THE JOURNAL OF PHYSICAL CHEMISTRY C

pubs.acs.org/JPCC Article

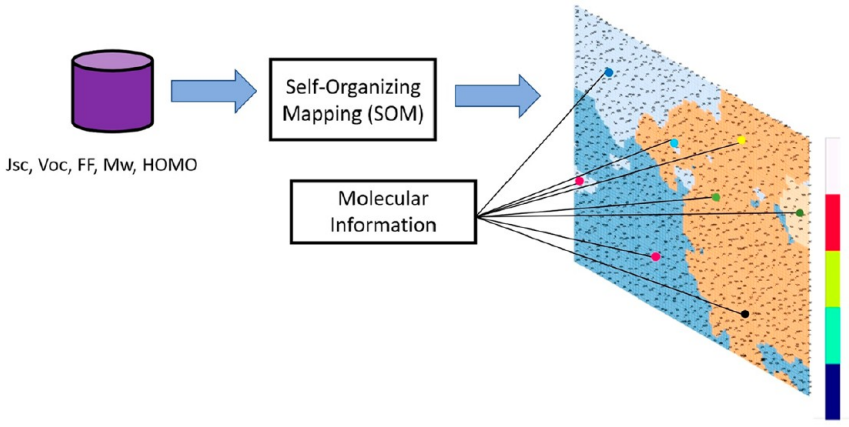
Structure–Property Correlation Study for Organic Photovoltaic Polymer Materials Using Data Science Approach

Published as part of The Journal of Physical Chemistry virtual special issue "Machine Learning in Physical Chemistry".

Yue Huang,*,# Jingtian Zhang,# Edwin S. Jiang, Yutaka Oya, Akinori Saeki, Gota Kikugawa, Tomonaga Okabe, and Fumio S. Ohuchi*

Cite This: *J. Phys. Chem. C* 2020, 124, 12871−12882 Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: A study workflow that utilizes several data science methods to apply on polymer materials databases is introduced to reveal correlations among their properties, structural information, and molecular descriptors. The data science methods used in this pipeline include the unsupervised machine learning (ML) method of self-organizing mapping (SOM) and the polymer molecular descriptor generator, both of which have been tailored to fit the polymer materials study. To demonstrate how this pipeline can be applied in this context, we used it on an organic photovoltaic (OPV) donor polymer database to investigate which properties or structural factors positively correlate with the power conversion efficiency (PCE) of OPV materials. This led us to discover that among the studied 8 properties and 11 molecular descriptors, only the photon energy loss (E_{loss}) and the number of fluorine atoms (nF) show strong positive correlations with PCE values, which is consistent with other verified studies. We also discovered that research trends can also be statistically visualized using our method. In our case study, we found that most of the studied OPV donor materials in the database have branched side chains and typically 7−12 non-hydrogen atoms, and high PCE materials usually have 6−9 aromatics rings as well. These results proved that the data science pipeline proposed in this study provides a fast and effective way to obtain research insights for polymer materials.

1. INTRODUCTION

We are at a time when data science is becoming applicable to almost any research fields. There are two major reasons for this: one, data science and computational developments have progressed greatly in the past years to the point where many tools and algorithms developed for other research fields have become very powerful, adaptable, and more easily accessible; two, the development of the research instrumentation makes science research more effective and efficient in generating data, and the size of the data in almost all research fields is increasing rapidly. New data science approaches now lead to many breakthroughs in scientific research, including in the field of materials research.

In particular, as it comes to polymer design research, molecular fingerprints and descriptors, such as molecular

Received: January 18, 2020
Revised: April 20, 2020
Published: April 24, 2020

https://dx.doi.org/10.1021/acs.jpcc.0c00517

electronegativity, molecular coordinates, number of atoms, etc., can be used as inputs to machine learning (ML) models to screen for well-matched polymer materials or generating polymer structures that can lead to improved performance. The ML methods used in such research can be classified into two categories, supervised ML or unsupervised ML.

In supervised ML, several intriguing researches[1−4] have been accomplished using the random forest (RF) classifier to generate the classification of polymers performance based on their molecular information. This pipeline is very successful in expediting the screening process for polymer design, synthesis, and characterization, but usually requires significant human efforts in conducting data training.

In unsupervised ML, recent progress has focused on utilizing dimension reduction methods such as principle component analysis (PCA),[5] t-distributed stochastic neighbor embedding (t-SNE),[6] and self-organizing map (SOM)[7] to analyze correlation across properties. With materials' properties having different levels of correlation and fundamental connection, their datasets naturally have a high dimensionality. Among the dimension reduction methods mentioned, t-SNE is popular for its nonlinear dimension reduction and visual presentation of high dimensional data in a low dimensional space. It was used in protein study to classify proteins based on simulation calculated properties,[9] and in another study, previously uncharted band structure space for thousands of materials are clustered based on the DFT calculated energy dispersion data.[10] Although limited publications can be found focusing on demonstrating the results of PCA analysis, this method is widely used as a reference in confirming the correlation between structure and particular properties mathematically.[5,8−10] However, most research found that PCA can cause unavoidable loss of physical meaning, whereas t-SNE and SOM are more favorable to study high-dimensional data for human-intuitive visualizations. In comparison with t-SNE, although relatively new to the community, SOM is more efficient in unraveling nonlinear and unordered data in real cases.[11] Therefore, it is more suitable for material science research to find correlations among different properties, or correlations between properties and structure of the materials. A detailed explanation of how to use SOM in materials science can be found in our earlier publication[12] in which we were able to cluster and discover correlations of properties among 20 properties simultaneously. None of these researches can be conducted manually without efficient data science methods.

Even though data science methods can efficiently classify hundreds, or even thousands materials based on their properties, not so many functions and methods have been developed to discover correlation patterns between particular properties and structure of the materials, particularly for the polymer materials, and the application of SOM on polymer study has been limited so far. In this study we will demonstrate how correlation among structure and properties of polymer can be discovered by introducing a new data science workflow which can be used to improve the efficiency of polymer design.

1.1. SOM and the Challenge of Its Application in Polymer Informatics. One of the difficulties and challenges in applying dimension reduction to polymer systems stems from the fact that there are many more distinctive factors differentiating one polymer structure from another than in inorganic materials.[13] In polymer informatics, these distinctive factors of the structures, or the mathematical and logical representations of the molecular configuration are called "molecular descriptors" and include mainly three categories of information about a polymer: (a) chemical, such as different chemical elements or different functional groups existing inside of the polymer, (b) geometrical, such as linear or ring structured, and (c) different types of branching for the side chains. The number of molecular descriptors of polymers can be in the thousands,[14] and the properties or performances of a polymer can change with any of them. Moreover, one molecular descriptor may impact the properties differently when the other molecular descriptors change. If all these molecular descriptors are included in the training set of SOM, this extremely high dimensionality training set will cause all materials to appear sparse and dissimilar in the high dimensional space, making "clustering and organization" almost impossible. This problem is usually described as the "curse of dimensionality".[15]

To avoid this problem, we adopted the "information projection" function of the SOM that was developed and described in our earlier study.[12] In that study, this function helped us by projecting categorical data onto the SOM cluster map to find the correlation between measured properties and categories of materials. In this study, we project the numerical data, including properties and molecular descriptors, onto the cluster map. This helps us effectively and efficiently identify which geometrical or chemical structure will likely give us the most desired property. This information projection process is illustrated in Figure 1.

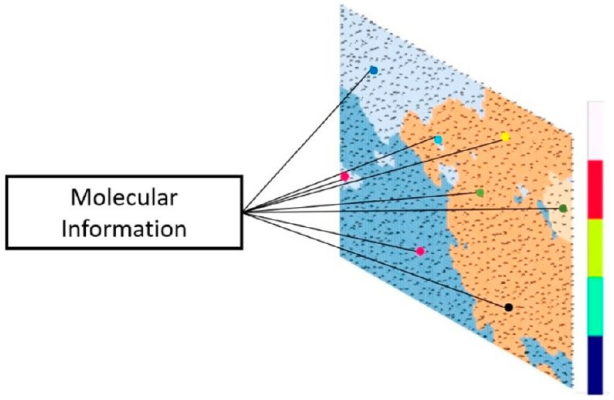


Figure 1. Molecular descriptors used for this project are projected on the cluster map by color coding the dots, which each correspond to a unique polymer ID. This enables visualization of their distributions on that map. Different colors correspond to various ranges of normalized values for the projected property.

