Computational Evolution of High-Performing Unfused Non-Fullerene Acceptors for Organic Solar Cells

Brianna L. Greenstein,[1] Danielle C. Hiener,[1] and Geoffrey R. Hutchison[1, a)]

Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh,

Pennsylvania 15260, United States

(Dated: 3 February 2022)

Materials optimization for organic solar cells (OSCs) is a highly active field, with many approaches using empirical experimental synthesis, computational brute-force approaches to screen candidates in a given subset of chemical space, or generative machine learning methods which often require significant training sets. While these methods may find high-performing materials, they can be inefficient and time-consuming. Genetic algorithms (GAs) are an alternative approach, allowing for the "virtual synthesis" of molecules and a prediction of their "fitness" for some property, with new candidates suggested based on good characteristics of previously generated molecules. In this work, a GA is used to discover high-performing unfused non-fullerene acceptors (NFAs) based on an empirical prediction of power conversion efficiency (PCE) and provides design rules for future work. The electron withdrawing/donating strength, as well as the sequence and symmetry of those units are examined. The utilization of a GA over a brute force approach resulted in speedups up to 1.8×10^{12}. New types of units not frequently seen in OSCs are suggested, and in total 5,426 NFAs are discovered with the GA. Of these, 1,087 NFAs are predicted to have a PCE greater than 18%, which is roughly the current record efficiency. While the symmetry of the sequence showed no correlation with PCE, analysis of the sequence arrangement revealed that higher performance can be achieved with a donor core and acceptor end groups. Future NFA designs should consider this strategy as an alternative to the current A-D-A′-D-A architecture.

---

a)geoffh@pitt.edu; https://hutchisonlab.org/; Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, United States

I. INTRODUCTION

Organic solar cells (OSCs) offer the vast tailorability of organic molecular, oligomer, and polymer motifs to convert light into electrons. To improve the performance of OSCs, there has been much research focusing on the optimization of p-type donor materials, with conjugated organic polymers being the most popular[1–4]. For a p-type donor material to perform well, however, it needs to be paired with a complementary n-type acceptor, frequently small molecules. Multiple properties can determine whether the materials are a good match, such as strong absorption of different wavelengths, energy level alignment for efficient exciton splitting and low energy loss, and optimal solubility to form suitable domains within the active layer.

Up until recently, fullerenes were the most frequently used acceptor due to their excellent electron transport and good energy alignment with commonly used donors[5]. They have drawbacks when acting as acceptors, however, including weak light absorption, complex synthesis, limited tunability, and poor photostability[6,7]. As an alternative, non-fullerene small molecules emerged, showing highly tunable optical and electronic properties, strong absorption of visible and near-infrared (NIR) light, and low production cost[6–8]. The most common design for NFAs is joining electron-withdrawing units and electron donating units, due to their "push-pull" effect and enhanced intramolecular charge transfer (ICT). This effect has been shown to improve the PCE by broadening the absorption spectrum and improving charge mobility[9,10].

Huge strides have been made in the last two years due to the emergence of the Y-series acceptors[11,12], with the PCE surpassing 18%[13,14], much higher than the highest PCE of 11.7% for fullerene binary devices[15,16]. Due to their extended conjugation, they have narrower optical band gaps, allowing the absorption of visible to NIR light[17]. The general configuration is A-DA$'$D-A, with the fused backbone allowing for planarity which increases the crystallinity and π orbital stacking within the device[18]. While fused NFAs have shown great promise, they have some drawbacks that limit their commercial potential. To form the fused backbone structure requires complex synthesis, often with 5-15 steps[19,20]. This difficult synthesis process increases the cost and results in low yield. Recent studies also show poor stability of fused ring NFA devices, due to their tendency to undergo photo-oxidation when exposed to light and air[21–25]. A recent strategy to mitigate some of these issues is to use unfused analogs of these acceptors. Unfused NFAs are similarly designed with electron-withdrawing (acceptor) and electron-donating (donor) organic conjugated units, but are separated by single bonds instead of fused rings. This allows a much

simpler synthesis (around 2-4 steps), higher yield, and generally better photostability[22,26,27].

It has been previously shown that the arrangement of acceptor and donor units in conjugated polymers affects their optoelectronic properties[28–30]. While A-DA′D-A is the most common sequence, attributed to the high success of this sequence in Y-series NFAs, other sequences are emerging as strong contenders, with asymmetry showing much promise[19]. Asymmetric NFAs may have larger dipole moments, broader absorption, higher electron mobility, and lower crystallinity than their symmetric counterparts[31–35].

The current highest PCE for organic solar cells is 19%[36], much lower than the theoretical limit of ~33% predicted by Shockley and Queisser[37] for single-junction devices. With the perfect combination of materials, we should be able to far exceed the current highest experimentally-produced PCE. However, finding these materials is similar to finding two needles in different haystacks. Not only do we need optimal acceptor and donor materials, but they need to be perfectly complementary for energy level alignment and light absorption to minimize energy losses. To discover new materials, the use of computational screening methods has emerged as a promising alternative to experimental screening.[38–50] The molecular chemical space for organic molecules is estimated to be 10^{18}-10^{200} compounds[51,52], which is impossible to evaluate experimentally. Computational screening allows for "virtual synthesis", a quick and relatively cheap method to explore promising materials and create *in silico* molecular libraries[41,45]. One type of computational screening is high-throughput virtual screening (HTVS), where every candidate in a library is screened[38,40,53]. The largest HTVS for organic photovoltaics (OPVs) is the Harvard Clean Energy Project, which used electronic structure calculations to evaluate the PCE for 2.6 million conjugated donor molecules and 51,000 NFAs using Scharber's model[47,49]. However, HTVS and other brute-force searches are inefficient and screen a massive number of molecules unlikely to be high-performing candidates.

