http://dx.doi.org/10.1590/0104-1428.2306

Structure-flammability relationship study of phosphoester dimers by MLR and PLS[a]

Luminita Crisan[1], Smaranda Iliescu[1] and Simona Funar-Timofei[1]*

[1]*Institute of Chemistry Timisoara, Romanian Academy, Timisoara, Romania*

***timofei@acad-icht.tm.edu.ro**

Abstract

Polyphosphonates and polyphosphates having good flame retardancy represent an important class of organophosphorus based polymer additives. In this analysis the flammability of 28 previously synthesized polyphosphoesters, modelled as dimmers, was explored using the multiple linear regression (MLR) and Partial Least Square (PLS) methodology. The statistical quality of the final MLR and PLS models was estimated using the following parameters: the squared correlation coefficient ($r^2_{training}$ = 0.917 and 0.976), the training root-mean-square errors ($RMSE_{tr}$ = 0.029 and 0.016) and the leave-seven-out cross-validation correlation coefficient (q^2_{L70} = 0.748 and 0.881), respectively. External validation was checked for a test set of seven compounds using several criteria. The MLR models had somewhat inferior fitting results. The final MLR and PLS models can be used for the estimation of limiting oxygen index (LOI) values of new polyphosphoester structures. The presence of phosphonate groups and increasing molecular branching in an isomeric series favour the dimer flammability.

Keywords: *quantitative structure-property relationships, polyphosphonate, polyphosphate, limiting oxygen index, flame retardancy.*

1. Introduction

An important feature of most commercial polymers is to be non-flammable or flame retardant[1]. Other polymer properties, like as: glass transition temperature, thermal decomposition temperature, etc., have been previously studied by quantitative structure-property relationships[2,3].

Flame retardant polymeric materials containing phosphorus, like poly(alkyl or aryl)phosphonates, display good flame retardancy[4].

Different polyphosphoesters with fire retardant properties were reported in the literature, being included in materials like: polycarbonates, polyamides, thermosets, etc[5]. The flammability of phosphorous polymers was investigated in order to determine structural–property relationships, too[6,7]. Two types (*R* and *S*) of chirality were found for the monomer polyphosphoesters, which were geometry optimized using the MMFF94s force field[6]. Multiple linear regression (MLR), artificial neural networks (ANNs) and support vector machines (SVMs) were applied to correlate the limiting oxygen index (LOI) values to the structural calculated descriptors. Good fitting results and predictable models were obtained using the MLR and ANN approaches, the SVM modelling providing the poorest results. It was concluded that the monomer geometry is important for flame retardancy.

Our goal was to develop robust multiple linear regression (MLR) and the partial least squares (PLS) models that select a set of variables that efficiently predict the limiting oxygen index (LOI) values and guide new information on the flammability mechanism of polyphosphoesters[6] dimers. This parallel approach gives the opportunity to compare the quality of results supplied by the two methodologies.

2. Materials and Methods

2.1 Data set

We used a series of 28 previously synthesized polyphosphoesters[6], which were modelled in the present study as dimers. The dataset in this investigation consisted of 28 RR, RS, SR and SS phosphoester dimers for compounds 1 to 14; compounds 15 to 28 had only one chiral centre, at the P2 phosphorous atom (see Figure 1).

Experimental data for the limiting oxygen index (LOI), expressed in % (Table 1), and used as dependent variable in this study, was previous reported in references[6] and[7]. Dimer molecular structures were built using the Marvin program[8], which was used for drawing, displaying and characterizing chemical structures. Dimer conformers were pre-optimized using the 94s variant of the MMFF (Merck Molecular force field)[9] with coulomb interactions and the attractive part of the van der Waals interactions, included in the OMEGA software[10-12]. The following parameters were used for the conformer generation: a maximum of 400 conformers per compound, an energy cut-off of 10 kcal/mol relative to a global minimum identified from the search. SMILES notation was used as program input. The stereoisomers were generated using the 'Flipper' utility inside the Omega program. To avoid redundant conformers, any conformer having a RMSD fit outside 0.5 Å to another conformer was removed.

[a] Dedicated to the 150th anniversary of the Romanian Academy.

