

Quantitative structure-property relationship model leading to virtual screening of fullerene derivatives: Exploring structural attributes critical for photoconversion efficiency of polymer solar cell acceptors

Supratik Kar^a, Natalia Sizochenko^a, Lucky Ahmed^b, Victor S. Batista^b, Jerzy Leszczynski^{a,*}

^a Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, Jackson, MS, USA

^b Department of Chemistry, Yale University, New Haven, CT, USA

ARTICLE INFO

Article history:

Received 23 January 2016

Received in revised form

30 May 2016

Accepted 7 June 2016

Available online 22 June 2016

Keywords:

Fullerene

Power conversion efficiency

Polymer solar cells

Photovoltaics

QSPR

Virtual screening

ABSTRACT

The power conversion efficiency (PCE) of pure polymer solar cells (PSCs) remains low, although significantly higher values could be achieved by using PSCs as carrier donors in conjunction with composite fullerene derivative (FD) acceptors. Significant resources, however, are required to experimentally develop and screen FDs that may serve as efficient acceptors in PSCs. Often, the materials are expensive, the methods are time consuming, and the production processes can generate toxic hazards. As an alternative approach, we introduce a quantitative structure-property relationship (QSPR) model for predicting the PCE of 59 FDs, including both C₆₀ and C₇₀ FDs. The QSPR model enables identification of the essential structural attributes necessary for quantifying the molecular prerequisites of diverse FDs, chiefly responsible for high PCE of PSC acceptors in composition with poly(3-hexylthiophene) (P3HT). The identified properties and structural fragments are particularly valuable for guiding future synthetic efforts for development of FDs with improved power conversion efficiency. Furthermore, a large number of FDs are collected to generate a database. Virtual screening of the database employing the developed QSPR model allows for identification of nine FDs with higher PCE than previously studied FDs.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Environmental pollution and resource depletion are serious problems that require immediate solutions. Solar cells are already producing a significant impact as viable solutions to sustainable energy. Replacing oil and coal energy sources by solar cell energy would provide multiple benefits to society, including increasing economic competitiveness in the global clean energy race, cutting carbon pollution to combat climate change and wide-spreading energy independence. In particular, solar technology already supports broader national priorities as a domestic energy source in the United States, including national security, economic growth and job creation [5]. Furthermore, it has great potential to modernize the global energy infrastructure since PSC panels on just 0.6% of total land area could supply electricity to power the entire nation [5]. An outstanding challenge is the development of materials for cost effective solar energy panels.

Significant research is currently directed at generating

materials for clean, renewable energy and photocatalytic approaches for sustainable chemistry [1]. Recently, organic photovoltaics have been the subject of intense research [2,3]. The power conversion efficiencies (PCEs) of organic photovoltaic materials, however, have remained significantly lower than those of silicon-based wafers or other semiconductor materials [4]. Therefore, there is great interest in the discovery of organic materials with high PCE to bypass some of the disadvantages of silicon-based solar cells technologies, such as high cost, heavy weight, limited silicon resources, and production methods that lead to high environmental pollution and high energy consumption during purification of silicon materials [4].

Recent studies have shown that combining electron donor polymer materials with composite of fullerene derivatives (FDs) as acceptors could provide attractive PCEs [4]. The efficiency conversion could be further improved by optimizing the excitation, dissociation, and charge transport processes in the device [4]. Therefore, polymer solar cells (PSCs) with a bulk-heterojunction (BHJ) active layer attract much attention and offer advantages in fabrication, cost, and flexibility, when compared to traditional inorganic semiconductor photovoltaics [6,7]. The general mechanism behind such PSCs is as follows: initially light irradiates the

* Correspondence to: Department of Chemistry and Biochemistry, Jackson State University, 1400 Lynch Street, Jackson, MS 39217-0510, USA.

E-mail address: jerzy@icnanotox.org (J. Leszczynski).

polymer donor through the transparent electrode to form excitons. The excitons diffuse to the donor/acceptor interface where they dissociate into electrons and holes. The electrons are typically localized on the acceptor LUMO and holes on the donor HOMO. Next, the dissociated electrons and holes are collected by the electrodes, resulting in overall photon-to-electron conversion [8].

Many conjugated polymers have been synthesized for use as donor materials in PSCs. Poly(3-hexylthiophene) (P3HT) is among the most widely used polymer in photovoltaic devices, typically providing PCE as high as 4.2 [8,9]. In addition, since [6,6]-phenyl-C61-butyric acid methylester (PCBM) was first introduced as PSC acceptor, a significant number of fullerene derivatives (FDs) have been synthesized and characterized to enhance the photovoltaic performance of PSCs, though relatively few have been able to outperform PCBM [10,11]. As a result, there is a constant effort to design more efficient FDs that might serve as PSC acceptors [12]. The required experimental work, however, requires a significant amount of resources. Experiments are expensive, time consuming, and may generate toxic hazards [10]. Therefore, there is significant interest in the use *in silico* modeling techniques that could predict PCE values for newly PSCs. Such predictions could significantly expedite experimental efforts, narrowing the range of promising acceptor materials by ruling out candidates that are doomed to fail experimental tests.

The use of *in silico* methods for chemical property prediction is well established and quantitative structure-property relationship (QSPR) methods emerged as an important computational tool with a diverse range of applications [13]. Robust and validated QSPR models can predict properties for new or untested molecular structures. In this way, QSPR approaches provide insights that expedite the design of novel compounds with desired properties [13]. When applied to property and risk assessment of chemical compounds, regulatory agencies worldwide have already accepted the implication of QSPR models. To date, however, only a single QSPR model has been built to predict the PCE of FDs for PSCs [14]. The model employed 45 FDs and was limited to C₆₀ FDs. Therefore, further work is required in the field to demonstrate the capabilities and limitations of QSPR models as applied to materials for energy research.

Here, we derive the first global QSPR model for predicting PCE values of 59 FDs, including both C₆₀ and C₇₀ FDs. Multiple linear regression (MLR) based on a genetic algorithm (GA) as a variable selection tool is employed to develop our predictive QSPR model. All of the models have been assessed according to the recommended Organization of Economic Co-operation and Development (OECD) principles [15]. In addition to identifying the structural features relevant for higher PCE value of FDs as acceptors for PSCs, the data analysis provides significant insights on the applicability of this type of statistical modeling. A large number of FDs (169) were collected to generate a database, which was subsequently followed to identify the most efficient FDs. Along with the predicting and quantifying the structural features responsible for higher PCE values of FDs, our study thus provides helpful insights regarding the use of virtual screening approaches to identification of lead FDs. A complete workflow of our study is shown in Fig. 1, addressing the issue of low efficiencies in PSCs which must be resolved to transform PSCs into competitively viable products.

2. Methodology

2.1. Dataset

Our data set includes 59 fullerene derivatives (52C₆₀ derivatives and 7C₇₀ derivatives) as solar cell acceptors with experimental PCE values collected from the literature [8,16,17]. The experimental % PCE values for all compounds are assessed according to BHJ devices, where P3HT is the donor material and FDs are the electron acceptor components. The BHJ devices were assembled by both solvent annealing and thermal annealing methods. Chemical structures and experimental %PCE values for all compounds are given in Table 1.

2.2. Descriptors calculation

2.2.1. Structure preparation: molecular and quantum mechanics

FDs structures were first prepared by molecular mechanics

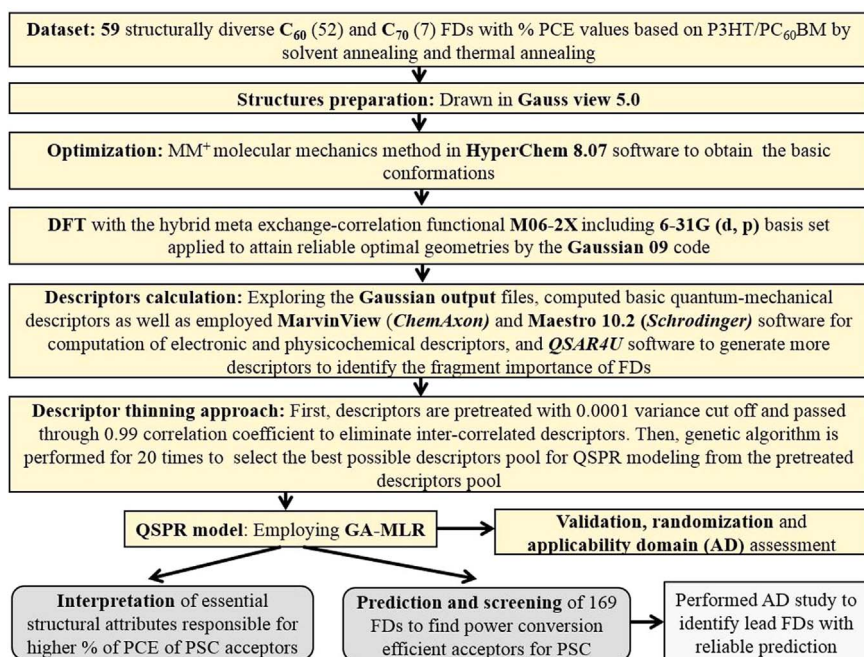


Fig. 1. A complete scheme of the present study.

