DOE BES Research Project Final Report

Aspuru-Guzik, Alán

Towards 3rd generation organic tandem solar cells with 20% efficiency: Accelerated discovery and rational design of carbon-based photovoltaic materials through massive distributed volunteer computing

Introduction

Fossil fuels now dominate human energy consumption. However, the burning of fossil fuels brings serious environmental pollution and exacerbates the greenhouse effect. Clean, affordable, and renewable energy sources are urgently needed to satisfy the 10s of terawatts (TW) energy need of human beings. Solar cells are one promising choice to replace traditional energy sources. Each year, the sun deposits 120,000 TW of power onto the surface of the Earth, far more than the 13 TW of total power currently used by the planet's population.[1] Current solar cell technology is expensive to assemble, and often results in fragile products. Additionally, the Fraunhofer Report has suggested that the average solar cell takes between 1 and 2.5 years to compensate for the energy that was used to build it. Assuming a three-fold reduction in manufacturing cost, it is unlikely that the cost will reduce to a truly marketable level. Current solar cells are required to be 100 μm thick and require high quality silicon wafers; these must be electrically connected in a batch process. Direct band gap semi-conductors represent an alternative approach because of their relatively low cost of manufacturing (3-5 fold cheaper). However, these materials require rare elements such as indium and tellurium. Ultimately, the scarcity of these metals will limit the ubiquity of their corresponding solar cells.

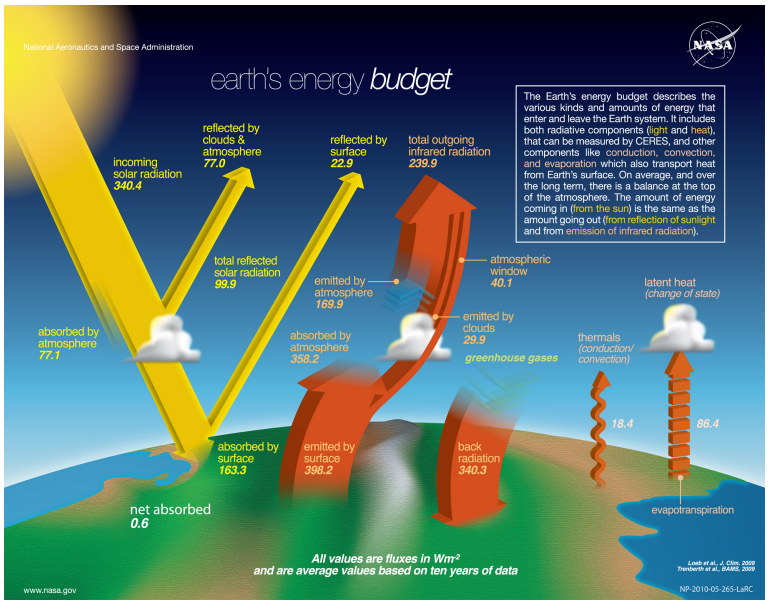

Figure 1. The Earth's energy budget – a graphical view on the various types of energy entering and leaving the Earth.[2]

A more attractive alternative is materials based upon carbon – organic photovoltaics (OPVs). Organic photovoltaic cells (OPVs) are an emerging technology because of their many advantages: easy fabrication, low cost, lightweight, diverse produced materials, and device flexibility.[3] The materials that make up OPVs are often crystalline small molecules or amorphous polymers. These materials are comprised of some of the most abundant elements on the planet. Deposition typically occurs quickly at room temperature, and atmospheric pressure, there is the potential for manufacturing them roll to roll. Optimizations of manufacturing processes and device performance could result in OPVs with 15% PCEs at a cost of *circa* \$30 m^{-2} in the next decade.

The next decade of experimental and theoretical research will overcome the major challenges of relatively poor efficiencies and limited lifetimes of OPVs. Our computational approach has paved the way towards the exploration of vast areas of chemical space via an *in silico* screening technique. We have fine-tuned molecular structure and frequently navigate areas of chemical space where undesirable properties are minimised, while desirable properties are enhanced. We combine chemical intuition with powerful theoretical techniques to identify top materials. We have established theory-experiment feedback loops with the Briseno (University of Massachusetts, Amherst) and Bao (Stanford) groups.

Achievement 1: Public Release of CEPDB

In June 2013, the CEPDB[3] – a database containing 2.3 million results – was released to the public as part of the White House Materials Genome Initiative.[4] Since that time, the database has been expanded to include 4.3 million molecules and 65 million calculations. The CEPDB is hosted on www.molecularspace.org and is a user-friendly interface for exploring the dataset by ranking, filtering, and querying molecular properties. The CEPDB can also accept new experimental results (*i.e.* crystal structures) to be queried by the user. Electronic structure calculations of CEPDB molecules can be benchmarked against the density functionals employed in the CEP. We have also evaluated the numerical stability of the algorithms underpinning their implementation can be undertaken. We have coordinated our efforts with the developers of the quantum chemistry program *Q-Chem,*[5] which allows us to have an influence upon the development of state of the art software.

Achievement 2: Update of CEPDB website

In 2015, the CEPDB website was updated to provide a better user experience and more capability. The new website provides better integration with ChemDoodle to allow users to draw molecules without having to input smiles strings. In addition, the site allows for similarity searches using cutting edge locality sensitive hashing to find similar molecules. Users can download optimized geometries to speed up higher-level calculations with the pre-optimized geometry starting point. Figure 2 shows four screenshots from our website, www.molecularspace.org and searching capabilities, therein.

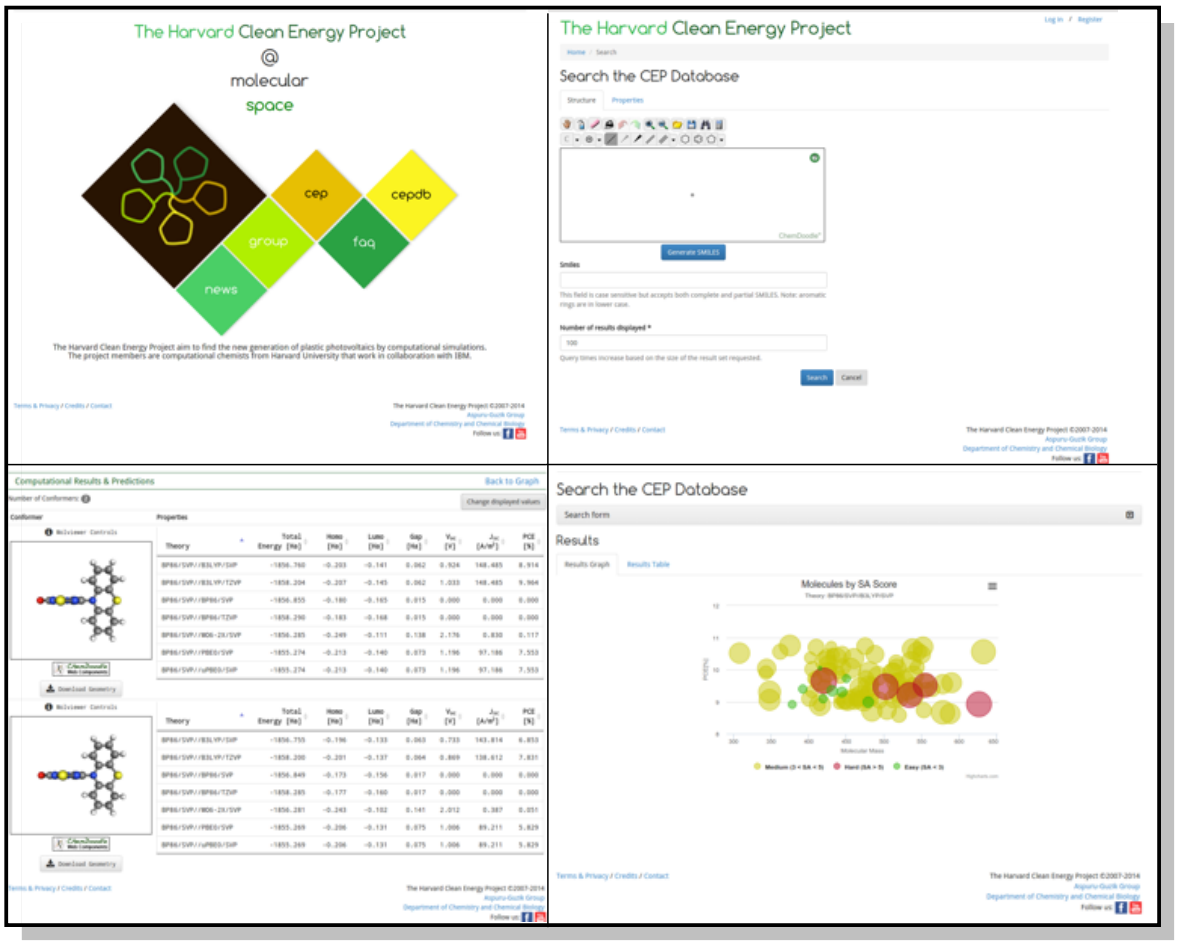

Figure 2. A typical screen-shot showing the CEPDB website front page (top left), search screen (top right), a detailed view of a molecule and its properties (bottom right), and an example of a unique plot of results (bottom right).

