Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

A Quantitative Structure-Property Study of Reorganization Energy for known p-Type Organic Semiconductors†

Sule Atahan-Evrenk*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Intramolecular reorganization energy (RE), which quantifies the electron-phonon coupling strength, is an important charge transport parameter for the theoretical characterization of molecular organic semiconductors (OSC). On a small scale, the accurate calculation of the RE is trivial; however, for large-scale screening, faster approaches are desirable. We investigate the structure-property relations and present a quantitative structure-property relationship study to facilitate the computation of RE from molecular structure. To this end, we generated a compound set of 171, which were derived from the known p-type OSCs built from moieties such as the acenes, thiophenes, and pentalenes. We show that simple structural descriptors such as the number of atoms, rings or rotatable bonds only weekly correlate with the RE. On the other hand, we show that regression models based on the more comprehensive representation of the molecules such as the SMILES-based molecular signatures and geometry-based molecular transforms can predict the RE with a coefficient of determination of 0.7 and mean absolute error of 40 meV in a library, in which the RE ranges from 76 to 480 meV. Our analysis indicates that a more extensive compound set for the training is necessary for more predictive models.

1 Introduction

Organic semiconductors (OSCs) show remarkable (opto)electronic properties such as semiconductivity, electroluminescence, or photovoltaic effect[1]. Owing to their potential for solution processibility and compatibility with flexible substrates, they are ideal for low cost, flexible electronics[2,3]. Moreover, the versatility of carbon allows for the discovery of new materials in a vast chemical compound space. Computational screening can facilitate the discovery of new OSCs by helping explore this chemical space at a low cost[4–7].

The understanding of the relationship between molecular/crystal structure and charge transport is crucial to facilitate the synthesis of high-performance organic semiconductors (OSCs). However, a thorough *de novo* multi-scale study of charge carrier mobility in OSCs is a formidable task, especially for screening a large library of compounds. An alternative approach adapts a computational funnel, in which the resources are gradually focused on more promising molecules[6,8]. For the preliminary screening, quantitative structure–property relationship (QSPR) models that can predict material properties according to easily calculated descriptors based on the ground-state molecular structure are indispensable.

Here we focus on the reorganization energy (RE) as a such parameter for large-scale screening[8–10]. The thermal hopping picture allows for a rapid evaluation of the RE at the molecular level. Unfortunately, it has only limited applicability (i.e. for materials with mobility values $< 10^{-2} cm^2/Vs$)[11]. Despite the fact, the magnitude of the electronic coupling term in comparison to the magnitude of the RE is important for the determination of the charge transport regime[12,13]. For single crystalline materials with strong electronic coupling, the decrease in the mobility as a function of increasing temperature, has been shown to be a result of the localization of the charges due to the modulation of the electronic coupling terms, not because of the reorganization energy. However, in these models the larger RE results in lower mobility values[12]. For these reasons, as well as for its importance in the charge carrier hopping, the strategies to adapt RE as a parameter for preliminary large-scale screening is of value. We should note however that charge transport performance of materials are ultimately decided by the crystal structure and ensuing charge transport mechanisms.

Although gas-phase quantum chemical intramolecular RE cal-

[a] *TOBB University of Economics and Technology, Faculty of Medicine, Sogutozu Cad No:43, Sogutozu Ankara, Turkey. Fax: +90 312 292 44 32; Tel: +90 312 292 44 26; E-mail: satahanevrenk@etu.edu.tr*

† Electronic Supplementary Information (ESI) available: The electronic data for the molecular library, the structural and electronic descriptors correlation matrix, explained variance by principle components and pentacene molecular transform plot. See DOI: 10.1039/b000000x/

culations are trivial on a small scale, for large-scale screening, faster calculation schemes are highly desirable. In this article, we present a new data set to explore the structure-property relationships as well as models for the prediction of the RE from molecular structure for the *p*-type OSCs. If successful, such QSPR models might enable the use of the RE as a screening parameter in high-throughput approaches to choose the best candidates for further higher level theoretical studies.

For data-driven materials research, reliable data sets for model *training* is crucial. However, most of the RE data for the experimentally realized molecules is thinly spread in the literature. To date, there is no comprehensive data set which would enable systematic studies based on the RE. Moreover, available quantum-chemical data are obtained at various levels of theory, therefore, it is hard to draw general trends in the structure-property relationship studies.

To the best of our knowledge, only two other attempts were made predicting the RE by using a QSPR methodology. In one study, Misra *et al.* developed QSPR models for structure-mobility predictions for a library including only polyaromatic hydrocarbons (PAHs)[10]. Another smaller scale study[14] used neural networks to predict the RE values for hole and electron hopping in carbon nanotubes. Both of the previous studies used compound libraries built only from fused benzene rings. Here we extend the structure-property study into a more diverse set of compounds which include many state-of-the-art molecular OSCs.

