

# The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid

Johannes Hachmann,<sup>\*,†</sup> Roberto Olivares-Amaya,<sup>†</sup> Sule Atahan-Evrenk,<sup>†</sup> Carlos Amador-Bedolla,<sup>†,‡</sup> Roel S. Sánchez-Carrera,<sup>†,‡</sup> Aryeh Gold-Parker,<sup>†</sup> Leslie Vogt,<sup>†</sup> Anna M. Brockway,<sup>§</sup> and Alán Aspuru-Guzik<sup>†,\*</sup>

<sup>†</sup>Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

<sup>‡</sup>Facultad de Química, Universidad Nacional Autónoma de México, México, DF 04510, México

<sup>§</sup>Department of Chemistry, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, United States

**ABSTRACT:** This Perspective introduces the Harvard Clean Energy Project (CEP), a theory-driven search for the next generation of organic solar cell materials. We give a broad overview of its setup and infrastructure, present first results, and outline upcoming developments. CEP has established an automated, high-throughput, *in silico* framework to study potential candidate structures for organic photovoltaics. The current project phase is concerned with the characterization of millions of molecular motifs using first-principles quantum chemistry. The scale of this study requires a correspondingly large computational resource, which is provided by distributed volunteer computing on IBM's World Community Grid. The results are compiled and analyzed in a reference database and will be made available for public use. In addition to finding specific candidates with certain properties, it is the goal of CEP to illuminate and understand the structure–property relations in the domain of organic electronics. Such insights can open the door to a rational and systematic design of future high-performance materials. The computational work in CEP is tightly embedded in a collaboration with experimentalists, who provide valuable input and feedback to the project.



The sun is an abundant source of energy, and its input on earth exceeds the global consumption by 4 orders of magnitude. It is hence an obvious alternative to fossil or nuclear energy supplies and will play an important role in safely and sustainably covering the rising demands of the future.<sup>1–4</sup> The current cost of electricity from commercial silicon-based solar cells is unfortunately still around 10 times higher than that of utility-scale electrical power generation.<sup>5,6</sup> These traditional inorganic photovoltaics come with further shortcomings, such as a complicated and energy-intensive manufacturing process that leads to high production costs. They can also contain rare or hazardous elements, and the devices tend to be heavy, bulky, rigid, and fragile.

Carbon-based solar cells have emerged as one of the interesting alternatives to this conventional technology.<sup>7–9</sup> Organic photovoltaics (OPVs) range from crystalline small molecule approaches<sup>10–12</sup> and certain dye-sensitized Grätzel cells<sup>13</sup> to amorphous polymers (plastics).<sup>14,15</sup> OPVs have great potential in two important areas: they promise simple, low-cost, and high-volume production<sup>16</sup> as well as the prospect of merging the unique flexibility and versatility of plastics with electronic features. OPVs can be processed via roll-to-roll printing,<sup>17,18</sup> and there is active research in sprayable and paintable materials.<sup>19–22</sup> Additionally, OPVs can be semitransparent,<sup>23</sup> variously colored,<sup>24</sup> lightweight,<sup>14</sup> and they can essentially be molded into any shape.<sup>25</sup> These properties make OPVs a promising candidate to achieve the ubiquitous harvesting of solar energy,<sup>26,27</sup> with

building-integrated<sup>28–30</sup> and ultraportable applications<sup>31–33</sup> as primary targets. The Equinox Summit international committee, for example, recently suggested the use of OPVs for the basic electrification of 2.5 billion people in rural areas without access to the power grid.<sup>34</sup>

There are, however, still significant issues to overcome in order to make plastic solar cells a viable technology for the future. The cardinal problems are their relatively low efficiency and limited lifetime:<sup>35,36</sup> the power conversion record has only reached 9.2%,<sup>37</sup> and current materials still degrade when exposed to the environment.<sup>38–41</sup> An increase in efficiency to about 10–15% in combination with lifetimes of over 10 years (for production materials) could push the power generation costs of OPVs below that of other currently available energy sources.<sup>42,43</sup>

In 2008, we started the Harvard Clean Energy Project (CEP)<sup>44</sup> to help find such high-efficiency OPV materials. This perspective article gives a general overview of CEP and provides the context for a series of detailed technical and result-oriented papers to follow. In section I we introduce the motivation and overall setup of the project, followed by a presentation of its different components: section II discusses our OPV candidates library, section III explains the use of cheminformatics descriptors

**Received:** June 27, 2011

**Accepted:** August 11, 2011

**Published:** August 22, 2011

to rapidly assess the potential of these candidates, and section IV is focused on the high-level calculation hierarchy in CEP. Section V is concerned with the calibration of the obtained results. The CEP database is introduced in section VI and the World Community Grid (WCG)—our primary computational resource—in section VII. Throughout sections I–VII we also indicate the next stages and extensions to the current setup. We summarize our discussion in the last two paragraphs of the paper.

**I. The Harvard Clean Energy Project.** The key parameters for the improvement of OPVs are essentially known; however, engineering materials that combine all these features is a hard problem.<sup>45–49</sup> Traditional experimental development is largely based on empirical intuition or experience within a certain family of systems, and only a few examples can be studied per year due to long turnaround times of synthesis and characterization.<sup>50</sup> Theoretical work is usually also restricted to a small set of candidates for which different aspects of the photovoltaic process are modeled.<sup>51–53</sup>

The Clean Energy Project stands out from other computational materials science approaches as it combines conventional modeling with strategies from modern drug discovery:<sup>54–61</sup> CEP features an automated, high-throughput infrastructure for a systematic screening of millions of OPV candidates at a first-principles electronic structure level.<sup>62</sup> It also adopts techniques from cheminformatics<sup>63,64</sup> and heavily relies on data mining.<sup>65,66</sup> Pioneering work on cheminformatics-type approaches and massive electronic structure calculations was, e.g., performed by Rajan et al.<sup>67–69</sup> and Ceder et al.,<sup>70</sup> respectively, in the context of inorganic solids. An *in silico* study combining the scale and level of theory found in CEP is, however, unprecedented.

**The Clean Energy Project stands out from other computational materials science approaches as it combines conventional modeling with strategies from modern drug discovery: CEP features an automated, high-throughput infrastructure for a systematic screening of millions of OPV candidates at a first-principles electronic structure level.**

As the starting point for CEP, we have chosen to investigate the molecular motifs at the heart of OPV materials.<sup>45,71</sup> A suitable motif is a *necessary* condition for a successful OPV development. Only a limited number of structural patterns have been explored so far, while the endless possibilities may well hold the key to overcoming the current material issues. We emphasize that a promising molecular structure is *not* a *sufficient* condition, as condensed matter and device considerations have to be addressed as well. CEP is set up with a calculation hierarchy in which we will successively characterize relevant electronic structure aspects of our OPV candidates. Eventually, we will go

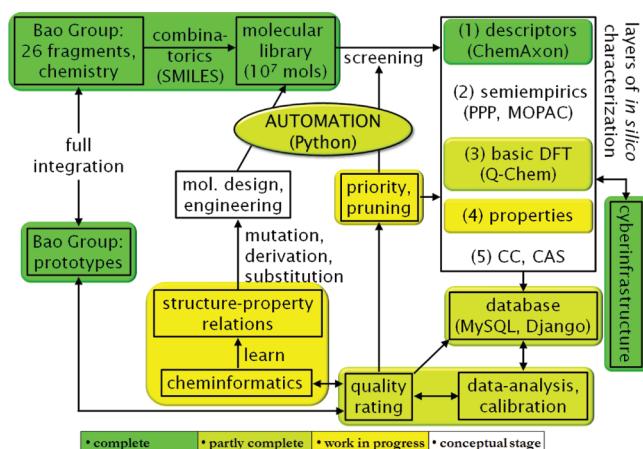


Figure 1. Structure and workflow of CEP.

beyond single molecule, gas-phase studies and consider intermolecular and bulk phase problems.

In addition to the search for specific structures with a desired set of properties, we also try to arrive at a systematic understanding of structure–property relationships.<sup>72–75</sup> Learning about underlying design principles is the key to moving from a screening effort toward an active engineering of novel organic electronics.<sup>76,77</sup>

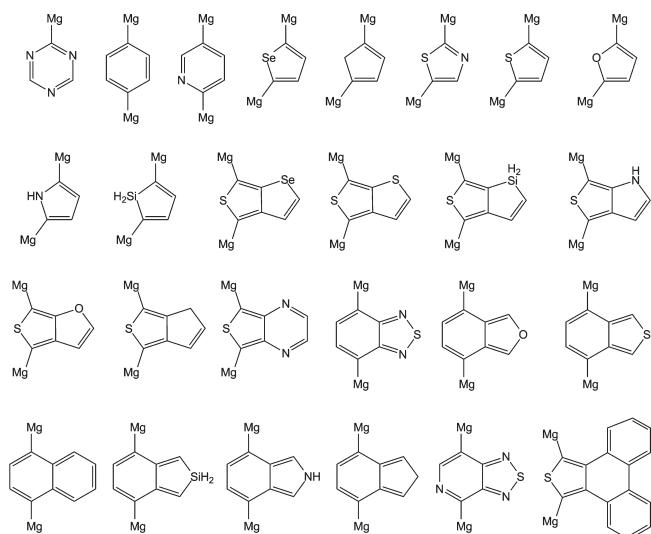
While its centerpiece is the *in silico* study of OPV candidates, we point out that an overarching theme of CEP is also the tight integration of experimental and theoretical work. The project is in part guided by input from experimentalist collaborators (in particular, the Bao Group at Stanford University), and our most promising candidates are subject to in-depth studies in their laboratories.<sup>78</sup> CEP is designed as a community tool and we invite and welcome joint ventures.