Molecular descriptors or geometrical and chemical information on monomers can be extracted by molecular descriptor generator from SMILES (Simplified Molecular-Input Line-Entry System), a widely used cheminformatics tool to describe molecular structure in a one-line short ASCII string.[16] It was initiated in the 1980s[17,18] and then commonly used in chemical informatics studies.[1,19,20] However, for experimental researchers or researchers who do not base their research on computational tools, the usage of SMILES is quite limited. Without a proper tool to help us organize and use information from SMILES, it cannot be used directly. Molecular descriptors can be expressed from SMILES by using the Mordred Descriptor Generator (MDG).[21] MDG is an open-source python library based on the cheminformatics toolkit Rdkit.[22] When compared to other available open-source software, such as BlueDesc,[23] PyPDI,[24]

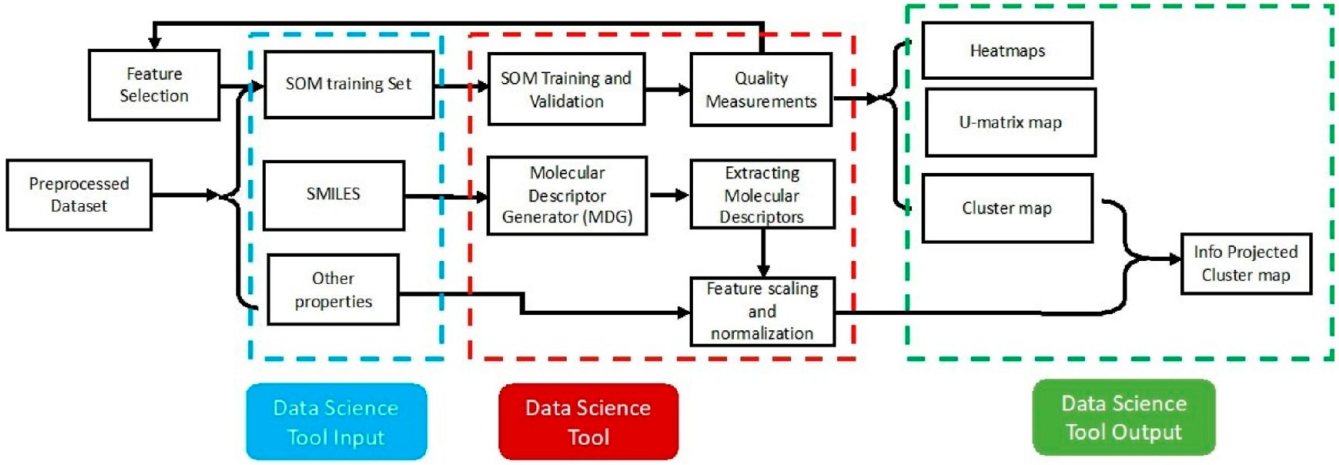


Figure 2. Data processing pipeline, including input, tools, and output. Information is projected on the cluster map to visualize the property−structure relationship after normalization. Heatmaps are used to visualize property−property correlations and are produced directly from SOM training.

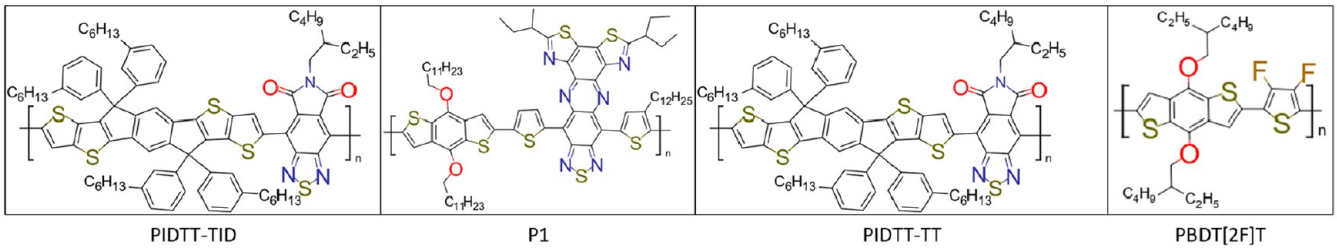


Figure 3. Samples of p-type donors' monomer structures in the data set. A total of 1203 polymers are included, which are the studied donor polymer in a fullerene acceptor OPV system.

Rcpi,[25] Dragon,[26] etc., it has advantages of being able to generate up to 1825 molecular descriptors and to support parallel computation, web interface, and command-line interface. The workflow adopted in this project is shown in Figure 2, where we combine both SOM and molecular descriptor generators to study OPV materials.

1.2. Database and Experimental Variables. To demonstrate how this study pipeline was utilized, we applied our methods to a data set of donor materials in the organic photovoltaic (OPV) devices with fullerene acceptor. Database used in the present study was built by Nagasawa et al.,[1] and the objective was to understand how the properties were correlated with each other, or to the molecular descriptors.

In the OPV devices, measurements commonly taken to accurately characterize functionality of the devices and materials include short-circuit current density (J_{sc}), open-circuit voltage (V_{oc}), fill factor (FF), various means of molecular weights (e.g., weight-averaged one, M_w, and number-averaged one, M_n), spectral absorption, internal and external quantum efficiencies (IQE and EQE),[1] and others. Performance of the devices is generally evaluated by the power conversion efficiency (PCE),[27] of which values were used as the main indicator of the performance of the OPV devices.

The data set used in this study contains a total of 1203 different types of organic photovoltaic polymers, of which experimental data were manually collected from more than 500 papers. Several examples of the p-type polymers in the data set are illustrated in Figure 3. This data set contains SMILES and 11 properties for each of these polymers, including the V_{oc}, J_{sc}, FF, molecular weights (M_w, M_n, and the weight of monomer unit), polydispersity index (PDI), principal energy levels (highest occupied molecular orbital, HOMO, and lowest unoccupied molecular orbital, LUMO), optical bandgap (E_g), photon energy loss (E_{loss}), and PCE.

In this paper, we investigated the correlation between 11 molecular descriptors and the PCE performance of the material in our data set. The selected 11 molecular descriptors are the existence of 4 elements (nitrogen (N), oxygen (O), sulfur (S), fluorine (F)), the number of side chains (nChain), the number of all aromatic rings (naRing), the number of five-membered aromatic rings (n5aRing), including pyrrole, pyridine, furan and thiophene, the number of six-membered aromatic rings (n6aRing), including cyclohexane and benzene, the ratio between nChain and naRing (nChain/naRing), and whether the material contains branched side chains and/or linear side chains. These are geometrical and chemical features that are possibly associated with PCE values of the device in reported studies.[11,28,29]

In this project we do not discuss 3D molecular descriptors from SMILES since they are mostly related to the molecular conformation and electronegativity. For example, 3D descriptors provided by MDG include radius and diameter of the entire molecule, fractional charge partial negative surface area, geometric shape index, etc. These descriptors are difficult to analyze since many factors have convoluted impacts on the outputs. We are focusing on investigating 2D descriptors in this work.

We also studied other material properties that could be correlated to the PCE values, including E_g and E_{loss}. Here, E_{loss}, as expressed in eq 1,[30] is preliminarily determined by variations in energetic offset between the E_g and V_{oc} with values ranging between 0.6 and 1.0 eV. Several theoretical and empirical studies

have demonstrated the importance of reducing E_{loss} to reach maximum possible V_{oc} for devices.[1]

$$E_{loss} = E_g - qV_{oc} \quad (1)$$

where q is the elementary charge.

In this paper, we constructed a research workflow that includes several data science methods to demonstrate how the data science methods can be utilized in searching for correlation between the structural or chemical descriptors and the polymer properties.

In the following sections, we describe data feature selection for ML, SOM training, and visualization of the data set, so that clustering of the materials reflects the PCE performance of the materials. We then describe the method used to extract molecular descriptors from SMILES and eventually discuss how the extracted molecular descriptors and other properties are projected to the 2D cluster maps in order to study the correlations between these information and material performances. Here, we point out that, even though very specific systems are studied in this case study, the method we describe in this study can be used in other data sets, such as nonfullerene acceptor systems that are attracting more research interests lately.[31−34]

2. METHODS

2.1. Feature Selection for SOM. In ML, feature selection is a critical step where features, or properties, are chosen to be used as input. In our study, the design matrix for model training is formed by 5 dominant features, J_{sc}, V_{oc}, FF, M_w, and HOMO. The first three, J_{sc}, V_{oc}, and FF, are direct contributors to the PCE values, whereas M_w and HOMO are chosen for their promising correlations with PCE. It has been known that M_w tends to give enhanced PCE for the same molecular structure and backbone and also creates a more fibrous structure of the polymer domain in a BHJ film.[1,35] In addition, higher M_w with a narrow distribution leads to a higher degree of crystallinity.[36−38] HOMO is chosen because deepening of HOMO results in an increased V_{oc}. With this combination, relative distances between the data points were preserved by ensuring the best preservation of the topology, therefore yielding better results.[7,11] More importantly, the properties we have chosen in the feature selection gave a good clustering of the materials in the PCE values. This will be discussed in the next section.

2.2. Training and Visualization by SOM. Both the data visualization and training come from an updated version of a python package called tfprop_sompy by Kikugawa and Nishimura[39] based on the Python package SOMPY.[40] The data processing and visualization works are done using multiple Python packages, including Pandas, Numpy, Scikit-learn, and Matplotlib.