2.2 Molecular descriptor calculation

Molecular descriptors were calculated for the optimized dimer structures, using the DRAGON[13] and InstantJchem (which was used for structure database management, search and prediction)[14] software. The 1511 Dragon molecular descriptors were divided into twenty-two logical blocks, as follows: constitutional descriptors, topological descriptors (MSD-mean square distance index (Balaban), PW4-path/walk 4 - Randic shape index), walk and path counts, connectivity indices, information indices (IC5-information content index (neighborhood symmetry of 5-order)), 2D autocorrelations (Gats5e-Geary autocorrelation - lag 5/weighted by atomic Sanderson electronegativities), edge adjacency indices (EEig09d-Eigenvalue 09 from edge adj. matrix weighted by dipole moments), BCUT descriptors, topological charge indices (GGI1-topological charge index of order 1, JGI2-mean topological charge index of order2), eigenvalue based indices, Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D-MoRSE descriptors (Mor15e-3D-MoRSE - signal 15/weighted by atomic Sanderson electronegativities, Mor13p-3D-MoRSE - signal 13/weighted by atomic polarizabilities Mor13m-3D-MoRSE - signal 13/weighted by atomic masses), WHIM descriptors, GETAWAY descriptors (R2m+ - R maximal autocorrelation of lag 2/weighted by atomic masses), functional group counts (nP(=O)O2R-number of phosphonates), atom-centered fragments, charge descriptors, molecular properties, 2D binary fingerprints, and 2D frequency fingerprints. Then the molecular descriptors were verified and constant or near-constant variables were eliminated. The calculated molecular descriptors play a fundamental role in transforming the chemical information into a numerical code suitable for application in computation[15].

2.3 Training and test set selection

The series of phosphoester dimers were divided into training and test set using several approaches: the partition against medoids (PAM) algorithm[16] ("cluster" package available in R[17] based on the Euclidian distance), the decreasing response order and the random splitting. In order to use same test set in both MLR and PLS approaches, seven out of twenty eight (25%) phosphoester dimers (compounds 2, 10, 11, 15, 17, 19 and 22, see Figure 1) were chosen as test set to validate the final models. The data structures and the LOI range values (in %), comprised in the test set (0.22-0.50) and the training set (0.18-0.55), are commensurate.

2.4 Multiple Linear Regression (MLR) and Partial Least Square (PLS)

Multiple linear regression (MLR)[18] has been applied after variable selection carried out by means of a genetic algorithm included in the QSARINS v. 2.2 program[19,20] using

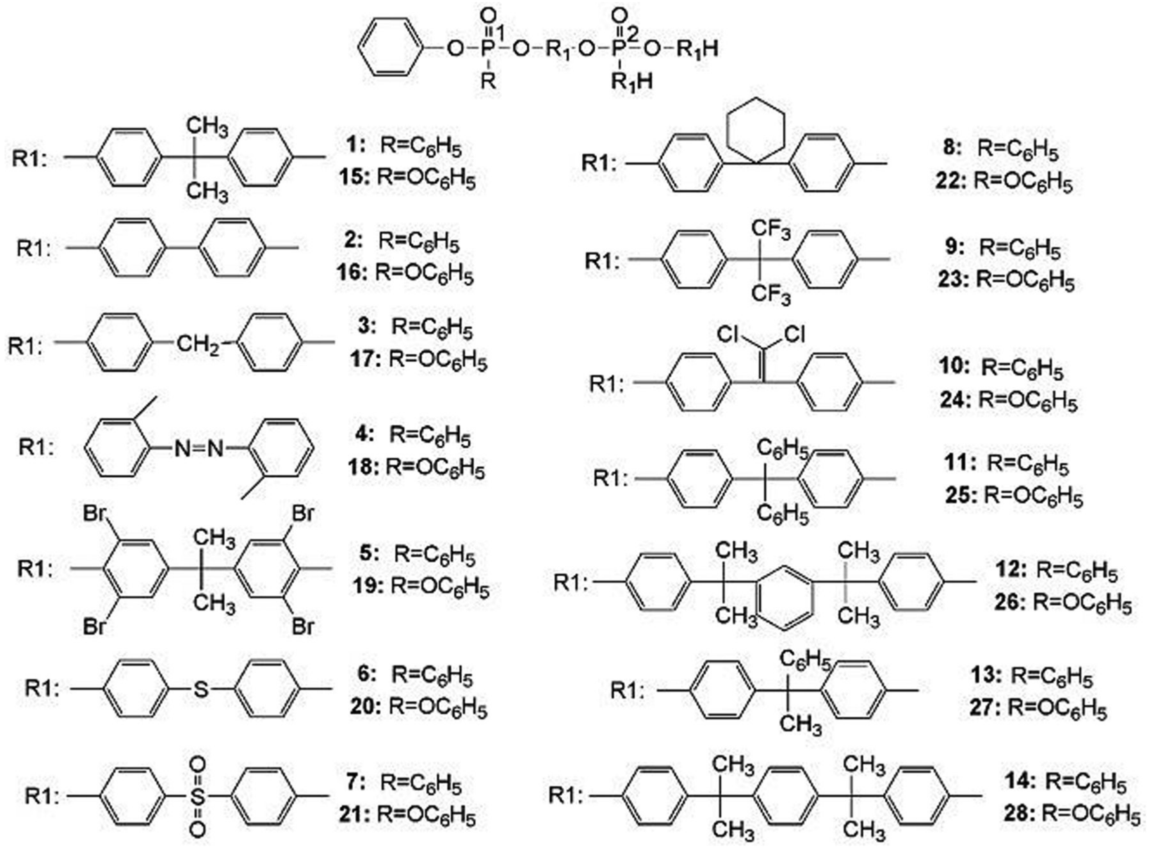

Figure 1. Dimer phosphoester structure. RR series: R chiral centre at P1, R chiral centre at P2; RS series: R chiral centre at P1, S chiral centre at P2; SR series: S chiral centre at P1, R chiral centre at P2; SS series: S chiral centre at P1, S chiral centre at P2; compounds 15 to 28 had only one chiral centre, at the P2 phosphorous atom.

