Toward Quantitative Structure−Property Relationships for Charge Transfer Rates of Polycyclic Aromatic Hydrocarbons

Milind Misra,[†] Denis Andrienko,[*,§,‡] Björn Baumeier,[§,‡] Jean-Loup Faulon,[∥] and O. Anatole von Lilienfeld[*,†,§,⊥]

[†]Advanced Device Technologies Department, Sandia National Laboratories, Albuquerque, New Mexico 87185-1322, United States

[‡]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

[§]Institute of Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095-7121, United States

[∥]Institute of Systems & Synthetic Biology, University of Evry, 5 rue Henri Desbruères, 91030, Evry cedex, France

S *Supporting Information*

ABSTRACT: Quantitative structure−property relationships (QSPRs) have been developed and assessed for predicting the reorganization energy of polycyclic aromatic hydrocarbons (PAHs). Preliminary QSPR models, based on a combination of molecular signature and electronic eigenvalue difference descriptors, have been trained using more than 200 PAHs. Monte Carlo cross-validation systematically improves the performance of the models through progressive reduction of the training set and selection of best performing training subsets. The final biased QSPR model yields correlation coefficients q^2 and r^2 of 0.7 and 0.8, respectively, and an estimated error in predicting reorganization energy of ± 0.014 eV.

I. INTRODUCTION

A key property of organic semiconducting materials is that their conducting properties can be tuned by optimizing their chemical structure.[1−5] A practical route to do this includes the synthesis of a new compound, optimization of its processing conditions, fabrication of the device, and measurement of its performance (properties). By repeating this procedure, one can formulate structure−processing−property relationships and proceed with the rational design of organic semiconductors.

It is desirable to assist the design by optimizing material properties using computer simulations. First, methods are required that are capable of predicting the property of interest starting from the chemical structure, preferably without fitting parameters. The second step consists of correlating these properties with the corresponding structures for a specified training set of compounds and formulating quantitative structure−property relationships (QSPRs). Finally, improved compounds are identified for a specific property range.

For organic semiconductors, already the first step in this scheme is nontrivial since charge carrier mobility depends on molecular geometry, electronic structure, and global percolation pathways for charge carriers. Without discussing any details, this represents a typical multiscale problem, and attempts to solve it constitute an entire research field.[6−22] Current experience suggests that it is very difficult to directly evaluate charge carrier mobility as a property of interest for an arbitrary chemical compound, since several assumptions are necessary regarding material morphology, the type of transport, and the model used to describe it. One could, however, ask whether it is possible to find adequate QSPRs that relate chemical structure to charge transport properties, the link between chemical structure and mobility being established first.

In this paper, we construct and assess the quality of several such QSPRs in the context of organic semiconductors. As a test

system, we use polycyclic aromatic hydrocarbons (PAHs). PAHs or, more specifically, discotic liquid crystals have already found application in organic solar cells and field effect transistors.[2,23,24] A typical chemical structure of a discotic liquid crystal consists of a flat conjugated core with side chains attached to its periphery. Discotics self-assemble into columnar structures with aromatic cores stacked on top of each other. Overlap of the π orbitals of these cores enables charge transport along columns, rendering these materials one-dimensional semiconductors. The efficiency of charge transport can be engineered by either varying the shape and size of the conjugated core or influencing their packing through the modification of side chains.

Due to structural, dynamic, and energetic disorder, charge transport in discotic liquid crystals occurs via charge carrier hopping between the neighboring molecules. The rate of hopping is given by the high-temperature nonadiabatic Marcus theory:[6,25,26]

$$\omega = \frac{J^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_B T}}\exp\left[-\frac{(\Delta G - \lambda)^2}{4 k_B T \lambda}\right] \tag{1}$$

where J is the electronic coupling matrix element between the donor and acceptor molecules, λ is the reorganization energy, ΔG is the free energy difference between the initial and final states, and T is the temperature.