Output of the SOM model consists of a set of 2D arrays including cluster map, heat maps, and U-matrix map. Among these, we focus on the visualization of cluster map and heatmaps, as described in the project workflow in Figure 2. The U-matrix maps were used only to benchmark the results of the clustering map to prove that our clustering map does represent the closeness of the studied properties of the polymer materials. The cluster map provides where the materials are located on the 2D array map, whereas the heatmaps are used to visualize the pattern of the properties' distribution in the design matrix. The PCE values are also projected from the database to confirm that our clustering map does put materials with high PCE together and, therefore, reflects the performance of the materials. The results of the PCE cluster map are shown in Figure 4, in which the high PCE value polymers are well clustered into several regions on the right side of the cluster map, and the low PCE value polymers in general are accumulated on the left side. In our polymer−fullerene OPV database, the PCE value ranges from 0.01% to 10.50%. To highlight the statistical significance of our method, we outline the polymers with a top 40% PCE value from our data set (PCE value higher than 6.3%) and focus our analysis on the correlation between the other factors with these polymers. For our study, we call these polymers high-PCE polymers.

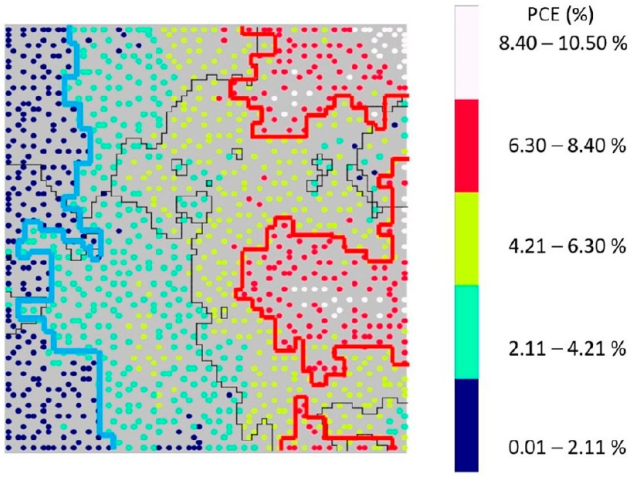


Figure 4. Projected PCE value map. The PCE values from the original data set are divided into 5 regions in different colors by min−max normalization. PCE boundary regions (red and blue line) are drawn out on the basis of the data distribution. The red line indicates the top 40% of the PCE values, and the blue line indicates the lowest 20% of the PCE value.

In this study we used a map size of 80 × 80, cluster numbers of 4 to incorporates all 1203 samples, and 5 selected features and generated the cluster map that no node accommodates more than one data point. With these parameters, the topographic error and quantization error of the cluster maps is minimized to 0.0116376 and 0.128816, respectively. SOM training was made on a local laptop with a 1.8 GHz eighth generation Intel Core quad-core processors of Intel (R) Core(TM) i7-8550U in 447.5 s.

2.3. Extraction of Information from SMILES. The python library Mordred Descriptor Generator includes most of the molecular descriptors that are studied in this project.[41] In our case study, a 1203 × 1825 matrix table was generated automatically with the numerical counts of every descriptor for each material from the 1203 instances using SMILES as the input file. The 11 molecular descriptors that are of interest were then selected manually for the information projection step of our workflow.

Other geometrical information of interest, such as the presence of a branched side chain or a linear side chain, were extracted by using data processing with some parsing rules directly from the SMILES ASCII code. All the data and the algorithm can be found at the GitHub address: https://github.com/DataScienceUWMSE/SOM_OPV.

2.4. Information Projection. The most important step in our method is to project information on the cluster maps to reveal correlations between structural information and perform-

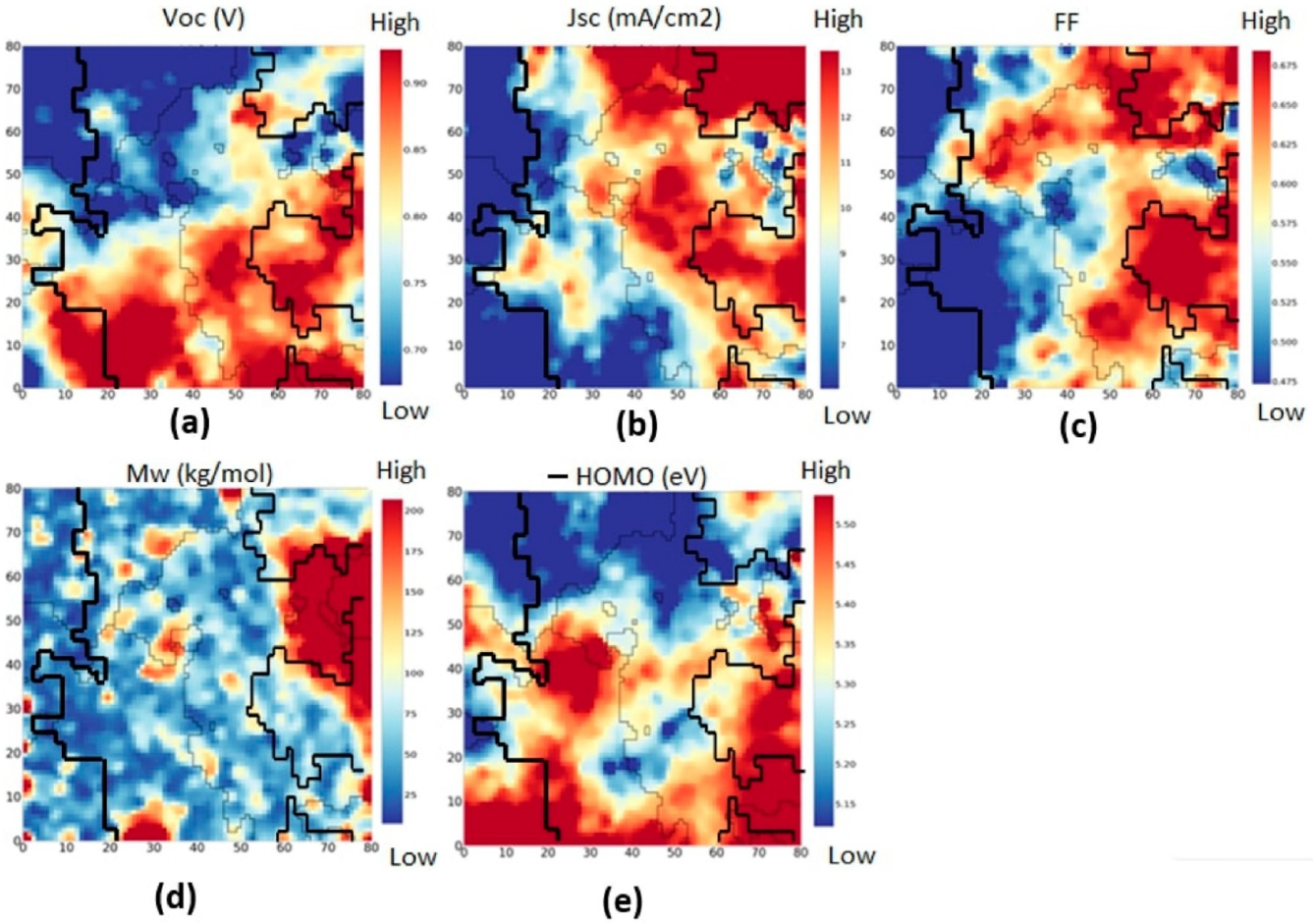


Figure 5. Heatmap for features, including (a) V_{oc}, (b) J_{sc}, (c) FF, (d) M_w, and (e) −HOMO. PCE boundaries are shown on the maps using the black bold lines.

ance of the materials. Since the cluster map has already grouped on the basis of the properties of interest (PCE performance in our study case), we are able to visually observe the distribution of the projected information. If high values of the projected information coincide well with high values of the properties presented in the cluster map, then a positive correlation can be concluded; otherwise, there is no correlation or a negative correlation between the projected information and the PCE value.

Through our study, we sometimes find that projection of the numerical data directly on the maps can be difficult to interpret. In the projection function, different numerical values are projected to the cluster map using different colors. If the range of values of the projected data is large, too many colors are used, which require users to differentiate colors that are very similar to each other or only have subtle gradient changes. Thus, feature scaling is required on all numerical data to make interpretation easier. In this case we applied the min−max normalization technique:[42]

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$
$$\text{label}_i = \text{round}(x' \times k + 0.499)$$
$$k \geq \text{label}_i \geq 0 \quad (2)$$

For every numerical data point of the property x, we compute a normalized data point x' and a label corresponding to it. k is the number of classes we want the property to be divided into, and 0.499 is a constant value to round up label values in order to make distribution of the results more uniform. The calculated label after rounding should always range between 0 and k. Then, a dot that represents each material is colored on the basis of the label corresponding to a range of the projected value this material belongs to. As shown in the illustration of the projection in Figure 1, different colors correspond to different ranges of normalized values for the projected property.

