THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2018, 9, 2639−2646

pubs.acs.org/JPCL

# Computer-Aided Screening of Conjugated Polymers for Organic Solar Cell: Classification by Random Forest

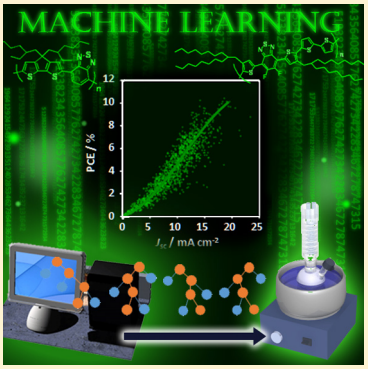
Shinji Nagasawa,[†] Eman Al-Naamani,[†] and Akinori Saeki*[,†,‡]

[†]Department of Applied Chemistry, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[‡]Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

Supporting Information

**ABSTRACT:** Owing to the diverse chemical structures, organic photovoltaic (OPV) applications with a bulk heterojunction framework have greatly evolved over the last two decades, which has produced numerous organic semiconductors exhibiting improved power conversion efficiencies (PCEs). Despite the recent fast progress in materials informatics and data science, data-driven molecular design of OPV materials remains challenging. We report a screening of conjugated molecules for polymer−fullerene OPV applications by supervised learning methods (artificial neural network (ANN) and random forest (RF)). Approximately 1000 experimental parameters including PCE, molecular weight, and electronic properties are manually collected from the literature and subjected to machine learning with digitized chemical structures. Contrary to the low correlation coefficient in ANN, RF yields an acceptable accuracy, which is twice that of random classification. We demonstrate the application of RF screening for the design, synthesis, and characterization of a conjugated polymer, which facilitates a rapid development of optoelectronic materials.




Materials informatics (MI) and cheminformatics are data-driven discovery approaches for a virtual high-throughput screening of a large molecular space.[1−3] Owing to the ever-growing efficiency of computer hardware and novel algorithms in machine learning, MI has been increasingly capturing the attention of a greater scientific community.[4−6] It is highly effective in those fields that exhibit a close relationship between structure and properties, such as drug discovery[7,8] and research on catalysis,[9,10] thermoelectrics,[11,12] and lithium ion batteries.[13,14] In organic electronics, virtual screening combined with density functional theory (DFT) has been successful in thermally activated delayed fluorescence[15,16] and in molecule-based field-effect transistors (FETs).[17] Combinatorial exploration of organic or inorganic metal halide perovskite solar cells[18−22] as well as analysis using data science techniques[23−26] are further notable examples.

Despite this progress, the MI-based development of organic photovoltaic (OPV) materials remains challenging,[27−30] as an OPV device consists of a p-type polymer (or molecule) and an n-type fullerene (or nonfullerene molecule), which forms a bicontinuous network of bulk heterojunctions (BHJs) as a result of complex factors, including solubility,[31,32] solvent additives,[33,34] crystallinity (local ordering),[35,36] and molecular orientation.[37,38] In particular, the charge separation at the p−n interface is largely affected by the electronic and morphological properties of molecular aggregates.[39,40] These properties include various mechanisms, for instance, coherent electron transfer to the fullerene crystallites,[41,42] entropic dissipation,[43,44] and spatiotemporal dynamics of charges,[45,46] which are difficult to predict from calculations involving a single molecule. Although chemists frequently use DFT calculations prior to the synthesis to estimate the bandgap ($E_g$), highest occupied molecular orbital (HOMO), and lowest unoccupied molecular orbital (LUMO), numerous trial-and-error experiments are required for the fine-tuning of the $\pi$-conjugated backbone and optimal alkyl chains,[47,48] where a universal approach to chemical design is yet to be established.[49,50] Device-less analysis of OPV materials using photoconductivity,[51] transient photoabsorption,[52] and photoemission[53] supports the screening of materials and processing conditions; however, they continue to require realistic materials.

A virtual survey of OPV molecules has been vigorously performed in the Harvard Clean Energy Project (CEP), which established an automated, high-throughput in silico framework using the World Community Grid by IBM.[54,55] The CEP provides online access to a few millions of molecular choices for OPV applications, where the power conversion efficiency (PCE) (0.01−10.7%), short-circuit current density ($J_{sc}$), open-circuit voltage ($V_{oc}$), and fill factor (FF), together with the HOMO/LUMO and $E_g$, are appended to each chemical structure.[56] It should be noted that these parameters are based on DFT and first-principles calculations.[54,57,58] Nevertheless, bridging the gap between numerical results and the actual device performance is a critical issue due to the above-

Received: February 28, 2018
Accepted: May 2, 2018
Published: May 7, 2018






DOI: 10.1021/acs.jpclett.8b00635
J. Phys. Chem. Lett. 2018, 9, 2639−2646

mentioned complexity of the photoelectric conversion in a BHJ film. Although, until now, thousands of polymers have been developed worldwide, machine learning, based on the experimental results, has rarely been performed, as these nondigitized data, including the chemical structure, are published in a diverse range of journals from several publishers.[59−66]

In this work, we perform supervised machine learning using artificial neural network (ANN) and random forest (RF) methods to find potential choices for a polymer−fullerene ([6,6]-phenyl-$C_{71}$-butyric acid methyl ester ($PC_{71}BM$) or [6,6]-phenyl-$C_{61}$-butyric acid methyl ester (PCBM)) OPV devices. Our models are constructed on the basis of approximately 1000 experimental parameters ($E_g$, HOMO, weight-averaged molecular weight ($M_w$), and device PCE) and relevant chemical structures obtained from the literature. A possible approach to the design and characterization of a new polymer is to combine machine learning, manual considerations, and experimental analysis.

Figure 1 shows a statistics of device parameters (~1200 points) collected from the literature (~500 papers). The data are provided in the Supporting Information, together with the histograms of journal title and publication year (Figure S1). Identical polymers with different $M_w$ values and side alkyl chains are included to consider their effects. The PCE is defined by $J_{sc}\cdot V_{oc}\cdot FF\cdot P_{in}^{-1}$, where $P_{in}$ is the incident illumination power. Although, from a mathematical point of view, $J_{sc}$, $V_{oc}$, and FF equally affect the PCE, $J_{sc}$ (0.1−24 mA $cm^{-2}$) is found to have the most crucial effect on the PCE (Figure 1a). The interpolated line suggests a sublinear dependence in the high $J_{sc}$ region, which is probably due to the trade-off relationship between $J_{sc}$ and $V_{oc}$. For example, the polymer with the highest $J_{sc}$ of 23.5 mA $cm^{-2}$ had a low $V_{oc}$ of 0.57 V, resulting in a PCE of 8.8%. The observed dominant contribution of $J_{sc}$ to the PCE is similar to that in the report by Jackson et al., who analyzed the metadata of ~150 polymer−fullerene OPV materials.[67] The FF (0.25−0.76) exhibits a mostly linear dependence on the PCE, while its correlation is less than that of $J_{sc}$ (Figure 1b). Indeed, it has been reported that $J_{sc}$ and the FF are also in a trade-off relationship through the film thickness, where face-on orientated polymers enable the thick active layer (~300 nm) to increase the $J_{sc}$ while maintaining a high FF.[37] The film thickness is a crucial parameter for the PCE; however, during the data collection, we noticed that the optimal thickness has a small variation (70−200 nm, typically ~100 nm). Similarly, the processing conditions including solvent (e.g., chlorobenzene and *o*-dichlorobenzene), solvent additive (e.g., 1,8-diiodooctane (DIO) and 1-chloronaphthalene (CN)), and p/n blend ratio (typically 1:1 to 1:3) are not varied so much. Therefore, we ignored these conditions in machine learning. On one hand, the dynamic range of the FF is just a factor of three, while that of $J_{sc}$ is two orders of magnitude. Accordingly, $J_{sc}$ is more dominant on the PCE than the FF. In contrast, the plot of the PCE and $V_{oc}$ (0.2−1.08 V) is rather scattered, where most of the data are lying in the range of 0.6 to 0.9 V (Figure 1c). Therefore, a high $V_{oc}$ is a necessary but not sufficient condition to achieve a high PCE.