Equation 1 identifies several parameters important for charge transport. The transfer integral J is related to the overlap of electronic orbitals, highest occupied molecular orbital (HOMO) for the hole and lowest molecular orbital (LUMO) for the electron transport. As such, it is very sensitive to the relative position and orientation of neighboring molecules.[13,17,27]
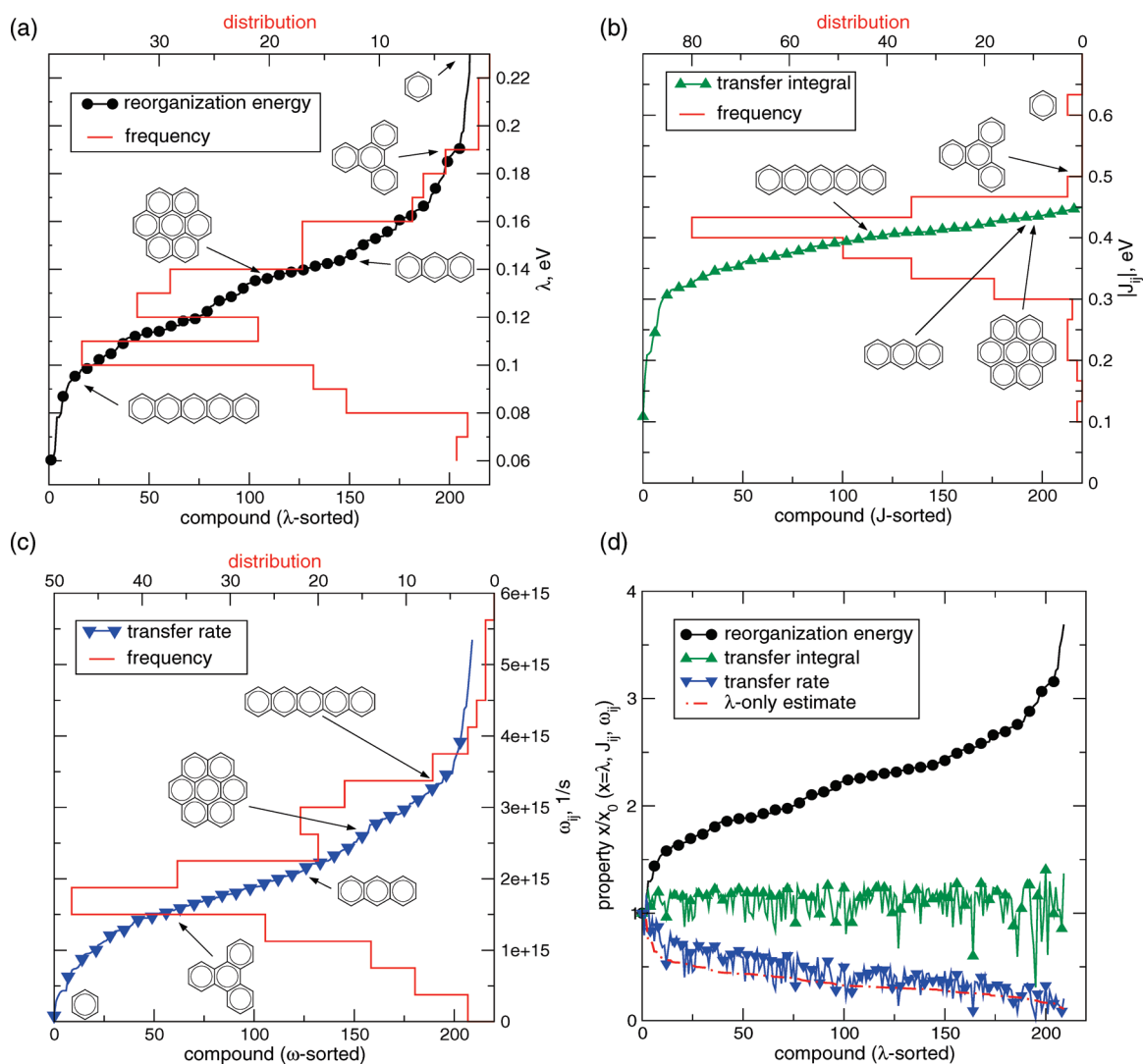
Figure 1. Analysis of reorganization energies, transfer integrals, and transfer rates in the PAH compound data set: Sorted values of (a) λ, (b) J, and (c) ω as functions of compound indices, as well as the respective compound distributions (red lines). For illustrative purposes, several compounds at λ values of 0.3 eV (benzene), 0.19 eV (triphenylene), 0.14 eV (anthracene), 0.13 eV (coronene), and 0.09 eV (pentacene) are shown. Panel d shows a combined plot of the three properties against the λ-sorted compound index. The red dash-dotted line indicates a λ-only prediction of the transfer rates based on eq 1, assuming a constant transfer integral for the entire data set.

In columnar phases of discotics, the maximum of the transfer integral is achieved in a face-to-face molecular arrangement, with the typical intermolecular distance of $d = 3.5$ Å.[24,28−30] In what follows, we assume such an "ideal" molecular arrangement, since it maximizes charge transport and hence provides an upper bound for the charge mobility which can be reached experimentally. We ignore the distribution in transfer integrals due to thermal fluctuations as well as static defects in morphology. Another parameter, ΔG, is the free energy difference between the states with charge localized on a donor or an acceptor of the charge transfer complex. For an ideal face-to-face arrangement, this contribution vanishes due to equivalence of the initial and final states. Finally, the internal reorganization energy, λ, expresses the strength of electron–phonon coupling and has an exponential impact on the transfer rate, with small λ favoring more efficient charge transport.

For an ideal face-to-face columnar alignment, the mobility of the charge carrier along the column is proportional to the hopping rate, eq 1, with $\Delta G = 0$:

$$\mu = \frac{\omega d^2}{k_B T} = \frac{J^2 d^2}{\hbar k_B T}\sqrt{\frac{\pi}{k_B T \lambda}}\exp\left[-\frac{\lambda}{4k_B T}\right] \qquad (2)$$

where d is the distance between neighboring sites. We can therefore argue that large hopping rates (that is, large transfer integrals, small reorganization energies) favor high charge mobilities. Hence, the potential descriptors shall link the chemical structure of a compound with the hopping rate, or, alternatively, J and λ.

In this study, we develop appropriate structure–mobility QSPRs. To do this, we first present how the PAH compound data set was generated and used to select the parameters dominating the charge transport in columnar phases of discotics. We then present two descriptors and assess their performance within preliminary QSPR models. Finally, a robust QSPR model is developed using Monte Carlo cross-validation for variable training/test set ratios.

