Computational Identification of Novel Families of Non-Fullerene Acceptors by Modification of Known Compounds

Zhi-Wen Zhao,[a] Ömer H. Omar,[b] Daniele Padula,[c] Yun Geng,[a] Alessandro Troisi*[b]

[a]Institute of Functional Material Chemistry, Faculty of Chemistry, Northeast Normal University, Changchun, 130024, Jilin, P. R. China.

[b]Department of Chemistry, University of Liverpool, Liverpool, L69 3BX, UK.

[c]Dipartimento di Biotecnologie, Chimica e Farmacia, Università di Siena, via A. Moro 2, 53100, Siena, Italy.

*Email: a.troisi@liverpool.ac.uk

Abstract

We considered a database of tens of thousands of known organic semiconductors and identified those compounds with computed electronic properties (orbital energies, excited state energies and oscillator strengths) that would make them suitable as non-fullerene electron acceptors in organic solar cells. The range of parameters for the desirable acceptors is determined from a set of experimentally characterized high-efficiency non-fullerene acceptors. This search leads to approximately 30 lead compounds never before considered for organic photovoltaic applications. We then proceed to modify these compounds to bring their computed solubility in line with that of the best small-molecule non-fullerene acceptors. A further refinement of the search can be based on additional properties like the reorganization energy for chemical reduction. This simple strategy, which relies on a few easily computable parameters and can be easily expanded to a larger set of molecules, enables the identification of completely new chemical families to be explored experimentally.

Many non-fullerene acceptors (NFAs) have been widely used to blend with a variety of donor materials and produced the best power conversion efficiency (PCEs) of approaching 18% and over 14% for polymer-based and small-molecule-based single-junction devices, respectively, far beyond those of the best fullerene-based counterparts.[1-3] The results raise the hope of reaching PCE of >20% for organic photovoltaic (OPV) cells in the not-so-distant future.[4] However, these achievements have mostly been driven by trial and error, naturally accompanied by high cost and the preference for molecular designs obtained by small modifications within the same families of compound that tend to have high efficiency.[5-8] The ideal contribution of theory should be the proposal of completely novel families of compounds that can introduce more radical changes in design to accelerate progress beyond what is achievable through incremental changes.

Different groups have explored the common properties for high-efficiency NFAs. Yi et al.[9] have found that a large oscillator strength ($f \approx 3$) and a moderate energy gap between singlet and triplet ΔE_{ST} (0.4–0.5 eV) are present in state-of-the-art A–D–A small molecules (ITIC, IT-4F, and Y6). Some of us proposed on the basis of physical arguments[10] and later verified on a diverse set acceptors[11] that a small energy gap between LUMO and LUMO+1 (< 0.3 eV) is beneficial for high PCE solar cells. Firdaus et al. found that balanced electron and hole mobilities of > 10^{-3} cm^{2}V^{-1}s^{-1} in combination with low nongeminate recombination rate constants of 10^{-12} cm^{3}s^{-1} could lead to relatively high PCE values in OPV cells.[12] It was reported that the balanced ambipolar charge transport and long exciton diffusion lengths are observed in high-performance NFAs compared to fullerene derivatives.[13] Also, Armin et al. have reported that NFAs show many different properties including narrower optical gap, lower energetic disorder and better planarity compared to fullerene derivatives.[14]

The main hypothesis of this work is that novel families of non-fullerene acceptors should be similar to the best known non-fullerene acceptors in terms of the most relevant electronic properties (assuming that similar donors are employed, for optimal energy level alignment). To enable the identification of novel molecules from a large set, we consider computed properties so that a molecule proposed as an acceptor for OPV has electronic properties within the same range found for known efficient electron acceptors. The electronic properties that we consider are the ground-state orbital energy of HOMO (E_{HOMO}) and LUMO (E_{LUMO}), the energy gap between LUMO and LUMO+1 (ΔE_{LL+1} whose relevance was proposed in ref. 11), the energy gap

between singlet (S1) and triplet (T1) state ΔE_{S1T1} (proposed as an important parameter in ref. 9), and the oscillator strength f_{max} of the excited state with strongest absorption of the three lowest states, as the ability to absorb radiation is one of the main advantages of novel acceptors with respect to their fullerene-based couterparts.[15] An additional property of interest is the reorganization energy λ for electron transport, as we expect materials with smaller λ to facilitate higher electron mobility.[16-18] Finally, possibly the most elusive yet important property is the solubility of the acceptor (and miscibility with the donor). This is not an electronic property and, if a molecule with suitable electronic properties is found, it is relatively easy to modify their solubility properties by chemical substitution with groups that do not alter these key parameters. We note that identifying candidates with electronic and physical properties close to the best candidates is akin to the intuitive experimental process of materials discovery and is, in essence, the underlying hypothesis of many recent papers that have explored machine learning to discover new OPV materials.[19-22] In this work, however, we will directly search for such "mimic" without building an explicit data-centric model.

Virtual screening has been used by many authors in the domain of organic electronics[22-28] and, specifically, in the identification of novel non-fullerene acceptors[26-28], sometime in combination with machine learning methods[22, 23, 29-31]. To the best of our knowledge, all previous attempts focused on the study of structures obtained as modification of known motifs[30, 32-35], combination of known oligomers[36] or more advanced generative models based on prior knowledge of the best candidates[26, 37]. In all these cases, the findings are bound to fall within the same class of molecules and, depending on the approach for generation, they are not guaranteed to yield molecules that are easy to synthesize. Here we follow an approach that is opposite in two ways: (i) we look into a dataset not based on organic electronics and we deliberately exclude from further consideration molecules belonging to known classes. We are able to identify very different lead compounds rather than promising modifications of what is known. (ii) The molecules in our screening have been synthesized (they have a documented synthetic pathway) and are sufficiently stable in solid state form to enable their crystallographic characterization. Our proposed molecules are therefore more likely to constitute a viable suggestion for experimental validation. An additional advantage in considering an unbiased dataset is the physical insight that this can provide. As we set different criteria for the ideal candidates based on experimental benchmarks, we can identify which criteria are more stringent (i.e. more rare) in a generic data set, an information that can be used in translating the

virtual screening results into design principles.