In this work, first we focus on the so called interpretable QSPR models. These models relate the molecular and electronic structural features of the molecules with the intramolecular RE. The intuition for these models usually stem from chemists' observation that certain structural features lead to predictable behavior of the target property in a small set of molecules. It is of interest how these trends are manifested in large and diverse compound sets. Second, we investigate the regression models, in particular partial least squares (PLS) and principle component regression (PCR), which are built using more systematic representation of molecules, where each atom, bond or a connection type is included in the structural coding. We showed that despite the small size of the molecular library, these models show promising predictive potential.

1.1 Molecular Library

Since the development of the first OFET with a polythiophene thin-film active layer[15], many heteroarene- and acene-based compounds have been synthesized as *p*-type OSCs for transistor applications[16]. Over the years, compounds such as the pentacene, oligothiophenes and their solution-processable forms gained a benchmark status. The thienoacenes, which are built from thiophene and acene units, emerged as high-performance and high-stability OSCs[17]. Therefore, we mostly restricted our library to the experimentally known acenes, thiophenes, and thienoacenes. A few compounds with the antiaromatic pentalene moiety were also included for variety[18]. In addition, we included building blocks, such as the smaller acenes and thiophenes, and a few molecules from published computational screening studies. ESI

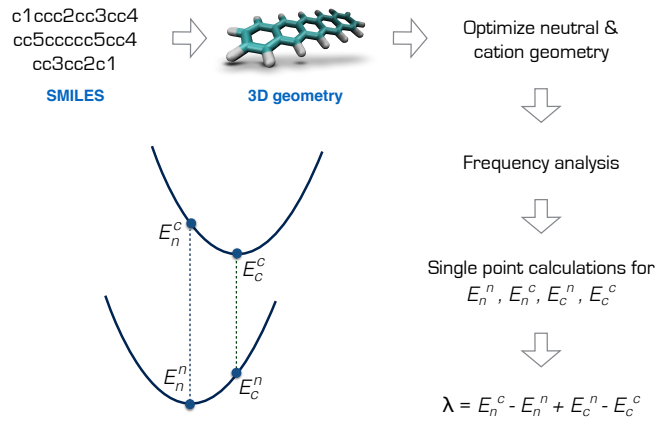

Fig. 1 Calculation of the reorganization energy from the neutral and cation potential energy surfaces for pentacene.

Table 1 lists all 171 molecules included in this study in the order of increasing molecular weight, along with the electronic data and previously available RE values for comparison. This work presents the most comprehensive RE data to date for the experimentally known $p-$type molecular OSCs.

1.2 Reorganization Energy

The RE is the total energy due to the deformation of the lattice and molecular structure as the charge moves. In the limit of the small electronic coupling, the largest deformation occurs in a molecular site; the charge spends a sufficient amount of time on the site so that the molecule can relax into its optimum geometry. As such, the RE can be calculated as the total of internal (*intra*molecular) and external (*inter*molecular) contributions as: $\lambda = \lambda_{int} + \lambda_{ext}$, where λ_{int} can be approximately calculated using gas-phase geometry deformations upon charging[19]. The external contribution λ_{ext} is more challenging to calculate or measure and highly dependent on the morphology[20–22]. Although it is not possible to accurately describe the charge transport without the inclusion of λ_{ext}[21], for the preliminary screening purposes, the λ_{int} can be sufficient[9,10].

By assuming a self-exchange hole transfer reaction, such as $A+A^+ \rightarrow A^+ + A$, the reorganization energy can be calculated as $\lambda = E_n^c - E_n^n + E_c^n - E_c^c$. Here, E_i^j refers to the energy of the charge state j calculated at the optimized geometry i such that E_n^c is the energy of the cation at the optimized neutral geometry. Those points over the potential energy surfaces are labeled in Figure 1, summarizing the computational scheme for pentacene molecule. This scheme requires the optimization of the ground and cation states of the molecules in the gas phase, and two additional single points on the ground and cation potential energy surfaces[19]. The most expensive step in this scheme is the calculation of the Hessian matrices of the optimized geometries necessary to ensure a true minima.‡

‡ For example, for the molecule number **109** in ESI Table 1 (232 electrons), out of 45 hours of computing time at the level of B3LYP/6-31G(d,p), approximately 8 were dedicated to the Hessian calculation for the neutral ground state and 31 were dedi-

It is well known that the density functional theory calculations are heavily influenced by the chosen density functional; for extended molecules, the range-separated tuned functionals gives better theoretical estimates for the RE. Nevertheless, B3LYP functional employed here usually produces the RE trends in the *p*-type OSCs correctly[23], hence widely used in charge transport studies of the OSC materials[24]. Although the experimental RE data for comparison is very limited, the values calculated from B3LYP/6-31G(d,p) level of theory, compares well with the gas-phase photoelectron spectroscopy experiments, for example, in pentacene and perfluorpentacene[22].