An alternative technique is the use of inverse design strategies, which set an ideal target value of a property as a starting point, and use algorithms to design new materials that optimize that property. One such technique is a genetic algorithm (GA), a type of evolutionary algorithm based on natural selection in which the fittest candidates pass to the next generation. In a GA designed for materials discovery, a population of potential molecules is generated that optimize a given property and their suitability is evaluated with a "fitness function". Considering many fitness functions require quantum chemical calculations, low cost semi-empirical methods should be used for fast and efficient chemical searches. Since the goal of this work is to evaluate potential NFAs, the Organic Photovoltaic Efficiency Predictor using simplified time-dependent density functional
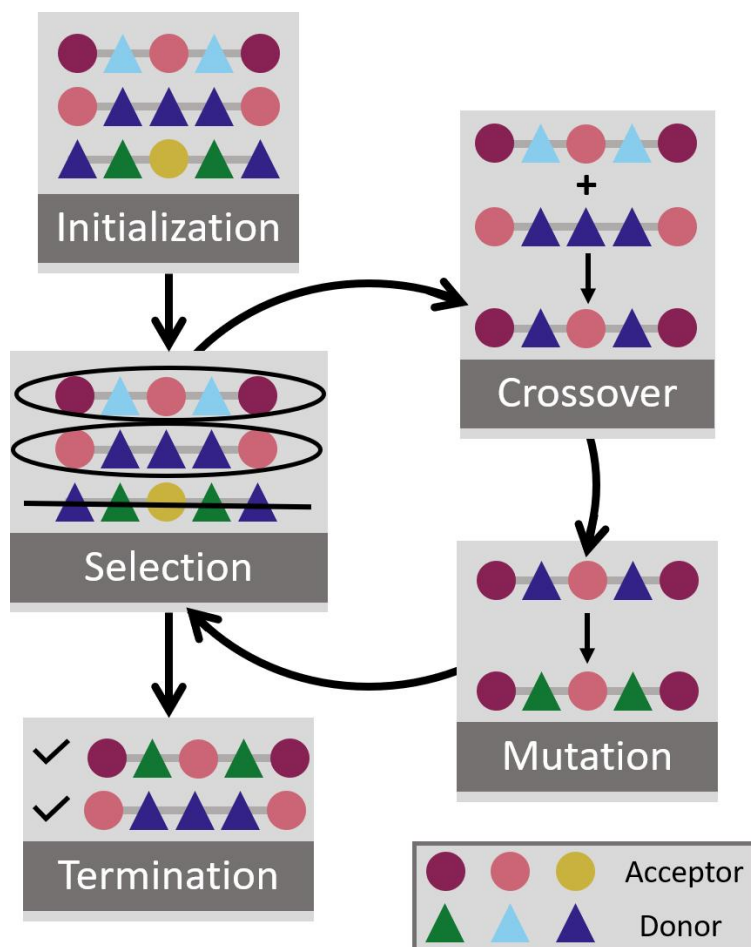
FIG. 1. A simplified genetic algorithm scheme displaying the selection, crossover, and mutation operations to optimize non-fullerene acceptors composed of electron-accepting (acceptor) and electron-donating (donor) units.

theory (OPEP/sTD-DFT) model proposed in our earlier work[54] is used as the fitness function to quickly evaluate the PCE of NFAs. This model requires the difference in energy between the HOMO and HOMO-1 and the sum of the oscillator strengths of the acceptor, obtained through sTD-DFT calculations. Other parameters for the model include properties of the donor material, with the polymer PBDB-T-SF producing the highest PCEs out of 20 donors. Thus, PBDB-T-SF is selected as the donor material to pair with new NFAs in the GA. The best candidates of each generation are selected to undergo crossover and mutation operations to ensure the exploration of

nearby chemical space, producing children for the next generation. This cycle of parent selection, crossover, mutation, and evaluation of fitness continues until the target property has been satisfied, or after a certain number of generations, as seen in Fig. 1. The main advantage of using a GA is to learn from each successive generation to design even better candidates. Our group previously used a GA to efficiently screen donor polymers for fullerene-based OPVs, producing hundreds of promising targets and discovering design principles related to the sequence of the monomers[30].

In this work, we report new design principles for NFAs discovered by screening thousands of molecules with a genetic algorithm. The types of building blocks, sequence, and symmetry are analyzed, and new NFAs are proposed that can reach PCEs over 22%.

II. RESULTS AND DISCUSSION

Four versions of the genetic algorithm were implemented to understand the structure-property relationships and provide design rules for optimized NFAs. In version 1, the GA searched for high-performing NFAs with an A-D-A′-D-A structure. This is a commonly used sequence, especially it being the unfused analogue of the high performing y-series acceptors. The acceptor and donor units comprising these NFAs were chosen from a library of units found in high-performing NFAs. This library was compiled by examining hundreds of NFAs previously published in literature. Within this library, the units were classified as either terminal or core units, where terminal refers to end-capping units and the core can be anywhere in NFA. The library for version 1 contains 104 acceptor units and 88 donor units (Fig. 2).

To expand the search space, the library was extended to small conjugated organic molecules not frequently employed in NFAs. Previous work in our group utilized a GA to search for donor materials for OSCs[30], which uses units similarly found in NFAs. In versions 2 and 3, the unit library was expanded to 949 donor units and 295 acceptor units (Fig. 2). In addition to expanding the search space, the arrangements of the units are analyzed. In version 2, the sequence was restricted to symmetrical arrangements and in version 3, any sequence was permitted. To ensure that most of the top candidates were found, a fourth version was performed whose dataset comprised of the units frequently selected in versions 1 and 2, and allowed any sequence of these units.

The GA was allowed to run for 100 generations, or until the best PCE remained steady for approximately 30 generations, whichever occurred later. Fig. 3 shows the best PCE among 32 NFAs from each generation. As expected, the PCE increases as the generations evolve, confirming

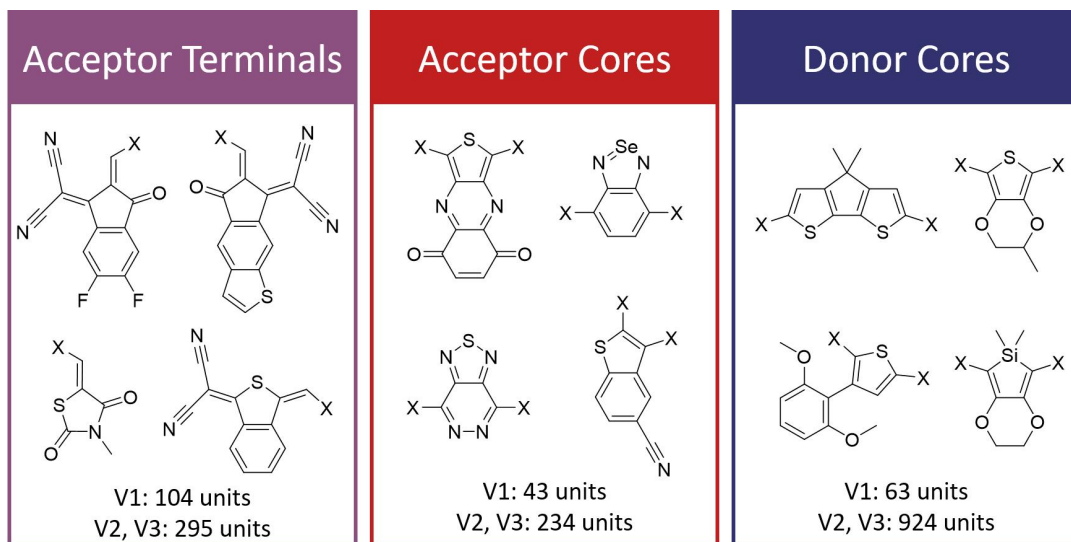| Acceptor Terminals	Acceptor Cores	Donor Cores
V1: 104 units V2, V3: 295 units	V1: 43 units V2, V3: 234 units	V1: 63 units V2, V3: 924 units

FIG. 2. Examples of building block units used in versions 1 (V1), 2 (V2), and 3 (V3), with terminal units referring to units used as end groups on the NFA, and cores referring to use anywhere in the molecule. Cores are included in the terminal subset.

that the GA works and is efficiently exploring the chemical space. Version 1, with the dataset containing units already used in NFAs and restricted to the sequence A-D-A′-D-A, found the best performing NFA with a PCE of 20.9% after 26 generations. Version 2, containing the expanded dataset and restricted to symmetrical sequences, found the top performing NFA with a PCE of 22.9% after 39 generations. Once the sequence was unrestricted in versions 3 and 4, the best PCE took longer to stabilize, mainly due to the increase in magnitude of the size of the search space. Version 3 found the best NFA after 84 generations with a PCE of 20.2%, while Version 4 found the best NFA after 70 generations of 22.3%. As the search space increases, the GA requires more generations to efficiently search for top candidates. With the current record PCE of 15.2% for unfused NFAs[55], these results show tremendous potential. Utilization of a GA instead of a brute-force approach shows a massive speedup, up to a magnitude of 10^{12}, seen in Table I. While the search space for each version is massive, on the order of 10^5, 10^7, 10^{15}, 10^{11}, for versions 1-4, respectively, tiny percentages of these areas had to be traversed to find high-performing NFAs.