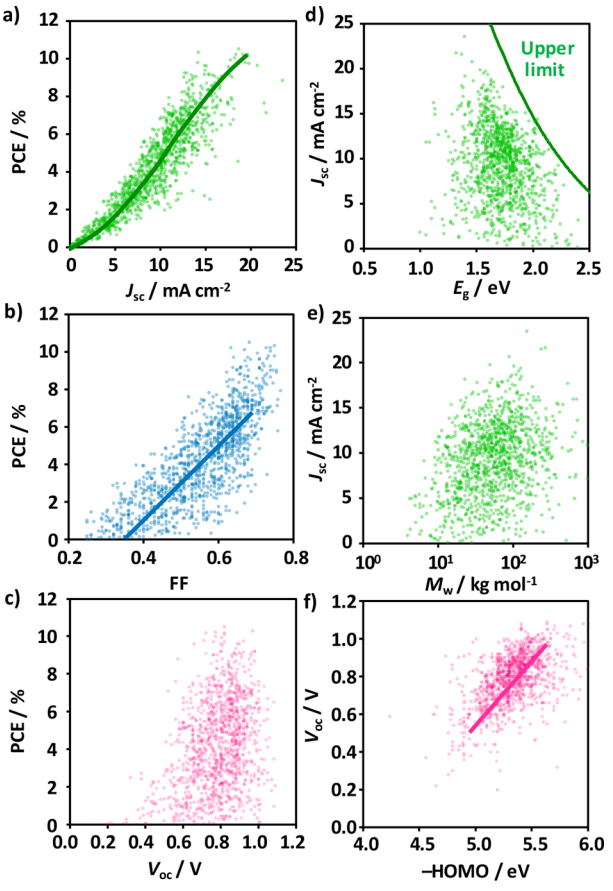



**Figure 1.** Statistics of experimental device parameters (~1200 points) collected from the literature. (See the Supporting Information.) (a) PCE versus $J_{sc}$, (b) PCE versus FF, (c) PCE versus $V_{oc}$, (d) $J_{sc}$ versus $E_g$, (e) $J_{sc}$ versus $M_w$, and (f) $V_{oc}$ versus −HOMO. The solid lines in panels a, b, and f are guides to the eye. The solid curve in panel d is the upper limit calculated by integrating the solar spectrum (AM1.5G) above $E_g$, where the internal quantum efficiency (IQE) is assumed to be 100% without loss of reflection and absorption in the glass/electrodes and interference.

To examine the connection between device parameters and material properties, $J_{sc}$ is plotted as a function of $E_g$ of a polymer in Figure 1d. The value of $J_{sc}$ is expected to increase with the decreasing $E_g$, as a narrow $E_g$ can harvest more energy from the sunlight. However, the expected negative dependence of $J_{sc}$ on $E_g$ is not obvious, although the experimental $J_{sc}$ is below the theoretical upper limit calculated by integrating the photon flux above $E_g$ (the solid curve in Figure 1d). We have not distinguished the experimental methods to determine $E_g$ (cyclic voltammetry (CV) or optical spectroscopy) and HOMO/LUMO (CV or photoelectron spectroscopy), which results in a certain fluctuation in the $J_{sc}$−$E_g$ plot.

The $M_w$ is one of the most important properties of a polymer, which generally affects the crystallinity and FET mobility,[68] interconnectivity of crystallites,[69] polymer orientation,[70] and fibrous structure of the polymer domain in a BHJ film.[71] Figure 1e shows an uneven, positive correlation of $J_{sc}$ with the logarithm of $M_w$ (3.9−1520 kg $mol^{-1}$). Increasing $M_w$ of an identical polymer backbone is a straightforward approach to improve the PCE;[70,71] however, $M_w$ in the different backbones does not universally govern $J_{sc}$ and the PCE.

As shown in Figure 1f, $V_{oc}$ appears to increase with the deepening HOMO levels of the polymer donors ($E_{HOMO}^{D}$), consistently with the empirical equation of $V_{oc} = e^{-1}(|E_{HOMO}^{D}| - |E_{LUMO}^{A}|) - 0.3$ V, where $e$ is the elementary charge and $E_{LUMO}^{A}$ is the LUMO level of the fullerene acceptor.[72,73] Therefore, simultaneous optimization of both $E_g$ and the