All the density functional theory calculations were performed using the Q-Chem software package[25] at the level of B3LYP[26,27] with a gaussian basis 6-31G(d,p)[28–30] without any symmetry restrictions. The spin contamination in the cations were always less than 7%. All minima was ensured with the absence of negative vibrational frequencies through frequency analysis.

1.3 Molecular representations

The molecules were first written as SMILES strings[31] and then represented as vectors either in the structural and electronic descriptor, graph-based signature descriptor[32] or molecular transform descriptor[33] spaces. The signatures were obtained from the SMILES, whereas the molecular transforms require 3*D* geometries. In particular, we explored the molecular transforms obtained from MMFF94[34] force field and density functional theory optimized geometries. In either case, the ground state neutral molecular structures were used. For the molecules with a rotatable bond, we chose only the lowest energy conformer.

The signature descriptors code the neighborhood of each atom in a molecule, starting with its immediate neighbors, and could be generated up to a desired height h. For example at $h = 0$, only the atom itself is included, and at height $h = 1$, its immediate neighbors are also included. Once all the atomic signatures of a molecular set are identified for a particular height, each molecule could be represented as an array storing the frequency of each atomic signature in the molecule. Hence, for a molecular set of size N, a matrix with size of $N \times N_{sig}$ is obtained, where N_{sig} is the total number of unique atomic signatures identified for the set. For simplicity, we call the total of number of signatures at a particular height n as σ_{h0n}.

The number of unique signatures necessary to describe the set increases rapidly. For example, in our molecular set, at σ_{h00}, there are only three signatures for sulphur, carbon, and hydrogen atom types. At $\sigma_{h01}, \sigma_{h02}, \sigma_{h03}$ and and σ_{h04}, there are 16, 115, 590 and 1604 signatures respectively. Previously, signatures up to height 3 (σ_{h03}) has been identified as the sufficient height to describe the RE[10]. Therefore, we explored σ_{h03} and σ_{h04} in our analysis.

Naturally, overfitting is a problem when a matrix with a size of 171×605 or 171×1604 to be solved. To combat overfitting, we explored dimensionality reduction algorithms such as the principle component analysis and partial least squares.

cated to the Hessian calculation of the cation (calculated with Q-Chem 4.0 on four Intel Xeon(R) CPU E3-1246 v3 3.50GHz processors).

As the 3*D* structure descriptor, we used the molecular transform descriptors introduced earlier by Soltzberg and Wilkins[33], and later adapted by Gasteiger and coworkers[35]. The molecular transforms, inspired by the X-ray diffraction intensities, are approximate functions obtained from the 3*D* atomic coordinates of the molecules. By assuming that the molecule is a rigid body and the atoms are point scatterers (no form factors), the 3*D* coordinates of the molecule with N atoms can be converted into a molecular transform as follows:

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} Z_i Z_j \frac{\sin sr_{ij}}{sr_{ij}} \quad (1)$$

where r_{ij} is the distance between the atoms i and j, Z_i is atomic number for the ith atom and the independent variable s measures the scattering angle in units of Å^{-1}. $I(s)$ is an oscillatory function storing geometry information as vector. (See ESI Figure 1)

One advantage of the molecular transform is the fixed length, independent of the library size. The length of the molecular transform descriptors depends how fine the parameter s is defined in the interval $[1,31]$. We determined that 100 points is a good length for the regression models we studied.

For generating the molecular signatures, we used scripts developed by Faulon and coworkers[32]. The molecular mechanics geometries were calculated with the ChemAxon *molconvert* utility. The structural descriptors such as the size, number of rings, type of atoms and bonds as well as the polarizability and van der Waals surface area were calculated with the *cxcalc* utility in ChemAxon suit of programs. The data were managed and analyzed with the modules such as the *statsmodels*[36] and *scikit learn*[37] available in Python language[38].

2 Results and Discussion

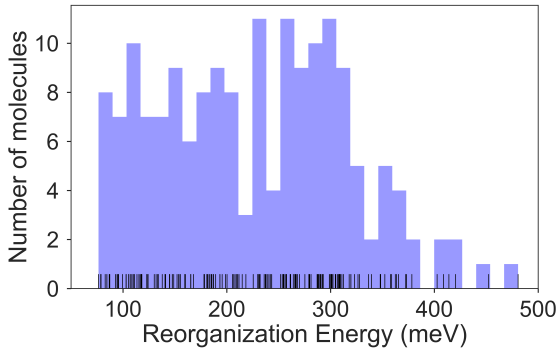

Fig. 2 Histogram of the intramolecular reorganization energies for the molecule set.