Table 1Chemical structure of C₆₀ and C₇₀ fullerene derivatives with their experimental and predicted % PCE properties.


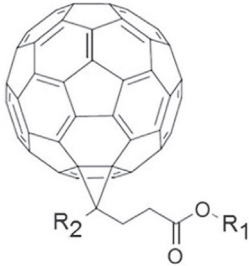
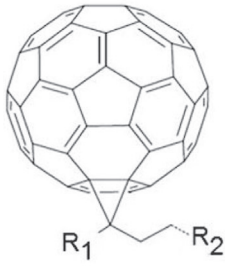
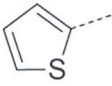
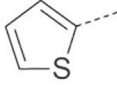
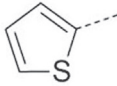
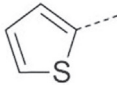
ID	Substituents						PCE (%)	
	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	Experimental	Predicted
								
1*	C ₄ H ₉	C ₆ H ₅	-	-	-	-	1.9	1.96
								
2	CH ₃	C ₆ H ₅	-	-	-	-	0.4	1.57
3	C ₃ H ₇	C ₆ H ₅	-	-	-	-	2.2	1.65
4	iso-Propyl	C ₆ H ₅	-	-	-	-	2.8	1.70
5	C ₄ H ₉	C ₆ H ₅	-	-	-	-	2.7	1.37
								
6*	-COOCH ₂ Ph	C ₆ H ₅	-	-	-	-	2.5	1.65
7	-CH ₂ COOCH ₂ CH ₃	C ₆ H ₅	-	-	-	-	2.7	3.11
8	-CH ₂ COOCH ₃		-	-	-	-	3.7	3.07
9	-COOCH ₃	4-OCH ₃ Ph	-	-	-	-	0.05	1.40
10	-COOC ₂ H ₅		-	-	-	-	2.5	2.73
11*	-COOC ₃ H ₇		-	-	-	-	3.4	2.34
12	C ₄ H ₉		-	-	-	-	2.9	2.06

Table 1 (continued)



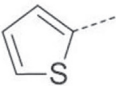
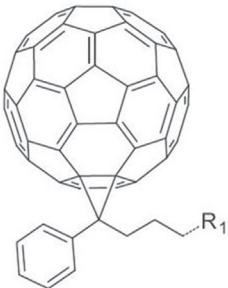
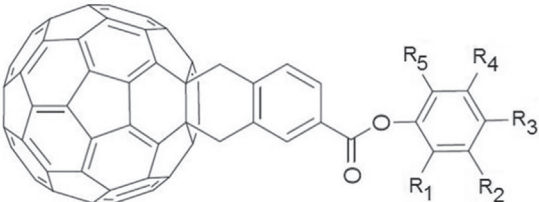
								
13	C ₆ H ₁₃	C ₆ H ₅	-	-	-	-	2.8	2.74
								
14	C ₄ H ₉		-	-	-	-	2.1	1.67
15	-COOCH ₃	(CH ₂) ₂ COOCH ₃	-	-	-	-	0.02	0.21
16 ^a	-CH ₂ CH ₂ COOCH ₂ CH ₂ OCH ₃	-COOC ₂ H ₅	-	-	-	-	0.9	0.16
17	H	COOC ₈ H ₁₇	-	-	-	-	0.3	0.11
								
18	-COOCH ₃	-	-	-	-	-	3.5	3.05
19 ^a	CH ₂ COOCH ₃	-	-	-	-	-	2.3	2.81
20	(CH ₂) ₂ COOCH ₃	-	-	-	-	-	3.6	2.56
21	(CH ₂) ₃ COOCH ₃	-	-	-	-	-	2.8	2.36
								
22	H	H	H	H	H	-	4.2	3.95
23 ^a	H	H	OCH ₃	H	H	-	3.6	3.85
24	H	OCH ₃	OCH ₃	OCH ₃	H	-	1.2	1.31
25	H	H	F	H	H	-	3.5	3.63

Table 1 (continued)

[illegible]

Table 1 (continued)

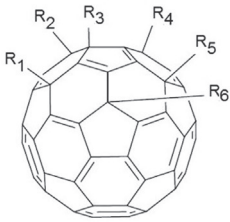
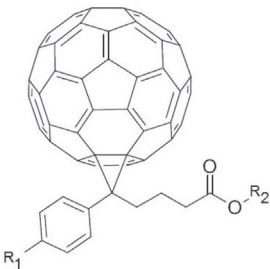
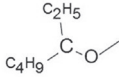
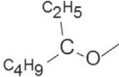
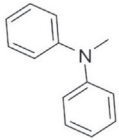
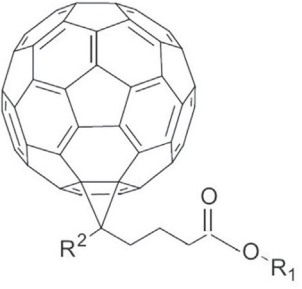
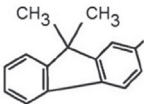
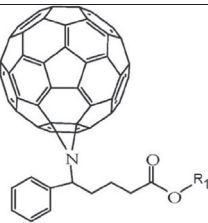
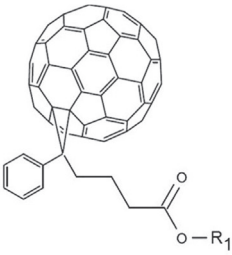
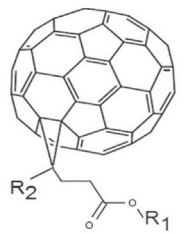


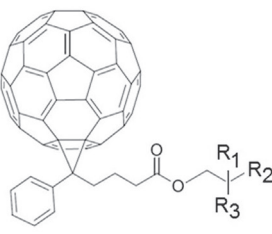
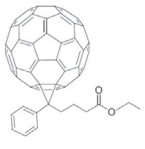
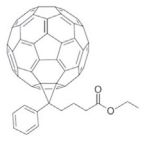
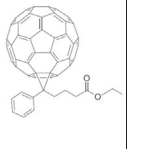
								
41	CH ₃	CH ₃	C ₈ H ₁₇	CH ₃	CH ₃	CH ₃	0.15	0.83
								
42*		H	-	-	-	-	1.73	1.77
43		CH ₃	-	-	-	-	1.01	2.03
44		CH ₃	-	-	-	-	4	3.06
								
45	C ₄ H ₉	C ₆ H ₅	-	-	-	-	2.45	2.86
46	C ₈ H ₁₇	C ₆ H ₅	-	-	-	-	1.27	1.96
47*	C ₁₂ H ₂₅	C ₆ H ₅	-	-	-	-	1.04	1.05
48*	C ₁₆ H ₃₃	C ₆ H ₅	-	-	-	-	0.11	0.15
49	CH ₃		-	-	-	-	4	4.05

Table 1 (continued)

								
50	CH ₃	-	-	-	-	-	2.3	2.32
								
51*	CH ₃	-	-	-	-	-	4.1	3.06
								
52	CH ₃	C ₆ H ₅	-	-	-	-	1.2	1.58
53	C ₂ H ₅	C ₆ H ₅	-	-	-	-	1.7	1.67
54	C ₄ H ₉	C ₆ H ₅	-	-	-	-	1.7	1.42
55	C ₃ H ₇		-	-	-	-	1.1	2.44
56*	C ₄ H ₉		-	-	-	-	1.1	2.20
								
57	CH ₃	CH ₃	CH ₃	-	-	-	2.74	3.23
58	CH ₃		CH ₃	-	-	-	3.7	3.18
59	CH ₃			-	-	-	2.46	2.70

*Compounds present in the test set.

