# QSTR STUDY OF ORGANIC PHOSPHONIUM SALTS BY MLR

**SIMONA FUNAR-TIMOFEI, ADRIANA POPA**

Institute of Chemistry of the Romanian Academy

24 Mihai Viteazul Bvd., 300223 Timisoara, Romania

e-mail: timofei@acad-icht.tm.edu.ro

# INTRODUCTION

- **Polyethylene glycols (PEGs) are polymers of ethylene oxide with the generalized formula HO(CH2CH2 O)*n*-H, "*n*" indicating the average number of oxyethylene groups are used as cleansing agents, emulsifiers, skin conditioners, and humectants [1].**

- **Many insoluble disinfectants reported are phosphonium salts grafted on polymer [2]**

**[1]. Fruijtier-Polloth, C. Toxicology 2005; 214: 1–38.**
**[2]. Kanazawa, A.; Ikeda, T.; Endo T. J. Polym. Sci. Pol. Chem. 1994; 32: 1997-2001.**

# INTRODUCTION

- Polymeric disinfectants have important applications, such as: antifouling coatings and fiber finishing, drugs with prolonged activity and less toxicity, water and air disinfection [3].

- According to the toxicity scale of Hodge and Steaner the poly(oxyethylene)s functionalized with quaternary phosphonium end groups can be considered as low toxic compounds [4]

[3]. Kanazawa, A.; Ikeda, T.; Endo, T. J. Appl. Polym. Sci. 1994; 53: 1237-1244. [4]. Popa, A. ; Trif, A. ; Curtui, V.G. ; Dehelean, G. ; Iliescu, S. ; Ilia G. Phosphorus Sulfur. 2002; 177: 2195-2196.

# *AIM:*

- **0D, 1D and 2D descriptors of organic phosphonium salts were related to their logarithm of oral mouse $LD_{50}$ values to find out structural features which influence their toxicity.**

# METHODS

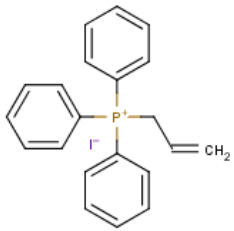
- Twenty eight quaternary phosphonium salts derivatives with known toxicity, the logarithm of the lethal oral dose for mouse $LD_{50}$ (taken from RTECS Database, MDL Information Systems, Inc. 14600 Catalina Street San Leandro, California U.S.A. 94577, http://www.ntis.gov/products/types/databases/rtecs.asp) were used

# Phosphonium salt training structures

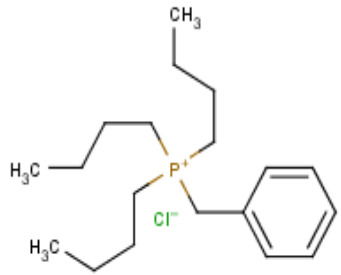
***Table 1*. Name and the logarithm of the LD50 values of phosphonium salt structures**

| No | Phosphonium salt name | No | Phosphonium salt name |
|---|---|---|---|
| 1 | Phosphonium, acetonyltriphenyl-, iodide | 16 | Phosphonium, (2,4-dimethylbenzyl)tributyl-, chloride |
| 2 | Phosphonium, tributyl-2-propen-1-yl-, chloride | 17 | Phosphonium, (2,4-dichlorobenzyl)triphenyl-, iodide |
| 5 | Phosphonium, benzyltriphenyl-, iodide | 18 | Phosphonium, (2,4-dichlorobenzyl)tri(p-tolyl)-, chloride |
| 6 | Phosphonium, bis(p-butylamino)benzylphenyl-, iodide | 19 | Phosphonium, (dichloromethyl)tripiperidino-, perchlorate |
| 7 | Phosphonium, bis (t-butylamino)methylphenyl-, iodide | 20 | Phosphonium, (ethoxycarbonylmethyl)triphenyl-, bromide |
| 9 | Phosphonium, (p-bromomethylbenzyl)triphenyl-, bromide | 21 | Phosphonium, (2-ethoxypropenyl)triphenyl-, iodide |
| 10 | Phosphonium, butyltriphenyl-, bromide | 22 | Phosphonium, ethyltriphenyl-, iodide |
| 11 | Phosphonium, butyltriphenyl-, iodide | 23 | Phosphonium, (o-methylbenzyl)triphenyl-, bromide |
| 12 | Phosphonium, carboxymethyltriphenyl-, chloride | 24 | Phosphonium, p-nitrobenzyltributyl-, iodide |
| 13 | Phosphonium, (p-chloromethylbenzyl)tris (dimethylamino)-, chloride | 27 | Phosphonium, (3-phenoxypropyl)triphenyl-, bromide |
| 14 | Phosphonium, chloromethyltriphenyl-, chloride | | |