The calculated RE values range from 76 to 480 meV, positively skewed as expected from high-performance OSCs (see histogram in Figure 2). The ground-state highest occupied molecular orbital (HOMO) energies show a distribution typical of the *p*-type OSCs with an average of −5.22 eV (See Figure 3).

The electronic data confirms the earlier HOMO eigenvalue difference descriptor derived from the neutral and cation HOMO en-

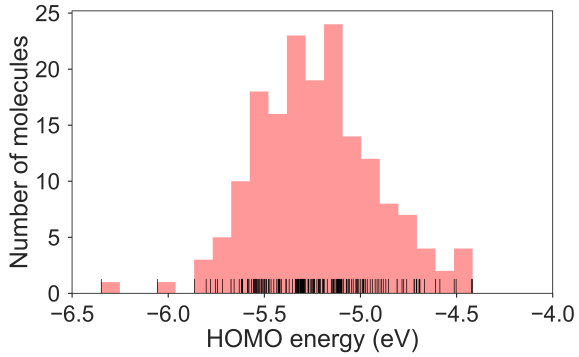

Fig. 3 Histogram of the HOMO energies for the molecule set. Mean= −5.22±0.31 eV.

ergies as $\varepsilon_c^{homo} - \varepsilon_n^{homo}$ [10] (Figure 4a). The cation HOMO energy, ε_c^{homo}, refers to the energy at the HOMO energy for the optimized cation. Although this descriptor shows a very strong correlation with the RE, since we would like to limit the set of descriptors we can use only to the ground neutral state of the molecules, it is unsuitable for our QSPR methodology.

Taherpour *et al.* [14] used both the adiabatic and vertical ionization potentials for the prediction of RE. Figure 4b shows the RE as a regression of the difference between these ionization potentials. This difference is equal to the reorganization of the nuclei as the cation species form and relaxes into the optimum cation geometry[39,40], which could be formulated as $E_n^c - E_c^c$ according to the potential energy surfaces shown in Figure 1. It is, approximately, the half of the RE as shown in Figure 4 when similar relaxation energies are observed for the neutral and cation species. As shown in Figure 4b, except for some of the high RE molecules which have an asymmetry in the relaxation of the charge donor and acceptor upon the vertical transitions and deviate slightly from the linear fit, the RE could be predicted from a linear relationship. Again, since we focus here only on the neutral state descriptors, this difference ($IP_{vert} - IP_{adia}$) is not useful for us as a descriptor.

First, we investigated the correlation of the electronic parameters with the RE. The total electronic energy, HOMO and LUMO energy values and the vertical IP present a weak correlation with the RE. These descriptors, which are also correlated with each other, were not enough by themselves to establish a model for the prediction of the RE. Another electronic descriptor, potentially important for molecular electronic materials, is the polarizability[41]. However, we show in Table 1 that the average molecular polarizability does not correlate with the RE. As explained above, we cannot use the adiabatic and vertical IP in the same model. Moreover, since the computation of the adiabatic IP requires the calculation of the optimum cation geometry, it does not satisfy our criteria that the descriptors should solely be calculated based on the neutral ground state of the compounds.

Owing to the importance of the RE, the structural factors affecting the magnitude of the RE drew considerable attention[24,42–46], for example, the length of the molecule[46,47] or the presence of rotatable bonds[46,48,49]. However, these observations are usually

Table 1 Pearson's r values for the correlation of the RE with electronic descriptors

Descriptor	Pearson's r	p-values
E_n	-0.20	0.0086
ε_n^{homo}	-0.20	0.0086
ε_n^{lumo}	0.16	0.03
ε_c^{homo}	0.076	0.32
ε_c^{lumo}	0.046	0.55
IP_{adia}	0.088	0.25
IP_{vert}	0.19	0.014
Average Polarizability	-0.053	0.49

Table 2 Pearson's r values for the structural descriptors

Descriptor	Pearson's r	p-values
Fused ring count	-0.46	1.5e-10
Rotatable bond count	0.44	2.1e-9
Sulphur atom count	0.34	4.4e-6
van der Waals surface area	-0.052	0.50
Ring count	-0.22	0.0046

limited to a small series of compounds. For example, the RE decreases in a series of compounds from the shorter to longer oligomers[46] or acenes[47,49]. In a compound set with diverse molecular shapes, these simple structural features are not descriptive enough by themselves for a highly predictive QSPR model[10]. Nevertheless, to observe the relationships in our data set, we investigated the correlations of the RE with the structural descriptors such as the number of (fused)rings, number of rotatable bonds, van der Waals surface area and sulphur atom count. Pearson's r values for these are tabulated in Table 2. The largest correlation belongs to the fused ring count with −0.46; the rotatable bond and sulphur atom count follow with 0.44 and 0.34, respectively.