Examination of the sequences selected in versions 2-4 shows a preference for donor cores and acceptor end groups. Over 80% of the NFAs in version 2 had either the A-D-D′-D-A or A-A′-D-A′-A sequence (Fig. S1). Version 3 shows over 80% of the NFAs have one central donor unit flanked by two acceptor units, while version 4 shows over 60% of the NFAs discovered have
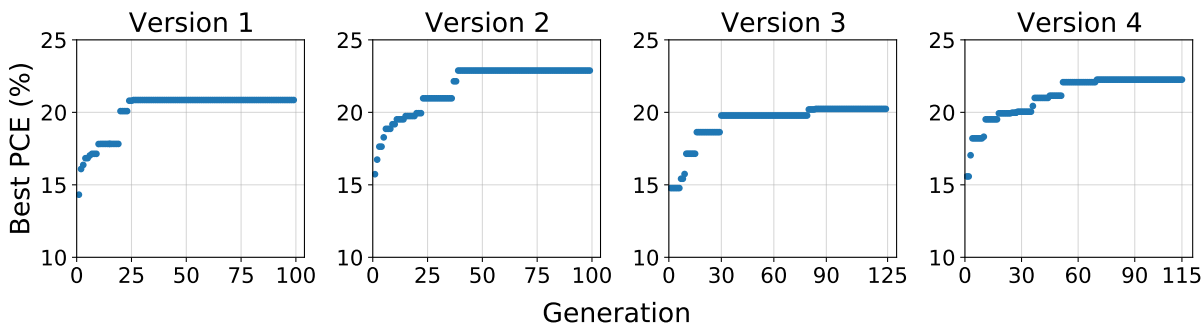
6

FIG. 3. Plots of the PCE of the best scoring candidate for each generation for versions 1-4. All PCEs were calculated using OPEP/sTD-DFT with the donor PBDB-T-SF

TABLE I. Performance of all four versions of the GA, including the highest PCE found, the percentage of search space examined before the best PCE was found, and the speedup of using a GA over a brute force approach (search space/ total NFAs analyzed)

Version	Best PCE	Search Space	Speedup
1	20.9%	2.8×10^5	733
2	22.9%	6.4×10^7	1.1×10^5
3	20.2%	2.1×10^{15}	1.8×10^{12}
4	22.3%	1.7×10^{11}	1.7×10^8

three donor core units and acceptor terminal units (Figs. S2a and S3a). Similar results within all three versions suggests a donor core and acceptor end groups are essential for improving the PCE. The symmetry of the sequence was also examined. Around 70% of the sequences in version 3 were asymmetric, while around 70% of sequences in version 4 were symmetric (Figs. S2b and S3b). While there is no clear correlation between the symmetry of the sequence and PCE, it is interesting to note that the top performers from version 3 were symmetric while the top performers from version 4 were asymmetric.

The types of units frequently selected were explored to examine if current experimental NFAs efficiently select promising building-blocks. Fig. 4 demonstrates that there are multiple units that appeared in hundreds or even thousands of NFAs throughout the GA. The top ten units most commonly selected as terminal groups all originated from the initial dataset used in version 1. This indicates that experimental NFAs currently employ the top high-performing end groups. However,

out of the top ten core units, eight of them are from the expanded dataset of units not currently utilized in NFAs. This suggests new types of units to be incorporated into future NFAs.
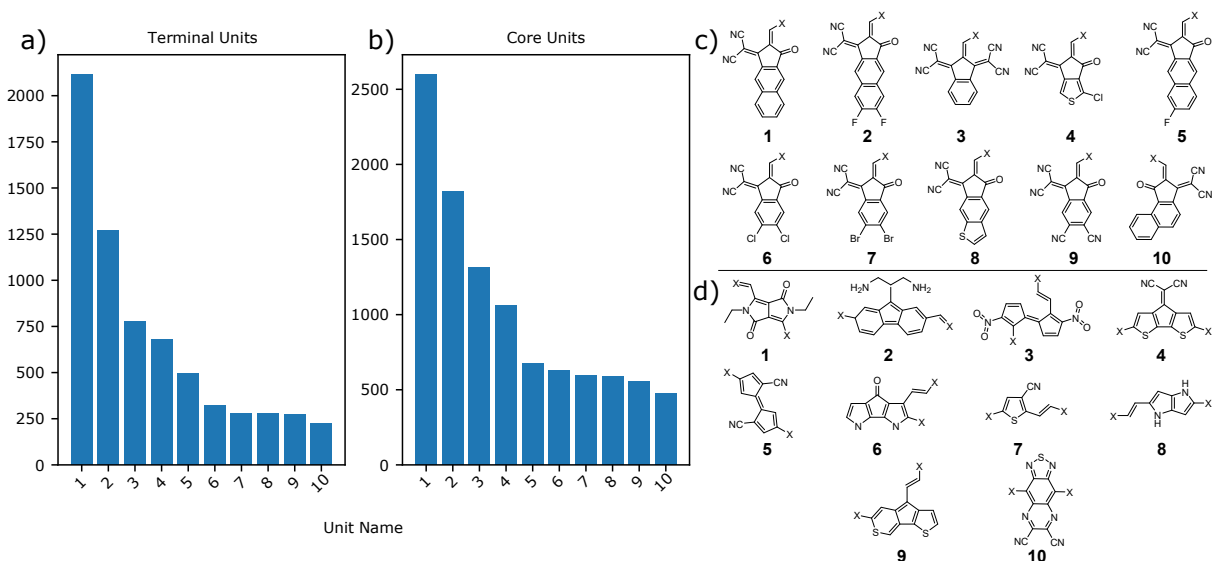

FIG. 4. The top occurring monomer units. a) Frequency of the top 10 terminal units (one point of attachment, can only be end units) throughout all four GA versions. b) Frequency of top 10 core units (two points of attachment, can be used anywhere in the molecule) throughout all four versions. c) Molecular structures of the top 10 terminal units seen in a). d) Molecular structures of the top 10 core units found in b).