the RQK fitness function, with leave-one-out cross-validation correlation coefficient, which constrained the function to be optimized. In MLR, the number of 1549 calculated descriptors is too high compared to the number of compounds (N = 28) and an appropriate variable selection method was required. MLR calculations were carried out separately for each dataset: RR, RS, SR, SS.

In MLR calculations the structural data was normalized based on the autoscaling method, which can be described as:

$$XT_{mj} = \frac{X_{mj} - \bar{X}_m}{S_m} \qquad (1)$$

where for each variable *m*, XT_{MJ} and X_{MJ} are the *j* values for the *m* variable after and before scaling, respectively, $\bar{X}_m$ is the mean and S_M the standard deviation of the variable.

The PLS methodology is a generalization of the MLR one, having as main advantage the possibility to analyze the data with correlated, noise, and large number of independent variables[21]. In the PLS equation the latent variables were transformed as function of the original X_{IJ} (i =1, 2,..., N; j=1, 2,..., K) variables, resulting following equation:

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + ... + b_j X_{ij} + ... + b_k X_{ik} \qquad (2)$$

where $\hat{Y}_I$ represents the calculated dependent variable, and b_J the PLS coefficients. The obtained models were optimized by a procedure of outlier detection and based on variables with significant coefficients different from zero. When the variable selection was achieved, only the significant descriptors with coefficients different from zero were preserved in the final models (for noise elimination).

Both methologies have as main goal to find out a mathematical model with minimum number of parameters and with good estimation capability.

2.5 Model validation

For the external validation of the MLR and PLS models several parameters were calculated: Q^2_{F1}[22], Q^2_{F2}[23], Q^2_{F3}[24] (models with values higher than 0.7 were considered acceptable), the CCC_{ext} (the concordance correlation coefficient, with satisfactory values higher than 0.85)[25], $RMSE_{ext}$ (root-mean-square errors) and MAE_{ext} (mean absolute error)[26] and R^2_{pred} (a higher limit than 0.5 was considered as acceptable)[27]. The comparable thresholds used in this study for different validation criteria have been rigorously previously determined[25,28]. Other statistical parameters[29] were used for the external test set: (i) squared correlation coefficient (r^2_{test}) between the predicted and observed activities as well as squared correlation coefficient by cross-validation (q^2); (ii) coefficient of determination for linear regressions with intercepts set to zero, i.e. r^2_0 (predicted versus observed activities), and r'^2_0 (observed versus predicted activities); (iii) slopes k and k' of the above mentioned two regression lines. The following conditions should be satisfied for a model with acceptable predictive ability:

$$q^2 > 0.5 \qquad (3)$$

$$r^2_{test} > 0.6 \qquad (4)$$

$$\frac{(r^2 - r_0^2)}{r^2} < 0.1 \quad and \quad 0.85 \le k \le 1.15 \qquad (5)$$

$$\frac{(r^2 - r_0'^2)}{r^2} < 0.1 \quad and \quad 0.85 \le k' \le 1.15 \qquad (6)$$

$$\left| r_0^2 - r_0'^2 \right| < 0.3 \qquad (7)$$

For the internal validation of the final models other parameters were employed: $r^2_{training}$ (determination coefficient), q^2_{L70} (leave-seven-out cross-validation coefficient; values higher than 0.7 were considered as acceptable), q^2_{LOO} (leave-one-out cross-validation coefficient), $RMSE_{tr}$, MAE_{tr} and CCC_{tr}, calculated for the training set.

In the mean time higher $r^2_{training}$ values must be accompanied by q^2 values as close to the $r^2_{training}$ ones as possible[30] (to avoid over fitting, which was, also, checked by the RMSE and MAE values).

The risk of chance correlation was, also, verified by the Y-scrambling procedure (r^2_{Scr} and q^2_{Scr}) and must have lower values than the original model. For calculation of r^2_{Scr} and q^2_{Scr} this process was repeated 999 times in case of PLS calculations and 2000 times in the MLR ones.

