

A Quantitative Structure - Permittivity Relationship Study of a Series of Polymers

Yevhenii Zhuravskiy,^{1,2} Kweeni Iduoku,² Meade Erickson,² Anas Karuth,² Durbek Usmanov,^{2,3} Gerardo Casanola-Martin,² Maqsud N. Sayfiyev,⁴ Dilshod A. Ziyayev,⁴ Zulayho Smanova,⁴ Bakhtiyor Rasulev^{2,4*}

¹ Department of Technology of Organic Products, Lviv Polytechnic National University, Lviv, Ukraine

² Department of Coatings and Polymeric Materials, North Dakota State University, Fargo, ND, United States

³ Institute of the Chemistry of Plant Substances AS RUz, Tashkent, Uzbekistan

⁴ Department of Chemistry, National University of Uzbekistan, Tashkent, 100174 Uzbekistan

*Correspondence: B.R., email: bakhtiyor.rasulev@nds.edu

Dielectric constant is a value, that is widely applied in many scientific fields, and characterizes the degree of polarization of substances under the external electric field. In this work, a structure-property relationship of dielectric constants (ϵ) for a diverse set of polymers was investigated. A transparent mechanistic model was developed with application of a genetic algorithm combined with multiple linear regression analysis (GA-MLRA), to get a mechanistically explainable and transparent model. Based on the evaluation conducted using various validation criteria, four- and eight-variable models were proposed. Obtained statistical performance results and selected descriptors in the best models were analyzed and discussed. With the validation procedures applied the models were proven to have a good predictive ability and robustness.

Keywords: dielectric permittivity, polarization, polymers, descriptors, models, validation.

Introduction

Polymeric properties related to electrical conductivity are useful in many applications, such as cable insulation [1], capsules for electrical components, interlayer dielectrics, charge-storage capacitors [2, 3], and printed circuit boards [4]. Dielectric permittivity is a value that is widely used and characterizes the degree of polarization of substances under the action of an external electric field. A larger dielectric constant means a larger polarization of the medium between two charges. Therefore, the dielectric constant is the ability of a substance to separate the charge and/or orient its molecular dipoles in an external electric field. The dielectric constant is an important basic molecular property that can also be a useful predictor of other electrical properties of polymers [4-6]. However, the exact experimental values of dielectric constant for

polymers are often unavailable. The prediction of dielectric constants computationally and by using theoretical approaches, such as machine learning predictive modeling is important in the molecular design of new polymeric materials with desired properties. The rapid and accurate implementation of predictions for a wide variety of chemical structures can significantly improve the performance and speed of phenomena investigation. However, the theoretical calculation of the property such as dielectric constant of the polymer is not an easy problem, since this property is a non-linear property and therefore a function of several factors, including polymer structure and composition, temperature, materials morphology, additives and plasticizers, impurities, and moisture in the volume of the polymer. Quantitative structure-activity relationship (QSAR) is a subsection of machine learning modeling and chemical informatics for revealing relationships

between chemical structures of molecules and their activity. QSAR modeling is a suitable approach for estimating the properties of polymers based on numerical features/descriptors derived from molecular structure to fit the experimental data [7-9]. The main idea of the QSAR approach is that the change in the desired property of a compound can be correlated with the structure-based properties that numerically expressed and called "molecular descriptors" [8-11]. In cheminformatics, molecular descriptors are numbers that formally represent a molecule, obtained by a well-defined algorithm and applied to a well-defined experimental procedure. In other words, molecular descriptor is the result of a mathematical expression that converts chemical structure to a numerical value [12]. Each molecular descriptor describes a molecular structure, by encoding a part of the structure or a whole molecular structure. Molecular descriptors play a fundamental role in the development of QSPR models. One of the main features of the QSPR approach is that it requires only knowledge of the chemical structure and is independent of any experimental properties. Once a correlation is found, it can be applied to predict the properties of new compounds that have not been synthesized or not found. Therefore, the QSPR approach can accelerate the development of new molecules and materials with the required properties. Using the QSPR approach, many different properties of polymers can be determined with sufficient accuracy, in particular, this approach is already used to determine such properties as refractive index [4, 13-21], glass transition temperature [14, 22-33], cohesive energy [34], thermal decomposition temperature [35], solubility parameter [36], as well as for fouling release properties [37]. Several QSPR models for the dielectric constants of small organic molecules have also been reported in the literature [6, 38-41]. But the number of attempts to predict the dielectric constants of the polymers was rather small [4, 42]. Liu et al. [42] introduced a model with a correlation coefficient of (R^2) 0.908 and a standard error (s) of 0.001 for 22 polyalkenes using three descriptors, but the values of ϵ in this case only cover the range from 2.154 to 2.165. Bicerano [4] developed a QSPR model with (R^2) 0.958 and (s) 0.087 to correlate ϵ with 32 topological and constitutional descriptors for 61 polymers. This model contains too many descriptors. Excessive correlation and randomness of correlations may be partly due to improved results by increasing the number of descriptors in the correlation. Moreover, the two models were not validated externally, using a

test set. In fact, validation is a crucial aspect of any QSPR/QSAR modeling [43].

The purpose of this study was to develop a reliable predictive QSPR model that could effectively be used to predict dielectric constant values, having mechanistically explainable descriptors for further design application. The model is developed using a set of 71 polymers with a large structural diversity, with further model validation applying specific validation approaches and an external set.

