DE-SC0015959

Final Technical Report for DOE/FOA

Project Title: Harnessing the power of ab initio calculations, distributed computing and machine learning to efficiently locate extreme molecules for use in carbon-based solar cells

Recipient: President and Fellows of Harvard College
Office for Sponsored Programs
Harvard University
Richard A. and Susan F. Smith Campus Center, Suite 600
1350 Massachusetts Avenue
Cambridge, MA 02138

PI: Alan Aspuru-Guzik,
Ph.D. Professor
Email: aspuru@utoronoto.ca

DE-FOA-0001204
Basic Energy Sciences – Predictive Theory and Modeling
Dr. James Davenport
Program Manager, Theoretical Condensed Matter Physics
Dr. Mark Pederson
Program Manager, Computational and Theoretical Chemistry

Executive Summary

The use of high-throughput virtual screening (HTVS) tools is a powerful tool to expedite the materials discovery of commercially relevant materials. In previous years, our group has developed a molecular discovery platform to generate libraries in order to obtain suitable candidates for different applications, starting from the Harvard Clean Energy Project [1,2]. This platform is suitable to test in-silico on traditional supercomputing clusters and shared resources, for example, in the IBM World Community Grid.

In this project, we used the molecular discovery platform to create and screen a library of candidates of organic photovoltaic (OPVs) molecules. Based on a set of candidates created with combinations of molecular moieties, we were able to filter, by conformation stability, the energy of electronic orbitals and approximated power conversion efficiencies (PCE). To improve the predictions of orbital energies calculated and the PCEs, we used Gaussian Process regression and two sets of molecules. These sets correspond to electronic structure calculations of a higher level of theory and experimental PCE values, respectively. Finally, we selected a subset of the best candidates (molecules with a PCE higher than 10%) to understand its absorbance properties with TD-DFT.

This project has demonstrated the capabilities of our molecular discovery platform for HTVS. Finally, machine learning can help us to introduce more complex effects included in bulk conditions and computational intensive calculations on models.

Index

Executive Summary 2

Index 3

Introduction 4

Previous work 5

Part 1 6

Achievement 1: Development of a novel molecular library of organic photovoltaics. 6

Achievement 2: Bayesian approach for correction of HOMO-LUMO energies. 7

Achievement 3: Correction of bulk effects with Gaussian Processes 8

Achievement 4: Excited state properties with quantum chemistry 9

Part 2 9

Achievement 5: Fluorination effect on conjugated polymers. 9

Future work 10

Conclusion 11

Outcome publications 11

References 12

Introduction

There is an increasing global demand for energy; for instance, in 2010, the estimated energy composition was 17TW, while it will be 27TW by 2020 [3]. However, the supply of traditional non-renewable sources such as fossil fuels dwindle and bring severe environmental problems, such as greenhouse effects and pollution [4]. Therefore, to stop the long-term consequences of climate change, we need a sustainable and cleaner source of energy. In consequence, the research of renewable energy is of substantial interest to governments, companies, and academics. A viable source of energy can be solar energy since 120,000 TW of power from the sun is deposited in the Earth. In this reward, the development of affordable photovoltaic (PV) cells is one of the most promising long-term solutions for clean, renewable energy.

Over the past three years, the use of PVs to generate electricity has increased [3]. Among the emerging PV technologies, the organic photovoltaics (OPVs) offers are low-cost, flexible, and easy to fabricated options. For instance, they can be manufactured with inexpensive techniques such as roll-to-roll printing[5]. Furthermore, they have shown long lifetimes [6,7] and high-efficiency, for example, a recent article by Zhao et al.[8] have reported a 13% efficiency with a non-fullerene acceptor indacenodithiophene derivative. These advances can be attributed to improvements in device architecture, processing, and materials design [9]. However, to enhance the performance and environmental stability of OPVs, the invention of novel materials is mandatory. One of the main challenges to improve the PV industry is the time that a novel molecule reaches the market. For instance, since on average it takes 25–35 years from the first report to being manufactured in relevant volume for commercial application [10]. Therefore, it is mandatory the speed up of the materials discovery.