In the next section we will first discuss the results and conclusion we can obtain from the heat maps and then discuss the results from projecting the molecular descriptors onto the cluster map.

3. RESULTS AND DISCUSSION

3.1. Heat Maps. Figure 5 presents heat maps of the properties of V_{oc}, J_{sc}, FF, M_w, and −HOMO, from which we can visually see how these property values are distributed on the cluster map. It is apparent that there is a trade-off effect between J_{sc} and V_{oc} on their heatmaps, whereas majority of the high PCE materials has a low-lying HOMO or medium HOMO. This could be rationalized by the fact that lower-lying HOMO levels contributes to larger V_{oc} in BHJ devices.[43] J_{sc} and FF have similar

https://dx.doi.org/10.1021/acs.jpcc.9b00517

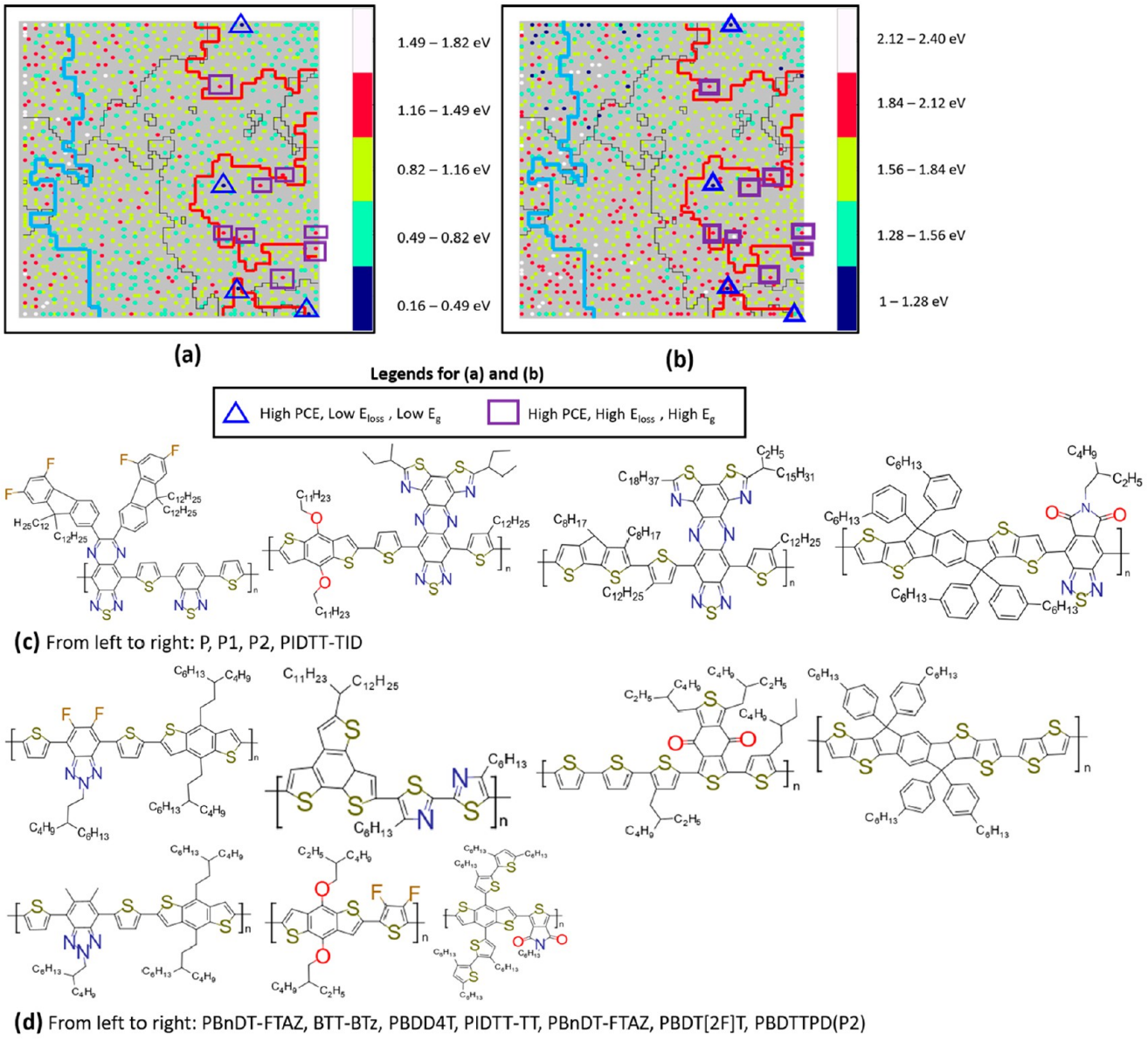


Figure 6. Information projected cluster map of (a) E_{loss} and (b) E_g. (c) Monomer structures generated from monomer SMILES of materials marked by triangles in (a) and (b) with high PCE, low E_{loss}, and low E_g. (d) Monomer structures marked by rectangles in (a) and (b) with high PCE, high E_g, and high E_{loss}. High PCE region are shown on the cluster maps in red, and low PCE region, in blue lines.

patterns of distribution, indicating a positive correlation between them.

Comparing the PCE values projected on the cluster map in Figure 4 with the other heat maps, we find that some materials with high molecular weight have PCE ranging between 4.21% and 6.30%, but in general, these high molecular weight polymers are not within the high PCE regions. This observation is not consistent with the general consensus that high-performance OPV materials have high M_W due to lower density of recombination centers (persistent radical defects revealed by EPR spectroscopy) and better photoactive layer morphology in the samples.[44] This may be an indication that molecular weight has competing contributions in getting high PCE values as some other molecular descriptors. This does not exclude the possibility that molecular weight can play an important role in achieving high PCE. Some research showed that even though the bandgap or energy levels might not significantly change, the absorption profile and the absorption intensity can be significantly influenced by the molecular weight.[45−47]

The heat maps provide visualizations displaying compressed information on the original high-dimensional data on an organized 2D mesh. The x axis and y axis of the 2D mesh, which are the same in the heat maps and in the cluster map, are unitless and without any physical meaning. They provide a distribution of the materials that are studied in the ML learning. The materials with the same ID locate at the exact same position on all heat maps and cluster map. The cluster map presents how materials cluster on the 2D mesh, and heatmaps relate their properties to it; hence, correlations among properties can be directly observed. This is a human intuitive way to observe high-dimensional features while their topology is relatively well-preserved. This has been demonstrated in Qian et al.'s work.[12]

3.2. Projection Maps. In order to study the correlation between the PCE performance and other factors, such as

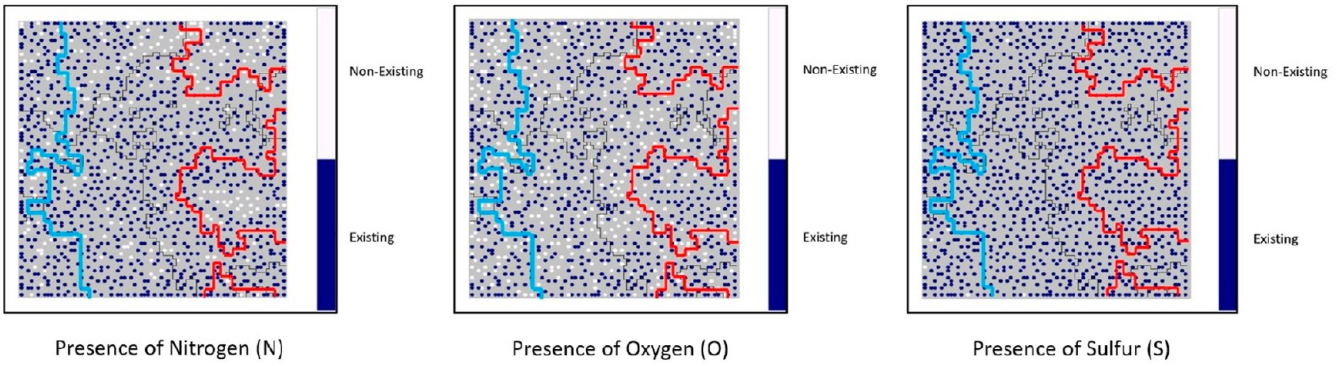


Figure 7. Presence of atoms for the following elements: N, O, S. PCE boundaries are shown on the cluster maps in red and blue lines.