Materials and Methods

Data set

The experimental data (polymers 1-56) were taken from the source that published by Bicerano [4], the remaining data (polymers 57-71) from the source published by Ku and Liepins [5], at room temperature (298K). In total, the data set for this study consists of 71 polymers with a diverse structure (see Table 1). The data set contains polymers of the following types: polyvinyles, polyethylenes, polyoxides, polystyrenes, polyethers, polysulfones, polyacrylonitrile, polyamides, polyacrylates, poly-siloxanes, polyxylylenes, and polycarbonates.

Computational Details

In this work, the structures of all polymers were computationally built, optimized and used for generating structural properties / features / descriptors calculation. Due to the fact that polymers are macromolecules with a large size and wide chain length distribution, the calculation of structural descriptors based on original structural formulas was not possible using current descriptors-generating software [23, 30]. Moreover, due to the high molecular weight of the polymers, the effect of the terminal groups on the overall structure of polymer is quite small, which allows us to neglect terminal structures contribution. In this regard, the structures of repeating monomer units of investigated polymers were used to calculate the structural features/descriptors (as shown in Fig. 1) [13, 15, 22-24, 30]. We assumed that main contributing factor to the polymer property is the structure of monomer units, and therefore the molecular descriptors are calculated based on the structure of repeating monomer units [44, 45].

The molecular structures of each polymer were drawn in ChemSketch software [46]. The optimization of monomeric units, i.e. a geometry optimization and finding the minimal energy

conformation is important step and provides a real conformation of the investigated structure for further QSAR modeling. The molecular modeling is often used for optimization and properties assessment of various chemical systems [47-50]. In this work the geometry optimization was carried out using HyperChem software, applying molecular mechanics force-field MM+ [51]. The criterion for energy optimization limit was chosen as the achieved gradient of 0.01 kcal/mol. The molecular descriptors for each polymer were calculated based on minimal energy conformation using DRAGON software [52]. Dragon 6.0 allows to generate about 5000 descriptors per structure [52]. The generated descriptors include the following categories: constitutional indices, 2D and 3D matrix-based descriptors, 2D autocorrelations, topological descriptors, connectivity indexes, information indices, atom-centered fragments, 3D-MoRSE descriptors, charge-based descriptors, 0D, 2D, and 3D descriptors, molecular properties, drug-like properties and etc [53]. Descriptors with high correlation, single variables, and non-informative information were discarded based on constant value, near constant, and pair correlation criteria ($R > 0.7$).

A total of 523 descriptors of different types were selected after initial filter criteria applied. Each descriptor is representing a molecular graph invariant and describes the particular property and in overall adds to chemical diversity of the monomeric unit.

The model development was performed with the following setup to find a best model. For the genetic algorithm (GA) based variable selection step the number of generations was set to 2000 and a mutation rate of 35% was used. For the best models' selection, the population size of the final models' list was set to 20. For validation purposes, multiple methods were applied, including leave-one-out (LOO) cross validation, y-scrambling, as well as internal and external validation protocols. After validation techniques were applied, the best model was chosen based on multiple criteria: 1) high statistical performance of R^2 and Q^2 variables (including $R^2 - Q^2 < 0.3$); 2) a low number of variables in the model; 3) low cross-correlation between descriptors in the selected model; and 4) best performance of R^2 for the external validation set (test set).

Table 1. A set of experimental and predicted dielectric constants data for the polymers involved in the experiment.

№	Name	Dataset Status	Exp.	Eq. 1.		Eq. 2.	
				Pred.	Residual	Pred.	Residual
1	Poly(1,4-butadiene)	Train.	2.51	2.4104	-0.0996	2.6006	0.0906
2	Poly[oxy(2,6-dimethyl-1,4-phenylene)]	Train.	2.6	2.9651	0.3651	2.7210	0.1210
3	Bisphenol-A polycarbonate	Train.	2.9	3.0325	0.1325	2.8725	-0.0275
4	Poly(ether ether ketone)	Train.	3.2	3.0998	-0.1002	3.0820	-0.1180
5	Poly(ethylene terephthalate)	Train.	3.25	3.1291	-0.1209	3.0958	-0.1542
6	Poly(chloro-p-xylylene)	Train.	2.95	2.8054	-0.1446	2.7932	-0.1568
7	Polyacrylonitrile	Train.	4	3.6164	-0.3836	3.9567	-0.0433
8	Polystyrene	Train.	2.55	2.4631	-0.0869	2.3794	-0.1706
9	Polypropylene	Train.	2.2	2.3304	0.1304	2.3763	0.1763
10	Poly(p-xylylene)	Train.	2.65	2.4154	-0.2346	2.3772	-0.2728
11	Polyisobutylene	Train.	2.23	2.1490	-0.0810	2.2123	-0.0177
12	Poly(p-chloro styrene)	Train.	2.65	2.8016	0.1516	2.7449	0.0949
13	Poly(N-vinyl carbazole)	Train.	2.9	2.9390	0.0390	2.7868	-0.1132
14	Poly(vinyl cyclohexane)	Train.	2.25	2.3931	0.1431	2.1312	-0.1188
15	Polyisoprene	Test	2.37	2.2119	-0.1581	2.4058	0.0358
16	Poly(p-hydroxybenzoate)	Train.	3.28	3.1280	-0.1520	3.1413	-0.1387
17	Poly(vinyl butyral)	Train.	2.69	2.9227	0.2327	3.0580	0.3680
18	Poly(cyclohexyl methacrylate)	Train.	2.58	2.9625	0.3825	2.7652	0.1852
19	Poly(vinyl acetate)	Train.	3.25	2.9128	-0.3372	3.1751	-0.0749
20	Poly(ϵ -caprolactam)	Train.	3.5	3.5411	0.0411	3.4218	-0.0782
21	Poly(3,4-dichlorostyrene)	Test	2.94	2.7643	-0.1757	2.9000	-0.0400
22	Poly(hexamethylene adipamide)	Train.	3.5	3.5852	0.0852	3.5226	0.0226
23	Poly(hexamethylene sebacamide)	Test	3.2	3.5443	0.3443	3.3880	0.1880
24	Poly(isobutyl methacrylate)	Train.	2.7	2.8675	0.1675	2.7456	0.0456
25	Poly(vinyl chloride)	Train.	2.95	3.1896	0.2396	2.9759	0.0259
26	Poly(m-chloro styrene)	Train.	2.8	2.6153	-0.1847	2.8629	0.0629
27	Polychlorotrifluoroethylene	Test	2.6	2.1061	-0.4939	2.2365	-0.3635