(MM), using the MM⁺ molecular mechanics method as implemented in the HyperChem 8.07 software package [18], and subsequently optimized by density functional theory (DFT) to evaluate properties and descriptors as implemented with the hybrid meta exchange-correlation functional M06-2X [19] and the 6-31G (d,p) basis set [20]. DFT calculations were carried out by using Gaussian 09 [21]. The Berny algorithm, using GEDIIS in redundant internal coordinates, was applied to obtain the local minima [22] following earlier studies where the selected combination of functional and basis set was found to be robust and reliable for calculations of fullerene derivatives [23].

2.2.2. Descriptors

Basic quantum-mechanical descriptors like energy of HOMO and LUMO, difference of HOMO and LUMO energy, valence band, conduction band, hardness, mulliken electronegativity, electronic chemical potential and dipole moment, were computed and extracted from the standard Gaussian output files. Next, *Maestro 10.2* (Schrodinger) [24] and *MarvinView* (ChemAxon) [25] software packages were employed to compute topological, semi-empirical and physicochemical descriptors as well as solvent accessible surface area descriptors, respectively. Additionally, *QSAR4U* software was utilized to generate descriptors derived from Simplex Representation of Molecular Structure (SiRMS) [26,27], identifying the FD fragment importance in the dataset. In the framework of SiRMS, molecules are represented as a system of simplexes (i.e., fragments of fixed composition and topology) and a 2D level of molecular representation is utilized to generate simplex fragments of size 3–5 atoms. Detailed description of HiT QSAR based on SiRMS could be found elsewhere [27]. Three fundamental physical properties were considered to describe simplex fragments: atom types, partial charges, or van-der-Waals interaction. On the first step, fullerene was represented as molecular graph. Next, each atom was labeled in accordance with certain physical property. For example, for partial charges, all atoms in molecular graph were labeled as A, B and C using range $-0.05 (A) \leq 0 (B) \leq 0.05 (C)$. Differentiation based on Van-der-Waals interactions was achieved by dividing types of atoms into six groups: $50 (A) \leq 100 (B) \leq 250 (C) \leq 400 (D) \leq 650 (E) \leq 2000 (F)$ [VdW units]. After differentiation, the molecules were represented as molecular graphs where the vertices were marked by the three properties listed above. Next, all molecules were fragmented to obtain all possible combinations of fragments (namely, simplexes). Finally, the numbers of simplexes of certain type (e.g., A-B-D-G) were used as descriptors. Therefore, a complete pool of 772 descriptors was computed to model the % PCEs of our 59 FDs.

2.3. Data pre-processing

Descriptors were initially pretreated with a 0.0001 variance cut off and passed through a 0.99 correlation coefficient to eliminate correlations between them and to reduce the noise level in the values of input descriptors and correlations. Then, a genetic algorithm (GA) was applied for 20 iterations to select the best possible subset of descriptors for QSPR modeling from the pretreated pool [28]. Such a 'descriptor-thinning' procedure led us to identify 23 descriptors from the pool of 772, which are most relevant for our study.

Before splitting the dataset into training and test subsets, it is vital to classify the dataset according to the calculated properties (descriptors) and to evaluate the number and size of distinctive groups of FDs. For this purpose we have applied a Kohonen's self-organizing map (SOM) algorithm to obtain a matrix dataset of 59×772 (45,548) data points. Application of artificial neural network analysis in the SOM algorithm reduced the activity space to a 6×6 neuron size map (Fig. 2). The weights were randomly initialized and the training was performed for a period of 200

epochs in an unsupervised manner. The resulting Kohonen's map was produced indicating the most frequent occupation, as shown in Fig. 2. Closely related FDs, based on the calculated descriptors, were grouped in the same basin with small interneuron distances colored in yellow. As shown in Fig. 2, most FDs with similar properties were projected into the same areas of the map. The distinctly diverse FDs were significantly far away from the main group and were separated by large interneuron distances. For example, neurons [2,3,5] consist of a higher number of FDs though they are considerably different from each other. The remaining neurons included fewer compounds (six neurons contain only 1 FD each) and were separated from the main two neurons by large interneuron distances, thus confirming that an extremely diverse group of FDs has been used for this study.

2.4. Dataset splitting

The data set was split into the training and test subsets by random selection considering three rules: (a) the range of the response values of both subsets should be covered from the lowest to the highest value; (b) the highest and lowest values of response were included in the training set, the response of the test set was similar to that of the training set, and (c) around 25% molecule contribution of C₆₀ and C₇₀ FDs separately were included in the test set. Therefore, out of 52C₆₀ FDs and 7C₇₀ FDs, 13 and 2 compounds, respectively were randomly selected as the test compounds, following these rules. Therefore, the training and test subsets consisted of 44 and 15 compounds, providing a ratio of around 75% and 25%, respectively. The test compounds were marked and listed in Table 1. The uniformity in the distribution of compounds is graphically shown in the principle component analysis (PCA) score plot (Fig. S1 in Supplementary material section). The plot shows that the testing molecules lie within the domain of the training subset, ensuring that the test subset thus selected captures all of the essential features of the entire dataset and all the 59 compounds are located in the close vicinity of each other in the 3D space.

2.5. Statistical model development and validation metrics

In the present work, we have built a regression-based QSPR model to quantify the contributions of the structural attributes and physicochemical properties to PCE values. We employed a genetic algorithm (GA) technique [29] as the selection statistical tool implemented in the *Genetic Algorithm 1.4 software package* [28]. Next, multiple linear regression (MLR) analysis was performed by *MLR Plus Validation GUI 1.2 software* [28], using the training set compounds to develop the QSPR model, which was followed by validation of the model using the test set compounds. We used the following steering parameters for the GA algorithm: total number of iterations 100, cross-over probability 1, mutation probability 0.5 and smoothing parameter (LOF calculation) 1.

Statistical metrics were computed to check the fitness of the QSPR model. Different internal, external and overall validation strategies were subsequently employed for model validation. The goodness-of-fit of the equations was judged by the quality metric R², as well as the internal validation metric leave-one-out cross-validation parameter Q_{LOO}² and external validation metric R_{pred}² or Q_{ext(F1)}². The calculation of the r_m² metrics for the test set data (r_m²(test), Δr_m²(test)) additionally estimated the closeness between the values of the predicted and the corresponding observed activity data of the testing set. It has been shown that for an acceptable model, the value of Δr_m²(test) should be lower than 0.2, provided that the value of r_m²(test) is more than 0.5 [30]. Similarly, r_m²(LOO) and Δr_m²(LOO) parameters were used for the training set and, r_m²(overall) and Δr_m²(overall) were used for the overall set [30].

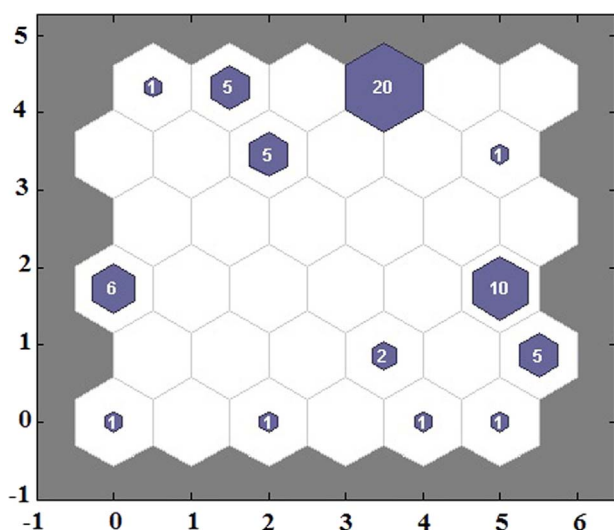


Fig. 2. Self-Organizing Map (SOM) of 169 FDs based on the computed properties.

Table 2

Obtained statistical data from the developed QSPR model.