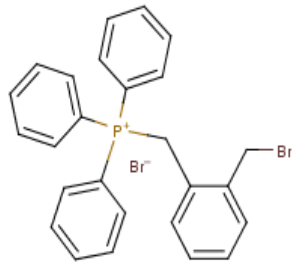
# Phosphonium salt test structures



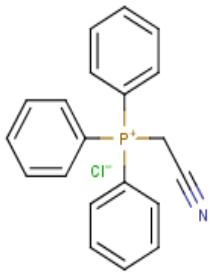

3



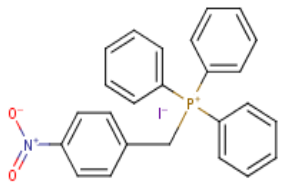

4



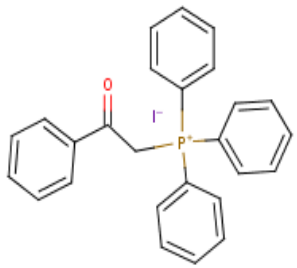

8



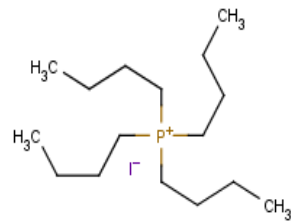

15




25




26




28

# METHODS

- **Phosphonium salts structure (modeled as cations) was built by the ChemOffice package (ChemOffice 6.0, CambridgeSoft.Com, Cambridge, MA, U.S.A.) and energetically optimized using the molecular mechanics approach.**

- **Twenty-two types of descriptors were calculated by the Dragon software (Dragon Professional 5.5/2007, Talete S.R.L., Milano, Italy)**

# METHODS

- **Multiple linear regression (MLR) calculations were performed by the STATISTICA (STATISTICA 7.1, Tulsa, StatSoft Inc, OK, USA) and MobyDigs [5] programs.**

- **The goodness of prediction of the MLR models was checked by the Akaike Information Criterion (AIC), the multivariate K correlation index , Y-scrambling and external validation parameters.**

**[5]. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. MobyDigs: software for regression and classification models by genetic algorithms, in: 'Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks'. (Leardi R., Ed.), Chapter 5, Elsevier, 2004, pp. 141-167.**

# RESULTS AND DISCUSSION

- Variable selection was carried out by the genetic algorithm included in the MobyDigs program, using the RQK fitness function [6], with leave-one-out crossvalidation correlation coefficient as constrained function to be optimised, a crossover/mutation trade-off parameter T = 0.5 and a model population size P = 50.

- The leave-one out cross-validation procedure was employed for the internal validation of models.

[6]. Todeschini R., Consonni V., Mauri A., Pavan M. Anal. Chim. Acta 2004; 515: 199-208.