We investigated the multiple linear regression (MLR) models built with the structural descriptors having the p values smaller than 0.01 as well as some of the electronic descriptors. In the ESI Table 2, we show the correlation diagram of the electronic and structural descriptors with each other and the RE. Many of the variables show a high correlation among themselves. Moreover, their correlation with the RE is weak. Only a few of these descriptors could be included in a model at a time with descriptive behaviour, a.k.a. with low p values.Therefore it was a challenge to obtain a predictive MLR model with these set of descriptor variables.

The best performing MLR model with the structural and electronic descriptors when all of the data points were fitted to the model had an R^2 of 0.49 ($R^2_{adj} = 0.48$) ($\log\lambda = 2.661 - 0.496\,\varepsilon_n^{homo}(eV) - 0.055$ Ring count $+ 0.077$ Sulphur atom count $+ 0.141$ Rotatable Bond Count). The test set performance was measured by randomly splitting the data into the test and training sets in the proportion of 20/80, respectively. The average is obtained from 100 runs. The root mean squared error for the prediction of the RE values for the test set was quite large: 72 ± 10meV and $R^2 = 0.31 \pm 0.24$. Due to this low performance, we investigated other descriptors which encode the molecular structures

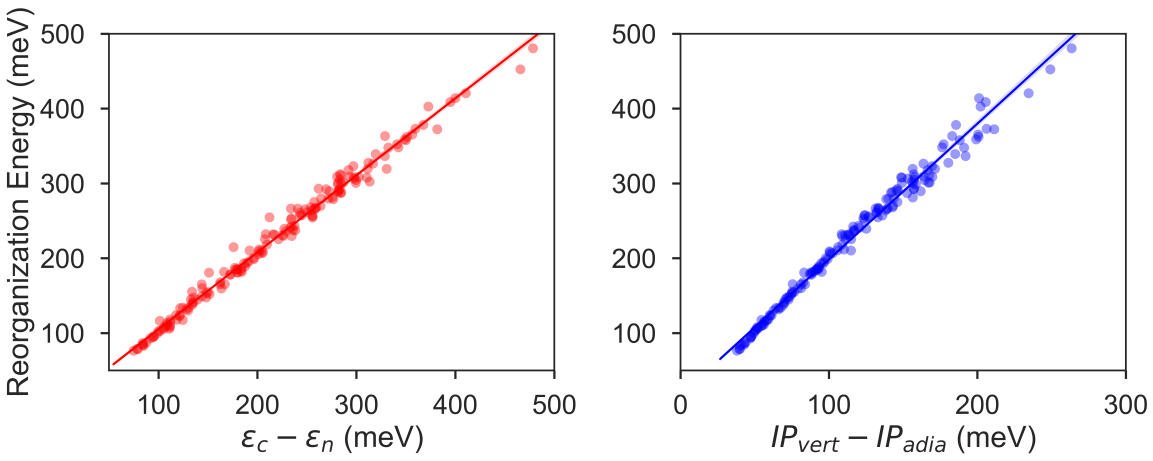

Fig. 4 Intramolecular reorganization energies as a function of (a) the HOMO eigenvalue difference descriptor and (b) the vertical and adiabatic ionization potential difference. The regression lines (a): $\lambda = 1.03 \times (\varepsilon_c^{homo} - \varepsilon_n^{homo}) + 2.16$; (b): $\lambda = 1.8 \times (IP_{vert} - IP_{adia}) + 16.8$.

more systematically.

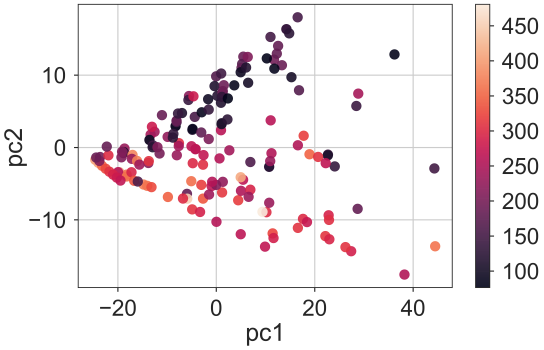

Fig. 5 The RE values (meV) of the compound set scattered in the first two principle components for the signatures up to the height 3, σ_{h03}.

Unlike the descriptors used in the MLR model, the signatures encode the structural features of the molecules systematically including all atoms and the bond types, in the case of the molecular transforms, also information about the $3D$ geometry to a certain extent. However, still two major issues emerge: 1) Large size of the descriptor space, especially when compared to the library size 2) Correlation/collinearity in the descriptors. A transformation of the descriptor space onto a set of orthogonal principle components, such as in the case of principle component analysis (PCA) might help combat both these issues at once. By careful analysis of the train and validation set errors, it is possible to determine the necessary number of the principle components for a model that does not overfit. We report the results from this type of regression as principle component regression, PCR.

