Structure-Flammability Relationships Study of Some Polyphosphonates

	SUZUKI Takahiro, FUNAR-TIMOFEI Simona, ILIESCU Smaranda
journal or publication title	Journal of Toyo University. Natural science
volume	57
page range	109-121
year	2013-03
URL	http://id.nii.ac.jp/1060/00005997/

Structure-Flammability Relationships Study of Some Polyphosphonates[1]

Takahiro Suzuki[2], Simona Funar-Timofei[3],
Smaranda Iliescu[3]

Abstract

It has been known that there is a quantitative relationship between the structure of polymers and their flammability. Knowledge in structure-flammability relationships is useful for developing the mechanism of flame retardancy and predicting the fire hazard from polymeric materials. This paper presents a structure-property study for a series of polyphosphonates, modelled by their monomers. Monomer structures were studied by molecular mechanics calculations, using the MMFF94s force field included in the Omega software. Two types of chiral structures were found. Structural parameters were derived by the InstantJChem from Chemaxon and Dragon programs from the structures of minimum energy thus obtained. The influence of calculated structural descriptors of the studied polymers on their flammability (expressed by the limiting oxygen index) was modelled by multiple regression analysis (MLR) and artificial neural networks (ANN). MLR calculations were combined with genetic algorithm for variable selection. Several two descriptor MLR stable models were obtained and comparable ANN modelling results were noticed. Information on the nature of structural descriptors which influence the flammability was discussed.

Keywords: flame retardancy, polyphosphonate, MLR, ANN, Chemaxon, Omega

1. Introduction

Flame retardants are added to plastic materials to inhibit or suppress the combustion process and to improve the ignition or burning performance. One of the most well known parameter of polymer flammability is the limiting oxygen index (LOI). It expresses the minimum percentage of oxygen required to sustain ignition and

combustion. Flammability of commercial polymers became important in the last decades because of the widespread use of the materials (Liaw & Wang, 1996).

The importance of structure of polymers deciding the flammability has been recognized in the literature. Knowledge in structure-flammability relationships is useful for developing the mechanism of flame retardancy and predicting the flammability of polymers. Aromatic polyesters have been found in the last years to be technically interesting as engineering plastics, especially the phosphoric polyesters (polyphosphonates and polyphosphates). The materials are excellent in mechanical, electrical and flame resistance properties and also because of their analogy with the nucleic acids (Sandler & Karo, 1974; Iliescu et al., 1999). The phosphorous, halogen and nitrogen-containing polymers are the most widely used as commercial fire-retardant polymers (Iliescu et al., 1999).

This paper presents a structure-flammability study for a series of polyphosphonates, modeled by their monomers. Nearly all properties of the solid, liquid, and dissolved states of polymers can be predicted from chemical structure of monomers (van Krevelen & Nijenhuis, 2009). The monomer molecular structures were energetically optimized by molecular mechanics calculations. Several 0D, 1D, 2D and 3D descriptors were calculated from the minimum energy conformations and two types of data mining approaches, multiple regression analysis (MLR) and artificial neural networks (ANN) methods, were applied to model polymer flammability.

2. Materials and Methods

A series of 14 polyphosphonates (Annakuty & Kishore, 1988) of various types (Fig. 1) was considered, having the limiting oxygen index (LOI) as dependent variable.

Structural parameters

The polyphosphonates which were studied (Fig. 1) were modeled by their monomers in a first approximation. The monomer molecular structures were built by the Marvin Sketch 5.10 program (Chemaxon Ltd., Budapest Hungary) and energetically optimized by molecular mechanics calculations (MMFF94s force field) using the OMEGA software (OMEGA (version 2.4.6), OpenEye Science Software, 3600 Cerrillos Road, Suite 1107, Santa Fe, USA, 2008). Several 0D, 1D, 2D and 3D descriptors were calculated from the minimum energy conformations by various programs: InstantJChem (InstantJChem 5.10, Chemaxon Ltd., Budapest Hungary) and Dragon (Dragon Professional 5.5/2007, Talete S.R.L., Milano, Italy).

