Published in final edited form as:
Small. 2011 April 18; 7(8): 1118–1126. doi:10.1002/smll.201002366.

Classification Nano-SAR Development for Cytotoxicity of Metal Oxide Nanoparticles

Dr. Rong Liu,
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095. Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095

Prof. Robert Rallo,
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095. Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095. Departament d'Enginyeria Informatica i Matematiques, Universitat Rovira i Virgili, Av. Paisos Catalans 26, 43007 Tarragona, Catalunya, Spain

Dr. Saji George,
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095. Department of Medicine - Div. of NanoMedicine, University of California, Los Angeles, CA 90095

Dr. Zhaoxia Ji,
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095

Dr. Sumitra Nair,
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095. Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095

Prof. André E. Nel, and
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095. Department of Medicine - Div. of NanoMedicine, University of California, Los Angeles, CA 90095

Prof. Yoram Cohen
Center for the Environmental Implications of Nanotechnology, California Nanosystems Institute, University of California, Los Angeles, CA 90095. Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095

Yoram Cohen: yoram@ucla.edu

Abstract

A classification based cytotoxicity nano-structure-activity-realtionship (nano-SAR) is presented based on a set of nine metal oxide nanoparticles to which transformed bronchial epithelial cells (BEAS-2B) were exposed over a range of concentrations of 0.375–200 $mg \cdot L^{-1}$ and exposure times up to 24 h. The nano-SAR is developed using cytotoxicity data from high throughput screening (HTS) assay that was processed to identify and label toxic (in terms of the Propidium Iodide

Correspondence to: Yoram Cohen, yoram@ucla.edu.

Supporting Information is available on the WWW under http://www.small-journal.com or from the author.

uptake of BEAS-2B cells) versus non-toxic events relative to unexposed control cell population. Starting with a set of fourteen intuitive but fundamental physicochemical nano-SAR input parameters, a number of models were identified which had classification accuracy above 95%. The best performing model had a 100% classification accuracy in both internal and external validation. This model is based on four descriptors including the atomization energy of the metal oxide, period of the nanoparticle metal, nanoparticle primary size, in addition to nanoparticle volume fraction (in solution). Notwithstanding the success of the present modeling approach with a relatively small nanoparticle library, it is important to recognize that a significantly larger data set would be needed in order to expand the applicability domain and increase the confidence and reliability of data-driven nano-SARs.

Keywords

nanoparticle; metal oxides; cytotoxicity; nano-SAR; classification

1. Introduction

Nano-sized materials are increasingly utilized as common elements in many modern industrial products and processes primarily due to their small size and unique nano-scale properties[1]. Engineered nanomaterials (eNMs) are estimated to be components of more than 1000 commercial products[2], and thus there is an increased public concern about the inherent adverse impacts of eNMs and the resulting exposure that may take place in the workplace, among consumers, and in the environment[3]. Although recent studies have identified that certain eNMs possess properties that may lead to biological hazard[4], the understanding of the general principles governing the toxicity potential and the long-term environmental health and safety impact of these products is in its infancy[5]. In this regard, toxicity screening is critical for characterization of the potential hazard of eNMs, which is in turn indispensable information for subsequent risk assessment and development of environmental and health regulatory policies. However, generation of the required *in vitro* and *in vivo* toxicity characterizations, which are necessary to cope with the expected growth in number and diversity of eNMs, is a formidable task. Therefore, in addition to experimental approaches for hazard assessment, there is a need for *in silico* methods in order to develop structure-activity relationships(SARs) that correlate toxicity end points[6]. Such structure-activity relationships, (i.e., nano-SARs) require sufficiently large experimental databases of reasonable diversity (.e.g., with respect to the heterogeneity of nanoparticles and biological receptors) and suitable nanoparticle descriptors[7].