After the check of all validation parameters, the applicability domain for the models is required, because robust and validated models cannot be expected to reliably predict the modelled property for any type of compounds. The applicability domain is a theoretical region in physicochemical of response and chemical structure space for which a QSAR model should make predictions with a given reliability[30]. In the applicability domain only the predictions for those compounds that fall within this domain can be considered as reliable, not extrapolations of the model. In the Williams plot the standardized residuals versus the leverages (hi) was exploited to visualize the applicability domain for our final MLR models.

3. Results and Discussions

The major objective of this paper was the estimation of limiting oxygen index (LOI) of phosphoester dimers using molecular descriptors that can be computed directly from molecular structure and guide new information on the flammability mechanism.

3.1 MLR results

The relationship between the molecular descriptors and LOI values of the dimer derivatives is illustrated by the following Equations 8-11:

RR model

$$\begin{gathered} LOI = 0.56(\pm 0.03) \text{ - } 0.23(\pm 0.03)MSD \text{ -} \\ 0.19(\pm 0.03)EEig09d + 0.20(\pm 0.03)R2m^{+} + \\ 0.05(\pm 0.02)nP(=O)O2R \\ SEE = 0.03 \quad r^2_{adj} = 0.896 \quad F = 44.09 \quad q^2_{LOO} = 0.864 \end{gathered} \qquad (8)$$

RS model

$$\begin{gathered} LOI = 0.55(\pm 0.05) + 0.18(\pm 0.04)PW4 \text{ -} \\ 0.26(\pm 0.05)GGI1 + \\ 0.16(\pm 0.06)JGI2 \text{ - } 0.35(\pm 0.06)Mor13m \\ SEE = 0.05 \quad r^2_{adj} = 0.787 \quad F = 19.48 \quad q^2_{LOO} = 0.745 \end{gathered} \qquad (9)$$

SR model

$$LOI = 0.38(\pm 0.02) + 0.19(\pm 0.04) IC5 + 0.13(\pm 0.05) Mor15e - 0.33(\pm 0.04) Mor13p \quad (10)$$
$$SEE = 0.04 \quad r^2_{adj} = 0.839 \quad F = 35.60 \quad q^2_{LOO} = 0.790$$

SS model

$$LOI = 0.52(\pm 0.04) + 0.18(\pm 0.04) IC5 - 0.15(\pm 0.04) GATS5e - 0.29(\pm 0.03) Mor13p \quad (11)$$
$$SEE = 0.04 \quad r^2_{adj} = 0.870 \quad F = 45.41 \quad q^2_{LOO} = 0.831$$

where SEE represents the standard error of estimates, r^2_{adj} - the adjusted r^2, F- the Fischer test, q^2_{LOO} -leave-one-out cross-validation coefficient. Other statistical results of models 8-11 are included in Tables 2, 3, 4.

The Williams (of the standardized residuals versus the leverage) plot was used to visualize the applicability domain of the final best MLR_RR model (Figure 2). This plot confirms the absence of outliers and influential points. All compounds were located within the applicability domain and were predicted accurately.

The MLR_RR model is completely satisfactory in the fitting and has high predictive power. The LOO (leave-one-out) cross-validation highlights that the model is stable, not obtained by chance, in fact the difference between $r^2_{training}$ and q^2_{LOO} is small: 5.3%. This model is internally predictive with differences between q^2_{LMO} and q^2_{LOO} of -4.5%, and between $r^2_{training}$ and r^2_{LMO} of 9.8%.

The risk of chance correlation was, also, verified by the Y-scrambling procedure. The extremely low calculated r^2_{Scr} and q^2_{Scr} scrambling values (Table 2) indicate no chance correlation for the chosen models.

The RMSE (root-mean-square error) values for the training and validation sets are similar. The chosen MLR_RR model demonstrate a satisfactory stability in internal validation, has high fitting, internal and external predictive power.

The high values of Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{ext} external validation parameters (see section 2.5) included in Table 3 and all calculated terms of Golbraikh and Tropsha (Table 4) confirm the predictive power of all MLR models.

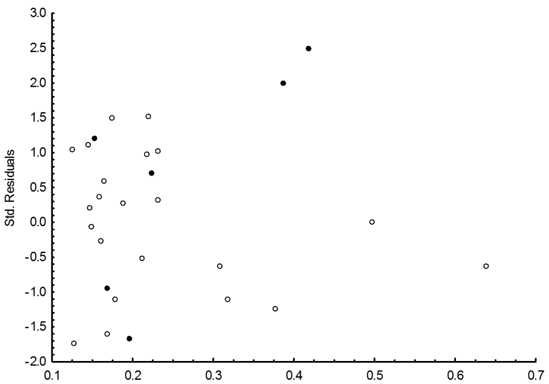

Figure 2. Williams plot: standardized residuals of the MLR_RR model versus leverages, predicted by fitting. Training compounds are marked by white circles and test compounds by black circles.

Better statistical results and a more stable model to simulate polymer flammability were noticed in case of the RR dataset model compared to the others.

