Task ID: 425.035

Task Title: High-Throughput Cellular-Based Toxicity Assays for Manufactured Nanoparticles and Nanostructure-Toxicity Relationships Models

Deliverable #7: Establish web portal for (secure) data deposition and sharing

Summary/Abstract:

We have established a web portal called "ChemBench" (http://chembench.mml.unc.edu/) as a public cheminformatics resource to provide the experimental and computational researchers with knowledge discovery tools and infrastructure that enables them to explore and exploit experimental datasets of chemical compounds with measured biological or physical/chemical properties. The portal enables researchers to build and explore rigorous data models, and reduce the amount of experimental effort required to identify novel biologically active compounds. The portal has modules for data deposition and preparation, model building, and prediction of external compounds. It has been used to add and procure the three datasets described in deliverables 4-6. Models described under Deliverable 6 can be used by interested experimental researchers to prioritize the selection of designed NPS for experimental investigation.

Technical Results and Data:

ChemBench has been designed to integrate translational cheminformatics research conducted in our group over a period of more than 15 years as well as in collaborating laboratories elsewhere. The prototype version of ChemBench developed in the last three years is publicly available (chembench.mml.unc.edu) and includes modules for:

- Quantitative Structure Property Relationship (QSPR) model development including QNAR modeling;
- Property/activity prediction from chemical structure(or descriptors of chemical structures) using models (predictors) developed externally or using the QSPR model development component of ChemBench;
- A database of compounds annotated by their published and/or predicted properties.

Chembench is built using only open source tools and software for data processing and visualization. Figure 3 summarizes our overall QSPR modeling strategy that is focused on delivering validated predictive models and ultimately, computational hits proposed for experimental validation. We start by randomly selecting a fraction of compounds (NPs in our case) (typically, 10-15%) as an external evaluation set. The remaining compounds are then divided rationally into multiple training and test sets that are used for model development and validation,

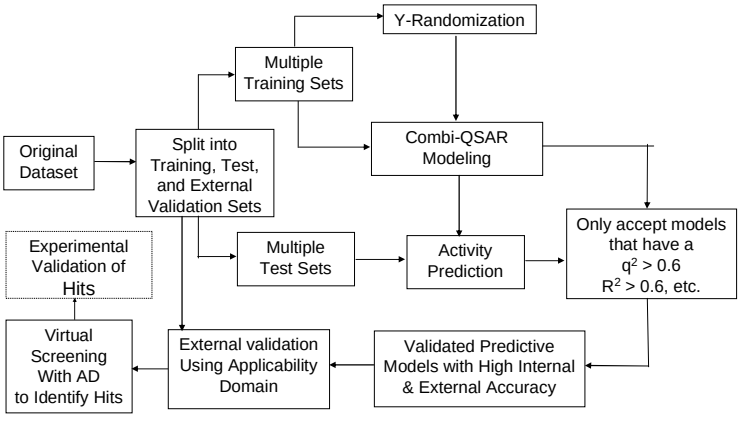

***Figure 3**. Flowchart of predictive QSPR modeling framework based on validated combi-QSAR models*

respectively using criteria discussed in more detail below. We employ multiple QSPR techniques based on combinatorial exploration of all possible pairs of descriptor sets coupled with various statistical data mining techniques and select models characterized by high accuracy in predicting both training and test sets data. Validated models are finally tested using the evaluation set. The critical step of the external validation is the use of applicability domains. If external validation demonstrates significant predictive power of the models, we use all such models for virtual screening of available chemical databases.

During the reporting period, we have uploaded the three dataset described under Deliverable 4-6 to the ChemBench (that provides tools for additional dataset uploading as they become available in the remaining years of the project). The models described under deliverables 4-6 have been built outside of the portal. We are in the process of re-developing models within the Chembench portal such that they can be used by external collaborators for estimating biological activity (or toxicity) of newly designed NPs.