Electronic Supplementary Material (ESI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2019

SUPPORTING INFORMATION

In silico profiling nanoparticles: predictive nanomodeling using universal nanodescriptors and various machine learning approaches

Xiliang Yan[a, b]*, Alexander Sedykh*[b, c]*, Wenyi Wang*[b]*, Xiaoli Zhao*[d]*, Bing Yan*[e, f]** and Hao Zhu*[b, g, h]***

[a] School of Chemistry and Chemical Engineering, Shandong University, Jinan 250100, China;

[b] The Rutgers Center for Computational and Integrative Biology, Camden, New Jersey 08102, USA;

[c] Sciome, Research Triangle Park, NC 27709, USA;

[d] Department of Physiological Science, Eastern Virginia Medical School, Norfolk, Virginia 23507, USA;

[e] Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Institute of Environmental Research at Greater Bay, Guangzhou University, Guangzhou 510006, China;

[f] School of Environmental Science and Engineering, Shandong University, Jinan 250100, China;

[g] Department of Chemistry, Rutgers University, Camden, New Jersey 08102, USA

[h] College of Life Science and Bio-engineering, Beijing University of Technology, Beijing, 100124, China

*Corresponding authors:
Hao Zhu
315 Penn St.
Department of Chemistry, Rutgers University
Camden, NJ 08102, USA
Telephone: (856) 225-6781
Email: hao.zhu99@rutgers.edu

Bing Yan
Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Institute of Environmental Research at Greater Bay, Guangzhou University
Guangzhou 510006, China
E-mail: drbingyan@yahoo.com

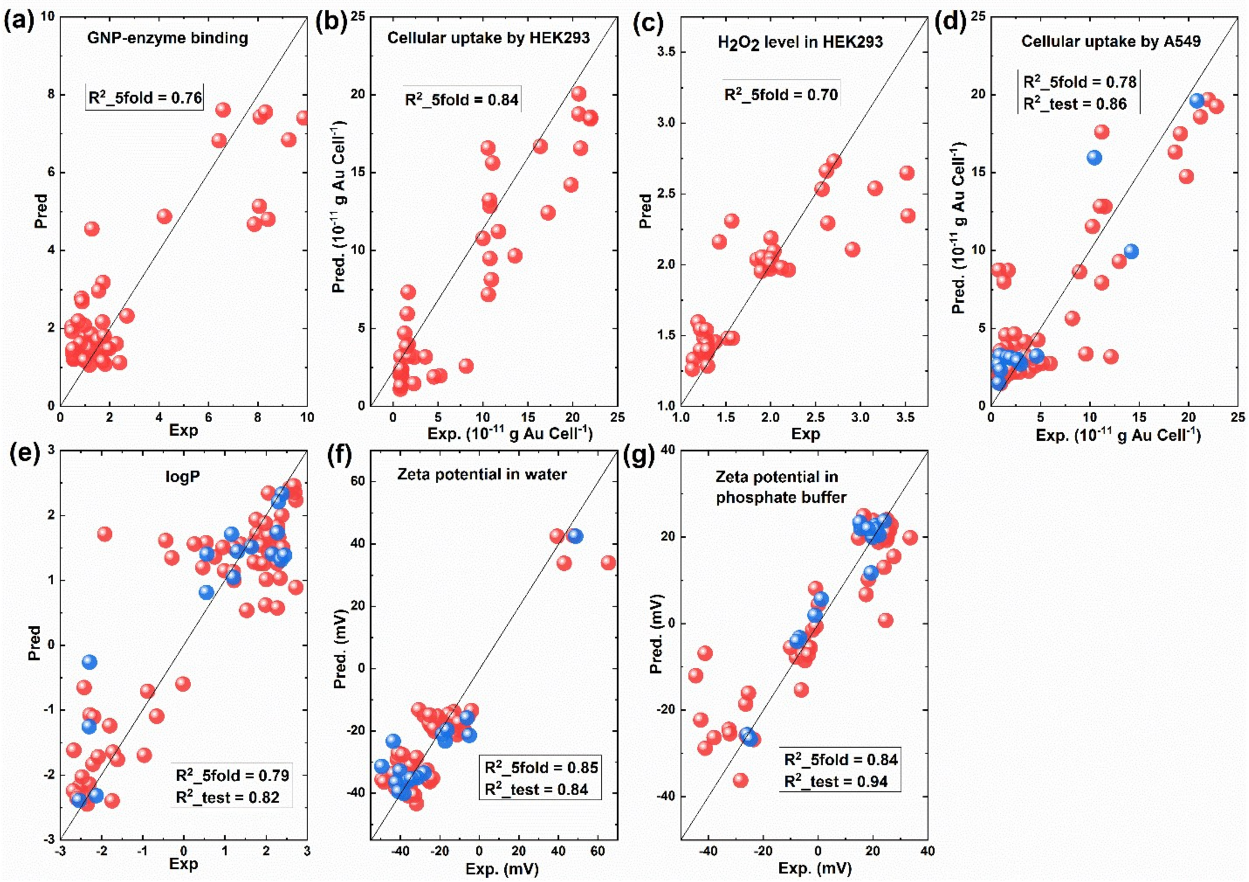

Fig. S1 QNAR model performance of consensus modeling results in (a) GNP-enzyme binding affinities (b) Cellular uptake potentials by HEK293 cell (c) ROS levels in HEK293 cell (d) Cellular uptake potentials by A549 cell (e) logP (f) Zeta potentials in water (g) Zeta potentials in phosphate buffer. Red dots are GNPs in the training set, and blue dots are those in the test set. The results were obtained from one-time simulation. Correlation coefficients (R^2) from consensus modeling results are also shown.