The distribution of the molecules labeled according to the RE values in the first three principle components for σ_{h03} is shown in Figure 5. (The ratio of the variance explained by each principle component is shown in ESI Figure 2). The color distribution indicates that the components can help organize data according to the magnitude of the RE. We observed a similar distribution for a deeper signature set, σ_{h04}. However, the principle components of the molecular transforms did not organize the molecules in a meaningful way related to the RE values. Therefore, the PCR models based on the molecular transforms are not included.

The principle components are chosen to explain the variance in the descriptor space only, thus not very effective in the prediction. As an alternative regression approach, we investigated partial least squares (PLS) algorithms. The advantage of the PLS method is that a set of vectors which capture most of the variance in the descriptors is found while the correlation between the descriptor space and target RE values also taken into consideration. For the PLS implementation we used the default version in python *scikit learn* (NIPALS algoritm).

The results from the both regression methods could be found in Table 3. The data for the test performances are obtained by 5-fold cross validation (20% test). The computations are repeated 100 times through random shuffling to gather enough statistics. In parenthesis are the number of principle component vectors (PCR) or latent variables (PLS) chosen with cross-validation. These numbers corresponds to the number of vectors included in the models for which the test set performance reported.

The best test performance belongs to the PLS models with the signature descriptors. The performance of the two signature levels, σ_{h03} and σ_{h04}, are close. These results improved upon our earlier MLR models and the prediction statistics were comparable to the previous models[10]. However, the discrepancy between the test and train performances especially in the PLS models are large and show that one avenue for the improvement of the models could be the expansion of the data set. On the other hand, the PCR models showed less predictive accuracy than the PLS models as expected. In those models, the train and test performances were similar. The molecular transforms based on the DFT optimized geometries, were significantly better both in the test and train sets than the MMFF94 predicted geometries. The Spearman rank correlation coefficients show that the ranking based on the

Table 3 Coefficient of determinations (R^2), Spearman rank correlation coefficients (ρ), and errors of the predicted RE from molecular signature and transform descriptors

Descriptor type	Model[a]	R^2_{train}	R^2_{test}	ρ^{rank}_{test}	RMSE[b]	MAE[b]
Signatures	σ_{03}					
	PLS (5)	0.96±0.00	0.69±0.09	0.81±0.06	55±8	41±6
	PCR (8)	0.62±0.02	0.57±0.09	0.78±0.06	57±8	43± 6
	σ_{04}					
	PLS (8)	0.99±0.00	0.70±0.09	0.82±0.06	54±8	39±7
	PCR (16)	0.67±0.02	0.58±0.10	0.79±0.06	56±7	42±6
Transforms	DFT					
	PLS (7)	0.85±0.01	0.66±0.12	0.81±0.06	60±15	43±7
	MM					
	PLS (5)	0.79±0.01	0.62±0.11	0.77±0.07	60±9	44±6

[a] Numbers in parenthesis represent the size of the descriptor space. [b] The root-mean-squared-error (RMSE) and mean absolute error (MAE) in meV. The statistics obtained from 100 runs, where data is shuffled randomly each time.

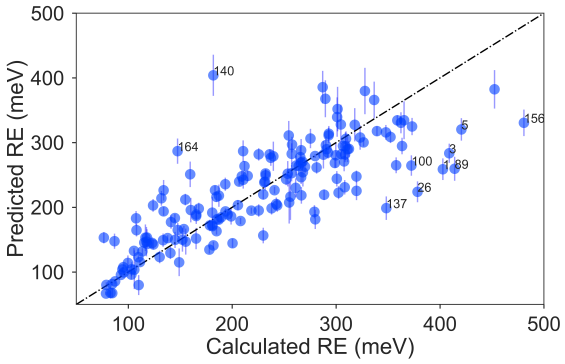

Fig. 6 Comparison of the average predicted and calculated values of RE with the PLS model at the σ_{h04} model. The data is collected from 5-fold CV with 100 runs. Error bars are the standard deviations. The molecules with the errors larger than 100 meV are marked with their number in ESI Table 1.

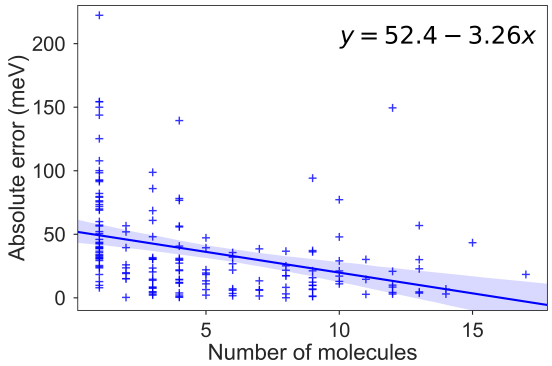

Fig. 7 Absolute error (meV) in the RE for a molecule, as a function of the number of molecules with Tanimoto coefficient larger than 0.85.