28	Poly(ethyl methacrylate)	Train.	3	2.8124	-0.1876	2,8927	-0,1073
29	Poly(n-butyl methacrylate)	Test	2.82	2.9877	0.1677	2,9430	0,1230
30	Poly(methyl methacrylate)	Train.	3.1	2.8846	-0.2154	2,8868	-0,2132
31	Poly[2,2'-(m-phenylene)-5,5'-bibenzimidazole]	Train.	3.3	3.3864	0.0864	3,4484	0,1484
32	Polyethylene	Test	2.3	2.4908	0.1908	2,3596	0,0596

Table 1 (continued)

№	Name	Status	Exp.	Eq. 1.		Eq. 2.	
				Pred.	Residual	Pred.	Residual
33	Poly(a-vinyl naphthalene)	Test	2.6	2.4277	-0.1723	2,4579	-0,1421
34	Poly(tetramethylene terephthalate)	Train.	3.1	3.2794	0.1794	3,1749	0,0749
35	Poly[thio(p-phenylene)]	Train.	3.1	3.4506	0.3506	3,2558	0,1558
36	Poly(4-methyl-1-pentene)	Train.	2.13	2.1958	0.0658	2,2527	0,1227
37	Poly(1-butene)	Train.	2.27	2.3378	0.0678	2,4632	0,1932
38	Poly(a,a,a',a'-tetrafluoro-p-xylylene)	Train.	2.35	2.4386	0.0886	2,4448	0,0948
39	Poly(o-methyl styrene)	Train.	2.49	2.4046	-0.0854	2,4385	-0,0515
40	Poly(b-vinyl naphthalene)	Train.	2.51	2.4622	-0.0478	2,5002	-0,0098
41	Poly(a-methyl styrene)	Test	2.57	2.4189	-0.1511	2,3614	-0,2086
42	Poly[oxy(2,6-diphenyl-1,4-phenylene)]	Train.	2.8	2.8875	0.0875	2,9843	0,1843
43	Poly(vinylidene chloride)	Train.	2.85	2.9919	0.1419	2,7289	-0,1211
44	Poly(p-methoxy-o-chloro styrene)	Train.	3.08	3.0777	-0.0023	3,1591	0,0791
45	Poly(ethyl a-chloroacrylate)	Test	3.1	3.1639	0.0639	3,4555	0,3555
46	Poly(methyl a-chloroacrylate)	Train.	3.4	3.2357	-0.1643	3,4685	0,0685
47	Poly(oxy-2,2-dichloromethyltrimethylene)	Train.	3	3.1478	0.1478	3,0166	0,0166
48	Ultem 1000	Test	3.15	3.4652	0.3152	3,3747	0,2247
49	Polyoxymethylene	Train.	3.1	2.9951	-0.1049	3,0158	-0,0842
50	Poly(1,4-cyclohexylidene dimethylene terephthalate)	Train.	3	3.1045	0.1045	3,0271	0,0271
51	Poly[N,N'-(p,p'-oxydiphenylene)pyromellitimide]	Train.	3.5	3.5482	0.0482	3,5032	0,0032
52	Poly[4,4'-diphenoxy di(4-phenylene)sulfone]	Train.	3.44	3.3943	-0.0457	3,4010	-0,0390
53	Poly[4,4'-isopropylidene diphenoxy di(4-phenylene)sulfone]	Test	3.18	3.3082	0.1282	3,4309	0,2509
54	Poly[4,4'-sulfone diphenoxy di(4-phenylene)sulfone]	Train.	3.8	3.5963	-0.2037	3,6851	-0,1149
55	Poly[1,1-cyclohexane bis(4-phenyl)carbonate]	Test	2.6	3.0918	0.4918	3,0117	0,4117
56	Poly[1,1-ethane bis(4-phenyl)carbonate]	Train.	2.9	3.0499	0.1499	2,9578	0,0578
57	Poly(cellulose propionate)	Train.	3.2	3.1174	-0.0826	3,4035	0,2035
58	Poly(amide-imide)	Train.	3.32	3.4811	0.1611	3,3832	0,0632
59	Poly(diallyl phthalate)	Train.	3.57	3.2808	-0.2892	3,3366	-0,2334
60	Poly(diallyl phenyl phosphonate)	Train.	3.84	3.6409	-0.1991	3,7696	-0,0704
61	Poly(2,5-dichlorostyrene)	Train.	2.61	2.7786	0.1686	2,8859	0,2759
62	Polyfumaronitrile	Excl.	8.5				
63	Poly(methyl cellulose)	Excl.	6.8				
64	Nylon 11	Train.	3.3	3.3744	0.0744	3,3483	0,0483
65	Nylon 12	Train.	3.6	3.3392	-0.2608	3,3367	-0,2633
66	Poly(vinyl fluoride)	Excl.	8.5				
67	Poly(2-vinyl pyridine)	Excl.	4.64				
68	Poly(vinyl toluene)	Train.	2.59	2.4524	-0.1376	2,4418	-0,1482
69	Poly(vinylidene fluoride)	Excl.	8.4				
70	Poly(dichloro-p-xylylene)	Test	2.82	2.9745	0.1545	2,8109	-0,0091
71	Poly(methyl-p-xylylene)	Train.	2.48	2.3989	-0.0811	2,4105	-0,0695