The edge adjacency matrix encodes information about the connectivity between graph edges[15]. EEig09d (eigenvalue 09 from edge adj. matrix weighted by dipole moments) takes into account the molecular polarity, being unfavourable for dimer flame retardancy.

The mean square distance index, denoted as MSD[15], is calculated from the second-order distance distribution moment[31]. The MSD index decreases with increasing molecular branching in an isomeric series, which is favourable for dimer flammability.

GETAWAYs (Geometry, Topology, and Atom-Weights Assembly) are geometrical descriptors encoding information on the effective position of substituents and fragments in the molecular space[15]. Moreover, they are independent of molecule alignment and they, also, account to some extent for information on molecular size and shape as well as for specific atomic properties. Increased R2m+ (R maximal autocorrelation of lag 2/weighted by atomic masses) values favour the dimer flammability. Compounds containing phosphonate groups are favourable for the dimer flame retardancy.

3.2 PLS results

PLS calculations were performed with SIMCA-P+12[32] program using 21 stereoisomers as a training set and 7 stereoisomers as a test set with the taken ratio of 75% for training set and 25% for test set in whole series of compounds. The large difference between the $r^2_{training}$ and q^2_{L70} values of the first calculated PLS model (lower than 0.3 is accepted) demonstrated the model over fit, and suggested the need for enhancement of the model quality. Therefore, the noise variables (with insignificant coefficient values) have been removed. Several PLS models were developed for the RR, RS, SR and SS datasets to increase their predictive power. In the final PLS_SS model compound 5 was omitted, being found as outlier, in accordance to the Hotelling's T^2 range plot[32].

The final (four-components for the RR, RS, SR datasets and two-components for the SS dataset) PLS models are satisfactory in the fitting (Table 2). The over fitting of the models was exceeded by the remarkable high and close values of $r^2_{training}$ and q^2_{L70}, and was, also, checked by the RMSE and MAE (mean absolute error) parameters. In the same time similar RMSE values for the training and validation sets are observed (Tables 2 and 3).

PLS models with predictive power were obtained (see Tables 3 and 4), except the PLS_SS one, as seen from the values of Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{ext} parameters. The predicted LOI values for the RR dataset are given in Table 1.

The PLS models were internally validated using, also, 999 permutations in Y-scrambling. The calculated r^2_{Scr} and q^2_{Scr} scrambling values (Table 2) indicate no chance correlation for the chosen models.

Table 1. Experimental and predicted LOI values, structural descriptors included in the final MLR_RR model.

No	Exp. LOI	Calc. LOI by MLR_RR	Calc. LOI by PLS_RR	MSD	EEig09d	R2m+	nP(=O)O2R
1	0.38[a]	0.38	0.36	0.218	2.266	0.022	2
2	0.35[a]	0.41	0.37	0.228	2	0.023	2
3	0.30[a]	0.33	0.33	0.235	2.245	0.026	2
4	0.48[a]	0.45	0.45	0.213	2.058	0.023	2
5	0.55[a]	0.55	0.55	0.191	2.652	0.187	2
6	0.28[a]	0.26	0.28	0.235	2.569	0.043	2
7	0.29[a]	0.32	0.29	0.218	2.576	0.044	2
8	0.42[b]	0.44	0.41	0.2	2.271	0.021	2
9	0.44[a]	0.42	0.44	0.182	2.744	0.062	2
10	0.50[a]	0.45	0.54	0.212	2.467	0.117	2
11	0.47[b]	0.52	0.54	0.169	2.361	0.022	2
12	0.32[a]	0.31	0.32	0.209	2.621	0.016	2
13	0.40[a]	0.45	0.44	0.191	2.361	0.021	2
14	0.33[a]	0.28	0.31	0.219	2.619	0.019	2
15	0.28[a]	0.29	0.25	0.216	2.434	0.018	0
16	0.25[a]	0.28	0.25	0.226	2.359	0.025	0
17	0.22[a]	0.24	0.27	0.232	2.431	0.026	0
18	0.36[a]	0.31	0.37	0.211	2.454	0.025	0
19	0.40[a]	0.37	0.47	0.19	2.722	0.082	0
20	0.18[a]	0.19	0.18	0.232	2.691	0.04	0
21	0.20[a]	0.23	0.19	0.216	2.695	0.032	0
22	0.31[a]	0.35	0.32	0.198	2.434	0.017	0
23	0.33[a]	0.37	0.35	0.181	2.749	0.063	0
24	0.50[a]	0.49	0.50	0.21	2.514	0.204	0
25	0.48[a]	0.45	0.47	0.168	2.445	0.017	0
26	0.23[a]	0.24	0.25	0.207	2.723	0.016	0
27	0.37[a]	0.38	0.37	0.189	2.441	0.019	0
28	0.24[a]	0.21	0.23	0.217	2.724	0.016	0

[a] from reference [6] and [b] from reference [7].