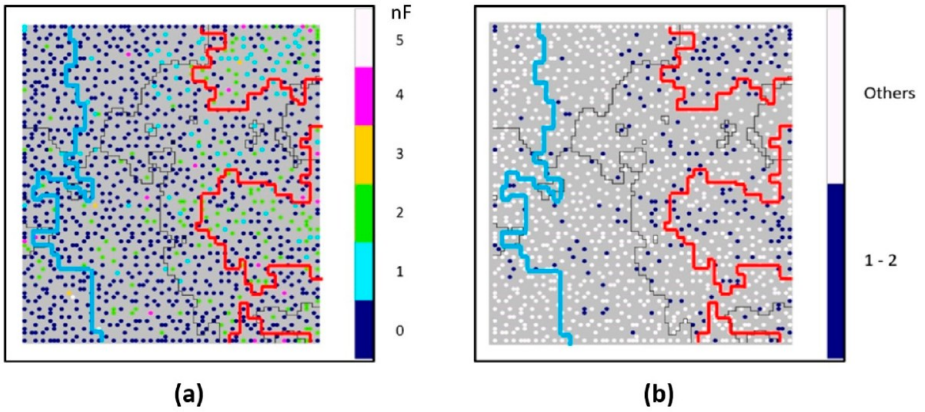


Figure 8. Information projected cluster map of (a) the number of F atoms contained in the monomers of the donor materials and (b) the existence of 1 or 2 F atoms in the monomers of the donor materials. PCE boundaries are shown on the cluster maps in red and blue lines.

chemical information, and geometrical information on the polymers, we now utilize the projection function developed in SOM on the cluster map. All the projection maps have a high PCE region outlined so that the correlation can be identified visually.

3.2.1. Projection of E_g and E_{loss}. First, we study the correlation of PCE values with E_g and E_{loss}. Note that these two properties are included in the data set but not used in the training set. Figure 6a shows the distribution of E_{loss} across the cluster map. It is found that materials with high E_{loss} values tend to be aggregated around the low-PCE regions, while the very few materials with low E_{loss} (as displayed in Figure 6c) all exist in the high PCE value region of the map. This observation is consistent with recent studies of the OPV materials[48−50] that a low E_{loss} (less than 0.5 eV) can improve the PCE value of the material by as much as 15% even though most materials typically hold E_{loss} values between 0.7 and 1 eV. Figure 6c lists the materials with a low bandgap and low energy loss. Using the unique ID of each polymer in SOM input, we can easily track these polymers to its original source publications.[51−53]

There are also a few high E_{loss} materials in the high-PCE region, which we mark with rectangles in Figure 6a. By comparing E_{loss} and E_g projected cluster maps (i.e., Figure 6a,b), we find that these materials also have wide bandgap values (high E_g). According to eq 1, high E_{loss} and high E_g might result in high V_{oc}, which increases the PCE value. By comparing with the heatmap of V_{oc} in Figure 5, we confirmed that all these low E_{loss} and wide-bandgap OPV materials (WBG) listed in Figure 6d have high V_{oc} values. It brings to light that for high E_{loss} materials obtain a high PCE value, high E_g is necessary. However, this does not conclude that high E_g leads to high PCE value. As shown in Figure 6b, on the contrary, very few high E_g polymers existing in the high PCE region, and we can see most of them are polymers with high E_{loss}.

3.2.2. Projection of Chemical Elements. Next, we study how the chemical elements influence on a trend in the PCE performance by projecting several elements, including F, N, O, and S, to the cluster map. From the results shown in Figures 7 and 8, we find that only the presence of F has evident patterns of the distribution on the map, while the presence of N, O, and S elements in the polymers shows little specific distribution on the map, indicating that those elements do not decisively influence the PCE value. In Figure 8, we also classify the polymers based on the number of the F atoms in the polymer chain and find that the distribution of the polymers that contain 1 to 2 F atoms aligns well with the high-PCE region, but the polymers with none or more than 2 F atoms do not show this pattern. Given this insight from the ML results, among the total of 253 low PCE materials, on the basis of Figure 8b, only 14 of them have 1 or 2 fluorine atoms attached in the monomers. We did some further data analyses and found that 21% of materials in our data set contain F atoms, but 44% of high-PCE materials have F atoms in their backbone structure, and among them more than 90% have 1 or 2 F atoms, which constitute 40% of high PCE materials.

The importance of the F functional group in a donor material is believed to be related to its electron-withdrawing nature.[54] Introduction of fluorine into the polymer backbones can simultaneously downshift HOMO and LUMO levels without

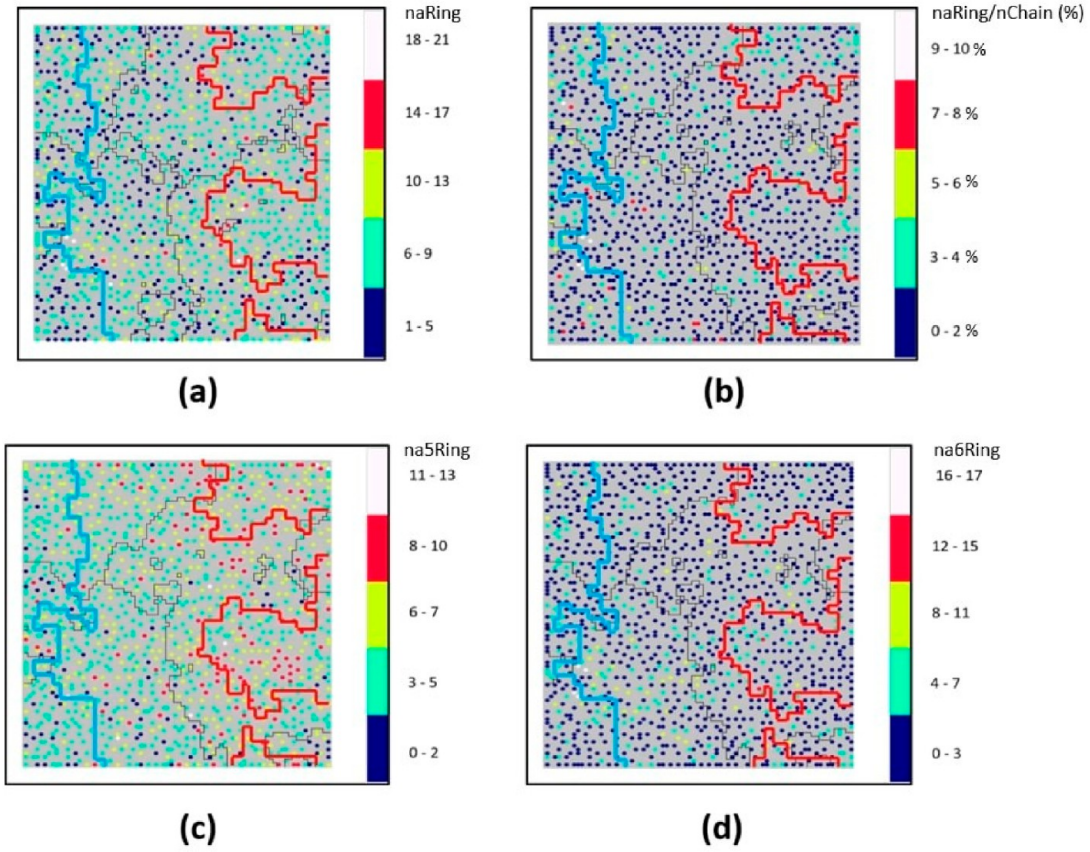


Figure 9. Information projected cluster map of (a) total number of all aromatic rings (naRing), (b) the ratio between naRing and nChain (naRing/nChain), (c) the number of five-membered aromatic rings (n5aRing), and (d) the number of six-membered aromatic rings (n6aRing) on the cluster map. PCE boundaries are shown on maps in red line.

causing strong steric hindrance of the resulting molecules,[54] thereby increasing V_{oc}. Enhancement of the inter/intramolecular interactions and localization of the LUMO density on the structure are known to increase the crystallinity, facilitating charge transfer and transport in fluorinated molecules.[54−57] Even though our data science approach does not give an explanation why polymers with 1 or 2 F atoms show better OPV properties, we statistically demonstrate that a majority of polymers containing 1 or 2 F atoms in the monomers have high PCE values when used in fullerene acceptor devices, which can provide some insight for researchers into engineering better OPV materials.

In the original data set, there are several pairs of comparisons that show the incorporation of fluorine can increase the performance: PTBF0[58] and PTBF1[58] (increase the PCE from 2.7% to 6.2%), PBT-0F[59] and PBT-3F[59] (increase the PCE from 4.5% to 8.6%), and PBnDT-HTAZ[60] and PBnDT-FTAZ[60] (increase the PCE from 4.36% to 7.1%). However, it is not a universal rule that the addition of fluorine will increase the performance. The addition of fluorine could cause the change of other conditions, thus leading to the decrease of performance.[61]

3.2.3. Projection of Number of Aromatic Rings. General existence of S shown in Figure 7 suggests the existence of thiophene rings inside of the polymers. Therefore, we project some geometrical information on the monomers on the cluster map in search for more patterns for high PCE polymers. Figure 9 presents the projection of number of chains, aromatic rings (total, five-membered and six-membered), and the ratio between the number of aromatic rings and that of side chains. The total aromatic rings include all kinds of aromatic ring structures, and five-membered or six-membered aromatic rings in the data set are mostly thiophene and benzene rings.