Validation	Metrics	Value	Threshold
Internal	N _{Training}	44	–
	R ²	0.76	> 0.5
	R _{adjusted} ²	0.70	> 0.5
	SEE	0.69	–
	F	13.398	–
		(DF:8, 35)	
	Q _{LOO} ²	0.68	> 0.5
	PRESS	22.87	–
	SDEP	0.72	–
	r _{m(LOO)Scaled} ²	0.57	> 0.5
External	Δr _{m(LOO)Scaled} ²	0.10	< 0.2
	N _{Test}	15	–
	RMSEP	0.61	–
	Q _{F1} ² or R _{pred} ²	0.70	> 0.5
	Q _{F2} ²	0.70	> 0.5
	r _{m(test)Scaled} ²	0.60	> 0.5
	Δr _{m(test)Scaled} ²	0.14	< 0.2
Overall	r _{m(overall)Scaled} ²	0.58	> 0.5
	Δr _{m(overall)Scaled} ²	0.11	< 0.2
Golbraikh and Tropsha's criteria	r ²	0.70	> 0.5
	r ₀ ² – r ₀ ²	0.06	< 0.3
	r ₀ ² – r ₀ ²	0.003	Any of them must be < 0.1
	r ₀ ² – r ₀ ²	0.09	
	k	1.01	0.85 ≤ k or k' ≤ 1.15
	k'	0.91	

The models were also subjected to additional external validation parameters, including the Q_{ext(F2)}² [31] and Golbraikh and Tropsha's [32] criteria to check the model reliability.

2.6. Y-randomization test

The robustness of the QSPR model was checked based on the Y-randomization technique. For a robust model, the determination

coefficient (R²) of the non-random model should exceed the squared average correlation coefficient of the randomized models (R_r²). The model randomization was performed 100 times via shuffling the dependent variables while maintaining the original independent variables. The average R² of 100 random models was computed and defined as R_r² followed by calculation of the 'R_p² parameter [33] that penalizes model R² for small differences in the values of R² and R_r².

$${}^cR_p^2 = R \times \sqrt{R^2 - R_r^2} \quad (1)$$

For an acceptable model, the value of 'R_p² should be more than 0.5.

2.7. Applicability domain test

According to the OECD principle #3, a QSPR model should have a defined domain of applicability (AD). Technically, AD represents the chemical space defined by the structural information of the chemicals used in model development, i.e., the training set compounds in a QSPR analysis. Here, we have tried two different approaches to assess the important issue of AD. The AD of the model was checked employing (a) the leverage approach [34] and (b) the Euclidean distance approach [35].

3. Results and discussion

3.1. Computational results

Based on the experimental PCE data and selected structural features (descriptors), we have developed a statistically significant QSPR model, employing a hybrid GA-MLR as the modeling method. The developed equation is

$$\begin{aligned} \text{PCE}(\%) = & 2.839(\pm 0.448) + 1.731(\pm 0.339) \times D1 \\ & - 0.814(\pm 0.151) \times D2 \\ & - 0.057(\pm 0.014) \times D3 - 0.806(\pm 0.192) \times D4 \\ & - 0.129(\pm 0.029) \times D5 \\ & + 0.488(\pm 0.140) \times D6 - 0.009(\pm 0.004) \times D7 \\ & - 0.081(\pm 0.050) \times D8 \end{aligned} \quad (2)$$

Extensive explanation of descriptors D1 to D8 are provided in the Section 3.2. The statistical data presented in Table 2 support that our QSPR model is well fitted and robust to reliably predict the PCE of untested compounds. Eq. (2) involves 8 descriptors and explains 76.0% of the variance. Eq. (2) was obtained in compliance with OECD principles, although the complete dataset is extremely diverse in terms of structural similarity of the molecules, satisfactory measures of goodness-of-fit, robustness and predictability. Moreover, least possible deviation of the predicted activity data from the corresponding observed ones is further implied from the satisfactory values of all the r_m² metrics. Identical values for the Q_{F1}² (0.70) and Q_{F2}² (0.70) metrics indicate that the test and training sets, selected for development of the QSPR model, have similar response distributions. External predictability was further assessed according to the Golbraikh and Tropsha's criteria, which are highly satisfactory. Both the very good fit and high predictability are confirmed by a scatter plot of experimentally determined (observed) versus predicted (employing QSPR model) PCE values of FDs (Fig. 3).

To avoid the possibility of "correlation-by-chance" and to confirm the statistical significance of the developed QSPR model, we have additionally performed a Y-randomization technique. We

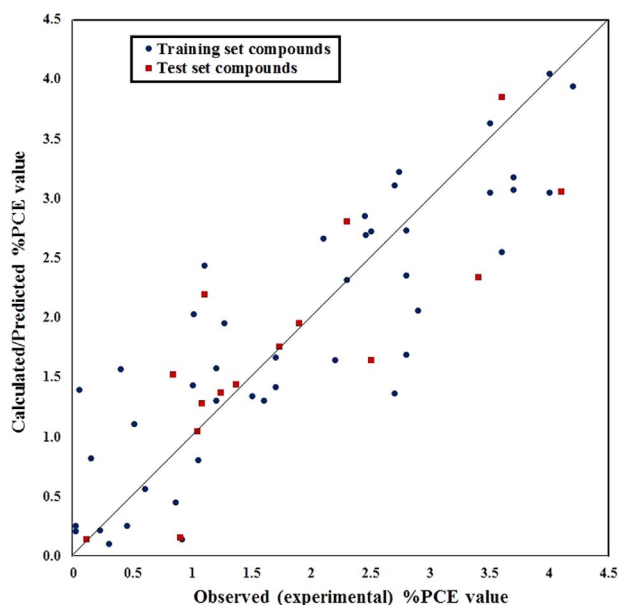


Fig. 3. Scatter plot of experimentally determined (observed) versus predicted PCE values of FDs.

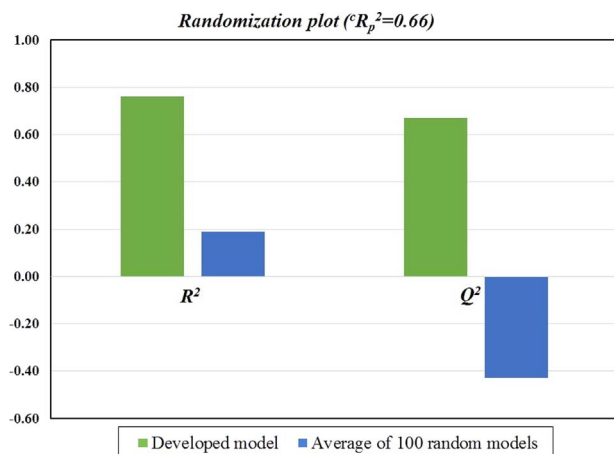


Fig. 4. Randomization plot of the developed GA-MLR model.

have built 100 random models using the same set of descriptors, however, in each of these models, the descriptors were correlated with randomly shuffled response values (i.e. PCE value of the FDs). The average R^2 and Q^2 of 100 random models are 0.19 (R_r^2) and -0.43 (Q_r^2); and the R_p^2 parameter is 0.66 which is significantly higher than the stipulated value of 0.5 (Fig. 4). All the numerical values strongly suggest that the obtained model is not derived by chance.

The leverage approach showed a critical HAT diagonal value (h^*) of 0.614. We have used a $\pm 3\sigma$ standard deviation unit of the predicted residual values for defining the domain of applicability of the model. All training set compounds bear values of standardized residue within the limit of $\pm 3\sigma$, indicating that there was no prediction outlier in the training set. However, two FDs of the training set have greater leverage values than the critical value of 0.614 ($h > h^*$). Therefore, these compounds behave as influential observations (X outlier), although they are not response outliers (not Y outlier). All testing compounds were found to be within the AD of the model. Additionally, in the case of the Euclidean distance validation approach (i.e., employing a Euclidean distance measure), all compounds were found to lie within the domain defined by the training set. Considering both of the AD tests, we conclude

that all test compounds were inside of the AD and their predictions are completely reliable. Therefore, we can confidently predict 100% of test compounds based on the developed model after rigorous testing for validation and AD. The standardized cross-validated residuals versus leverage values (Williams plot) and Euclidean distance plot are provided in Fig. 5(a) and (b), respectively.