II. COMPOUND DATA SET

Even for an "ideal" molecular packing, it is not immediately obvious which of the two physical parameters, J or λ, is more important for charge transfer rate prediction. In order to identify the dominant physical parameter and set up the reference values for QSPRs, we have generated a compound data set of PAHs and analyzed its properties. Starting from benzene, we have appended additional aromatic rings at random available bonds. We have used standard carbon—carbon and carbon—hydrogen bond lengths and angles, checking for atom overlaps as well as aromaticity (Hückel rule) and discarding multiple copies of the same PAH. Thus, a data set of 211 closed shell aromatic PAHs with up to nine benzene rings has been generated.

For hole transport, the reorganization energy can be written as a sum of the relaxation energies in neutral and positively charged states

$$\lambda = E_n^+ - E_n^0 + E_c^0 - E_c^+ \qquad (3)$$

where E_g^q is an energy of the compound in charge state q and geometry g. $q = 0$ corresponds to a neutral molecule and $q = +$ to a cation. $g = n$ indicates optimized geometry of a neutral molecule, while $g = c$ corresponds to an optimized cation geometry. Hence, four calculations per compound are necessary, two geometry optimizations for the neutral (E_n^0) and cationic (E_c^+) species and two single point energy calculations for the cationic species in the neutral geometry (E_n^+) and for the neutral species in the cationic geometry (E_c^0).

Reorganization energies were computed using density functional theory (DFT; B3LYP functional,[31] the 6-311++g(d,p) basis set) using the Gaussian 03 package.[32] Figure 1a shows the values of λ in ascending order together with their distribution in the data set. The reorganization energies of 211 compounds are spread from 0.06 to 0.30 eV.

Transfer integrals J were calculated for a cofacial geometry and molecular separation of 3.5 Å using Zerner's Intermediate Neglect of Differential Overlap method as implemented in the Molecular Orbital Overlap package.[33] Figure 1b shows the resulting values of J in ascending order and their distribution within the data set. The transfer integrals span a range of 0.1 to 0.5 eV, which is relatively large due to the assumed columnar stacking of the molecules. This distribution is sharply peaked around 0.4 eV, indicating that there are only small variations of J within the data set. The corresponding distribution of transfer rates ω is shown in Figure 1c.

Figure 1d combines the three parameters $x = \lambda$, J, and ω plotted as a function of the λ-sorted compound index. All values are shown relative to the value of the compound with index zero (x_0). This representation illustrates that among the three parameters the reorganization energy has the largest relative variance and that the transfer integrals only slightly fluctuate around a constant value. To further support this conclusion, we have included a λ-only estimate of the transfer rates, with a constant transfer integral for the entire data set. The result, shown in Figure 1d, corroborates the assumption that, for an "ideal" packing considered here, the reorganization energy is a dominant factor influencing the transfer rates and thereby the charge carrier mobilities. Henceforth, we will concentrate on developing QSPRs for the reorganization energy λ.

III. DESCRIPTORS

QSPR relies on the definition of descriptors that characterize the chemical structure, for instance, the number of atoms,
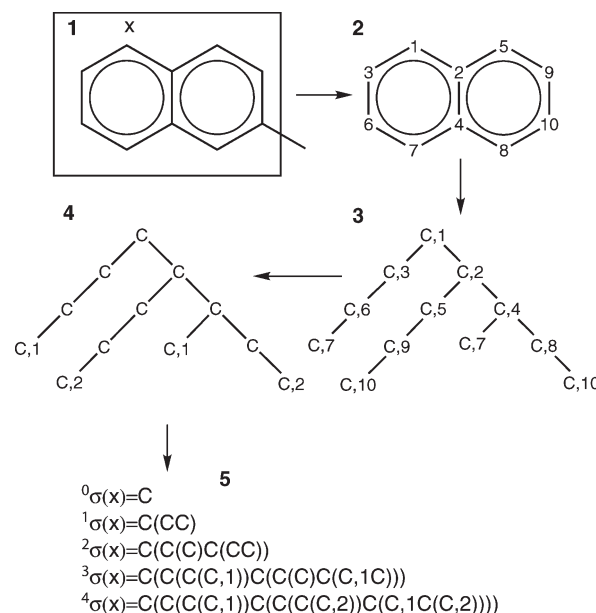

Figure 2. Atomic signatures ${}^h\sigma$ from height $h = 0-4$ for an atom X in 2-methyldecahydronaphtalene. ${}^h\sigma(X)$ is determined as follows: (1) The subgraph containing all atoms at distance 4 from atom X is extracted. (2) This subgraph is canonicalized with atom X having label 1. (3) A tree spanning all edges of the subgraph is constructed. (4) All labels appearing only once are removed, and the remaining labels are renumbered in the order they appear. (5) The atomic signature is determined after reading the tree in a depth-first order, the depth corresponding to height h.

molecular mass, and deviations of the molecular shape from planarity or linearity. Several scalar descriptors that have been investigated but rejected due to their low correlation with the reorganization energy are described in the Supporting Information.

