Task ID: 425.035

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

Deliverable #6: Build initial QNTR models for selected end points

Summary/Abstract:

This report summarizes the results of Quantitative Nanostructure-Toxicity Relationship (QNTR) analysis of nanoparticles tested against multiple endpoints. Three datasets collected and curated from the literature were used for QNTR modeling (cf. deliverables 4 and 5). Statistically significant QNTR models with acceptable external predictivity were developed by implementing support vector machine and *k* nearest neighbors algorithms. These models could be potentially useful in prioritizing nanoparticles for biological/toxicological testing.

Technical Results and Data:

QNTR models were developed individually for three datasets as follows.

(1) In the first case study, nanoparticles were composed of very different core structures for which conventional chemical descriptor can not be computed. Instead, experimentally determined properties of nanoparticles, such as size, zeta potential and relaxitivities, were used as descriptors of nanostructures. Nanoparticles were categorized into two classes based on their averaged activity values in 64 bioassays (arbitrary threshold was chosen for binary classification purpose). Results showed that the averaged external prediction accuracy was 73% (Figure 1).

(2) In the second case study, nanoparticles had the same core structure (CLIO, cross-linked iron oxide) but different small molecules were used as surface modifiers. Therefore, each nanoparticle in the dataset was represented by the chemical structure of its surface modifier; conventional chemical descriptors were calculated with the MOE software and *k* nearest neighbors algorithm was used for model building. QNTR models were built for PaCa2 cell uptake data. Results of 5-fold external cross validation showed that the averaged R_o^2 (coefficient of determination, which is a squared correlation coefficient for regression through the origin) was 0.72; using applicability domain threshold (which restricts predictions to those particles that are relatively similar to those in the training set) raised this correlation coefficient to 0.77 (Figure 2A). Analysis of MOE descriptors found to be used most frequently in QNTR models with the highest prediction power demonstrated that they could explain most of the variance of the biological/toxicological effect (Figure 2B). For example, lipophilicity-related descriptor GCUT_SLOGP_0 significantly discriminate nanoparticles with highest vs. lowest PaCa2 cell uptake (Figure 2C).

(3) The third case study was relatively similar to the second. Nanoparticles in this dataset have the same core structure (MWNT, multi-walled nanotube) but different surface modifiers so each nanoparticle was represented by the modifier. Dragon descriptors were calculated for each nanoparticle and QNTR models were developed against two protein

binding endpoints (chymotrypsin and hemoglobin) using *k* nearest neighbors algorithm. An external dataset was set aside and models were built using remaining compounds. The external accuracy of those models was found to be relatively high (Figure 3A) with R_o^2 values of 0.76 and 0.73 for CT binding and HB binding, respectively. Descriptors that were used most frequently by models with high prediction power (Figures 3C and 3D), e.g., GATS5m and a_nCl could explain most of the variance of CT (Figure 3E) and HB (Figure 3F) binding by nanoparticles.

	Effect	Size	Zeta pot.	Relaxitivities	
NP-01	High	0.4865	0.5278	0.2941	0.3986
NP-02	Low	0.4054	0.7222	0.4837	0.6476
NP-03	High	0.4324	0.5833	0.3529	1.0000
NP-04	Low	1.0000	0.5833	1.0000	0.7991
NP-05	High	0.3649	0.4722	0.2353	0.9403
NP-06	High	0.3919	0.6111	0.3333	0.9079
NP-07	High	0.5135	0.5833	0.4052	0.6270

⇩

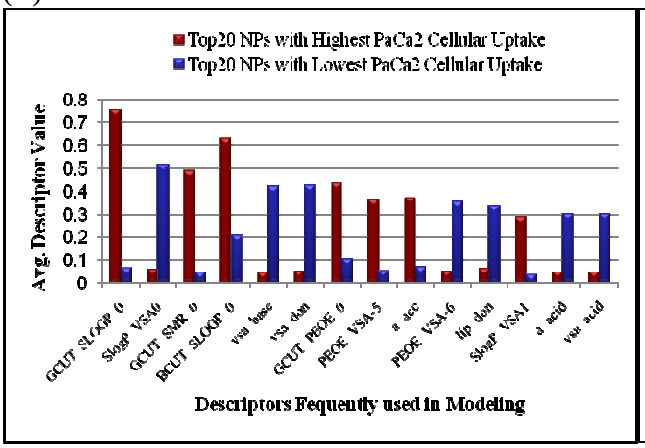
Fold	MODELING SETS				EXTERNAL SETS				
	n	# models	% accuracy internal 5-fold CV	% accuracy	*n*	% accuracy	% CCR[a]	% Sensitivity (SE)	% Specificity (SP)
1	35	11	51.4 – 60.0	71.4 – 82.9	9	**78**	83	67	100
2	35	13	51.4 – 60.0	71.4 – 77.1	9	**78**	75	50	100
3	35	16	57.1 – 62.9	74.3 – 82.9	9	**78**	78	80	75
4	35	11	60.0 – 62.9	77.1 – 88.6	9	**56**	55	50	60
5	36	4	66.7	83.3 – 86.1	8	**75**	67	33	100
[a]CCR – Correct Classification Rate; CCC = ½ (SE +SP)					**44**	**73**	**73**	**60**	**86**

Figure 1. QNTR modeling of the biological effect for 44 MNPs in the first dataset.