# RESULTS AND DISCUSSION

## *Table 2*. MLR results (selection)*

| No | Descriptors | $r^2$ | $q^2_{LOO}$ | $q^2_{boot}$ | $q^2_{ext}$ | $r^2_{Y-scrambling}$ | $q^2_{Y-scrambling}$ | AIC | Kx | Kxy | SDEP | SDEC | F | s |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | P2e<br>HATS3m<br>HATS6m<br>REIG | 0.863 | 0.782 | 0.707 | 0.951 | 0.237 | -0.498 | 0.138 | 0.26 | 0.40 | 0.321 | 0.254 | 25.26 | 0.291 |
| 2 | PW5<br>RDF030u<br>RDF045u<br>Mor05e | 0.862 | 0.763 | 0.717 | 0.690 | 0.379 | -0.488 | 0.139 | 0.47 | 0.56 | 0.334 | 0.255 | 24.9 | 0.293 |
| 3 | E3m<br>HATS3m<br>H1e<br>R7v+ | 0.860 | 0.777 | 0.712 | 0.757 | 0.341 | -0.302 | 0.141 | 0.29 | 0.45 | 0.325 | 0.257 | 24.57 | 0.294 |
| 4 | P2p<br>HATS3m<br>HATS6m<br>REIG | 0.856 | 0.768 | 0.684 | 0.943 | 0.28 | -0.39 | 0.145 | 0.28 | 0.42 | 0.331 | 0.261 | 23.68 | 0.299 |
| 5 | PW5<br>RDF045u<br>Mor05e<br>HATS6m | 0.855 | 0.749 | 0.690 | 0.699 | 0.338 | -0.226 | 0.146 | 0.45 | 0.54 | 0.344 | 0.261 | 23.62 | 0.299 |
| 6 | P2e<br>HATS6m<br>REIG<br>R4m+ | 0.854 | 0.794 | 0.718 | 0.911 | 0.314 | -0.32 | 0.146 | 0.27 | 0.37 | 0.312 | 0.262 | 23.46 | 0.3 |

* $r^2$ – correlation coefficient, SDEP – standard deviation error in prediction ($RMSE_{test}$), SDEC – standard deviation error in calculation ($RMSE_{training}$), F- Fischer test, s – standard error of estimate, AIC - Akaike Information Criterion, the multivariate K correlation index (Kx and Kxy), Y-scrambling variables ($r^2_{Y-scrambling}$ and $q^2_{Y-scrambling}$), $q^2_{ext}$ - external $q^2$, $q^2_{boot}$ - bootstrapping parameter, $q^2_{LOO}$ - leave-one out cross-validation parameter

# RESULTS AND DISCUSSION

- **Starting from the descriptor matrix containing all variables, following descriptors were found to be significant and were included in the final MLR models: topological, walk and path count, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, topological charge indices, eigenvalue-based indices, RDF descriptors, 3D-MoRSE, WHIM descriptors, Getaway descriptors, and molecular properties**

# RESULTS AND DISCUSSION

- **Model 1 (Table 2) was selected as the best single model:**

$$\log LD_{50} = 2.36(\pm 0.99) + 2.36(\pm 1.31)P2e - 12.37(\pm 2.88)HATS3m + 4.87(\pm 1.15)HATS6m - 11.28(\pm 1.6)REIG$$

$$N_{training} = 21\ N_{test} = 7$$

$$r^2_{training} = 0.863\ q^2_{LOO} = 0.782\ q^2_{ext} = 0.951\ r^2_{Y-scrambling} = 0.251\ q^2_{Y-scrambling} = -0.348$$

$$K_{XY} = 0.402\ K_X = 0.258\ RMSE_{training} = 0.254\ RMSE_{test} = 0.321$$

- **where P2e-2nd component shape directional WHIM index / weighted by atomic Sanderson electronegativities, HATS3m-leverage-weighted autocorrelation of lag 3 / weighted by atomic masses, HATS6m-leverage-weighted autocorrelation of lag 6 / weighted by atomic masses; REIG-first eigenvalue of the R matrix**