It has been suggested that the addition of thiophene rings can improve the fill factor and morphology while designing donor−acceptor copolymers.[54] The prior research shows that the number of benzenes has a relationship with the performance of OPV materials.[62] This is in fact shown in Figure 9a, where we find that most of high-PCE materials have 6−9 aromatics rings within their backbone structures. We therefore projected five-membered rings (usually thiophene rings) to the cluster map and then discovered that most of the polymers investigated had 3−7 thiophene rings in their monomer structures. This is clearly shown in Figure 9c. Even though we did not observe a correlation between the number of aromatic rings and high PCE values, we did expose the trend in current research in this field to focus on polymers containing 3−7 thiophene rings in their monomer structures.

The increase of aromatic rings generally leads to a lower solubility and worse BHJ network, which can be facilitated by increasing the number of side alkyl chains. The fraction of naRing over the nChain represented as naRing/nChain in percent was projected in the cluster map in Figure 9b. The results of that projection, however, show little or no correspondence with the PCE value distribution, suggesting that the scaling law is not preserved. Namely, the high-performance polymers with many aromatic rings do not require many alkyl chains to ensure good solubility, which is a prerequisite for a strong stacking effect and high hole mobility.

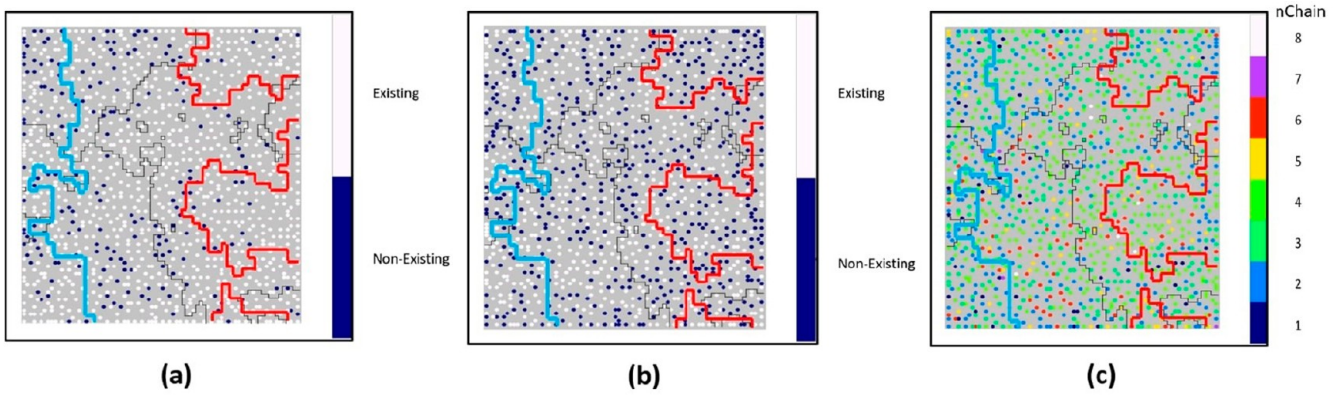


Figure 10. Information projected cluster map of (a) the number of branched chains, (b) the number of linear chains, and (c) the number of side chains in the monomer (nChain). PCE boundaries are shown on maps in red lines.

Thus, the nonlinear effect of alkyl chains (branched and linear) and backbone aromatics on the solubility and crystallinity of polymers should be clarified in due course.

3.2.4. Projection of Chain Length. For materials that share the same backbone structure, the optimal alkyl chain structure would be a combination of both branched and linear side alkyl chains that provide improved morphology of the donor−acceptor mix, thereby leading to better photophysical properties.[63−66] Even though it is not meaningful to compare the number and length of side chains (linear or branched) across materials with different backbone structures, we are getting insight into the type of polymers researchers tend to study.

Besides the number of alkyl side chains, the length of each chain can also be extracted from canonicalized SMILES. We designed a function to extract the number and length of backbone and side chains from SMILES by recognizing the consecutively connected carbon atoms, including their numbers and lengths. We regard more than 3 carbon atoms as a side chain, and a side chain with additional attached chains will be considered as a branch.

The generated result is the combination of two arrays, such as [0,2,2][8,7,7]; the number of elements in each array represents the number of chains, and each value in the list represents the length of each chain. For example, [0,2,2][8,7,7] should be interpreted as this monomer contains 3 chains, the first one is an 8-atom long straight chain, and the second and the third one are both branched, each with a short side chain of 2 atoms and a long backbone of 7 atoms. This SMILES parsing algorithm is also included in the GitHub repo.

With some simple data processing, we can visualize the distribution of the number of branches and length of chains on the cluster map by projecting them to the cluster map, as shown in Figure 10. By visually comparing Figure 10a,b, we may conclude that most of the polymers studied have branched side chains, whereas there is no obvious preference with respect to the existence of linear side chains in the materials being studied. By projecting the number of the side chains to the cluster map (Figure 10c), we found that most materials have 2−4 side chains attached to the backbone structure in their monomers. Even though we cannot find a pattern for the exact number or length of side chains on simple visualizations, at least we discovered that the majority of researchers who studied the fullerene OPV system reported polymers with 2−4 side chains containing at least one branched side chain. To further study the impact of length of side chains, specific studies should be done for each type of backbone with varying side chains lengths.

We also studied the average number of heavy atoms on each side chain's impact to the properties, as shown in Figure 11. Most of high-PCE materials have around 7−12 non-hydrogen atoms on each side chain.

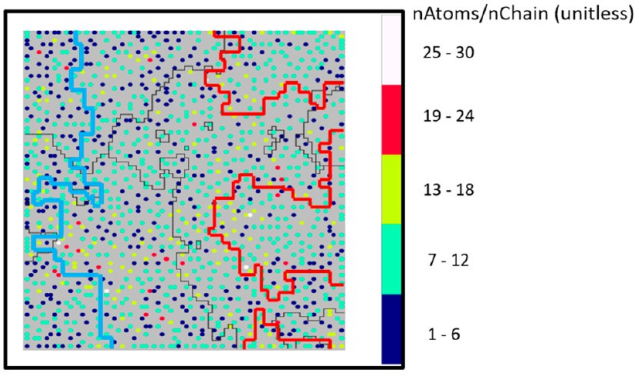


Figure 11. Information projected cluster map of average number of non-hydrogen atoms on each side chain in the monomer.

4. CONCLUSION

We demonstrated that SOM in combination with molecular descriptor generation techniques can be utilized to study the correlation between the structural information on the polymer materials and the performance of the polymer materials effectively and efficiently. In our case study, by visually comparing the PCE cluster maps and projected information, including other properties, chemical information, and structural information, we discovered that for fullerene acceptor OPV systems:

1. The donor polymers with 1 or 2 fluorine atoms in their monomer structures tend to have higher PCE values; more than 90% of polymers with 1 or 2 fluorine atoms have PCE values higher than 6.3%. Polymers containing none or more than 2 fluorine atoms do not exhibit the same pattern of having high PCE values.
2. Molecular weight of the donor polymers does not play a dominant role in influencing the PCE performance when M_w is compared across all polymers with different

structures, different chemical components, and different geometries.

3. Most systems that have low photon energy loss also have high PCE values. For high photon energy loss polymers, only the one with wide bandgap can achieve high PCE.
4. Most donor polymers that are currently under study by materials researchers have monomer structures containing branched side chain, and 2−4 side chains. Each side chain, on average, has 7−12 non-hydrogen atoms on it.

Even though these results were achieved purely from a statistical approach using data science methods, they are consistent with physical approach research results. Therefore, we can conclude that our proposed data science workflow, or pipeline, can be successfully utilized to extract useful information even from data sets containing data coming from a variety of sources across different research laboratories.

Although our case study focuses on the fullerene acceptor system, the proposed study workflow we introduced is generic and can be utilized to study other systems, such as nonfullerene small molecule acceptor organic solar cells systems, or other research topics related to polymer materials. As the amount of data available for polymer materials increases, such data processing pipelines as the one we proposed here will become ever more relevant and will help guide researchers in designing materials presenting the characteristics they desire.

AUTHOR INFORMATION

Corresponding Authors

Yue Huang − *Department of Materials Science and Engineering, University of Washington, Seattle 98195-1750, Washington, United States;* orcid.org/0000-0002-9927-1243; Email: huangyue@uw.edu

Fumio S. Ohuchi − *Department of Materials Science and Engineering, University of Washington, Seattle 98195-1750, Washington, United States; School of Engineering, Tohoku University, 980-8579 Sendai, Japan*; Email: ohuchi@uw.edu

Authors

Jingtian Zhang − *Chemical Engineering Department, University of Washington, Seattle, Wahsington 98195-2120, United States*

Edwin S. Jiang − *Department of Materials Science and Engineering, University of Washington, Seattle 98195-1750, Washington, United States*

Yutaka Oya − *School of Engineering, Tohoku University, 980-8579 Sendai, Japan*

Akinori Saeki − *Department of Applied Chemistry, Graduate School of Engineering, Osaka University, Suita 565-0871, Osaka, Japan;* orcid.org/0000-0001-7429-2200

Gota Kikugawa − *Institute of Fluid Science, Tohuku University, 980-8577 Sendai, Japan*

Tomonaga Okabe − *Department of Materials Science and Engineering, University of Washington, Seattle 98195-1750, Washington, United States; School of Engineering, Tohoku University, 980-8579 Sendai, Japan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.0c00517

Author Contributions

#Y.H. and J.Z. share equal contribution to this project.