3.2. Interpretation of the developed model

The descriptors appearing in Eq. (2) are listed in descending order of significance, based on their standardized coefficients (Fig. 6). To comply with the OECD Principle 5, mechanistic interpretation should be given for any predictive QSPR model. Here we provide the interpretation, with suitable examples, and justify the importance of each descriptor appearing in the GA-MLR equation:

[D1] The $S_A(chg)/A_D_D_D/1_2s, 1_3s, 3_4a/6$ descriptor is related to the positive effects on the PCE value contributed by the partial-charge. This descriptor represents a four-atomic fragment labeled by partial charges. In this case, partial charges are induced by *-ortho* directing groups in the substituted benzene rings. The specific types of fragments on example molecules are presented below (blue and red circles). As indicated in Box 1 of Fig. 7, compounds numbered 31, 33 and 35 have fragment type (a) and 26 has another form of fragment type (b). Both are *-ortho* directing groups in the substituted benzene ring, so the value of this descriptor for the mentioned molecules would be 2. In the same way, compound 28 has two (c) type fragments, so its value will be double i.e. 4. As mentioned earlier, the feature has a positive contribution to the PCE value, and as expected, the experimental PCE value of compound 28 is greater than the values of compounds 26, 31, 33, and 35.

[D2] The second significant descriptor $Fr5(chg)/B_C_C_C_D/1_4s, 2_3s, 2_4s, 3_4s/$ also defines partial charges and has negative impact on the modeled equation. Such a descriptor, based on partial charges, is represented by the following molecular fragment (d) in Box 2 of Fig. 7, where, A is an aromatic ring, R_1 an alkane substituent, and the cross line shows the bond attaching the fullerene.

[D3] Feature $Fr5(type)/C_3_C_3_C_3_C_3_H/1_2s, 2_3s, 3_4s, 4_5s/$ defines the presence of specific types of atoms (here, saturated carbon chains like $[C(sp^3)-C(sp^3)-C(sp^3)-C(sp^3)-H]$) in a particular compound. It has a negative effect on the PCE values. For instance, the value of this descriptor is the highest for the compound 28, which has three long saturated carbon chains ($-C_8H_{17}$) attached to the benzene ring. Compounds 31, 33, and 35 have two long saturated carbon chains ($-C_8H_{17}$), generating higher values for this descriptor leading to lower PCE values. On the contrary, lower values of this descriptor for compounds 22 and 25 lacking saturated side chains, generate higher PCE values.

[D4] The $Fr5(att)/C_C_E_E_E/1_3s, 2_4s, 3_5a, 4_5a/$ descriptor is related to the van der Waals attraction between 3 substituents in benzene rings, when those are located close to each other. It also has a negative impact on the PCE. Therefore, the van-der-Waals attraction between three or more *-ortho* substituents in benzene ring decreases the PCE value. This descriptor represents a five-atomic fragment $[X-C(benz)-C(benz)-C(benz)-X]$, as shown in Box 3 of Fig. 7 (fragments (e) and (f)). Fragment type (e) is found in compounds 24 ($X=CH_3$) and 28 ($X=C_8H_{17}$), with descriptor values of 3, and 1, respectively, while fragment type (f) is found in compound 26, with a descriptor value of 6. Due to the negative impact of this feature, compound 26 has a very poor PCE value of 0.6. Compounds 24 and 28 also showed lower PCE values of 1.2 and 1.5, respectively, but higher than the PCE value of compound 26, due to the lower value of this descriptor.

[D5] $Fr5(type)/C_3_C_3_C_C_AR_C_AR_C_AR/1_4s, 2_3s, 2_5s, 4_5a/$ is an atom type SiRMS-based descriptor that has a negative effect on the PCE.

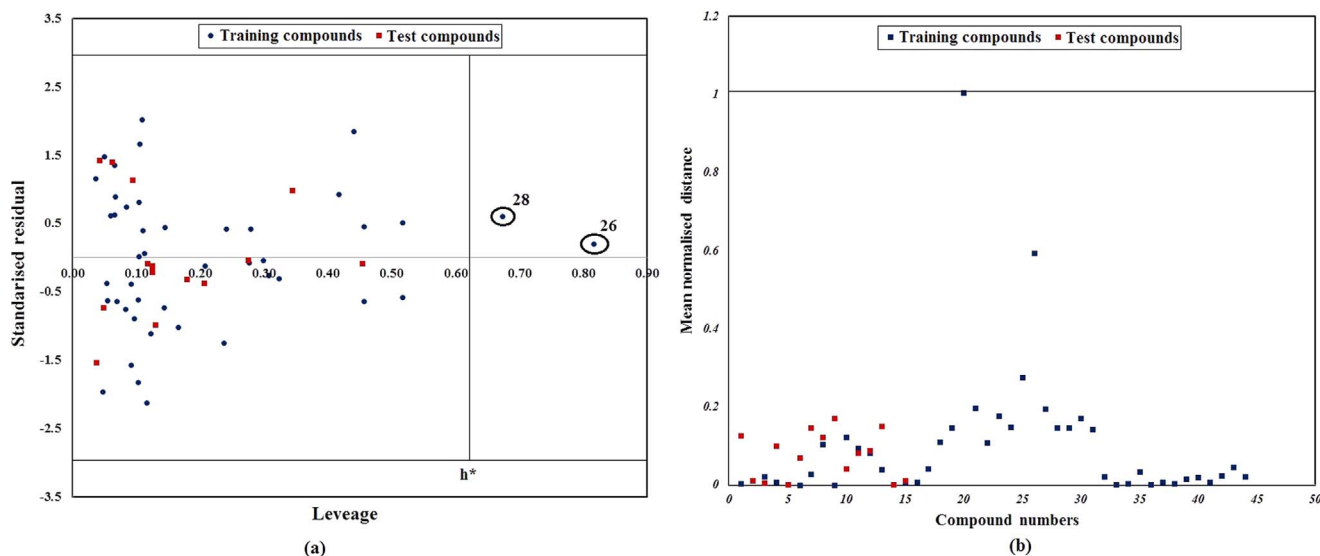


Fig. 5. (a) Williams plot and (b) Euclidean distance plot of the developed QSPR model.

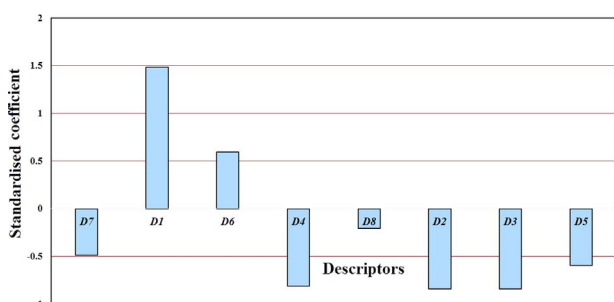


Fig. 6. Standardized coefficients plot for the modeled descriptors.

This descriptor reflects substituents attached near a pentagon of the fullerene core. The feature is mainly present in penta(organo) C_{60} fullerene structures, which are chemically and electrochemically more stable than two-point substituted FDs. Due to the large number of attachments in the fullerene core itself, aromaticity and unsaturation properties decrease, reducing the acceptor property of FDs. Compounds 37–41, for example, have this particular feature which results in a high value for the D5 descriptor. Due to its negative impact, it reduces the PCE values. Therefore, it is advisable to avoid this particular type of FD in polymer solar cell acceptors.

[D6] $S_A(type)/C.3_C.3_C.3_C. AR/1_3s, 2_3s_3_4s/5$ is another atom type SiRMS descriptor with a positive influence on the PCE value. Descriptor **D6** defines two types of fullerene substituent linkages (Box 4 of Fig. 7), including fragment type (g) which corresponds to the presence of aromatic rings (X=phenyl, pyrrole, thiophene) attached via a linker to the fullerene, and fragment type (h) corresponds to the presence of only a phenyl ring fused to fullerene via a linker. It is important to mention that the value of the descriptor for case (g) is lower than that for case (h). Due to the positive impact of these fragments, FDs bearing these fragments show higher PCE values. FDs such as 8, 11, 18 and 20 have type (g) fragments leading to higher PCE values: 3.7, 3.4, 3.5, and 3.6 respectively. FDs such as 22, 23 and 25 have type (h) fragments and generate higher PCE values: 4.2, 3.6, and 3.5, respectively.