The structure of this work is summarized in Figure 1. We first characterize the electronic properties of several benchmark NFAs and we use their properties as reference to select molecules from a large database of molecular semiconductors derived from the Cambridge Structural Database (CSD[38]). After removing known NFAs and performing more accurate evaluation of their properties, we tune the solubility of the best candidates to match those of the known high efficiency NFAs.
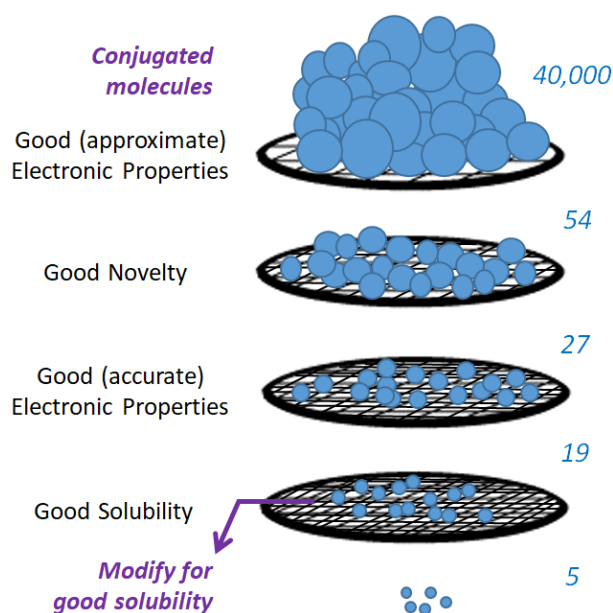

Figure 1. Outline of the virtual screening protocol followed in this work.

To identify an initial range of electronic properties for NFAs we considered the 35 molecules with PCE > 5% reported in a 2017 work,[11] and computed the electronic properties at the M062X/def2-TZVP level, on the optimized geometry at the BLYP35/def2-SVP level (these choice are to be consistent with available data on a large dataset of molecular semiconductors,[39] see below). The E_{LUMO}, f_{max} and E_{HOMO} values of these acceptors are in the range of (-3.32 eV, -2.13 eV), (0.09, 3.54), and (-7. 48 eV, -5.88 eV), respectively. The ΔE_{S1T1} and ΔE_{LL+1} values of NFAs are in the range of (0.65 eV, 1.30 eV) and (0, 0.59 eV), respectively. To verify these criteria, the properties of seven recent high-performance molecules (Y6,[2] ITIC-0F,[40] IDIC,[41] ITIC-Th,[42] ITIC-4F,[40] IDTBR[43] and IDT2Se-4F[44]) were calculated at the same level. The results, summarized in Table 1, show the following ranges for these top examples: -3.21 eV < E_{LUMO} < -2.81 eV, -6.91 eV < E_{HOMO} < -6.36 eV, 2.45 < f_{max} < 3.32, 0.70 eV < ΔE_{S1T1} < 0.79 eV and 0.17 eV < ΔE_{LL+1} < 0.44 eV. These ranges are consistent with those determined from the 2017 dataset but overall narrower. We define novel acceptors to be sufficiently similar to

known efficient electron NFAs if all their five electronic parameters (E_{LUMO}, ΔE_{S1T1}, E_{HOMO}, ΔE_{LL+1}, f_{max}) are within the same range found in acceptors collected in the dataset of ref. 11 and with experimental PCEs larger than a 5% cut-off (a range that also includes all recent acceptors in Table 1). This definition is convenient because one can easily broaden or narrow the search, and therefore the criteria of similarity, by changing the value of this PCE cut-off. In general, a dataset of computed properties for experimentally characterized solar cell is very valuable regardless of the specific virtual screening approach.

Table 1. The calculated properties (E_{HOMO}, E_{LUMO}, ΔE_{LL+1}, $E(S1)$ and $E(T1)$, ΔE_{S1T1} and λ_{int} in eV) of latest reported high-performance NFAs.

	E_{HOMO}	E_{LUMO}	ΔE_{LL+1}	$E(S1)$	f_{max}	$E(T1)$	ΔE_{S1T1}	consensus Log $P_{o/w}$	λ_{int}
Y6	-6.79	-3.21	0.30	2.21	2.45	1.51	0.70	9.27	0.24
IDT2Se-4F	-6.60	-3.09	0.17	2.15	3.32	1.44	0.72	18.80	0.24
ITIC-0F	-6.69	-2.94	0.27	2.33	2.97	1.59	0.75	18.93	0.23
IDIC	-6.91	-3.06	0.44	2.35	2.53	1.56	0.79	13.27	0.21
IDTBR	-6.36	-2.81	0.20	2.19	2.72	1.42	0.77	19.96	0.25
ITIC-4F	-6.81	-3.11	0.27	2.30	3.01	1.57	0.74	20.07	0.24
ITIC-Th	-6.72	-2.96	0.27	2.34	2.91	1.60	0.75	21.21	0.24