Table 2. Internal validation parameters of the MLR and PLS models (training set).

Model	$N_{training}$	R_X^2	$r^2_{training}$	q^2_{L70}	$RMSE_{tr}$	MAE_{tr}	CCC_{tr}	r^2_{Scr}	q^2_{Scr}
MLR_RR	21	-	0.917	0.748	0.029	0.025	0.957	0.198	-0.453
MLR_RS	21	-	0.830	0.658	0.042	0.032	0.907	0.199	-0.422
MLR_SR	21	-	0.863	0.743	0.037	0.029	0.926	0.149	-0.309
MLR_SS	21	-	0.889	0.800	0.034	0.025	0.941	0.152	-0.340
PLS_RR	21	0.726	0.976 (4)**	0.881	0.016	0.012	0.988	0.635	-0.510
PLS_RS	21	0.701	0.965 (4)**	0.754	0.019	0.014	0.982	0.627	-0.557
PLS_SR	21	0.702	0.972 (4)**	0.792	0.017	0.013	0.986	0.556	-0.571
PLS_SS	20*	0.461	0.885 (2)**	0.656	0.031	0.026	0.939	0.324	-0.398

* Compound 5 was found as outlier and omitted from the final model; ** Number of components is given in parenthesis.

Table 3. External validation parameters of the MLR and PLS models (test set).

Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	MAE_{ext}	CCC_{ext}
MLR_RR	0.811	0.808	0.833	0.041	0.037	0.900
MLR_RS	0.814	0.811	0.836	0.041	0.033	0.901
MLR_SR	0.713	0.708	0.747	0.051	0.044	0.856
MLR_SS	0.787	0.783	0.812	0.044	0.039	0.896
PLS_RR	0.777	0.773	0.803	0.045	0.039	0.912
PLS_RS	0.727	0.723	0.759	0.050	0.040	0.902
PLS_SR	0.716	0.711	0.749	0.051	0.042	0.898
PLS_SS	0.554	0.529	0.514	0.050	0.056	0.799

Table 4. Golbraikh and Tropsha criteria[29] calculated for external validation of the MLR and PLS models (test set).

Model	r^2_{test}	$\frac{r^2 - r_0^2}{r^2}$	$\frac{r^2 - r_0'^2}{r^2}$	k	k'	$\left\| r_0^2 - r_0'^2 \right\|$
MLR_RR	0.832	0.000	0.039	0.963	1.028	0.032
MLR_RS	0.867	0.004	0.050	0.945	1.049	0.040
MLR_SR	0.732	0.032	0.022	0.994	0.987	0.008
MLR_SS	0.812	0.032	0.002	1.014	0.973	0.025
PLS_RR	0.941	0.022	0.006	0.912	1.091	0.015
PLS_RS	0.970	0.037	0.019	0.896	1.115	0.018
PLS_SR	0.940	0.050	0.018	0.903	1.100	0.030
PLS_SS	0.687	0.073	0.018	0.923	1.059	0.038

In the PLS modelling the terms having VIP values greater than 1 are the most relevant for explaining the dependent variable, and usually only these descriptors were interpreted. The descriptors showing the largest VIP values can simulate polymer flammability and are discussed below.

For all models higher values of the Randic shape index (-path/walk 4 and path/walk 5 - PW4 and PW5) are favourable for the flammability, while the MSD (Balaban mean square distance index) descriptor is unfavourable for flammability. They are topological descriptors obtained from molecular graph[15].

Another group of significant descriptors is the class of 2D autocorrelation descriptors, which are computed from molecular graph as the sum of products of atom weights of the terminal atoms of all the paths for the considered path length (the so called lag)[15]. The most important 2D autocorrelation descriptors involved in our model are the Geary parameters. The positive coefficients of GATS6m - Geary autocorrelation of lag 6 weighted by mass, increase the flame retardancy of RR, RS and SR series, while for SS dimers, the same effect was observed for descriptor GATS5v - Geary autocorrelation of lag 5 weighted by van der Waals volume.

The 3D-MoRSE descriptors provide 3D information from atomic coordinates using the same transform as in electron diffraction (which uses them to prepare theoretical scattering curves)[15]. For the RR and SR datasets, Mor13m- signal 13/weighted by mass, decrease the flame retardancy, for the RS dataset Mor15m- signal 15/weighted by mass is favourable for flammability, while for SS dataset these descriptors are insignificant.

Class of topological and frequency fingerprints descriptors are expressed as sum of topological distances between two elements or frequency of two atoms at a topological distance. Descriptors T(O..P) - the sum of topological distances between O..P, F07[O-S] – the frequency of O - S at topological distance 7, and F10[C-S] - the frequency of C - S at topological distance 10, with negative coefficients are unfavourable for flammability for RR, RS and SS datasets.