A powerful tool to accelerate the materials and molecular discovery in different areas is High-throughput virtual screening (HTVS) [11,12]. Whereas, the number of molecules, materials, and devices that can be tested experimentally is limited by the synthesis, characterization and analytical testing procedures; the possibility of automatically prescreen in-silico before synthesis is promising [13]. In this way, we can reduce some orders of magnitude a set of candidates to just a few promising molecules based on quantum-chemistry or molecular mechanics simulations to experimentally test them.

In this work, we used in-house developed HTVS tools and Machine learning approaches to determine suitable candidates for organic photovoltaics materials.

Previous work

Our group has worked on the development of the appropriate platform and libraries for HTVS. The molecular discovery platform is suitable for traditional resources (like supercomputing clusters) and shared resources (like the IBM World Community Grid). We have employed this platform in screening different materials: organic flow redox battery electrolytes screening for cheap, clean grid storage [14–17]; OLED screenings in collaboration with Samsung [18,19]; and the Harvard Clean Energy Project [1,2].

The Harvard Clean Energy Project (CEP) has coupled ab initio calculations and located tens of thousands of molecules with projected power conversion efficiencies (PCEs) near the expected maximum for single-junction organic solar cells. So far, the CEP has screened 4.3 million molecules with more than 300 million DFT calculations, with computer time donated by volunteers to virtually screen p-type, and later n-type, organic photovoltaic oligomers. They have led to 35,000 molecules with predicted conversion efficiency over 10%, and a database with a size of 2.3 Million has been included in the White House Material Genome Initiative.

Additionally, we have reported a curated data set. It contains quantum chemistry and experimental results for 266 donor materials from bulk heterojunction devices, the Harvard Organic Photovoltaic Dataset (HOPV15)[20].

In each of these projects, we closely collaborated with experimental groups to establish a feedback loop among experiments and theory to increase the chances of the utility of the molecules we screened (Figure 1).

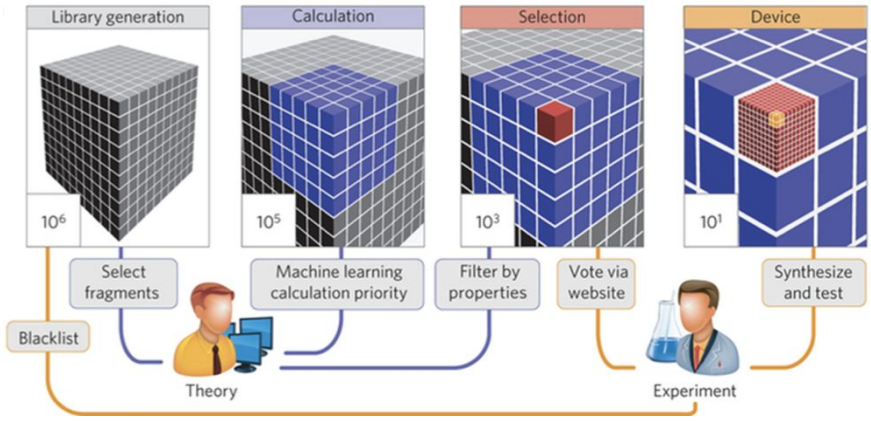

Figure 1: Diagram of the collaborative discovery approach. The distinct screening stages, from left to right, involve different theoretical and computational approaches as well as experimental input and testing.

Part 1

In this work [9], we used the CEP infrastructure (Figure 2) to screen over non-fullerene acceptors materials. Here, we i) created a set of candidates, ii) used our automatic workflow to explore the conformational space of each molecule, iii) performed quantum mechanical calculations to obtain electronic structure information, and iv) calibrated the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) with Gaussian processes (GP) technique.

Achievement 1: Development of a novel molecular library of organic photovoltaics.

