

COMPARISON OF QSPR METHODS AND DESCRIPTORS FOR THE PREDICTION OF EXPERIMENTAL HALF-LIVES OF BROMINATED FLAME RETARDANTS, THEIR METABOLITES AND CHLORO/BROMOPHENOLS

Harju Mikael¹, Bastos Patricia², Bergman Åke² and Heimstad Eldbjørg Sofie¹

¹ NILU, Polar Environmental Centre, 9296 Tromsø, Norway; ² Department of Environmental Chemistry, Stockholm University, 10691 Stockholm, Sweden

Introduction

The higher the persistency is of a chemical the longer the compound will stay in the environment. Persistency is made up of the intrinsic properties of a compound, i.e. its physicochemical and chemical reactivity characteristics. The persistency is the abiotic parameter of a chemical that relate to the occurrence, partitioning and distribution of a chemical in each of the environmental compartments. Persistency is a key property for assessing the environmental fate of chemicals. The chemical reactivity of a compound is strongly influencing its bioaccumulative potential. The use of theoretical methods as alternative methods to estimate chemical characteristics has grown and become of importance for regulation, evaluation and risk assessment of chemicals¹. Multimedia fate models are increasingly applied to understand and predict the environmental fate of organic contaminants². Lately, such models have based their parameterisation of chemicals on linear solvation energy relationships (LSER)³ that account for molecular size and various solubility descriptors.

We aim at developing user-friendly QSPRs where molecular descriptors are easily accessible and can be interpreted and understood on a chemical basis. In addition, the equations should have relevance and be attractive for multimedia modellers. Applying this integrated predictive methodology with the use of experimental reactivity and *in silico* methods should lead to major improvements in risk assessments of chemicals and to provide a tool for improved management of chemicals.

Materials & Methods

Experimental methods

The experimental data provided for this study have been measured experimentally as described elsewhere^{4,5}. The experimental data, shown in **Table 1**, was created by a single laboratory with the same method of oxidation and at a pH of 7.6. The differences in temperatures and solvent mixture (methanol and water) were present likely due to the large differences in reaction rates and solubility of these compounds. To normalize the data set we extrapolated the data to the same temperature (298K) using the Arrhenius equation and verified by duplicate experiments made for a few compounds at different temperatures while possible effects due to differences in solvent mixture could not be normalized.

Theoretical methods

Molecular structures were generated in the molecular modelling software Cache WorkSystem pro (Fujitsu Limited) or downloaded from ChemIdPlus (<http://chem.sis.nlm.nih.gov/chemidplus/>). The 3-dimensional structures of each chemical were geometry optimised with the use of energy minimisation with the semi-empirical quantum mechanics MOPAC AM1 method. Three main methods were applied to calculate molecular descriptors; the Cache MOPAC AM1, the fragment based Linear Solvation Energy Relationship (LSER) method⁶ with the software ADME boxes v. 3.5 (Pharma Algorithms, Inc.) and the semi-empirical quantum mechanics based method TLSER⁷ using the software of MOPAC⁸ and MADCAP⁷ (kindly provided by Leland Y. Wilson).

Table 1. QSPR data-base with temperature normalized experimental data^a containing logarithmically transformed oxidation half-lives, calculated physicochemical data for all variables included in MLR models and predicted results for the regressions based on all variables^b and TLSE^c variables.