The majority of published nano-SAR models[8–14] focus primarily on the predictions of physicochemical properties of nanoparticles (e.g. solubility and Young's modulus). In a recent study[15], nano-SAR classifiers based on nanoparticle properties such as primary size, aggregate size in different media, and zeta potential were developed to predict cell membrane damage induced by TiO_2 and ZnO. More recently, two quantitative nanostructure-activity relationships (nano-SARs) were proposed for *in vitro* biological effects of a range of different nanoparticles[16]. One was a *k*NN-based regression model[17], for nanoparticle uptake by pancreatic cancer cells (PaCa-2), developed based on 109 nanoparticles of the same core (iron oxides/NH_2 core based) but with different organic chemical surface modifications[18]. This model made use of a set of commonly used molecular descriptors for the organic coatings and thus did not explicitly consider the intrinsic nanoparticles properties. The highest and average reported model performances were quantified by a square correlation coefficient of 0.8 and 0.72, respectively. A second model was a Support Vector Machine (SVM[19]) classifier for cytotoxicity, based on 51 iron-oxide and quantum dots based nanoparticles with different polymeric coatings[20],

utilizing nanoparticle primary size, two magnetic properties and zeta potential as descriptors. The model was developed for averaged toxicity responses relative to control (i.e., unexposed) cell population and which were then grouped into two halves being below or above an arbitrary threshold. A model classification accuracy of 73% was reported, but without a definitive identification of true toxic end points.

It is noted that in recent years there has been a significant increase in efforts to identify physicochemical characteristics of nanoparticles that drive their toxic activity[21–23]. These studies have demonstrated that the dependence of cytotoxicity on physicochemical nanoparticle properties can vary significantly for different nanoparticle classes and cell lines. A recent review[7] on the challenges of developing nano-QSARs concluded that, due to the complexity of chemical and morphological structures of nanoparticles, it would be most beneficial to develop nano-SARs for individual nanoparticle classes along with appropriate validation of the applicability domain of such models and selection of suitable NP descriptors. The generation of detailed descriptors of the structure and chemistry of nanoparticles requires the use of advanced computational methods that include, but are not limited to, quantum mechanics calculations[24, 25], molecular dynamics and Monte Carlo simulations[26, 27]. However, these approaches are only feasible when the target nanoparticles consist of a relatively small number of atoms given the significant increase in computational complexity and overhead as the number of atoms increases. Recent studies have argued that a preferred practical approach is to utilize a small set of fundamental nanoparticle descriptors, especially given the yet limited nanoparticle characterization and toxicity data databases[7] relative to the chemical world. Accordingly, based on a small set of fundamental physicochemical descriptors, the current work presents an approach for the development of a classification nano-SAR for cytotoxicity of a small set of metal oxide nanoparticles as quantified by high throughput screening (HTS) of the Propidium Iodide (PI) uptake (i.e., an indicator of plasma membrane damage) by BEAS-2B cells.

2. Results and Discussion

The present approach for the development of nano-SAR classifiers for cytotoxicity of metal oxide nanoparticles is summarized in Figure 1. *In vitro* cytotoxicity data for nine metal oxide nanoparticles were generated for BEAS-2B and subsequently processed to remove outliers in the control population, and identify and label measurable cytotoxicity effects in the nanoparticle exposed cell population. Using an initial set of fundamental nanoparticle physicochemical descriptors and appropriate concentration measures (i.e., nano-SAR input parameters), the most suitable subset of input parameters (i.e., belonging to the nano-SAR of the highest accuracy) was identified via model training and validation (internal and external) following the OECD guideline[28].

Cytotoxicity of the metal oxide nanoparticles (Al_2O_3, CeO_2, Co_3O_4, TiO_2 (80% anatase and 20% rutile), ZnO, CuO, SiO_2 (amorphous), Fe_3O_4, and WO_3) in the size range of 8–19 nm was assessed by quantifying the loss of plasma membrane integrity (PI uptake assay[29]) for a concentration range of 0.375–200 $mg \cdot L^{-1}$ and exposure times up to 24 h. The experimental results, expressed in terms of the percentage of membrane-damaged cells, are provided in Table S1. The associated dose-response curves are shown in Figure S1 (Supporting Information), along with error bars depicting the experimental variability. After identification and removal of outliers (Figure 2), the Strictly Standardized Mean Difference (SSMD[30–32]) was employed (Section 4) to quantify the difference in the percentage of membrane-damaged cells (indicated by PI uptake) by the cell population exposed to nanoparticles relative to the unexposed population (Figure 3). A nanoparticle, at a given concentration, was labeled toxic (i.e., plasma membrane damage indicated by PI uptake) if the average of the estimated SSMD (for the different exposure times) was greater than 2;

otherwise the nanoparticle was considered nontoxic. The resulting labels (i.e., the cytotoxicity endpoints for building the nano-SAR model) along with the estimated SSMD are provided in Table S2 (Supporting Information), respectively. It is noted that, with the labeling approach selected in the present work, as described in Section 4.4 and Table S2 (Supporting Information), the risk of mislabeling nanoparticles as nontoxic (i.e., false negatives) was kept below 5%, while at the same time ensuring that only nanoparticles that exhibit strong cytotoxic effects are identified as toxic (i.e., reducing the level of false-positives).