Figure 1 summarizes the overall structure and workflow of CEP, and in the following sections we discuss its various components.

**II. Molecular Candidate Libraries.** The molecular structure of essentially all organic electronic materials features a conjugated  $\pi$ -backbone.<sup>79</sup> Modern semiconducting compounds are often composed of linked or fused (hetero)aromatic scaffolds.<sup>80</sup>

We have developed a combinatorial molecule generator to build the primary CEP library. It contains  $\sim 10\,000\,000$  molecular motifs of potential interest ( $\sim 3\,600\,000$  distinct connectivities, each with a set of conformers), which cover small molecule OPVs and oligomer sequences for polymeric materials. They are based on 26 building blocks and bonding rules (see Figure 2), which were chosen following advice from the Bao Group considering promise and feasibility. The fragments were linked and fused up to a length of 4–5 units according to the given rules. The generator is graph-based and employs a SMILES (simplified molecular input line entry specification)<sup>81</sup> string representation of the molecules as well as SMARTS (SMILES arbitrary target specification) in its engine. It is built around Marvin Reactor,<sup>82</sup> and the Corina code<sup>83</sup> provides force-field optimized three-dimensional structures. A detailed description of the library generator is in preparation.

The construction of the primary library was tailored toward OPV donor candidates, but the obtained structures are of interest for organic electronics in general.<sup>84,85</sup> Our generator approach is flexible and can readily produce additional libraries (e.g., geared



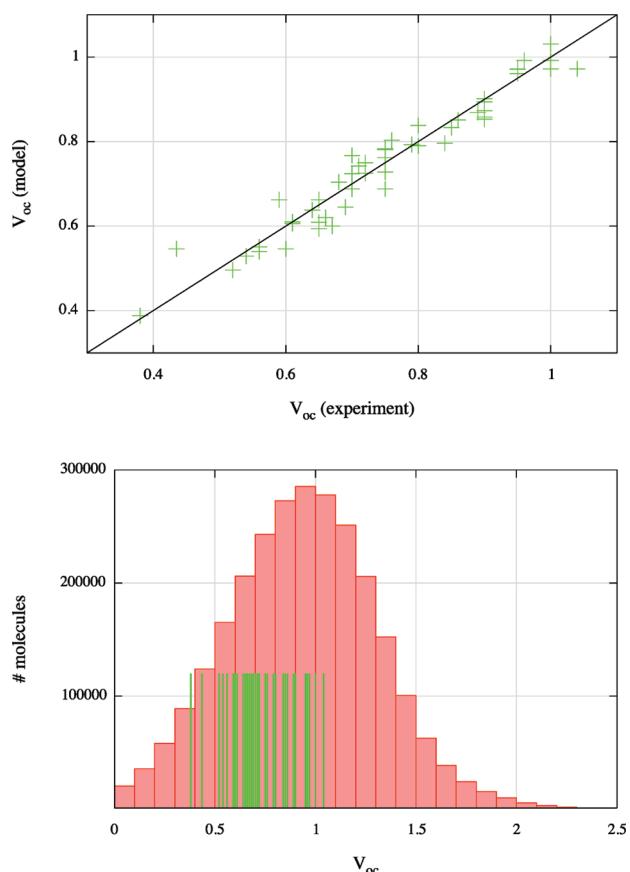
**Figure 2.** The 26 building blocks used for generating the CEP molecular library. The Mg atoms represent chemical handles, i.e., the reactive sites in the generation process. We introduce simple links between two moieties (by means of substituting two Mg for a single C–C bond) as well as the fusion of two rings.

toward acceptor materials or Grätzel cell dyes) by using a different set of fragments and connection rules. Substituents can be incorporated in a similar fashion. Combinatorial libraries allow for an exhaustive and systematic exploration of well-defined chemical spaces, but the number of generated molecules grows exponentially. At a later stage we plan to use a genetic algorithm<sup>80–90</sup> as a complement to the current approach.<sup>62</sup> Its fitness function will be based on the information gathered from the screening of the primary library. Finally, CEP also provides the facility (e.g., to collaborators) to manually add structures to the screening pool.

**III. Cheminformatics Descriptors.** While the main objective of CEP is the first-principles characterization of OPV candidates, we also explore the use of cheminformatics descriptors<sup>91,92</sup> and ideas from machine learning,<sup>93,94</sup> pattern recognition,<sup>95,96</sup> and drug discovery<sup>54,56,64</sup> to rapidly gauge their quality. We have devised descriptor models for parameters such as the short-circuit current density ( $J_{sc}$ ), the open-circuit voltage ( $V_{oc}$ ), and the power conversion efficiency (PCE). Our models are the first step in a successive ladder of approximations for these key quantities associated with photovoltaic performance.

### A suitable molecular motif is a necessary condition for a successful OPV development.

The basic strategy behind this approach is to identify and exploit correlations between certain physicochemical or topological descriptors and the properties of interest. Suitable descriptors are combined into models which are then empirically parameterized using a training set of experimentally well-characterized reference systems. As descriptors are easily computed, we can



**Figure 3.** Top: Linear regression model for  $V_{oc}$  with the training set data. Bottom: Histogram of the projected  $V_{oc}$  values in the molecular library with the position of the training set marked in green (with arbitrary height). See ref 97 for details.

quickly (i.e., within a few days on a single workstation) obtain an initial assessment and preliminary ranking of the entire molecular library.

In Figure 3 we present the linear regression model for  $V_{oc}$  along with the histogram of the calculated values for the molecular library. We note that our model for  $V_{oc}$  shows a very good correlation. This can be rationalized considering the principal dependence of  $V_{oc}$  on intramolecular properties, which are apparently well reflected in the employed molecular descriptors. The  $J_{sc}$  model behaves similarly well although  $J_{sc}$  is also linked to bulk effects. For the fill factor—a quantity primarily determined by morphology and device characteristics—we could, in contrast, only obtain poor models.

The early stages of this work<sup>97</sup> utilized descriptors from the Marvin code by ChemAxon,<sup>82</sup> and for the modeling we employed the R statistics package.<sup>98</sup> Recently, we started using the more comprehensive descriptor set from Dragon<sup>99</sup> and the specialized modeling code StarDrop.<sup>100</sup> A focus of our current work is to utilize the quantum chemical results discussed in the following section as a descriptor basis in our models.

As in biomedical applications of cheminformatics, we do not expect quantitative results. Rather, this technique can yield valuable trends, which we use to prioritize and prune the high-level screening and to uncover molecular motifs of particular interest. In ref 97 we give an introduction to this approach with a detailed discussion of the systematic construction and optimization of descriptor models.

**IV. First-Principles Screening Hierarchy.** Electronic structure theory offers a way to probe the properties of OPV materials and the photophysical processes in organic solar cells.<sup>101,102</sup> The complexity of these problems, however, poses severe methodological challenges. Multiscale simulations have improved considerably in recent years,<sup>51,103</sup> but in practice we still commonly choose to model, approximate, or deduce the different aspects of the problem separately. CEP adopts such a divide-and-conquer approach and combines it with a calculation hierarchy to screen for promising material candidates. This multilevel setup is designed to successively address the relevant issues in OPVs and provide results at an increasing level of theory. At each stage, the candidates are rated with respect to the investigated parameters. The scoring is freely customizable to reflect different research priorities. The most promising candidates and related structures from the library receive priority in the CEP hierarchy, and their characterization is expedited. On the other hand, if a candidate is unfit with respect to a tested criterium and hence unlikely to be successful overall, it has lower priority and may be characterized in less detail.

The early CEP stages concentrate on various molecular properties of our material candidates and are hence most useful to assess macroscopic quantities that primarily depend on them, such as the  $V_{oc}$ . The later stages will shift the focus to intermolecular and condensed phase characteristics. The latter are central to, e.g., exciton and charge transport processes,<sup>46,51,52,104–106</sup> for which molecular properties alone are clearly of limited value. Bulk structure considerations impact quantities such as the external quantum efficiency and thus  $J_{sc}$  as well as the overall PCE. Since the cost and complexity of such studies increase significantly, they can only be performed for a subset of the highest rated candidates.