[D7] ASA_P defines the solvent accessible surface area of polar atoms ($|q| \geq 0.125$), where $|q|$ is the absolute value of the partial charge of the atom. The accessible solvent surface area is important when calculating free energy changes due to transferring

the molecule from a polar to a non-polar solvent during formation of PSCs with BHJ layers. The D7 descriptor has a negative effect on the outcome, so increasing its value results in a reduction of PCE.

[D8] $Fr5(chg)/B_B_C_C/1_4S, 1_5s, 2_3a, 4_5s/$ is related to partial charges and has a negative impact on the PCE values. The D8 descriptor represents the transmission of charge through a chain of saturated carbons induced by a $C=O$ group. On the contrary, when the carbonyl fragment is attached to an aromatic ring, it can induce a mesomeric effect that is stronger than an inductive effect. Thus, in the case of compounds 15, and 17, for example, the $C=O$ group in the saturated carbon chain produces an inductive effect and leads to lower PCE values. On the contrary, when the $C=O$ group is connected to an aromatic ring, the mesomeric effect leads to higher PCE values, as in compounds 22 and 25.

Based on developed model one can conclude that results are in agreement with modern theories of optical excitation of a polymer-fullerene solar cell [36,37]. Summarizing the overall interpretation of descriptors, we conclude that the descriptor (**D1**), related to partial charge, and the descriptor (**D4**) related to van der Waals reflect the impact of functional groups in aromatic rings. In particular, *-ortho* oriented substitution imposes a positive impact on the PCE value, while van der Waals attraction between three or more *-ortho* substituents decreases the PCE. Aromatic rings are known electron acceptors. Optical excitation is related to the charge-transfer transition at the donor-acceptor interface. It leads to generation of charge carriers. Thus, the mesomeric transmission of charge through the aromatic ring plays an important role in power conversion.

However, as it was previously demonstrated, the photoinduced electron transfer between polymer and fullerene is not essential for the charge carrier generation [37]. The delocalisation of charge carriers over a fullerene may be important in driving charge separation. Intramolecular (descriptors **D1** and **D8**) or intermolecular delocalisation of charge may result in a higher effective electron-hole separation, and reduced Coulomb binding in the charge transfer state [36]. In contrast, the transmission of charge induced by a $C=O$ group (**D8**) through saturated carbon chains is related to a typical inductive effect. A similar effect is observed in the molecular fragment assigned to descriptor **D2**. Based on the developed model, both features (**D2** and **D8**) have a negative impact on the PCE.

Considering the influence of simple atomic descriptors (**D3**, **D5** and **D6**), one concludes that aromatic rings (phenyl, pyrrole, thiophene) (**D6**) attached to the fullerene by a linker increase the

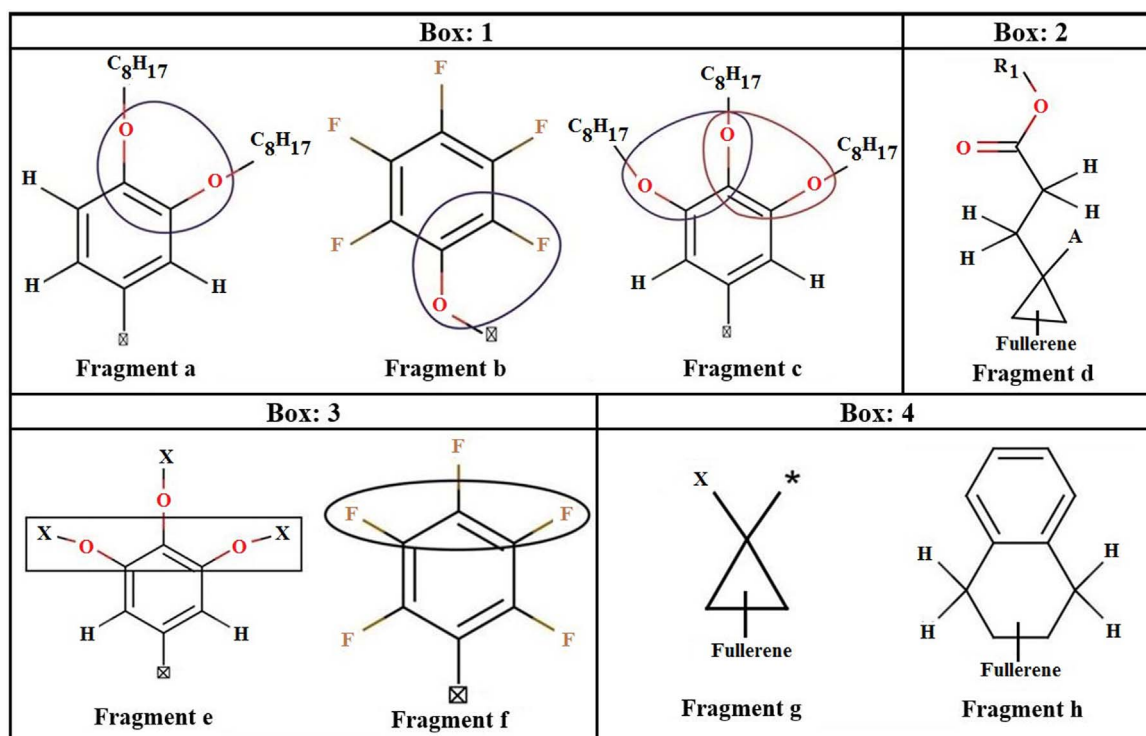


Fig. 7. Important structural fragments affecting the % PCE values.

electron withdrawing properties of the fullerene (as aromatic rings are also electron acceptors) and have a positive effect on the PCE value. It could be related to mechanism, when electronic state is formed by “hybridizing” a P3HT π^* state with a unoccupied state of the functional groups [38]. At the same time, saturated carbon chains (**D3**) introduce inductive effects and quench the mesomeric effects of aromatic rings. Attachment of substituents near one pentagon of the fullerene core (**D5**) can also deteriorate electron-accepting properties and decrease the PCE.

4. Screening of external datasets

4.1. External dataset collection

We have generated a set of 169 FDs from extensive literature review (structures and references are given in the [Supplementary material section, Table S1](#)) as an external data set of ligands. Structures were prepared and the modeled descriptors were calculated following the schemes mentioned in [Sections 2.2.1](#) and [2.2.2](#), respectively.

4.2. Screening result

Our QSPR equation was applied to predict the % PCE values of the complete set of 169 FDs. The major aim was to identify compounds with better PCE than those of the compounds used for development of the model. Here, we must mention that the resulting predictions for such a large number of FDs may not be sufficiently accurate for all FDs. Therefore, we have verified those results by performing AD studies based on Euclidean distance and leverage methods to check whether the FDs fall under the AD of the modeled compounds.

Out of 169 FDs, 10 were found to be out of the AD (including compounds 18, 41, 54, 59, 60, 67, 77, 107, 162, and 163). The leverage values larger than the stipulated HAT (h^*) value of 0.614 from the developed model ([Fig. 8a](#)). In addition, three FDs

(compound numbers 124, 162, and 163) were identified to be out of the AD by the Euclidean distance method ([Fig. 8b](#)). Therefore, combining the results of both methods, we can conclude that, except for 11 FDs (2 FDs are common in both cases), predictions for the remaining 158 FDs are completely reliable.

In our main dataset, compound22 has the highest PCE value (4.2). Interestingly, employing the QSPR equation, we have identified twelve FDs from the true external set for which the predicted PCE is higher than 4.2. Out of these twelve compounds, nine are within the AD of the main dataset. Three compounds (numbers 18, 41, and 77) are outside of the AD and their predictions (4.39, 12.11 and 4.95) are not reliable. Therefore, we suggest that the remaining nine FDs (compounds 17, 19, 20, 28, 40, 45, 71, 73, and 126, with predicted PCE values of 4.23, 5.99, 5.53, 4.23, 4.93, 6.29, 4.62, 4.29, and 4.22, respectively) should be tested as *lead* molecule experimentally as polymer solar cell acceptors ([Fig. 9](#)).