The two descriptors, molecular signature and $\Delta\varepsilon$, had the largest correlation with λ. We first discuss their use for the preliminary QSPR models that are based on the full 211-compound PAH data set. We then present the development of "biased" QSPR models after partitioning the data set into a 188-compound training set and a 23-compound test set. Finally, we address the predictive power of the biased models for the test set compounds.

A. Molecular Signature. The molecular signature is a compilation of a set of atomic signatures, $\{\sigma\}$, that occur in a molecule. It was first presented and applied in the context of structure elucidation[34] and later defined for acyclic compounds and used in QSPR analyses.[35] An atomic signature describes the extended covalent bonding neighborhood of an atom within a molecule up to a certain "height", h. Figure 2 illustrates how atomic signatures, ${}^h\sigma$, are generated. The molecular signature for a given height is a vector that contains the frequencies of all of the ${}^h\sigma$'s occurring in the molecule. As such, it represents a methodical codification system over an alphabet of atom types.

The MolConverter program from ChemAxon[36] was used to convert the *xyz*-coordinate files of the structures in our PAH data set to corresponding simplified molecular input line entry specification (SMILES) strings. SMILES describe chemical structures and topologies using short textual strings.[37] From the SMILES strings, molecular signatures have been determined for individual heights. The correlation between these molecular signatures alone and the reorganization energy is, however, insufficient for making predictions of λ. For example, the correlation
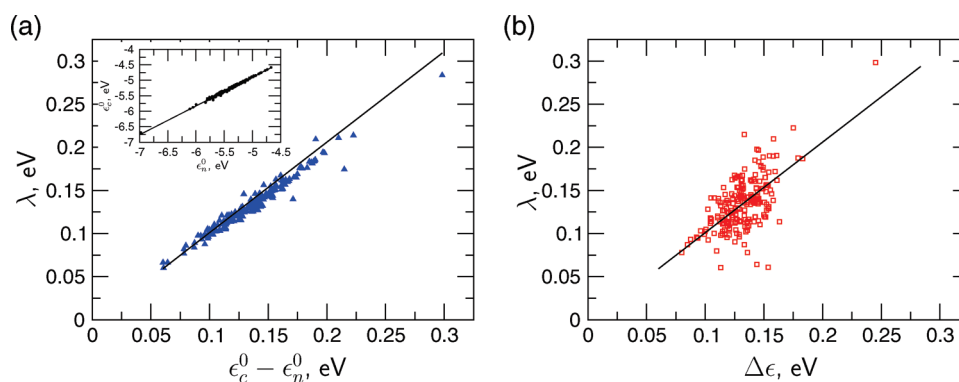
2551

dx.doi.org/10.1021/ct200231z |*J. Chem. Theory Comput.* 2011, 7, 2549–2555

Figure 3. Correlations of calculated reorganization energies λ of the PAH data set with differently predicted values. In panel a, the correlation to $\varepsilon_c^0 - \varepsilon_n^0$ according to eq 4 results in $\lambda = 1.05 \times (\varepsilon_c^0 - \varepsilon_n^0) - 0.004$ [eV] with $r^2 = 0.96$ (blue triangles). The inset shows the correlation between the highest occupied electronic eigenvalue of the neutral molecule in cationic geometry, ε_c^0, with the respective value for the neutral geometry ε_n^0, resulting in $\varepsilon_c^0 = 0.93 \times \varepsilon_n^0 - 0.25$ [eV] and $r^2 = 0.99$. Using this relation to predict λ from ε_n^0 only according to $\Delta \varepsilon$ in eq 6 yields the correlation shown in panel b with $\lambda = 1.01 \times \Delta \varepsilon_n^0 + 0.001$ [eV], $r^2 = 0.39$ (red squares).

coefficient, r^2, of molecular signatures with $h \in (0, 1, 2, 3)$ versus λ does not exceed 0.39.

B. HOMO Eigenvalue Difference, $\Delta\varepsilon$. As mentioned above, the reorganization energy λ expresses the strength of electron−phonon coupling in the molecule. Thus, a weak correlation of λ and the descriptor based solely on structural features, such as the molecular signature, is not surprising. In order to add electronic properties to descriptors, we note that eq 3 can be rearranged in terms of the difference between vertical excitation energies, linking states of the same geometry but a different number of electrons N_e, yielding

$$\lambda = \delta_n - \delta_c,$$
$$\delta_n = E_n^+(N_e - 1) - E_n^0(N_e)$$
$$\delta_c = E_c^+(N_e - 1) - E_c^0(N_e) \qquad (4)$$

Here, δ_n is the iso-nuclear change in energy due to removal of an electron from the neutral species in its relaxed geometry, while δ_c is the iso-nuclear change in energy due to addition of an electron to the cationic species in its relaxed geometry.