DOI: 10.1021/acs.jpclett.8b00635
J. Phys. Chem. Lett. 2018, 9, 2639−2646

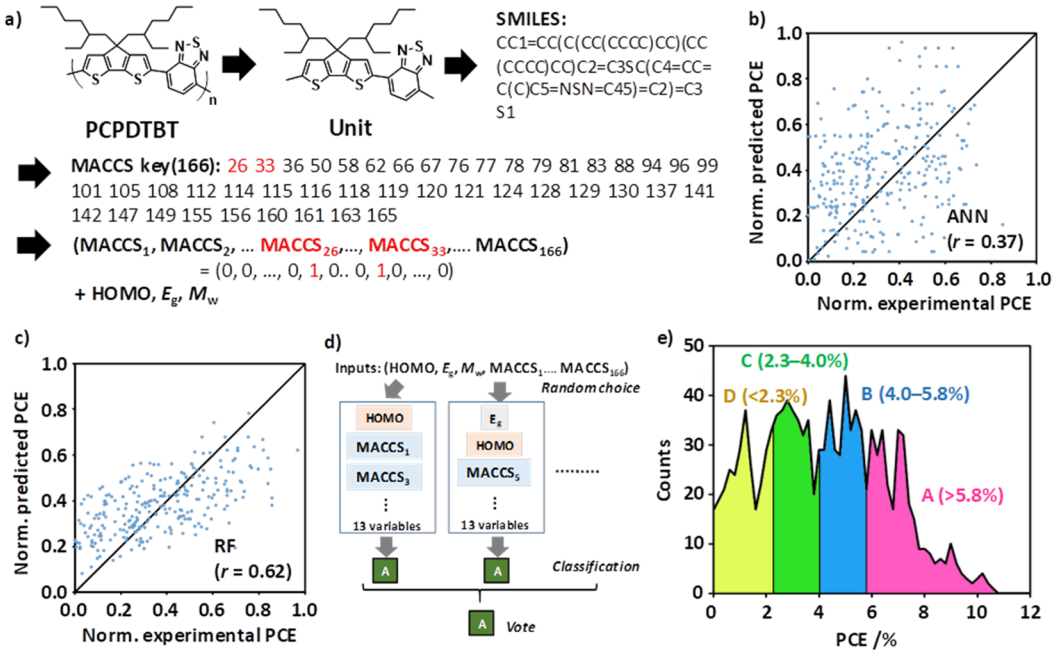



**Figure 2.** (a) Scheme of converting a chemical structure to digitized data. (b) Results of ANN and (c) results of RF, where the horizontal and vertical axes represent the normalized experimental PCE and predicted PCE, respectively, and $r$ is the correlation coefficient. The diagonal black line indicates the perfect positive correlation ($r$ = 1). (d) Scheme of a classification using RF. (e) Histogram of the collected experimental PCE data (~1200) and classification of $n$ = 4. For example, label "A" corresponds to the highest PCE group.

HOMO level with a sufficient LUMO−LUMO offset of donor and acceptor is a reasonable method of molecular engineering, where DFT calculation can intuitively support the design. However, the poor correlation between $J_{sc}$ and $E_g$ greatly disturbs the direct relationship between structure and properties, which suggests the need for data science in addition to computer science.

To transform chemical structures compatible with machine learning, we converted the repeating unit of a polymer to a simplified molecular input line entry system (SMILES) code and then to a molecular access system (MACCS)[74] or an extended connectivity fingerprint (ECFP6) key[75] using the RStudio software[76] with the RDKit toolkit (Figure 2a). MACCS is a primitive 2D fingerprint, which counts the presence of chemical fragments, such as S−N and alkaline metal, providing 166 digital keys. We also considered a larger sized structural fragment dictionary fingerprint of PubChem (881 bit); however, the results were similar to those of the MACCS and thus are not discussed here. In contrast, the ECFP6 is a high-resolution fingerprint, which considers the neighboring connectivity of atoms (1064 bit). Note that regioregularity of polymer backbone is not expressed in these fingerprints and thus is ignored in our model. We applied the digitized keys together with the normalized HOMO level, $E_g$, and $M_w$ to ANN machine learning to predict PCE. After optimizing the number of the hidden layer and nodes, a correlation coefficient ($r$) of 0.37 was obtained (Figure 2b), which is far below the perfect positive correlation (= 1) and insufficient for practical use. In contrast, the RF gave an improved $r$ of 0.62 (Figure 2c). The ANN is a powerful model in image and language recognition, where the data contain multiple dimensions, which cannot be separated linearly. Nevertheless, the RF is known to be robust even in the presence of several explanatory variables, which is similar to the conditions in this work. In addition, an ANN usually requires millions of data entries to construct an accurate model, while the present polymer PCE data have only 1000 data sets. Such a condition (a small number of data sets and a relatively large number of independent input parameters) is supposed to be responsible for the higher correlation coefficient of the RF than that of the ANN.

However, the direct prediction of the PCE using RF is unsatisfactory for the polymer screening. Thus we modified the RF model to provide a rank of the PCE instead of a value (Figure 2d). Regarding this modification, it should be recalled that a problem in our collected metadata is the small number of data in the PCE range of 8−11%, which is still rare in experimental terms (Figure 2e). Accordingly, the histogram of the PCE was divided into $n$ groups, so as to have the same number in each group. The colored area in Figure 2e is an example of a four-group fraction, where the rank "A" is the highest PCE group (>5.8%). The range of the rank for $n$ = 2−10 is listed in Table S2. RF machine learning was performed with changing $n$, where 13 parameters [= square root of the total parameters (166 + 3 = 169)] were randomly chosen and subjected to a decision tree (Figure 2d). On the basis of the classification by hundreds of decision trees, the answer (output) was determined by voting. The RF model was tuned by using 900 training data.

Figure 3a shows the result of the test data (the total number is 62) obtained by using RF with a MACCS key and four-group ($n$ = 4) fraction. The row and column represent the experimental and predicted classifications of the PCE, respectively, such that the diagonal line corresponds to the correct answer. The percentage of correct answer was 30/62 = 48%, which is approximately twice the random classification (1/4 = 25%). Figure 3b shows the dependence of the correct answers obtained on the number of group fractions ($n$ = 2−10). For each $n$, cross-validation (interchange of supervised data and test data) and averaging was applied. Notably, the percentage of


DOI: 10.1021/acs.jpclett.8b00635
J. Phys. Chem. Lett. 2018, 9, 2639−2646

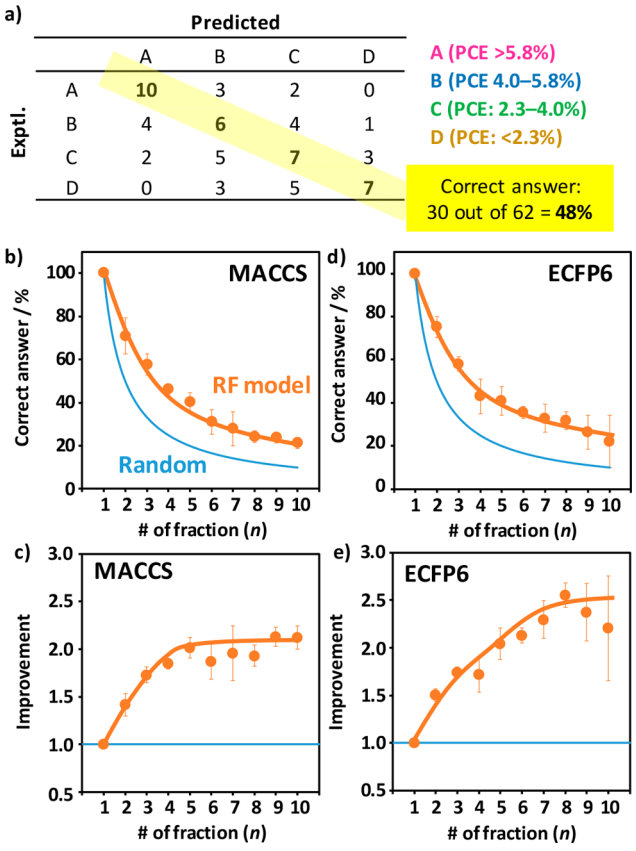



**Figure 3.** (a) Results of classification using the RF model (MACCS) and $n$ = 4. (b,d) Correct answers versus the number of fractions ($n$) obtained by the RF model with (b) MACSS and (d) ECFP6. The solid blue line is the random classification, for example, 100/2 = 50% for $n$ = 2. (c,e) Improvement factor of correct answers of the RF model using (c) MACCS and (e) ECFP6.

correct answers obtained by RF classification (orange circles) is always higher than that of the random classification (blue solid line). The relative improvement of the RF prediction compared to that of the random prediction is shown in Figure 3c. The improvement factor is increased from $n$ = 2 to 4 and mostly saturated at ∼2.1 for a large $n$. Similar results were obtained when the ECFP6 key was used instead of the MACCS key (Figure 3d). The improvement factor was increased to ∼2.5 and saturated above $n$ = 8, indicating that the use of ECFP6 resulted in a more accurate prediction than that of MACCS (Figure 3e). If the test data are identical to the training data, then the percentage of correct answers exceeds 80% for the MACCS key and 95% for the ECFP6 key (Figure S2). Specifically, the RF model gives correct answers with excellent accuracy for the polymers that have already been learned.