Results and discussion

In this work, a dataset of 71 polymers was used to develop a quantitative structure-permittivity relationship model. For the model validation, the set was split into training and test sets consisting of 57 and 14 polymers, respectively. The splitting was

performed with care to ensure that at least one compound of each structural class in the training set was represented in the test set. After GA-MLRA computation iterations, the best models were found. After a first round of GA-MLRA it was found that five compounds are outliers, with high experimental values error. The outliers are: 62, 63, 66, 67 and 69.

After elimination of outliers the GA-MLRA iteration was repeated. The set with a total of 66 components was split into training and test sets, containing 53 and 13 polymers, respectively. In the process of finding the best model, several options were selected that best

correlate with the dielectric constant of the selected polymers. Two models with four and eight variables are proposed, the statistical characteristics of which are given in Table 2.

The following equations represent the proposed models with four (1) and eight (2) variables:

$$\varepsilon = -3.839(\pm 0.559)\text{Me} + 3.478(\pm 0.281)\text{AAC} - 0.477(\pm 0.138)\text{JGI1} - 0.616(\pm 0.115)\text{R5p} + 2.491(\pm 0.085) \quad (1)$$

$$N = 53; R^2_{\text{train}} = 0.842; s = 0.187; F = 64.124; Q^2 = 0.813; R^2_{\text{test}} = 0.715$$

$$\varepsilon = -2.487(\pm 0.489)\text{Me} + 2.285(\pm 0.289)\text{AAC} + 1.044(\pm 0.151)\text{GATS1p} - 0.345(\pm 0.107)\text{ESpm11u} - 0.343(\pm 0.154)\text{Mor22v} - 1.308(\pm 0.178)\text{RARS} + 1.048(\pm 0.225)\text{R1v} + 0.254(\pm 0.120)\text{nCt} + 2.863(\pm 0.182) \quad (2)$$

$$N = 53; R^2_{\text{train}} = 0.905; s = 0.151; F = 52.542; Q^2 = 0.865; R^2_{\text{test}} = 0.812$$

The four-variable model shows a good performance, with $R^2_{\text{train}} = 0.842$ and $R^2_{\text{test}} = 0.715$, which is a satisfactory level. A graphical representation of the model for the training and test sets is represented in Figure 1 (A). Compared to the 4-variable model, the eight-variable model shows better R^2_{train} and Q^2 performance values for training set, less standard deviation s , and better predictive performance due to higher R^2_{test} for the test set, 0.812. In comparing to four-variable model, the 8-variable model having larger number of variables, which can lead to some level of overfitting. A graphical representation of the model for the training and test sets is presented in Figure 1 (B).

Both equations: (1) and (2), show satisfactory statistical results that confirms the robustness of these models. However, considering the combined productivity for both training and test sets, the second model provides better performance.

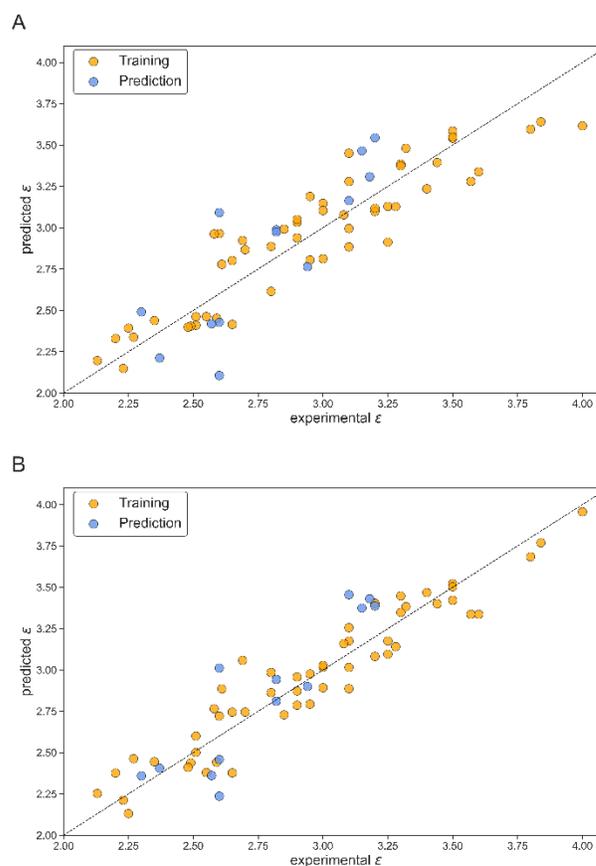


Fig.1. Plots of experimental and predicted values of the dielectric constants for the entire data set. Yellow dots are the training set, blue dot are the test set (A – for Eq. 1.; B – for Eq. 2).