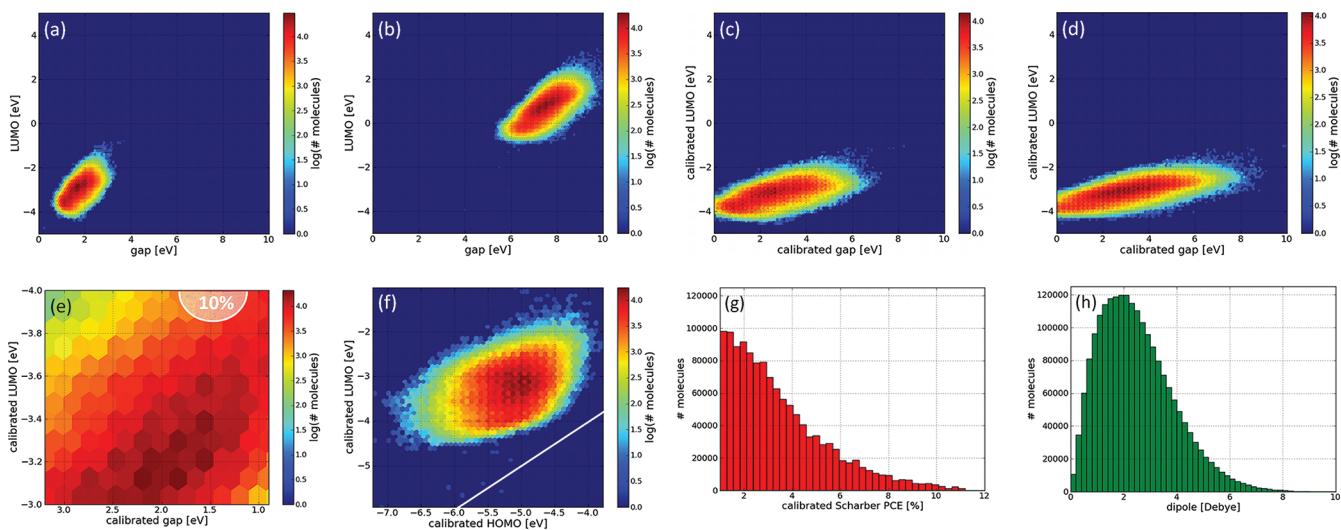
In the first CEP phase, we perform a set of density functional theory (DFT) calculations<sup>108,109</sup> employing the BP86,<sup>110,111</sup> B3LYP,<sup>110,112,113</sup> PBE0,<sup>114–119</sup> BH&HLYP,<sup>110,112,120</sup> and M06-2X<sup>121,122</sup> functionals as well as the Hartree–Fock (HF) theory in combination with the single- $\zeta$  STO-6G,<sup>123,124</sup> double- $\zeta$  def2-SVP, and triple- $\zeta$  def2-TZVP<sup>125</sup> basis sets. We test and compare both restricted and spin-polarized settings. Our selection of functionals covers both generalized gradient approximation (GGA) and hybrid designs with a progression in the amount of exact exchange<sup>113</sup> (BP86 and HF can be seen as the limiting cases). The latter has a systematic influence on orbital localization and thus eigenvalues.<sup>126</sup> The GGA BP86 is a cost-effective way to obtain good geometries, B3LYP is arguably the most widely used functional in molecular quantum mechanics, and PBE0 has shown favorable performance in a variety of areas, as has the highly parametrized M06-2X. (We will further test meta-GGAs such as TPSS<sup>127</sup> and double-hybrids such as B2PLYP<sup>128</sup> when they become available for CEP.) This range of model chemistries was chosen to put our analysis on a broader footing, but also to assess the performance of the different theoretical methods.<sup>109</sup> In total, each molecule is characterized by a BP86/def2-SVP geometry optimization and 14 single-point ground state calculations. We obtain geometries, total energies (including their decomposition into different contributions), molecular orbitals (MOs) and their energy eigenvalues, electron and spin densities, electrostatic potentials, multipole moments, Mulliken<sup>129,130</sup> and natural populations,<sup>131,132</sup> as well as natural atomic, localized molecular, and bond orbital analyses<sup>133–135</sup> for the different model chemistries.<sup>136</sup>

These basic electronic quantities can be used as a first approximation to the following points. The MO energies and

their differences can be related to ionization potentials, electron affinities, gaps, and partial density of states (we note that the application of Koopmans' theorem is problematic in DFT).<sup>126,137–142</sup> The electronic levels have to be tuned for an efficient light absorption, for the necessary interplay between donor, acceptor, and lead materials, as well as for atmospheric stability.<sup>45</sup> The delocalization of the frontier orbitals can be associated with (intramolecular) exciton and charge carrier mobility.<sup>143–147</sup> Their spatial overlap indicates the transition character and probability of the corresponding excitation.<sup>148–150</sup> Excess spin densities reflect the inadequacy of simple closed-shell solutions and can also point to a complex and potentially unstable electronic situation.<sup>151</sup> Charge maps can identify chemically unstable sites in these highly unsaturated molecules as well as patterns that may have an impact on their packing in the condensed phase. The molecular multipole moments can correspondingly be correlated to the organization in the bulk structure and also to transport properties.<sup>152–155</sup> The wave function analysis techniques are of interest for the study of structure–property relations. The obtained data can furthermore be utilized as quantum chemical descriptors for the models described in section III. The results of a consecutive oligomer series can be used to extrapolate to the polymer limit.<sup>156–159</sup>

One example for how this basic information can be employed is the model developed by Scharber et al. in ref 107. The authors assume (based on observations on actual OPV materials) that condensed matter aspects in such systems can be tuned to a certain viable level (i.e., a fill factor and external quantum efficiency of each 65% can be achieved). If that is the case, their PCE can be approximated using only gap and lowest unoccupied molecular orbital (LUMO) energy values. This model can readily be applied to the current CEP data as shown in Figure 4. Our preliminary analysis reveals that only about 0.3% of the screened compounds (i.e., presently resulting in ~3000–5000 distinct structures, depending on the model chemistry) have the necessary energetic levels to realize OPVs with 10% or higher efficiency. This underscores the importance of carefully selecting the compounds to be synthesized and tested, and at the same time the value of the fast theoretical characterization and extensive search that CEP can provide toward this task. An unaided search has only a small chance of success, while CEP finds several thousand suitable structures. Figure 4 also displays the dynamic gap range and dipole moment distribution of the screened candidates as examples for the wide range of electronic properties found in the CEP library. This versatility will be vital to a successful discovery of materials with specifically adjusted features (e.g., for different acceptors or tandem devices<sup>160</sup>).

We again emphasize that this first phase of CEP only addresses a subset of the important material issues<sup>161</sup> and has the inherent accuracy of the employed model chemistries and calibrations, if taken at face value. There are, however, four factors that add considerable value to these calculations: (i) we correlate the computed results to actual experimental quantities to provide insights into their relationship; (ii) we use the electronic structure data as a source for new cheminformatics descriptors, which will put our modeling efforts on a more physical foundation; (iii) the analysis of the aggregated data from millions of molecules in combination with structural similarity measures can reveal guiding trends, even if the absolute result for an individual candidate might be inaccurate due to a particular limitation of its electronic structure; (iv) by employing a variety of different model chemistries, we compensate for the chance of such a failure in any particular



**Figure 4.** (a–e) Distribution of CEP hits in the gap–LUMO plane suggested by Scharber et al.<sup>107</sup> Note that the density plots are on a logarithmic scale, and the red entries correspond to  $\mathcal{O}(10\,000)$  compounds and  $\mathcal{O}(150\,000)$  individual electronic structure calculations. A total of 2 600 000 compounds are represented in these plots: (a,b) raw data from BP86/def2-SVP and BP86/def2-SVP//HF/def2-SVP (i.e., calculations which incorporate 0% vs 100% exact exchange), respectively; (c,d) corresponding data after preliminary calibration; (e) OPV relevant parameter space with the 10% PCE region (with respect to a PCBM acceptor). About 0.3% of the screened compounds fall in this high-efficiency region. (f) Dynamic gap range, i.e., the range of available LUMO energies given a particular HOMO and vice versa; (g) PCE histogram according to the Scharber model; (h) dipole moment histogram. Panels e–h are all at the BP86/def2-SVP//PBE0/def2-TZVP level of theory (calibrated).

method. We do not rely on any single result but use a composite scoring with many contributions.

The calculations allow for the elimination of unfit candidates and provide a first set of predictions for promising structural trends. For these we can carry out the subsequent levels of increasingly more sophisticated calculations, which are planned as follows: we will calculate (i) vibrational spectra and partition functions to gauge phonon-scattering and trapping in vibrational modes;<sup>162,163</sup> (ii) anionic/cationic states for improved electron affinity/ionization potential values (a Dyson orbital analysis<sup>164</sup> can improve our insights into the charge transfer processes);<sup>165</sup> (iii) optimized geometries in the ionic states to deduce the reorganization during charge migration;<sup>163</sup> (iv) triplet states and gaps to indicate potential for singlet fission processes;<sup>166</sup> (v) linear response properties to assess the charge mobility;<sup>167,168</sup> and (vi) excited states employing the maximum overlap method<sup>169</sup> and/or (range-separated) time-dependent DFT<sup>170,171</sup> with an electron attachment/detachment density<sup>172</sup> and natural transition orbital analysis<sup>173</sup> for a more sound description of the absorption process. Each higher level result can be used to assess the interpretations at the lower levels. Further studies could involve the calculation of transfer integrals between oligomer pairs,<sup>163</sup> packing and interactions in the bulk phase, as well as the use of high-level wave function theory for more reliable results in complicated bonding situations. We also want to consider the opposite approach, i.e., how well simple semiempirical and model Hamiltonian calculations (which are popular in other communities) perform compared to first-principles DFT.<sup>101,174</sup> Possible stages of the CEP hierarchy were recently vetted in a successful proof-of-principle study: we predicted exceptional charge mobility in a novel organic semiconductor, and this prediction could be confirmed experimentally by the Bao Group with values of up to  $16\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ .<sup>178</sup> We also used quantum chemical calculations in an earlier study to explain the observed mobility values in another system.<sup>175</sup>