Achievement 3: Continued Expansion and Updating of the Harvard Clean Energy Project and Database

The Harvard CEP is our vehicle for the high throughput virtual screening of materials to identify donor materials for high efficiency OPV devices.[6] In addition to computed properties, output files from the calculation (in plain-text and binary formats) are stored and linked to the relevant entries in the database. This large database utilizes an extensive file store, featuring a capacity well over 800 Tb. The IBM World Community Grid powers our vast computational undertaking.[7] It is a distributed computer framework, where volunteers donate computational resources via a screensaver mechanism. We processed 70,000 conformations a day, which is equivalent to 25,000 CPU years. Figure 3 shows graphical representations of runtime (left) and number of results (right) as a daily rate, a 30-day rolling average, and a cumulative figure.

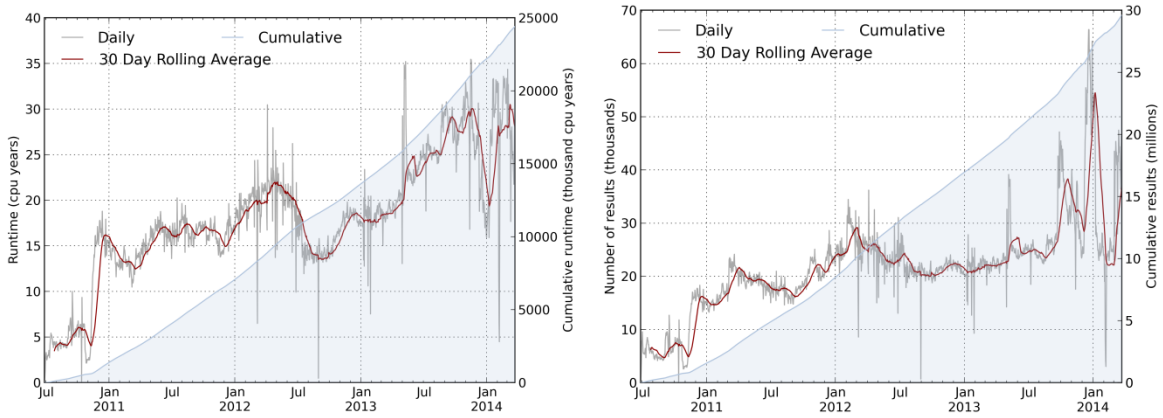

Figure 3. Performance of the World Community Grid with respect to runtime (left) and number of results returned (right)

Figure 4 shows a graphical representation of our approach to this research. We depend on a tight feedback loop between our theoretical results and experimental collaborators. Our calculated electronic structures are related to observable physical properties with the Shockley-Queisser model known as the Scharber model.[9] To minimize systematic errors introduced by the functional and basis set chosen, we average over a range of calculations, and adjust these averaged results based upon a calibration scheme derived from experimentally reported data.

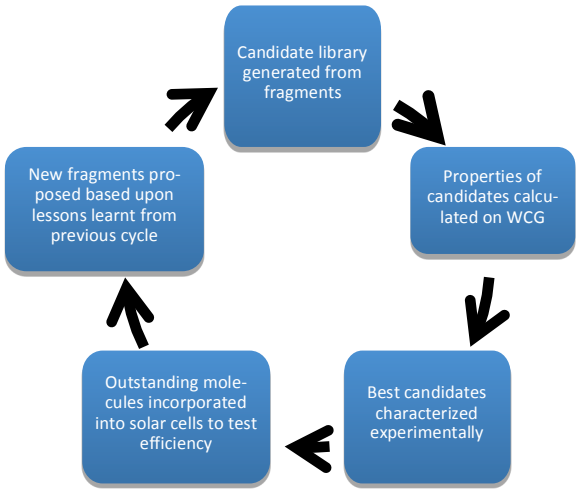

Figure 4. A positive feedback loop describing the principles used in the design of novel OPV materials

The Scharber model predicts power conversion efficiency (PCE) as a function of the frontier molecular orbital energies of electron donor and acceptor materials. As employed here, it is a metric by which the ability of a molecule to donate electrons can be assessed. Choosing which region of chemical space to navigate is clearly challenging due to the nearly infinite possible molecules. We seek to improve the infrastructure to facilitate this form of discovery.

The current SQL database is not conducive to highly structured queries, especially when moving between tables. The return time to query the best performing molecule is ~40 minutes. Machine learning requires a greater diversity of information, and a means for fast querying of data structures. In order for this to be made possible, a restructuring of the database backend to the CEP was necessary. We are now in the final stages of implementing a much more sophisticated scheme based upon a noSQL framework, powered by MongoDB. This more flexible system allows a great deal more data to be stored and queried quickly, since it is easy to add additional fields when needed without a significant restructure. This increased flexibility will also aid in the full utilization of our data, not just by us, but also by collaborations; our ability to alter the structure in an agile manner will allow easy facilitation of a range of queries and storage request. Our new database will, in addition to an increase in the amount of information stored from calculations, allow the storing of a large number of molecular descriptors and fingerprints which will aid our development of pre-screening models to increase efficiency.

In contrast to the relational database model (RDBM) of SQL solutions, MongoDB uses a document-orientated approach for storing data. Documents are comprised of key-value pairs, and can be thought of as similar to JSON objects. The value returned from a key lookup can be another document; this is known as embedding. This support for embedding greatly reduces I/O activity on a database since it reduces the need for JOIN calls, and hence can greatly increase performance. It is clear that the choice of when to embed a document will play a large role in the overall performance of the solution; the overall structure of the database will be discussed in detail. In determining when to embed, and how to embed, the nature and frequency of querying were examined, with preference given to high-performance solutions, with helper functions implemented to aid the user in cases when the solution seems counter intuitive. Figure 5 shows a view of a molecule stored within our MongoDB storage solution; each document has a related embedded metadata document, and this has two main uses.