To identify novel NFAs with similar electronic characteristics we considered the database of 40k molecular semiconductors presented in ref. 39. Such database was constructed using the crystal structure geometry[38] of molecules with small HOMO-LUMO gap and, in this work, the electronic properties have been computed at the M06-2X/def2-TZVP level, the same chosen to identify a reasonable electronic property range. Recent applications of the database for the identification of molecules for singlet fission[39, 45] and temperature activated delayed fluorescence[46] showed that the geometry of the molecule within the crystal is an excellent starting point for an initial screening, but geometry optimization should be performed for the most promising candidates. The criteria of similarity with existing NFAs are applied sequentially (the end result does not depend on the order):

(i) the condition -3.32 eV < E_{LUMO} < -2.13 eV results in 5159 molecules from the initial 40k molecules. This major reduction is expected as only a minority of molecules have electron accepting characteristics (a data repository[47] gives access to the list and key electronic properties of these acceptors);

(ii) the condition f_{max} > 0.09 further reduced to 2697 molecules;

(iii)   the condition 0.65 eV < ΔE_{S1T1} < 1.30 eV further reduced the set to 1793 molecules (it should be noted that the last two conditions require a very computationally demanding evaluation of excited states properties of both singlet and triplet.);

(iv)   the condition -7.48 eV < E_{HOMO} < -5.88 eV, like the previous two, also has a small effect reducing the set to 1116 molecules;

(v)   the condition 0 < ΔE_{LL+1} < 0.59 has a substantial effect, reducing the set to just 54 molecules and supporting the idea that the gap between LUMO and LUMO+1 energy can help discriminating between good and bad electron acceptors.[11]

The chemical structures of these 54 molecules, about 0.14% of the total number of 40k semiconductor molecules in the semiconductors database derived from the CSD, are depicted in Table S1. Among these 54 molecules, two molecules have been employed as OSC acceptor materials (identified as PHTLBC02 and IQIDIW02 in Table S1). 27 out of 54 molecules identified as promising contain BODIPY, TPA, indole, naphthalimide, sub-phthalocyanine, porphyrin, azulene, Benzothiadiazol, and other building blocks common in OSC materials (the SI provides a list of references for such materials). This is very reassuring from the point of view of the computational protocol. The database of 40k molecules does not focus on organic electronic materials and 54 molecules represent just 0.14% of the database. The 27 molecules that are either know OPV acceptors or chemically very similar to known OPV acceptors represent just 0.1% of the database. The identification of so many "known" acceptors from a database that contains very few of them is what provides experimental support to the predictions. The more interesting molecules from the point of view of materials discovery are 27 molecules that, to the best of our knowledge, do not contain in the building blocks reported in OSC (a list is given in Table S2).

The next step in the screening procedure is to perform geometry optimization of the 27 potentially novel NFAs at BLYP35/def2-SVP[45] level and re-compute their properties again at the M062X/def2-TZVP level. 18 molecules continue to satisfy all five criteria after optimization, 9 miss one criterion as seen in Table 2. A diagram illustrating how the electronic properties of these 27 candidates are aligned with those of known NFA molecules in given in Figure 2. The optimized geometry of these molecules are available via ref. 47.
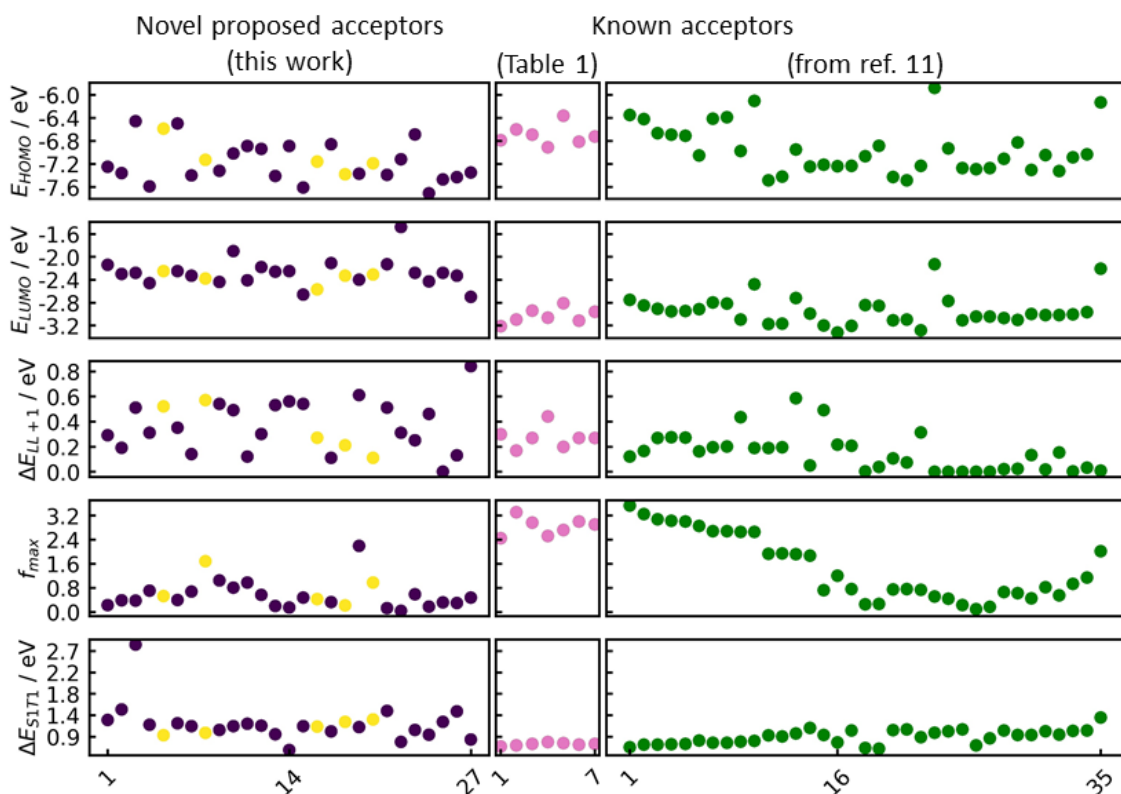
Figure 2. Summary of the electronic properties for 27 promising acceptor molecules in Table 2 (left panel), 7 NFA molecules in Table 1 (central panel) and 35 NFA molecules from ref. 11 (right panel). The left panel highlights in yellow the 5 molecules selected for optimal solubility which in this diagram are EMONUP, EPOHEY, LOKPOS, MOTSIX and MUJTOB01 from left to right respetively.