Chemical	Abbreviation	Physicochemical variables							Oxidation half-lives (log(1/t _{1/2}))					
		pKa	B	Vm	Ea	Qh+	TotDip	H _{charge}	ΔE _{OH}	Exp. ^a	MLR ^b	PLS ^b	MLR ^c	PLS ^c
Training set														
2-bromophenol	2-BP	8.1	0.31	1.03	0.12	0.10	6.19	0.08	-8.0	-1.96	-1.80	-2.02	-1.87	-1.88
2,4-dibromophenol	24-DBP	8	0.23	1.27	0.18	0.23	4.65	0.10	-8.5	-2.36	-2.55	-2.68	-2.62	-2.69
2,6-dibromophenol	26-DBP	6.8	0.22	1.24	0.13	0.10	6.43	0.10	-8.4	-1.96	-2.59	-2.57	-2.57	-2.75
2,4,6-tribromophenol	246-TBP	6.7	0.15	1.45	0.13	0.12	6.77	0.12	-8.8	-2.79	-3.31	-3.16	-2.41	-3.06
2,3,4,6-tetrabromophenol	2346-TeBP	6.3	0.07	1.69	0.18	0.24	5.21	0.12	-9.0	-4.14	-3.54	-3.45	-2.96	-3.68
2,3,4-trichlorophenol	234-TCP	8.1	0.15	1.24	0.13	0.12	5.53	0.10	-8.5	-2.96	-2.71	-3.00	-2.45	-2.39
Bisphenol-A	BPA	10.1	0.91	2.08	0.17	0.22	9.35	0.08	-13.7	-1.96	-2.00	-1.63	-2.29	-1.99
2,2',4-trichlorobisphenol-A	TCBPA	6.5	0.4	2.64	0.15	0.22	10.6	0.05	-8.6	-1.15	-1.07	-1.17	-2.02	-1.81
3,3'-dibromodiphenyl ether	bde11	14	0.31	1.98	0.18	0.17	5.18	0.14	-14.2	-5.38	-4.69	-4.85	-4.90	-4.40
2,4,4'-tribromodiphenyl ether	bde28	14	0.32	2.18	0.18	0.17	6.54	0.15	-14.3	-5.22	-4.94	-5.05	-4.81	-4.87
2,2',4,5'-tetrabromodiphenyl ether	bde49	14	0.25	2.38	0.18	0.17	6.32	0.16	-14.3	-4.96	-5.33	-5.25	-5.22	-5.34
2,3',4,4'-tetrabromodiphenyl ether	bde66	14	0.25	2.37	0.18	0.17	6.40	0.16	-14.3	-4.87	-5.32	-5.29	-5.17	-5.23
2,2',4,4',5-pentabromodiphenyl ether	bde99	14	0.18	2.61	0.19	0.18	6.95	0.16	-14.4	-5.33	-5.48	-5.48	-5.62	-5.72
2,2',4,4',5,5'-hexabromodiphenyl ether	bde153	14	0.11	2.76	0.19	0.18	6.80	0.17	-14.5	-5.70	-5.89	-5.90	-5.93	-6.09
6-OH-2,2',4-tribromodiphenyl ether	6-OH-bde17	7.7	0.38	2.24	0.15	0.17	8.26	0.12	-9.0	-2.51	-2.89	-2.76	-3.10	-2.64
6-OH-2,2',4,4'-tetrabromodiphenyl ether	6OH-BDE47	7.6	0.37	2.43	0.15	0.18	11.4	0.13	-9.1	-3.51	-3.19	-3.14	-2.21	-2.75
6'-OH-2,2',4,5'-tetrabromodiphenyl ether	6'-OH-BDE49	6.8	0.25	2.38	0.15	0.15	9.06	0.11	-8.9	-3.18	-2.90	-2.77	-3.49	-3.23
2'-OH-2,3',4,4'-tetrabromodiphenyl ether	2'-OH-BDE66	6.7	0.25	2.4	0.15	0.18	9.34	0.12	-9.1	-1.87	-3.20	-2.95	-2.82	-2.78
6-OH-2,2',3,4,4'-pentabromodiphenyl ether	6-OH-BDE85	7.6	0.23	2.62	0.15	0.17	9.15	0.14	-9.4	-4.21	-3.77	-3.58	-3.46	-3.45
6-OH-2,2',3,4,4',5-hexabromodiphenyl ether	6-OH-BDE137	6.3	0.04	2.81	0.15	0.17	9.47	0.14	-9.6	-4.53	-4.23	-3.94	-3.68	-3.88
Prediction set														
2,4,6-trichlorophenol	246-TCP	6.3	0.15	1.24	0.13	0.12	4.71	0.12	-8.48	-2.88	-3.24	-3.07	-2.71	-2.32
2,2',4,4'-tetrabromobisphenol-A	TBBPA	6.9	0.40	2.95	0.15	0.23	13.43	0.05	-8.83	-1.26	-1.12	-1.26	-1.42	-1.91
2,2',4-tribromodiphenyl ether	bde17	14	0.32	2.23	0.18	0.16	6.51	0.14	-14.2	-5.46	-4.67	-4.83	-5.11	-4.95
2,2',4,4'-tetrabromodiphenyl ether	bde47	14	0.26	2.41	0.18	0.17	7.01	0.16	-14.3	-4.92	-5.30	-5.25	-5.05	-5.33
4'-OH-2,2',4-tribromodiphenyl ether	4'-OH-BDE17	8.9	0.53	2.24	0.15	0.16	11.07	0.10	-8.75	-1.84	-2.02	-2.23	-2.41	-2.89
6-OH-2,2',3,4,5-pentabromodiphenyl ether	6-OH-BDE90	6.7	0.14	2.67	0.15	0.17	9.65	0.13	-9.38	-4.07	-3.73	-3.44	-3.38	-3.48

pKa = Acidic equilibrium constant; B = hydrogen bonding basicity parameter; Vm = McGowan's molecular volume; Ea = Covalent acidity; Qh+ = Electrostatic acidity
 Totdip = Total dipolar momentum; H_{charge} = hydrogen with the lowest positive charge; ΔE_{OH} = Energy difference of HOMO (substrate) and LUMO (OH radical)

The properties generated were variables describing molecules polarity, reactivity properties based on the energies of the frontier molecular orbitals (E_{HOMO} and E_{LUMO}) and atom specific properties such as charges, electron density, and electrophilic/nucleophilic/radical susceptibility calculated with semi-empirical quantum mechanics calculations. Quantum mechanical molecular properties based on the theory of TLSER⁷ included properties describing polarity, covalent/ electrostatic acidity and basicity, total, maximum, average dipole moments and molecular volume. Finally, LSER solute descriptors consist of the excess molar refraction, the polarizability, the effective hydrogen bond acidity and basicity, McGowan's volume.