Next, we attempted to design and synthesize a new polymer to demonstrate the application of the RF model. The overall scheme is illustrated in Figure 4a. First, data for 1000 molecules were downloaded from the Web site of CEP, including their structures (SMILES) and calculated $E_g$/HOMO/LUMO values.[56] These molecules were sequentially collected from the groups that exhibited a calculated PCE of >10%. Next, the molecules were screened by the RF with a MACCS key ($n$ = 4), where the calculated $E_g$/HOMO, structural fingerprints, and a typical $M_w$ of 50 kg mol$^{-1}$ were used. Consequently, 149 molecules were obtained as the highest PCE group. (The structures are shown in Figure S3.) Finally, one molecule was selected by considering the possibilities of synthesis and π-conjugation (repeatability of alternating single and double bonds) in a polymeric structure. In particular, molecules that have an unsymmetric large heteroacene are removed because of the difficulty of synthesis. The picked-up molecule is an alternating copolymer consisting of phenylene (Ph), thiadiazolo[3,4-*c*]pyridine (TP), and thieno[3,4-*b*]thiophene (TT). Each component is a well known, primitive unit that has been incorporated into high-performing polymers, such as Ph as the donor in Ph-thiophene-fluorobenzothiadiazole copolymer (PPDT2FBT),[77] TP as the acceptor unit of a low-bandgap polymer with controlled regioregularity,[78] and TT as the acceptor unit of benzodithiophene-fluorinated TT copolymer (PTB7).[79] However, their combinations have not been reported, probably due to the small size of the electron-donating Ph unit and the synthetic difficulty in coupling the two acceptor units of TP and TT. By further considering the availability of chemicals and the synthetic route using Stille coupling, the molecule was modified to a copolymer of alkoxy-Ph, TP, alkyl-thieno[3,4-*b*]thiophene-2-carboxylate, and thiophene (T). The length and shape of the alkyl chains (methyl, hexyl, octyl, dodecyl, 2-ethylhexyl, and 2-buthyloctyl) at the alkoxy-Ph and alkyl-TT units were examined by the RF with a ECFP6 key (Figure S4). Eleven combinations of side chains were subjected to the screening, owing the commercial availability of the corresponding brominated monomers ready for polymerization. Three types of side chains were ranked at the highest group (C, 5.0−5.8%) out of the 11 polymers. As a result, the polymer structure with di(*n*-octyloxy)-Ph and 2-ethylhexyl-TT was determined (Figure 4a). Note that the MACCS key cannot distinguish the difference between side chains, while the ECFP6 key allows the screening of side alkyl chains.

The details of the polymer synthesis are provided in the Supporting Information. For a synthetic reason, fluorinated TT was used, which improved the PCE rank from C (5.0−5.8%) to A (>7.0%). The resultant random copolymer P1 (Figure 4b) exhibits the broad photoabsorption spectrum with a peak at 676 nm, $E_g$ of 1.36 eV, HOMO level of −5.03 eV, and LUMO level of −3.67 eV in the film state (Figure 4c and Figure S5). The synthetic yield and $M_w$ were as low as 17% and 1.2 kg mol$^{-1}$, respectively, although we have thoroughly purified the stannyl-thiophene and attempted two batches of synthesis. The difference in reactivity of the brominated monomers or insufficient solubility can be a reason for an unsuccessful synthesis. The latter is confirmed by the gel permeation chromatography (GPC) profile accompanied by an extremely high-molecular-weight peak (∼10$^4$ kg mol$^{-1}$) attributed to the aggregation (Figure S5). The PCE prediction using the experimental $E_g$, HOMO, and $M_w$ was degraded to rank C (5.0−5.7%).

The optoelectronics of the P1/PCBM blend films were analyzed by an Xe-flash-photolysis time-resolved microwave conductivity (Xe-TRMC) technique using a white-light pulse as excitation.[52] The drop-cast film of P1/PCBM 1:2 exhibited the largest photoconductivity maximum ($\Delta\sigma_{max}$), which is 10−16-fold compared with P1 and PCBM pristine films, indicating an efficient charge separation in the BHJ films (Figure S6). The spin-coated film of P1/PCBM 1:2, processed from a chlorobenzene solution with 3 vol % DIO, exhibited a $\Delta\sigma_{max}$ value of 2.7 × 10$^{-9}$ S cm$^{-1}$, corresponding to the PCE $V_{oc}^{-1}$ of 5.4%V$^{-1}$ in the correlation between the Xe-TRMC transient maxima and device efficiencies. The linear interpolation of $V_{oc}$