DFT geometry derived molecular transform is as accurate as the PLS models for the σ_{h03}, although the average R^2_{test} is smaller than the PLS model.

Finally, we show the pair plot for the best performing model in Figure 6. Some of the outliers (errors larger than 100 meV) are marked with the numbers of the molecules according to ESI Table 1. These outliers could be explained to a certain extent in terms of molecular similarity. For example, the molecule **140** stands out with a large overestimation. This is not surprising since this molecule has a sulfur atom with bonding pattern which does not exist in any of the other compounds. Therefore, the training of the model does not cover the pattern of this molecule. The same could be said for the molecule **26** with the unusual annulene pattern. There is again a large error for the molecule number **1**, the thiophene ring, the only monocycle in the library. The predictions for the molecules with two rings, thienothiophene (**3**), and dithienyl (**5**) are also poor. On the other hand the prediction error for the diphenyl or naphthalene is not large, although there is not many molecules with two rings in the library.

Due to the difficulty of the analysis of each molecule one-by-one and conflicting observations such as mentioned above, we systematically investigated the hypothesis that a molecular (dis)similarity is correlated with the errors as follows. First, we calculated a similarity matrix for all of the molecules based on the Tanimoto metric. Then we counted the number of similar molecules for each molecule with Tanimoto index cutoff of 0.85. We show a regression of the absolute error over this counts in the Figure 7. Albeit small, we observe a negative slope for the fit, indicating that for most molecules with a high count of similar molecules, we observe smaller errors.

3 Conclusions

We presented a new library for the computed reorganization energy values for the experimentally known OSCs and investigated several regression models for the prediction of the RE from molecular structure. Our best model was the PLS regression based on the molecular signature descriptors. We observed that the size and the diversity of the training set is crucial for the establishment of the predictive and generalizable models. The discrepancy between the test and train performances in the PLS models indicates that to reduce the model bias, a larger molecular library will be necessary. We estimate that the library size need to be at least in the order of thousands of compounds. The construction of a library of this size restricted to the known OSC molecules can be challenging, and hence a combinatorially generated training set

with potential OSC molecules might be necessary.

Nevertheless, for the present molecular library, the prediction accuracy of the models (R^2_{test} up to 0.7) with the descriptors based on the ground-state properties of the compounds is remarkable as the RE is a parameter that measures the difficulty of a molecule to go through exchange of holes. Therefore, any higher accuracy prediction should include the effects of the adjustment of the nuclei on the charging process. Work is underway in our laboratory in the direction of the extension of the library and investigation of higher level molecular descriptors for the representation of the molecules. This work also illustrates the potential of the present approach for the prediction of the RE for electron and exciton transport materials.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

SAE acknowledges financial support from TUBITAK BIDEB 2232 (Grant No:114C153) and software support from ChemAxon Ltd.

Notes and references

1 A. Kohler and H. Bassler, *Electronic Processes in Organic Semiconductors: An Introduction*, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany, 1st edn, 2015.

2 H. Klauk, *Organic Electronics: Materials, Manufacturing, and Applications*, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany, 1st edn, 2006.

3 H. Klauk, *Organic Electronics II: More Materials and Applications*, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany, 1st edn, 2012.

4 B. Baumeier, F. May, C. Lennartz and D. Andrienko, *J. Mater. Chem.*, 2012, **22**, 10971–10976.

5 I. Y. Kanal, S. G. Owens, J. S. Bechtel and G. R. Hutchison, *J. Phys. Chem. Lett.*, 2013, **4**, 1613–1623.

6 E. O. Pyzer-Knapp, C. Suh, R. Gomez-Bombarelli, J. Aguilera-Iparraguirre and A. Aspuru-Guzik, *Annu. Rev. Mater. Res.*, 2015, **45**, 195–216.

7 L. H. Nguyen and T. N. Truong, *ACS Omega*, 2018, **3**, 8913–8922.

8 C. Schober, K. Reuter and H. Oberhofer, *J. Phys. Chem. Lett.*, 2016, **7**, 3973–3977.

9 A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C. B. Mannsfeld, A. P. Zoombelt, Z. Bao and A. Aspuru-Guzik, *Nat. Commun.*, 2011, **2**, 437–8.

10 M. Misra, D. Andrienko, B. Baumeier, J.-L. Faulon and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2011, **7**, 2549–2555.