Three GETAWAY descriptors: one in the RR set: R5m- the R autocorrelation of lag 5/weighted by mass, and two in the SR set: HATS6v – the leverage-weighted autocorrelation of lag 6/weighted by van der Waals volume and HATS6p- leverage-weighted autocorrelation of lag 6/weighted by polarizability increase the dimer flame retardancy.

Better fitting and predictivity results were obtained by PLS calculations compared to the MLR ones. From MLR and PLS models better statistical results were observed in case of the RR series. Therefore R chirality of phosphorous atom is significant for dimer flammability. The final selected structural descriptors included in the MLR_RR model have VIP values > 1 in the PLS_RR model: EEig09d, VIP = 1.358, CoeffCS = -0.0086 (±0.0022), MSD, VIP = 1.670, CoeffCS = -0.0096 (±0.0021), R2m, VIP = 1.058, CoeffCS = 0.0025 (±0.0022) and nP(=O)O2R, VIP = 0.994, CoeffCS = 0.0059 (±0.0030).

Compared to the MLR previously published monomer models[6], the statistical results for fitting are improved in case of MLR and PLS dimer models. Additional structural information which influences the flame retardancy was included in the final dimer MLR (e.g. the number of phosphonates) and PLS (e.g. 2D frequency fingerprints) models.

4. Conclusions

The MLR and PLS models developed for this series of dimer phosphoesters will be helpful to predict the LOI values of new untested compounds. Better statistical results and a more stable model to simulate polymer flammability were noticed in case of the RR dataset compared to the others, the presence of R chiral centre at the phosphorous atom being important for the dimer flammability. The mean square distance index and GETAWAY descriptors favour the dimer flammability, as well as increased number of phosphonates included in the dimer structure, as derived from both MLR and PLS methodologies. Better PLS fitting and predictivity results were obtained compared to the MLR ones for all datasets, except for the SS one.

Dimers including structures with R chiral centres gave more stable and predictive models compared to the previously published MLR monomer ones.

New structural information which influences the flame retardancy was included in the final MLR and PLS dimer models.

5. Acknowledgements

This project was financially supported by Project 1.1 of the Institute of Chemistry Timisoara of the Romanian Academy. The authors are indebted to Chemaxon Ltd., OpenEye Ltd. and Prof. Paola Gramatica from The University of Insubria (Varese, Italy) for giving access to their software.