Table 2. The calculated properties (E_{HOMO}, E_{LUMO}, ΔE_{LL+1}, ΔE_{S1T1} and λ_{int} in eV) of the 27 screened molecules. The ID number is that which is used in the CSD. The values underlined are those that miss the criteria after optimization.

ID	E_{HOMO}	E_{LUMO}	ΔE_{LL+1}	f_{max}	ΔE_{S1T1}	consensus Log $P_{o/w}$	λ_{int}
AQIHAJ	-7.25	-2.14	0.29	0.22	1.25	2.67	0.56
CIYWIQ	-7.36	-2.30	0.19	0.38	<u>1.47</u>	4.38	1.69
CUVTAO	-6.46	-2.28	0.51	0.37	<u>2.83</u>	4.90	1.23
DAWHUH	-7.59	-2.46	0.31	0.70	1.15	2.47	0.36
EMONUP	-6.59	-2.25	0.52	0.52	0.93	11.92	0.35
ENBILO	-6.50	-2.25	0.35	0.39	1.18	5.50	0.31
ENEWIF	-7.40	-2.33	0.14	0.67	1.12	3.14	1.29
EPOHEY	-7.13	-2.38	0.57	1.68	0.98	7.39	0.38
FEYVIQ	-7.32	-2.44	0.54	1.04	1.04	3.42	0.27
HOYVEX	-7.02	<u>-1.90</u>	0.49	0.80	1.12	1.51	0.73
HUMXIX	-6.89	-2.41	0.12	0.97	1.17	2.05	0.32
IDOGAL	-6.94	-2.18	0.30	0.56	1.13	3.73	0.39
IRUQOB	-7.41	-2.26	0.53	0.19	0.95	4.13	0.92
JASVII	-6.89	-2.25	0.56	0.14	0.62	7.19	0.87

LAFWEV	<u>-7.61</u>	-2.66	0.54	0.47	1.12	4.13	1.01
LOKPOS	-7.16	-2.57	0.27	0.42	1.11	13.18	0.31
MOKCEW	-6.86	-2.11	0.11	0.32	1.01	5.20	0.60
MOTSIX	-7.38	-2.33	0.21	0.21	1.21	9.24	0.26
MUBTEJ	-7.37	-2.40	0.61	2.19	1.10	4.52	0.37
MUJTOB0 1	-7.19	-2.31	0.11	0.97	1.26	7.76	0.25
NEFLIU	-7.39	-2.13	0.51	0.12	<u>1.44</u>	3.00	0.53
NEPQIL	-7.12	<u>-1.48</u>	0.31	0.03	0.79	5.56	0.36
PARDAO	-6.69	-2.28	0.25	0.58	1.04	3.63	0.68
UBIKEV	<u>-7.71</u>	-2.43	0.46	0.17	0.94	5.00	0.67
XALVIS	-7.47	-2.28	0.00	0.31	1.21	6.49	0.19
YEGBIW	-7.43	-2.33	0.13	0.29	<u>1.43</u>	1.99	0.99
ZOPBAJ	-7.35	-2.70	<u>0.84</u>	0.47	0.84	3.41	0.97

The film-forming properties of the molecules are very critical; however, the related simulation methods are very expensive. Therefore, this work considers a compromise method to characterize solubility, which is to calculate the lipophilicity of the molecule measured by the logarithm of octanol/water partition function (Log $P_{o/w}$), as common in medicinal chemistry applications.[48, 49] In the past years, the developments of Log $P_{o/w}$ prediction programs have attracted growing interest[49-56]; however, predictive Log P methods themselves have differences in prediction accuracy for different molecules[49]. Thus, the consensus Log $P_{o/w}$ model (generally described as consensus Log $P_{o/w}$) has been adopted to improve the prediction accuracy, which was the average of five performing models[50] (iLOGP,[51] XLOGP3,[52] WLOGP,[53] MLOGP[54] and SILICOS-IT[55]). The values used in this work are the consensus Log $P_{o/w}$ evaluated from the SMILES representation of the molecule using the SwissADME tool[56]. As shown in Table 1 and Table 2, among these 27 molecules, there are molecules with lipophilicity very close to Y6 and IDIC. However, there are no molecules with similar lipophilicity to other NFAs (much more hydrophobic). The number of molecules similar to Y6 (the difference is (-2,2)) is 3, and the number of molecules similar to IDIC is 2 in term of lipophilicity. The chemical structures of these 5 molecules with good lipophilicity are shown in Figure 3. The molecule EPOHEY[57] was reported as an n-type field-effect transistor. And the molecules MUJTOB01,[58] EMONUP,[59] MOTSIX[60] and LOKPOS[61] have not been used in OPV devices.
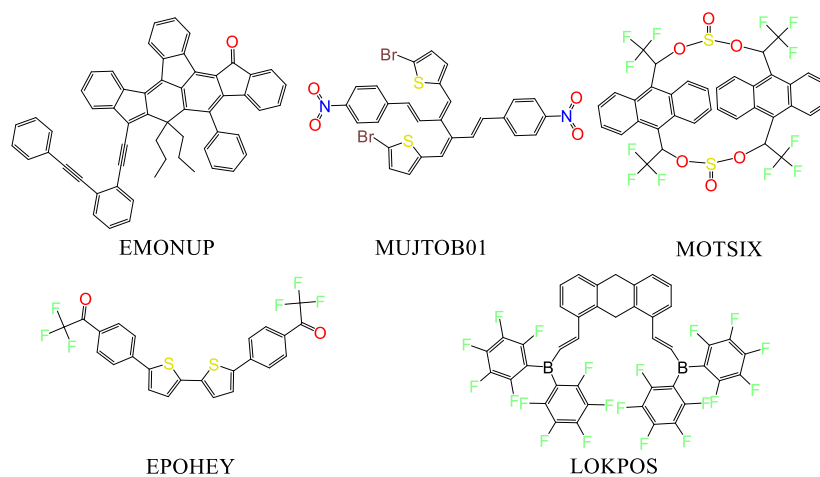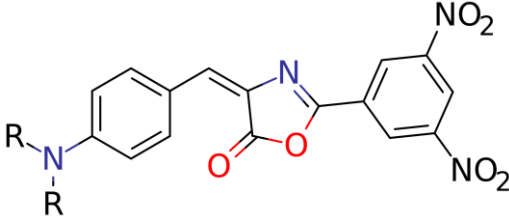
Figure 3. Chemical structures of 5 molecules with similar electronic and solubility parameters of known high efficiency NFAs.