The usual successful acceptor materials for OPVs based on derivatives of fullerenes[19], for instance, many consider the standard electron acceptor is [6,6]-Phenyl-C61-butyric acid methyl ester (PCBM) [20]. Therefore, the donor materials have been designed to have an optimal overlap with it. However, the derivatives of fullerenes have their challenges: the production cost, limited absorbance of visible light, and the modulation of their electronic properties. For this reason, there have been some efforts to design and test new non-fullerene acceptor materials with similar energy gaps than PCBM or its derivatives but cheaper to manufacture.

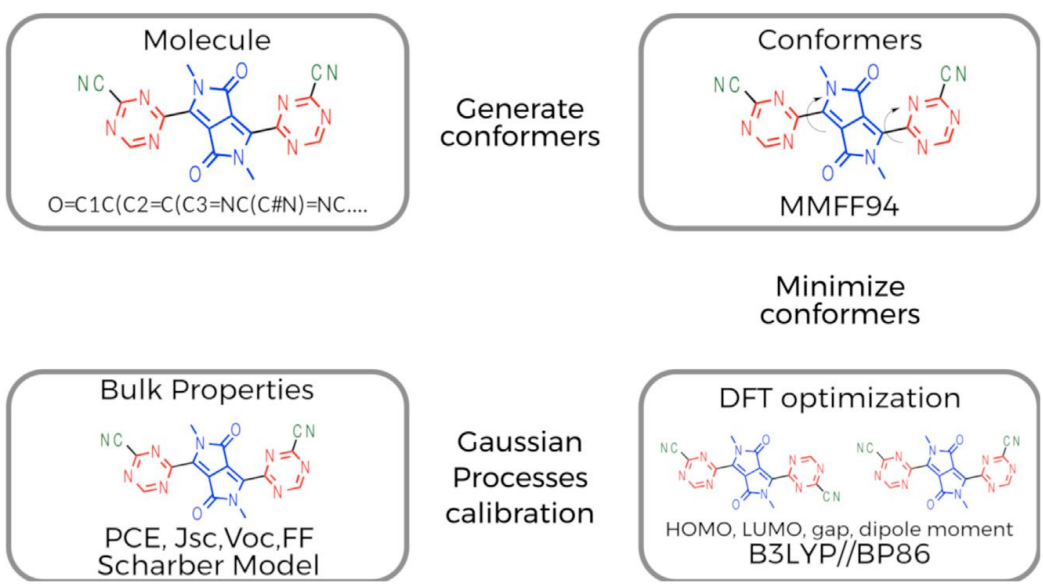

Figure 2. Scheme of the Workflow of the HTVS Process in the Harvard Clean Energy Project

To generate suitable candidates, we selected a set of molecules based on di and monosubstituted fragments. To try improving the chances of synthesizability, each of these fragments and substitutions has been reported in the literature of having a commercial precedent. The full screening is done with different combinations of 13 cores, 49 spacers, and 45 terminal groups (Figure 3). These combinations were restricted to molecules with 306 electrons of less, due to computational resources on the World Community Grid. With each candidate, we obtain 1,500 molecular conformations obtained with RDKit [21] that were minimized with a force field. Only the lowest energy conformers with a certain threshold up to 20 samples were kept, to later being optimized with a DFT method, BP86/def2-SVP. Finally, to improve the accuracy of HOMO and LUMO energies, we used a Gaussian process regression based on the experimental HOMO and LUMO energies of 94 molecules.