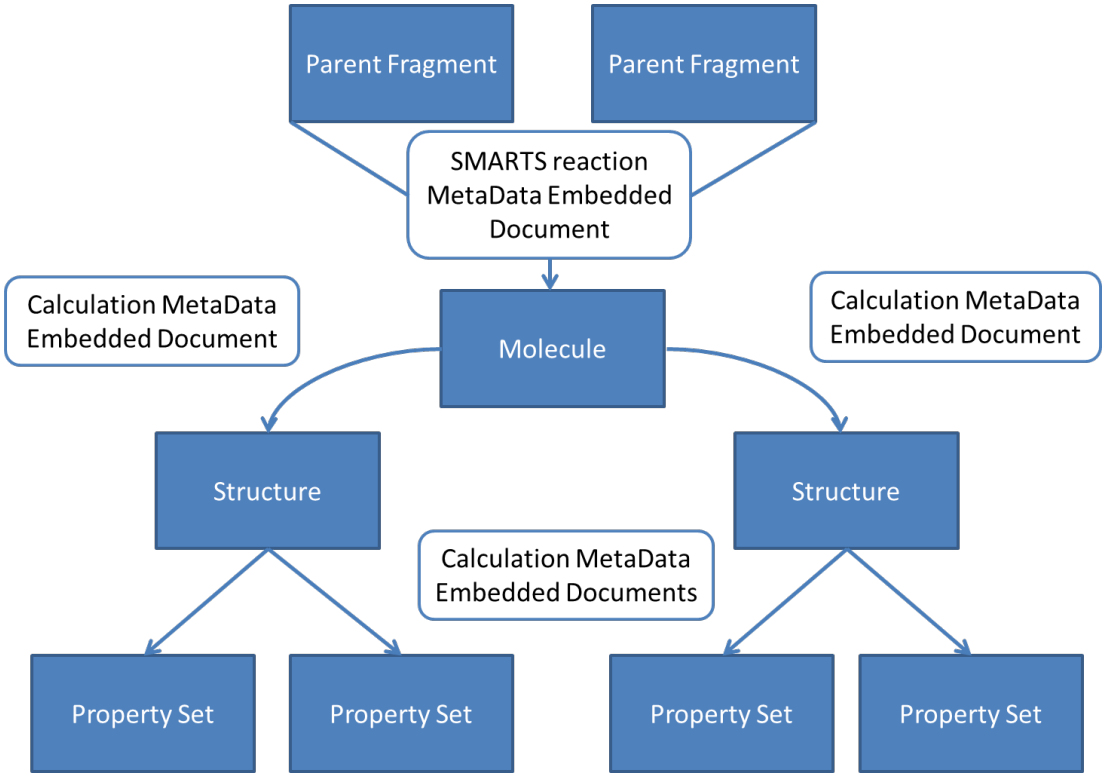

Figure 5. A simplified view of a molecule, as held in the MongoDB structure. Arrows represent bi-directional links between documents, and are embedded within the parent and child document. Additional MetaData is embedded within documents, which improves performance by reducing the number of steps required in a query.

This metadata stores information relevant to the calculation or structure, (e.g. level of theory used, program version the calculation was performed with) which allows for easy reproduction and validation of results. Furthermore, the metadata allows a query to relate different documents to each other, thus improving amalgamated querying. The parent fragments document represents a good example of tuning the structure of the database to the nature of the querying. Initially these documents were embedded within the Molecule document, which represents an intuitive storage solution. It is common to try and link these fragments, which are produced in the molecular generation, to their corresponding molecule documents. Searching within documents embedded within a list represented a major performance hit, and so these fragments were promoted from embedded document to document, and links to these documents were stored within the original molecule documents. This solution has shown significant speedup, for all queries, which require the engine to return a full document, since the size of each of these is now reduced.

Achievement 4: Improvement to SMARTS based molecular generation methods, and integration to new database structure.

The ability to move from fragments to libraries of molecules is an essential part of the CEP workflow.[6] We have implemented a python based solution based around the RDKIT module[8] to allow the rapid combinatorial generation of molecules from fragment sets. Projected timings based upon generation of approximately half a million fragments suggest that our new implementation reduces the time for generation from approximately one month to just one day. In addition to an improvement in speed of the solution previously employed, we have implemented a scheme that allows 'protection' of reactive sites. This affords a much finer control over how the combinatorial molecule generator explores space – for instance we can now easily generate molecules, which are functionalized in a symmetric manner, which has lead to improved synthesizability of generated molecules.

Our molecular generator is also capable of storing a wide range of meta-data associated with each stage of the generation. This affords the potential for more sophisticated data analysis, leading to a greater understanding of factors that affect performance of the materials in OPV devices. This data is integrated into our new database structure, demonstrating how the flexibility of our data model can improve understanding of the factors, which influence the performance of our candidate molecules. We anticipate further improving the molecular generator with a Monte Carlo type routine with an acceptance criteria based upon statistics built from our current database.

Achievement 5: Framework for linking and tracking calculation lifecycle

Currently, the CEPDB provides a framework, which facilitates linking of calculations (*i.e.* using the geometry from one method/program and directly feeding it into another for additional calculations) and the automated parsing and storage of the results. This represents a great improvement to the semi-manual approach. The framework has been developed within the group as a framework to automate a large proportion of the jobs (Figure 6).

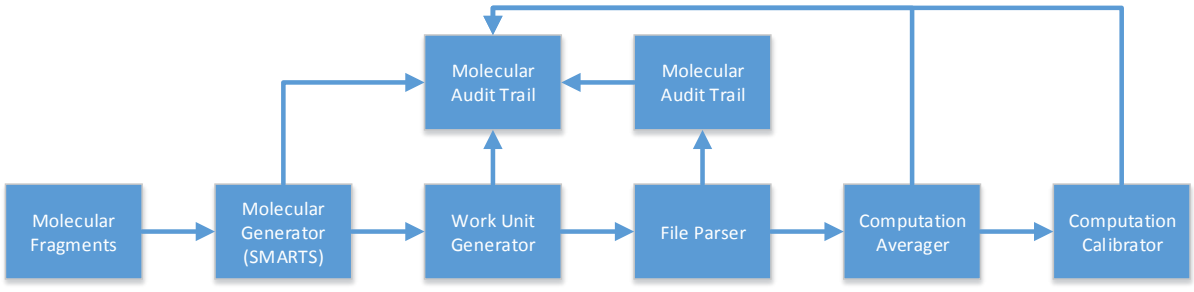

Figure 6. An overview of the CEPDB framework for automation.

This will also greatly increase the ease with which advances in implementation and analysis tools with the scientific community since we can now track and integrate new information and elements into our calculation pipeline.

The following section provides summaries of our publication record for DOE grant number DE-SC0008733. The publications fall into two classes of research: 1) Harvard Clean Energy project advances and 2) investigations of structure-device property relationships for organic semiconducting materials, and are presented in this order.

Part I

Achievement 6: What is High Throughput Virtual Screening? A Perspective from Organic Materials Discovery.

Edward O. Pyzer-Knapp, Changwon Suh, Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, and Alán Aspuru-Guzik *Annu. Rev. Mat. Res.* **2015**, *45*, 195-216.

This perspective described the concept of virtual screening (HTVS) to the community and how data-driven analyses can expand the scope of our libraries to identify the next generation of materials for organic electronic devices. This work dissects the process of HTVS into four philosophies that can be taken when attempting to screen millions of molecules: **significant timescale**, **automated techniques**, **data-driven discovery**, and **computational funnels**.

The first aspect of HTVS that will be addressed is molecular diversity of libraries. We lack absolute axes to survey molecular space and we do not have universal metrics to assess similarity. The similarity metric that is most useful – ultimately how diverse the library is– depends on chemical intuition. The generation of molecular libraries requires the consideration of: 1) all possible combinations of a given set of fragments in a combinatorial way according to a set of rules and 2) a size limit for the molecules (maximum atom and electron count and/or molecular mass), where the growth of the molecule will cease. The computational efficiency can be improved by hard coding some constraints into the library generation software, such as the presence of specific functional groups, moieties that are fundamentally unstable and/or limit device performance. Figure 7 shows how fragments are linked and fused to form molecules that comprise our libraries. These libraries have been generated for organic photovoltaics, organic-based flow batteries, and blue organic light-emitting materials.