From the conformational search of each molecule the minimum energy structure

was used to derive structural descriptors. Thus, twenty-two types of descriptors were calculated by the Dragon software, like: constitutional, functional groups counts, topological descriptors, Burden eigenvalues, eigenvalue-based indices, Galvez descriptors (topological charge indicies), Getaway descriptors: ISH - standardized information content on the leverage equality, Randic descriptors (Randic molecular profiles), RDF descriptors (radial distribution function descriptors; MWC (Molecular walk counts path counts – atomic and molecular descriptors) and 3D-MoRSE (3D-molecule representation of structure based on electron diffraction descriptors): Mor02m - 3D-MoRSE - signal 02 / weighted by atomic masses (3D MoRSE descriptor), Mor13u - 3D-MoRSE - signal 13 / unweighted, atom-centred fragments, information indices, edge adjacency indices, topological charge indices, connectivity indices, 2D-autocorrelations, molecular properties, 2D binary fingerprints, and 2D frequency fingerprints.

Several structural descriptors were calculated by using InstantJChem software: Maximal Projection Area, Minimal Projection Radius, Average Polarizability, Van der Waals Surface Area, ASA Hydrophobic - solvent accessible surface area of all hydrophobic ($|q_i| < 0.125$) atoms ($|q_i|$ is the absolute value of the partial charge of the atom), ASA - solvent accessible surface area calculated using the radius of the solvent (1.4 Å for the water molecule), refractivity, logP – logarithm of octanol/water partition coefficient.

Multiple linear regression (MLR)

MLR calculations (Wold & Dunn III, 1983) were performed by the STATISTICA (STATISTICA 7.1, Tulsa, StatSoft Inc, OK, USA) and combined with genetic algorithm (Rogers & Hopfinger, 1994) in MobyDigs (Todeschini et al., 2004a) programs. The leave-one-out fitness function was used in our study as RQK fitness function (Todeschini et al., 2004b).

Artificial neural networks (ANNs)

The artificial neural networks have an inherent ability to provide non-linear and cross product terms for QSAR modeling. The ANNs are especially useful when a rigid theoretical basis and/or mathematical relationship to describe a phenomenon to be modeled are not available in advance.

The three-layer ANNs with the back-propagation of errors (Zupan & Gasteiger, 1999) were employed in this study. ANNs calculations were carried out by using our inhouse program. The most commonly used log sigmoid function and the delta rule for the error correction formula were used in the networks.
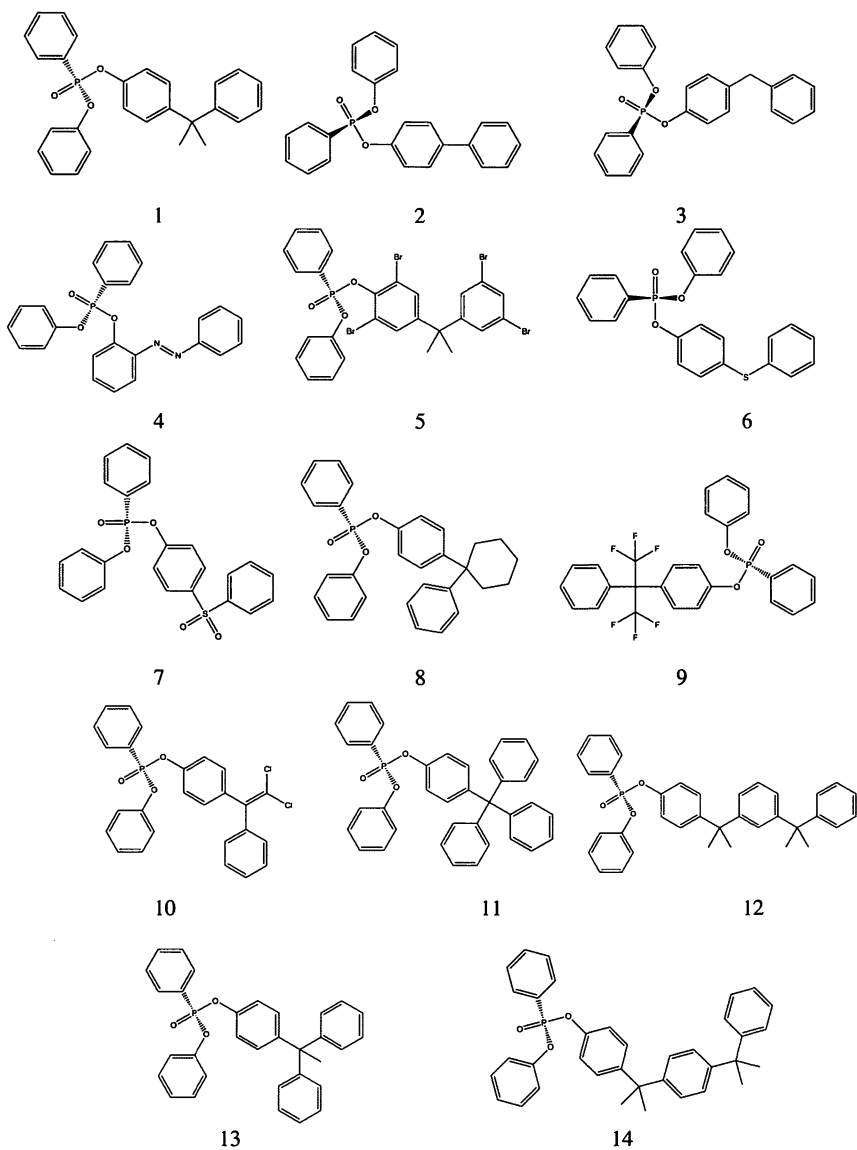
Fig. 1. Polyphosphonate structures

