Predictive QSAR models to screen potential hazardous chemicals under REACH and applicability within the industrial context

Lunghini, F.,[*,1,2] Marcou, G.,[1] Azam, P.,[2] Varnek, A.[1]
*lead presenter, f.lunghini@outlook.com
[1]*Laboratory of Chemoinformatics, University of Strasbourg, 4 Rue Blaise Pascal, 67081, Strasbourg, France*
[2]*Toxicological and Environmental Risk Assessment unit, Solvay S.A., 20 Rue Marcel Etienne Sembat, 69190, Saint-Fons, France*

The REACH (Registration Evaluation Authorization and restriction of Chemicals) [1] regulation compels industries to report potential sources of hazards for substances to be put on the market. REACH supports the use of alternative methods, in particular the employment of predictive statistical models, in order to decrease the need of expensive or ethically debatable assays (animal testing). However, state of the art models, based on currently-available public data, fail to meet the industrial needs. This is due to a limited overlap between industry specialties and public domain data, mostly reducing the applicability domain of the models. This observation is based on a substantial amount of data coming from an industrial context to challenge models built on public data-only.

The following three properties were considered: Bioconcentration Factor, Environmental Biodegradability and Rodent Oral Acute Toxicity. The Generative Topographic Mapping [2] approach was used to detect industrial-specific chemotypes. Many structural features resulted to be consistently under sampled for more than one property, suggesting that the application of currently available models could be limited for some chemical families of industrial interest.

Finally, all data sources have been merged, including the industrial datasets. As expected, the applicability domain of the resulting models is extended, covering important chemotypes of the industrial context.

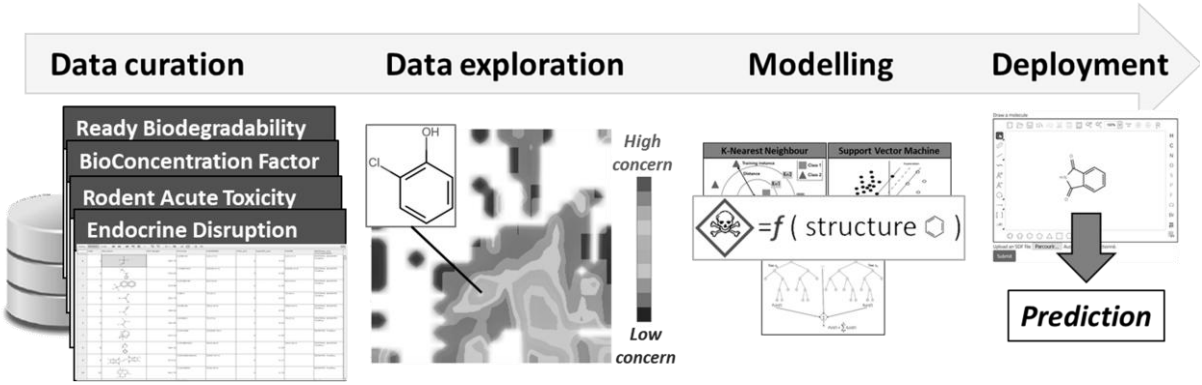

Figure 1. Graphical abstract.

Bibliography:

[1] Regulation (EC) No 1907/2006 - EUR-Lex

[2] Kireeva N.; Baskin I.I. et al. Mol. Inf., 2012, 31, 301-312