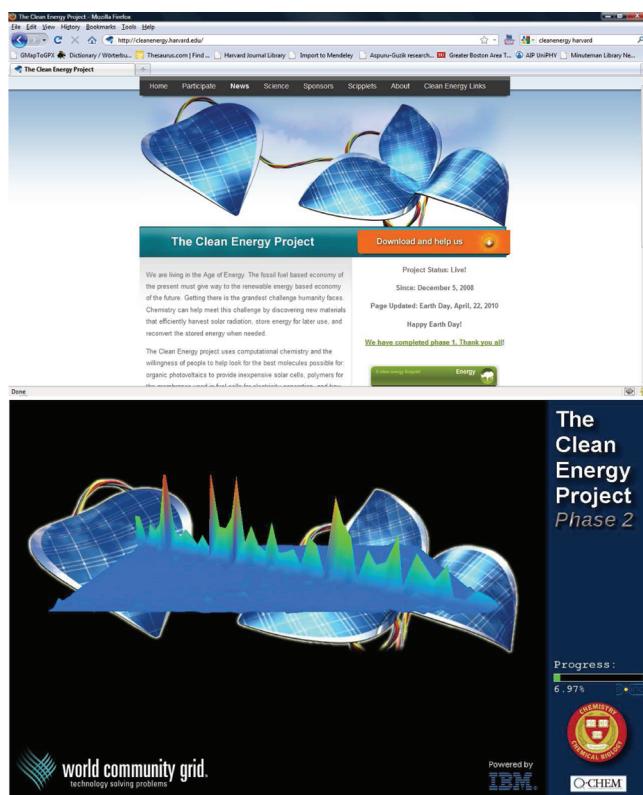
**V. Empirical Result Calibration.** To bridge the gap between theory and experiment, we have introduced an empirical calibration

of the computational results. Such a calibration is a pragmatic way to approximately account and correct for differences in experimental and theoretical property definitions, as well as *in vacuo* versus bulk, and oligomer versus polymer results.

We have established the organic electronics 2011 (OE11) training set of experimentally well-characterized organic electronic materials for the calibration and aligned the theoretical findings with the corresponding data from experiment. The current calibration is largely based on data from bulk-heterojunction setups with PCBM as acceptor and introduces a corresponding bias. A different focus can be chosen provided appropriate reference data. The use of training sets is a common technique in other areas of quantum chemistry.<sup>121,176</sup> The details of this work will be presented elsewhere. A preliminary calibration of the current CEP results was used in the analysis shown in Figure 4, and panels a–d in particular demonstrate the success of this approach.

**VI. Database and Data Mining.** The results of all calculations are used to build up a reference database: the Clean Energy Project Database (CEPDB). This data collection is comparable to the more general but much smaller NIST Computational Chemistry Comparison and Benchmark Database (CCCBDB).<sup>177</sup> As mentioned before, the information accumulated in CEPDB is not only relevant to OPVs but to organic electronics in general. It is also designed to provide benchmarks for the performance of various theoretical methods in this family of systems as well as a parameter repository for other calculations (e.g., for model Hamiltonians<sup>178–180</sup> or custom force fields<sup>181</sup>). It will be available to the public by 2012.

The primary purpose of CEPDB is to store and provide access to the CEP data. Candidates with a certain set of desired parameters can readily be identified. CEPDB also serves as the hub for all data mining, analysis, and scoring operations to facilitate the study of global trends, correlations, and OPV design rules. Finally, CEPDB is also responsible for bookkeeping, archiving the raw data (including the binary MO eigenvectors for subsequent



**Figure 5.** Top: CEP project homepage and hub for project participants. Bottom: The CEP “screensaver” as it is displayed on volunteer hosts.

calculations), keeping track of the project progress, and prioritizing the study, which are clearly important tasks considering the scope of the project and the volume of data.

**VII. The World Community Grid.** The massive demand in computing time for CEP is largely provided by the WCG,<sup>182,183</sup> a distributed volunteer computing platform for philanthropic research organized by IBM. Presently, ~560 000 users have signed up ~1 800 000 computers to the various WCG projects. Participants can donate computing time by running the supported science applications on their personal computers, either on low priority in the background or in screensaver mode during idle times (see Figure 5). In CEP, we use a custom version of the Q-Chem 3.2 program package,<sup>184</sup> which was ported to the Berkeley Open Infrastructure for Network Computing (BOINC)<sup>185</sup> environment. From the user perspective, the participation in a WCG project is fully automated and usually does not require any input or maintenance beyond the initial setup.

The WCG is a powerful resource and provides us with the means for our high-throughput investigation. It is, however, also unusual due to the nonspecialized hardware and host demands. These limitations have to be taken into consideration in the design of suitable tasks. In addition to the WCG, CEP utilizes the Harvard FAS Odyssey cluster as well as accounts at NERSC<sup>186</sup> and TeraGrid<sup>187</sup> for problems that are outside the scope of volunteer grid computing (e.g., due to their size and computational complexity). CEP currently characterizes about 20 000 oligomer sequences per day, and so far we have studied ~2 600 000 structures in ~36 000 000 calculations, utilizing ~4500 years of CPU time. Close to 100% of these calculations were performed on the grid, but the use of cluster resources will become more important in the more demanding stages of the

project. Their contribution in terms of volume, however, will remain very limited.

One important aspect in a computational study at this scale is that all processes have to be automated to keep it feasible, efficient, and reduce human error. We created the necessary facilities using the Python scripting language.<sup>188,189</sup>

## CEP applies a modern cyberinfrastructure paradigm to computational materials science and, in particular, to renewable energy research.

This initial presentation of the Harvard Clean Energy Project has outlined its overall architecture, the machinery we put in place, and upcoming extensions. We pointed to the first results, and more detailed reports on the various aspects of the project—tied together by this Perspective—will be given in subsequent publications.

CEP applies a modern cyberinfrastructure paradigm to computational materials science and, in particular, to renewable energy research. Engineering successful OPV materials is a complex challenge as a number of optical and electronic requirements have to be met. We showed that CEP with its large-scale screening is well equipped for a knowledge-based search of systems with suitable features. It provides a unique access to data for a wide array of potential compounds with diverse electronic structures, and it is ideally suited to identify highly promising donor or acceptor candidates in the infinite space of organic electronics. Our work can thus complement and accelerate traditional research approaches, and it can help develop an understanding of structure–property relationships to facilitate the rational design of new materials. We hope that joint efforts with experimental collaborators can contribute to overcoming the current limitations of OPV materials in order to provide a clean source of electricity that can compete with conventional power production.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Electronic address: jh@chemistry.harvard.edu (J.H.); aspuru@chemistry.harvard.edu (A.A.-G.).

### Present Addresses

<sup>†</sup>Research and Technology Center North America, Robert Bosch LLC, Cambridge, MA 02142.

## ■ BIOGRAPHIES

**Johannes Hachmann** obtained his Dipl.-Chem. degree (2004) after undergraduate studies at the University of Jena, Germany, and the University of Cambridge, U.K. He received an M.Sc. (2007) and Ph.D. (2010) from Cornell University under the supervision of Prof. G.K.-L. Chan working on density matrix renormalization group theory and computational transition metal chemistry. In 2009 he joined the Aspuru-Guzik Group and the Clean Energy Project at Harvard University.

**Roberto Olivares-Amaya** is a Ph.D. Chemical Physics student in the Aspuru-Guzik Group at Harvard University and was a 2010–2011 Giorgio Ruffolo Sustainability Science Doctoral Fellow. He obtained his B.Sc. degree (2007) from the National Autonomous University of Mexico (UNAM). His research interests include the study of structure and response of OPV materials by means of first-principles electronic structure and machine learning methods.

**Sule Atahan-Evrdenk** is a postdoctoral research fellow in the Aspuru-Guzik Group at Harvard University. She received a B.S. degree in chemistry from Bilkent University, Turkey, and a Ph.D. in physical chemistry from the University of Maryland College Park. Her research interests include the theoretical characterization of organic electronic materials by quantum chemistry and molecular dynamics methods.

**Carlos Amador-Bedolla** is a professor at the Facultad de Química of the National Autonomous University of Mexico in Mexico City. He obtained his doctorate in chemistry at UNAM (1989) and has worked with several leading groups at Case Western Reserve University, UC Berkeley, and Harvard University. His current research focuses on renewable energy materials (see [www.quimica.unam.mx/ficha\\_investigador.php?ID=77&tipo=2](http://www.quimica.unam.mx/ficha_investigador.php?ID=77&tipo=2)).

**Roel S. Sánchez-Carrera** is currently a research staff at the Robert Bosch Research and Technology Center North America. He obtained his B.S. in chemistry from the Monterrey Institute of Technology and Higher Education, México, and a Ph.D. in physical chemistry from the Georgia Institute of Technology under the supervision of Prof. J.-L. Brédas. He then moved to Harvard University, where he worked in the group of Prof. Aspuru-Guzik as a Mary-Fieser Postdoctoral Fellow.

**Aryeh Gold-Parker** is an undergraduate student at Harvard University and will complete his B.A. degree in chemistry and physics in 2012. His research interests are in physical chemistry, specifically in renewable energy materials. He joined the Aspuru-Guzik Group in 2009 to work on the Clean Energy Project.

**Leslie Vogt** completed a B.S. in physical chemistry at Juniata College in 2005 and spent 2006–2007 working with Prof. R. Zimmermann as a Fulbright Fellow in Germany. She is completing her Ph.D. at Harvard University in the Aspuru-Guzik Group studying the electronic properties of organic materials as an NSF Graduate Research Fellow.

**Anna M. Brockway** is a chemistry senior at Haverford College where she works with Prof. J. Schrier in the field of computational theoretical chemistry. She joined the Clean Energy Project in 2010 during her time as an REU summer student at Harvard University.