To understand the impact of the electron-withdrawing/donating strength of the building units on the PCE, the HOMO energy of each unit is examined. In Fig. 5, the HOMO energy levels of the top five candidates from each version are examined. Thiophene was selected as a middle-ground with a monomer HOMO of -6.55 eV, allowing the classification of acceptors to be units with a HOMO below thiophene, and donors as above. The deeper the color, the stronger the electron-withdrawing or donating strength. When the sequence was not restricted (versions 2-4), the top performing NFAs all contained similar electronic sequences. The results suggest that having NFAs with acceptor end groups and donor cores can result in better performing NFAs. To further verify this, 50 high-performing NFAs from version 1 with the A-D-A′-D-A sequence were compared to their A-D-D-D-A analogues (Fig S4). If there were multiple NFAs from version 1 that would result in the same A-D-D-D-A sequence (same terminal acceptor and donor core), then the NFA with the highest PCE was selected for comparison, since this is the theoretical highest PCE analyzed for that sequence. When comparing the PCE, 57.1% of the NFAs showed better performance with the A-D-D-D-A sequence. This is an encouraging result, since it negates the need to limit NFAs to the

8

alternating A-D sequence and instead attach three donor units together as a synthetic alternative. Additionally, the overall top performers from version 2 all have a strong donating core, suggesting charge transfer to the end units. This agrees with our understanding of the current design rules of $\pi - \pi$ stacking with the end units[56]. Results from versions 3 and 4 support this as well and show that it is not necessary to have strongly withdrawing/donating units to obtain a high PCE.

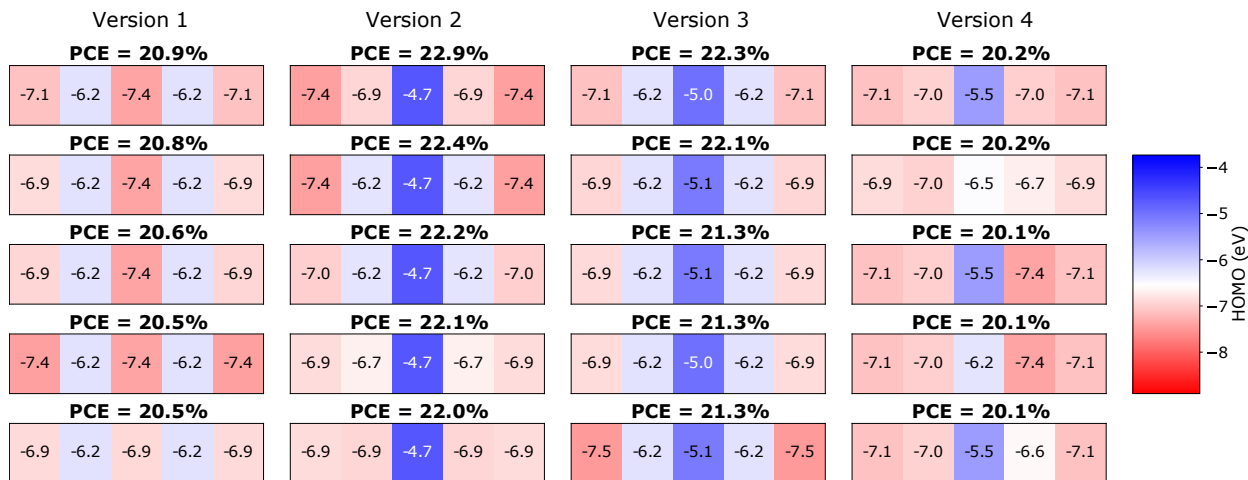|	Version 1					Version 2					Version 3					Version 4			
PCE = 20.9%					**PCE = 22.9%**					**PCE = 22.3%**					**PCE = 20.2%**				
-7.1	-6.2	-7.4	-6.2	-7.1	-7.4	-6.9	-4.7	-6.9	-7.4	-7.1	-6.2	-5.0	-6.2	-7.1	-7.1	-7.0	-5.5	-7.0	-7.1
PCE = 20.8%					**PCE = 22.4%**					**PCE = 22.1%**					**PCE = 20.2%**				
-6.9	-6.2	-7.4	-6.2	-6.9	-7.4	-6.2	-4.7	-6.2	-7.4	-6.9	-6.2	-5.1	-6.2	-6.9	-6.9	-7.0	-6.5	-6.7	-6.9
PCE = 20.6%					**PCE = 22.2%**					**PCE = 21.3%**					**PCE = 20.1%**				
-6.9	-6.2	-7.4	-6.2	-6.9	-7.0	-6.2	-4.7	-6.2	-7.0	-6.9	-6.2	-5.1	-6.2	-6.9	-7.1	-7.0	-5.5	-7.4	-7.1
PCE = 20.5%					**PCE = 22.1%**					**PCE = 21.3%**					**PCE = 20.1%**				
-7.4	-6.2	-7.4	-6.2	-7.4	-6.9	-6.7	-4.7	-6.7	-6.9	-6.9	-6.2	-5.0	-6.2	-6.9	-7.1	-7.0	-6.2	-7.4	-7.1
PCE = 20.5%					**PCE = 22.0%**					**PCE = 21.3%**					**PCE = 20.1%**				
-6.9	-6.2	-6.9	-6.2	-6.9	-6.9	-6.9	-4.7	-6.9	-6.9	-7.5	-6.2	-5.1	-6.2	-7.5	-7.1	-7.0	-5.5	-6.6	-7.1

FIG. 5. Plot of the HOMO (eV) of each unit in the top 5 NFAs discovered in each version. Thiophene monomer is used as the midpoint (HOMO = -6.55 eV), with blue indicating electron donors and red as electron acceptors. The deeper the color, the stronger the electron withdrawing/donating strength. All PCEs were calculated using OPEP/sTD-DFT with the donor PBDB-T-SF

Throughout all four versions, 5,426 unique NFAs were analyzed. Of these, 1,087 have a PCE above 18% and of those, 159 have a PCE above 20%. Fig. 6a is a histogram of all NFAs discovered, showing the average PCE of 15.8%. This distribution can be split up into each version's contributions, as seen in the kernel density estimate plot in Fig. 6b. Version 3 had the widest range of PCEs, which was anticipated due to having the largest chemical search space. Version 1 had the most narrow distribution, as expected due to the smallest search space and less drastic modifications in the GA. Version 4 discovered the largest percentage of the NFAs above 18%, validating the purpose of this version to find the top candidates using the top performing units from previous versions.