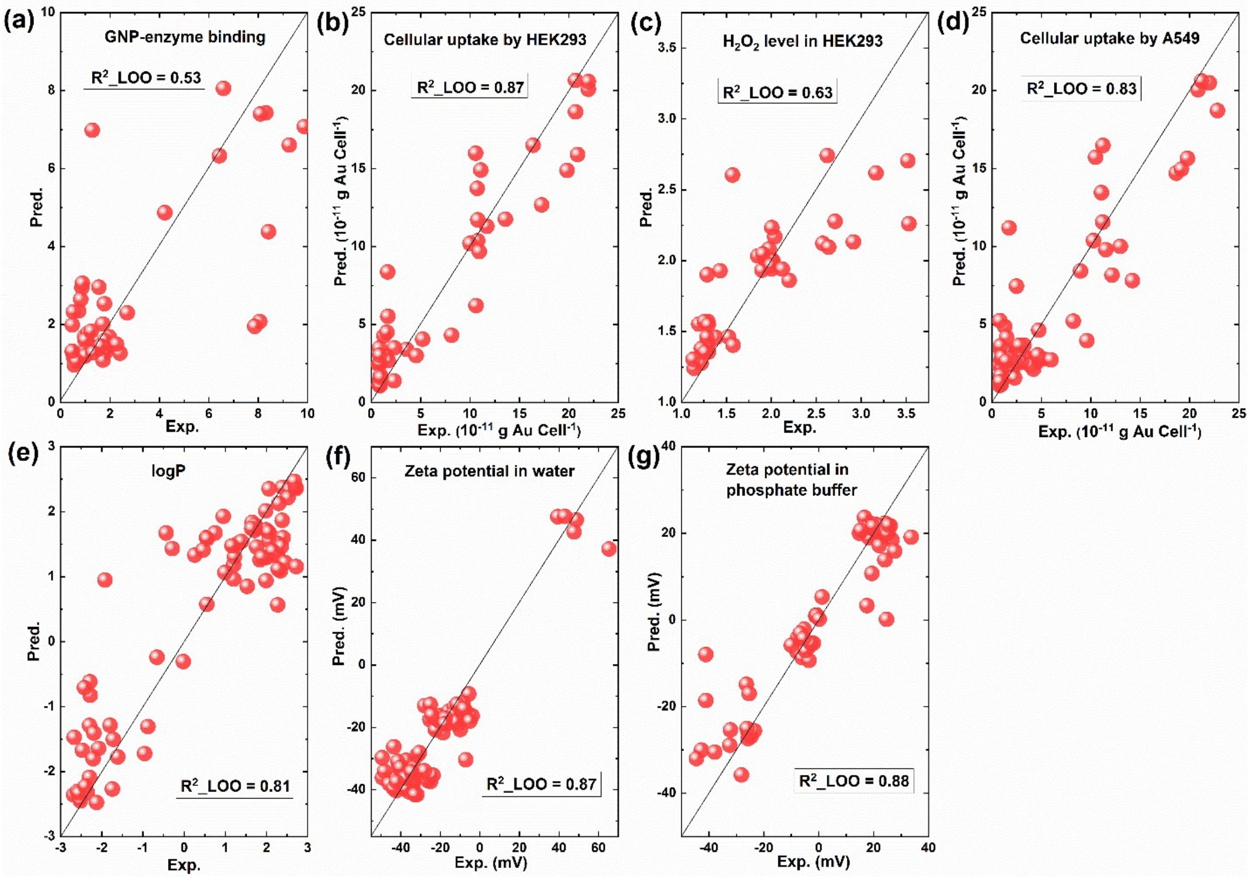

Fig. S2 QNAR model performance of consensus modeling results in (a) GNP-enzyme binding affinities (b) Cellular uptake potentials by HEK293 cell (c) ROS levels in HEK293 cell (d) Cellular uptake potentials by A549 cell (e) logP (f) Zeta potentials in water (g) Zeta potentials in phosphate buffer. Red dots are GNPs in the training set, and blue dots are those in the test set. The results were obtained from leave-one-out (LOO) validation. Correlation coefficients (R^2) from consensus modeling results are also shown.

Table S1. Correlation coefficients (R^2) from *k*NN, RF and consensus modeling results of seven modeling sets. The results were obtained from one-time simulation.

Modeling set	***k*NN**		**RF**		**Consensus**	
	5-fold cross validation	Eternal validation	5-fold cross validation	External validation	5-fold cross validation	External validation
1	0.85	-	0.57	-	0.76	-
2	0.86	-	0.76	-	0.84	-
3	0.80	-	0.51	-	0.70	-
4	0.78	0.87	0.68	0.81	0.78	0.86
5	0.81	0.86	0.69	0.76	0.79	0.82
6	0.89	0.81	0.76	0.86	0.85	0.84
7	0.81	0.93	0.83	0.93	0.84	0.94

Table S2. Correlation coefficients (R^2) from *k*NN, RF and consensus modeling results of seven modeling sets. The results were obtained from leave-one-out (LOO) validation.

Modeling set	***k*NN**	**RF**	**Consensus**
1	0.51	0.51	0.53
2	0.90	0.81	0.87
3	0.67	0.54	0.63
4	0.84	0.76	0.83
5	0.86	0.73	0.81
6	0.86	0.86	0.87
7	0.85	0.89	0.88