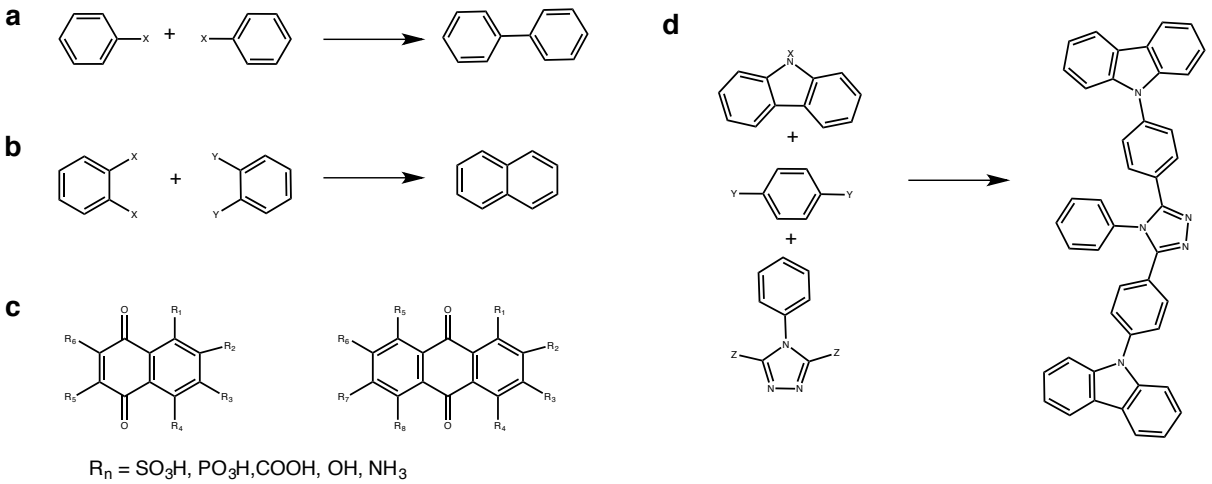

Figure 7. Reactions and combinations in virtual library enumeration. (*a*) The linking procedure used in the blue organic light-emitting diode (OLED) project. (*b*) The fusion procedure additionally utilized in the Clean Energy Project. (*c*) The enumeration of the different substitution positions considered in the organic-based flow battery project. (*d*) Combination of a donor molecule, a bridge molecule, and an acceptor molecule to give a potential blue OLED material.

The selection of simulation techniques must be guided by clearly defined objectives in terms of the desired physical and chemical properties. It is extremely important to select computational techniques that have the desired accuracy, but do not become computationally prohibitive when applied a library containing millions of molecules. This article describes how the method and computed properties are chosen for our three major screening projects: **The Materials Project**, **Harvard Clean Energy Project**, and **Organic-Based Flow Batteries**. The types of databases that are typically used for data volume of this size fall into two categories: Structured Query Language (SQL) and nonrelational database (NoSQL). The databases ultimately should be selected with care for the nature of the data being handled. Figure 8 demonstrates the basic differences between SQL and NoSQL databases.

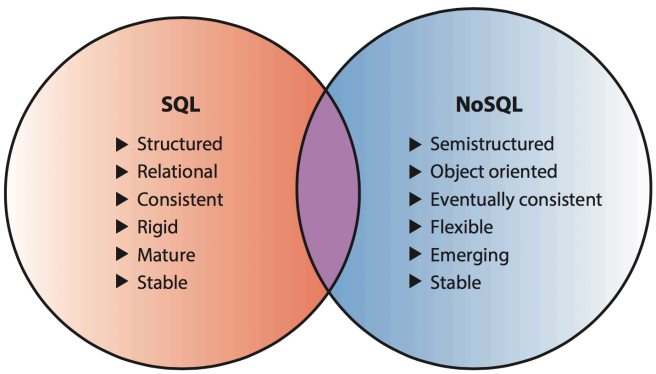

Figure 8. A comparison between SQL and NoSQL architectures.

Although SQL databases allow for transactional integrity, we believe that the enhanced flexibility of NoSQL databases makes them the ideal choice for high-throughput virtual screening because they can be easily modified to adapt to changing data and requirements. Clearly, the substantial volume of data generated in a screening requires innovative approach to identify trends and target materials. Calibrating results to experimental data can

overcome deficiencies in specific methods. Exploratory, as well as exploitative, experimental results are crucial for the continued success of high-throughput virtual screening. This will allow for screening methods to be brought to many more communities beyond the current group of scientists that use them. Achievements 8 and 9 describe our successful efforts in applying exciting machine learning techniques to the CEP.

Achievement 7: Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project.

Johannes Hachmann, Roberto Olivares-Amaya, Adrian Jinich, Anthony L. Appleton, Martin A. Blood-Forsythe, László R. Seress, Carolina Román-Salgado, Kai Trepte, Sule Atahan-Evrenk, Süleyman Er, Supriya Shrestha, Rajib Mondal, Anatoliy Sokolov, Zhenan Bao, and Alán Aspuru-Guzik *Energy Environ. Sci.* **2014**, *7*, 698-704.

We utilized the intrastructure of the Harvard CEP (described above) to explore the chemical space of 2.3 million molecules using 150 million DFT calculations. The 2.3 million molecules are electron-donor materials for OPV devices and were combinatorially generated from an initial 26 fragments, identified by the Bao group. The PCE values, computed using the Scharber model were used to rank the donor ability of each material. The PCE values reported in this work correspond to a standard phenyl-C_{61}-butyric acid methyl ester (PCBM) acceptor counterpart. The theoretical maximum of PCE based on the Scharber model is 11.1%. Of the 2.3 million molecules, 0.04% show a PCE of 11% or higher, and 1.5% over 10%; most are predicted to have a value below 4%. Three of the top candidates are shown in Figure 9.

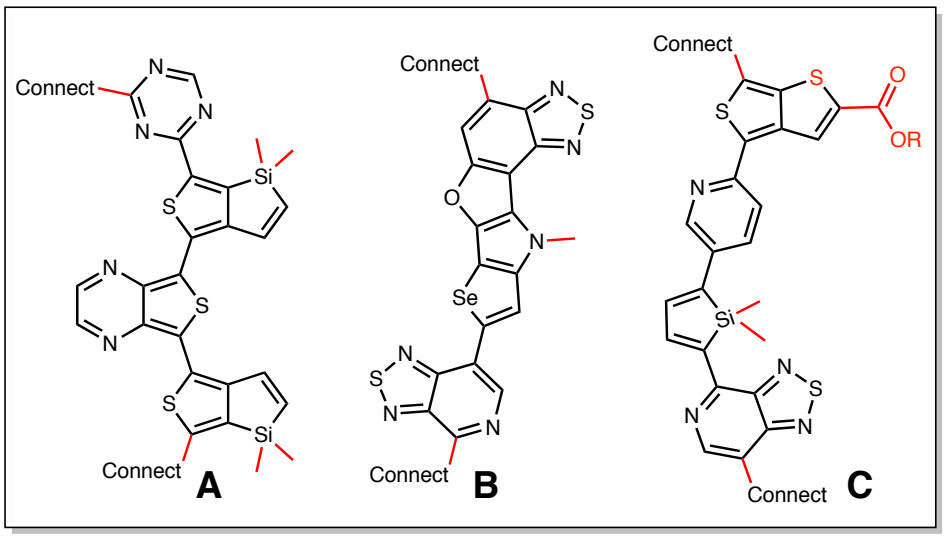

Figure 9. Example structures from the top candidates list (each with potential modifications marked in red).

Molecule A is ranked #77 and has multiple inter-monomer nitrogen-sulfur interactions. These tend to feature highly planarized solid-state structures and can enhance the electronic coupling between monomers. Molecule B is ranked #5. It has a five-ring fused heterocyclic co-monomer structure, which may reduce reorganization and relaxation energies. Molecule C is the top ranked molecule; minor modifications marked in red, results in Yu's highly

efficient thienothiophene co-monomer.[9] This co-monomer has been utilized in organic photovoltaic materials that have consistently surpassed 7.0% power conversion efficiency.

A statistical analysis with respect to the occurrence of the molecular building blocks used in the library generation was performed using a hypergeometric distribution analysis to assess the prevalence of the 26 fragments in the top candidates. Figure 10 shows the fragments that appear more often in top candidates (green) and those fragments that appear more often in poor candidates (red).

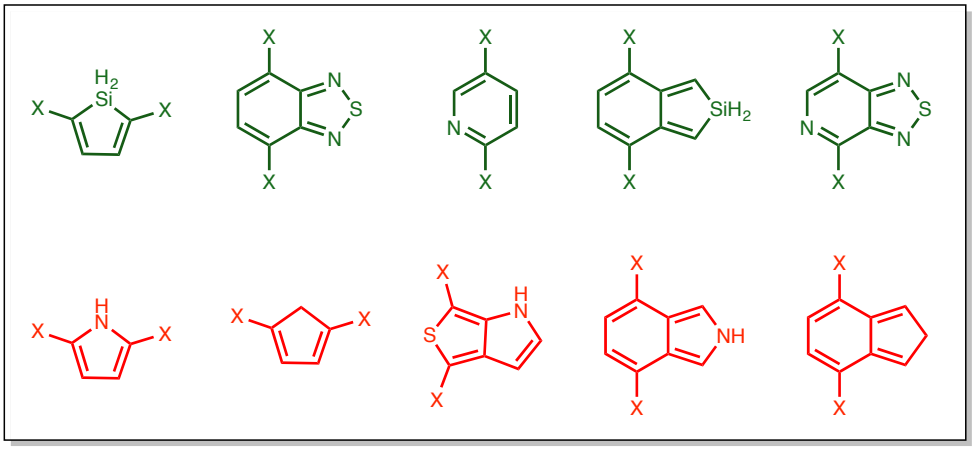

Figure 10. Moieties with the most amplified occurrence in the top candidates (relative to the statistical expectation) are highlighted in green, and red indicates the ones with the most decreased occurrence

The insights will aid in the transition from a brute-force searching approach towards the active design and engineering of new materials.