11 H. Oberhofer, K. Reuter and J. Blumberger, *Chemical Reviews*, 2017, **117**, 10319–10357.

12 A. Troisi, *Chem. Soc. Rev.*, 2011, **40**, 2347–2358.

13 S. Giannini, A. Carof and J. Blumberger, *The Journal of Physical Chemistry Letters*, 2018, **9**, 3116–3123.

14 A. A. Taherpour, A. Aghagolnezhad-Gerdroudbari and S. Rafiei, *Int. J. Electrochem. Sci.*, 2012, **7**, 2468–2486.

15 A. Tsumura, H. Koezuka and T. Ando, *Appl. Phys. Lett.*, 1986, **49**, 1210–1212.

16 C. Wang, H. Dong, W. Hu, Y. Liu and D. Zhu, *Chem. Rev.*, 2012, **112**, 2208–2267.

17 K. Takimiya, S. Shinamura, I. Osaka and E. Miyazaki, *Adv. Mater.*, 2011, **23**, 4347–4370.

18 S. Kato, S. Kuwako, N. Takahashi, T. Kijima and Y. Nakamura, *J. Org. Chem.*, 2016, **81**, 7700–7710.

19 V. Coropceanu, J. Cornil, D. S. Filho, D. A, Y. Olivier, R. Silbey and J.-L. Brédas, *Chem. Rev*, 2007, **107**, 926–952.

20 J. E. Norton and J.-L. Bredas, *J. Am. Chem. Soc.*, 2008, **130**, 12377–12384.

21 D. P. McMahon and A. Troisi, *J. Phys. Chem. Lett.*, 2010, **1**, 941–946.

22 S. Kera, S. Hosoumi, K. Sato, H. Fukagawa, S.-i. Nagamatsu, Y. Sakamoto, T. Suzuki, H. Huang, W. Chen, A. T. S. Wee, V. Coropceanu and N. Ueno, *J. Phys. Chem. C*, 2013, **117**, 22428–22437.

23 C. Brückner and B. Engels, *Journal of Computational Chemistry*, 2016, **37**, 1335–1344.

24 H. Geng, Y. Niu, Q. Peng, Z. Shuai, V. Coropceanu and J. L. Brédas, *J. Chem. Phys.*, 2011, **135**, 104703.

25 Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. W. III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. D. Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. G. III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. S. III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E.

Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill and M. Head-Gordon, *Mol. Phys.*, 2015, **113**, 184–215.
26 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
27 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
28 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
29 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.
30 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.
31 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
32 J.-L. Faulon, D. P. Visco and R. S. Pophale, *J. Chem. Inf. Comput. Sci*, 2003, **43**, 707–720.
33 L. J. Soltzberg and C. L. Wilkins, *J. Am. Chem. Soc.*, 1977, **2**, 439–443.
34 T. A. Halgren, *Journal of Computational Chemistry*, 1996, **17**, 553–586.
35 J. Schuur and J. Gasteiger, *Analytical chemistry*, 1997, **69**, 2398–405.
36 J. Seabold and J. Perktold, Proceedings of the 9th Python in Science Conference, 2010.
37 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.
38 *Python*, 2016, `https://www.python.org/`.
39 M. Meot-Ner, S. F. Nelsen, M. F. Willi and T. B. Frigo, *Journal of the American Chemical Society*, 1984, **106**, 7384–7389.
40 S. F. Nelsen, S. C. Blackstock and Y. Kim, *Journal of the American Chemical Society*, 1987, **109**, 677–682.
41 E. A. Silinsh, *Organic Molecular Crystals: Their Electronic States*, Springer-Verlag, Berlin Heidelberg, 1st edn, 1980.
42 R. Zhu, Y.-A. Duan, Y. Geng, C.-Y. Wei, X.-Y. Chen and Y. Liao, *Comp. Theor. Chem.*, 2016, **1078**, 16–22.
43 M. Mamada and Y. Yamashita, in *S-Containing Polycyclic Heteroarenes: Thiophene-Fused and Thiadiazole-Fused Arenes as Organic Semiconductors*, Wiley-VCH Verlag GmbH & Co. KGaA, 2015, pp. 277–308.
44 H.-Y. Chen and I. Chao, *ChemPhysChem*, 2006, **7**, 2003–2007.
45 V. Coropceanu, O. Kwon, B. Wex, B. R. Kaafarani, N. E. Gruhn, J. C. Durivage, D. C. Neckers and J.-L. Brédas, *Chem. Eur. J.*, 2006, **12**, 2073–2080.
46 G. R. Hutchison, M. A. Ratner and T. J. Marks, *J. Am. Chem. Soc.*, 2005, **127**, 2339–2350.
47 W.-Q. Deng and W. A. Goddard, *The Journal of Physical Chemistry B*, 2004, **108**, 8614–8621.
48 D. A. da Silva Filho, V. Coropceanu, D. Fichou, N. E. Gruhn, T. G. Bill, J. Gierschner, J. Cornil and J.-L. Brédas, *Philos. Trans. R. Soc., A*, 2007, **365**, 1435–1452.
49 S. Atahan-Evrenk and A. Aspuru-Guzik, *Top. Curr. Chem.*, 2014, **345**, 95–138.