**Alán Aspuru-Guzik** holds a Ph.D. in physical chemistry from UC Berkeley, and is currently an associate professor of Chemistry and Chemical Biology at Harvard University. His research lies at the intersection of quantum information/computation and theoretical chemistry. He is interested in energy transfer dynamics and renewable energy materials. (See <http://aspuru.chem.harvard.edu/> and the project webpage <http://cleanenergy.harvard.edu/>.)

## ACKNOWLEDGMENT

We wish to thank IBM for organizing the World Community Grid and the WCG members for their computing time donations. Additional information about the project as well as links to join the WCG and CEP can be found in refs 44 and 182. We are

grateful to our experimental collaborators Zhenan Bao, Rajib Mondal, and Anatoliy N. Sokolov at Stanford University; for support from the Q-Chem development team by Yihan Shao, Alex J. Sodt, and Zhengting Gan; for support from the IBM WCG team by Jonathan D. Armstrong, Kevin N. Reed, Keith J. Uplinger, Al Seippel, Viktors Berstis, and Bill Bovermann; for support from Harvard FAS Research Computing by Suvendra Dutta, Christopher Walker, Gerald I. Lotto, and James Cuff; and to Aidan Daly, Erica Lin, and Lauren A. Kaye at Harvard. The CEP is supported by the National Science Foundation through Grant Nos. DMR-0820484 and DMR-0934480 as well as the Global Climate and Energy Project (Stanford Grant No. 25591130-45282-A. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of Stanford University, the Sponsors of the Global Climate and Energy Project, or others involved with the Global Climate and Energy Project). R.O.-A. was supported by a Giorgio Ruffolo Fellowship in the Sustainability Science Program at Harvard University's Center for International Development, C.A.-B. acknowledges financial support from CONACyT México and computer access to DGTIC UNAM, A.G.-P. acknowledges funding by the Harvard College Research Program and the Harvard University Center for the Environment Summer Research Fund, R.S.-C. thanks the Mary Fieser Postdoctoral Fellowship at Harvard University, L.V. was supported by an NSF Graduate Research Fellowship, and A.M.B. was supported by the Harvard REU program. We acknowledge software support from Q-Chem Inc., Optibrium Ltd., Dotmatics Ltd., Molecular Networks GmbH, and ChemAxon Ltd. Additional computing time was provided on the Odyssey cluster supported by the FAS Research Computing Group, NERSC, and the National Science Foundation TeraGrid.

## REFERENCES

- (1) Wöhrle, D.; Meissner, D. *Organic Solar Cells. Adv. Mater.* **1991**, 3, 129–138.
- (2) Conti, J. J.; Holtberg, P. D.; Beamon, J. A.; Schaal, A. M.; Ayoub, J. C.; Turnure, J. T. *Annual Energy Outlook 2011*; U.S. Energy Information Administration: Washington, DC, 2011.
- (3) Zweibel, K.; Mason, J.; Fthenakis, V. *A Solar Grand Plan. Sci. Am.* **2008**, 298, 64–73.
- (4) Jacobson, M. Z.; Delucchi, M. A. *A Plan to Power 100% of the Planet with Renewables. Sci. Am.* **2009**, 301, 58–65.
- (5) Lewis, N. S. *Toward Cost-Effective Solar Energy Use. Science* **2007**, 315, 798–801.
- (6) Boxwell, M. *Solar Electricity Handbook – 2011 Edition*, 4th ed.; Greenstream Publishing: Ryton on Dunsmore, U.K., 2011.
- (7) Brabec, C. J.; Dyakonov, V.; Parisi, J.; Sariciftci, N. S. *Organic Photovoltaics: Concepts and Realization*; Springer: Berlin, 2010.
- (8) Brabec, C. J.; Dyakonov, V.; Scherf, U. *Organic Photovoltaics: Materials, Device Physics, and Manufacturing Technologies*; Wiley-VCH: Weinheim, Germany, 2008.
- (9) Sun, S.-S.; Sariciftci, N. S. *Organic Photovoltaics: Mechanisms, Materials, and Devices*; CRC Press: Boca Raton, FL, 2005.
- (10) Tang, C. W. *Two-Layer Organic Photovoltaic Cell. Appl. Phys. Lett.* **1986**, 48, 183–185.
- (11) Beaumont, N.; Hancox, I.; Sullivan, P.; Hatton, R. A.; Jones, T. S. *Increased Efficiency in Small Molecule Organic Photovoltaic Cells through Electrode Modification with Self-Assembled Monolayers. Energy Environ. Sci.* **2011**, 4, 1708–1711.
- (12) Riede, M.; Mueller, T.; Tress, W.; Schueppel, R.; Leo, K. *Small-Molecule Solar Cells – Status and Perspectives. Nanotechnology* **2008**, 19, 424001.