The database used for the screening of alternative NFAs is based on crystalline molecular materials, so it tends to contain more rigid molecules as those with long saturated chains (which increase solubility and lipophilicity) are more difficult to crystallize. On the other hand, the screening of such database remains very useful because the main electronic properties are not modified by adding saturated chains to the structure and these modifications are generally easy from the synthetic point of view. It is therefore possible to design suitable candidates by starting from a molecule with the correct electronic properties and modify it by adding side chains (alkyl, alkoxy were chosen according to refs 62-64) to approach the desired (computed) solubility parameters. For example, the consensus Log $P_{o/w}$ of molecule HUMXIX is 2.05, which is the farthest one from that of Y6 (9.27) or any other NFA molecules. However, after alkylation, comparable lipophilicity with Y6 and IDIC (9.33 vs. 9.27, 13.10 vs. 13.27) could be obtained, as illustrated in Figure 4. Using this approach, all 27 molecules of Table 2 can be tuned to have appropriate Log $P_{o/w}$. Of course, this approach is unable to guarantee that the bulk heterojunction has the correct microstructure in terms of domain size of donor and acceptor phase but it alleviates the problem present in virtual screenings ignoring properties that are not deriving from the electronic structure.
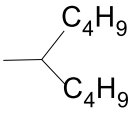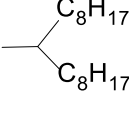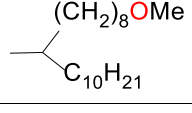
R	Log $P_{o/w}$
Me	2.05
C_4H_9 C_4H_9	9.33
C_8H_{17} C_8H_{17}	13.10
$(CH_2)_8OMe$ $C_{10}H_{21}$	14.12

Figure 4. The calculated consensus Log $P_{o/w}$ of molecule HUMXIX and several modified molecules.

A final property that can be considered for the selection of the best NFAs from this screening is the internal reorganization energy for the chemical reduction, defined as $\lambda_{int} = \left[E\left(M^-\right) - E(M)\right] + \left[E^-(M) - E^-\left(M^-\right)\right]$, where $E(M)$ and $E\left(M^-\right)$ are the neutral state energies at the optimal ground geometry (M) and anion geometry (M^-), respectively, while $E^-\left(M^-\right)$ and $E^-(M)$ refer to the energies of the anion at the optimal anion and neutral geometries, respectively. It is expected that λ_{int} is small for materials with larger electron mobilities.[16-18] This property is computationally too expensive to evaluate on an initial dataset of tens of thousands of molecules but it can be used to further select the best candidates. It was reported in Table 1, for reference NFA compounds, and Table 2 for all 27 novel NFAs found after the first layer of the screening. The values of λ_{int} in Table 1 spans a narrow range between 0.21 and 0.25 eV (also because that data set is quite limited). Of the 5 molecules with ideal solubility and electronic properties in Figure 3, two values are very close to this range (0.25 and 0.26 eV for MUJTOB01 and MOTSIX), and three have slightly larger but still comparable values (0.31, 0.35 and 0.38 eV for LOKPOS, EMONUP and EPOHEY). Of the 27 molecules in Table 2 whose Log $P_{o/w}$ can be adjusted to match the best NFAs, 8 have $\lambda_{int} \leq 0.35$ eV, i.e., approximately 30% of the molecules identified with this screening

method have suitable internal reorganization energy. It is therefore appropriate to apply this more expensive but not very selective criterion at the end of the procedure.

In conclusion, we have proposed an approach to identify completely new acceptors to be used in organic solar cells based on selecting, from an unbiased set of known 40k molecular semiconductors, a set of molecules sharing all key electronic properties with the best known non-fullerene acceptors. We show that it is possible to either select molecules which also share similar solubility character or to modify the candidates to tune their solubility. A final refinement of the prediction was introduced by considering the role of the reorganization energy. The approach can be generalized to identify candidates for any property and is naturally expanded starting from a larger database of initial compounds. If the initial database is not biased toward a certain property, like in this case, it is more likely to find completely new lead compounds to explore.

Supporting information

Computational detail for the identification of the 54 molecules. Chemical structures of the 54 and 27 molecules selected in two stages of the screening. References to works reporting acceptors deemed not novel. Data for 5159 acceptor molecules and optimized geometry of 27 candidates available in machine readable format from ref. 47.