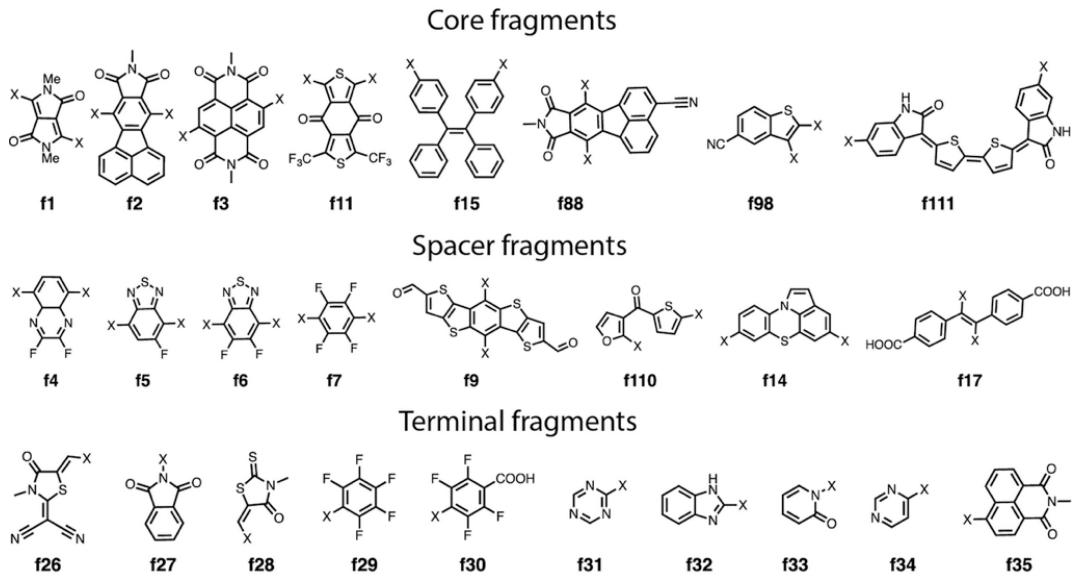

Figure 3: A few samples of molecular fragments used to build candidates.

Achievement 2: Bayesian approach for correction of HOMO-LUMO energies.

We used a Bayesian approach to calibrate the correction GP regression of HOMO-LUMO energies [22]. In the non-fullerene candidate, we use this approach with molecular fingerprints based as a metric for molecular similarity. We used Morgan fingerprints. They capture the topological information of molecules by taking into account the local vicinity of an atom by bonds. And the molecular similarity is measured by the Tanimoto similarity index (Jaccard index). The GPs predictions could be interpreted as weighted averages of training data based on molecular similarity, where the weights are probabilistic.

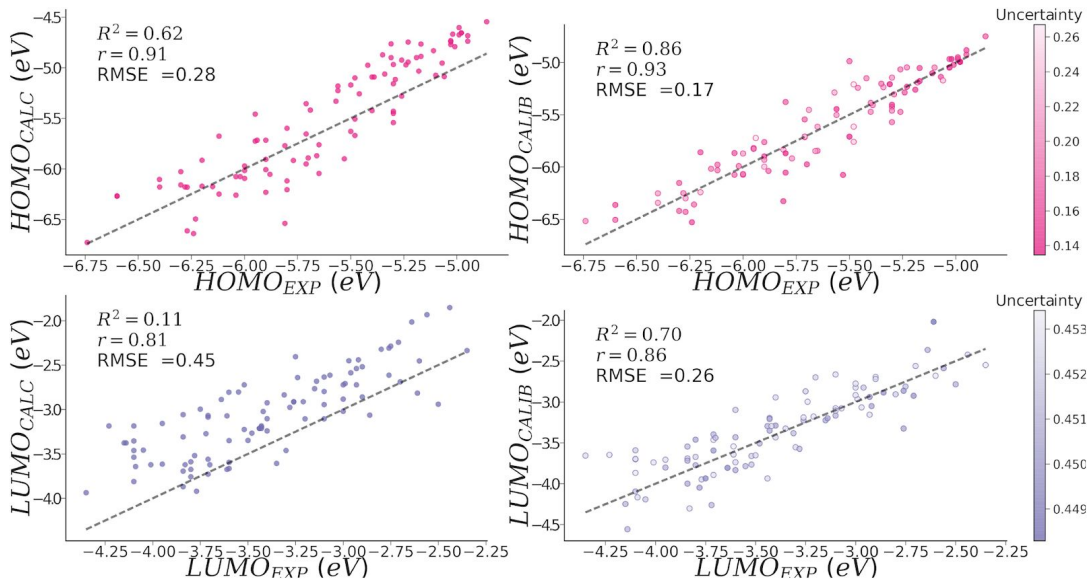

Figure 4. Calculated (CALC left) and GP-Calibrated predictions (CALIB) compared with experimental (EXP), HOMO - LUMO values (eV).