- (13) Kalyanasundaram, K. *Dye-Sensitized Solar Cells*; CRC Press: Boca Raton, FL, 2010.
- (14) Krebs, F. C. *Polymer Photovoltaics: A Practical Approach*; SPIE Publications: Bellingham, WA, 2008.
- (15) Pagliaro, M.; Palmisano, G.; Ciriminna, R. *Flexible Solar Cells*; Wiley-VCH: Weinheim, Germany, 2008.
- (16) Green, M. A. Third Generation Photovoltaics: Solar Cells for 2020 and Beyond. *Physica E* **2002**, *14*, 65–70.
- (17) Galagan, Y.; de Vries, I. G.; Langen, A. P.; Andriessen, R.; Verhees, W. J.; Veenstra, S. C.; Kroon, J. M. Technology Development for Roll-to-Roll Production of Organic Photovoltaics. *Chem. Eng. Process* **2011**, *50*, 454–461.
- (18) Krebs, F. C.; Tromholt, T.; Jørgensen, M. Upscaling of Polymer Solar Cell Fabrication Using Full Roll-to-Roll Processing. *Nanoscale* **2010**, *2*, 873–886.
- (19) Vak, D.; Kim, S.; Jo, J.; Oh, S.; Na, S.; Kim, J.; Kim, D. Fabrication of Organic Bulk Heterojunction Solar Cells by a Spray Deposition Method for Low-Cost Power Generation. *Appl. Phys. Lett.* **2007**, *91*, 081102.
- (20) Green, R.; Morpha, A.; Ferguson, A. J.; Kopidakis, N.; Rumbles, G.; Shaheen, S. E. Performance of Bulk Heterojunction Photovoltaic Devices Prepared by Airbrush Spray Deposition. *Appl. Phys. Lett.* **2008**, *92*, 33301.
- (21) Steirer, K. X.; Reese, M. O.; Rupert, B. L.; Kopidakis, N.; Olson, D. C.; Collins, R. T.; Ginley, D. S. Ultrasonic Spray Deposition for Production of Organic Solar Cells. *Sol. Energy Mater. Sol. Cells* **2009**, *93*, 447–453.
- (22) O'Neill, M.; Kelly, S. M. Ordered Materials for Organic Electronics and Photonics. *Adv. Mater.* **2011**, *23*, 566–584.
- (23) Bailey-Salzman, R. F.; Rand, B. P.; Forrest, S. R. Semitransparent Organic Photovoltaic Cells. *Appl. Phys. Lett.* **2006**, *88*, 233502.
- (24) Colsmann, A.; Puetz, A.; Bauer, A.; Hanisch, J.; Ahlsweide, E.; Lemmer, U. Efficient Semi-Transparent Organic Solar Cells with Good Transparency Color Perception and Rendering Properties. *Adv. Energy Mater.* **2011**, *1*, 599–603.
- (25) Pagliaro, M.; Ciriminna, R.; Palmisano, G. Flexible Solar Cells. *ChemSusChem* **2008**, *1*, 880–891.
- (26) Heeger, A. In *Global Sustainability - A Nobel Cause*; Schellnhuber, H. J., Molina, M., Stern, N., Huber, V., Kadner, S., Eds.; Cambridge University Press: Cambridge, U.K., 2010.
- (27) Forrest, S. R. The Path to Ubiquitous and Low-Cost Organic Electronic Appliances on Plastic. *Nature* **2004**, *428*, 911–918.
- (28) Marsh, G. BIPV: Innovation Puts Spotlight on Solar. *Renewable Energy Focus* **2008**, *9*, 62–67.
- (29) Henemann, A. BIPV: Built-In Solar Energy. *Renewable Energy Focus* **2008**, *9* (14), 16–19.
- (30) Tipnis, R.; Bernkopf, J.; Jia, S.; Krieg, J.; Li, S.; Storch, M.; Laird, D. Large-Area Organic Photovoltaic Module – Fabrication and Performance. *Sol. Energy Mater. Sol. Cells* **2009**, *93*, 442–446.
- (31) Schubert, M. B.; Werner, J. H. Flexible Solar Cells for Clothing. *Mater. Today* **2006**, *9*, 42–50.
- (32) Brabec, C. J. Organic Photovoltaics: Technology and Market. *Sol. Energy Mater. Sol. Cells* **2004**, *83*, 273–292.
- (33) Bedeloglu, A.; Demir, A.; Bozkurt, Y.; Sariciftci, N. S. A Flexible Textile Structure Based on Polymeric Photovoltaics Using Transparent Cathode. *Synth. Met.* **2009**, *159*, 2043–2048.
- (34) [http://wgsi.org/files/EquinoxCommunique\\_June9\\_2011.pdf](http://wgsi.org/files/EquinoxCommunique_June9_2011.pdf)
- (35) Shrotriya, V. Organic Photovoltaics: Polymer Power. *Nat. Photonics* **2009**, *3*, 447–449.
- (36) Peters, C. H.; Sachs-Quitana, I. T.; Kastrop, J. P.; Beaupré, S.; Leclerc, M.; McGehee, M. D. High Efficiency Polymer Solar Cells with Long Operating Lifetimes. *Adv. Energy Mater.* **2011**, *1*, 491–494.
- (37) Service, R. F. Outlook Brightens for Plastic Solar Cells. *Science* **2011**, *332*, 293.
- (38) Rivaton, A.; Chambon, S.; Manceau, M.; Gardette, J.-L.; Lemaitre, N.; Guillerez, S. Light-Induced Degradation of the Active Layer of Polymer-Based Solar Cells. *Polym. Degrad. Stab.* **2010**, *95*, 278–284.
- (39) Manceau, M.; Rivaton, A.; Gardette, J.-L.; Guillerez, S.; Lemaitre, N. Light-Induced Degradation of the P3HT-Based Solar Cells Active Layer. *Sol. Energy Mater. Sol. Cells* **2011**, *95*, 1315–1325.
- (40) Bertho, S.; Haeldermans, I.; Swinnen, A.; Moens, W.; Martens, T.; Lutzen, L.; Vanderzande, D.; Manca, J.; Senes, A.; Bonfiglio, A. Influence of Thermal Ageing on the Stability of Polymer Bulk Heterojunction Solar Cells. *Sol. Energy Mater. Sol. Cells* **2007**, *91*, 385–389.
- (41) Bertho, S.; Janssen, G.; Cleij, T. J.; Conings, B.; Moens, W.; Gadisa, A.; D'Haen, J.; Goovaerts, E.; Lutzen, L.; Manca, J.; et al. Effect of Temperature on the Morphological and Photovoltaic Stability of Bulk Heterojunction Polymer:Fullerene Solar Cells. *Sol. Energy Mater. Sol. Cells* **2008**, *92*, 753–760.
- (42) Gaudiana, R.; Brabec, C. Organic Materials: Fantastic Plastic. *Nat. Photonics* **2008**, *2*, 287–289.
- (43) Shaheen, S. E.; Ginley, D. S.; Jabbour, G. E. Organic-Based Photovoltaics: Toward Low-Cost Power Generation. *MRS Bull.* **2005**, *30*, 10–15.
- (44) Harvard CEP Home Page. <http://cleanenergy.harvard.edu> (accessed August 10, 2011).
- (45) Brédas, J.-L.; Norton, J. E.; Cornil, J.; Coropceanu, V. Molecular Understanding of Organic Solar Cells: The Challenges. *Acc. Chem. Res.* **2009**, *42*, 1691–1699.
- (46) Heremans, P.; Cheyns, D.; Rand, B. P. Strategies for Increasing the Efficiency of Heterojunction Organic Solar Cells: Material Selection and Device Architecture. *Acc. Chem. Res.* **2009**, *42*, 1740–1747.
- (47) Chen, J.; Cao, Y. Development of Novel Conjugated Donor Polymers for High-Efficiency Bulk-Heterojunction Photovoltaic Devices. *Acc. Chem. Res.* **2009**, *42*, 1709–1718.
- (48) Schlenker, C. W.; Thompson, M. E. The Molecular Nature of Photovoltage Losses in Organic Solar Cells. *Chem. Commun.* **2011**, *47*, 3702–3716.
- (49) Borchert, H. Elementary Processes and Limiting Factors in Hybrid Polymer/Nanoparticle Solar Cells. *Energy Environ. Sci.* **2010**, *3*, 1682–1694.
- (50) Blouin, N.; Michaud, A.; Gendron, D.; Wakim, S.; Blair, E.; Neagu-Plesu, R.; Belletête, M.; Durocher, G.; Tao, Y.; Leclerc, M. Toward a Rational Design of Poly(2,7-carbazole) Derivatives for Solar Cells. *J. Am. Chem. Soc.* **2008**, *130*, 732–742.
- (51) Nelson, J.; Kwiatkowski, J. J.; Kirkpatrick, J.; Frost, J. M. Modeling Charge Transport in Organic Photovoltaic Materials. *Acc. Chem. Res.* **2009**, *42*, 1768–1778.
- (52) Zhu, X.-Y.; Yang, Q.; Muntwiler, M. Charge-Transfer Excitons at Organic Semiconductor Surfaces and Interfaces. *Acc. Chem. Res.* **2009**, *42*, 1779–1787.
- (53) Wang, L.; Nan, G.; Yang, X.; Peng, Q.; Li, Q.; Shuai, Z. Computational Methods for Design of Organic Materials with High Charge Mobility. *Chem. Soc. Rev.* **2010**, *39*, 423–434.
- (54) Kerns, E. H. High Throughput Physicochemical Profiling for Drug Discovery. *J. Pharm. Sci.* **2001**, *90*, 1838–1858.
- (55) Bajorath, J. Integration of Virtual and High-Throughput Screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882–894.
- (56) McInnes, C. Virtual Screening Strategies in Drug Discovery. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494–502.
- (57) Gibbon, P.; Andreas, S. High-Throughput Drug Discovery: What Can We Expect from HTS?. *Drug Discovery Today* **2005**, *10*, 17–22.
- (58) Appell, K.; Baldwin, J. J.; Egan, W. J. In *Handbook of Modern Pharmaceutical Analysis: Separation Science and Technology*; Ahuja, S., Scypinski, S., Eds.; Academic Press: San Diego, CA, 2001; Vol. 3, pp 23–56.
- (59) Gordon, E. M.; Gallop, M. A.; Patel, D. V. Strategy and Tactics in Combinatorial Organic Synthesis. Applications to Drug Discovery. *Acc. Chem. Res.* **1996**, *29*, 144–154.
- (60) Ellman, J. A. Design, Synthesis, and Evaluation of Small-Molecule Libraries. *Acc. Chem. Res.* **1996**, *29*, 132–143.
- (61) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions. *J. Med. Chem.* **1994**, *37*, 1385–1401.