Achievement 8. Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery

Edward O. Pyzer-Knapp, Kewei Li, and Alan Aspuru-Guzik *Adv. Funct. Mater.* **2015**, *25*, 6495-6502.

The Harvard Clean Energy Project is a rich resource containing the electronic structure of many donor materials for OPVs. In this study, we applied the machine learning technique of neural nets to calibrate our data to predictably replicate experimental data. Neural nets have been used extensively in medicinal chemistry.[10] Artificial neural networks are comprised of interconnected layers of neurons, which are connected to each other along paths (See Figure 11).

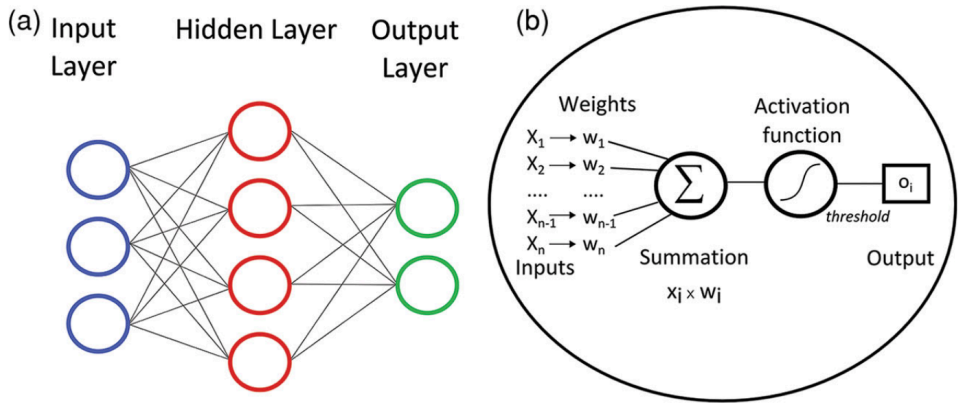

Figure 11. a) In a neural network, features typically enter through a linear input layer (blue) pass through a nonlinear hidden layer (red) before being combined by a linear output layer (green), which produces the target estimation. b) The operation of a single neuron, where inputs are combined with weights, pass through an activation function and produce an output that can be connected with further neurons.

We have further optimized his process by implementing a parallelized training of the neural net using the Hogwild algorithm[11] as implemented by Microsoft's Project Adam.[12] This lead to a reduction of training time from nearly two weeks to approximately four hours. This led to a significant reduction of the mean absolute error and fewer epochs. We selected a set of 200,000 molecules from the CEP to train with and tested on 50,000 CEP molecules. The resulting mean average errors for the HOMO, LUMO, and PCE are shown in Figure 12.

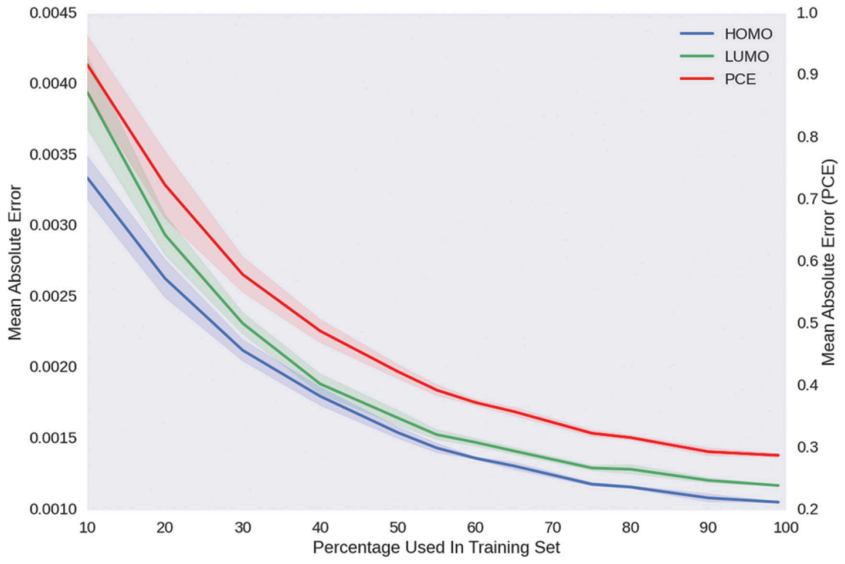

Figure 12. The MAE of the network as the data size is increased from 20,000 as a function of dataset proportion.

Figure 12 shows that reasonable performance can be obtained with relatively small sets, and that relative improvements decrease as 100% of the library size is approached. We have demonstrated that within this restriction in the context of focused HTVS efforts, multilayer perceptrons can be trained to predict frontier molecular orbital energies and PCEs to a very good accuracy. Additionally, nonthread locking techniques significantly reduce the training time of the MLP, allowing the training of significantly larger data sets, which further increase the accuracy.

Achievement 9: A Bayesian Approach to Calibrating High-Throughput Virtual Screening Results and Application to Organic Photovoltaic Materials.

Edward O. Pyzer-Knapp, Gregor N. Simm, and Alán Aspuru-Guzik. arXiv:1510.00388

In this work, we present an advance upon this calibration technique, which takes into account both quantum chemical information, and information about the molecular graph. In addition, this technique reports an uncertainty alongside each calibration – providing a confidence that the method is being used appropriately. We recently reported the Harvard Organic Photovoltaic Dataset (HOPV15), which contains experimental results for 266 donor materials from bulk heterojunction devices, alongside corresponding quantum-chemical calculations performed using various functionals and basis sets.

We sought to determine if systematic failings were related to the chemical structures of the casting this problem into molecular space could afford a method for applying appropriate corrections, which take into account the chemical makeup of the molecules in question. Gaussian processes were used to learn the deviations of computational results from their experimental analogues. Figure 13 shows a plot of experimental vs. theoretical HOMO and LUMO energies before and after calibration.

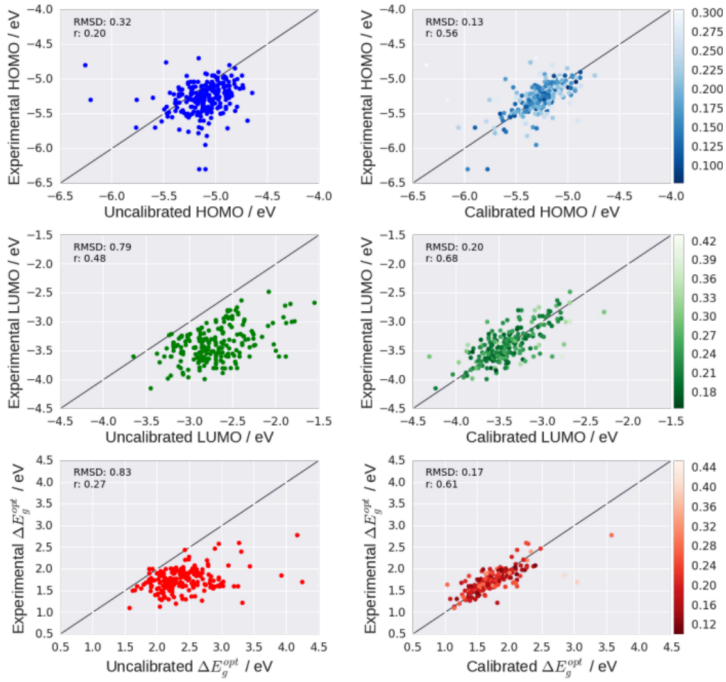

Figure 13. The results of calibrating B3YP/def2-SVP quantum-chemical results for the HOMO, LUMO energies and optical gap to the experimental HOPV15 data set. The uncertainty in the calibrated values is represented in the fill color. Lighter colors represent more uncertain values.