Based on molecular grand-canonical ensemble DFT,[38−40] we can Taylor-expand δ_n and δ_c in number of electrons, N_e

$$E(N_e + \Delta N_e) = E(N_e) + \frac{\partial E(N_e)}{\partial N_e}\Delta N_e + O(\Delta N_e^2) \qquad (5)$$

For an exact expression for the exchange-correlation potential within density-functional theory, all higher order terms would vanish for $0 \leq \Delta N_e \leq 2$ because the total potential energy of a molecule with fixed external potential changes only linearly as one varies the number of electrons.[41,42] Since the derivative of the energy with respect to N_e is the eigenvalue of the highest occupied molecular orbital (HOMO),[43,44] we can combine eqs 4 and 5 and express λ as

$$\delta_n = \frac{\partial E_n^0(N_e)}{\partial N_e}\Delta N_e = \varepsilon_n^0(N_e)\Delta N_e$$

$$\delta_c = \frac{\partial E_c^0(N_e)}{\partial N_e}\Delta N_e = \varepsilon_c^0(N_e)\Delta N_e$$

$$\lambda = \varepsilon_c^0 - \varepsilon_n^0 \qquad (6)$$

where $\Delta N_e = -1$ and $\varepsilon_n^0(N_e)$ and $\varepsilon_c^0(N_e)$ denote the eigenvalues of the highest occupied molecular Kohn−Sham orbitals of the

neutral molecule in the respective optimal neutral and cationic geometries.

The exact form of the exchange-correlation functional is, however, unknown. Moreover, the self-interaction error increases for fractional occupation within widely used functionals.[45] The difference between electronic eigenvalues of the HOMOs in the neutrally and cationically relaxed geometries, $\varepsilon_c^0 - \varepsilon_n^0$, yields therefore only an estimate of λ. In our case, we have tested the quality of this approximation for the B3LYP hybrid functional by correlating the λ obtained from the eigenvalues as in eq 6 with the λ obtained from the energies according to eq 3. As shown in Figure 3a, the correlation is very strong with a correlation coefficient r^2 of 0.96. This could be further improved by using functionals that correctly account for fractional occupation numbers.[42]

The (approximate) determination of the reorganization energy according to eq 6 still requires the optimizations of neutral and cationic geometries, as well as a single-point calculation for the neutral molecule in the cationic geometry. While this is one calculation less than in eq 3, it is inconvenient since ideally one would like to predict λ from ground-state properties of the neutral molecule alone, i.e., without having to calculate ε_c^0. We have therefore probed whether ε_c^0 correlates with ε_n^0 in the PAH data set. The inset in Figure 3a shows ε_c^0 plotted versus the respective ε_n^0. The linear regression yields $\varepsilon_c^{pred} = 0.93\varepsilon_n^0 - 0.25$ [eV] with a remarkable correlation of $r^2 = 0.99$. On the basis of these observations, we have used

$$\Delta\varepsilon \equiv \varepsilon_c^{pred} - \varepsilon_n^0 \qquad (7)$$

as an additional scalar descriptor for λ.

Figure 3b shows the correlation of the actual λ from eq 3 with the estimated $\Delta\varepsilon$, $\lambda \approx 1.01\Delta\varepsilon + 0.001$ [eV]. The regression for this expression, however, yields a rather low correlation coefficient of only $r^2 = 0.39$. Thus, solely an electronic descriptor cannot reliably predict reorganization energies.

IV. QSPR MODELS

From the two preceding sections, it is apparent that when used separately neither the structural molecular signature nor the electronic eigenvalue descriptor $\Delta\varepsilon$ are sufficient for reliable quantitative estimates of the reorganization energy in our set of PAHs. Since λ is a measure of the coupling of structural and

Table 1. Preliminary QSPR Models, i—viii, and Corresponding q^2 Values for Multiple Linear Regression (MLR) and Partial Least Squares (PLS), Respectively[a]

h	type	#σ	i	ii	iii	iv	v	vi	vii	viii
0—3	MLR	63	0.29	**0.62**	0.20	0.28	0.29	0.29	0.63	0.64
0—3	PLS	63	0.47	**0.47**	0.47	0.46	0.47	0.47	0.46	0.45
0—4	PLS	431	0.50	**0.50**	0.50	0.49	0.50	0.50	0.50	0.26
0—5	PLS	1635	0.50	**0.50**	0.50	0.50	0.50	0.50	0.50	0.31

[a] See section IV.A and the Supporting Information for more details. Here, h is the signature height; #σ refers to the number of atomic signatures, i.e., the dimension of the molecular signature vector. These models were generated using the data set of 211 PAHs. Highlighted model ii has been used for the construction of the "biased" QSPR. (i) molecular signatures. (ii) molecular signatures + $\Delta\varepsilon$. (iii) molecular signatures + dM (molecular distance). (iv) molecular signatures + dL (molecular linearity). (v) molecular signatures + dP (molecular planarity). (vi) molecular signatures + dH (hydrogen repulsion). (vii) molecular signatures + dM + dL + dP + dH + $\Delta\varepsilon$. (viii) vii redundant descriptors removed based on UFS.

electronic degrees of freedom in a molecule, it is natural to attempt a combination of the two descriptors.

A. Preliminary QSPR Models. For the different heights of molecular signatures (see Section IIIA), we have set up different preliminary QSPR models using signatures of heights 0—3 through 0—5 for the PAH compound data set (without outliers). Specifically, leave-one-out cross-validated correlation coefficients (q^2) have been calculated using multiple linear regression (MLR) and partial least-squares (PLS). These coefficients, together with the preliminary models, are listed in Table 1. For the sake of completeness, we also present the results for additional models that are not based on $\Delta\varepsilon$ but that combine molecular signature with various other scalar structural descriptors. More technical details can be found in the Supporting Information.