Optical and electronic properties of all NFAs discovered were analyzed. Fig. S5 shows the optical band gap and its predicted oscillator strength of every NFA examined, with NFAs with a PCE above 20% colored. There is a wide range of band gaps with varying oscillator strengths,
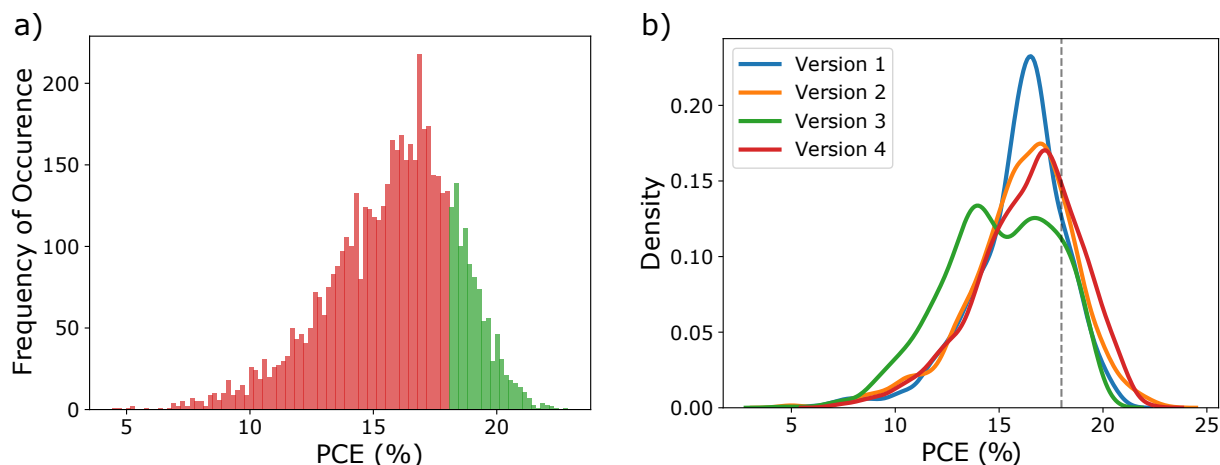
FIG. 6. (a) Histogram of all NFAs discovered, with green bars indicative of above 18% PCE (b) Kernel density estimate plot of the PCE in all four versions, with the dashed line indicating PCE of 18%. All PCEs were calculated using OPEP/sTD-DFT with the donor PBDB-T-SF

indicating there is no optimal band gap for high performance. The fundamental band gap of all NFAs was analyzed in Fig S6. All high-performing NFAs have a fundamental band gap between 2-3.5 eV. This small band gap range is consistent with the idea of small band gap NFAs absorbing in lower energy regions and complementary to a donors absorption, therefore increasing the PCE. This range although is still relatively large, indicating there is leeway in the ideal fundamental band gap range for the design of new NFAs.

The top six performing NFAs were examined in Fig. 7. The NFAs displayed in Figs. 7 a, b, d, and e were discovered in version 2, while the NFAs in Figs. 7c and f were discovered in version 4. The OPEP fitness function scored these NFA candidates highly on two properties: the sum of the oscillator strengths and the difference in energy between the HOMO and HOMO-1 energies. Examination of the absorption spectra shows strong absorption in the UV and high-energy visible regions for all top NFAs, with NFAs c and d showing broad absorption across the visible and near-IR region, a highly desired property in photovoltaics. The difference in energy between HOMO and HOMO-1 is conceptually connected to the density of states and delocalization of holes, and is maximized in each of these NFAs as well. Finally, we note that the central core unit seen in Figures 7 a, b, c, and d, breaks the aromaticity with an sp^3-hybridized carbon and a non-conjugated side-chain. At first, we believed the monomer was incorrectly specified to the GA, but we note that several other top candidates exhibit broken aromaticity or conjugation. These motifs may modulate the electronic structure, and the side chain appears to serve as a conformational lock on
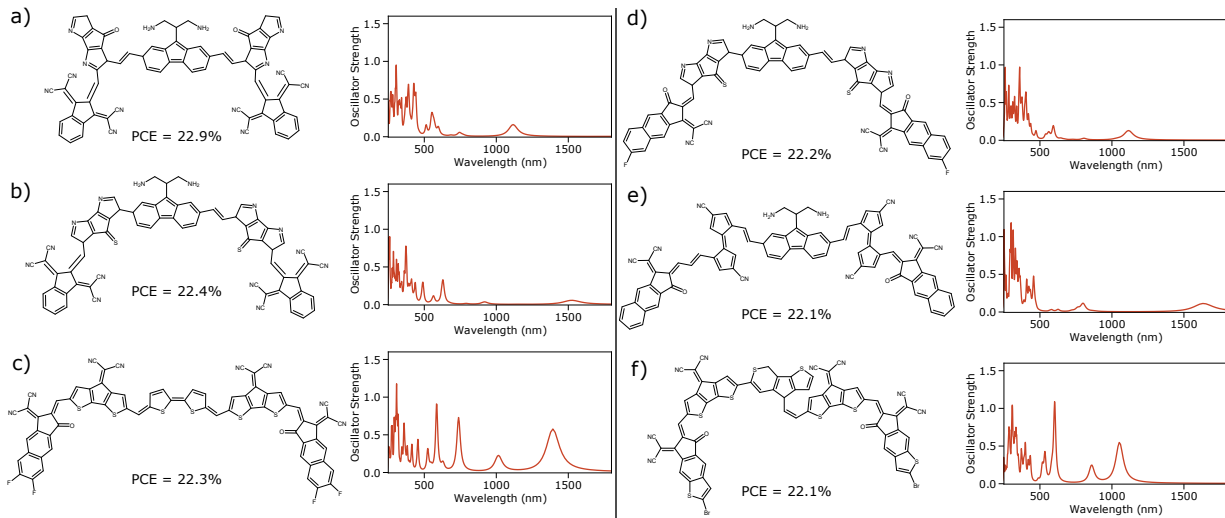
the neighboring monomers.

FIG. 7. Best 6 NFAs discovered using the GA, with the PCE calculated using OPEP/sTD-DFT with the donor PBDB-T-SF. Note that surprisingly, most of these structures do not retain aromaticity in all units.

III. CONCLUSION

In this work, the chemical design rules for high-performing NFAs were established. Through the utilization of a genetic algorithm, 5,426 NFAs were discovered, of which 1,087 are predicted to have a PCE above 18% and 159 above 20%. The results indicate that the best sequence for high-PCE is A-D-D-D-A or A-A-D-A-A, with a donor core and acceptor end groups. There is no significant difference in performance if the sequence is symmetric or asymmetric. When selecting individual building blocks, current terminal acceptor units employed in common NFAs are found to be top performers in the GA candidates. For core units, however, further exploration is needed, with the units suggested here providing some interesting starting points. The acceptor terminal groups can be weakly electron-withdrawing, and the donor core performs well when it is strongly electron-donating. Overall, this work suggests that new design rules for unfused NFA materials can yield further improvements in power conversion efficiency.

IV. COMPUTATIONAL DETAILS

A. Dataset

For version 1, a dataset of 192 organic conjugated electron-withdrawing/accepting units with three fused rings or less was constructed by examining high-performing NFAs reported in the literature. Versions 2 and 3 used the dataset from version 1 with the addition of 1,221 units from a previous GA project in our group[57]. Structures were generated for each unit and optimized with MMFF94[58] or UFF[59] force fields before additional optimization with GFN2-xTB[60–62]. Further geometry optimization was performed with Orca 4.2.0.[63] using B3LYP-D3BJ/def2-SVP[64–67]. To classify the units as electron acceptors or donors, their HOMO values were compared to thiophene's HOMO. If the unit's HOMO is lower than thiophene's, it is considered an acceptor; if the HOMO is higher, it is a donor. Thiophene was chosen as an arbitrary midpoint since it is a common organic semiconductor unit and is typically found as π-bridges in OPV materials, where its HOMO is not too deep or shallow. The dataset was grouped into four categories: acceptor terminal, acceptor core, donor terminal, and donor core, where the terminal refers to one point of attachment, so it must be an end unit, and the core refers to two points of attachment and can be used as either a terminal unit or a central unit.