Acknowledgements

References

1.   Cui, Y.; Yao, H.; Zhang, J.; Xian, K.; Zhang, T.; Hong, L.; Wang, Y.; Xu, Y.; Ma, K.; An, C., et al., Single-Junction Organic Photovoltaic Cells with Approaching 18% Efficiency. *Adv. Mater.* **2020,** *32*, 1908205.

2.   Zhou, R.; Jiang, Z.; Yang, C.; Yu, J.; Feng, J.; Adil, M. A.; Deng, D.; Zou, W.; Zhang, J.; Lu, K., et al., All-small-molecule organic solar cells with over 14% efficiency by optimizing hierarchical morphologies. *Nat. Commun.* **2019,** *10*, 5393.

3.   Qin, J.; An, C.; Zhang, J.; Ma, K.; Yang, Y.; Zhang, T.; Li, S.; Xian, K.; Cui, Y.; Tang, Y., et al., 15.3% efficiency all-small-molecule organic solar cells enabled by symmetric phenyl substitution. *Sci. China Mater.* **2020**, *63*, 1142-1150.

4.   Chen, L. X., Organic Solar Cells: Recent Progress and Challenges. *ACS Energy Lett.* **2019,** *4*, 2537-2539.

5.   Wan, X.; Li, C.; Zhang, M.; Chen, Y., Acceptor-donor-acceptor type molecules for high performance organic photovoltaics - chemistry and mechanism. *Chem. Soc. Rev.* **2020,** *49*, 2828-2842.

6.   Wang, J.; Zhan, X., Fused-Ring Electron Acceptors for Photovoltaics and Beyond. *Acc. Chem. Res.* **2021,** *54*, 132-143.

7.   Yao, H.; Ye, L.; Zhang, H.; Li, S.; Zhang, S.; Hou, J., Molecular Design of Benzodithiophene-Based Organic Photovoltaic Materials. *Chem. Rev.* **2016,** *116*, 7397-457.

8.   Zhang, G.; Zhao, J.; Chow, P. C. Y.; Jiang, K.; Zhang, J.; Zhu, Z.; Zhang, J.; Huang, F.; Yan, H., Nonfullerene Acceptor Molecules for Bulk Heterojunction Organic Solar Cells. *Chem. Rev.* **2018,** *118*, 3447-3507.

9.   Han, G.; Hu, T.; Yi, Y., Reducing the Singlet-Triplet Energy Gap by End-Group pi-pi Stacking Toward High-Efficiency Organic Photovoltaics. *Adv. Mater.* **2020,** *32*, e2000975.

10.   Ma, H.; Troisi, A., Modulating the Exciton Dissociation Rate by up to More than Two Orders of Magnitude by Controlling the Alignment of LUMO + 1 in Organic Photovoltaics. *J. Phys. Chem. C* **2014,** *118*, 27272-27280.

11.   Kuzmich, A.; Padula, D.; Ma, H.; Troisi, A., Trends in the electronic and geometric structure of non-fullerene based acceptors for organic solar cells. *Energy Environ. Sci.* **2017,** *10*, 395-401.

12.   Firdaus, Y.; Le Corre, V. M.; Khan, J. I.; Kan, Z.; Laquai, F.; Beaujuge, P. M.; Anthopoulos, T. D., Key Parameters Requirements for Non-Fullerene-Based Organic Solar Cells with Power Conversion Efficiency >20%. *Adv. Sci.* **2019,** *6*, 1802028.

13.   Firdaus, Y.; Le Corre, V. M.; Karuthedath, S.; Liu, W.; Markina, A.; Huang, W.; Chattopadhyay, S.; Nahid, M. M.; Nugraha, M. I.; Lin, Y., et al., Long-range exciton diffusion in molecular non-fullerene acceptors. *Nat. Commun.* **2020,** *11*, 5220.

14.   Armin, A.; Li, W.; Sandberg, O. J.; Xiao, Z.; Ding, L.; Nelson, J.; Neher, D.; Vandewal, K.; Shoaee, S.; Wang, T., et al., A History and Perspective of Non-Fullerene Electron Acceptors for Organic Solar Cells. *Adv. Energy Mater*. **2021**, 20003570

15.   Ip, C. M.; Eleuteri, A.; Troisi, A., Predicting with confidence the efficiency of new dyes in dye sensitized solar cells. *Phys. Chem. Chem. Phys.* **2014,** *16*, 19106-10.

16.   Mas-Torrent, M.; Hadley, P.; Bromley, S. T.; Ribas, X.; Tarrés, J.; Mas, M.; Molins, E.; Veciana, J.; Rovira, C., Correlation between Crystal Structure and Mobility in Organic Field-Effect Transistors Based on Single Crystals of Tetrathiafulvalene Derivatives. *J. Am. Chem. Soc.* **2004,** *126*, 8546-8553.

17.   Cheung, D. L.; Troisi, A., Modelling charge transport in organic semiconductors: from quantum dynamics to soft matter. *Phys. Chem. Chem. Phys*., **2008**, *10*, 5941-5952

18.   Lee, H.; Jeong, K.; Cho, S. W.; Yi, Y., Theoretical study on the effects of nitrogen and methyl substitution on tris-(8-hydroxyquinoline) aluminum: an efficient exciton blocking layer for organic photovoltaic cells. *J. Chem. Phys.* **2012,** *137*, 034704.

19.   Mahmood, A.; Wang, J.-L., Machine learning for high performance organic solar cells: current scenario and future prospects. *Energy Environ. Sci*. **2021,** *14*, 90-105.