For the creation of QSPR models we used two software suites, ADME works model builder (Fujitsu Ltd) for variable selection using an genetic algorithm and for the creation of multiple linear regression equations (MLR) and Simca (ver 11.5, Umetrics, Inc) using the partial least square projection to latent structures (PLS). A training and prediction set was selected 20 + 6 chemicals where the prediction set was randomly selected based on activity sampling.

Results and discussion

To evaluate the persistency of a chemical in the environment one should take into account the major degradation routes in different environmental compartments such as oxidation, radical reactions and UV degradation in air and water, reduction in soil and sediments. In this work we have focused on oxidation and if the use of the linear solvation energy relationship (LSER) could be valuable as a relatively simple means for multimedia modelers to evaluate the reactivity of different chemicals. We have here calculated a large suit of different physicochemical properties based on semi-empirical quantum calculations to compare with properties calculated using LSER and TLSER by using two multiple regression methods to find correlations between physicochemical properties and the oxidation half-lives. MLR and PLS was used on all semi-empirically calculated variables and the TLSER variables as seen in Figure 1. The training set of 20 compound shows that the difference in regression results between using all calculated variables compared to only using TLSER variables is only slightly in favour for the modelling of all variables. This suggests that TLSER variables are enough to make some fair estimates of the oxidation potential of compounds when experimental data is unavailable. However, the LSER variables did not perform as well as the TLSER and gave no viable MLR or PLS models. The different regression methods used, MLR and PLS show no greater difference and give comparable results. Results from the prediction set, seen as triangles in Figure 1, well fits the models, giving models with a goodness of prediction Q^2_{ext} between 0.82-0.93 which is an excellent result. With the use of MLR we constructed a linear mathematical model that relates the oxidation half-lives to the calculated physicochemical properties:

$$\text{Log}(1/t_{1/2}) = 0.9060 - 1.7016 \times \mathbf{Vm} - 42.1222 \times \mathbf{Ea} + 20.5695 \times \mathbf{Qh+} + 0.3186 \times \mathbf{Totdip} \quad (1)$$

(n = 20, $R^2 = 0.81$, F = 16.2, $p < 2.6\text{E-}5$, $S^2 = 0.70$, $R^2_{\text{ext}} = 0.88$)

$$\text{Log}(1/t_{1/2}) = 1.3871 - 24.4328 \times \mathbf{H}_{\text{charge}} + 0.2517 \times \Delta\mathbf{E}_{\text{OH}} + 1.9741 \times \mathbf{B} + 0.0221 \times \mathbf{pKa} \quad (2)$$

(n = 20, $R^2 = 0.91$, F = 36.5, $p < 1.2\text{E-}7$, $S^2 = 0.49$, $R^2_{\text{ext}} = 0.86$)

Equation 1 shows the MLR model based on the TLSER variables while eq. 2 shows the results for the MLR model based on all calculated physicochemical variables.

Results show that the most important physicochemical properties being, in falling order of magnitude, the molecular volume (Vm), covalent acidity (Ea), electrostatic acidity (Qh+) and the total dipolar momentum (Totdip) the variable with the lowest importance. Eq. 2 on the other hand show an decrease in oxidation half-times strongly correlated (with an correlation coefficient of $R^2=0.85$) to the molecules that have an hydrogen connected to the phenyl ring with the lowest positive charge (H_{charge}) suggesting a point where the molecule is most easily oxidized on the phenyl ring supplanting a hydrogen with a hydroxyl group (Ph-H to Ph-OH). The second most important variable is the energy difference between the E_{HOMO} of the substrate and the E_{LUMO} energies of a hydroxyl group (ΔE_{OH}) or the potential for the substrate to have a nucleophilic reaction with water. Additional variables but with a decrease in importance is the hydrogen bonding basicity parameter (B) and pKa as ionized species at a pH of 7.6 are more easily oxidized.