B. Fitness Function

The OPEP/sTD-DFT PCE[54] model was selected as the fitness function due to the high accuracy and low computational cost of the sTD-DFT method. This model has six terms, four of which are from the donor and two from the acceptor. For evaluation of fitness, 22 donors were analyzed with sTD-DFT to optimize the donor term, resulting in PBDB-T-SF being selected as the best performer. The two terms for the acceptor are the sum of the oscillator strengths and the difference in energy between the HOMO and HOMO-1 eigenvalues. To evaluate these properties for each candidate acceptor, force field and GFN2-xTB geometry optimization were performed, and the electronic and optical properties are calculated with sTD-DFT[68–70] using CAM-B3LYP with def2-SVP[66] basis set in Orca 4.2.0. The maximum energy threshold is set to 5.0 eV, since the majority of the solar spectrum we are interested in for solar cell applications is below this energy.

C. Version 1

The GA is constructed to design high-performing unfused NFAs with an A-D-A′-D-A sequence, meaning the central A' is a different unit than the terminal A units and results in a symmetrical structure. In this GA, 32 NFAs are chosen as the initial population, with each NFA constructed from random A and D units. In each generation, the PCE of each candidate is evaluated and the population is ranked from the highest to lowest PCE with the top 50% selected as parents for the next generation. Each successive generation contains 32 NFAs, 16 are the parents from the previous generation, and 16 are newly formed NFAs created using crossover and mutation operators on the parents. These children are constructed through a crossover operation that randomly chooses two units to keep from one random parent and one to keep from another. The crossover operation is repeated until 16 new children are constructed and added to the generation. The mutation rate is set to 40%, and if selected the mutation operator will replace one of the three types of units with a new building block, chosen at random. Termination occurs when the highest PCE remains the top performer for 30 generations or 100 generations have occurred, whichever is later.

D. Versions 2 and 3

To further expand the search space, the sequence restriction of A-D-A′-D-A was lifted. Version 2 allowed for symmetrical sequences while version 3 allowed any sequence. Both versions used the expanded dataset. To initialize the first generation, sequences and units were chosen randomly. For version 3, sequences are first randomly selected to be symmetric or asymmetric, as required due to a large number of possible asymmetric sequences, and then a random sequence is selected from within the given symmetry subgroup. During the crossover operation, the units and sequences of two parents are crossed, however, if there is not enough of the required unit types for that sequence, then the units are reused. For the mutation operator, either the sequence or unit type is altered. If the sequence is chosen, the units are rearranged to fit the sequence, however, if there are not enough units to satisfy the sequence, random units are pulled from the initial dataset. If swapping a unit is chosen, a randomly chosen unit is replaced with a random unit from the initial dataset, ensuring the unit is of the same type as the one it is replacing so that the sequence does not change.

E. Version 4

The dataset of units were created by selecting units from versions 1 and 2 that were used in a NFA for more than 5 times. The sequence was unrestricted and the GA operations were the same as version 3. Unfortunately, an error occurred during this version due to a mismatch in labeling between the units and their corresponding designation of acceptor or donor. This resulted in the GA occasionally selecting a different sequence than intended and it being mislabeled. For example, if during the crossover step, the sequence A-D-A′-D-A was selected to pass to the next generation. However, the unit it selected as D may have been mislabeled and is actually an acceptor. Therefore, while the GA was not perfect, the results still indicate high-performing NFAs were discovered, and all figures shown have the true sequence.

V. DATA AVAILABILITY

The data that supports the findings of this study are openly available at: `https://github.com/hutchisonlab/GA_unfused_NFAs`

VI. SUPPLEMENTARY MATERIAL

See supplementary material for analysis of sequences in versions 2-4, symmetry of sequences for versions 3 and 4, comparison of the A-D-A′-D-A sequence to A-D-D-D-A, optical and fundamental band gaps of every NFA discovered.

CONFLICTS OF INTEREST

There are no conflicts to declare.

ACKNOWLEDGMENTS

REFERENCES

[1] Z. Li, K. Jiang, G. Yang, J. Y. L. Lai, T. Ma, J. Zhao, W. Ma, and H. Yan, "Donor polymer design enables efficient non-fullerene organic solar cells," Nature Communications **7**, 13094 (2016).

[2] C. An, Z. Zheng, and J. Hou, "Recent progress in wide bandgap conjugated polymer donors for high-performance nonfullerene organic photovoltaics," Chemical Communications **56**, 4750–4760 (2020).

[3] C. Cui and Y. Li, "High-performance conjugated polymer donor materials for polymer solar cells with narrow-bandgap nonfullerene acceptors," Energy & Environmental Science **12**, 3225–3246 (2019).

[4] L. Sun, X. Xu, S. Song, Y. Zhang, C. Miao, X. Liu, G. Xing, and S. Zhang, "Medium-Bandgap Conjugated Polymer Donors for Organic Photovoltaics," Macromolecular Rapid Communications **40**, 1900074 (2019).

[5] A. Pivrikas, N. S. Sariciftci, G. Juška, and R. Österbacka, "A review of charge transport and recombination in polymer/fullerene organic solar cells," Progress in Photovoltaics: Research and Applications **15**, 677–696 (2007).

[6] A. Armin, W. Li, O. J. Sandberg, Z. Xiao, L. Ding, J. Nelson, D. Neher, K. Vandewal, S. Shoaee, T. Wang, H. Ade, T. Heumüller, C. Brabec, and P. Meredith, "A History and Perspective of Non-Fullerene Electron Acceptors for Organic Solar Cells," Advanced Energy Materials **11**, 2003570 (2021).

[7] C. Duan and L. Ding, "The new era for organic solar cells: non-fullerene small molecular acceptors," Science Bulletin **65**, 1231–1233 (2020).

[8] P. Meredith, W. Li, and A. Armin, "Nonfullerene Acceptors: A Renaissance in Organic Photovoltaics?" Advanced Energy Materials **10**, 2001788 (2020).

[9] W. Li, L. Ye, S. Li, H. Yao, H. Ade, and J. Hou, "A High-Efficiency Organic Solar Cell Enabled by the Strong Intramolecular Electron Push-Pull Effect of the Nonfullerene Acceptor," Advanced Materials **30**, 1707170 (2018).

[10] H. Lai, L. Liu, N. Zheng, L. Han, and F. He, "Push or Pull Electrons: Acetoxy and Carbomethoxy-Substituted Isomerisms in Organic Solar Cell Acceptors," The Journal of Physical Chemistry Letters **12**, 4666–4673 (2021).

[11] S. Li, C.-Z. Li, M. Shi, and H. Chen, "New Phase for Organic Solar Cell Research: Emergence of Y-Series Electron Acceptors and Their Perspectives," ACS Energy Letters **5**, 1554–

1567 (2020).

[12] X. Li, H. Lu, and W. Zhu, "Recent Advances in Y6-Based Semiconductors: Performance in Solar Cells, Crystallography, and Electronic Structure," ChemPlusChem **86**, 700–708 (2021).