Model validity

All the statistical tests were performed at a significance level of 5 % or less. In GA-MLR calculations outliers were tested by estimating the standardized residuals of less than -3.0 or more than +3.0 (Frank & Althoen, 1995) and by the value of residual greater than three times the value of standard error in calculation (Todeschini & Consonni, 2000), as implemented in the MobyDigs program. The Kubinyi fitness

function (Kubinyi, 1994) was, also, used to check the goodness of fit of the obtained GA-MLR models.

To avoid models with collinearity without prediction power, regression models were calculated only for variable subsets with an acceptable multivariate correlation applying the QUIK (Q Under Influence of K) procedure based on the K multivariate correlation index (Todeschini et al., 1999). Only models with a global correlation of [XY] block (K_{XY}) greater than the global correlation of the X block (K_{XX}) variable can be accepted, where X is the descriptor matrix and Y is the dependent variable. To each model, the K_{XY} and K_{XX} values were calculated. Several commonly used statistic terms were adopted to check the reliability, robustness and stability of the proposed model such as correlation coefficient (r^2), leave-one-out (LOO) cross-validated q^2, root mean square error for the training set (*SDEC*) and predictive set (*SDEP*) (Consonni et al., 2009). The goodness of prediction of the GA-MLR models was, also, checked by the Akaike Information Criterion (*AIC*) (Gentleman & Wilk, 1975), and Y-scrambling (Lindgren et al., 1996). Y scrambling was applied to exclude the possibility of chance correlation and to check for reliability and robustness by permutation testing: new models were recalculated for randomly reordered responses (Y scrambling). The resulting models obtained with randomized responses should have significantly lower q^2 values than the proposed ones because the relationship between the structure and response is broken. Y scrambling was performed by response scrambling with maximum iterations of 500, and then the mean values of $R^2_{Yscrambling}$ ($a(r^2)$) and $Q^2_{Yscrambling}$ ($a(q^2)$) were reported. All these calculations were performed by the MobyDigs software. The leave-one-out cross-validation procedure (Wold, 1978) was also employed to check the robustness of the model.

Tools of regression diagnostic as residual plots and Williams plots were used to check the quality of the best models and define their applicability domain using the Mobydigs software. Residual plot shows validated residuals versus response values and enables the search for outliers and to verify the assumption of the GA-MLR method on the normal error distribution, therefore this plot is a tool to evaluate the existence of a linear relation between variables and response. Leverages of test compounds were calculated to check their distance from the model experimental space; the greater the distance the more unreliable the predicted response (Frank & Todeschini, 1994).

