Y.Z. Akbasheva[1]
Y.N. Biglova[1]
V.R. Khayrullina[1]
A.Ja. Gerchikov[1]

QSPR - MODELING OF FULLERENE DERIVATIVES USING GUSAR: EXPLORING STRUCTURAL ATTRIBUTES CRITICAL FOR PHOTOCONVERSION EFFICIENCY OF POLYMER SOLAR CELL ACCEPTORS

[1]Department of Chemistry, Bashkir State University, Zaki Validi Street, 32, Ufa, Russia;
yulliietta@gmail.com

A quantitative analysis of relationships between structures of fullerene derivative and efficiency of transformation of solar energy to electrical (PCE) was made by the program GUSAR 2013 (General Unrestricted Structure Activity Relationships) [1-2]. The experimental data from PCE [3-4] were used for creation of QSPR models. In general 33 statistically significant QSPR-models ($R_{train\ set}^2 > 0.6$, $R_{test\ set}^2 > 0.5$, $Q^2 > 0.6$) for prediction of PCE values for various monosubstituted methanofullerenes were created based on MNA- and QNA-descriptors, as well as consensus of their combinations. The statistical characteristics of some of the models we constructed are presented in the Table1. Training set TrS10 and test set TS included 66 and 14 structures of fullerene derivatives, respectively. They were obtained by dividing the pre-sorted in ascending order of PCE values in ratio 5:1, i.e. excluded from TrS1 every sixth compound to the effectiveness of transformations. These models can be used for quantitative prediction of potential polymer solar cells based on derivatives of matanofullerenes. Atoms and structural fragments of the studied structures influencing on increase and decrease of PCE were identified by GUSAR 2013 visualization of quantitative "structure-activity" relationships in the created models. The results of structural analysis of the contribution of the different functional groups in the efficiency of transformation of solar energy to electrical can be considered in the molecular design of potential polymer solar cells in order to enhance the efficiency of transformation of solar energy to electrical.

Table 1. Characteristics and prediction accuracy of PCE values for consensus models M1 – M33. PCE activity in TrS1 and TrS10 lies in the range 0.00 – 4.00.

Training set	Models	N	R^2_{OB}	R^2_{TB}	F	S.D.	Q^2	V
QSAR model based on MNA-descriptors								
TSet1	M2	80	0.783	-	12.159	0.592	0.701	10
TSet10	M29	66	0.734	0.724	7.402	0.657	0.621	9
QSAR model based on QNA-descriptors								
TSet1	M1	80	0.723	-	12.190	0.640	0.639	9
TSet10	M28	66	0.674	0.600	8.978	0.678	0.574	8
QSAR model based on MNA- and QNA-descriptors								
TSet1	M3	80	0.797	-	12.719	0.576	0.719	11
TSet10	M30	66	0.779	0.753	8.946	0.609	0.687	9

N – number of structures in the training set; R^2_{TrS} - a multiple coefficient of determination calculated for compounds from the training set; R^2_{TS} - a multiple coefficient of determination calculated for compounds from the test set; Q^2 – a cross-validated R^2 calculated during leave-one-out cross-validation procedure on data of the training set; F – Fisher's coefficient; SD – standard deviation; V- the number of variables in the final regression equation.

1. Filimonov D.A. et al. *SAR and QSAR in Environmental Research*, 2009,. **20 (**7–8): 679–709.
2. Khayrullina V.R. et al. *Biochemistry (Moscow)*, 2015, **80** (1): 74–86.
3. He Yo. Li Yo. *Physical Chemistry Chemical Physics*, 2011. **13**: 1970-1983.
4. Troshin P.A. et al. *Advanced Functional Materials*, 2009. **19**: 779-788.