[13] Q. Liu, Y. Jiang, K. Jin, J. Qin, J. Xu, W. Li, J. Xiong, J. Liu, Z. Xiao, K. Sun, S. Yang, X. Zhang, and L. Ding, "18% Efficiency organic solar cells," Science Bulletin **65**, 272–275 (2020).

[14] C. He, Z. Bi, Z. Chen, J. Guo, X. Xia, X. Lu, J. Min, H. Zhu, W. Ma, L. Zuo, and H. Chen, "Compromising charge generation and recombination with asymmetric molecule for high-performance binary organic photovoltaics with over 18% certified efficiency," Advanced Functional Materials , 2112511 (2022).

[15] J. Zhao, Y. Li, G. Yang, K. Jiang, H. Lin, H. Ade, W. Ma, and H. Yan, "Efficient organic solar cells processed from hydrocarbon solvents," Nature Energy **1**, 1–7 (2016).

[16] F. Zhang, O. Inganäs, Y. Zhou, and K. Vandewal, "Development of polymer–fullerene solar cells," National Science Review **3**, 222–239 (2016).

[17] Y. Chen, T. Liu, L.-K. Ma, W. Xue, R. Ma, J. Zhang, C. Ma, H. K. Kim, H. Yu, F. Bai, K. S. Wong, W. Ma, H. Yan, and Y. Zou, "Alkoxy substitution on idt-series and y-series non-fullerene acceptors yielding highly efficient organic solar cells," Journal of Materials Chemistry A **9**, 7481–7490 (2021).

[18] J. f. Hai, S. Luo, H. Yu, H. Chen, Z. Lu, L. Li, Y. Zou, and H. Yan, "Achieving ultra-narrow bandgap non-halogenated non-fullerene acceptor via vinylene π-bridges for efficient organic solar cells," Materials Advances , 2132–2140 (2021).

[19] J. Cao, S. Qu, L. Yang, H. Wang, F. Du, J. Yu, and W. Tang, "Asymmetric simple unfused acceptor enabling over 12% efficiency organic solar cells," Chemical Engineering Journal **412**, 128770 (2021).

[20] J. Yuan, Y. Zhang, L. Zhou, G. Zhang, H.-L. Yip, T.-K. Lau, X. Lu, C. Zhu, H. Peng, P. A. Johnson, M. Leclerc, Y. Cao, J. Ulanski, Y. Li, and Y. Zou, "Single-Junction Organic Solar Cell with over 15% Efficiency Using Fused-Ring Acceptor with Electron-Deficient Core," Joule **3**, 1140–1151 (2019).

[21] Y. Li, T. Li, and Y. Lin, "Stability: next focus in organic solar cells based on non-fullerene acceptors," Materials Chemistry Frontiers **5**, 2907–2930 (2021).

[22] Z.-X. Liu, Z.-P. Yu, Z. Shen, C. He, T.-K. Lau, Z. Chen, H. Zhu, X. Lu, Z. Xie, H. Chen, and C.-Z. Li, "Molecular insights of exceptionally photostable electron acceptors for organic photovoltaics," Nature Communications **12**, 3049 (2021).

[23] R. Lv, S. Geng, S. Li, F. Wu, Y. Li, T. R. Andersen, Y. Li, X. Lu, M. Shi, and H. Chen, "Influences of Quinoid Structures on Stability and Photovoltaic Performance of Nonfullerene Acceptors," Solar RRL **4**, 2000286 (2020).

[24] P. Cheng and X. Zhan, "Stability of organic solar cells: challenges and strategies," Chemical Society Reviews **45**, 2544–2582 (2016).

[25] L. Duan and A. Uddin, "Progress in Stability of Organic Solar Cells," Advanced Science **7**, 1903259 (2020).

[26] X. Zhang, L. Qin, J. Yu, Y. Li, Y. Wei, X. Liu, X. Lu, F. Gao, and H. Huang, "High-Performance Noncovalently Fused-Ring Electron Acceptors for Organic Solar Cells Enabled by Noncovalent Intramolecular Interactions and End-Group Engineering," Angewandte Chemie **133**, 12583–12589 (2021).

[27] L. Zhang, S. Tu, W. Wang, and Q. Ling, "Brominated small-molecule acceptors with a simple non-fused framework for efficient organic solar cells," ACS Applied Energy Materials **4**, 4805–4814 (2021).

[28] I. Y. Kanal, J. S. Bechtel, and G. R. Hutchison, "Sequence Matters: Determining the Sequence Effect of Electronic Structure Properties in π-Conjugated Polymers," **1170**, 379–393 (2014).

[29] S. Zhang, G. R. Hutchison, and T. Y. Meyer, "Sequence Effects in Conjugated Donor-Acceptor Trimers and Polymers," Macromolecular Rapid Communications **37**, 882–887 (2016).

[30] I. Y. Kanal, S. G. Owens, J. S. Bechtel, and G. R. Hutchison, "Efficient Computational Screening of Organic Polymer Photovoltaics," The Journal of Physical Chemistry Letters **4**, 1613–1623 (2013).

[31] Q. Guo, J. Lin, H. Liu, X. Dong, X. Guo, L. Ye, Z. Ma, Z. Tang, H. Ade, M. Zhang, and Y. Li, "Asymmetrically noncovalently fused-ring acceptor for high-efficiency organic solar cells with reduced voltage loss and excellent thermal stability," Nano Energy **74**, 104861 (2020).

[32] W. Hu, X. Du, W. Zhuang, W. Su, J. Cao, A. M. Pourrahimi, N. Li, X. Shen, M. Zhang, D. Yu, W. Gruber, T. Unruh, E. Wang, and C. J. Brabec, "Axisymmetric and Asymmetric Naphthalene-Bisthienothiophene Based Nonfullerene Acceptors: On Constitutional Isomerization and Photovoltaic Performance," ACS Applied Energy Materials **3**, 5734–5744 (2020).

[33] X. Song, L. Hou, R. Guo, Q. Wei, L. Yang, X. Jiang, S. Tu, A. Zhang, Z. Kan, W. Tang, G. Xing, and P. Müller-Buschbaum, "Synergistic Interplay between Asymmetric Backbone Conformation, Molecular Aggregation, and Charge-Carrier Dynamics in Fused-Ring Electron Acceptor-Based Bulk Heterojunction Solar Cells," ACS Applied Materials & Interfaces **13**, 2961–2970

(2021).

[34] Y. Zhang, Y. Wang, T. Shan, Q. Wei, Y. X. Xu, and H. Zhong, "Non-fullerene acceptors with an optical response over 1000 nm toward efficient organic solar cells," ACS Applied Materials & Interfaces **13**, 51279–51288 (2021).

[35] J. Cai, X. Zhang, C. Guo, Y. Zhuang, L. Wang, D. Li, D. Liu, and T. Wang, "Asymmetric and halogenated fused-ring electron acceptor for efficient organic solar cells," Advanced Functional Materials **31**, 2102189 (2021).