6. References

1. Irvine, D. J., McCluskey, J. A., & Robinson, I. M. (2000). Fire hazards and some common polymers. *Polymer Degradation & Stability*, *67*(3), 383-396. http://dx.doi.org/10.1016/S0141-3910(99)00127-5.
2. Le, T., Epa, V. C., Burden, F. R., & Winkler, D. A. (2012). Quantitative structure–property relationship modeling of diverse materials properties. *Chemical Reviews*, *112*(5), 2889-2919. http://dx.doi.org/10.1021/cr200066h. PMid:22251444.
3. Barbosa-da-Silva, R., & Stefani, R. (2013). QSPR based on support vector machines to predict the glass transition temperature of compounds used in manufacturing OLEDs. *Molecular Simulation*, *39*(3), 234-244. http://dx.doi.org/10.1080/08927022.2012.717282.
4. Troev, K. D. (2012). *Polyphosphoesters: chemistry and application* (pp. 263-320). London: Elsevier Insights.
5. Chen, L., & Wang, Y. Z. (2010). Aryl polyphosphonates: useful halogen-free flame retardants for polymers. *Materials*, *3*(10), 4746-4760. http://dx.doi.org/10.3390/ma3104746.
6. Funar-Timofei, S., Iliescu, S., & Suzuki, T. (2014). Correlations of limiting oxygen index with structural polyphosphoester features by QSPR approaches. *Structural Chemistry*, *25*(6), 1847-1863. http://dx.doi.org/10.1007/s11224-014-0474-7.
7. Iliescu, S., Avram, E., Visa, A., Plesu, N., Popa, A., & Ilia, G. (2011). New technique for the synthesis of polyphosphoesters. *Macromolecular Research*, *19*(11), 1186-1191. http://dx.doi.org/10.1007/s13233-011-1111-6.
8. ChemAxon. (2015). *Marvin sketch 15.2.16 software.* Záhony: ChemAxon. Retrieved in 27 April 2015, from http://www.chemaxon.com
9. Halgren, T. A. (1999). MMFF VI.MMFF94s option for energy minimization studies. *Journal of Computational Chemistry*, *20*(7), 720-729. http://dx.doi.org/10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X.
10. OpenEye Scientific. (2013). *OMEGA version 2.5.1.4 software*. Santa Fe: OpenEye Scientific. Retrieved in 29 April 2015, from http://www.eyesopen.com
11. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., & Stahl, M. T. (2010). Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, *50*(4), 572-584. http://dx.doi.org/10.1021/ci100031x. PMid:20235588.
12. Hawkins, P. C. D., & Nicholls, A. (2012). Conformer generation with OMEGA: learning from the data set and the analysis of failures. *Journal of Chemical Information and Modeling*, *52*(11), 2919-2936. http://dx.doi.org/10.1021/ci300314k. PMid:23082786.
13. Talete SRL. (2007). *Dragon professional 5.5 software*. Milano: Talete SRL. Retrieved in 4 May 2015, from http://www.talete.mi.it
14. ChemAxon. (2015). *Instant JChem 15.2.23 software*. Záhony: ChemAxon. Retrieved in 4 May 2015, from http://www.chemaxon.com
15. Todeschini, R., Consonni, V., Mannhold, R., Kubinyi, H., & Folkers, G. (Eds.). (2009). *Molecular descriptors for chemoinformatics*. Weinheim: Wiley – VCH.
16. Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. New York: Wiley.
17. R Development Core Team. (2011). *R: A language and environment for statistical computing. Version 2.13.1*. Vienna: R Foundation for Statistical Computing. Retrieved in 11 May 2015, from www.r-project.org
18. Wold, S., &Dunn, W. J.3rd (1983). Multivariate quantitative structure-activity relationships (QSAR):conditions for their applicability. *Journal of Chemical Information and Computer Sciences*, *23*(1), 6-13. http://dx.doi.org/10.1021/ci00037a002.
19. Chirico, N., Papa, E., Kovarich, S., Cassani, S., & Gramatica, P. (2012). *QSARINS, software for QSAR MLR model development and validation*. Varese: University of Insubria/QSAR Res Unit in Environ Chem and Ecotox., DiSTA. Retrieved in 11 May 2015, from http://www.qsar.it
20. Gramatica, P., Chirico, N., Papa, E., Cassani, S., & Kovarich, S. (2013). A new software for the development, analysis, and validation of QSAR MLR models. *Journal of Computational Chemistry*, *34*(24), 2121-2132. http://dx.doi.org/10.1002/jcc.23361.
21. Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109-130. http://dx.doi.org/10.1016/S0169-7439(01)00155-1.
22. Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. L., & Sheehan, D. M. (2001). QSAR models using a large diverse set of estrogens. *Journal of Chemical Information and Modeling*, *41*(1), 186-195. http://dx.doi.org/10.1021/ci000066d. PMid:11206373.
23. Schüürmann, G., Ebert, R. U., Chen, J., Wang, B., & Kuhne, R. (2008). External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *Journal of Chemical Information and Modeling*, *48*(11), 2140-2145. http://dx.doi.org/10.1021/ci800253u. PMid:18954136.
24. Consonni, V., Ballabio, D., & Todeschini, R. (2009). Comments on the definition of the Q2 parameter for QSAR validation. *Journal of Chemical Information and Modeling*, *49*(7), 1669-1678. http://dx.doi.org/10.1021/ci900115y. PMid:19527034.
25. Chirico, N., & Gramatica, P. (2011). Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*, *51*(9), 2320-2335. http://dx.doi.org/10.1021/ci200211n. PMid:21800825.
26. Goodarzi, M., Deshpande, S., Murugesan, V., Katti, S. B., & Prabhakar, Y. S. (2009). Is feature selection essential for ANN modeling? *QSAR & Combinatorial Science*, *28*(11-12), 1487-1499. http://dx.doi.org/10.1002/qsar.200960074.
27. Roy, P. P., Paul, S., Mitra, I., &Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules*, *14*(5), 1660-1701. http://dx.doi.org/10.3390/molecules14051660. PMid:19471190.
28. Chirico, N., & Gramatica, P. (2012). Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *Journal of Chemical Information and Modeling*, *52*(8), 2044-2058. http://dx.doi.org/10.1021/ci300084j. PMid:22721530.
29. Tropsha, A., & Golbraikh, A. (2010). *Predictive quantitative structure–activity relationships modeling: development and validation of QSAR models*. In J. L. Faulon & A. Bender (Ed).

Handbook of chemoinformatics algorithms (pp. 213-233). London: Chapman & Hall/CRC.

30. Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science*, *26*(5), 694-701. http://dx.doi.org/10.1002/qsar.200610151.
31. Balaban, A. T. (1983). Topological indices based on topological distances in molecular graphs. *Pure and Applied Chemistry*, *5*(2), 199-206. http://dx.doi.org/10.1351/pac198855020199.
32. Umetrics AB. (2013). *SIMCA-P+ version 12.0 software*. Umea: Umetrics. Retrieved in 18 May 2015, from http://www.umetrics.com

Received: July 20, 2015
Revised: Sept. 23, 2015
Accepted: Nov. 06, 2015