Predicted % PCE values are provided in the [Supplementary material section \(Table S1\)](#). Final descriptors for all modeled and screened compounds are included in the excel file of [Supplementary material](#). It is worth mentioning that since the developed model was based on a large number of C_{60} and C_{70} FDs, it can be applied for prediction of PCE value for untested or newly developed FD that could serve as a potential polymer solar cell acceptor.

5. Conclusions

We have introduced a QSPR model that allowed us to identify fragments and structural features of FD acceptors most responsible for high PCE of polymer solar cells (PSCs) with a bulk-heterojunction (BHJ) active layer (where P3HT is the donor). As there is no complete theory of charge generation at organic heterojunctions yet, different parts of the problem have been addressed. Developed model is useful tool for rationalization of experimental conditions.

Additionally, virtual screening of a large FD database has enabled us to identify nine lead FDs with high PCE values, suggesting

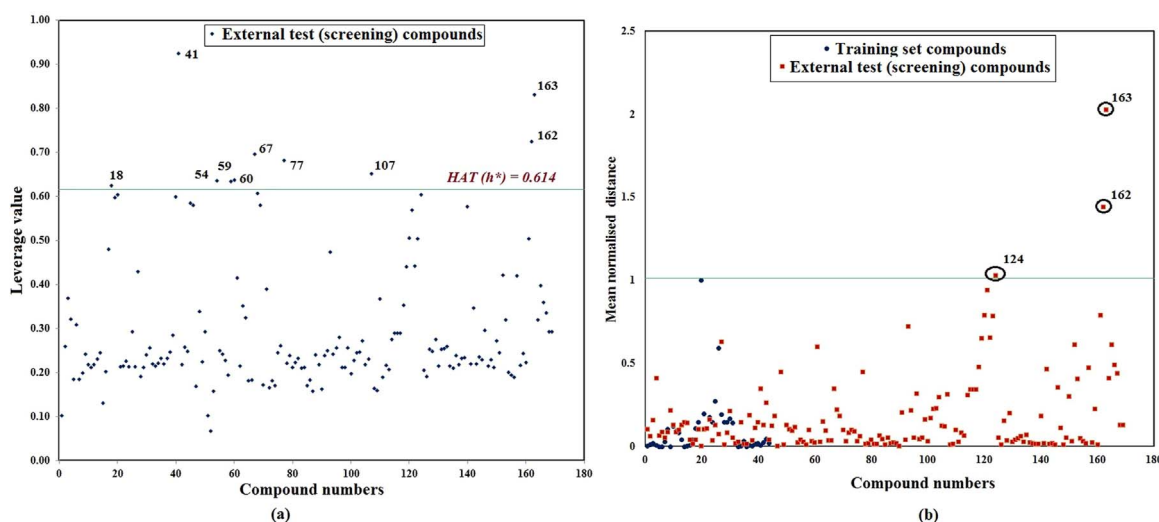


Fig. 8. (a) Leverage plot and (b) Euclidean distance plot for the checking of applicability domain of the screened compounds.

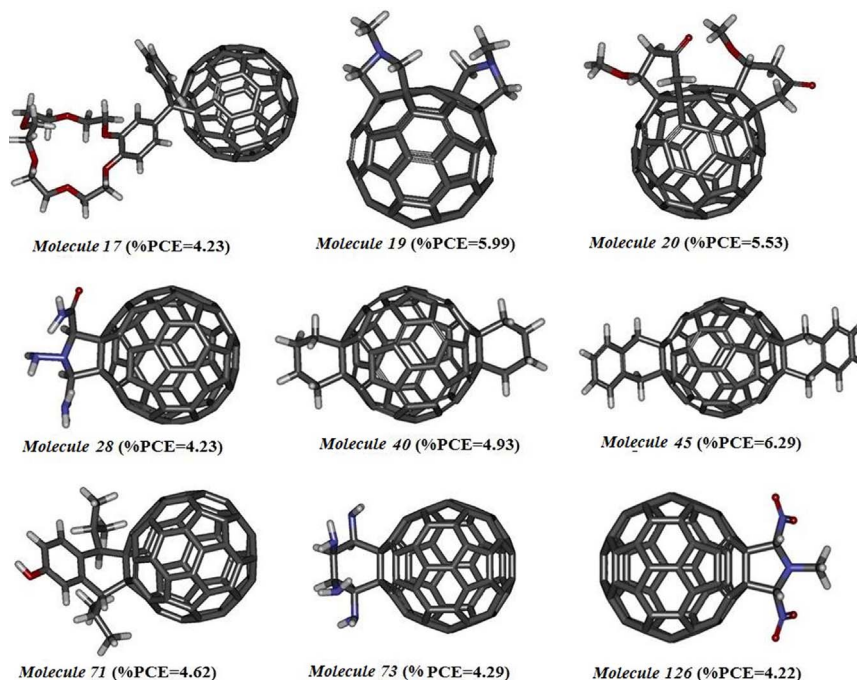


Fig. 9. Structure of 9 FDs as 'lead molecule' identified from true external set screening for power conversion efficient acceptor for PSCs.

that further computational and experimental studies of these molecules are worthwhile pursuing. Among the nine lead FDs, the most promising FD shows PCE value of 6.29, predicting a 50% increase in PCE when compared to current FDs for which the highest PCE value has been 4.2 – no doubt an encouraging outcome. At the same time, it is important to mention that one of the screened FDs (compound 41) that failed the AD study, and therefore was not included as a projected lead FD, has actually shown the most promising PCE value of 12.11, which would provide a 200% increase in PCE compared to existing FDs. Although such an FD failed the AD test, based on the acceptable QSPR prediction criteria, the large predicted PCE should motivate experimental studies, along with the projected nine lead FDs.

Our findings can be summarized, as follows:

- Our QSPR model enables identification of the essential structural attributes necessary for quantifying the prime molecular prerequisites of diverse FDs, including molecular requirements for high PCE of PSC acceptors in composition with P3HT. The identified structural fragments could guide the design and synthesis of more efficient FDs.
- Our QSPR model, developed from a set of 59 diverse FDs, is an efficient tool to screen a wide range of C_{60} and C_{70} FDs, allowing for identification of FDs with high PCE in a time and cost effective way. Using the QSPR model developed in this study to virtually screening a set of 169 FDs, we have already identified nine FDs with promising PCE values. We suggest the FDs identified here for further experimental investigations as PSC acceptors.
- Our results showed that SiRMS-based descriptors can be used efficiently for QSPR modeling.
- The developed QSPR model is particularly valuable to predict and characterize the nature of donor-acceptor relationships critical for photoconversion.

Our calculations provide a set of data for experimentalists to reduce the experimental efforts, time and resources by many folds. In addition, the exploratory features may assist in designing more efficient units. In our future work, we are working to calculate the electronic structure and molecular dynamics aspects of all the screened compounds at P3HT/FD bulk heterojunction systems with emphasis on electron transfer rates and the experimental analysis that could further validate our QSPR model and the virtual screening approach.

Acknowledgements

S.K., N.S. and J.L. thank the National Science Foundation (NSF/CREST HRD-1547754) and EPSCoR (Award #: 362492-190200-01/NSFEP-0903787) for financial support. V.S.B acknowledges support from the Argonne-Northwestern Solar Energy Research (ANSER) Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0001059 and computing facilities from NERSC. The authors are grateful to the Mississippi Center for Supercomputing Research (MCSR) for providing state-of-the-art high performance computing facilities and excellent services for supporting this research. We thank Prof. Kuz'min V.E. for providing copy of software generating SiRMS-based descriptors.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.nanoen.2016.06.011>.