[36] Y. Cui, Y. Xu, H. Yao, P. Bi, L. Hong, J. Zhang, Y. Zu, T. Zhang, J. Qin, J. Ren, Z. Chen, C. He, X. Hao, Z. Wei, and J. Hou, "Single-Junction Organic Photovoltaic Cell with 19% Efficiency," Advanced Materials **33**, 2102420 (2021).

[37] W. Shockley and H. J. Queisser, "Detailed balance limit of efficiency of p-n junction solar cells," Journal of Applied Physics **32**, 510–519 (1961).

[38] H. Sahu, F. Yang, X. Ye, J. Ma, W. Fang, and H. Ma, "Designing promising molecules for organic solar cells *via* machine learning assisted virtual screening," Journal of Materials Chemistry A **7**, 17480–17488 (2019).

[39] H. Sahu and H. Ma, "Unraveling Correlations between Molecular Properties and Device Parameters of Organic Solar Cells Using Machine Learning," The Journal of Physical Chemistry Letters **10**, 7277–7284 (2019).

[40] Y. Wen, B. Yan, T. Gaudin, J. Ma, and H. Ma, "Simultaneous Optimization of Donor/acceptor Pairs and Device Specifications for Non-Fullerene Organic Solar Cells Using a QSPR Model with Morphological Descriptors," (2021), 10.26434/chemrxiv.14379065.v1, publisher: ChemRxiv.

[41] S. Nagasawa, E. Al-Naamani, and A. Saeki, "Computer-Aided Screening of Conjugated Polymers for Organic Solar Cell: Classification by Random Forest," The Journal of Physical Chemistry Letters **9**, 2639–2646 (2018).

[42] D. Padula, J. D. Simpson, and A. Troisi, "Combining electronic and structural features in machine learning models to predict organic solar cells properties," Materials Horizons **6**, 343–349 (2019).

[43] D. Padula and A. Troisi, "Concurrent Optimization of Organic Donor–Acceptor Pairs through Machine Learning," Advanced Energy Materials **9**, 1902463 (2019).

[44] H. Sahu, W. Rao, A. Troisi, and H. Ma, "Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors," Advanced Energy Materials **8**, 1801032 (2018).

[45]Z.-W. Zhao, O. H. Omar, D. Padula, Y. Geng, and A. Troisi, "Computational Identification of Novel Families of Nonfullerene Acceptors by Modification of Known Compounds," The Journal of Physical Chemistry Letters , 5009–5015 (2021).

[46]Z.-W. Zhao, M. del Cueto, Y. Geng, and A. Troisi, "Effect of Increasing the Descriptor Set on Machine Learning Prediction of Small Molecule-Based Organic Solar Cells," Chemistry of Materials **32**, 7777–7787 (2020).

[47]J. Hachmann, R. Olivares-Amaya, A. Jinich, A. L. Appleton, M. A. Blood-Forsythe, L. R. Seress, C. Román-Salgado, K. Trepte, S. Atahan-Evrenk, S. Er, S. Shrestha, R. Mondal, A. Sokolov, Z. Bao, and A. Aspuru-Guzik, "Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project," Energy & Environmental Science **7**, 698–704 (2014).

[48]F. Häse, L. M. Roch, P. Friederich, and A. Aspuru-Guzik, "Designing and understanding light-harvesting devices with machine learning," Nature Communications **11**, 4587 (2020).

[49]S. A. Lopez, B. Sanchez-Lengeling, J. de Goes Soares, and A. Aspuru-Guzik, "Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics," Joule **1**, 857–870 (2017).

[50]E. O. Pyzer-Knapp, G. N. Simm, and A. A. Guzik, "A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials," Materials Horizons **3**, 226–233 (2016), publisher: The Royal Society of Chemistry.

[51]T. Fink, H. Bruggesser, and J.-L. Reymond, "Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons," Angewandte Chemie International Edition **44**, 1504–1508 (2005).

[52]R. van Deursen and J.-L. Reymond, "Chemical Space Travel," ChemMedChem **2**, 636–640 (2007).

[53]K. Kranthiraja and A. Saeki, "Experiment-Oriented Machine Learning of Polymer:Non-Fullerene Organic Solar Cells," Advanced Functional Materials **31**, 2011168 (2021), _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.202011168.

[54]B. Greenstein and G. Hutchison, "Organic photovoltaic efficiency predictor: Data-driven predictions of power conversion efficiencies of non-fullerene acceptor organic solar cells," ChemRxiv (2021), 10.26434/chemrxiv-2021-nvcx9.

[55]L. Ma, S. Zhang, J. Zhu, J. Wang, J. Ren, J. Zhang, and J. Hou, "Completely non-fused electron acceptor with 3D-interpenetrated crystalline structure enables efficient and stable organic solar

cell," Nature Communications **12**, 5093 (2021).

[56]W. Zheng, J. Liu, Y. Guo, G. Han, and Y. Yi, "Regulation of molecular orientations of a–d–a nonfullerene acceptors for organic photovoltaics: The role of end-group π–π stacking," Advanced Functional Materials , 2108551 (2021).

[57]I. Y. Kanal and G. R. Hutchison, "Rapid computational optimization of molecular properties using genetic algorithms: Searching across millions of compounds for organic photovoltaic materials," (2017), arXiv:1707.02949.

[58]T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94," Journal of Computational Chemistry **17**, 490–519 (1996).

[59]A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," Journal of the American Chemical Society **114**, 10024–10035 (1992).

[60]C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions," Journal of Chemical Theory and Computation **15**, 1652–1671 (2019).

[61]S. Grimme, C. Bannwarth, and P. Shushkov, "A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1$–86)," Journal of Chemical Theory and Computation **13**, 1989–2009 (2017).

[62]P. Pracht, E. Caldeweyher, S. Ehlert, and S. Grimme, "A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules," (2019), 10.26434/chemrxiv.8326202.v1.

[63]F. Neese, "The ORCA program system," Wily interdisciplinary Reviews - Computational Molecular Science **2**, 73–78 (2012).

[64]S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," Journal of Computational Chemistry **32**, 1456–1465 (2011).

[65]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," The Journal of Chemical Physics **132**, 154104 (2010).

[66]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Physical Chemistry

Chemical Physics **7**, 3297–3305 (2005).

[67] F. Weigend, "Accurate Coulomb-fitting basis sets for H to Rn," Physical Chemistry Chemical Physics **8**, 1057–1065 (2006).

[68] M. de Wergifosse and S. Grimme, "Nonlinear-response properties in a simplified time-dependent density functional theory (sTD-DFT) framework: Evaluation of excited-state absorption spectra," The Journal of Chemical Physics **150**, 094112 (2019).

[69] U. Ekström, L. Visscher, R. Bast, A. J. Thorvaldsen, and K. Ruud, "Arbitrary-Order Density Functional Response Theory from Automatic Differentiation," Journal of Chemical Theory and Computation **6**, 1971–1980 (2010).

[70] M. de Wergifosse and S. Grimme, "Nonlinear-response properties in a simplified time-dependent density functional theory (sTD-DFT) framework: Evaluation of the first hyperpolarizability," The Journal of Chemical Physics **149**, 024108 (2018).