DOI: 10.1021/acs.jpclett.8b00635
J. Phys. Chem. Lett. 2018, 9, 2639−2646

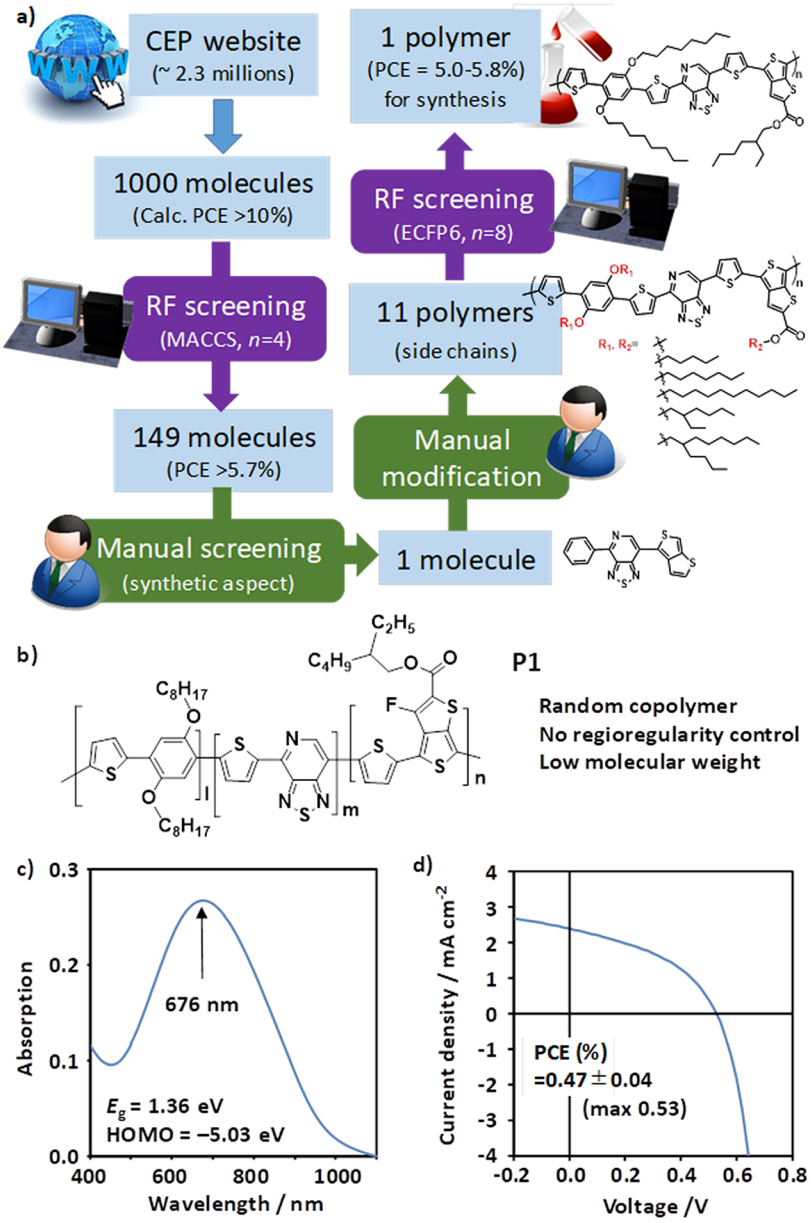



**Figure 4.** (a) Scheme of polymer design by combining the RF screening and manual screening/modification. The picked-up molecule or polymer in each stage is shown. For the details of the structures, refer to the Supporting Information. (b) Synthesized random copolymer for OPV analysis. (c) Photoabsorption spectrum of P1 in the film state. (d) Current density−voltage curve of the best performing device.

versus −HOMO (Figure 1f) suggests a $V_{oc}$ of 0.5 to 0.6 V by considering the HOMO level of P1 (−5.03 eV). Accordingly, the PCE of P1/PCBM is expected to be ∼3.0% from the electrode-less Xe-TRMC. This PCE predicted from the empirical, experimental analysis is lower than that obtained from the RF prediction (5.0−5.7%). This is owing to the random structure, the uncontrolled regioregularity (direction of TP and FTT), and the low $M_w$ of the synthesized polymer. We fabricated an OPV device of ITO/ZnO/(P1/PCBM 1:2)/$MoO_x$/Ag, which resulted in the low PCE of 0.52% (0.47 ± 0.04% averaged over 15 devices) together with $J_{sc}$ = 2.34 mA $cm^{-2}$, $V_{oc}$ = 0.55 V, and FF = 0.40 (Figure 4d and Figure S7 and Table S1). The film indicates a smooth morphology, observed by atomic force microscopy (AFM) and weak X-ray diffraction (XRD), attributed to the interlamellar distance of 19.5 Å (Figure S7). Although the obtained $V_{oc}$ is in the expected range of 0.5 to 0.6 V, the overall PCE is about one-sixth of that expected from the Xe-TRMC due to the low $J_{sc}$ and FF. The processing parameters such as thickness, solvent, and thermal annealing, have not yet been analyzed rigorously because of the small amount of the obtained polymer. However, by considering the RF prediction and Xe-TRMC analysis, we expect that the synthesized random copolymer of P1 is not a high-performing polymer, and thus another polymer needs to be chosen. Without the RF screening, we might synthesize and examine other alky chains for P1; however, this might lead to unnecessary efforts and consume time without satisfactory outcomes. Although the RF screening comprises an uncertain probability, it can support decision making of polymer design, which facilitates a smooth development of new conjugated materials.

Finally, we recall the remaining issues in machine learning of polymer−fullerene OPV. One of them is the lack of specific processing conditions including film thickness, solvent, solvent



DOI: 10.1021/acs.jpclett.8b00635
J. Phys. Chem. Lett. 2018, 9, 2639−2646

additive, p/n blend ratio, and thermal/solvent annealing, although their variations are relatively small compared with $M_w$ and electrochemical properties. Regioregularity (regiorandom or regioregular backbone) is not included in the digitized fingerprints. Other missing factors are purity of a polymer, surface free energy, solubility, polymer orientation, exact shape and intensity of a photoabsorption spectrum (not rectangular below $E_g$), and BHJ morphology associated with miscibility. Inclusion of these factors leads to the reduced number of available data set and increased complexity of input parameters and facile screening of polymers. Automatic morphology analysis of transmittance electron microscope (TEM) and AFM images by using machine learning may improve the prediction accuracy. Another primitive and critical issue is the fact that machine learning is based on the existing data. A further evolution of MI-based machine learning with extrapolation/interpolation of molecular structures with serendipity aspect is anticipated.

In conclusion, we introduced supervised learning methods into the design and synthesis of a conjugated polymer for OPV applications for the first time. ANN and RF models were constructed based on ∼1000 experimental parameters ($E_g$, HOMO, $M_w$, and molecular structure digitized using MACCS or ECFP6 keys). Contrary to the low correlation coefficient in ANN, RF exhibited an improved accuracy, in particular, for the PCE classification. By combining the screening with the RF model, manual considerations, and TRMC analysis, we demonstrated an alternative approach to the design of polymers (backbone and side chains), which is believed to be beneficial for decision making of molecular design.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.8b00635.

Experimental; Synthesis of P1; Summary of OPV device performance (Table S1); Group ranking of PCE used for RF classification (Table S2); Statistics of collected experimental device data (Figure S1); Results of RF modeling obtained for the identical test and training data (Figure S2); Chemical structures of 146 molecules extracted from 1000 molecules by using the RF screening with a MACCS key ($n$ = 4) (Figure S3); Screening of an alkyl side chain by using RF with a ECPF6 key ($n$ = 8) (Figure S4); Optical, electronic, and polymeric data of P1 (Figure S5); Photoconductivity of drop-cast P1/PCBM (1:0 to 0:1) blend films evaluated using Xe-TRMC (Figure S6); Photoconductivity of spun-coat P1/PCBM (1:2) blend film analyzed using Xe-TRMC (Figure S7); and OPV device characterization of P1/PCBM 1:2 (Figure S8). (PDF)

Text data of the collected experimental device parameters (PCE, $J_{sc}$, $V_{oc}$, FF) along with SMILES, nickname, $M_w$, number-averaged molecular weight ($M_n$), polydispersity index (PDI), HOMO, LUMO, $E_g$, molecular weight of a repeating unit, and reference number. (TXT)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: saeki@chem.eng.osaka-u.ac.jp.

### ORCID

Akinori Saeki: 0000-0001-7429-2200


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Prof. Ichigaku Takigawa at Hokkaido University for his advice on fingerprints and Mr. Kenta Aoshima at Osaka University for their support of polymer synthesis. This work was supported by the PRESTO program (grant no. JPMJPR15N6) from the Japan Science and Technology Agency (JST) of Japan and the Japan Society for the Promotion of Science (JSPS), with the KAKENHI Grant-in-Aid for Scientific Research (A) (grant no. JP16H02285).

## ■ REFERENCES


(1) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013**, *12*, 191−201.

(2) Barnard, A. S. In Silico Veritas. *ACS Nano* **2014**, *8*, 6520−6525.

(3) Audus, D. J.; de Pablo, J. J. Polymer Informatics: Opportunities and Challenges. *ACS Macro Lett.* **2017**, *6*, 1078−1082.