20.   Padula, D.; Troisi, A., Concurrent Optimization of Organic Donor–Acceptor Pairs through Machine Learning. *Adv. Energy Mater*. **2019,** *9*, 1902463.

21.   Padula, D.; Simpson, J. D.; Troisi, A., Combining electronic and structural features in machine learning models to predict organic solar cells properties. *Mater. Horizons* **2019,** *6*, 343-349.

22.   Wu, Y.; Guo, J.; Sun, R.; Min, J., Machine learning for accelerating the discovery of high-performance donor/acceptor pairs in non-fullerene organic solar cells. *NPJ Comput. Mater*. **2020,**

6, 120.

23. Saeki, A.; Kranthiraja, K., A high throughput molecular screening for organic electronics via machine learning: present status and perspective. *Jpn. J. Appl. Phys.* **2020,** *59*, SD0801.

24. Kranthiraja, K.; Saeki, A., Experiment-Oriented Machine Learning of Polymer:Non-Fullerene Organic Solar Cells. *Adv. Funct. Mater.* **2021**, 2011168

25. Pollice, R.; Dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z., et al., Data-Driven Strategies for Accelerated Materials Design. *Acc. Chem. Res.* **2021,** *54*, 849-860.

26. Peng, S. P.; Zhao, Y., Convolutional Neural Networks for the Design and Analysis of Non-Fullerene Acceptors. *J. Chem. Inf. Model.* **2019,** *59*, 4993-5001.

27. Lopez, S. A.; Sanchez-Lengeling, B.; de Goes Soares, J.; Aspuru-Guzik, A., Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics. *Joule* **2017,** *1*, 857-870.

28. Imamura, Y.; Suganuma, M.; Hada, M., Computational Study on the Search for Non-Fullerene Acceptors, Examination of Interface Geometry, and Investigation of Electron Transfer. *J. Phys. Chem. C* **2019,** *123*, 17678-17685.

29. 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.

30. Olivares-Amaya, R.; Amador-Bedolla, C.; Hachmann, J.; Atahan-Evrenk, S.; Sánchez-Carrera, R. S.; Vogt, L.; Aspuru-Guzik, A., Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics. *Energy Environ. Sci.* **2011,** *4*, 4849-4861

31. Jorgensen, P. B.; Mesta, M.; Shil, S.; Garcia Lastra, J. M.; Jacobsen, K. W.; Thygesen, K. S.; Schmidt, M. N., Machine learning-based screening of complex molecules for polymer solar cells. *J. Chem. Phys.* **2018,** *148*, 241735.

32. Sahu, H.; Yang, F.; Ye, X.; Ma, J.; Fang, W.; Ma, H., Designing promising molecules for organic solar cells via machine learning assisted virtual screening. *J. Mater. Chem. A* **2019,** *7*, 17480-17488.

33. Shin, Y.; Liu, J.; Quigley, J. J.; Luo, H.; Lin, X., Combinatorial Design of Copolymer Donor Materials for Bulk Heterojunction Solar Cells. *ACS Nano* **2014,** *8*, 6089-6096.

34. Hachmann, J.; Olivares-Amaya, R.; Jinich, A.; Appleton, A. L.; Blood-Forsythe, M. A.; Seress, L. R.; Román-Salgado, C.; Trepte, K.; Atahan-Evrenk, S.; Er, S., et al., Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project. *Energy Environ. Sci.* **2014,** *7*, 698-704.

35. Zanlorenzi, C.; Akcelrud, L., Theoretical studies for forecasting the power conversion efficiencies of polymer-based organic photovoltaic cells. *J. Polym. Sci., Part B: Polym. Phys.* **2017,** *55*, 919-927.

36. Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R., Efficient Computational Screening of Organic Polymer Photovoltaics. *J. Phys. Chem. Lett.* **2013,** *4*, 1613-23.

37. Yuan, Q.; Santana-Bonilla, A.; Zwijnenburg, M. A.; Jelfs, K. E., Molecular generation targeting desired electronic properties via deep generative models. *Nanoscale* **2020,** *12*, 6744-6758.

38. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallogr., Sect. B* **2016,** *72*, 171-179.

39. Padula, D.; Omar, Ö. H.; Nematiaram, T.; Troisi, A., Singlet fission molecules among known compounds: finding a few needles in a haystack. *Energy Environ. Sci.* **2019,** *12*, 2412-2416.

40. 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*, 7148-7151.

41. Chandrabose, S.; Chen, K.; Barker, A. J.; Sutton, J. J.; Prasad, S. K. K.; Zhu, J.; Zhou, J.; Gordon, K. C.; Xie, Z.; Zhan, X., et al., High Exciton Diffusion Coefficients in Fused Ring Electron Acceptor Films. *J. Am. Chem. Soc.* **2019,** *141*, 6922-6929.

42. Bristow, H.; Thorley, K. J.; White, A. J. P.; Wadsworth, A.; Babics, M.; Hamid, Z.; Zhang, W.; Paterson, A. F.; Kosco, J.; Panidi, J., et al., Impact of Nonfullerene Acceptor Side Chain Variation on Transistor Mobility. *Adv. Electron. Mater*., **2019**, *5*, 1900344

43. Che, Y.; Zhang, Y.; Yang, Y.; Liu, C. H.; Izquierdo, R.; Xiao, S. S.; Perepichka, D. F., Understanding the Photovoltaic Behavior of A-D-A Molecular Semiconductors through a Permutation of End Groups. *J. Org. Chem.* **2020,** *85*, 52-61.

44. Liang, Z.; Li, M.; Zhang, X.; Wang, Q.; Jiang, Y.; Tian, H.; Geng, Y., Near-infrared absorbing non-fullerene acceptors with selenophene as π bridges for efficient organic solar cells. *J. Mater. Chem. A* **2018,** *6*, 8059-8067.