References

- [1] European Commission, Regulation (EC) (No. 1907/2006) of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency, 2006.
- [2] H. Hoppe, N.S. Sariciftci, *Photoresponsive Polymers II*, in: S.R. Marder, K.-S. Lee (Eds.), *Advances in Polymer Science*, 214, Springer-Verlag, Berlin, 2008, pp. 1–86.
- [3] B.C. Thompson, J.M.J. Frechet, *Angew. Chem. Int. Ed.* 47 (2008) 58–77.
- [4] G. Li, V. Shrotriya, J. Huang, Y. Yao, T. Moriarty, K. Emery, Y. Yang, *Nat. Mater.* 4 (2005) 864–868.
- [5] National Renewable Energy Lab and U.S. Department of Energy, Sunshot Vision Study, Feb. 2012, pp. 4–5 (<http://energy.gov/eere/sunshot/sunshot-vision-study>).
- [6] G. Dennler, M.C. Scharber, C.J. Brabec, *Adv. Mater.* 21 (2009) 1323–1338.
- [7] J.L. Delgado, P.A. Bouit, S. Filippone, M.A. Herranzand, N. Martin, *Chem. Commun.* 46 (2010) 4853–4865.
- [8] Y. He, Y. Li, *Phys. Chem. Chem. Phys.* 13 (2011) 1970–1983.
- [9] Y.-J. Cheng, S.-H. Yang, C.-S. Hsu, *Chem. Rev.* 109 (2009) 5868–5923.
- [10] G. Yu, J. Gao, J.C. Hummelen, F. Wudl, A.J. Heeger, *Science* 270 (1995) 1789–1791.
- [11] M. Seri, E. Rossi, T. Carofiglio, S. Antonello, G. Ruani, M. Magginiand, M. Muccini, *J. Mater. Chem.* 21 (2011) 18308–18316.
- [12] W. Ma, C. Yang, X. Gong, K. Lee, A.J. Heeger, *Adv. Funct. Mater.* 15 (2005) 1617.
- [13] K. Roy, S. Kar, R.N. Das, *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, USA 2015, pp. 1–478.
- [14] E. Pourbasheer, A. Banaei, R. Aalizadeh, M.R. Ganjali, P. Norouzi, J. Shadmanesh, C. Methenitis, *J. Ind. Eng. Chem.* 21 (2015) 1058–1067.
- [15] OECD Document, Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SARs] Models, ENV/JM/MONO(2007)2, 2007.
- [16] P.A. Troshin, H. Hoppe, J. Renz, M. Egginger, J.Y. Mayorova, A.E. Goryachev, A. S. Peregodov, R.N. Lyubovskaya, G. Gobsch, N.S. Sariciftci, V.F. Razumov, *Adv. Funct. Mater.* 19 (2009) 779–788.
- [17] J. Liu, X. Guo, Y. Qin, S. Liang, Z.-X. Guo, Y. Li, *J. Mater. Chem.* 22 (2012) 1758–1761.
- [18] HyperChem(TM), 1115 NW 4th Street, Gainesville, Florida 32601, USA.
- [19] Y. Zhao, D. Truhlar, *Theor. Chem. Acc.* 120 (2008) 215–241.
- [20] V.A. Rassolov, M.A. Ratner, J.A. Pople, P.C. Redfern, L.A. Curtiss, *J. Comput. Chem.* 22 (2001) 976–984.
- [21] Gaussian 09: M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian Inc., Wallingford CT, 2009.
- [22] X. Li, M.J. Frisch, *J. Chem. Theor. Comput.* 2 (2006) 835–839.
- [23] L. Ahmed, B. Rasulev, M. Turabekova, D. Leszczynska, J. Leszczynski, *Org. Biomol. Chem.* 11 (2013) 5798–5808.
- [24] Maestro, version 10.2, Schrödinger, LLC, New York, NY, 2015.
- [25] (<https://www.chemaxon.com>).
- [26] (<http://qsar4u.com/pages/sirms.php>).
- [27] A.G. Artemenko, E.N. Muratov, V.E. Kuz'min, N.N. Muratov, E.V. Varlamova, A. V. Kuz'mina, L.G. Gorb, A. Golius, F.C. Hill, J. Leszczynski, *SAR QSAR Environ. Res.* 22 (2011) 575–601.
- [28] (http://teqip.jdvu.ac.in/QSAR_Tools/).
- [29] C.L. Waller, M.P. Bradley, *J. Chem. Inf. Comput. Sci.* 39 (1999) 345–355.
- [30] K. Roy, I. Mitra, S. Kar, P. Ojha, R.N. Das, H. Kabir, *J. Chem. Inf. Model.* 52 (2012) 396–408.
- [31] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang, R. Kühne, *J. Chem. Inf. Model.* 48 (2008) 2140–2145.
- [32] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20 (2002) 269–276.
- [33] S. Kar, A. Gajewicz, T. Puzyn, K. Roy, *Toxicol. In Vitro* 28 (2014) 600–606.
- [34] P. Gramatica, *QSAR Comb. Sci.* 26 (2007) 694–701.
- [35] EUCLIDEAN (a program written in Java) is developed and validated on known data sets by Pravin Ambure (Email: ambure.pharmait@gmail.com) of Drug Theoretics and Cheminformatics Laboratory, Jadavpur University, 2013.
- [36] S. Few, J.M. Frostab, J. Nelson, *Phys. Chem. Chem. Phys.* 17 (2015) 2311–2325.
- [37] T.G.J. van der Hofstad, D. Di Nuzzo, M. van den Berg, R.A.J. Janssen, S.C. J. Meskers, *Adv. Energy Mater.* 2 (2012) 1095–1099.
- [38] Y. Kanai, J.C. Grossman, *Nano Lett.* 7 (2007) 1967–1972.



Dr. Supratik Kar is a post-doctoral research associate in Interdisciplinary Center for Nanotoxicity at Jackson State University, Mississippi, USA. He has completed his B.Pharm. (2008) and M.Pharm. (2010) degree from Jadavpur University, India securing first position in both degrees. He has earned his PhD (2015) from the Jadavpur University, India. Former visiting researcher at the University of Gdańsk (Poland). Has experience in QSAR and chemometric modeling studies for seven years. Author of 32 research and review articles, 4 book chapters till date. He has also coauthored 2 QSAR related books in Academic press and Springer. His current h-index is 15 (Scopus).



Mrs. Natalia Sizochenko is a research associate at the Jackson State University and a PhD candidate (expected Spring 2016) at the University of Gdansk, Poland. She received her B.Sc. and M.Sc. degrees in Chemistry from the Odessa I.I. Mechnikov National University, Ukraine, in 2010 and 2011, respectively. Her areas of research are Computational chemistry, Chemoinformatics and Drug Design. The bulk of her research interests include development of molecular descriptors; implementing machine learning and causal inference algorithms for databases with incomplete data; QSAR analysis; pharmacophore search; molecular docking and molecular dynamics studies.



Dr. Lucky Ahmed is a postdoctoral associate in the Department of Chemistry at Yale University, Connecticut, USA. She has received her PhD in Chemistry from Jackson State University, Mississippi, USA in 2015. She was awarded her M.S and B.Sc. degrees in Chemistry from the University of Dhaka, Bangladesh in 2007 and 2005 respectively. Her areas of research focus on quantum chemical calculations, quantum mechanics/molecular mechanics (QM/MM) simulation, molecular docking, molecular dynamics simulation, quantitative structure–activity relationship (QSAR) analysis, and Drug Design.



Prof. Victor S. Batista received his Lic. Ciencias Químicas degree from Universidad de Buenos Aires, Argentina (1989), and the Sugata Ray Award (1995) and a PhD degree in Theoretical Chemistry (1996) from Boston University. After completing postdoctoral programs with William H. Miller at the University of California, Berkeley (1997–1999) and Paul Brumer at the University of Toronto (2000), he joined the Yale faculty in 2001, where he has received the ACS PRF-G6 Award (2002), the Research Corporation Innovation Award (2002), the NSF Career Award (2004), the Sloan Fellowship (2005–2006) and the Camille Dreyfus Teacher-Scholar Award (2005).



Prof. Jerzy Leszczynski, Professor of Chemistry and President's Distinguished Fellow at the Jackson State University, directs Interdisciplinary Nanotoxicity NSF CREST Center. Dr. Leszczynski is a computational quantum chemist whose vast areas of interest include nature of chemical bonds, theoretical predictions of molecular potential energy surfaces and vibrational spectra, structures and properties of molecules with heavy elements, properties and structure of DNA fragments, and characteristics of nanomaterials. He has published over 800 referred papers and 60 book chapters. Dr. Leszczynski is editor of a book series: Computational Chemistry: Reviews of Current Trends (World Scientific); editor of a book series Challenges and Advances in Computational Chemistry and Physics (Springer); editor of Handbook of Computational Chemistry (Springer); editor of book series Practical Aspects of Computational Chemistry (Springer); and series editor for Lecture Notes in Chemistry (Springer); and editor and member of editorial boards of eight journals.