Figure 14 shows the HOMO energies for the molecules within the HOPV15 data set, as computed by various density functionals before (top) and after (bottom) calibration.

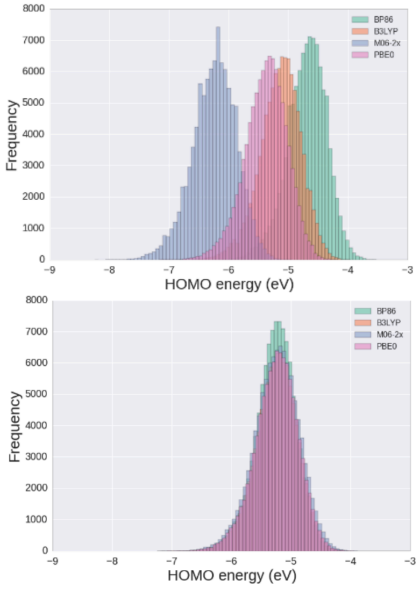

Figure 14. HOMO energies Boltzmann averaged over conformers, as calculated by BP86, B3LYP[18] PBE0[13] and M06-2X[14] with the def2-SVP basis calculated for 100,000 molecules from the Clean Energy Project Database (top) and the values for the same set of molecules after calibration (bottom).

The plots in Figure 14 shows that Bayesian calibration eliminates the functional choice variable compute the HOMO energy. The calibrated energies are much more congruent with experiment, and afford a greater confidence that the calculated property. Gaussian process with a prior based upon relevant experimental observations, is a robust method for relating the results of quantum chemical calculations to experiment. The Bayesian nature of our proposed calibration results in a confidence in each calibration point being returned. This is an invaluable tool, since it can inform the user that the scheme is being used for systems for which it is not designed, or for which the prior is not informative.

Part II

In this final part of the report, we describe results obtained outside of the CEP. The theme is on the relationship between molecular structure and morphological effects and device performance.

Achievement 10: Hydrogen-bonded diketopyrrolopyrrole (DPP) pigments as organic semiconductors.

Eric Glowacki, Halime Coskun, Martin A. Blood-Forsythe, Uwe Monkowius, Lucia Leonat, Marek Grzybowski, Daniel Gryko, Matthew S. White, Alán Aspuru-Guzik, and Niyazi S Sariciftci. *Organic Electronics* **2014**, *15*, 3521–3528.

A key challenge in the search for new high performance organic electronic semiconducting materials, involves understanding the relationship between molecular structure and solid-state packing. The crystal structure has a significant impact on the charge carrier transport. Hydrogen bond mediated crystal engineering has gained attention as a possible route to explore of a chemical space while maintaining some confidence in the likely topology of the crystal environment. Recent examples have shown that H-bond forming

pigments such as perylene bisimides,[11,1215] indigos,[13,1416] and quinacridone[1517] are air-stable and are high-mobility ambipolar transistor materials. We have focused a small library of H-bond forming pigment materials built around the "H-chromophore" motif. We predicted the charge carrier mobilities of a family of three diketopyrrolopyrrole (DPPs): diphenyl-DPP (DPP), di(*p*-chlorophenyl)-DPP (p-Cl DPP), and di(*p*-bromophenyl)-DPP (p-Br DPP). Many properties of DPPs originate from the interplay of intermolecular hydrogen bonding and π-π stacking.

Table 1. Comparison of experimental and theoretical values of hole and electron mobilities of DPP derivatives. Band gaps and mobilities were calculated with density functional theory using the B3LYP[18]-D3BJ/ def2-svp.

Material	Computed HOMO-LUMO Gap (eV)	Measured Optical Band Gap (eV)	Reorganization Energy (meV)	Computed Mobility ($cm^2/V{\cdot}s$)	Measured Mobility ($cm^2/V{\cdot}s$)
DPP	1.6 (1.6)	2.1	λ_e= 184 (241)	m_e = 0.09	m_e = 0.01
			λ_h= 317 (353)	m_h = 0.08	m_h = 0.01
p-Cl-DPP	1.4 (1.5)	2.1	λ_e= 182 (192)	m_e = 0.09	m_e = 0.03
			λ_h= 328 (354)	m_h = 0.17	m_h = 0.006
p-Br-DPP	1.6 (1.6)	2.1	λ_e= 175 (186)	m_e = 0.14	m_e = 0.06
			λ_h= 316 (344)	m_h = 0.05	m_h = 0.02

Using the semi-classical Marcus theory[19] of electron transfer, we predicted the electron and hole carrier mobilities. Table 1 shows that our computed hole and electron mobilities are within an order of magnitude of those measured experimentally.

Achievement 11: Effects of Odd-Even Side Chain Length of Alkyl-Substituted Diphenyl-bithiophenes on First Monolayer Thin Film Packing Structure.

Hylke B. Akkerman, Stefan Mannsfeld, Ananth Kaushik, Eric Verploegen, Luc Burnier, Arjan Zoombelt, Jonathan Saathoff, Sanghyun Hong, Sule Atahan-Evrenk, Xueliang Liu, Alán Aspuru-Guzik, Michael Toney, Paulette Clancy, and Zhenan Bao *J. Am. Chem. Soc.* **2013**, *135*, 11006-11014.

We undertook a collaborative approach to explore the dependence of morphology on the length of the alkyl (solubilizing) side chains for a series of diphenylbithiophenes. The Bao group observed a substantial difference in the molecular tilt angle in a monolayer of 5,5'bis(4-alkylphenyl)-2,2'-bithiophenes (P2TPs), depending on whether the alkyl chain was of odd or even length (Figure 15). We utilized a multiscale computational approach including Molecular Dynamics simulations (TINKER[20] package), semi-empirical (PM3[21]) and density functional theory (B3LYP[18]) calculations to identify the origin of this effect. Figure 15 shows the equilibrated MD structure of P2TP deposited on a self-assembled monolayer (SAM).

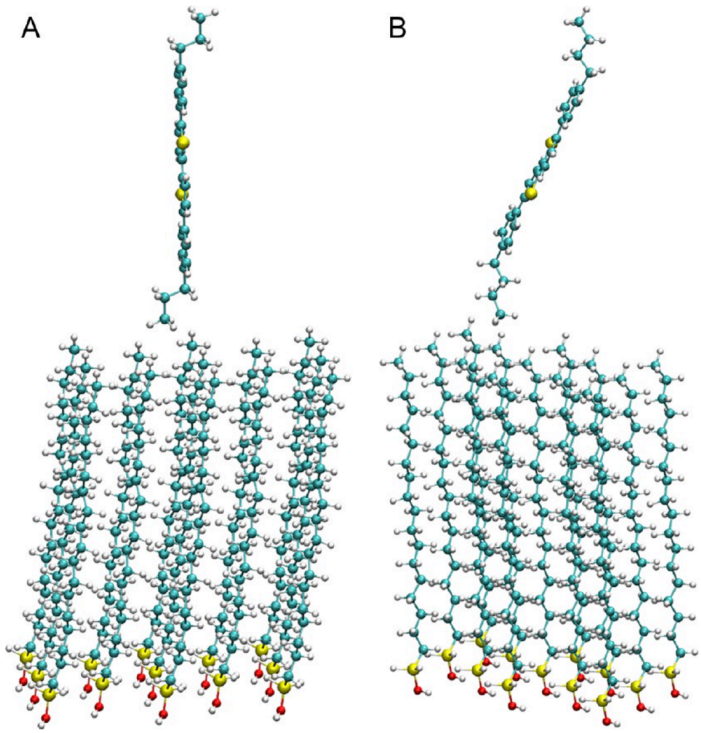

Figure 15. Lowest energy position of P2TP on top of ODTS: (A) C_3– P2TP–C_3 molecules with a tilt angle of 5°; (B) C_4–P2TP–C_4 under a tilt angle of 19°.