(4) Agrawal, A.; Choudhary, A. Perspective: Materials Informatics and Big Data: Realization of the "Fourth Paradigm" of Science in Materials Science. *APL Mater.* **2016**, *4*, 053208.

(5) Hu, J. − M.; Duan, C. − G.; Nan, C. − W.; Chen, L. − Q. Understanding and Designing Magnetoelectric Heterostructures Guided by Computation: Progresses, Remaining Questions, and Perspectives. *npj Computational Materials* **2017**, *3*, 18.

(6) Pulido, A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B. J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; et al. Functional Materials Discovery Using Energy−Structure−Function Maps. *Nature* **2017**, *543*, 657−664.

(7) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci.* **2017**, *3*, 283−293.

(8) Korotcov, A.; Tkachenko, V.; Russo, D. P.; Ekins, S. Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol. Pharmaceutics* **2017**, *14*, 4462−4475.

(9) Nørskov, J. K.; Bligaard, T. The Catalyst Genome. *Angew. Chem., Int. Ed.* **2013**, *52*, 776−777.

(10) Li, N.; Chen, X.; Ong, W. − J.; MacFarlane, D. R.; Zhao, X.; Cheetham, A. K.; Sun, C. Understanding of Electrochemical Mechanisms for $CO_2$ Capture and Conversion into Hydrocarbon Fuels in Transition-Metal Carbides (MXenes). *ACS Nano* **2017**, *11*, 10825−10833.

(11) Madsen, G. K. H. Automated Search for New Thermoelectric Materials: The Case of LiZnSb. *J. Am. Chem. Soc.* **2006**, *128*, 12140−12146.

(12) Mori, T. Novel Principles and Nanostructuring Methods for Enhanced Thermoelectrics. *Small* **2017**, *13*, 1702013.

(13) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22*, 3762−3767.

(14) Jalem, R.; Kimura, M.; Nakayama, M.; Kasuga, T. Informatics-Aided Density Functional Theory Study on the Li Ion Transport of Tavorite-Type $LiMTO_4F$ ($M^{3+}$−$T^{5+}$, $M^{2+}$−$T^{6+}$). *J. Chem. Inf. Model.* **2015**, *55*, 1158−1168.

(15) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D. − G.; Wu, T.; et al. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120−1127.

(16) Penfold, T. J. On Predicting the Excited-State Properties of Thermally Activated Delayed Fluorescence Emitters. *J. Phys. Chem. C* **2015**, *119*, 13535−13544.