Several statistical measures to test the model fitting quality were derived from the calculated LOI values (e.g. *RMSE*-the root mean square error (Goodarzi, 2009), *NRMSE*-the normalized root mean square, *RMSECoef*-the coefficient of variation of the *RMSE*, *RSE*(%)-the relative standard error of prediction(Goodarzi, 2009), and *MAE*(%)- the mean absolute error (Goodarzi, 2009) and, respctively from the leave-one-out predicted LOI values (the corresponding *RMSEP*,

NRMSEP, RMSEPCoef, RSEP(%), MAEP(%) criteria):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{n}} \tag{1}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{2}$$

$$RMSECoef = \frac{RMSE}{\frac{\sum_{i=1}^{n} y_{obs}}{n}} \tag{3}$$

$$RSE(\%) = 100 \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{\sum_{i=1}^{n}(y_{obs})^2}} \tag{4}$$

$$MAE(\%) = \frac{100}{n} \sqrt{\sum_{i=1}^{n}|(y_{pred} - y_{obs})|} \tag{5}$$

where y_{obs} is the observed value of the dependent variable in the sample, y_{pred} the calculated/predicted value of the dependent variable (either as internal, cross-validated, or external, test set prediction), n the number of samples in the set, y_{max} and y_{min} the maximum respectively minimum value of the dependent variable.

3. Results and Discussion

Two types of chiral structures were found by conformational analysis for each compound and conformers of minimum energy were further used. Structural parameters were derived by InstantJChem and Dragon programs from the structures of minimum energy thus obtained. MLR calculations were performed for each type of isomer. Variable selection was carried out by the genetic algorithm, using the leave-one-out fit criterion as constrained function to be optimized. Satisfactory MLR models was obtained for R (Table 1) and for S isomer (Table 2).

Best MLR model for R isomer was MLR9R and for S isomer: MLR3S. Better fitting results and stable models were noticed in case of R isomers models.

The applicability domain of the best MLR9R model was checked by the Williams plot (Figure 2). This plot confirms the absence of outliers and influential points (the leverage average value = 0.214).

Figures 3 and 4 presents the experimental versus calculated, respectively, predicted LOI values derived from model MLR9R.
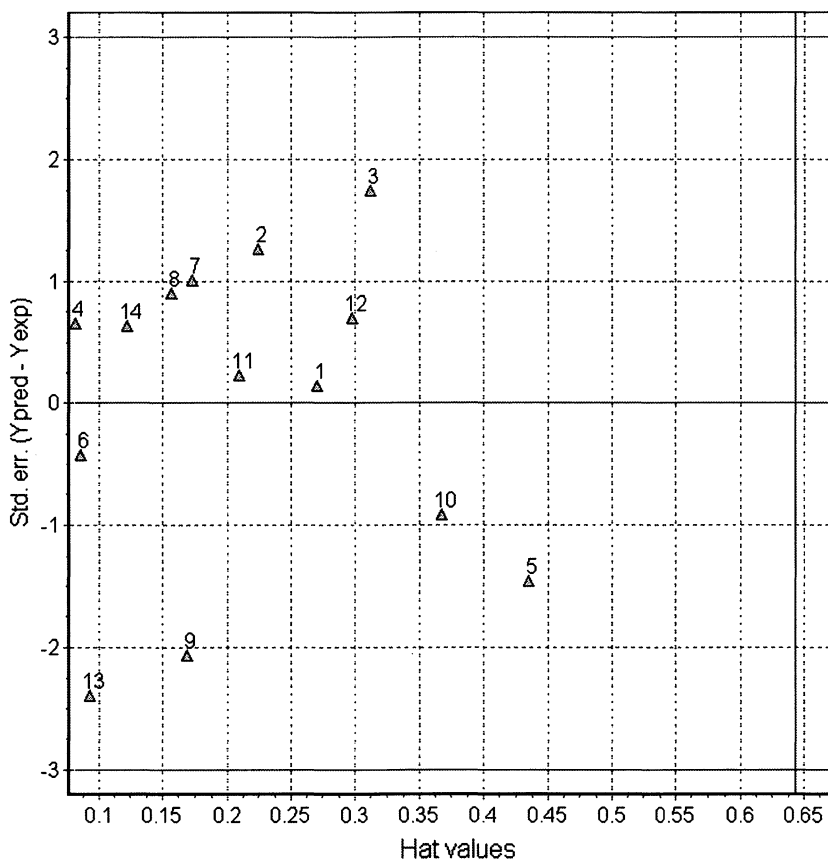