The estimated SSMD of the cytotoxicity data suggests that labeling of ZnO (50–200 $mg{\cdot}L^{-1}$), CuO (150–200 $mg{\cdot}L^{-1}$), and SiO_2 (150–200 $mg{\cdot}L^{-1}$) nanoparticles as toxic (Figure 3), while the remainder of the nanoparticles induced no notable impacts. The toxicity labeling of ZnO, CuO and SiO_2 nanoparticles are consistent with previous studies. Studies with ZnO nanoparticles have demonstrated that the generation of Zn^{2+} ions due to dissolution of zinc oxide, causes lysosomal injury, mitochondrial damage, and triggers subsequent cellular events leading to cell death[29, 33]. The cytotoxicity of CuO nanoparticles in cellular and animal based studies[34, 35] has been documented in various studies. It has also been reported that amorphous SiO_2 can lead to the generation of reactive oxygen species (ROS) that may trigger pro-in ammatory responses both *in vivo* and *in vitro*[36].

The nano-SAR was developed based on a small set of ten fundamental nanoparticle descriptors (Table 1) which was selected consistent with the recommendations of a recent comprehensive nano-SARs review[7]. The selected set included descriptors relating to constitutional information (N_M, N_O, m_{Me}, m_{MeO}), encoding the periodic properties of the metal atom (P_{Me}, G_{Me}), stability and reactivity of the metal oxide (E_{MeO}), surface charge (Z_w, *IEP*), and primary size (d). In addition, four different concentration measures were evaluated (C_m, C_s, θ_v, C_n, Table 2). Different nano-SARs were then constructed using a logistic regression model[37, 38], which estimates the probability of a nanoparticle being toxic (or non-toxic) (Section 4), based on the labeled data. Data for the Fe_3O_4, WO_3 and SiO_2 nanoparticles (Table 2) were reserved for model validation with the remaining six were used for model training.

Given the relatively small set of fourteen model input parameters, an exhaustive exploration for the 2^{14} (i.e., 16384) possible subsets of nano-SAR input parameters was carried out. Of the various possible models, a total of 27 (Table S3, Supporting Information) were identified with a classification accuracy (based on internal cross-validation) higher than 95% and zero false negatives. These models were further assessed using the reserved external validation set resulting in only one model, having four input parameters (P_{Me}, E_{MeO}, θ_v, d), that performed with 100% classification accuracy (M1, Table 3). Examples of other models that yielded ~96% or higher classification accuracy in the internal cross-validation are listed in Table3. For example, model M2 included P_{Me}, E_{MeO}, d as nanoparticle descriptors in addition to two different concentration measures (C_n and C_s) but not the primary size. The selection of the above two concentrations introduces redundancy since $C_s/C_n = (d/2)^2$. In contrast, model M1 (i.e., the best performing model) also included the P_{Me}, E_{MeO}, d descriptors, but with θ_v as the concentration measure (M1, Table 3). Models M3-M6, which include either the isoelectric point (*IEP*) or the zeta potential (Z_w), demonstrated lower classification accuracy in internal validation and false negatives during external validation. It is noted that these models did not include the primary particle size as a descriptor. Interestingly, a global search for the optimal set of input parameters, omitting the primary particle size, resulted in reasonably good performance (e.g., M7, Table 3) of 96% and 93% accuracy for the internal and external validation, respectively, with no false negatives, where

three of the four model input parameters (P_{Me}, E_{MeO}, θ_v) were the same as for best performing model (M1).

The analytical expression for the best performing classification nano-SAR (M1, Table 3) is given by:

$$\ln\left(\frac{\mathrm{P}(NP \in T)}{\mathrm{P}(NP \in N)}\right) = 3600.6 + 103.5 \times d + 9.5 \times C_v + 97.6 \times P_{Me} - 58.5 \times E_{MeO} \quad (1)$$