(A)

Fold	# comp. model	# comp. external	w/o AD		w/ AD		
			R_o^2	MAE	R_o^2	MAE	% cov
1	87	22	0.65	0.18	0.67	0.18	86
2	87	22	0.67	0.14	0.73	0.13	91
3	87	22	0.72	0.22	0.75	0.21	82
4	87	22	0.75	0.19	0.90	0.14	64
5	88	21	0.80	0.16	0.78	0.17	76
Average	87	22	**0.72**	**0.18**	**0.77**	**0.17**	**80**

(B)



(C)

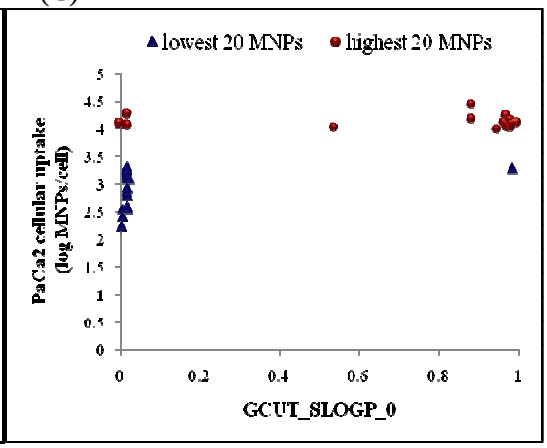


Figure 2. QNTR modeling of PaCa2 cell uptake for 109 MNPs with different surface attachment in the second dataset (A). (B) Average descriptor values in MNPs with highest and lowest PaCa2 cellular uptakes. (C) Example of a lipophilicity related descriptor (GCUT_SLOGP_0) significantly discriminating particles with highest and lowest PaCa2 cellular uptakes.

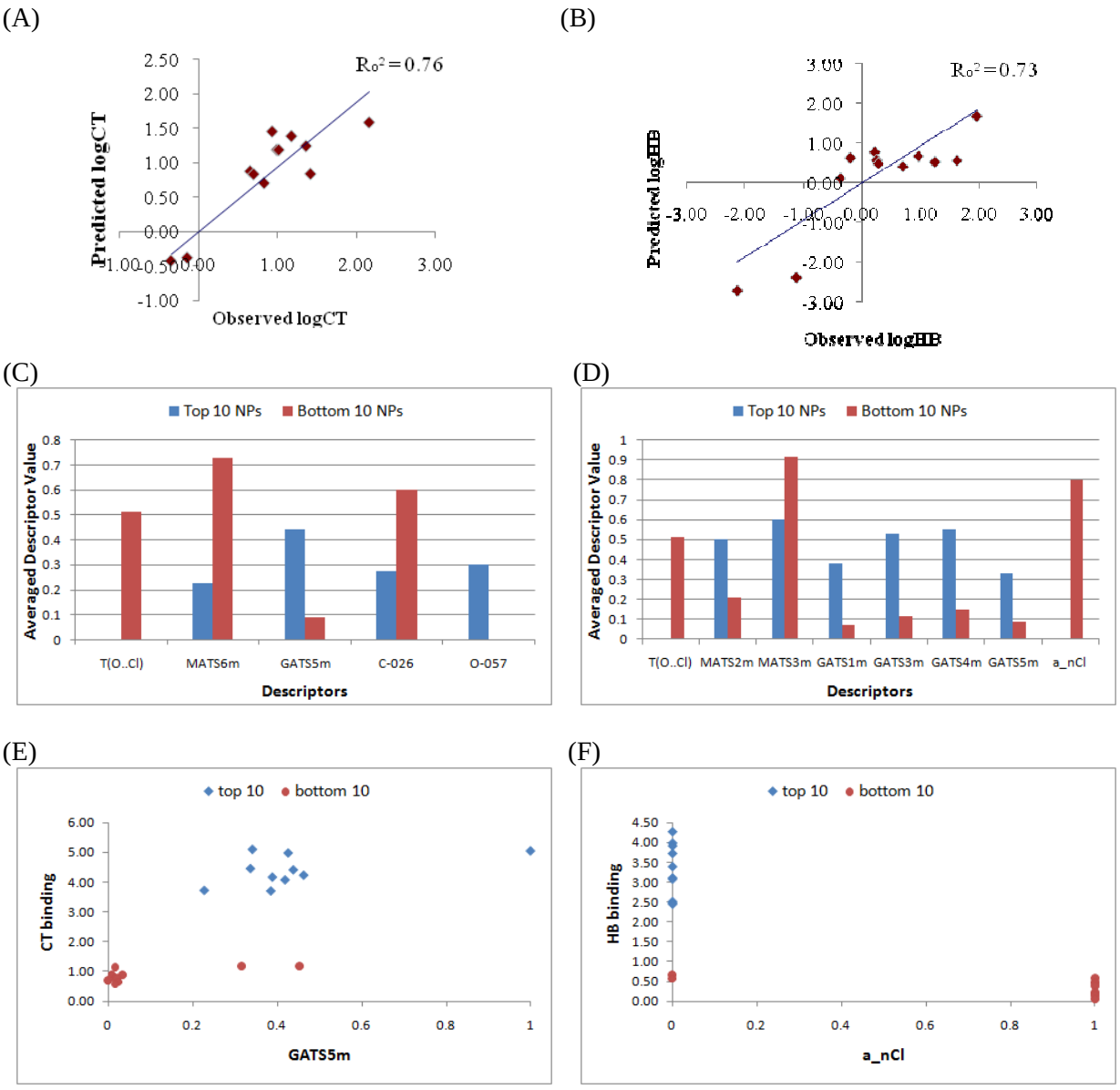


Figure 3. Results of QNTR analysis for the third dataset comprised of ca. 80 NPs with the same core but different surface modifiers. Models with acceptable external predictivity were developed using CT and HB binding data. Prediction for an external dataset is shown for CT (A) and HB (B) binding, respectively. The most frequently used descriptors and their averaged values for NPs with the highest and lowest CT (C) and HB (D) binding are shown. Examples of descriptors with the highest discriminatory power with respect to strongest vs. weakest CT (E) and HB (F) binding are also shown.