To avoid overfitting, we used leave-one-out cross-validations, where the training is done removing one point of our data, and later the model tries to predict the value of that point. (Figure 4).

Achievement 3: Correction of bulk effects with Gaussian Processes

To compare experimental PCE, we used GPs to take into account the effects not covered by the Scharber model [23]. For this, we used 49 molecules reported with experimental and theoretical PCE values. And instead of fingerprints, we used atomic orbitals energy and molecular weights, whereas instead of Tanimoto distances, we use a radial basis function kernel for euclidean distances (Figure 5).

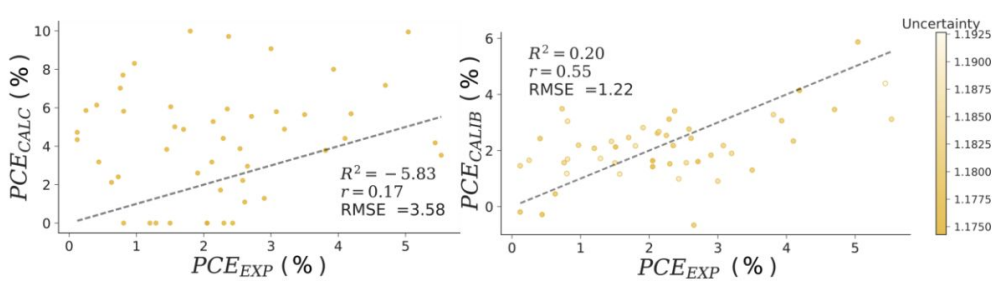

Figure 5. Calculated (CALC left) and GP-Calibrated predictions (CALIB) compared with experimental (EXP), PCE values (%).

Achievement 4: Excited state properties with quantum chemistry

Finally, following the hierarchical computational scheme or down-selection approach facilitated by the HTVS, we selected a subset of the molecules with a higher predicted PCE and compared it with a computational more expensive method, TD-DFT (Figure 6). We computed the S0→S1 transitions in the range of visible light (1.4–2.2 eV) to understand their underlying absorption properties. For instance, the oscillator strength of two reported diketopyrrolopyrroles derivatives agreed with the literature [24]. Since the oscillator strength indicates the likelihood of electronic transition, we can assess their efficiency as a chromophore in fullerene-free OPVs. We found diketopyrrolopyrroles and quinoidal thiophene derivatives are a good promise.

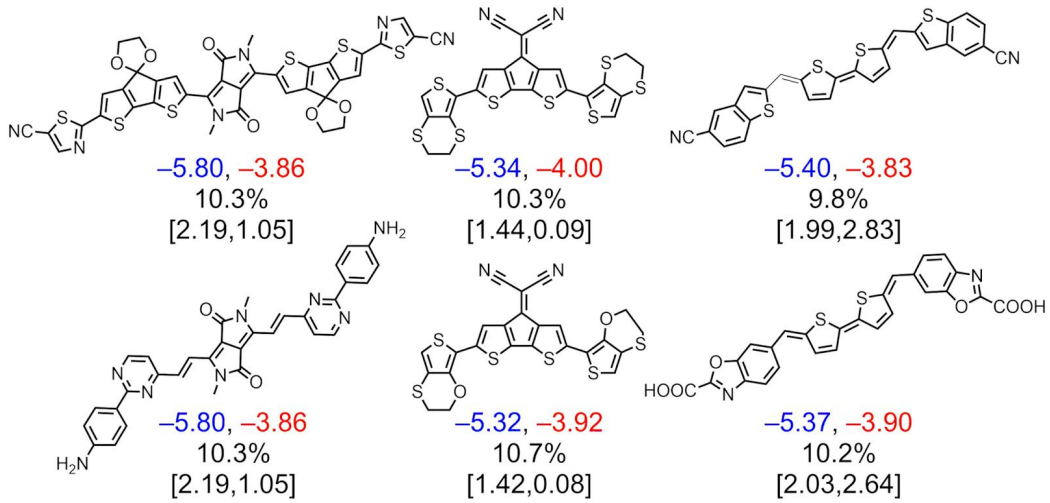

Figure 6: Representative selection of the molecules with the best-predicted PCE molecules from the Quinoidal Thiophene, Dicyanocylopentadienyl, and diketopyrrolopyrroles Fragments. First row, the calibrated HOMO (blue) and LUMO (red) energies. Second row, the estimated PCE. Third row, S0→S1 transition energy (reported in eV).