where the probabilities of a nanoparticle being toxic or non-toxic are denoted by P($NP \in T$) and P($NP \in N$), respectively, with the units of the variables (P_{Me}, E_{MeO}, d and θ_v in Eq. (1)) provided in Table 1 and Table 2. It is noted that although the SiO_2 nanoparticle data at 150 and 200 mg·L^{-1} were outside the model applicability domain (Figure S2, Supporting Information), these were correctly classified as toxic (Table 2). It is noted that for a given nanoparticle the metal (P_{Me}) and the atomization energy (E_{MeO}) are fixed, hence the ratio P($NP \in T$)/P($NP \in N$) in Eq. (1) only depends on the nanoparticle volume concentration (θ_v) and primary size (d). Therefore, for each of the study nanoparticles, the nano-SAR classifier (Eq. (1)) can be used to determine a decision boundary (i.e., P($NP \in T$) = P($NP \in N$)) that divides the input parameter space (i.e., primary size and exposure volume fraction) into toxic and non-toxic regions. Accordingly, the studied nanoparticles can be ranked in terms of decreasing hazard as ZnO > CuO > {$WO_3 \approx Co_3O_4 > Fe_3O_4$>} SiO_2> {$CeO_2 > Al_2O_3$> TiO_2}. The above nano-SAR analysis indicates (as per Eq. (1)) that there may be a range of concentrations and particle sizes above which the nanoparticles within the parentheses may trigger toxic response. The above exploration of cytotoxicity data suggests that experimental verification may be warranted to explore if the nano-SAR approach can potentially guide the development of cytotoxicity studies that encompass an expanded range of nanoparticle properties.

It is interesting to note that recent work on cytotoxicity, induced by TiO_2 nanoparticles of 30–125 nm size range, showed that membrane damage (for L2 lung epithelial cells cells) increased with primary particle size[15, 39]. Another recent study[40] with mammalian cells also showed increase uptake of protein-coated gold nanoparticles with increasing primary nanoparticle size from 20 nm up to 50 nm and then decreased uptake with further increase in particle size (up to ~200 nm). Consistent with the above studies, Eq. (1) also predicts that, within the narrow domain of the dataset (8–19 nm), cytotoxicity would increase with increasing primary nanoparticle size. While plausible mechanisms for the above behavior are yet unclear, one must recognize that, for a suspension of a nanoparticle of a given size, there is a distribution of nanoparticle sizes due to aggregation[41, 42]. As acknowledged in a recent review on the correlation of physicochemical with toxicological properties of nanoparticles[43], the interpretations of nanoparticle toxicity data have to carefully consider the aggregation behavior of nanoparticles and associated particle size distribution. Clearly, more data are needed over a wide range of primary nanoparticle sizes and types in order to elucidate their fundamental interactions (over their entire size distribution) with various cell lines.

Inspection of the range of magnitude of the variables in Eq. (1) (Table 1 and Table 2), after normalization of the data for each variable over the range [0, 1] (i.e., $(x-x_{min})/(x_{max}-x_{min})$), revealed that the term containing E_{MeO} makes the greatest contribution to the model (i.e., RHS of Eq. (1)). As indicated by Eq. (1), the probability of a nanoparticle being classified as toxic increases as the atomization energy of the metal oxide (E_{MeO}) decreases. The above behavior could be attributed to the decrease of metal-oxide nanoparticle stability (e.g., dissolution potential) and increased reactivity with decreasing energy of atomization. It has been suggested that, the oxidative stress potential of metal oxide nanoparticles (i.e.,

production of ROS) is related to their conduction and valence band energy levels[44]. The band energy levels are related to nanoparticle's bond strength and thus the atomization energy (or total bond energy) of the metal oxide also provides information that is related to ROS[44]. Increased likelihood of cytotoxicity is also expected (Eq. (1)) with increasing magnitude of the parameters P_{Me}, θ_v, and d. For metals that are at a higher period (P_{Me}) the increased atom size could result in reduced nuclear attraction for the electrons in the outermost shell contributing to reduction of the ionization energy and thus increased tendency for dissolution (i.e., formation of cations in solution). As the nanoparticle volume fraction (θ_v) increases there may be increased cell-particle interactions which appears to be consistent with the dependence of colloidal interactions on corresponding increased total particle surface area in suspension[21, 22] as well as increased osmotic pressure[45].