Descriptor selection was performed by applying variable selection GA algorithm, followed by MLRA approach together with cross-validation LOO procedure. Based on the size of the dataset and the correlation coefficients of the training and test sets (R^2_{train} and R^2_{test}), the significance criterion F and the standard error s , the number of descriptors in the final QSPR model was determined.

Table 2. Statistical characteristics of the four- and eight-variable models.

Model, No. of descriptors	R^2_{train}	R^2_{adj}	s	F	Q^2	R^2_{test}
---------------------------	----------------------	--------------------	-----	-----	-------	---------------------

1	4	0.842	0.829	0.187	64.124	0.813	0.715
2	8	0.905	0.888	0.151	52.542	0.865	0.812

A very important step is to check an applicability domain (AD). Predictions of compounds can be considered reliable only if the dataset's chemical space of applicability is within the predictive chemical space of the developed model, before the model can be applied for predictions. The AD check was performed by application of leverage approach, i.e. William's plot evaluation for the final models. All points were within the three standardized residues ($\pm 3\sigma$) and within the HAT index, where h^* - critical value of leverage h . If the errors of estimation would exceed the values of the standardized residues, then the predicted values could go out of AD and give inaccurate predictions as they go beyond reasonable extrapolation. If the value of h of the resulted data is higher than h^* , then they are considered as structurally significant contributors to the model [54].

As can be seen in the Williams plots (Fig. 2) for both equations, in the first model (A) there only two polymers, and in the second (B) only one polymer that have values h higher than h^* . However, these polymers have low residual values, which means that the model is stable enough to make reliable predictions for all polymers structurally similar to the ones in the dataset.

The obtained models contain the following descriptors: Me - mean atomic Sanderson electronegativity (scaled on Carbon atom); AAC - mean information index on atomic composition; R5p+ - R maximal autocorrelation of lag 5 / weighted by polarizability; JGI1 - mean topological charge index of order 1; GATS1p - Geary autocorrelation of

lag 1 weighted by polarizability; Mor22v - signal 22 / weighted by van der Waals volume; RARS - R matrix average row sum; ESpm11u - Spectral moment 11 from edge adj. matrix; R1v+ - R maximal autocorrelation of lag 1 / weighted by van der Waals volume; nCt - number of total tertiary C(sp³).

More information about these descriptors can be found in Dragon software user's guide [52, 53] and the references therein.

As a rule, the value of the coefficient F indicates the ability of the model to predict the value of the properties in the training set. The large F ratio values in both equations (64.124 and 52.542 for the first and second, respectively) indicate that both equations do an excellent job with predicting ϵ values. Each equation has an adjusted value of R^2_{adj} : 0.829 and 0.888, which denotes a very good correspondence between correlation and data variation. The cross-validated correlation coefficient (Q^2 for Eq.1. = 0.813; Q^2 for Eq.2. = 0.865) demonstrates the robustness of the models. The model was further validated using a y -randomization test. The obtained R^2Y_{scr} against the correlation coefficient between the original and shuffled data is shown in Fig. 3. It can be seen from the Fig. 3 that the original models are not due to random correlations, values of R^2Y_{scr} are significantly low. It is worth noting that the model 1 showed much stronger robustness at y -scrambling test than model 2. The calculated results of the values of ϵ from equations 1 and 2 for the training and test sets are shown in Table 1 and Fig. 1.

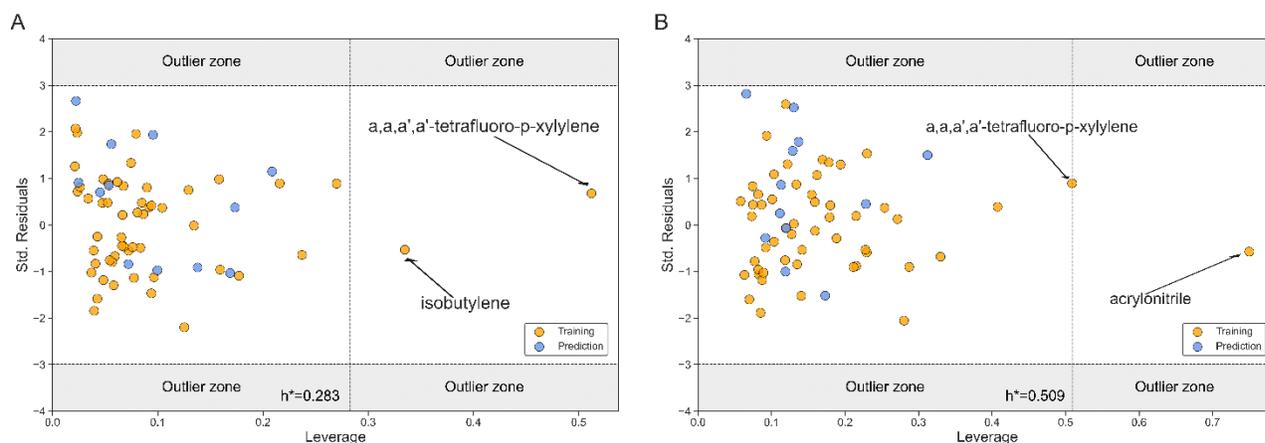


Fig.2. Williams plots for Eq.1. (A) and Eq.2. (B): yellow balls — training set; blue balls — test set.