Figure 15 shows that P2TP with a propyl sidechain is nearly linear, while P2TP with a butyl sidechain is tilted by 19°. These simulations were extended for side chains of length 3-8; crystal parameters and the title angles are summarized in Table 2.

Table 2. Unit cell geometries and P2TP off-normal title angle by simulations on SAMs.

Side chain length N	*a* (Å)	*b* (Å)	Tilt angle (deg)
3	5.77 ± 0.3	7.90 ± 0.3	3.5 ± 1.5
4	6.12 ± 0.3	8.07 ± 0.3	19 ± 1.5
5	5.85 ± 0.3	7.60 ± 0.3	4.2 ± 1.5
6	5.61 ± 0.3	9.37 ± 0.3	19 ± 1.5
7	5.85 ± 0.3	7.65 ± 0.3	4.5 ± 2
8	5.63 ± 0.3	9.64 ± 0.3	22 ± 0.3

To isolate the origin of the odd-even effect, it was necessary to remove the P2TP-SAM interaction. Therefore, 2D crystals of the P2TP derivatives with alkyl side chains of length 3-8 were simulated. We found that despite the additional freedom of the P2TP molecules, the crystals continued to pack in a herringbone fashion for both odd and even length chains. This suggests an inherent odd-even effect that is enhanced in the presence of a SAM.We then estimated the impact of the odd-even effect on the intrinsic coupling between the molecules in the P2TP layer by computing transfer integrals for the neighboring dimers at

the B3LYP[18]/6-31G(d) level of theory for the obtained packing pattern. Table 3 summarizes the results of these calculations.

Table 3. Transfer Integral Values for Chain Lengths 3–7, Calculated at the B3LYP/6-31G(d,p) Level

N	P (meV)	T1 (meV)	T2 (meV)
3	1.64	4.23	4.20
4	2.32	7.25	7.25
5	1.21	0.62	1.02
6	2.76	4.23	4.34
7	1.23	2.90	2.76

Table 3 shows a clear odd-even effect in all three dimer transitions, most pronounced in the diagonal transfer elements T1 and T2.

This work demonstrates that the tilting of molecules occurs as a composite of inter and intramolecular interactions. The difference is increased by SAMs whose close interactions with P2TP alter lattice parameters and tilting magnitude. The use of molecular dynamics simulations could be useful to study similar problems in the future.

Achievement 12: Understanding Polymorphism in Organic Semiconductor Thin Films Through Nanoconfinement.

Ying Diao, Kristina M. Lenn, Wenya Lee, Martin A. Blood-Forsythe, Jie Xu, Yisha Mao, Yeongin Kim, Julia A. Reinspach, Steve Park, Alán Aspuru-Guzik, Gi Yue, Paulette Clancy, Zhenan Bao, and Stefan C.B. Mansfield. *J. Am. Chem. Soc.* **2014**, *136*, 17046-17057

In line with our general interest to understand the relationship between morphology and device performance, we undertook a highly collaborative approach to understand the known polymorphism of 6,13-bis(triisopropylsilylethynyl)pentacene (TIPS-pentacene). Controlling polymorphism of organic semiconductors in thin films is particularly important because of their ubiquity in organic electronics.[22] Our collaborators first mapped the structural phase space of TIPS-pentacene using *in situ* Grazing Incidence X-ray Diffraction (GIXD) and molecular simulation. A new polymorph of the extensively studied TIPS-pentacene was found, which is metastable at room temperature and featured a different side chain conformation. This polymorph was shown to exhibit very close π-π stacking. We used Molecular Dynamics simulations to predict the energy of the system as a function of the unit cell parameters [a, b, γ] to locate the lowest energy arrangements. We then computed the charge transport properties of these polymorphs using quantum mechanical calculations. Figure 16 shows the three major polymorph families of TIPS-pentacene.

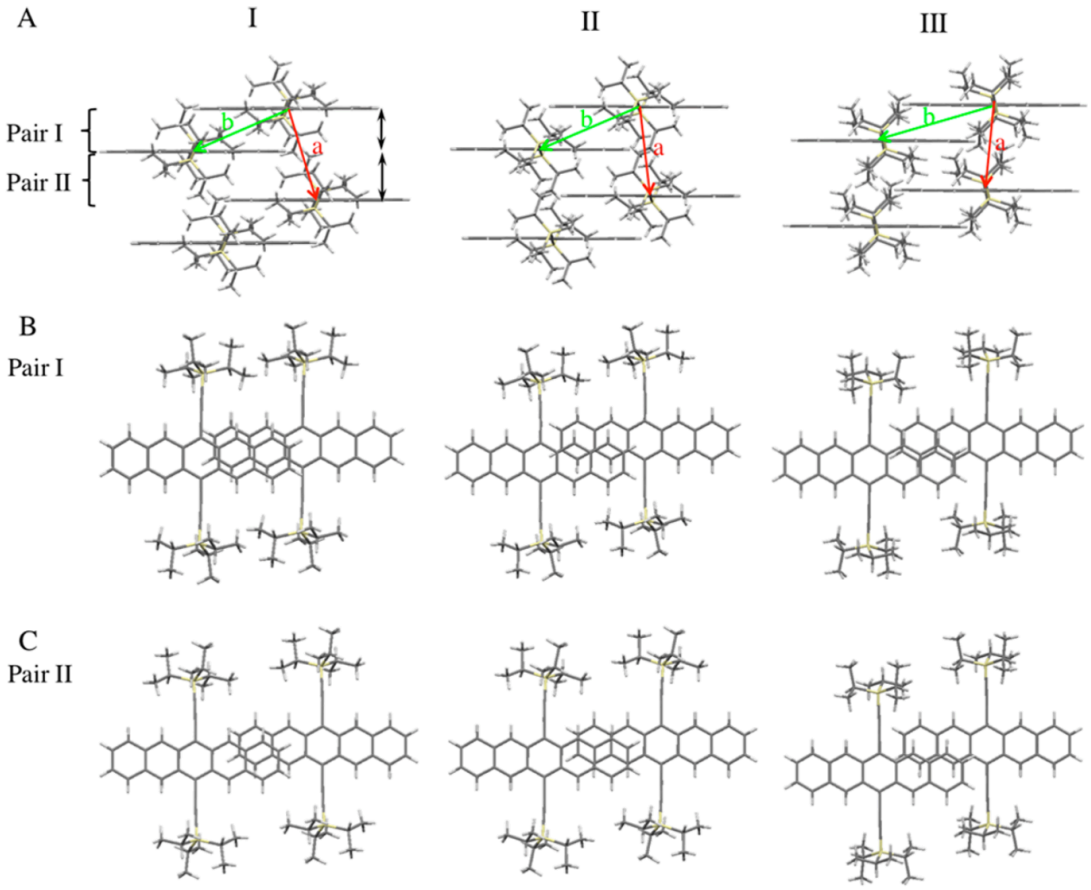

Figure 16. Comparison of the three major polymorphs of TIPS-pentacene in their π–π stacking (A) and molecular offset along the conjugated backbone (B,C) as obtained from the crystallographic refinement calculations.

We modelled the TIPS-pentacene using Avogadro[23] and TINKER using the MM3[24] semiempirical potential to minimize the unit cell containing four TIPS-pentacene molecules. Many more 'metabasins' emerged than the five polymorphs observed experimentally, but our results allowed us to distinguish them into the same three families of polymorphs. This presented the opportunity to study the impact of molecular packing on charge transport properties. The differences in electronic coupling are substantial, likely due to the sensitivity of charge transport to molecular packing.

This study shows that the structural diversity of TIPS-pentacene offers an intriguing example for studying structure-property relationships in organic semiconductors. We find that even with the same packing motif, small changes in molecular packing can have a large impact on the electronic coupling and charge mobilities for these materials. The room temperature equilibrium molecular packing is not necessarily the one with the best charge transport properties. In this case, the highest hole mobility belonged to Form II, which is a metastable form at ambient conditions.

Conclusion

The DOE grant number DE-SC0008733 has allowed for major advances in the intrastructure of the Harvard Clean Energy Project and substantial applications of exciting neural nets and Bayesian approaches have blazed the trail of calibration results for future screening of non-fullerene acceptor molecules. Our studies of structure-morphology relationships have allowed us to computationally explore well-known organic semiconducting materials, where experimental data is available. The insightful computational results are encouraging. In the future, novel materials identified by the CEP can be explored in an ad hoc fashion.