Our results show that while the PLS calculations yield a q^2 of around 0.50, indicating predictability in general for all of the models, they do not suggest a preference for a particular descriptor combination. MLR results, in contrast, indicate a clear preference for the model combination of the height 0—3 molecular signature with $\Delta\varepsilon$, which has a q^2 of 0.62 (see model ii in Table 1). For the alternative combinations of molecular signature and various structural scalar descriptors, the corresponding r^2 ranges only from 0.20 to 0.29. An additional model which combines all descriptors considered in this study (see, e.g., model vii in Table 1) does not improve the performance of preliminary QSPR model ii, even when using unsupervised forward selection of the descriptors (see the Supporting Information) to further eliminate redundancy among descriptors (model viii).

Thus, the combination of height 0—3 signatures and $\Delta\varepsilon$ in model ii is identified as the optimal starting combination for developing the "biased" QSPR model in the next section. More specifically, the logarithm of λ is estimated by

$$\log_{10}\lambda = c_\lambda\Delta\varepsilon + c_0 + \sum_i c_i\sigma_i \qquad (8)$$

where i runs over all 63 signatures and where c_0, c_λ, and $\{c_i\}$ are the trained QSPR coefficients of the intercept and the descriptors $\Delta\varepsilon$ and signatures, respectively (see the Supporting Information).

B. Biased QSPR Models. We found that the previously identified optimal preliminary QSPR model can further be

developed into a "biased" QSPR model with the help of Monte Carlo cross-validation. To this end, the PAH compound data set was first split into a total 188-compound training set and a test set of 23 compounds, enabling the validation of the biased models. The test set was determined using dissimilarity-based compound selection, as described in the Supporting Information. The resulting test set compounds are shown in Table 2.

Thereafter, out of the total 188-compound training set, subsets with varying percentages x were defined, where $x \in (5, 10, 15, ..., 90, 95)$%. For each x, 10 000 random partitions from among the 188 compounds were generated. All of the random partitions were subjected to training using the preliminary QSPR model ii, i.e., height 0—3 molecular signatures combined with $\Delta\varepsilon$ based on PLS.

The models obtained, dubbed M_x^k ($k \in 1, 2, ..., 10 000$), were subsequently ranked according to their performance as measured by q^2. For M_{100}, $q^2 = 0.44$ and $r^2 = 0.53$, which is below the conventional predictive threshold of 0.50. As described in more detail in the Supporting Information, q^2 can be improved by reducing the training subset size x, followed by the QSPR model training of 10 000 random partitions for each of these reduced training subsets.

Figure 4 illustrates the results for varying percentage x. The average q^2, i.e., the average of the cross-validated correlation coefficients over all randomly chosen partitions, declines progressively as the training set size decreases. The standard deviation around that average, however, increases even more, thereby enabling us to identify "biased" models $M_x^I, M_x^{II}, M_x^{III}$, etc., namely models that yield the respective best, second best, third best, etc. q_x^2 out of all 10 000 models that have been trained for each partition at a particular x. This behavior is in line with ref 46.

C. Test Set Results of Biased QSPR Models. Figure 4 demonstrates overfitting, namely, that predicting the reorganization energy based on the biased (best performing) Monte Carlo models will always be the more favorable the smaller the subset is. From the behavior of q^2 versus x, one could therefore be tempted to deduce that the optimal model should be based on the smallest training subset. Obviously, q^2 is not a sufficient requirement for the predictability of a model, and only external validation provides a sound assessment of a QSPR model.[47] Thus, to determine the optimal size of the training subset, we compute the root-mean-square (RMS) deviation of predicted λ from actual λ for the 23 test set molecules using the biased M_x^I, where $x = 5, 10, ..., 100$. Note that we excluded two outliers from the test set since they had the largest residuals and corresponded to extreme λ values (maximum and minimum) within the entire compound data set.

As shown in Figure 4, as x decreases from 100 to 60%, RMS remains roughly constant ($\sim$ 16 meV) and starts to strongly increase in oscillatory fashion for subsets smaller than 55%. Since RMS is minimal at $x = 40\%$ (14 meV), we define the corresponding biased model M_{40} as our best QSPR model for predicting λ of PAHs. In contrast, model M_{100} has a higher RMS of 17 meV. Biased QSPR model M_{40} does not only have a lower RMS deviation but also exhibits improved correlation coefficients, $q^2 = 0.70$ and $r^2 = 0.80$. Table 2 lists the residuals for the predictions of λ based on models M_{100} and M_{40}.

In summary, model M_{100} predicts the reorganization energy of more than 75% compounds within a reasonable margin of error (± 20 meV). The biased model M_{40}, however, predicts a larger number (>85%) of test set compounds within the same error margin of ± 20 meV.