# RESULTS AND DISCUSSION

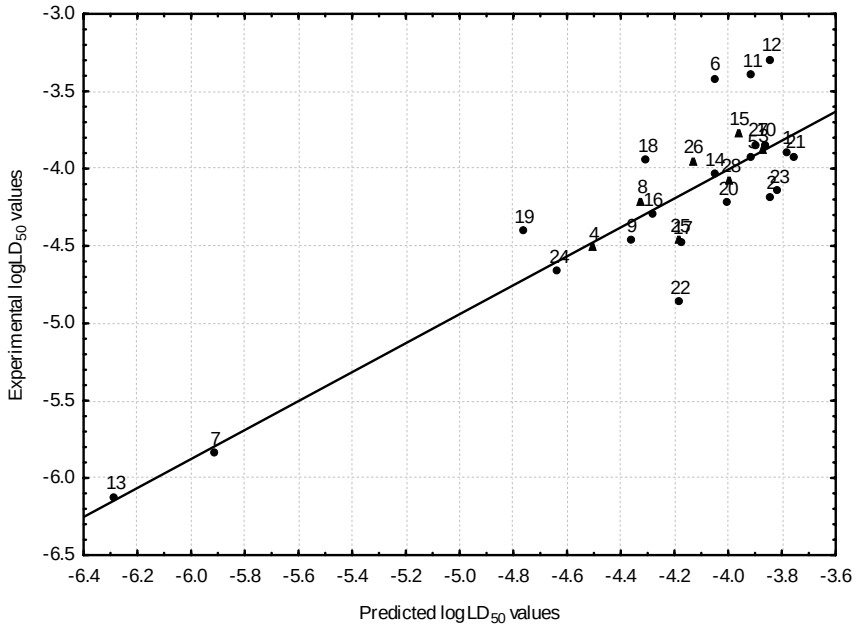



***Figure 1*. Experimental *versus* predicted $logLD_{50}$ values of the final MLR model 1 (Table 2). Training set is marked by circles, test set marked by blue triangles.**

# RESULTS AND DISCUSSION

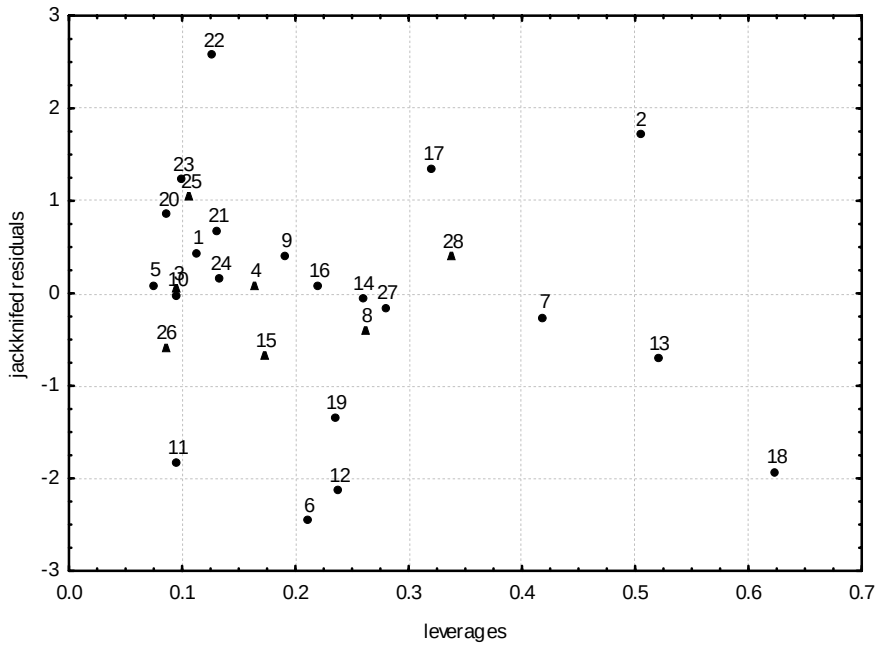



***Figure 2*. Williams plot: jackknifed residuals *versus* leverages of the MLR model 1 (Table 2). Training set is marked by circles, test set marked by triangles (leverage control value of 0.714)**

# CONCLUSIONS

- The quaternary phosphonium salts toxicity was modeled by MLR combined with genetic algorithm for variable selection, with acceptable statistical results
- Electronic distribution is very important for the phosphonium salts toxicity.
- Steric factors of phosphonium salts can be considered to influence the toxicity.