Consensus QSAR modeling for the toxicity of organic chemicals against *Pseudokirchneriella subcapitata* using 2D descriptors
Kabiruddin Khan1, Kunal Roy2

1Drug Theoretical and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, 188 Raja S C Medlick Road, 700032, Kolkata, India

**Ecotoxicity of organic chemicals (OCs)**
- OCS-Most extensively used chemical class
- Potent PBT potential candidates
- Consist of Pharmaceuticals, UV filters, hormones, biocides and endocrine disrupting chemicals etc
- Environmental risk assessment of many OCs became mandatory in the US (since 1995) and EU (since 1993)
- NOEC and EC50 are most widely used in ecotoxicity measurement

**What is NOEC?**
- No observed-effect concentration
- Highest concentration at which there is no significant change in control population against the measured endpoint
- Represents lower threshold for toxicity

**What is EC50?**
- Denotes % effective concentration where % is change in the populations of studied species (x can be 5-100%)
- OECD and SETAC recommend replacement of NOEC with EC(x = 5-20)
- In USEPA, effluent testing relies on NOEC and EC0.25
- In EU-REACH regulation, NOEC and EC50 are fundamental information for indicating toxicity
- Consensus on need of replacing NOEC with EC50 in workshop conducted by OECD in 1996

**Protocol followed**
1. To propose robust QSAR models to calculate EC0 and EC50 concentrations in algae
2. To demonstrate advantage of consensus modeling in error reduction
3. To identify features responsible for algae immobility
4. To fill the data gap of untested/unknown compounds using developed models
5. To analyze the predictive efficiency of the developed models

**Results and Analysis**
**Data curation for model development**
- CrippenMR (molecular refractivity), liposolubility, presence of nitrogen/hydrogen (due to presence of CrippenMR, MLOGP, Rg25(NA); Rg12(NA); PHLG(NA), PHLG(NA) enhanced pEC50 values, while presence of more polar atoms like oxygen (MLOGP, Rg25(NA) are not predicted EC50 values. Molecular weight was a significant descriptor appearing in all the five (QSAR models with positive contribution to pEC50 values. Among the other descriptors, Rg12(NA), PHLG(NA) and PHLG(NA) contributed positively while 3NA contributed negatively to pEC50.

**Applicability domain assessment (pEC50)**
- The models developed using the pEC50 data gave 22 outliers in the training set, and 5 chemicals in the test set were outside the domains of applicability.
- The outliers were retained in the final models as they were predicted with moderate precision by the respective models (absolute predicted residual being 32 log units)

**Conclusions**
- Consensus models (CMs) were more able to predict 64 additional synthetic organic compounds, leading to an absolute observed values reported in the paper of Kuan et al.
- Initially, individual five PLS models (pEC50) were used to derive the data predicted values of 64 compounds, the models could include 64 compounds with confidence satisfactorily. Prediction reliability index (PRI) tool.
- In the second round, the obtained consensus model was employed in the estimation of all predicted toxicities of 64 compounds, and the predictions were compared with the range values provided in the manuscript.
- The model could predict accurately 56 (85%) of 64 compounds with deviations of plus minus 2 log unit (for the lower range values) and 51 (80%) of 64 compounds with deviation of plus minus 2 log unit (for the higher range values)
- In the extended set, 37 molecules (55%) were predicted with deviation of less than 1 log unit
- Seven molecules were outside the domain of prediction as given by the PRI tool
- The precise predictions of additional data points proves the significance of the developed QSAR models, consensus approach and practical application of the PRI tool in predictions of unsubstituted chemicals.

**References**

**Acknowledgement**
- K.K. thanks Indian Council of Medical Research, New Delhi for financial support in the form of a senior research fellowship.