Effects of delocalized charge carriers in organic solar cells: predicting nanoscale device performance from morphology. Adv. Funct. Mater. 25, 1996–2003.

45. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36.

46. Landrum, G. RDKit: Open-source cheminformatics. http://www.rdkit.org.

47. Halgren, T.A. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J. Comput. Chem. 17, 490–519.

48. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: an open chemical toolbox. J. Cheminform. 3, 33.

49. Perdew, J.P. (1986). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys. Rev. B 33, 8822–8824.

50. Becke, A.D. (1988). Density-functional exchange energy approximation with correct asymptotic behavior. Phys. Rev. A 38, 3098–3100.

51. Weigend, F., and Ahlrichs, R. (2005). 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, 3297–3305.

52. Seeger, M. (2004). Gaussian processes for machine learning. Int. J. Neural Syst. 14, 69–106.

53. Lopez, S.A., Pyzer-Knapp, E.O., Simm, G.N., Lutzow, T., Li, K., Seress, L.R., Hachmann, J., and Aspuru-Guzik, A. (2016). The Harvard organic photovoltaic dataset. Sci. Data 3, 160086, Additional references can be found in the supporting information.

54. Matthews, A.G.G., van der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., Leon-Villagra, P., Ghahramani, Z., and Hensman, J. (2016).

GPflow: a Gaussian process library using TensorFlow. arXiv, arXiv:1610.08733.

55. Pyzer-Knapp, E.O., Simm, G.N., and Aspuru-Guzik, A. (2016). A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials. Mater. Horiz. 3, 226–233.

56. Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. J. Chem. Inf. Model. 50, 742–754.

57. Snoek, J., Larochelle, H., and Adams, R.P. (2016). Practical Bayesian optimization of machine learning algorithms. machine learning, learning. arXiv, arXiv:1206.2944.

58. Rasmussen, C.E., and Williams, C.K.I. (2016). Gaussian Processes for Machine Learning 2006, Sec 5.2 (MIT Press), pp. 108–111.

59. Scharber, M.C., Muhlbacher, D., Koppe, M., Denk, P., Heeger, A.J., Waldauf, C., and Brabec, C.J. (2006). Design rules for donors in bulk-heterojunction solar cells–towards 10% energy-conversion efficiency. Adv. Mater. 18, 789–794.

60. Ameri, T., Dennler, G., Lungenschmied, C., and Brabec, C.J. (2009). Organic tandem solar cells: a review. Energy Environ. Sci. 2, 347–363.

61. Shockley, W., and Queisser, H.J. (1961). Detailed balance limit of efficiency of p-n junction solar cells. J. Appl. Phys. 32, 510–519.

62. Blouin, N., Michaud, A., and Leclerc, M. (2007). A low-bandgap poly(2,7-carbazole) derivative for use in high-performance solar cells. Adv. Mater. 19, 2295–2300.

63. Brédas, J.L., Beljonne, D., Coropceanu, V., and Cornil, J. (2004). Charge-transfer and energy-transfer process in π–conjugated oligomers and polymers: a molecular picture. Chem. Rev. 104, 4971–5003.

64. Dang, M.T., Hirsch, L., and Wantz, G. (2011). P3HT: PCBM, best seller in polymer photovoltaic research. Adv. Mater. 23, 3597–3602.

65. Luber, E.J., and Buriak, J.M. (2013). Reporting performance in organic photovoltaic devices. ACS Nano 7, 4708–4714.

66. Rice, J.A. (2006). Mathematical Statistics and Data Analysis, Third Edition (Duxbury Press).

67. Komatsu, M., Nakazaki, J., Uchida, S., Jubo, T., and Segawa, H. (2013). A donor–acceptor type organic dye connected with a quinoidal thiophene for dye-sensitized solar cells. Phys. Chem. Chem. Phys. 15, 3227–3232.

68. Anthony, J.E., Facchetti, A., Heeny, S.R., Marder, S.R., and Zhan, X. (2010). N-type organic semiconductors in organic electronics. Adv. Mater. 22, 3876–3892.

69. Lin, Y.Z., Fan, H.J., Li, Y.F., and Zhan, X.W. (2012). Thiazole-based organic semiconductors for organic electronics. Adv. Mater. 24, 3087–3106.

70. Wong, K.-T., Chao, T.-C., Liang-Chen, C., Chu, Y.-Y., Balaiah, A., Chiu, S.-F., Liu, Y.-H., and Wang, Y. (2006). Synthesis and structures of novel heteroarene-fused coplanar π-conjugated chromophores. Org. Lett. 8, 5033–5036.

71. Wang, E., Yao, Z., Zhang, Y., Shao, G., Zhang, M., and Wang, P. (2016). Significant influences of elaborately modulating electron donors on

light absorption and multichannel charge-transfer dynamics for 4-(benzo[c][1,2,5]thisdiazol-4-ylethynyl)benzoic acid dyes. ACS Appl. Mat. Inter. *8*, 18292–18300.

72. Vezie, M.S., Few, S., Meager, I., Pieridou, G., Dörling, B., Ashraf, R.S., Goñi, A.R., Bronstein, H., McCulloch, I., Hayes, S.C., et al. (2016). Exploring the origin of high optical absorption in conjugated polymers. Nat. Mater. *15*, 746–753.

73. Alharbi, F.H., Rashkeev, S.N., El-Mellouhi, F., Luthi, H.P., Tabet, N., and Kais, S. (2015). An efficient descriptor model for designing materials for solar cells. Nat. Comput. Mater. *1*, 15003.

74. Yanai, T., Tew, D.P., and Handy, N.C. (2004). A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). Chem. Phys. Lett. *393*, 51–57.

75. Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. J. Comput. Chem. *32*, 1456–1465.

76. Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. *132*, 154104.