Fig. 2. Williams plot: jackknifed residuals of the MLR9R model versus leverages.

Table 1. MLR and ANN models obtained for the R isomer

ID	Models	r^2	q^2	q^2_{boot}	$a(r^2)$	$a(q^2)$	r^2_{adj}	AIC	K_x	K_{xy}	$SDEP$	$SDEC$	F	s
MLR1R	Maximal Projection Area, ISH	0.883	0.825	0.802	0.364	-0.019	0.862	15.764	40.93	47.65	3.462	2.831	41.52	3.194
MLR 2R	Average Polarizability, ISH	0.848	0.765	0.751	0.249	-0.293	0.820	20.547	25.75	46.53	4.013	3.232	30.57	3.646
MLR 3R	Van der Waals Surface Area, ISH	0.819	0.716	0.687	0.446	0.091	0.786	24.352	42.17	46.29	4.41	3.519	24.94	3.97
MLR 4R	ASA Hydrophobic, ISH	0.819	0.711	0.703	0.338	0.037	0.786	24.369	36.37	46.12	4.449	3.52	24.92	3.971
MLR 5R	ASA, ISH	0.811	0.712	0.690	0.245	-0.197	0.776	25.492	32.36	45.81	4.445	3.6	23.58	4.061
MLR 6R	Refractivity, ISH	0.855	0.762	0.721	0.300	-0.218	0.828	19.59	29.21	46.78	4.037	3.156	32.34	3.56
MLR 7R	Minimal Projection Radius, Mor02m	0.802	0.722	0.716	0.255	-0.194	0.767	26.623	27.02	45.6	4.366	3.679	22.34	4.151
MLR 8R	Average Polarizability, Mor13u	0.846	0.762	0.723	0.381	-0.012	0.818	20.785	47.08	47.18	4.04	3.251	30.16	3.667
MLR 9R	logP, ISH	0.887	0.830	0.822	0.166	-0.444	0.866	15.281	18.81	47.35	3.417	2.787	43.01	3.144
ANN9R	ANN	0.870	-	-	-	-	-	-	-	-	-	-	-	-

* r^2 – Correlation coefficient, q^2 – leave-one-out crossvalidation parameter, q^2_{boot} – bootstrapping parameter, $a(r^2)$ and $a(q^2)$ – Y-scrambling variables, r^2_{adj} – adjusted r^2, SDEP – standard deviation error in prediction, $SDEC$– standard deviation error in calculation, F– Fischer test, s – standard error of estimate, AIC – Akaike Information Criterion, the multivariate K correlation indices (K_x – the multivariate correlation index of the matrix of X descriptors and K_{xy} – the multivariate correlation index of the matrix of X descriptors and Y response variable), FIT– the Kubinyi fitness function,

**Table 2.** MLR and ANN models obtained for the S isomer

ID	Models	r^2	q^2	q^2_{boot}	$a(r^2)$	$a(q^2)$	r^2_{adj}	AIC	K_x	K_{xy}	$SDEP$	$SDEC$	F	s
MLR1S	Average Polarizability, ISH	0.778	0.648	0.579	0.113	-0.460	0.737	29.973	7.71	44.26	4.909	3.904	19.23	4.404
MLR2S	Minimal Projection Radius, GATS1m	0.678	0.510	0.463	0.104	-0.567	0.620	43.354	3.11	42.49	5.792	4.695	11.6	5.296
MLR3S	Minimal Projection Radius, Mor02m	0.841	0.790	0.777	0.249	-0.200	0.812	21.499	27.71	46.5	3.793	3.306	28.98	3.73
ANN3S	ANN	0.82	-	-	-	-	-	-	-	-	-	-	-	-

