HiT QSAR Modeling of Chemical Toxicants Tested against *Tetrahymena pyriformis*

Eugene N. Muratov[a], Kun Wang[c], Alexander Golbraikh[c], Victor E. Kuz'min[a,b], Anatoly G. Artemenko[a,b], Leonid Gorb[a,d], Mohammad Quasim[d], Jerzy Leszczynski[a], Alexander Tropsha[c].

[a] Laboratory on Theoretical Chemistry, A.V. Bogatsky Physical-Chemical Institute National Academy of Sciences of Ukraine, Odessa, Ukraine;
[b] Computational Center for Molecular Structure and Interactions, Jackson State University, Jackson, Mississippi, USA;
[c] Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA;
[d] US Army ERDC, Vicksburg, Mississippi, USA;

International virtual collaboratory consisting of six independent groups with shared interests in computational chemical toxicology has been formed in order to unite the efforts in obtaining predictive QSAR models in the mentioned area. The Hierarchical QSAR Technology (HiT QSAR) has been used for consensus QSAR analysis of chemical toxicants tested against *Tetrahymena pyriformis* with the purpose of subsequent addition to this project. Simplex representation of molecular structure (SiRMS) QSAR approach has been used for descriptors generation and statistical models have been obtained by Partial Least Squares (PLS) and K-Nearest Neighbors methods. Training set consists of 644 compounds and 2 external test sets consist of 349 and 110 compounds respectively.

The aims of investigation were: development of successful consensus QSAR model for *Tetrahymena pyriformis* set, it subsequent validation on two external test sets and inclusion of our models into united consensus model obtained by initial six teams.

Successful united consensus model (R^2_{test1} = 0.84; R^2_{test2} = 0.72) based on PLS and KNN consensus QSAR models has been obtained (100% coverage of the test sets). Leverage and ellipsoid domain applicability (DA) approaches in PLS modeling and sphere-based DA procedure have been used for additional estimation of the reliability of prognosis. However, application of DA procedures doesn't improve the quality and leads only to decreasing of test sets coverage.

Acknowledgement: Authors are indebted to organizers of the School for their kind support of Dr. Muratov visit.