3. Conclusions

Using high throughput toxicity screening (HTS) data for BEAS-2B cells exposed to nine metal oxide nanoparticles, a logistic-regression based nano-SAR classifier was developed and validated following accepted QSAR validation guidelines[28]. The HTS data were first processed to identify and label toxic versus non-toxic events (at the specific nanoparticle exposure concentrations) via statistical analysis for the standardized difference of the biological responses of cells exposed to nanoparticles relative to unexposed control cell population. A relatively small set of intuitive but fundamental physicochemical descriptors was selected and through an exhaustive search the best performing set of models were identified. The best performing model (100% classification accuracy) was based on three nanoparticle descriptors that included period of the nanoparticle metal, atomization energy of the metal oxide, and nanoparticle primary size, in addition to the nanoparticle volume fraction. The model demonstrated remarkable prediction accuracy and low false negative classification rates as confirmed in both the internal and external validation. Given the established nano-SAR, it is possible to establish metal-oxide nanoparticle hazard ranking (on the basis of the HTS data). However, in order to improve the reliability of and increase confidence in the nano-SAR approach expanded experimental data sets are required.

4. Materials and Methods

4.1 Nano-SAR Development Workflow

The present approach for the development of nano-SAR classifiers followed the workflow summarized in Figure 1 (section 2.1). First, cytotoxicity of the nine metal oxide nanoparticles was assessed via screening of the plasma membrane leakage of the BEAS-2B cells. Toxic events (plasma membrane damage) were identified from the HTS data of nanoparticle exposed cells relative to unexposed control cells. Subsequently, an initial set of fundamental model input parameters were selected (ten nanoparticle descriptors and four different concentration measures). An exhaustive exploration of the model input parameter space was then performed, with different parameter subsets, in order to identify the best performing nano-SAR corresponding to the most suitable parameter subset.

4.2. Experimental Data Generation

The nine metal oxide nanoparticles were characterized in terms of their physicochemical properties as well as their *in vitro* cytotoxicity. Three nanoparticles (Al_2O_3, SiO_2, and Fe_3O_4) were acquired from commercial sources (Meliorum nanotechnologies, Rochester, NY, USA) and the remaining six (CeO_2, Co_3O_4, TiO_2, ZnO, CuO, and WO_3) were synthesized via flame spray pyrolysis (by Lutz et al., University of Bremen, Germany)[29, 33]. The nanoparticles were spherical with primary sizes in the range of 8–18 nm and density in the range of 2.2–7.22 $g \cdot cm^{-3}$ [46]. The nanoparticles' surface charge was

determined via zeta potential measurements (ZetaPALS, Brookhaven Instruments Corporation, Holtsville, NY) in water and as a function of pH and these measurements served to determine the isoelectric point (*IEP*, the pH at which a nanoparticle suspension has zero zeta potential) of the nanoparticles[47]. All measurements were conducted using 1.5 mL of 50 $mg \cdot L^{-1}$ aqueous nanoparticle dispersion and for each measurement five replicate runs of 10 cycles were collected. In addition, the *IEP* was also determined by measuring the zeta potential of each nanoparticle over a wide pH range (Table 1).

The cytotoxicity induced in BEAS-2B cells exposed to nanoparticle concentrated in the range of 0.375–200 $mg \cdot L^{-1}$ was assessed by measuring plasma membrane leakage via PI uptake[29] with the results quantified in terms of the percentage of membrane-damaged cells. Briefly, the cytotoxicity-screening assay was carried out using a set of six 384 well plates (16 rows by 24 columns; Greiner bio-one, Frickenhausen, Germany) containing both cells in a BEGM medium[29] exposed to nanoparticles (i.e., sample) in a range of concentrations and unexposed cells (i.e., control). In order to improve the reliability of toxic response identification[48], replicate samples and controls were used within each plate to estimate experimental variability (Table S1, Supplementary Material). Exposure of the cell population to six of the nanoparticles (Al_2O_3, CeO_2, Co_3O_4, TiO_2, ZnO, and CuO) was carried out at concentrations of 1, 5, 10, 15, 20, 25, 50, 100, 150, and 200 $mg \cdot L^{-1}$ with exposure times of 15 and 24 h. In addition, three different nanoparticles, Fe_3O_4, WO_3, and SiO_2, were selected for model validation. The concentration range and exposure times for SiO_2 (labeled as toxic at two concentrations; Table 2) was the same as for the six nanoparticles in the training set. In order to reasonably differentiate the validation from the training data set, the exposure times for Fe_3O_4, and WO_3 (labeled as non-toxic, Table 2) were set at 6 h and 24 h and the concentration range was set as 0.375–200 $mg \cdot L^{-1}$ at two fold increases.