DOI: 10.1021/acs.jpclett.8b00635
*J. Phys. Chem. Lett.* 2018, 9, 2639−2646

(17) Sokolov, A. N.; Atahan-Evrenk, S.; Mondal, R.; Akkerman, H. B.; Sánchez-Carrera, R. S.; Granados-Focil, S.; Schrier, J.; Mannsfeld, S. C. B.; Zoombelt, A. P.; Bao, Z.; et al. From Computational Discovery to Experimental Characterization of a High Hole Mobility Organic Crystal. *Nat. Commun.* **2011**, *2*, 437.
(18) Filip, M. R.; Eperon, G. E.; Snaith, H. J.; Giustino, F. Steric Engineering of Metal-Halide Perovskites with Tunable Optical Band Gaps. *Nat. Commun.* **2014**, *5*, 5757.
(19) Kieslich, G.; Sun, S.; Cheetham, A. K. An Extended Tolerance Factor Approach for Organic−Inorganic Perovskites. *Chem. Sci.* **2015**, *6*, 3430−3433.
(20) Filip, M. R.; Giustino, F. Computational Screening of Homovalent Lead Substitution in Organic−Inorganic Halide Perovskites. *J. Phys. Chem. C* **2016**, *120*, 166−173.
(21) Jain, A.; Voznyy, O.; Sargent, E. H. High-Throughput Screening of Lead-Free Perovskite-like Materials for Optoelectronic Applications. *J. Phys. Chem. C* **2017**, *121*, 7183−7187.
(22) Nakajima, T.; Sawada, K. Discovery of Pb-Free Perovskite Solar Cells via High-Throughput Simulation on the K Computer. *J. Phys. Chem. Lett.* **2017**, *8*, 4826−4831.
(23) Persson, N. E.; Rafshoon, J.; Naghshpour, K.; Fast, T.; Chu, P. − H.; McBride, M.; Risteen, B.; Grover, M.; Reichmanis, E. High-Throughput Image Analysis of Fibrillar Materials: A Case Study on Polymer Nanofiber Packing, Alignment, and Defects in Organic Field Effect Transistors. *ACS Appl. Mater. Interfaces* **2017**, *9*, 36090−36102.
(24) Perea, J. D.; Langner, S.; Salvador, M.; Sanchez-Lengeling, B.; Li, N.; Zhang, C.; Jarvas, G.; Kontos, J.; Dallos, A.; Aspuru-Guzik, A.; et al. Introducing a New Potential Figure of Merit for Evaluating Microstructure Stability in Photovoltaic Polymer-Fullerene Blends. *J. Phys. Chem. C* **2017**, *121*, 18153−18161.
(25) Du, X.; Jiao, X.; Rechberger, S.; Perea, J. D.; Meyer, M.; Kazerouni, N.; Spiecker, E.; Ade, H.; Brabec, C. J.; Fink, R. H.; et al. Crystallization of Sensitizers Controls Morphology and Performance in Si-/C-PCPDTBT-Sensitized P3HT:ICBA Ternary Blends. *Macromolecules* **2017**, *50*, 2415−2423.
(26) Ishida, N.; Wakamiya, A.; Saeki, A. Quantifying Hole Transfer Yield from Perovskite to Polymer Layer: Statistical Correlation of Solar Cell Outputs with Kinetic and Energetic Properties. *ACS Photonics* **2016**, *3*, 1678−1688.
(27) Olivares-Amaya, R.; Amador-Bedolla, C.; Hachmann, J.; Atahan-Evrenk, S.; Sánchez-Carrera, R. S.; Vogt, L.; Aspuru-Guzik, A. Accelerated Computational Discovery of High-Performance Materials for Organic Photovoltaics by Means of Cheminformatics. *Energy Environ. Sci.* **2011**, *4*, 4849−4961.
(28) Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. Efficient Computational Screening of Organic Polymer Photovoltaics. *J. Phys. Chem. Lett.* **2013**, *4*, 1613−1623.
(29) Shin, Y.; Liu, J.; Quigley, J. J., IV; Luo, H.; Lin, X. Combinatorial Design of Copolymer Donor Materials for Bulk Heterojunction Solar Cells. *ACS Nano* **2014**, *8*, 6089−6096.
(30) Imamura, Y.; Tashiro, M.; Katouda, M.; Hada, M. Automatic High-Throughput Screening Scheme for Organic Photovoltaics: Estimating the Orbital Energies of Polymers from Oligomers and Evaluating the Photovoltaic Characteristics. *J. Phys. Chem. C* **2017**, *121*, 28275−28286.
(31) Troshin, P. A.; Susarova, K.; Khakina, E. A.; Goryachev, A. A.; Borshchev, O. V.; Ponomarenko, S. A.; Razumov, V. F.; Sariciftci, N. S. Material Solubility and Molecular Compatibility Effects in The Design of Fullerene/Polymer Composites for Organic Bulk Heterojunction Solar Cells. *J. Mater. Chem.* **2012**, *22*, 18433−18441.
(32) Mikie, T.; Saeki, A.; Yamazaki, Y.; Ikuma, N.; Kokubo, K.; Seki, S. Stereochemistry of Spiro-Acetalized [60]Fullerenes: How the Exo and Endo Stereoisomers Influence Organic Solar Cell Performance. *ACS Appl. Mater. Interfaces* **2015**, *7*, 8915−8922.
(33) Peet, J.; Kim, J. Y.; Coates, N. E.; Ma, W. L.; Moses, D.; Heeger, A. J.; Bazan, G. C. Efficiency Enhancement in Low-Bandgap Polymer Solar Cells by Processing with Alkane Dithiols. *Nat. Mater.* **2007**, *6*, 497−500.
(34) Güldal, N. S.; Berlinghof, M.; Kassar, T.; Du, X.; Jiao, X.; Meyer, M.; Ameri, T.; Osvet, A.; Li, N.; Destri, G. L.; Fink, R. H.; Ade, H.; Unruh, T.; Brabec, C. J. Controlling Additive Behavior to Reveal an Alternative Morphology Formation Mechanism in Polymer: Fullerene Bulk-Heterojunctions. *J. Mater. Chem. A* **2016**, *4*, 16136−16147.
(35) Jackson, N. E.; Kohlstedt, K. L.; Savoie, B. M.; Olvera de la Cruz, M.; Schatz, G. C.; Chen, L. X.; Ratner, M. A. Conformational Order in Aggregates of Conjugated Polymers. *J. Am. Chem. Soc.* **2015**, *137*, 6254−6262.
(36) Yoshikawa, S.; Saeki, A.; Saito, M.; Osaka, I.; Seki, S. On the Role of Local Charge Carrier Mobility in the Charge Separation Mechanism of Organic Photovoltaics. *Phys. Chem. Chem. Phys.* **2015**, *17*, 17778−17784.
(37) Osaka, I.; Saito, M.; Koganezawa, T.; Takimiya, K. Thiophene−Thiazolothiazole Copolymers: Significant Impact of Side Chain Composition on Backbone Orientation and Solar Cell Performances. *Adv. Mater.* **2014**, *26*, 331−338.
(38) Shimata, Y.; Ide, M.; Tashiro, M.; Katouda, M.; Imamura, Y.; Saeki, A. Charge Dynamics at Heterojunction between Face-on/Edge-on PCPDTBT and PCBM Bilayer: Interplay of Donor/Acceptor Distance and Local Charge Carrier Mobility. *J. Phys. Chem. C* **2016**, *120*, 17887−17897.
(39) Heeger, A. J. 25th Anniversary Article: Bulk Heterojunction Solar Cells: Understanding the Mechanism of Operation. *Adv. Mater.* **2014**, *26*, 10−28.
(40) Pfannmöller, M.; Kowalsky, W.; Schröder, R. R. Visualizing Physical, Electronic, and Optical Properties of Organic Photovoltaic Cells. *Energy Environ. Sci.* **2013**, *6*, 2871−2891.
(41) Falke, S. M.; Rozzi, C. A.; Brida, D.; Maiuri, M.; Amato, M.; Sommer, E.; De Sio, A.; Rubio, A.; Cerullo, G.; Molinari, E.; Lienau, C. Coherent Ultrafast Charge Transfer in an Organic Photovoltaic Blend. *Science* **2014**, *344*, 1001−1005.
(42) Jamieson, F. C.; Domingo, E. B.; McCarthy-Ward, T.; Heeney, M.; Stingelin, N.; Durrant, J. R. Fullerene Crystallisation as a Key Driver of Charge Separation in Polymer/Fullerene Bulk Heterojunction Solar Cells. *Chem. Sci.* **2012**, *3*, 485−492.
(43) Gregg, B. A. Entropy of Charge Separation in Organic Photovoltaic Cells: The Benefit of Higher Dimensionality. *J. Phys. Chem. Lett.* **2011**, *2*, 3013−3015.
(44) Kobori, Y.; Miura, T. Overcoming Coulombic Traps: Geometry and Electronic Characterizations of Light-Induced Separated Spins at the Bulk Heterojunction Interface. *J. Phys. Chem. Lett.* **2015**, *6*, 113−123.
(45) Vithanage, D. A.; Devižis, A.; Abramavičius, V.; Infahsaeng, Y.; Abramavičius, D.; MacKenzie, R. C. I.; Keivanidis, P. E.; Yartsev, A.; Hertel, D.; Nelson, J.; et al. Visualizing Charge Separation in Bulk Heterojunction Organic Solar Cells. *Nat. Commun.* **2013**, *4*, 2334.
(46) Causá, M.; De Jonghe-Risse, J.; Scarongella, M.; Brauer, J. C.; Buchaca-Domingo, E.; Moser, J.-E.; Stingelin, S.; Banerji, N. The Fate of Electron−Hole Pairs in Polymer:Fullerene Blends for Organic Photovoltaics. *Nat. Commun.* **2016**, *7*, 12556.
(47) Graham, K. R.; Cabanetos, C.; Jahnke, J. P.; Idso, M. N.; El Labban, A.; Ngongang Ndjawa, G. O.; Heumueller, T.; Vandewal, K.; Salleo, A.; Chmelka, B. F.; et al. Importance of the Donor:Fullerene Intermolecular Arrangement for High-Efficiency Organic Photovoltaics. *J. Am. Chem. Soc.* **2014**, *136*, 9608−9618.
(48) Al-Naamani, E.; Gopal, A.; Ide, M.; Osaka, I.; Saeki, A. Exploring Alkyl Chains in Benzobisthiazole-Naphthobisthiadiazole Polymers: Impact on Solar-Cell Performance, Crystalline Structures, and Optoelectronics. *ACS Appl. Mater. Interfaces* **2017**, *9*, 37702−37711.
(49) Mei, J.; Bao, Z. Side Chain Engineering in Solution-Processable Conjugated Polymers. *Chem. Mater.* **2014**, *26*, 604−615.
(50) Yao, H.; Ye, L.; Zhang, H.; Li, S.; Zhang, S.; Hou, J. Molecular Design of Benzodithiophene-Based Organic Photovoltaic Materials. *Chem. Rev.* **2016**, *116*, 7397−7457.
(51) Saeki, A.; Yoshikawa, S.; Tsuji, M.; Koizumi, Y.; Ide, M.; Vijayakumar, C.; Seki, S. A Versatile Approach to Organic Photovoltaics Evaluation Using White Light Pulse and Microwave Conductivity. *J. Am. Chem. Soc.* **2012**, *134*, 19035−19042.