Table 2. Actual λ's [eV] and Residuals [meV] of Predicted λ's of the Test Set of Compounds for the Two Biased Models, M_{100} and M_{40}

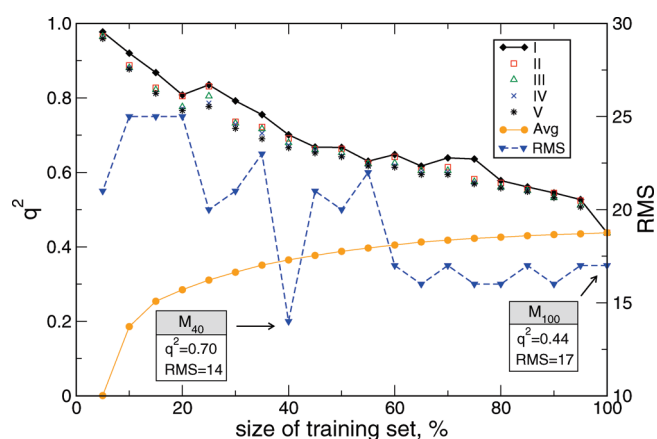| ID	λ	M_{100}	M_{40}	structure	ID	λ	M_{100}	M_{40}	structure	ID	λ	M_{100}	M_{40}	structure
0	0.30	-111	-100		1	0.19	1	8		2	0.22	-43	-27	
7	0.11	4	5		41	0.13	-5	5		43	0.13	-9	-10	
55	0.20	-21	-25		83	0.11	8	13		86	0.14	-10	-15	
107	0.11	11	-11		116	0.10	7	2		124	0.12	-3	6	
129	0.09	-11	-9		143	0.13	6	8		158	0.12	-2	2	
166	0.11	13	13		185	0.13	-13	-4		193	0.10	6	8	
195	0.09	25	18		207	0.06	72	77		209	0.10	31	20	
211	0.16	-34	-29		215	0.15	-3	0						

Figure 4. $q^2(M_x^k)$ of the best five partitions $k \in (I, II, III, IV, V)$ (out of 10 000) as a function of training subset size x in steps of 5%. Yellow represents the average q^2 of 10 000 partitions. Root mean square (RMS) deviation [meV] of actual λ from λ predicted by $M_x^{k=1}$ for test set compounds in Table 2. M_{40} refers to the best partition (I) and is dubbed the biased QSPR model.

V. CONCLUSIONS

On the basis of conceptual density functional theory, we have developed a frontier orbital eigenvalue descriptor $\Delta\varepsilon$ for the empirical prediction of reorganization energies, λ. For a compound data set of over 200 polycyclic aromatic hydrocarbons, we have investigated the performance of various QSPR models aimed at predicting reorganization energies of PAHs based on a combination of a structural and an electronic descriptor, molecular signature, and $\Delta\varepsilon$, respectively. For the entire data set, we find that preliminary QSPR models yield at best a correlation coefficient of $q^2 = 0.5$. Monte Carlo cross-validation with training subsets

permits the definition of a "biased" model with significantly better performance, yielding a q^2 and r^2 of 0.70 and 0.80, respectively, and a root-mean-square deviation of predicted from actual λ of 0.014 eV. Additional scalar structural descriptors, such as average interatomic distance, deviation from linearity, or deviation from planarity yielded only negligible improvement when combined with molecular signature. Furthermore, we have confirmed the basic assumption of selection algorithms based on dissimilarity, which requires that compounds spanning structure/descriptor space also span property/activity space. The main drawback of the proposed descriptor is that it does not account accurately enough for the changes of the molecular geometry upon charging/discharging.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Details of scalar descriptors, molecular signatures, methodology of the QSPR analysis, PLS coefficients for the QSPR models M_{40} and M_{100}, and complete reference Gaussian 03. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: denis.andrienko@mpip-mainz.mpg.de; anatole@alcf.anl.gov.

Present Addresses
⊥Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, United States

■ ACKNOWLEDGMENT

2554

dx.doi.org/10.1021/ct200231z |J. Chem. Theory Comput. 2011, 7, 2549–2555

■ REFERENCES

(1) Mcculloch, I.; Heeney, M.; Bailey, C.; Genevicius, K.; Macdonald, I.; Shkunov, M.; Sparrowe, D.; Tierney, S.; Wagner, R.; Zhang, W.; Chabinyc, M.; Kline, R.; Mcgehee, M.; Toney, M. *Nat. Mater.* **2006**, *5*, 328–333.

(2) Wu, J. S.; Pisula, W.; Müllen, K. *Chem. Rev.* **2007**, *107*, 718–747.

(3) Feng, X.; Marcon, V.; Pisula, W.; Hansen, M. R.; Kirkpatrick, J.; Grozema, F.; Andrienko, D.; Kremer, K.; Müllen, K. *Nat. Mater.* **2009**, *8*, 421.

(4) Nelson, J.; Kwiatkowski, J.; Kirkpatrick, J.; Frost, J. *Acc. Chem. Res.* **2009**, *42*, 1768–1778.

(5) Yan, H.; Chen, Z.; Zheng, Y.; Newman, C.; Quinn, J.; Dotz, F.; Kastler, M.; Facchetti, A. *Nature* **2009**, *457*, 679–U1.

(6) Bredas, J.; Calbert, J.; da Silva, D.; Cornil, J. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5804–5809.

(7) Nagata, Y.; Lennartz, C. *J. Chem. Phys.* **2008**, *129*, 034709.

(8) Kwiatkowski, J. J.; Nelson, J.; Li, H.; Bredas, J. L.; Wenzel, W.; Lennartz, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 1852–1858.

(9) Cheung, D. L.; Troisi, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5941–5952.

