Physico-chemical property prediction of emerging pollutants: PFC and (B)TAZ for environmental distribution

Barun Bhhatarai, Paola Gramatica

QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Structural and Functional Biology, University of Insubria, Varese (Italy)

E-mail: bhhataba@gmail.com, paola.gramatica@uninsubria.it

Department of Structural and Functional Biology, University of Insubria, via J.H. Dunant, 3, 21100 Varese (Italy)

Perfluorinated compounds (PFCs) are a family of chemicals with a long carbon chain which is predominantly substituted by fluorine. They are used in different materials as non-adhesives, waterproof fabrics, fire-fighting foams, etc. (Benzo)triazoles (B/TAZs) are another class of chemicals with multi-nitrogenated aromatic ring system. They are widely used in industrial processes, deicing agents (1H–B/TAZs), pharmaceuticals and pesticides. These chemicals are considered as "emerging pollutants" as they are broadly distributed in the environment because of their extensive use and are considered to be hazardous as they cause adverse effects to humans and other non-target species. Their high concern as pollutants, lack of experimental data and crucial Authorisation under REACH legislation urges for a need to maximize the information inherited in the existing data. Internally robust and externally validated QSPR models were developed for the endpoints, as also recommended under the REACH regulation, to predict large set of unknown properties for existing or not yet synthesized chemicals. The predicted compounds within the structural applicability domain (AD) were projected in a multivariate PCA plot to understand their leaching and volatility behaviour along with soil sorption partition coefficient (K_{OC}) predicitons for B/TAZs and to understand the relationship between the studied endpoints with Critical Micelle Concentration (CMC) and Bioconcentration Factors (BCF) for PFCs.

Regression Models

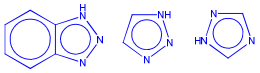
DATA SET: Non-ionic compounds, with experimental data were taken from ChemID Plus [1] and PERFORCE [2]. The dataset was divided into training and prediction sets using the Random by response approach and by Kohonen Maps-Artificial Networks (k–ANN). QSAR models developed were used to define the structural AD on large set of compounds.

MOLECULAR DESCRIPTORS: DRAGON [3] theoretical molecular descriptors (0D–3D).

QSAR MODELING: Multiple Linear Regression (MLR) and Genetic Algorithm Variable Selection (GA–VSS) using MOBY DIGS software [4] using the Ordinary Least Square (OLS) regression.

VALIDATION OF THE MODELS: Internal predictive ability: leave-one-out cross-validation (Q^2_{LOO}) and bootstrap (Q^2_{boot}). Reliability and robustness: Y permutation testing ($R^2_{Y\text{-}scrambling}$). External validations: Schüürmann et al. (Q^2_{extF2}) [5], and Todeschini et al. (Q^2_{extF3}) [6].

(Benzo)triazoles & Perfluorinated chemicals



95-14-7 1,2,3-Benzotriazole | 37306-44-8 Triazole | 288-88-0 1H-1,2,4-Triazole

Response	$RMSE_{TR}$	$RMSE_{EPI}$
Water solubility (WS)	0.54	0.86
Hydrophobicity ($logK_{OW}$)	0.64	0.65
Vapor pressure (VP)	0.88	1.832
Melting Point (MP)	29.63	73.29

$logWS = 13.803(\pm1.003) - 2.415(\pm0.167)ClC0 - 0.444(\pm6.43E\text{-}02)AMW + 1.655(\pm0.397)MATS7e$
$n=49, R^2=83.81, Q^2=81.15, Q^2_{Ext}=69.19\text{-}88.14, s=0.52, F=77.64$ T1

$logK_{OW} = 1.706(\pm 0.438) + 2.802(\pm0.167)B08[C\text{-}C] - 0.689(\pm 0.103)nN + 1.159(\pm 0.199)\ GATS3m + 1.527(\pm 0.317)MATS1v$
$n=64, R^2=88.63, Q^2=86.71, Q^2_{Ext}=80.90\text{-}94.49, s=0.62, F=115$ T2

$logVP = 17.302(\pm5.364) - 15.668(\pm1.527)BELp2 + 0.441(\pm0.087)RBN + 1.378(\pm0.314)B09[N\text{-}Cl]$
$n=33, R^2=80.91, Q^2=75.08, Q^2_{Ext}=63.65\text{-}73.61, s=0.82, F=40.96$ T3

$MP = 1098.25(\pm 171.06) - 162.83(\pm 17.6)R2e + 53.225(\pm 10.126)GGI4 + 26.823(\pm3.891)F03[N\text{-}N] - 1693.79(\pm 381.339)\chi1A$
$n=56, R^2=81.32, Q^2=77.34, Q^2_{Ext}=71.93\text{-}87.58, s=0.28, F=55.5$ T4

$logWS = -0.418\ (\pm1.940) - 0.003\ (\pm0.001)T(F..F) + 5.185\ (\pm3.849)SIC1$
$n=20, R^2=76.31, Q^2=69.13, Q^2_{Ext}=79.21\text{-}92.76, s=0.91, F=27.38$ P1

$logVP = -0.16(\pm0.02)F03[C\text{-}F] - 3.16(\pm0.92)AAC - 0.64(\pm0.40)nDB + 7.97(\pm1.26)$
$n=35, R^2=90.93, Q^2=88.21, Q^2_{Ext}=80.36\text{-}87.78, s=0.883, F=103.6$ P2