45. Omar, Ö. H.; Padula, D.; Troisi, A., Elucidating the Relationship between Multiradical Character and Predicted Singlet Fission Activity. *ChemPhotoChem* **2020,** *4*, 5223-5229.

46. Zhao, K.; Omar, Ö. H.; Nematiaram, T.; Padula, D.; Troisi, A., Novel thermally activated delayed fluorescence materials by high-throughput virtual screening: going beyond donor–acceptor design. *J. Mater. Chem. C* **2021,** *9*, 3324-3333.

47. https://github.com/amiswen/Virtual-Screening-of-Non-Fullerene-Acceptor

48. Arnott, J. A.; Planey, S. L., The influence of lipophilicity in drug discovery and design. *Expert Opin. Drug Discov.* **2012,** *7*, 863-875.

49. Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V., Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharm. Sci.* **2009,** *98*, 861-93.

50. Plante, J.; Werner, S., JPlogP: an improved logP predictor trained using predicted data. *J. Cheminform.* **2018,** *10*, 61.

51. Daina, A.; Michielin, O.; Zoete, V., iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* **2014,** *54*, 3284-301.

52. Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L., Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007,** *47*, 2140-2148.

53. Wildman, S. A.; Crippen, G. M., Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999,** *39*, 868-873.

54. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y., Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992,** *40*, 127-130.

55. http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html. (Accessed by June 2020)

56. Daina, A.; Michielin, O.; Zoete, V., SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep*. **2017,** *7*, 42717.

57. Ie, Y.; Nitani, M.; Uemura, T.; Tominari, Y.; Takeya, J.; Honsho, Y.; Saeki, A.; Seki, S.; Aso, Y., Comprehensive Evaluation of Electron Mobility for a Trifluoroacetyl-Terminated Electronegative Conjugated Oligomer. *J. Phys. Chem. C* **2009,** *113*, 17189-17193.

58. Kanibolotsky, A. L.; Forgie, J. C.; McEntee, G. J.; Talpur, M. M.; Skabara, P. J.; Westgate, T. D.; McDouall, J. J.; Auinger, M.; Coles, S. J.; Hursthouse, M. B., Controlling the conformational changes

in donor-acceptor [4]-dendralenes through intramolecular charge-transfer processes. *Chemistry* **2009,** *15*, 11581-93.

59. Yang, Y.; Petersen, J. L.; Wang, K. K., Cascade Radical Cyclizations of Benzannulated Enyne–Allenes. Unusual Cleavage of a Benzene Ring Leading to Twisted 1,1'-Dialkyl-9,9'-bifluorenylidenes and Spiro[1H-cyclobut[a]indene-1,9'-[9H]fluorenes]. *J. Org. Chem.* **2003,** *68*, 5832-5837.
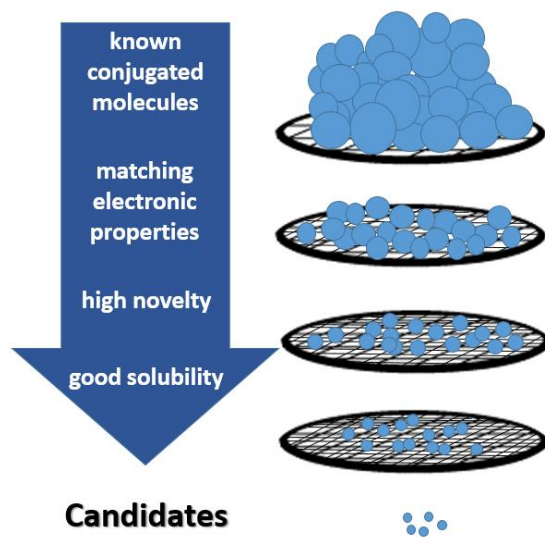
60. Sánchez-Aris, M.; Maestre, I.; Parella, T.; Jaime, C.; Virgili, A.; Alvarez-Larena, A.; Piniella, J. F., Preparation, structural and conformational study of cyclobis[(R,R)-α,α'-bis(trifluoromethyl)-9,10-anthracenedimethylsulphite]. *Tetrahedron: Asymmetry* **2002,** *13*, 1231-1236.

60. Lamm, J.-H.; Horstmann, J.; Nissen, J. H.; Weddeling, J.-H.; Neumann, B.; Stammler, H.-G.; Mitzel, N. W., Poly-Boron, -Silicon, and -Gallium Lewis Acids by Hydrometallation of 1,5- and 1,8-Dialkynylanthracenes. *Eur. J. Inorg. Chem.* **2014**, 4294-4301.

62. Liu, Y.; Zhao, J.; Li, Z.; Mu, C.; Ma, W.; Hu, H.; Jiang, K.; Lin, H.; Ade, H.; Yan, H., Aggregation and morphology control enables multiple cases of high-efficiency polymer solar cells. *Nat. Commun.* **2014,** *5*, 5293.

63. Chen, Y.; Zhang, S.; Wu, Y.; Hou, J., Molecular design and morphology control towards efficient polymer solar cells processed using non-aromatic and non-chlorinated solvents. *Adv. Mater.* **2014,** *26*, 2744-9, 2618.

64. Chen, S.; Zhang, L.; Ma, C.; Meng, D.; Zhang, J.; Zhang, G.; Li, Z.; Chow, P. C. Y.; Ma, W.; Wang, Z., et al., Alkyl Chain Regiochemistry of Benzotriazole-Based Donor Polymers Influencing Morphology and Performances of Non-Fullerene Organic Solar Cells. *Adv. Energy Mater.* **2018,** *8*, 1702427.

TOC graphics