* r^2 – Correlation coefficient, q^2 – leave-one-out crossvalidation parameter, q^2_{boot} – bootstrapping parameter, $a(r^2)$ and $a(q^2)$ – Y-scrambling variables, r^2_{adj} – adjusted r^2, $SDEP$ – standard deviation error in prediction, $SDEC$ – standard deviation error in calculation, F – Fischer test, s – standard error of estimate, AIC – Akaike Information Criterion, the multivariate K correlation indices (K_x – the multivariate correlation index of the matrix of X descriptors and K_{xy} – the multivariate correlation index of the matrix of X descriptors and Y response variable), FIT – the Kubinyi fitness function.

Table 3. Calculated criteria for MLR and ANN model goodness of fitting

Model	RMSE	RMSEP	RMSENorm	RMSEPNorm	SMSECoef	RMSEPCoef	RSE	RSEP	MAE	MAEP
MLR1R	2.83	3.46	0.10	0.13	0.07	0.09	7.04	8.60	42.48	47.28
MLR 2R	3.23	4.01	0.12	0.15	0.08	0.10	8.04	9.97	43.99	49.57
MLR 3R	3.52	4.41	0.13	0.15	0.09	0.10	8.75	10.97	46.62	52.68
MLR 4R	3.52	4.45	0.13	0.16	0.09	0.11	8.75	11.06	47.20	36.44
MLR 5R	3.60	4.44	0.13	0.16	0.09	0.11	8.95	11.05	47.14	52.96
MLR 6R	3.16	4.04	0.12	0.15	0.08	0.10	7.85	10.04	43.28	49.31
MLR 7R	3.68	4.37	0.14	0.16	0.09	0.11	9.15	10.85	46.12	51.43
MLR 8R	3.25	4.04	0.12	0.15	0.08	0.10	8.08	10.04	45.31	51.03
MLR 9R	2.79	3.42	0.10	0.13	0.07	0.09	6.93	8.50	40.30	45.29
ANN9R	2.94	3.64	0.11	0.13	0.07	0.10	7.32	9.05	41.59	46.07
MLR1S	3.90	4.91	0.14	0.18	0.10	0.12	9.71	12.20	47.42	53.61
MLR2S	4.70	5.79	0.17	0.21	0.12	0.15	11.67	14.40	53.61	60.49
MLR3S	3.31	3.79	0.12	0.14	0.08	0.10	8.22	9.43	41.75	45.42
ANN3S	3.51	4.53	0.13	0.17	0.09	0.12	8.72	11.27	45.79	52.39

* *RMSE* – root mean squared error, *RMSEP* – root mean squared error of prediction (from LOO cross-validation), *RMSENorm* - Normalized *RMSE*, *RMSEPNorm* - Normalized *RMSEP*, *RMSECoef* – coefficient of variation of the *RMSE*, *RMSEPCoef* – coefficient of variation of the *RMSEP*, *RSE* (%) – relative standard error of fitting, *RSEP* (%) – relative standard error of prediction, *MAE* (%) – mean absolute error of fitting, *MAEP* (%) – mean absolute error of prediction
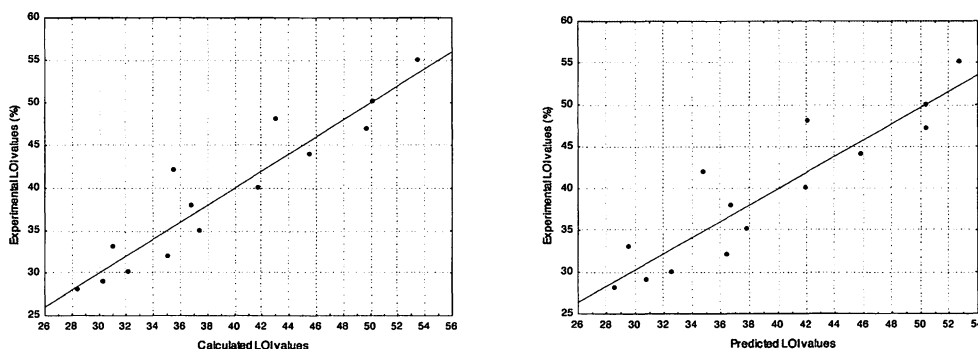
Fig. 3. Experimental versus calculated (left), respectively predicted (right) LOI values for the MLR9R model.