Part 2

We have closely collaborated and encouraged discussions with experimental laboratories working on photoactive organic materials in order to gain experience and increase the utility of our tools for future applications.

Achievement 5: Fluorination effect on conjugated polymers.

Among one of the projects funded with this grant, we explored collaboration with Lerlerc's group from Université Laval. In this work

[25], we used DFT calculations to compute the optical and electronic properties of different conformations of fluorinated conjugated polymers.

One way to fine-tune the properties of conjugated systems has been the subject of much investigation by introducing functional groups. One approach can be incorporating fluorine to modulate the electronic properties, whereas it favors π–π stacking[26] since it doesn't introduce steric effects. Furthermore, the fluorine-containing materials enhanced the crystallinity by introducing noncovalent intermolecular interactions [27].

In this work, we studied the effect of the fluorination of 3-hexylthiophene-based polymer (P3HT), by varying the substitution degree. We reported 3 different fluorinated ratios, P3HT-50F, P3HT-33F, and P3HT-25F. They were synthesized from monomers prepared iteratively by Migita−Stille couplings from 2-bromo-4-fluoro-3-hexylthiophene. Later, they were polymerized with the Kumada catalyst transfer polymerization.

We studied different conformations with DFT calculations to understand the effect of fluoride substitution with simplified versions of mono-, di-, tri- and tetramers. These results suggest there are fluorine−sulfur interactions that could contribute to the planar conformation (Figure 7) on these truncated models.

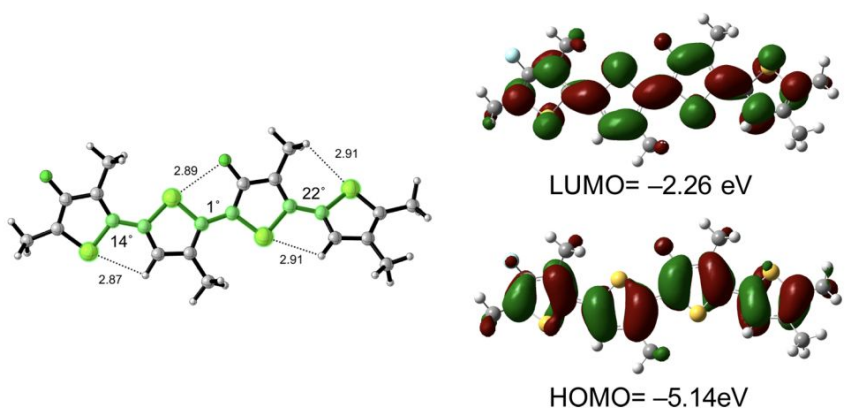

Figure 7. Structure of the lowest energy conformation of P3HT-50F-dimer. The calculations were done using B3LYP/6-311+G(d,p)-D3BJ)-IEFPCM(ε = 4)//B97D/6-31G(d,p)-IEFPCM(ε=4).

Future work

Recently, our group has presented two types of generative machine learning methods to design novel molecules: autoencoders [28] and generative adversarial networks[29]. These methods have the goal to generate molecules with specific target properties, rather than predict properties of a given molecule. These techniques, together with our virtual screening tools, could

be used to inverse design novel molecules. In this way, we could be able based on a given set of molecules, to obtain new ones with a given set of desired properties, rather than building and screen large libraries. This approach could expedite orders of magnitude faster than our old methods once further developed.