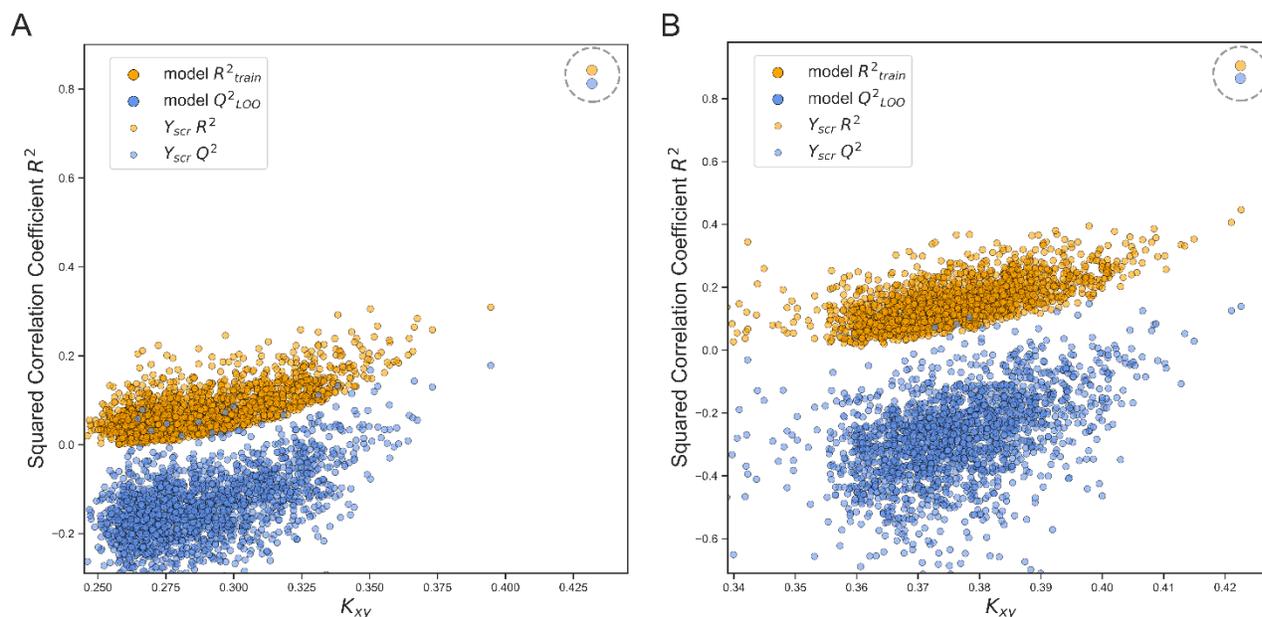


Fig.3. Y-scrambling plots of selected 4-descriptor Eq.1 (A) and 8-descriptor Eq.2 (B) models.

Based on the model selection procedure described earlier, the relative contribution of descriptors to the respective models was determined and shown in Fig. 4. The descriptors involved in the model are having the reducing contribution to the model in the following order: for the Eq.1: Me > AAC > R5p+ > JGI1; for the Eq.2: Me > AAC > RARS > R1v+ > GATS1p > ESpm11u > Mor22v > nCt.

One of the most important descriptors involved in both equations is the AAC information index. This descriptor contains information about each atom in a molecule by its own atom type and its bond type and the atom types of its first neighbors. AAC is a measure of atomic composition associated with molecular complexity. When a molecule is larger and its elemental composition is more complex, the value of the descriptor increases. The positive value of this descriptor indicates that polymers with a more complex structure and, accordingly, with a larger value for this descriptor would have larger values of ϵ .

ESpm11u is based on the use of bond distances as weights in the diagonal entries of the edge matrix.

Conclusions

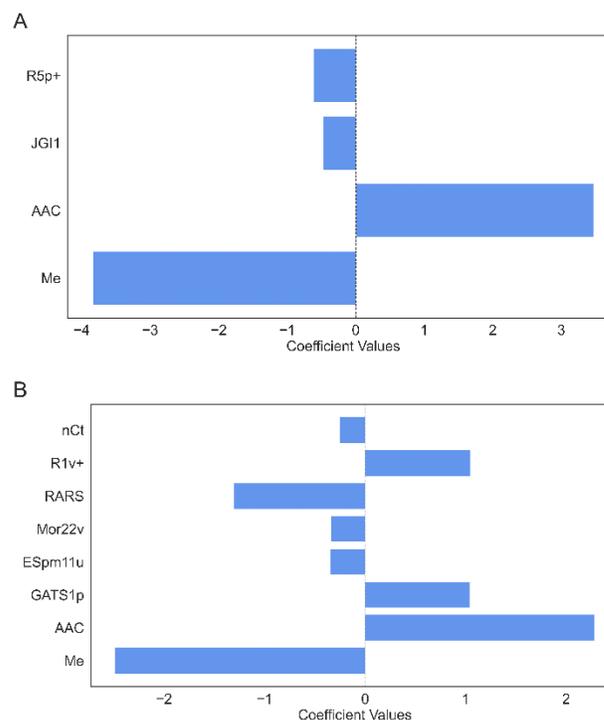


Fig.4. Descriptors contributions to Eq.1 (A) and Eq.2 (B).

