

Supplementary material

Metal and metal oxide nanoparticle toxicity: Moving towards a more holistic structure-activity approach

G.P. Gakis, I.G. Aviziotis, C.A. Charitidis*

Research Lab of Advanced, Composite, Nano-Materials and Nanotechnology, Materials Science and Engineering Department, School of Chemical Engineering, National Technical University of Athens, 9 Heroon Polytechniou Street, Zografos, Athens 15780, Greece

Corresponding authors: Tel: +30 2107724046. E-mail: charitidis@chemeng.ntua.gr

S1. Descriptor selection

The absolute value of the Pearson correlation coefficient is presented in Figure S1, below.

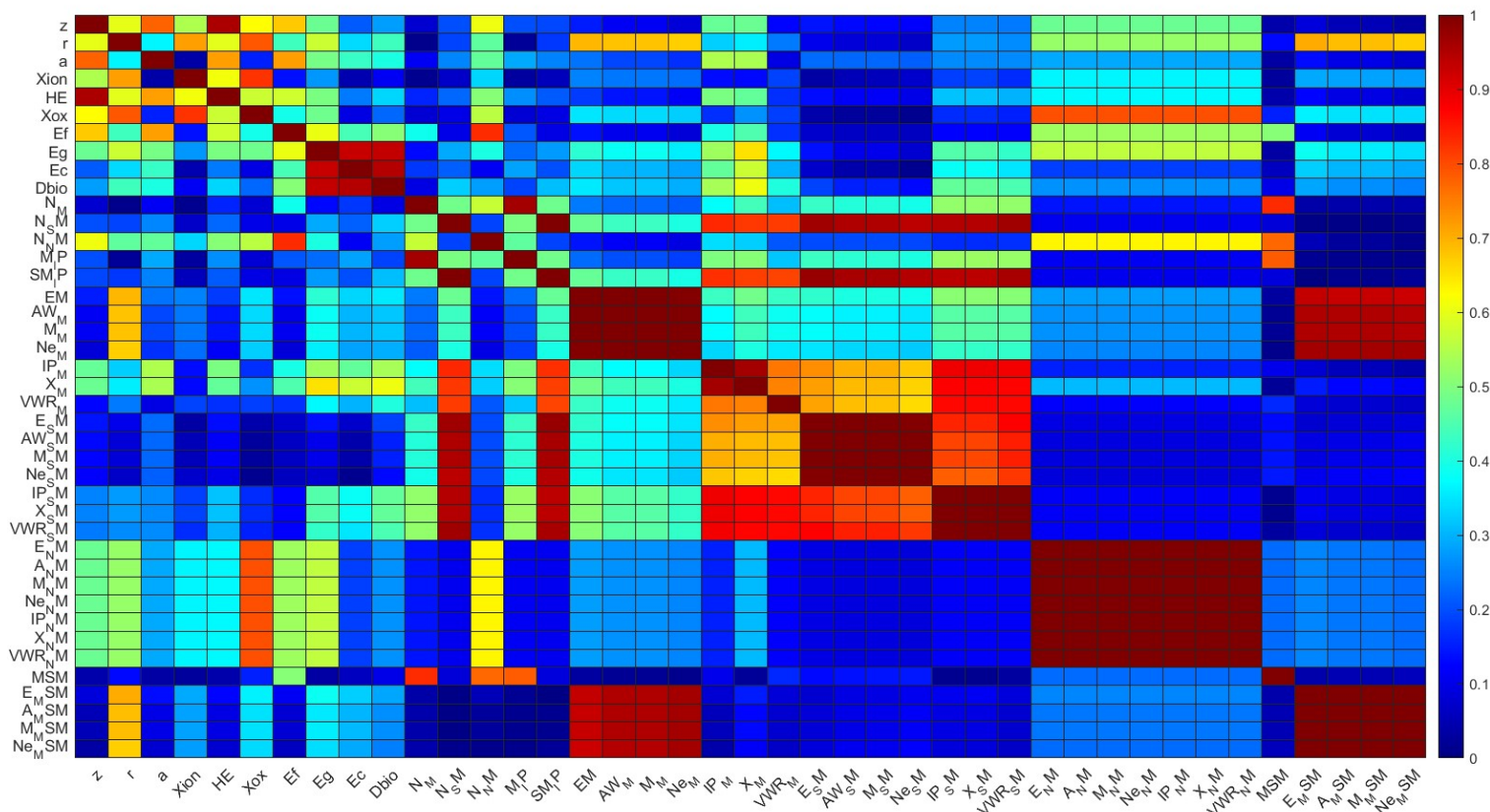


Figure S1. Pearson correlation coefficient for the initial descriptor set.

Following the results of Figure S1, the descriptors with an intercorrelation above 0.95 are deemed to be intercorrelated, and only one of the intercorrelated descriptors is kept. Following this procedure, the descriptors that are omitted are presented below, in Table S1.

Table S1. List of eliminated descriptors from the intercorrelation analysis

Descriptor removed	Intercorrelation with
A MSM	E_MSM
M MSM	
Ne MSM	
AW M	
M M	
Ne M	
E M	
A NM	E_NM
Ne NM	
M NM	
IP NM	
X NM	
VWR NM	
E SM	SM_IP
A SM	
Ne SM	
M SM	
IP SM	
X SM	
VWR SM	
N SM	
A SM	E_M
Ne SM	
M SM	
X M	IP_M
N M	M_IP
Eg	Dbio
Ec	
Z	HE

S2. Optimal descriptor combination

Based on the descriptor selection procedure, the four most relevant descriptors (Dbio, HE, IP_M, E_MSM) were tested in different combinations. Models were built using different descriptor combinations, and the respective model accuracy was estimated. The results towards the most populous individual cell line dataset (A549 cells) are presented in Table S2

Table S2. Model accuracy with the different descriptor combinations towards the A549 dataset

Dbio	HE	E_MSM	IP_M	Accuracy
Single descriptor models				
✓	✗	✗	✗	88.2
✗	✓	✗	✗	85.3
✗	✗	✓	✗	91.2
✗	✗	✗	✓	88.2
Two descriptor models				
✓	✓	✗	✗	94.1
✓	✗	✓	✗	82.4
✓	✗	✗	✓	82.4
✗	✓	✓	✗	85.3
✗	✓	✗	✓	91.2
✗	✗	✓	✓	88.2
Three descriptor models				
✓	✓	✓	✗	88.2
✓	✓	✗	✓	94.1
✓	✗	✓	✓	85.3
✗	✓	✓	✓	85.3
Four descriptor models				
✓	✓	✓	✓	88.2

Results of Table S2 show that the optimal model accuracy is obtained using the Dbio-HE and the Dbio-HE-IP_M descriptor sets. Since the addition of IP_M does not increase the model accuracy, the set of Dbio-HE descriptors is selected as the optimal model.