$logCMC = 1.351(\pm0.183) - 0.30(\pm0.018)\ \chi3$
$n=10, R^2=97.35, Q^2=95.93, s=0.164, F=293.5$ P3

Response	$RMSE_{TR}$	$RMSE_{EPI}$
Water solubility	1.98	0.96
Vapor pressure	1.13	0.95

335-67-1 PFOA

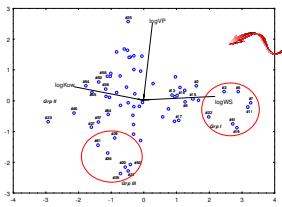
1763-23-1 PFOSA

Environmental partitioning of (Benzo)triazoles & Perfluorinated chemicals using PCA

Bhhatarai B., P. Gramatica, Water Research (2011) 45, 1463-1471

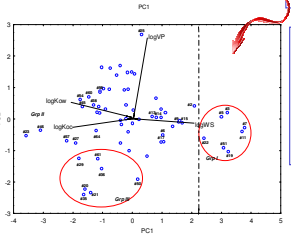
– 66 (18.8%) B/TAZ within the AD of models (T1–T4) and for which at least an experimental data available were studied in PCA plot. PC1: 59.9% and PC2: 33.4%.

– K_{OC} values were calculated when not available experimentally by using earlier published models [7]: Leaching index (LIN) where the solubility and the sorption coefficients plays opposite role, and Volatility index (VIN) which differentiates between the volatile and non-volatile compounds were studied.

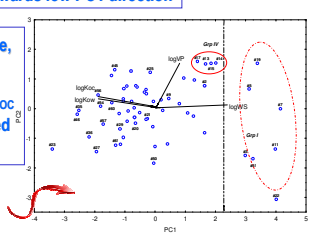


Group I: short chain TAZs → More leachable
Group II: Azocyclotrin (#23), Imibenconazole (#46) → More sorbed, Group III → Less volatile.
Isazofos (#25) → most volatile
1H–BTAZs (#2, 9, 15, 17) located near to Group I.

Introduction of K_{OC} changes the distribution for (#20, 21, 27, 29, 35, 57) towards low PC1 direction



EPI: Group I: More sparse, Group IV: more volatile. No difference when $logK_{OC}$ prediction was introduced ($logK_{OW}$ dependent)

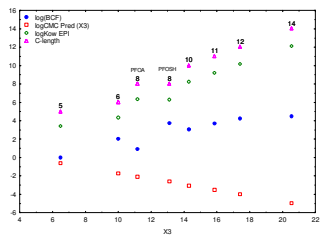


Bhhatarai, B.; Gramatica, P., Environ. Sci. Technol. (2010) Online first

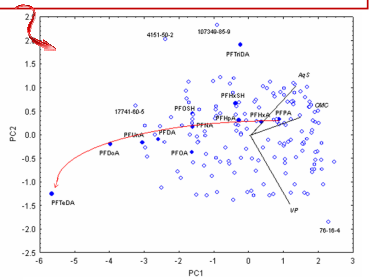
– 174 (78.7%) PFCs within the AD of models (P1–P3) for all end-points were studied using PCA. PC1: 81.3% and PC2: 16.2%.

– Bioconcentration factors (BCF) values were collected [8] and the relationship between the end-points were studied.

– The increasing trend of BCFs is in opposite direction to that of WS and CMC and the trend is found different for carboxylates and sulfonates.



Relation of K_{OW}, C-length and BCF with $\chi3$



Increasing trend of BCF and its relation with WS, VP and CMC

– Parameter $\chi3$ a connectivity index descriptor representing molecular branching and complexity was able to correlate not only CMC but also differentiate the trend between logBCF value of sulfonates and carboxylates and isomers of PFCs

– logBCF is found proportional to $logK_{OW}$ and the C-length as expected, but inversely proportional to the CMC value.

[1]. ChemID Plus, http://chem.sis.nlm.nih.gov/chemidplus.
[2]. Perforce dataset, H. Krop, P. de Voogt, PERFORCE 2, Task 1: Physicochemical parameters and source markers of PFAS. 2008, Interfaculty Environmental Science Department (IVAM), Universiteit van Amsterdam, ISO doc O/0818.
[3]. Dragon - Todeschini R., Consonni V. and Pavan M., 2001. DRAGON – Software for the calculation of molecular descriptors, rel. 5.5 for Windows.
[4] MobyDigs - Todeschini, R. et al., 2004, Moby Digs – MOdels BY Descriptors In Genetic Selection, ver. 1 beta for Windows, Talete srl, Milan (Italy).

References

[5] Q^2F2- Schüürmann,G.; et al, 2008, J. Chem. Inf. Model.,48, 2140–2145.
[6] Q^2F3- Consonni, V., 2009, J. Chem. Inf. Model., 49, 1669–1678
[7] Gramatica, P. et al., 2007, J. Mol. Graph. Model. 25, 755-766.
[8] Martin, J.W. et al, 2003, Environ. Toxicol. Chem., 22, 196-204.

Financial support by European Union through the project CADASTER FP7-ENV-2007-1-212668 is gratefully acknowledged.