In this work, a structure-property relationship model for dielectric constants (ϵ) based on a diverse set of polymers is developed. A transparent model

was obtained with application of GA-MLRA approach, to get a mechanistically explainable model. This work represents two QSPR models developed based on descriptors computed from monomeric polymer structures. The reliability of the models was validated by several verification methods. The best overall performance is achieved by a four- and eight-descriptor QSAR models, with R^2 values of 0.842/0.715 and 0.905/0.812 for training/test sets, respectively per each model. The models are suitable for further development of polymers with desired dielectric constants based on chemical structure information of monomers.

Acknowledgments

A.K. and B.R. thank the Department of Coatings and Polymeric Materials at North Dakota State University (NDSU) for providing the Lowell Wood funding. The authors thank Prof. Paola Gramatica for generously provided a free license for the QSARINS software. M.E. and B.R. acknowledge the support from the North Dakota Established Program to Stimulate Competitive Research (ND EPSCoR Award #IIA-1355466) through the New Faculty Award. Supercomputing support from CCAST Thunder HPC System at NDSU and Extreme Science and Engineering Discovery Environment (XSEDE) through the award allocation (TG-DMR110088) is acknowledged.

References

- Huang X, Jiang P (2015) Core-shell structured high-k polymer nanocomposites for energy storage and dielectric applications. *Adv Mater* 27(3):546–554
- Rao Y, Wong CP (2004) Material characterization of a high-dielectric-constant polymer-ceramic composite for embedded capacitor for RF applications. *J Appl Polym Sci* 92(4): 2228–2231
- Dang ZM, Yuan JK, Yao SH et al (2013) Flexible nanodielectric materials with high permittivity for power energy storage. *Adv Mater* 25(44):6334–6365
- J. Bicerano, Prediction of Polymer Properties, Marcel Dekker, New York (1996).
- C. C. Ku and R. Liepins, *Electrical Properties of Polymers, Chemical Principles*. Hanser, Munich (1987).
- G. Hougham, G. Tesoro, and A. Viehbeck, *Macromolecules*, 29, 3453 (1996).
- R.C. Schweitzer and J.B. Morris, *J. Chem. Inf. Comput. Sci.*, 40, 1253 (2000).
- J. Devillers and A.T. Balaban, Eds., *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, The Netherlands (1999).
- M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley Online Library, New York (2000).
- X.J. Yao, Y.W. Wang, X.Y. Zhang, R.S. Zhang, M.C. Liu, Z.D. Hu, and B.T. Fan, *Chemom. Intell. Lab. Syst.*, 62, 217 (2002).
- J. Xu, B. Guo, B. Chen, and Q. Zhang, *J. Mol. Model.*, 12, 65 (2005).
- R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim (2000).
- A. Katritzky, S. Sild, and M. Karelson, *J. Chem. Inf. Comput. Sci.*, 38, 1171 (1998).
- R. Garcí'a-Domenech and J.V. Julia'n-Ortiz, *J. Phys. Chem. B*, 106, 1501 (2002).
- J. Xu, B. Chen, Q. Zhang, and B. Guo, *Polymer*, 45, 8651 (2004).
- X. Yu, B. Yi, and X. Wang, *J. Comput. Chem.*, 28, 2336 (2007).
- J. Gao, J. Xu, B. Chen, and Q. Zhang, *J. Mol. Model.*, 13, 573 (2007).
- J. Xu, H. Liang, B. Chen, W. Xu, X. Shen, and H. Liu, *Chemom. Intell. Lab. Syst.*, 92, 152 (2008).
- F. Jabeen, M. Chen, B. Rasulev, M. Ossowski, P. Boudjouk, Refractive indices of diverse data set of polymers: A computational QSPR based study, *Computational Materials Science*, 2017, 137, 215-224
- Erickson M.E., Ngongang M., Rasulev B. A Refractive Index Study of a Diverse Set of Polymeric Materials by QSPR with Quantum-Chemical and Additive Descriptors, *Molecules*, 2020, 25 (17), 3772
- Khan M., Rasulev B., Roy K., QSPR modeling of the refractive index for diverse polymers using 2D descriptors, *ACS Omega*, 2018, 3 (10), 13374-13386
- Katritzky, S. Sild, V. Lobanov, and M. Karelson, *J. Chem. Inf. Comput. Sci.*, 38, 300 (1998).
- B.E. Mattioni and P.C. Jurs, *J. Chem. Inf. Comput. Sci.*, 42, 232 (2002).
- C. Cao and Y. Lin, *J. Chem. Inf. Comput. Sci.*, 43, 643 (2003).
- Afantitis, G. Melagraki, K. Makridima, A. Alexandridis, H. Sarimveis, and O. Iglessi-Markopoulou, *J. Mol. Struct. (Theochem)*, 716, 193 (2005).
- X. Yu, X. Wang, H. Wang, A. Liu, and C. Zhang, *J. Mol. Struct. (Theochem)*, 766, 113 (2006).
- X. Yu, B. Yi, X. Wang, and Z. Xie, *Chem. Phys.*, 332, 115 (2007).
- C. Bertinetto, C. Duce, A. Micheli, R. Solaro, A. Starita, and M.R. Tine', *Polymer*, 48, 7121 (2007).
- C. Duce, A. Micheli, A. Starita, M.R. Tine', and R. Solaro, *Macromol. Rapid Commun.*, 27, 711 (2006).
- X. Yu, *Fiber Polym.*, 11, 757 (2010).
- Chen M., Jabeen F., Rasulev B., Ossowski M., Boudjouk P. A computational structure–property relationship study of glass transition temperatures for a diverse set of polymers, *Journal of Polymer Science, Part B: Polymer Physics*, 2018, 56, 877–885
- Petrosoyan L.S., Sizochenko N., Leszczynski J., Rasulev B. Modeling of Glass Transition Temperatures for Polymeric Coating Materials: Application of QSPR Mixture-based Approach, *Molecular Informatics*, 2019, 38 (8-9), 1800150
- Anas K., Amirhadi A., Xia W., Rasulev B. Predicting Glass Transition of Amorphous Polymers via Integration of Cheminformatics and Molecular Dynamics Simulations, *Polymer*, 2021, in review
- J. Xu, B. Chen, H. Liang, W. Xu, and W. Cui, *Polimery*, 54, 19 (2009).
- D. Ajloo, A. Sharifian, and H. Behniafar, *Bull. Korean Chem. Soc.*, 31, 2009 (2008).