References

Financial Information

Approved budget period 1: 9/1/12-8/31/13

Awarded amount: $321,690

Actual expenses incurred through 8/31/13: $155,897.41

Approved budget period 2: 9/1/13-8/31/14

Awarded amount $328,598

Actual expenses incurred through 3/31/14: $199,150.01

Budget period 3: 9/1/14-8/31/15

Anticipated amount: $341,056

Total project period: 9/1/12-8/31/15

Total Award: $984,344

Publications

1. Edward O. Pyzer-Knapp, Changwon Suh, Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, and Alán Aspuru-Guzik.What is High Throughput Virtual Screening? A Perspective from Organic Materials Discovery. *Annu. Rev. Mat. Res.* 2015, *45*, 195-216.
2. Edward O. Pyzer-Knapp, Kewei Li, and Alan Apuru-Guzik Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery *Adv. Funct. Mater.* 2015, *25*, 6495-6502.
3. Eric Glowacki, Halime Coskun, Martin A. Blood-Forsythe, Uwe Monkowius, Lucia Leonat, Marek Grzybowski, Daniel Gryko, Matthew S. White, Alán Aspuru-Guzik, and Niyazi S Sariciftci. Hydrogen-bonded diketopyrrolopyrrole (DPP) pigments as organic semiconductors. Organic Electronics 15, no. 12 (2014): 3521-3528.
4. Ying Diao, Kristina M. Lenn, Wenya Lee, Martin A. Blood-Forsythe, Jie Xu, Yisha Mao, Yeongin Kim, Julia A. Reinspach, Steve Park, Alán Aspuru-Guzik, Gi Yue, Paulette Clancy, Zhenan Bao, and Stefan C.B. Mansfield. Understanding Polymorphism in Organic Semiconductor Thin Films Through Nanoconfinement. Journal of the American Chemical Society 136, no. 49 (2014): 17046-17057.

5. Johannes Hachmann, Roberto Olivares-Amaya, Adrian Jinich, Anthony L. Appleton, Martin A. Blood-Forsythe, László R. Seress, Carolina Román-Salgado, Kai Trepte, Sule Atahan-Evrenk, Süleyman Er, Supriya Shrestha, Rajib Mondal, Anatoliy Sokolov, Zhenan Bao, and Alán Aspuru-Guzik. Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project. Energy & Environmental Science 7, no. 2 (2014): 698-704.
6. Hylke B. Akkerman, Stefan Mannsfeld, Ananth Kaushik, Eric Verploegen, Luc Burnier, Arjan Zoombelt, Jonathan Saathoff, Sanghyun Hong, Sule Atahan-Evrenk, Xueliang Liu, Alán Aspuru-Guzik, Michael Toney, Paulette Clancy, and Zhenan Bao. Effects of Odd-Even Side Chain Length of Alkyl-Substituted Diphenyl-bithiophenes on First Monolayer Thin Film Packing Structure. Journal of the American Chemical Society 135, no. 30 (2013): 11006-11014

References

[1] Lewis, N. S.; *Solar Energy Workshop*, US Department of Energy, Washington DC, 2005.
[2] NASA. The Earth's Energy Budget http://science-edu.larc.nasa.gov/energy_budget/.

[3] Dennler, G.; Scharber, M. C.; Brabec, C. J.; *Adv. Mater.* **2009**, *21*, 1323-1338.
[4] Materials Genome Initiative http://www.whitehouse.gov/mgi.

[5] Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; Jr, R. A. D.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Voorhis, T. V.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Iii, H. L. W.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Iii, H. F. S.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.

[6] (a) 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.; Shrestha, S.; Mondal, R.; Sokolov, A.; Bao, Z.; Aspuru-Guzik, A. *Energy Environ. Sci.* **2014**, *7*, 698.

[7] World Community Grid http://www.worldcommunitygrid.org/.

[8] Landrum, G. RDKit: Open-source cheminformatics www.rdkit.org.

[9] Chen, H.-Y.; Hou, J.; Zhang, S.; Liang, Y.; Yang, G.; Yang, L.; Wu, Y.; Li, G. *Nat. Photonics* **2009**, *3*, 649-653.
[10] (a) Agarwal, S.; Dugar, D.; Sengupta, S. *J. Chem. Inf. Model* **2010**, *45*, 195. (b) Gelernter, H.; Rose, J. R.; Chen, C. *J. Chem. Inf. Model* **1990**, *30,* 492. (c) Vert, J.-P.; Jacob, L. *Comb. Chem. High Throughput Screening* **2008**, *11*, 677. (d) Kayala, M. A.; Baldi, P. *J. Chem. Inf. Model* **2012**, *52*, 2526.
[11] Recht, B.; Re. C.;Wright, S.; Niu, F. *Advances in Neural Information Processing Systems 24* (Eds: J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, K. Q. Weinberger), Curran Associates, Inc. Red Hook, NJ **2011**, pp. 571-582.
[12] Chilumbi, Suzue, Y.; Apacible, J.; Kalyanaraman, K. *11th USENIX Symp. Oper. Syst. Design Implementation (ODSI 14)*, USENIX Association, Brromfield, CO, **2014**, pp. 571-582.
[13] (a) Perdew, J. P.; Burke, K.; Ernzerhof, M.; *Phys. Rev. Lett.*, **1996**, *77*, 3865-68.
(b) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.*, **1997** *78*, 1396.
[14] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

[15] (a) Schmidt, R.; Ling, M. M.; Oh, J. H.; Winkler, M.; Könemann, M.; Bao, Z.; Würthner, F. *Adv. Mater.* **2007**, *19*, 3692. (b) Chen, Z.; Debije, M. G.; Debaerdemaeker, T.; Osswald, P.; Würthner, F. *ChemPhysChem* **2004**, *5*, 137.
[16] (a) Głowacki, E. D.; Leonat, L.; Voss, G.; Bodea, M.-A.; Bozkurt, Z.; Ramil, A. M.; Irimia-Vladu, M.; Bauer, S.; Sariciftci, N. S. *AIP Adv.* **2011**, *1*, 042132. (b) Kanbur, Y.;

Irimia-Vladu, M.; Głowacki, E. D.; Voss, G.; Baumgartner, M.; Schwabegger, G.; Leonat, L.; Ullah, M.; Sarica, H.; Erten-Ela, S.; Schwödiauer, R.; Sitter, H.; Küçükyavuz, Z.; Bauer, S.; Sariciftci, N. S. *Org. Electron.* **2012**, *13*, 919.
[17] Głowacki, E. D.; Irimia-Vladu, M.; Kaltenbrunner, M.; Gsiorowski, J.; White, M. S.; Monkowius, U.; Romanazzi, G.; Suranna, G. P.; Mastrorilli, P.; Sekitani, T.; Bauer, S.; Someya, T.; Torsi, L.; Sariciftci, N. S. *Adv. Mater.* **2013**, *25*, 1563.
[18] (a) Becke, A. D. J. Chem. Phys. 1993, 98, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. Phys. Rev. B 1988, 37, 785–789. (c) Francl, M. M. J. Chem. Phys. 1982, 77, 3654.

[19] Marcus, R. A.; Sutin, N. *Biochem. Biophys. Acta* **1985**, *811*, 265–322.

[20] TINKER Molecular Modeling. http://dasher.wustl.edu/tinker/, accessed January 18, 2014.

[21] J. J. P. Stewart, *J. Comp. Chem.* **1989**, *10,* 209-20.
[22] (a) Cheng, H,-L.; Mai, Y.-S.; Chou, W.-Y.; Chang, L. –R.; Liang, X-W. *Adv. Funct. Mater.* **2007**, *17*, 3639. (b) Mannsfeld, S. C. B.; Virkar, A.; Reese, C.; Toney, M. F.; Bao, Z. *Adv. Mater.* **2009**, *21*, 2294.
[23]Avogadro-Free Cross-Platform Molecular editor, http://avogadro.openmolecules.net/wiki/Main)_Page, accessed January 18, 2014.

[24] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.