(10) Bredas, J. L.; Beljonne, D.; Coropceanu, V.; Cornil, J. *Chem. Rev.* **2004**, *104*, 4971–5003.

(11) Coehoorn, R.; Pasveer, W. F.; Bobbert, P. A.; Michels, M. A. J. *Phys. Rev. B* **2005**, *72*, 155206.

(12) Coropceanu, V.; Cornil, J.; da Silva, D. A.; Olivier, Y.; Silbey, R.; Bredas, J. L. *Chem. Rev.* **2007**, *107*, 2165–2165.

(13) Kirkpatrick, J.; Marcon, V.; Nelson, J.; Kremer, K.; Andrienko, D. *Phys. Rev. Lett.* **2007**, *98*, 227402.

(14) Kirkpatrick, J.; Marcon, V.; Kremer, K.; Nelson, J.; Andrienko, D. *J. Chem. Phys.* **2008**, *129*, 094506.

(15) Marcon, V.; Kirkpatrick, J.; Pisula, W.; Andrienko, D. *Phys. Status Solidi B* **2008**, *245*, 820–824.

(16) Marcon, V.; Breiby, D.; Pisula, W.; Dahl, J.; Kirkpatrick, J.; Patwardhan, S.; Grozema, F.; Andrienko, D. *J. Am. Chem. Soc.* **2009**, *131*, 11426–11432.

(17) Olivier, Y.; Muccioli, L.; Lemaur, V.; Geerts, Y.; Zannoni, C.; Cornil, J. *J. Phys. Chem. B* **2009**, *113*, 14102–14111.

(18) Troisi, A.; Cheung, D. L.; Andrienko, D. *Phys. Rev. Lett.* **2009**, *102*, 116602.

(19) Vehoff, T.; Baumeier, B.; Troisi, A.; Andrienko, D. *J. Am. Chem. Soc.* **2010**, *132*, 11702–11708.

(20) Vehoff, T.; Chung, Y. S.; Johnston, K.; Troisi, A.; Yoon, D. Y.; Andrienko, D. *J. Phys. Chem. C* **2010**, *114*, 10592–10597.

(21) Vehoff, T.; Baumeier, B.; Andrienko, D. *J. Chem. Phys.* **2010**, *133*, 134901.

(22) Lukyanov, A.; Andrienko, D. *Phys. Rev. B* **2010**, *82*, 193202.

(23) Schmidt-Mende, L.; Fechtenkötter, A.; Müllen, K.; Moons, E.; Friend, R. H.; MacKenzie, J. D. *Science* **2001**, *293*, 1119–1122.

(24) Li, J.; Kastler, M.; Pisula, W.; Robertson, J.; Wasserfallen, D.; Grimsdale, A.; Wu, J.; Müllen, K. *Adv. Funct. Mater.* **2007**, *17*, 2528–2533.

(25) Marcus, R. A. *Rev. Mod. Phys.* **1993**, *65*, 599.

(26) Hutchison, G. R.; Ratner, M. A.; Marks, T. J. *J. Am. Chem. Soc.* **2005**, *127*, 2339.

(27) Olivier, Y.; Lemaur, V.; Bredas, J.; Cornil, J. *J. Phys. Chem. A* **2006**, *110*, 6356–6364.

(28) Pisula, W.; Kastler, M.; Wasserfallen, D.; Pakula, T.; Müllen, K. *J. Am. Chem. Soc.* **2004**, *126*, 8074–8075.

(29) Pisula, W.; Tomovic, Z.; Simpson, C.; Kastler, M.; Pakula, T.; Müllen, K. *Chem. Mater.* **2005**, *17*, 4296–4303.

(30) Kastler, M.; Pisula, W.; Wasserfallen, D.; Pakula, T.; Müllen, K. *J. Am. Chem. Soc.* **2005**, *127*, 4286–4296.

(31) Stevens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1993**, *98*, 11623.

(32) Frisch, M. J. et al. *Gaussian 03*, revision B.05; Gaussian, Inc.: Wallingford, CT, 2003.

(33) Kirkpatrick, J. *Int. J. Quantum Chem.* **2008**, *108*, 51.

(34) Faulon, J. L. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1204–1218.

(35) Visco, J.; Pophale, R. S.; Rintoul, M. D.; Faulon, J. L. *J. Mol. Graphics Modell.* **2002**, *20*, 429–438.

(36) http://www.chemaxon.com (accessed July 2011).

(37) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

(38) von Lilienfeld, O. A.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 154104.

(39) Marcon, V.; von Lilienfeld, O. A.; Andrienko, D. *J. Chem. Phys.* **2007**, *127*, 064305.

(40) von Lilienfeld, O. A. *J. Chem. Phys.* **2009**, *131*, 164102.

(41) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. *Phys. Rev. Lett.* **1982**, *49*, 1691.

(42) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *Phys. Rev. Lett.* **2009**, *102*, 066403.

(43) Janak, J. F. *Phys. Rev. B* **1978**, *18*, 7165.

(44) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Phys. Rev. B* **2008**, *77*, 115123.

(45) Zhang, Y.; Yang, W. *J. Chem. Phys.* **1998**, *109*, 2604.

(46) Shao, J. *J. Am. Stat. Assoc.* **1993**, *88*, 486–494.

(47) Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.