4.3. Nano-SAR Descriptors

An initial set of ten nanoparticle descriptors was selected in order to describe the characteristics of the nanoparticles at reasonably diverse levels (i.e., molecular, chemical and physical information). The selection of simple constitutional descriptors included number of metal and oxygen atoms in the metal oxide (N_{Me}, N_O), atomic mass of the metal (m_{Me} ($g \cdot mol^{-1}$)) and the metal oxide molecular weight of (m_{MeO} ($g \cdot mol^{-1}$))[7]. Information regarding stability and the reactivity of the metal oxide was provided via the atomization energy[49, 50] (E_{MeO} ($kcal \cdot eqv^{-1}$)), which is a structure-dependent electronic property that is indicative of the bonding forces (i.e., energies) holding together the metal oxide elements in a standard state. Finally, the group and period (G_{Me} and P_{Me}) in the periodic table served to encode the periodic properties of the metal atom as these are often used to classify metals according to their catalytic and electronic properties.

The nanoparticle primary size (d (nm)) was also used as a simple geometric descriptor as it has been reported to be a key parameter impacting nanoparticle aggregation state and toxicity[51]. The nanoparticles isoelectric point (*IEP*) and the zeta potential in water (Z_w, mV) at the pH of the HTS assay (~7.4) were also selected as indicators of surface charge and aggregation tendency (e.g., as the operational pH approaches the *IEP*). Although the above descriptors do not provide direct information on the expected aggregation behavior in biological media, such descriptors do provide information for distinguishing among nanoparticles for classification purposes when for the same environmental conditions a single biological medium is utilized. It is also interesting to note that the edges of the energy bands of a metal oxide (related to its capacity to induce redox processes inside the cell) shift to higher or lower energy levels with respect to the difference between *IEP* and solution's pH[44]. It is noted that, relevant information regarding the particle aggregate size or size

distribution[7, 51] in the biological media was not utilized since such measurement are of low reliability in biological suspensions.

Finally, four different concentrations were also evaluated for model input, whereby the nanoparticle loading was expressed as concentrations on the basis of mass (C_m, mg·L^{-1}), surface area (C_s, m^2·L^{-1}), number (C_n, #·L^{-1})) and volume fraction (θ_v). Given the significant differences on metal oxide nanoparticle density (2.2–7.2 g·cm^{-3}), for the same mass loading the number, surface area and volume fractions will differ. The above concentration measures are related through the metal oxide nanoparticle density, ρ (g·cm^{-3}), and primary size (d) (i.e., $\theta_v=C_m/\rho$, $C_s=2\theta_v/d$ and $C_n=8\theta_v/d^3$) and hence for the purpose of establishing nano-SARs they provide different weighting of the importance of the nanoparticle primary size and density.

4.4. Data Processing

Cell-based toxicity screening often shows high variability that needs to be carefully handled[48, 52]. Therefore, the quality of the control data were first assessed, via the box plot method[53] for each HTS plate and the identified outliers (average of ~5% per plate) were removed for the subsequent analysis. Subsequently, the Strictly Standardized Mean Difference (SSMD[30, 31]) was used to identify and label measurable cytotoxicity effects in the nanoparticle exposed relative to the unexposed cell population. The SSMD (β) defined as:

$$\beta=\frac{\mu_{sample}-\mu_{control}}{\sqrt{\sigma^2_{sample}+\sigma^2_{control}}} \quad (2)$$

measures the magnitude of the (standardized) difference between the cell response for the sample and control (in which the mean and variance are denoted by μ_{sample}, σ^2_{sample}, and $\mu_{control}$, $\sigma^2_{control}$, respectively). In the present analysis a practical rule was developed (Table S2, Supporting Information) in order to control False-Negative labeling via use of the Maximum Likelihood Estimation (MLE,$\hat{\beta}$) of the SSMD.

4.5. Model Development, Validation and Interpretation

The model development approach followed a scheme that integrates the selection of relevant input parameters (feature selection), from an initial pool of input parameters, with model training and validation in order to identify the most suitable subset of input parameters that produces the optimal nano-SAR classifier (Figure 4). Both internal and external model validations were carried out in order to avoid over-fitting and assess model performance. Different nano-SAR classification models were developed and tested based on a logistic regression[54–56] expressed by:

$$\ln\left(\frac{\mathrm{P}(NP\in T)}{\mathrm{P}(NP\in N)}\right)=b+\sum\nolimits_i a_i NP_i \quad (3)$$

