Journal of Cheminformatics

POSTER PRESENTATION

Open Access

Updating existing QSAR models: selection and weighting of new data

Tomas Öberg[*], T Liu

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

Computational chemistry and quantitative structure-activity relationships (QSAR) are foreseen to be extensively used in the implementation of the new REACH regulation for chemicals in Europe. However, for some compound groups the data are too few in number to permit both calibration and testing of a new model. Usage and previously developed or updated models are then viable alternatives.

Perfluorocarboxylic acids (PFCAs) and fluoroteleomer alcohols (FTOHs) are two groups of environmentally relevant compounds, with unique physical and chemical properties. The subcooled liquid vapour pressure (pL) is one such property, where experimental determinations are limited and far from consistent [1]. Updating is, however, challenging when the new compounds are far outside of the original calibration domain space. But by carefully selecting and weighting only three new compounds, we have been able to update a previously developed general QSAR model [2], to cover the new domain while maintaining predictive performance for the earlier calibration and test data. The optimal weighting scheme was determined from the sample leverages and residuals in the calibration phase [3].

The performance of this re-calibrated model greatly surpassed previous modelling attempts [4], when applied to an external test set of two PFCAs and four FTOHs with pL in the range 0.2-200 Pa; with Q2Ext = 0.994 and RMSEP = 0.190 units of log Pa. The domain coverage also increased from 1% to 51%, for 426 perfluoroalkylated compounds selected from the REACH registration list, the PhysProp database, and the OECD 2006 survey [5]. Selection and weighting of new calibration data can thus facilitate the extension and use of existing QSAR models.

References
1. Goss K-U, Bronner G, Harner T, Hertel M, Schmidt TC: *Environ Sci Technol* 2006, **40**:3572.
2. Öberg T, Liu T: *QSAR Comb Sci* 2008, **27**:273.
3. Stork CL, Kowalski BR: *Chemometr Intell Lab Syst* 1999, **48**:151.
4. Arp HP, Niederer C, Goss K-U: *Environ Sci Technol* 2006, **40**:7298.
5. Lists of PFOS, PFAS, PFOA, PFCA, related compounds and chemicals that may degrade to PFCA. ENV/JM/MONO(2006)16, OECD. 2007.

University of Kalmar, Sweden