GETAWAY (Geometry, Topology, and Atom-Weights AssemblY) descriptors encode both geometrical information given by the Molecular Influence Matrix (which in turn takes into account the relative position of atoms in a molecular structure optimized in some way) and the topological information given by the molecular graph, weighted by chemical information encoded in selected atomic weights.

The best set of molecular descriptors included in the above MLR model was used to develop the nonlinear models by ANNs. ANNs gave worse results in comparison to the MLR model, indicating a preferred linear fitting. The obtained "tentative" models allow a rough estimation and factors that influence the polymer flammability.

The calculated goodness of fit criteria presented in Tables 1 and 2 indicated the MLR9R as a stable model to simulate polymer flammability.
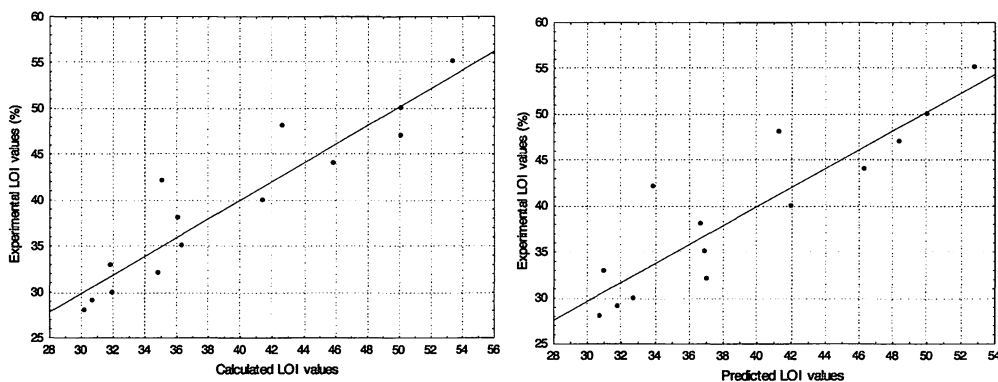

Fig. 4. Experimental versus calculated (left), respectively predicted (right) LOI values for the ANN9R model.

4. Conclusion

A structure-flammability study was developed for a series of 14 polyphosphonates. Their monomer structures were studied by molecular mechanics calculations, using the MMFF94s force field. Two types of chirality were found by conformational analysis performed for each compound. Several descriptors derived from the minimum energy structures were related by multiple linear regression and neural networks to their flammability, expressed by the LOI values. Several criteria of goodness of fit were calculated for these models. Better statistical results and stable models were obtained for the R isomers. Increased polymer hydrophobicity favors higher flammability. The geometry of R isomers is favorable for the flammability too.

Acknowledgements

References

Annakuty, K.S.; Kishore, K. (1988) A novel approach to structure in polyphosphonate ester, *Polymer,* **29**, 1273-1276.

Consonni, V.; Ballabio, D.; Todeschini, R. (2009) Comments on the definition of the Q2 parameter for QSAR validation, *J. Chem. Inf. Model.,* **49**, 1669–1678.

Frank I. E.; Todeschini R. (1994) *The Data Analysis Handbook.* Elsevier, Amsterdam, The Netherlands.

Frank, H.; Althoen, S. C. (1995) Outliers. In *Statistics: Concepts and Applications,* Cambridge University Press, Cambridge, Great Britain, pp. 142-143.