DOI: 10.1021/acs.jpclett.8b00635
*J. Phys. Chem. Lett.* 2018, 9, 2639−2646

(52) Clarke, T. M.; Ballantyne, A.; Shoaee, S.; Soon, Y. W.; Duffy, W.; Heeney, M.; McCulloch, I.; Nelson, J.; Durrant, J. R. Analysis of Charge Photogeneration as a Key Determinant of Photocurrent Density in Polymer: Fullerene Solar Cells. *Adv. Mater.* **2010**, *22*, 5287−5291.
(53) Grancini, G.; Maiuri, M.; Fazzi, D.; Petrozza, A.; Egelhaaf, H.-J.; Brida, D.; Cerullo, G.; Lanzani, G. Hot Exciton Dissociation in Polymer Solar Cells. *Nat. Mater.* **2013**, *12*, 29−33.
(54) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2011**, *2*, 2241−2251.
(55) Pyzer-Knapp, E. O.; Li, K.; Aspuru-Guzik, A. Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery. *Adv. Funct. Mater.* **2015**, *25*, 6495−6502.
(56) *Harvard Clean Energy Project Database*. http://cepdb.molecularspace.org.
(57) Pyzer-Knapp, E. O.; Simm, G. N.; Aspuru-Guzik, A. A Bayesian Approach to Calibrating High-Throughput Virtual Screening Results and Application to Organic Photovoltaic Materials. *Mater. Horiz.* **2016**, *3*, 226−233.
(58) Lopez, S. A.; Pyzer-Knapp, E. O.; Simm, G. N.; Lutzow, T.; Li, K.; Seress, L. R.; Hachmann, J.; Aspuru-Guzik, A. Data Descriptor: The Harvard Organic Photovoltaic Dataset. *Sci. Data* **2016**, *3*, 160086.
(59) Gendron, D.; Leclerc, M. New Conjugated Polymers for Plastic Solar Cells. *Energy Environ. Sci.* **2011**, *4*, 1225−1237.
(60) Duan, C.; Huang, F.; Cao, Y. Recent Development of Push−Pull Conjugated Polymers for Bulk-Heterojunction Photovoltaics: Rational Design and Fine Tailoring of Molecular Structures. *J. Mater. Chem.* **2012**, *22*, 10416−10434.
(61) Nielsen, C. B.; Turbiez, M.; McCulloch, I. Recent Advances in the Development of Semiconducting DPP-Containing Polymers for Transistor Applications. *Adv. Mater.* **2013**, *25*, 1859−1880.
(62) Gao, C.; Wang, L.; Li, X.; Wang, H. Rational Design on D−A Conjugated P(BDT−DTBT) Polymers for Polymer Solar Cells. *Polym. Chem.* **2014**, *5*, 5200−5210.
(63) Wu, J. − S.; Cheng, S. − W.; Cheng, Y. − J.; Hsu, C. − S. Donor−Acceptor Conjugated Polymers Based on Multifused Ladder-Type Arenes for Organic Solar Cells. *Chem. Soc. Rev.* **2015**, *44*, 1113−1154.
(64) Hemavathi, B.; Ahipa, T. N.; Pai, R. K. Polymer Design for Solar Cell − Current Trend and Future Scenario. *Eur. Polym. J.* **2015**, *72*, 309−340.
(65) Yao, H.; Ye, L.; Zhang, H.; Li, S.; Zhang, S.; Hou, J. Molecular Design of Benzodithiophene-Based Organic Photovoltaic Materials. *Chem. Rev.* **2016**, *116*, 7397−7457.
(66) Cai, Y.; Huo, L.; Sun, Y. Recent Advances in Wide-Bandgap Photovoltaic Polymers. *Adv. Mater.* **2017**, *29*, 1605437.
(67) Jackson, N. E.; Savoie, B. M.; Marks, T. J.; Chen, L. X.; Ratner, M. A. The Next Breakthrough for Organic Photovoltaics? *J. Phys. Chem. Lett.* **2015**, *6*, 77−84.
(68) Kline, R. J.; McGehee, M. D.; Kadnikova, E. N.; Liu, J.; Fréchet, J. M. J. Controlling the Field-Effect Mobility of Regioregular Polythiophene by Changing the Molecular Weight. *Adv. Mater.* **2003**, *15*, 1519−1522.
(69) Vakhshouri, K.; Smith, B. H.; Chan, E. P.; Wang, C.; Salleo, A.; Wang, C.; Hexemer, A.; Gomez, E. D. Signatures of Intracrystallite and Intercrystallite Limitations of Charge Transport in Polythiophenes. *Macromolecules* **2016**, *49*, 7359−7369.
(70) Osaka, I.; Saito, M.; Mori, H.; Koganezawa, T.; Takimiya, K. Drastic Change of Molecular Orientation in a Thiazolothiazole Copolymer by Molecular-Weight Control and Blending with $PC_{61}BM$ Leads to High Efficiencies in Solar Cells. *Adv. Mater.* **2012**, *24*, 425−430.
(71) Bartelt, J. A.; Douglas, J. D.; Mateker, W. R.; Labban, A. E.; Tassone, C. J.; Toney, M. F.; Fréchet, J. M. J.; Beaujuge, P. M.; McGehee, M. D. Controlling Solution-Phase Polymer Aggregation with Molecular Weight and Solvent Additives to Optimize Polymer-Fullerene Bulk Heterojunction Solar Cells. *Adv. Energy Mater.* **2014**, *4*, 1301733.
(72) Scharber, M. C.; Mühlbacher, D.; Koppe, M.; Denk, P.; Waldauf, C.; Heeger, A. J.; Brabec, C. J. Design Rules for Donors in Bulk-Heterojunction Solar Cells—Towards 10% Energy-Conversion Efficiency. *Adv. Mater.* **2006**, *18*, 789−794.
(73) Janssen, R. A. J.; Nelson, J. Factors Limiting Device Efficiency in Organic Photovoltaics. *Adv. Mater.* **2013**, *25*, 1847−1858.
(74) *MACCS*. http://rdkit.org/Python_Docs/rdkit.Chem.MACCSkeys-pysrc.html.
(75) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(76) *R studio*. https://www.rstudio.com/.
(77) Nguyen, T. L.; Choi, H.; Ko, S. J.; Uddin, M. A.; Walker, B.; Yum, S.; Jeong, J. E.; Yun, M. H.; Shin, T. J.; Hwang, S.; et al. Semi-Crystalline Photovoltaic Polymers with Efficiency Exceeding 9% in a ∼ 300 nm Thick Conventional Single-Cell Device. *Energy Environ. Sci.* **2014**, *7*, 3040−3051.
(78) Ying, L.; Huang, F.; Bazan, G. C. Regioregular Narrow-Bandgap-Conjugated Polymers for Plastic Electronics. *Nat. Commun.* **2017**, *8*, 14047.
(79) Liang, Y.; Xu, Z.; Xia, J.; Tsai, S. − T.; Wu, Y.; Li, G.; Ray, C.; Yu, L. For the Bright Future—Bulk Heterojunction Polymer Solar Cells with Power Conversion Efficiency of 7.4%. *Adv. Mater.* **2010**, *22*, E135−E138.


DOI: 10.1021/acs.jpclett.8b00635
*J. Phys. Chem. Lett.* 2018, 9, 2639−2646