For instance, one of our current work is the use of Phoenics[30]. Phoenics is a Bayesian optimizer algorithm than includes generative models. Together with our virtual screening tools, is being applied to the discovery of new redox flow battery electrolytes. Since organic flow batteries require further development to increase their performance stability, we could use the combination of generative models, discriminative models, and virtual screening, to explore new regions of chemical space. Merely employing a discriminative model has been shown to accelerate the virtual screening process by an order of magnitude.

Finally, the ultimate goal is to accelerate the discovery of new materials. Our group is currently working on integrating both HVTS and high-throughput experimentation. Furthermore, going beyond traditional HTVS, our group has been actively working on the development of self-driving laboratories, as the next-generation facilities.

Self-driving laboratories combine machine learning algorithms and automated experimentation platforms. Since the ML models can identify complex relations between variables, they can be used for decision-making without human intervention [31]. For instance, recently, our group has collaborated with an experimental laboratory to increase the efficiency and stability of OPVs [32] by high-throughput and autonomous experimentation methods. Using our experimentation platform, ChemOS an active learning algorithm [33], we were able to optimize films in an automated fashion with machine learning-enabled autonomous experimentation that is able to identify the stable compositions.

Conclusion

The DOE grant number DE-SC0015959 has allowed us major advances in the infrastructure of our molecular discovery platform. Our studies in novel libraries of OPVs have shown its potential and readiness to use in other applications. Additionally, they have enabled us to encourage further collaborations with experimental groups, to engage in the conversation of understanding, predicting, and designing novel materials with a target property.

Outcome publications

1. Lopez SA, 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.

2. Blaskovits JT, Bura T, Beaupré S, Lopez SA, Roy C, de Goes Soares J, et al. A Study of the Degree of Fluorination in Regioregular Poly(3-hexylthiophene). Macromolecules. 2017;50: 162–174.
3. Ly JT, Lopez SA, Lin JB, Kim JJ, Lee H, Burnett EK, et al. Oxidation of rubrene, and implications for device stability. J Mater Chem. 2018;6: 3757–3761.

References

1. Hachmann J, Olivares-Amaya R, Jinich A, Appleton AL, Blood-Forsythe MA, Seress LR, et al. Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project [Internet]. Energy Environ. Sci. 2014. pp. 698–704. doi:10.1039/c3ee42756k

2. Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Sánchez-Carrera RS, Gold-Parker A, et al. 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.

3. Capuano L. International energy outlook 2018 (IEO2018). US Energy Information Administration (EIA): Washington, DC, USA. 2018;2018: 21.

4. Solomon S, Qin D, Manning M, Others. Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC (Vol. 4). Cambridge university press Cambridge; 2007.

5. Krebs FC, Espinosa N, Hösel M, Søndergaard RR, Jørgensen M. 25th Anniversary Article: Rise to Power - OPV-Based Solar Parks [Internet]. Advanced Materials. 2014. pp. 29–39. doi:10.1002/adma.201302031

6. Mazzio KA, Luscombe CK. The future of organic photovoltaics. Chem Soc Rev. 2015;44: 78–90.

7. Lu L, Zheng T, Wu Q, Schneider AM, Zhao D, Yu L. Recent Advances in Bulk Heterojunction Polymer Solar Cells. Chem Rev. 2015;115: 12666–12731.

8. Zhao W, Li S, Yao H, Zhang S, Zhang Y, Yang B, et al. Molecular Optimization Enables over 13% Efficiency in Organic Solar Cells. J Am Chem Soc. 2017;139: 7148–7151.

9. Lopez SA, 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.

10. Tabor DP, Roch LM, Saikin SK, Kreisbeck C, Sheberla D, Montoya JH, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. Nature Reviews Materials. 2018;3: 5–20.

11. Pyzer-Knapp EO, Suh C, Gómez-Bombarelli R, Aguilera-Iparraguirre J, Aspuru-Guzik A. What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery. Annu Rev Mater Res. 2015;45: 195–216.

12. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;432: 862–865.

13. Gómez-Bombarelli R, Aspuru-Guzik A. Machine Learning and Big-Data in Computational Chemistry. In: Andreoni W, Yip S, editors. Handbook of Materials Modeling : Methods: Theory and Modeling. Cham: Springer International Publishing; 2018. pp. 1–24.