Gentleman, J.F.; Wilk, M.B. (1975) Detecting outliers. II. Supplementing the direct analysis of residuals. *Biometrics.,* **31**, 387-410

Goodarzi, M.; Deshpande, S; Murugesan, V.; Katti, S.B.; Prabhakar, Y.S. (2009) Is feature selection essential for ANN modeling? *QSAR Comb. Sci.,* **28**, 1487-1499.

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, *J. Chem. Inf. Model.,* **50**, 572-584.

Hawkins P. C. D.; Nicholls A. (2012) Conformer generation with OMEGA: Learning from the data set and analysis of failures, *J. Chem. Inf. Model.,* **52**, 2919-2936.

Iliescu, S.; Ilia, G.; Dehelean, G.; Popa, A.; Macarie, L. (1999) *Polimeri organici cu fosfor,* Ed. Omega, Bucureşti.

Kubinyi, H. (1994) Variable selection in QSAR studies. I. An evolutionary algorithm. *Quant. Struct-Act. Rel.,* **13**, 285–294

Liaw, D.; Wang, D.-W. (1996) Synthesis of fluorine-containing polyphosphates: low-temperature solution polycondensation of bisphenol AF and aryl phosphorodichloridates. *React. Func. Polym.,* **30**, 309-315.

Lindgren, F.; Hansen, B.; Karcher, W.; Sjöström, M.; Eriksson L. (1996) Model validation by permutation tests: Applications to variable selection. *J. Chemometr.,* **10**, 521-532.

Rogers, D.; Hopfinger, A.J.J. (1994) Application of Genetic Function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, *J. Chem. Inf. Comput. Sci.,* **34**, 854-866.

Sandler, S. R.; Karo, W.(1974) *Polymer Synthesis,* Vol. 1, Academic press, New York.

Todeschini, R.; Consonni, V.; Maiocchi A. (1999) The K correlation index theory development and its application in chemometrics. *Chemom. Intell. Lab. Syst.,* **46**, 13-29.

Todeschini, R.; Consonni V. (2000) *Handbook of Molecular Descriptors.* Wiley, Weinheim, pp. 369

Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. (2004a) In Leardi, R. editor. *Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks.* Chapter 5, Elsevier, Amsterdam, pp. 141-167

Todeschini. R.; Consonni, V.; Mauri, A. ; Pavan, M. (2004b) Detecting "bad" regression models: multicriteria fitness functions in regression analysis. *Anal. Chim. Acta,* **515**, 199-208.

van Krevelne, D.W.; Nijenhuis, Klaas te (2009) *Properties of Polymers: Their Correlation with Chemical Structure; their Numerical estimation and Prediction from Additive Group Contributions,* Elsevier, Amsterdam.

Wold, S.; Dunn Ⅲ, W. J. (1983) Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comp. Sci.,* **23**, 6-13

Zupan, J.; Gasteiger, J. (1999) *Neural Networks for Chemistry and Drug Design,* 2nd Edition, Wiley-VCH, Weinheim

要　　旨

ポリホスホン酸エステル類の化学構造と燃焼性の定量的な相関の検討

鈴 木 孝 弘 , Simona FUNAR-TIMOFEI, Smaranda ILIESCU

高分子化合物材料の難燃化の設計にあたっては、ポリマーの構造と燃焼性との相関を明らかにする必要がある。本研究では、14種類のポリホスホン酸エステル類の構造と燃焼性との相関をモノマーの構造から検討した。モノマーの構造は、MMFF94力場を用いて分子力学法によってエネルギー的に安定な構造を見出した。その結果、2種類のキラルな構造が存在することが分かった。ポリホスホン酸エステル類の分子構造に関するパラメータ（分子記述子）は、ChemaxonとDragonプログラムによって計算し、燃焼性との相関を重回帰分析とニューラルネットワークを用いて検討した。2種の分子記述子を用いた重回帰分析モデルが統計的に有意であることを見出し、非線形のニューラルネットワークモデルと比較した。さらに、燃焼性に影響を及ぼす分子記述子についてその物理化学的な意味を検討した。本研究によって、ポリマーの疎水性や立体化学的な特性が、燃焼性と関係していることが明らかになった。