- (62) The authors are aware of and acknowledge related work by Geoffrey Hutchison (University of Pittsburgh), Michel Côté (University of Montreal), and Josef Michl (University of Colorado). Their respective research efforts each have a certain degree of overlap with CEP.
- (63) Bajorath, J. *Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery*; Humana Press: Totowa, NJ, 2010.
- (64) Ghose, A. K.; Herbertz, T.; Salvino, J. M.; Mallamo, J. P. Knowledge-Based Chemoinformatic Approaches to Drug Discovery. *Drug Discovery Today* **2006**, *11*, 1107–1114.
- (65) Gaber, M. M. *Scientific Data Mining and Knowledge Discovery: Principles and Foundations*; Springer: Berlin, 2009.
- (66) Grossman, R. L.; Kamath, C.; Kegelmeyer, P.; Kumar, V.; Namburu, R. *Data Mining for Scientific and Engineering Applications*; Springer: Berlin, 2001.
- (67) Suh, C.; Rajan, K. Invited Review: Data Mining and Informatics for Crystal Chemistry: Establishing Measurement Techniques for Mapping Structure-Property Relationships. *Mater. Sci. Technol.* **2009**, *25*, 466–471.
- (68) Rajan, K. Combinatorial Materials Sciences: Experimental Strategies for Accelerated Knowledge Discovery. *Annu. Rev. Mater. Res.* **2008**, *38*, 299–322.
- (69) Suh, C.; Rajan, K. Virtual Screening and QSAR Formulations for Crystal Chemistry. *QSAR Comb. Sci.* **2005**, *24*, 114–119.
- (70) Jain, A.; Hautier, G.; Moore, C. J.; Ong, S. P.; Fischer, C. C.; Mueller, T.; Persson, K. A.; Ceder, G. A High-Throughput Infrastructure for Density Functional Theory Calculations. *Comput. Mater. Sci.* **2011**, *50*, 2295–2310.
- (71) Colella, S.; Mazzeo, M.; Grisorio, R.; Fabiano, E.; Melcarne, G.; Carallo, S.; Angione, M. D.; Torsi, L.; Suranna, G. P.; della Sala, F.; et al. Monodispersed Molecular Donors for Bulk Hetero-junction Solar Cells: From Molecular Properties to Device Performances. *Chem. Commun.* **2010**, *46*, 6273–6275.
- (72) Zhan, X.; Zhu, D. Conjugated Polymers for High-Efficiency Organic Photovoltaics. *Polym. Chem.* **2010**, *1*, 409–419.
- (73) Liang, Y.; Feng, D.; Wu, Y.; Tsai, S.-T.; Li, G.; Ray, C.; Yu, L. Highly Efficient Solar Cell Polymers Developed via Fine-Tuning of Structural and Electronic Properties. *J. Am. Chem. Soc.* **2009**, *131*, 7792–7799.
- (74) Chochos, C. L.; Choulis, S. A. How the Structural Deviations on the Backbone of Conjugated Polymers Influence Their Optoelectronic Properties and Photovoltaic Performance. *Prog. Polym. Sci.* **2011**, *36*, 1326–1414.
- (75) Liu, J.; Mikhaylov, I. A.; Zou, J.; Osaka, I.; Masunov, A. E.; McCullough, R. D.; Zhai, L. Insight into How Molecular Structures of Thiophene-Based Conjugated Polymers Affect Crystallization Behaviors. *Polymer* **2011**, *52*, 2302–2309.
- (76) Mishra, A.; Fischer, M. K. R.; Bäuerle, P. Metal-Free Organic Dyes for Dye-Sensitized Solar Cells: From Structure:Property Relationships to Design Rules. *Angew. Chem., Int. Ed.* **2009**, *48*, 2474–2499.
- (77) Beaujuge, P. M.; Amb, C. M.; Reynolds, J. R. Spectral Engineering in  $\pi$ -Conjugated Polymers with Intramolecular Donor–Acceptor Interactions. *Acc. Chem. Res.* **2010**, *43*, 1396–1407.
- (78) Sokolov, A. N.; Atahan-Evrnek, S.; Mondal, R.; Akkerman, H. B.; Sánchez-Carrera, R. S.; Granados-Focil, S.; Schrier, J.; Mannsfeld, S. C. B.; Zoombelt, A. P.; Bao, Z.; Aspuru-Guzik, A. From Computational Discovery to Experimental Characterization of a High Hole Mobility Organic Crystal. *Nat. Commun.* **2011**, *2*, 437.
- (79) Facchetti, A.  $\pi$ -Conjugated Polymers for Organic Electronics and Photovoltaic Cell Applications. *Chem. Mater.* **2011**, *23*, 733–758.
- (80) Pron, A.; Rannou, P. Processible Conjugated Polymers: From Organic Semiconductors to Organic Metals and Superconductors. *Prog. Polym. Sci.* **2002**, *27*, 135–190.
- (81) Weininger, D. SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (82) Marvin 5.3.2 by ChemAxon, 2010, <http://www.chemaxon.com>
- (83) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
- (84) Farchioni, R.; Gross, G. *Organic Electronic Materials: Conjugated Polymers and Low Molecular Weight Organic Solids*; Springer: Berlin, 2001.
- (85) Meller, G.; Grasser, T. *Organic Electronics*; Springer: Berlin, 2009.
- (86) Paszkowicz, W. Genetic Algorithms, A Nature-Inspired Tool: Survey of Applications in Materials Science and Related Fields. *Mater. Manuf. Processes* **2009**, *24*, 174–197.
- (87) Arora, V.; Bakhshi, A. Molecular Designing of Novel Ternary Copolymers of Donor–Acceptor Polymers Using Genetic Algorithm. *Chem. Phys.* **2010**, *373*, 307–312.
- (88) Giro, R.; Cyrillo, M.; Galvão, D. S. Designing Conducting Polymers Using Genetic Algorithms. *Chem. Phys. Lett.* **2002**, *366*, 170–175.
- (89) Martins, B. V. C.; Brunetto, G.; Sato, F.; Coluci, V. R.; Galvão, D. S. Designing Conducting Polymers Using Bioinspired Ant Algorithms. *Chem. Phys. Lett.* **2008**, *453*, 290–295.
- (90) Lameijer, E.-W.; Bäck, T.; Kok, J.; IJzerman, A. D. Evolutionary Algorithms in Drug Design. *Nat. Comput.* **2005**, *4*, 177–243.
- (91) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2002.
- (92) Sykora, V. J.; Leahy, D. E. Chemical Descriptors Library (CDL): A Generic, Open Source Software Library for Chemical Informatics. *J. Chem. Inf. Model.* **2008**, *48*, 1931–1942.
- (93) Mitra, S.; Datta, S.; Perkins, T.; Michailidis, G. *Introduction to Machine Learning and Bioinformatics*; CRC Press: Boca Raton, FL, 2008.
- (94) Yang, Z. R. *Machine Learning Approaches to Bioinformatics*; World Scientific: Singapore, 2010.
- (95) Bishop, C. M. *Pattern Recognition and Machine Learning*, 2nd ed.; Springer: Berlin, 2007.
- (96) Fukunaga, K. *Introduction to Statistical Pattern Recognition*, 2nd ed.; Academic Press: San Diego, CA, 1990.
- (97) Olivares-Amaya, R.; Amador-Bedolla, C.; Hachmann, J.; Atahan-Evrnek, 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**, accepted for publication.
- (98) R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.
- (99) Dragon 6 by Talete, 2011, <http://www.talete.mi.it/index.htm>
- (100) StarDrop 5.0 by Optibrium, 2011, <http://www.optibrium.com/stardrop/>
- (101) André, J.-M.; Delhalle, J.; Brédas, J.-L. *Quantum Chemistry Aided Design of Organic Polymers: An Introduction to the Quantum Chemistry of Polymers and Its Applications*; World Scientific: Singapore, 1991.
- (102) Risko, C.; McGehee, M. D.; Brédas, J.-L. A Quantum-Chemical Perspective into Low Optical-Gap Polymers for Highly-Efficient Organic Solar Cells. *Chem. Sci.* **2011**, *2*, 1200–1218.
- (103) Wang, L.; Li, Q.; Shuai, Z.; Chen, L.; Shi, Q. Multiscale Study of Charge Mobility of Organic Semiconductor with Dynamic Disorders. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3309–3314.
- (104) Benanti, T.; Venkataraman, D. Organic Solar Cells: An Overview Focusing on Active Layer Morphology. *Photosynth. Res.* **2006**, *87*, 73–81.
- (105) Gregg, B. A. In *Molecules as Components of Electronic Devices*; Marya, L., Ed.; ACS: Washington, DC, 2003; Chapter 19, pp 243–257.
- (106) Xiao, S.; Price, S. C.; Zhou, H.; You, W. In *Functional Polymer Nanocomposites for Energy Storage and Conversion*; Qing, W., Lei, Z., Eds.; ACS: Washington, DC, 2010; Chapter 7, pp 71–80.
- (107) Scharber, M. C.; Mühlbacher, D.; Koppe, M.; Denk, P.; Heeger, A. J.; Waldauf, C.; Brabec, C. J. Design Rules for Donors in Bulk-Heterojunction Solar Cells – Towards 10% Energy-Conversion Efficiency. *Adv. Mater.* **2006**, *18*, 789–794.
- (108) Parr, R.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989; p 333.

- (109) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2001.
- (110) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (111) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- (112) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (113) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (114) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (115) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for Mixing Exact Exchange with Density Functional Approximations. *J. Chem. Phys.* **1996**, *105*, 9982–9985.
- (116) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. **77**, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396.
- (117) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (118) Perdew, J. P.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. Meta-Generalized Gradient Approximation: Explanation of a Realistic Nonempirical Density Functional. *J. Chem. Phys.* **2004**, *120*, 6898–6911.
- (119) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6169.
- (120) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (121) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06 Functionals and Twelve Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (122) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- (123) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (124) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52*, 2769–2773.
- (125) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (126) Zhang, G.; Musgrave, C. B. Comparison of DFT Methods for Molecular Orbital Eigenvalue Calculations. *J. Phys. Chem. A* **2007**, *111*, 1554–1561.
- (127) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (128) Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
- (129) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- (130) Wiberg, K. B.; Rablen, P. R. Comparison of Atomic Charges Derived via Different Procedures. *J. Comput. Chem.* **1993**, *14*, 1504–1518.
- (131) Reed, A. E.; Weinhold, F. Natural Bond Orbital Analysis of near-Hartree–Fock Water Dimer. *J. Chem. Phys.* **1983**, *78*, 4066–4073.
- (132) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (133) Reed, A. E.; Weinhold, F. Natural Localized Molecular Orbitals. *J. Chem. Phys.* **1985**, *83*, 1736–1740.
- (134) Foster, J. P.; Weinhold, F. Natural hybrid orbitals. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
- (135) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular Interactions from a Natural Bond Orbital, Donor–Acceptor Viewpoint. *Chem. Rev.* **1988**, *88*, 899–926.
- (136) Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed.; Gaussian Inc.: Pittsburgh, PA, 1996.
- (137) Stowasser, R.; Hoffmann, R. What Do the Kohn–Sham Orbitals and Eigenvalues Mean? *J. Am. Chem. Soc.* **1999**, *121*, 3414–3420.
- (138) Hamel, S.; Duffy, P.; Casida, M. E.; Salahub, D. R. Kohn–Sham Orbitals and Orbital Energies: Fictitious Constructs but Good Approximations All the Same. *J. Electron Spectrosc. Relat. Phenom.* **2002**, *123*, 345–363.
- (139) Chong, D. P.; Gritsenko, O. V.; Baerends, E. J. Interpretation of the Kohn–Sham Orbital Energies as Approximate Vertical Ionization Potentials. *J. Chem. Phys.* **2002**, *116*, 1760–1772.
- (140) Luo, J.; Xue, Z. Q.; Liu, W. M.; Wu, J. L.; Yang, Z. Q. Koopmans' Theorem for Large Molecular Systems within Density Functional Theory. *J. Phys. Chem. A* **2006**, *110*, 12005–12009.
- (141) Zhan, C.-G.; Nichols, J. A.; Dixon, D. A. Ionization Potential, Electron Affinity, Electronegativity, Hardness, and Electron Excitation Energy: Molecular Properties from Density Functional Theory Orbital Energies. *J. Phys. Chem. A* **2003**, *107*, 4184–4195.
- (142) Ayers, P.; Melin, J. Computing the Fukui Function from Ab Initio Quantum Chemistry: Approaches Based on the Extended Koopmans' Theorem. *Theor. Chem. Acc.* **2007**, *117*, 371–381.
- (143) Shuai, Z.; Wang, L.; Li, Q. Evaluation of Charge Mobility in Organic Materials: From Localized to Delocalized Descriptions at a First-Principles Level. *Adv. Mater.* **2011**, *23*, 1145–1153.
- (144) Scholes, G. D.; Rumbles, G. Excitons in Nanoscale Systems. *Nat. Mater.* **2006**, *5*, 683–96.
- (145) Jaiswal, M.; Menon, R. Polymer Electronic Materials: A Review of Charge Transport. *Polym. Int.* **2006**, *55*, 1371–1384.
- (146) Köhler, A.; dos Santos, D. A.; Beljonne, D.; Shuai, Z.; Brédas, J.-L.; Holmes, A. B.; Kraus, A.; Müllen, K.; Friend, R. H. Charge Separation in Localized and Delocalized Electronic States in Polymeric Semiconductors. *Nature* **1998**, *392*, 903–906.
- (147) Pohl, H. A. Theories of Electronic Behavior in Macromolecular Solids. *J. Polym. Sci. C: Polym. Symp.* **1967**, *17*, 13–40.
- (148) Maitra, N. T. Undoing Static Correlation: Long-Range Charge Transfer in Time-Dependent Density-Functional Theory. *J. Chem. Phys.* **2005**, *122*, 234104.
- (149) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128*, 044118.
- (150) Fabian, J.; Mann, M.; Petiau, M. The Origin of the Color of 1,2-Dithiins - Interpretation by Kohn–Sham Orbitals. *J. Mol. Model.* **2000**, *6*, 177–185.
- (151) Hachmann, J.; Dorando, J. J.; Avilés, M.; Chan, G. K.-L. The Radical Character of the Acenes: A Density Matrix Renormalization Group Study. *J. Chem. Phys.* **2007**, *127*, 134309.
- (152) Dunitz, J. D.; Gavezzotti, A. How Molecules Stick Together in Organic Crystals: Weak Intermolecular Interactions. *Chem. Soc. Rev.* **2009**, *38*, 2622–2633.
- (153) Williams, J. H. The Molecular Electric Quadrupole Moment and Solid-State Architecture. *Acc. Chem. Res.* **1993**, *26*, 593–598.
- (154) Borsenberger, P. M.; Fitzgerald, J. J. Effects of the Dipole Moment on Charge Transport in Disordered Molecular Solids. *J. Phys. Chem.* **1993**, *97*, 4815–4819.
- (155) Young, R. H.; Fitzgerald, J. J. Dipole Moments of Hole-Transporting Materials and Their Influence on Hole Mobility in Molecularly Doped Polymers. *J. Phys. Chem.* **1995**, *99*, 4230–4240.
- (156) Zade, S. S.; Zamoshchik, N.; Bendikov, M. From Short Conjugated Oligomers to Conjugated Polymers. Lessons from Studies on Long Conjugated Oligomers. *Acc. Chem. Res.* **2011**, *44*, 14–24.

