Final project report September/2013

18 November 2013

Chemical grouping and computational methods to assess biologic activity and toxicity

Early stage researcher:

Mr. Swapnil Chavan

Project supervisor:

Prof. Ian A. Nicholls

Research Institution:

Linneaus University, Sweden

Duration of fellowship:

28 months

Introduction:

The known traditional physico-chemical groupings, as performed by environmental scientists, may not fulfil the requirements for use of the similarity of compounds in predicting biological response. Instead statistical approaches based on principal component analysis and supervised or unsupervised clustering using functional and structural descriptors could be beneficial. Thus the chemical categories and read-across approaches will be investigated within ECO by employing multivariate statistical methods as well as unsupervised clustering methods, which could benefit the analysis by splitting the chemical space into suitable blocks prior to grouping. A number of physical-chemical and toxicological endpoints will be studied to clarify variability and the need for adoption of the approaches to chemical category assignment.

Planned action:

Workplan-1: Literature search

1. ECHA, Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals, European Chemicals Agency, 2008.

2. ECHA, Guidance on information requirements and chemical safety assessment Chapter R.10: How to avoid unnecessary testing on animals, European Chemicals Agency, 2008.

3. OECD, Guidance on Grouping of Chemicals, OECD Environment Health and Safety Publications, Series on Testing and Assessment No. 80, Organisation for Economic Co-operation and Development, 2007.

4. Munro, I., et al., Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food and Chemical Toxicology, 1996, 34(9) 829-867.

5. Stochero, M., Bassan, A., Fioravanzo, E., Pavan, M., Applicability of physicochemical data, QSARs and read-across in Threshold of Toxicological Concern assessment. SCIENTIFIC REPORT submitted to EFSA, 2011, 1-135.

Workplan-2: Setting up database for descriptor and endpoint

The 596 Munro chemicals for which 4885 descriptors were calculated using Dragon 6. NOEL values will be used as endpoints. Hands on and tutorial study have been carried out on OECD-QSAR toolbox2.2.

Results: Successful hands on tutorial study on software OECD QSAR toolbox2.2 and Dragon 6.

Status: Completed (Expected time:0-4 months, Time consumed:0-4 month)

Workplan-3: Setting up database and descriptor selection

a). Authentication of all the Munro database chemicals were done based on CAS registry, IUPAC name and structure. The salts, mixtures were removed. The acute toxicity data (LD_{50}) was searched for all the chemicals.

Results: 441 chemicals were found to have correct structure and exact LD_{50}

Workplan-4: Use of Principal Component Analysis approach

a) Principle Component Analysis (PCA) was used for outlier detection.

Results: 5 chemicals were identified as potential outliers in PCA score plot.

b) Descriptor selection was performed using genetic algorithm

Results: 25 significant descriptors found for 436 Munro chemical dataset

Status: Completed (Expected time:0-9 months, Time consumed:0-15 month)

Milestone-2: Investigation of unsupervised approaches for chemical categorization

Workplan-5: Classification by k-Nearest Neighbor (kNN) approach

The classes of chemicals were formed on the basis of Globally Harmonized Scheme (GHS). The optimum k (no. of neighbors) value were obtained by cross validation. The kNN model was developed and validated.

Results: Successfully estimated GHS scheme with 67% accuracy for classification of Munro database chemicals.

Workplan-6: PCA study for Chemical cluster analysis and mechanistic interpreatation

The score plot and loading plot were analyzed in details to understand role of descriptors in clustering chemicals. The Interpretation was made to link the specific structural clusters formed with corresponding descriptor of high loading value. The correlation of significant descriptors with LD_{50} values was also calculated to support the mechanistic interpratation.

Results: 16 descriptors were found to have more than 95% significance level with respect to LD_{50}. Major role of few significant descriptors in forming structural clusters on PCA score plot was confirmed.

Status: Completed (Expected time:4-20 months, Time consumed:4-27 month, comment: Delayed because of change of supervisor and time spent on Industrial training)

Expected publication: "**Towards Global QSAR Model Building for Acute Toxicity: Munro Database Case Study",** Swapnil Chavan, Ian A. Nicholls, Björn C. G. Karlsson , Annika M. Rosengren, Davide Ballabio, Viviana Consonni, Roberto Todeschini (under submission)

Milestone-3: Investigation of unsupervised approaches for Read-across

Workplan-7: Use of kNN approach for category formation and Read-across

Database used: NEDO database of 136 chemicals

Endpoint: LD50 and LOEL

The kNN classification model was developed based on classes formed on the basis of LD50 values. The optimum k (no. of neighbors) value was obtained by cross-validation. The kNN classification model was developed using training set of 107 chemicals and validated on 27 test set chemicals. The Read-across studies were carried out to predict LOEL values of each target chemicals using arithmetic mean of LOEL values of k neighbors.

Results: The Read-across prediction shows 71% accuracy.

Status: Completed (Expected time:18-24 months, Time consumed:18-21 month, comment: Faster than proposed time span)

Expected publication: "Prediction of repeated dose toxicity from acute toxicity: K-nearest neighbor classification based read-across approach**",** Swapnil Chavan, Ian A. Nicholls (under submission)

Workplan-8: Industrial training

Company: eADMET gmbH, Germany

Duration: 3 months

Status: Successfully completed 3 month Industrial training