36. X. Yu, X. Wang, H. Wang, X. Li, and J. Gao, *QSAR Comb. Sci.*, 25, 156 (2006).
37. B. Rasulev, F.J. Jabeen, S. Stafslie, B.J. Chisholm, J. Bahr, M. Ossowski, P. Boudjouk, Polymer coating materials and their fouling release activity: A cheminformatics approach to predict properties, *ACS Applied Materials & Interfaces*, 2017, 9 (2), 1781–1792
38. M. Cocchi, P.G.D. Benedetti, R. Seeber, L. Tassi, and A. Ulrici, *J. Chem. Inf. Comput. Sci.*, 39, 1190 (1999).
39. R.C. Schweitzer and J.B. Morris, *Anal. Chim. Acta*, 384, 285 (1999).
40. S. Sild and M. Karelson, *J. Chem. Inf. Comput. Sci.*, 42, 360 (2002).
41. J.-P. Liu, W.V. Wilding, N.F. Giles, and R.L. Rowley, *J. Chem. Eng. Data*, 55, 41 (2010).
42. Liu, X. Wang, L. Wang, H. Wang, and H. Wang, *Eur. Polym. J.*, 43, 989 (2007).
43. Golbraikh and A. Tropsha, *J. Mol. Graph. Model.*, 20, 269 (2002).
44. Sun Lili, Zhou Liping, Yu Yu, Lan Yukun, Li Zhiliang. QSPR study of polychlorinated diphenyl ethers by molecular electronegativity distance vector (MEDV-4). *Chemosphere* 2007;66:1039.
45. Marrero-Ponce Yovani. Total and local (atom and atom type) molecular quadratic indices:significance interpretation, comparison to other molecular descriptors, and QSPR/ QSAR applications. *Bioorg Med Chem* 2004;12:6351.
46. Advanced Chemistry Development, I. ChemSketch. (2014).
47. Turabekova, M.A., Rasulev, B.F., Dzhakhangirov, F.N., Leszczynska, D., Leszczynski, J., Aconitum and Delphinium alkaloids of curare-like activity. QSAR analysis and molecular docking of alkaloids into AChBP, *European Journal of Medicinal Chemistry*, 2010, 45 (9), 3885-3894
48. Juretic D., Kusic H., Dionysiou D.D., Rasulev B., Bozic A.L., Modeling of photooxidative degradation of aromatics in water matrix; combination of mechanistic and structural-relationship approach, *Chemical Engineering Journal*, 2014, 257, 229-241
49. Han, Y., Meng, Q., Rasulev, B., Kilin, D.S. Photofragmentation of Gas-Phase Lanthanum Iso-propyl-cyclopentadienyl Complex: Computational Modeling vs. Experiment, *Journal of Physical Chemistry A*, 2015, 119(44), 10838
50. Rasulev B.F., Kušic H., Leszczynska D., Leszczynski J., Koprivanac N. QSAR modeling of acute toxicity on mammals for aromatic compounds: the case study using oral LD50 for rats, *Journal of Environmental Monitoring*, 2010, 12 (5), 1037-1044
51. HyperChem® for Windows. Hypercube. HyperChem Release Version 8.0.7 <http://www.hyper.com/>
52. Todeschini, R., Consonni, V., Mauri, A. & Pavan, M. DRAGON-Software for the calculation of molecular descriptors. *Web version 3*, (2004)
53. R. Todeschini, V. Consonni: *Handbook of Molecular Descriptors*. WILEY-WCH Publishers, Weinheim, 2000.
54. F. Gharagheizi, A. Eslamimanesh, P. Ilani-Kashkouli, et al., QSPR molecular approach for representation /prediction of very large vapor pressure dataset, *Chem. Eng. Sci.* 76 (2012) 99–107.