- (157) Gierschner, J.; Cornil, J.; Egelhaaf, H.-J. Optical Bandgaps of  $\pi$ -Conjugated Organic Materials at the Polymer Limit: Experiment and Theory. *Adv. Mater.* **2007**, *19*, 173–191.
- (158) Zade, S. S.; Bendikov, M. From Oligomers to Polymer: Convergence in the HOMO–LUMO Gaps of Conjugated Oligomers. *Org. Lett.* **2006**, *8*, 5243–5246.
- (159) Ladik, J. J. Recent Developments in the Quantum Theory of Organic Polymers. *J. Mol. Struct. THEOCHEM* **1991**, *230*, 127–142.
- (160) Ameri, T.; Dennler, G.; Lungenschmied, C.; Brabec, C. J. Organic Tandem Solar Cells: A Review. *Energy Environ. Sci.* **2009**, *2*, 347–363.
- (161) Heimel, G.; Salzmann, I.; Duhm, S.; Koch, N. Design of Organic Semiconductors from Molecular Electrostatics. *Chem. Mater.* **2011**, *23*, 359–377.
- (162) Troisi, A. Charge Transport in High Mobility Molecular Semiconductors: Classical Models and New Theories. *Chem. Soc. Rev.* **2011**, *40*, 2347–2358.
- (163) Brédas, J.-L.; Beljonne, D.; Coropceanu, V.; Cornil, J. Charge-Transfer and Energy-Transfer Processes in  $\pi$ -Conjugated Oligomers and Polymers: A Molecular Picture. *Chem. Rev.* **2004**, *104*, 4971–5004.
- (164) Bethe, H. A.; Salpeter, E. E. *Quantum Mechanics of One and Two Electron Atoms*; Plenum: New York, 1977.
- (165) Singh, R.; Mishra, M. Electronic Structure Analysis and Vertical Ionization Energies of Thiophene and Ethynylthiophenes. *J. Chem. Sci.* **2009**, *121*, 867–872.
- (166) Smith, M. B.; Michl, J. Singlet Fission. *Chem. Rev.* **2010**, *110*, 6891–6936.
- (167) Houili, H.; Picon, J. D.; Zuppiroli, L.; Bussac, M. N. Polarization Effects in the Channel of an Organic Field-Effect Transistor. *J. Appl. Phys.* **2006**, *100*, 023702.
- (168) She, C.; Easwaramoorthi, S.; Kim, P.; Hiroto, S.; Hisaki, I.; Shinokubo, H.; Osuka, A.; Kim, D.; Hupp, J. T. Excess Polarizability Reveals Exciton Localization/Delocalization Controlled by Linking Positions on Porphyrin Rings in Butadiyne-Bridged Porphyrin Dimers. *J. Phys. Chem. A* **2010**, *114*, 3384–3390.
- (169) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM). *J. Phys. Chem. A* **2008**, *112*, 13164–13171.
- (170) Jacquemin, D.; Perpète, E. A.; Scalmani, G.; Frisch, M. J.; Kobayashi, R.; Adamo, C. Assessment of the Efficiency of Long-Range Corrected Functionals for Some Properties of Large Compounds. *J. Chem. Phys.* **2007**, *126*, 144105.
- (171) Le Bahers, T.; Pauporte, T.; Scalmani, G.; Adamo, C.; Ciofini, I. A TD-DFT Investigation of Ground and Excited State Properties in Indoline Dyes Used for Dye-Sensitized Solar Cells. *Phys. Chem. Chem. Phys.* **2009**, *11*, 11276–11284.
- (172) Head-Gordon, M.; Grana, A. M.; Maurice, D.; White, C. A. Analysis of Electronic Transitions as the Difference of Electron Attachment and Detachment Densities. *J. Phys. Chem.* **1995**, *99*, 14261–14270.
- (173) Martin, R. L. Natural Transition Orbitals. *J. Chem. Phys.* **2003**, *118*, 4775–4777.
- (174) Hutchison, G. R.; Ratner, M. A.; Marks, T. J. Accurate Prediction of Band Gaps in Neutral Heterocyclic Conjugated Polymers. *J. Phys. Chem. A* **2002**, *106*, 10596–10605.
- (175) Sánchez-Carrera, R. S.; Atahan, S.; Schrier, J.; Aspuru-Guzik, A. Theoretical Characterization of the Air-Stable, High-Mobility Dinaphtho[2,3-b:2'3'-f]thieno[3,2-b]-thiophene Organic Semiconductor. *J. Phys. Chem. C* **2010**, *114*, 2334–2340.
- (176) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and Density Functional Theories for the Computation of Enthalpies of Formation. *J. Chem. Phys.* **1997**, *106*, 1063–1079.
- (177) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database No 101, Rel. 15a, April 2010; Russell D. Johnson III, Editor; <http://cccbdb.nist.gov/> (accessed August 10, 2011).
- (178) 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.
- (179) San-Fabián, E.; Guijarro, A.; Vergés, J.; Chiappe, G.; Louis, E. PPP Hamiltonian for Polar Polycyclic Aromatic Hydrocarbons. *Eur. Phys. J. B* **2011**, *81*, 253–262.
- (180) Vergés, J. A.; San-Fabián, E.; Chiappe, G.; Louis, E. Fit of Pariser–Parr–Pople and Hubbard model Hamiltonians to Charge and Spin States of Polycyclic Aromatic Hydrocarbons. *Phys. Rev. B* **2010**, *81*, 085120.
- (181) Trevethan, T.; Shluger, A.; Kantorovich, L. Modelling Components of Future Molecular Devices. *J. Phys.: Condens. Matter* **2010**, *22*, 084024.
- (182) World Community Grid Home Page. [www.worldcommunitygrid.org](http://www.worldcommunitygrid.org) (accessed August 10, 2011)
- (183) Clery, D. IBM Offers Free Number Crunching for Humanitarian Research Projects. *Science* **2005**, *308*, 773.
- (184) Shao, Y.; Fusti Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; et al. Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (185) Anderson, D. P. BOINC: A System for Public-Resource Computing and Storage, 2004; <http://dx.doi.org/10.1109/GRID.2004.14>.
- (186) National Energy Research Scientific Computing Center Home Page. [www.nersc.gov](http://www.nersc.gov) (accessed August 10, 2011).
- (187) Catlett, C.; Allcock, W. E.; Andrews, P.; Aydt, R.; Bair, R.; Balac, N.; Banister, B.; Barker, T.; Bartelt, M.; Beckman, P.; et al. TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications. *Adv. Parallel Comput.* **2008**, *16*, 225–249.
- (188) Beazley, D. M. *Python Essential Reference*, 4th ed.; Addison-Wesley: Upper Saddle River, NJ, 2009.
- (189) Langtangen, H. P. *Python Scripting for Computational Science*, 3rd ed.; Springer: Berlin, 2010.