14. Huskinson B, Marshak MP, Suh C, Er S, Gerhardt MR, Galvin CJ, et al. A metal-free organic–inorganic aqueous flow battery. Nature. 2014;505: 195.

15. Er S, Suh C, Marshak MP, Aspuru-Guzik A. Computational design of molecules for an all-quinone redox flow battery. Chem Sci. 2015;6: 885–893.

16. Gerhardt MR, Tong L, Gómez-Bombarelli R, Chen Q, Marshak MP, Galvin CJ, et al. Anthraquinone Derivatives in Aqueous Flow Batteries. Adv Energy Mater. 2017;7: 1601488.

17. Lin K, Gómez-Bombarelli R, Beh ES, Tong L, Chen Q, Valle A, et al. A redox-flow battery with an alloxazine-based organic electrolyte [Internet]. Nature Energy. 2016. doi:10.1038/nenergy.2016.102

18. Gómez-Bombarelli R, Aguilera-Iparraguirre J, Hirzel TD, Duvenaud D, Maclaurin D, Blood-Forsythe MA, et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. Nat Mater. 2016;15: 1120–1127.

19. Gómez-Bombarelli R, Aspuru-Guzik A. Computational discovery of organic LED materials. Computational Materials Discovery. Royal Society of Chemistry; 2018. pp. 423–446.

20. Lopez SA, Pyzer-Knapp EO, Simm GN, Lutzow T, Li K, Seress LR, et al. The Harvard organic photovoltaic dataset. Sci Data. 2016;3: 160086.

21. Landrum G, Others. RDKit: Open-source cheminformatics. 2006.

22. Pyzer-Knapp EO, Simm GN, Guzik AA. A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials [Internet]. Materials Horizons. 2016. pp. 226–233. doi:10.1039/c5mh00282f

23. Scharber MC, Mühlbacher D, Koppe M, Denk P, Waldauf C, Heeger AJ, et al. Design Rules for Donors in Bulk-Heterojunction Solar Cells—Towards 10 % Energy-Conversion Efficiency. Adv Mater. 2006;18: 789–794.

24. Vezie MS, Few S, Meager I, Pieridou G, Dörling B, Ashraf RS, et al. Exploring the origin of high optical absorption in conjugated polymers [Internet]. Nature Materials. 2016. pp. 746–753. doi:10.1038/nmat4645

25. Blaskovits JT, Bura T, Beaupré S, Lopez SA, Roy C, de Goes Soares J, et al. A Study of the Degree of Fluorination in Regioregular Poly(3-hexylthiophene). Macromolecules. 2017;50: 162–174.

26. Zhang Y, Yao H, Zhang S, Qin Y, Zhang J, Yang L, et al. Fluorination vs. chlorination: a case study on high performance organic photovoltaic materials. Sci China Chem. 2018;61: 1328–1337.

27. Reichenbächer K, Süss HI, Hulliger J. Fluorine in crystal engineering—“the little atom that could.” Chem Soc Rev. 2005;34: 22–30.

28. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent Sci. 2018;4: 268–276.

29. Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik A. Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models [Internet]. arXiv [stat.ML]. 2017. Available: http://arxiv.org/abs/1705.10843

30. Häse F, Roch LM, Kreisbeck C, Aspuru-Guzik A. Phoenics: A Bayesian Optimizer for Chemistry [Internet]. ACS Central Science. 2018. pp. 1134–1145. doi:10.1021/acscentsci.8b00307

31. Häse F, Roch LM, Aspuru-Guzik A. Next-Generation Experimentation with Self-Driving Laboratories. Trends in Chemistry. 2019;1: 282–291.

32. Langner S, Häse F, Perea JD, Stubhan T, Hauch J, Roch LM, et al. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multi-Component Systems [Internet]. arXiv [physics.app-ph]. 2019. Available: http://arxiv.org/abs/1909.03511

33. Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LPE, Hein JE, et al. ChemOS: Orchestrating autonomous experimentation. Science Robotics. 2018;3: eaat5559.