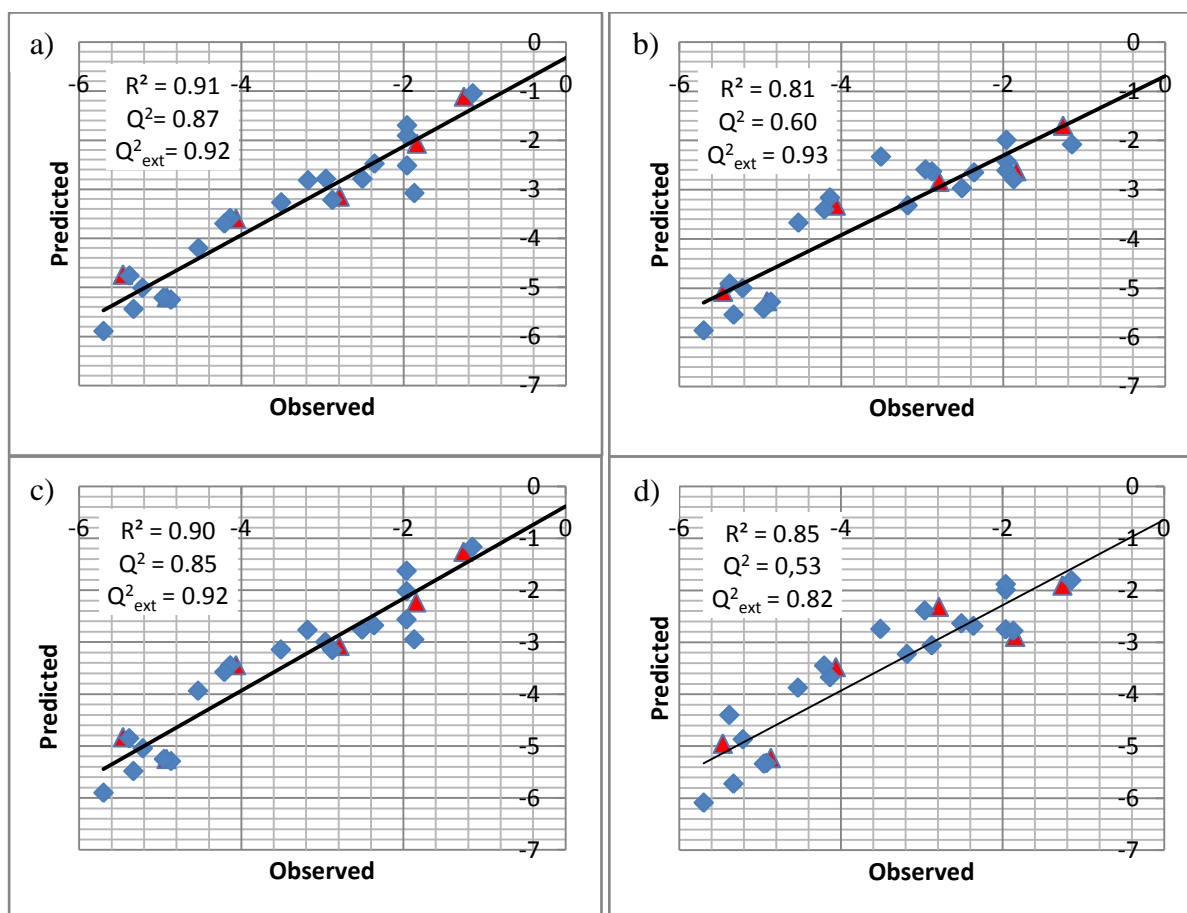


Figure 1. Observed and predicted oxidation half-lives ($\log 1/t_{1/2}$) and the correlation coefficients for the compounds in the training set (\diamond) R^2 , the goodness of prediction Q^2 and Q^2_{ext} for the prediction set (\blacktriangle). a) MLR model using all calculated variables, b) MLR model using only TLSE variables, c) PLS model using all calculated variables, d) PLS model using TLSE variables.

References

1. EC (2006). [Official Journal of the European Union L 396](http://eurlex.europa.eu/LexUriServ/site/en/oj/2006/l_396/l_39620061230en00010849.pdf).
2. OECD (2007). OECD Guidelines for the Testing of Chemicals. http://www.oecd.org/document/40/0,2340,en_2649_34377_37051368_1_1_1_1,00.html
3. Breivik K. and Wania F. (2003). *Environ Sci Technol.* 37, 4934-43.
4. Bastos P. M., Eriksson J., Green N., Bergman Å (2008). *Chemosphere* 70, 1196-1202.
5. Bastos P. M., Eriksson J., Vidarsson J., Bergman Å (2008). Submitted to *Environ. Sci. Pollut. Res.*
6. Platts JA, Butina D, Abraham MH, Hersey A (1999). *J. Chem. Inform. & Computer Sci.* 39(5), 835-845.
7. Wilson LY, Famini GR (1991) *J. Med. Chem.* 34(5):1668-1674.
8. Stewart, J.J.P. (1990). *J. Comput.-Aided Mol. Design* 4, 1-45.