Notes

The authors declare no competing financial interest.

All the data and the algorithm can be found at the GitHub address: https://github.com/DataScienceUWMSE/SOM_OPV.

ACKNOWLEDGMENTS

A.S. acknowledges the PRESTO program (Grant No. JPMJPR15N6) from the Japan Science and Technology Agency (JST) and the Japan Society for the Promotion of Science (JSPS) with the KAKENHI Grant-in-Aid for Scientific Research (A) (Grant No. JP16H02285). The authors also acknowledge the vitally important encouragement and support made through the University of Washington-Tohoku University: Academic Open Space (UW-TU: AOS). Special thanks to Prof. Samson A. Jenekhe (UW-ChemE) for mentoring JZ for his Master of Science degree.

REFERENCES

(1) Nagasawa, S.; Al-Naamani, E.; Saeki, A. Computer-Aided Screening Of Conjugated Polymers For Organic Solar Cell: Classification By Random Forest. *J. Phys. Chem. Lett.* **2018**, *9*, 2639−2646.

(2) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning In Materials Informatics: Recent Applications And Prospects. *npj Comput. Mater.* **2017**, *3*, 54.

(3) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sanchez-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.

(4) Shi, Z.; Yang, W.; Deng, X.; Cai, C.; Yan, Y.; Liang, H.; Liu, Z.; Qiao, Z. Machine-Learning-Assisted High-Throughput Computational Screening Of High Performance Metal−Organic Frameworks. *Mol. Syst. Des. Eng.* **2020**, *5*, 725−742.

(5) Rajan, K.; Suh, C.; Mendez, P. Principal Component Analysis And Dimensional Analysis As Materials Informatics Tools To Reduce Dimensionality In Materials Science And Engineering. *Statistical Analysis and Data Mining* **2009**, *1* (6), 361−371.

(6) Maateen, L.; Hinton, G. Visualizing High-Dimensional Data Using T-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579−2605.

(7) Kohonen, T. The Self-Organizing Map. *Proc. IEEE* **1990**, *78* (9), 1464−1480.

(8) Karlov, D. S.; Sosnin, S.; Tetko, I. V.; Fedorov, M. V. Chemical Space Exploration Guided By Deep Neural Networks. *RSC Adv.* **2019**, *9*, 5151−5157.

(9) Zhou, H.; Wang, F.; Tao, P. T-Distributed Stochastic Neighbor Embedding Method With The Least Information Loss For Macromolecular Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 5499−5510.

(10) Nuñez, M. Exploring Materials Band Structure Space With Unsupervised Machine Learning. *Comput. Mater. Sci.* **2019**, *158*, 117−123.

(11) Rajan, K. *Informatics for Materials Science And Engineering*; Elsevier, Inc., 2013, pp 423−442.

(12) Qian, J.; Nguyen, N.; Oya, Y.; Kikugawa, G.; Okabe, T.; Huang, Y.; Ohuchi, F. Introducing Self-Organized Maps (SOM) As A Visualization Tool For Materials Research And Education. *Results in Materials* **2019**, *4*, 100020.

(13) Audus, D.; de Pablo, J. Polymer Informatics: Opportunities And Challenges. *ACS Macro Lett.* **2017**, *6* (10), 1078−1082.

(14) Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*; Wiley, 2000.

(15) Donoho, D. L. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. In *AMS Conference on Math Challenges of the 21st Century*; AMS, 2000.

(16) *Daylight Theory: SMILES*. https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html (accessed March 16, 2020).

(17) Weininger, D. SMILES, A Chemical Language And Information System. 1. Introduction To Methodology And Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28* (1), 31−36.
(18) Weininger, D. SMILES. 3. DEPICT. Graphical Depiction Of Chemical Structures. *J. Chem. Inf. Model.* **1990**, *30* (3), 237−243.
(19) Gasteiger, J.; Engel, T. *Applied Chemoinformatics: Achievements and Future Opportunities*; Wiley-VCH Verlag GmBH, 2018.
(20) Quirós, M.; Gražulis, S.; Girdzijauskaitė, S.; Merkys, A.; Vaitkus, A. Using SMILES Strings For The Description Of Chemical Connectivity In The Crystallography Open Database. *J. Cheminf.* **2018**, *10* (1), 23.
(21) Moriwaki, H.; Tian, Y.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminf.* **2018**, *10* (1), 4.
(22) *RDKit*. http://www.rdkit.org/ (accessed March 16, 2020).
(23) *University of Tübingen: BlueDesc*. http://www.ra.cs.uni-tuebingen.de/software/bluedesc/ (accessed March 16, 2020).
(24) Cao, D.; Liang, Y.; Yan, J.; Tan, G.; Xu, Q.; Liu, S. Pydpi: Freely Available Python Package For Chemoinformatics, Bioinformatics, And Chemogenomics Studies. *J. Chem. Inf. Model.* **2013**, *53* (11), 3086−3096.
(25) Cao, D.; Xiao, N.; Xu, Q.; Chen, A. Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. *Bioinformatics* **2015**, *31*, 279−281.
(26) Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. DRAGON software: an easy approach to molecular descriptor calculations. *MATCH Commun. Math Comput. Chem.* **2006**, *56* (2), 237−248.
(27) Tress, W. *Organic Solar Cells*; Springer Verlag, 2016; pp 67−215.
(28) Kanal, I. Y.; Hutchison, G. R. Rapid Computational Optimization of Molecular Properties using Genetic Algorithms: Searching Across Millions of Compounds for Organic Photovoltaic Materials. *arXiv:1707.02949*. https://arxiv.org/abs/1707.02949 (accessed March 16, 2020).
(29) Olivares-Amaya, R.; Amador-Bedolla, C.; Hachmann, J.; Atahan-Evrenk, S.; Sánchez-Carrera, R.; Vogt, L.; Aspuru-Guzik, A. Accelerated Computational Discovery Of High-Performance Materials For Organic Photovoltaics By Means Of Cheminformatics. *Energy Environ. Sci.* **2011**, *4* (12), 4849.
(30) Liu, W.; Zhang, J.; Xu, S.; Zhu, X. Efficient Organic Solar Cells Achieved At A Low Energy Loss. *Science Bulletin* **2019**, *64* (16), 1144−1147.
(31) Zhao, W.; Li, S.; Yao, H.; Zhang, S.; Zhang, Y.; Yang, B.; Hou, J. Molecular Optimization Enables Over 13% Efficiency In Organic Solar Cells. *J. Am. Chem. Soc.* **2017**, *139* (21), 7148−7151.
(32) Zhang, Y.; Samuel, I.; Wang, T.; Lidzey, D. Current Status Of Outdoor Lifetime Testing Of Organic Photovoltaics. *Advanced Science* **2018**, *5* (8), 1800434.
(33) Zhan, X.; Marder, S. Non-Fullerene Acceptors Inaugurating A New Era Of Organic Photovoltaic Research And Technology. *Materials Chemistry Frontiers* **2019**, *3* (2), 180−181.
(34) Nielsen, C.; Holliday, S.; Chen, H.; Cryer, S.; McCulloch, I. Non-Fullerene Electron Acceptors For Use In Organic Solar Cells. *Acc. Chem. Res.* **2015**, *48* (11), 2803−2812.
(35) Kim, J.; Gadisa, A.; Schaefer, C.; Yao, H.; Gautam, B.; Balar, N.; Ghasemi, M.; Constantinou, I.; So, F.; O'Connor, B.; Gundogdu, K.; Hou, J.; Ade, H. Strong Polymer Molecular Weight-Dependent Material Interactions: Impact On The Formation Of The Polymer/Fullerene Bulk Heterojunction Morphology. *J. Mater. Chem. A* **2017**, *5* (25), 13176−13188.
(36) Kline, R.; McGehee, M.; Kadnikova, E.; Liu, J.; Fréchet, J. Controlling The Field-Effect Mobility Of Regioregular Polythiophene By Changing The Molecular Weight. *Adv. Mater.* **2003**, *15* (18), 1519−1522.
(37) Vakhshouri, K.; Smith, B.; Chan, E.; Wang, C.; Salleo, A.; Wang, C.; Hexemer, A.; Gomez, E. Signatures Of Intracrystallite And Intercrystallite Limitations Of Charge Transport In Polythiophenes. *Macromolecules* **2016**, *49* (19), 7359−7369.
(38) 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 PC61BM Leads To High Efficiencies In Solar Cells. *Adv. Mater.* **2012**, *24* (3), 425−430.
(39) Oya, Y.; Kikugawa, G.; Okabe, T. Clustering Approach For Multidisciplinary Optimum Design Of Cross-Linked Polymer. *Macromol. Theory Simul.* **2017**, *26* (2), 1600072.
(40) Moosavi, I.; Packmann, S.; Valles, I. sevamoo/SOMPY. https://github.com/sevamoo/SOMPY.git (accessed November 30, 2019).
(41) *mordred 1.2.1a1 documentation*. https://mordred-descriptor.github.io/documentation/master/index.html (accessed March 16, 2020).
(42) Patro, S.; sahu, K. Normalization: A Preprocessing Stage. *IARJSET* **2015**, 20−22.
(43) Chen, H.; Hou, J.; Zhang, S.; Liang, Y.; Yang, G.; Yang, Y.; Yu, L.; Wu, Y.; Li, G. Polymer Solar Cells With Enhanced Open-Circuit Voltage And Efficiency. *Nat. Photonics* **2009**, *3* (11), 649−653.
(44) Ding, Z.; Kettle, J.; Horie, M.; Chang, S.; Smith, G.; Shames, A.; Katz, E. Efficient Solar Cells Are More Stable: The Impact Of Polymer Molecular Weight On Performance Of Organic Photovoltaics. *J. Mater. Chem. A* **2016**, *4* (19), 7274−7280.
(45) Coffin, R. C.; Peet, J.; Rogers, J.; Bazan, G. C. Streamlined Microwave-Assisted Preparation Of Narrow-Bandgap Conjugated Polymers For High-Performance Bulk Heterojunction Solar Cells. *Nat. Chem.* **2009**, *1*, 657−661.
(46) 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 PC61BM Leads To High Efficiencies In Solar Cells. *Adv. Mater.* **2012**, *24*, 425−430.
(47) Li, W.; Yang, L.; Tumbleston, J. R.; Yan, L.; Ade, H.; You, W. Controlling Molecular Weight Of A High Efficiency Donor-Acceptor Conjugated Polymer And Understanding Its Significant Impact On Photovoltaic Properties. *Adv. Mater.* **2014**, *26*, 4456−4462.
(48) Menke, S.; Ran, N.; Bazan, G.; Friend, R. Understanding Energy Loss In Organic Solar Cells: Toward A New Efficiency Regime. *Joule* **2018**, *2* (1), 25−35.
(49) Ran, N.; Love, J.; Takacs, C.; Sadhanala, A.; Beavers, J.; Collins, S.; Huang, Y.; Wang, M.; Friend, R.; Bazan, G.; Nguyen, T. Harvesting The Full Potential Of Photons With Organic Solar Cells. *Adv. Mater.* **2016**, *28* (7), 1482−1488.
(50) Staple, D.; Oliver, P.; Hill, I. Derivation Of The Open-Circuit Voltage Of Organic Solar Cells. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89* (20), 205313.
(51) Keshtov, M.; Kuklin, S.; Khokhlov, A.; Osipov, S.; Radychev, N.; Godovskiy, D.; Konstantinov, I.; Chen, F.; Koukaras, E.; Sharma, G. Polymer Solar Cells Based Low Bandgap A1-D-A2-D Terpolymer Based On Fluorinated Thiadiazoloquinoxaline And Benzothiadiazole Acceptors With Energy Loss Less Than 0.5 Ev. *Org. Electron.* **2017**, *46*, 192−202.
(52) Keshtov, M.; Kuklin, S.; Radychev, N.; Nikolaev, A.; Ostapov, I.; Krayushkin, M.; Konstantinov, I.; Koukaras, E.; Sharma, A.; Sharma, G. New Low Bandgap Near-IR Conjugated D−A Copolymers For BHJ Polymer Solar Cell Applications. *Phys. Chem. Chem. Phys.* **2016**, *18* (12), 8389−8400.
(53) Wang, C.; Xu, X.; Zhang, W.; Bergqvist, J.; Xia, Y.; Meng, X.; Bini, K.; Ma, W.; Yartsev, A.; Vandewal, K.; Andersson, M.; Inganäs, O.; Fahlman, M.; Wang, E. Low Band Gap Polymer Solar Cells With Minimal Voltage Losses. *Adv. Energy Mater.* **2016**, *6* (18), 1600148.
(54) Li, S.; Ye, L.; Zhao, W.; Yan, H.; Yang, B.; Liu, D.; Li, W.; Ade, H.; Hou, J. A Wide Band Gap Polymer With A Deep Highest Occupied Molecular Orbital Level Enables 14.2% Efficiency In Polymer Solar Cells. *J. Am. Chem. Soc.* **2018**, *140* (23), 7159−7167.
(55) Leclerc, N.; Chávez, P.; Ibraikulov, O. A.; Heiser, T.; Lévêque, P. Impact of Backbone Fluorination on π-Conjugated Polymers in Organic Photovoltaic Devices: A Review. *Polymers* **2016**, *8* (1), 11.
(56) Zhang, S.; Qin, Y.; Uddin, M. A.; Jang, B.; Zhao, W.; Liu, D.; Woo, H. Y.; Hou, J. A Fluorinated Polythiophene Derivative with Stabilized Backbone Conformation for Highly Efficient Fullerene and Non-Fullerene Polymer Solar Cells. *Macromolecules* **2016**, *49* (8), 2993−3000.

(57) Ma, Y.; Kang, Z.; Zheng, Q. Recent advances in wide bandgap semiconducting polymers for polymer solar cells. *J. Mater. Chem. A* **2017**, *5* (5), 1860−1872.
(58) Son, H.; Wang, W.; Xu, T.; Liang, Y.; Wu, Y.; Li, G.; Yu, L. Synthesis Of Fluorinated Polythienothiophene-Co-Benzodithiophenes And Effect Of Fluorination On The Photovoltaic Properties. *J. Am. Chem. Soc.* **2011**, *133* (6), 1885−1894.
(59) Zhang, M.; Guo, X.; Zhang, S.; Hou, J. Synergistic Effect Of Fluorination On Molecular Energy Level Modulation In Highly Efficient Photovoltaic Polymers. *Adv. Mater.* **2014**, *26* (7), 1118−1123.
(60) Price, S.; Stuart, A.; Yang, L.; Zhou, H.; You, W. Fluorine Substituted Conjugated Polymer Of Medium Band Gap Yields 7% Efficiency In Polymer−Fullerene Solar Cells. *J. Am. Chem. Soc.* **2011**, *133* (20), 8057−8058.
(61) Carsten, B.; Szarko, J.; Son, H.; Wang, W.; Lu, L.; He, F.; Rolczynski, B.; Lou, S.; Chen, L.; Yu, L. Examining The Effect Of The Dipole Moment On Charge Separation In Donor−Acceptor Polymers For Organic Photovoltaic Applications. *J. Am. Chem. Soc.* **2011**, *133* (50), 20468−20475.
(62) Wang, Jiuxing; Xiao, Manjun; Chen, Weichao; Qiu, Meng; Du, Zhengkun; Zhu, Weiguo; Wen, Shuguang; Wang, Ning; Yang, Renqiang Extending π-Conjugation System with Benzene: An Effective Method To Improve the Properties of Benzodithiophene-Based Polymer for Highly Efficient Organic Solar Cells. *Macromolecules* **2014**, *47* (22), 7823−7830.
(63) Duan, C.; Willems, R. E. M.; van Franeker, J. J.; Bruijnaers, B. J.; Wienk, M. M.; Janssen, R. A. J. Effect of side chain length on the charge transport, morphology, and photovoltaic performance of conjugated polymers in bulk heterojunction solar cells. *J. Mater. Chem. A* **2016**, *4* (5), 1855−1866.
(64) Dyer-Smith, C.; Howard, I. A.; Cabanetos, C.; Labban, A. E.; Beaujuge, P. M.; Laquai, F. Interplay Between Side Chain Pattern, Polymer Aggregation, and Charge Carrier Dynamics in PBDTTPD: PCBM Bulk-Heterojunction Solar Cells. *Adv. Energy Mater.* **2015**, *5* (9), 1401778.
(65) Heckler, I.; Kesters, J.; Defour, M.; Madsen, M.; Penxten, H.; D'haen, J.; Mele, B. V.; Maes, W.; Bundgaard, E. The Influence of Conjugated Polymer Side Chain Manipulation on the Efficiency and Stability of Polymer Solar Cells. *Materials* **2016**, *9* (3), 181.
(66) Kim, Y. J.; Park, K. H.; jin Ha, J.; Chung, D. S.; Kim, Y.-H.; Park, C. E. The effect of branched versus linear alkyl side chains on the bulk heterojunction photovoltaic performance of small molecules containing both benzodithiophene and thienopyrroledione. *Phys. Chem. Chem. Phys.* **2014**, *16* (37), 19874−19883.