where P($NP \in T$) and P($NP \in N$) are the probabilities that a nanoparticle will be classified as toxic (T) or non-toxic (N), respectively, and NP_i is the i-th model input parameter (i.e., nanoparticle descriptor or concentration measure). If P($NP \in T$) > P($NP \in N$) (i.e., In(P($NP \in T$)/P($NP \in N$)) > 0), the nanoparticle will be classified as toxic, otherwise, it is considered non-toxic. The model development made use of an integrated approach (i.e., a wrapper approach[57]) in which the complete space of nano-SAR input parameters (i.e., 2^{14} different subsets) was explored to identify the best performing subset. Performance of the different logistic regression models (obtained with the different parameter subsets) was assessed with

a combined criterion encompassing classification accuracy (i.e., percent of correct classified samples) and number of false negatives (i.e., number of toxic nanoparticles identified as nontoxic). In order to assess the performance of each input parameter subset, the cytotoxicity data set was divided in two parts to quantify the internal and external validation performances. The internal validation data set (i.e., the training set) included Al_2O_3, CeO_2, Co_3O_4, TiO_2, ZnO, and CuO was used for model screening (i.e., detection of candidate models with reasonable accuracy). The average performance of each model was determined using a Leave-One-Out strategy[58] where each of the six nanoparticles was used in turn to test a model developed from the remaining five nanoparticles. The models of prediction accuracy higher than 95% and no false negatives were then evaluated via external validation data set (containingSiO_2, Fe_3O_4, and WO_3) to identify the best overall performing nano-SAR.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This material is based upon work supported by the National Science Foundation and the Environmental Protection Agency under Cooperative Agreement Number DBI-0830117. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the Environmental Protection Agency. This work has not been subjected to EPA review and no official endorsement should be inferred. Robert Rallo also acknowledges support provided by CICYT (Project CTQ2009-14627), Generalitat de Catalunya (2009SGR-01529) and the EU Commission (OSIRIS, Contract No. 037017). Support was also provided by the US Public Health Service Grants U19 ES019528 (UCLA Center for NanoBiology and Predictive Toxicology) and R01 ES 016746.

References

1. Guo, Z.; Tan, L. Fundamentals and Applications of Nanomaterials. 1. Artech House Publishers; 2009.
2. The Project on Emerging Nanotechnologies: Consumer Products Inventory. Woodrow Wilson International Center; 2010. Available from: http://www.nanotechproject.org/inventories/consumer/
3. Colvin VL. The potential environmental impact of engineered nanomaterials. Nature Biotechnology. 2003; 21(10):1166–70.
4. Ray PC, Yu HT, Fu PP. Toxicity and Environmental Risks of Nanomaterials: Challenges and Future Needs. Journal of Environmental Science and Health Part C-Environmental Carcinogenesis & Ecotoxicology Reviews. 2009; 27(1):1–35.
5. Nel A, Xia T, Madler L, Li N. Toxic potential of materials at the nanolevel. Science. 2006; 311(5761):622–627. [PubMed: 16456071]
6. Cattaneo AG, Gornati R, Sabbioni E, Chiriva-Internati M, Cobos E, Jenkins MR, Bernardini G. Nanotechnology and human health: risks and benefits. Journal of Applied Toxicology. 2010; 30(8): 730–44. [PubMed: 21117037]
7. Puzyn T, Leszczynska D, Leszczynski J. Toward the development of “nano-QSARs”: advances and challenges. Small. 2009; 5(22):2494–509. [PubMed: 19787675]
8. Toropov AA, Leszczynska D, Leszczynski J. Predicting water solubility and octanol water partition coefficient for carbon nanotubes based on the chiral vector. Computational Biology and Chemistry. 2007; 31(2):127–8. [PubMed: 17363328]
9. Martin D, Maran U, Sild S, Karelson M. QSPR modeling of solubility of polyaromatic hydrocarbons and fullerene in 1-octanol and n-heptane. Journal of Physical Chemistry B. 2007; 111(33):9853–7.
10. Sivaraman N, Srinivasan TG, Rao PRV, Natarajan R. QSPR modeling for solubility of fullerene (C-60) in organic solvents. Journal of Chemical Information and Computer Sciences. 2001; 41(4): 1067–1074. [PubMed: 11500126]

11. Toropov AA, Leszczynska D, Leszczynski J. QSPR study on solubility of fullerene C60 in organic solvents using optimal descriptors calculated with SMILES. Chemical Physics Letters. 2007; 441(1–3):119–122.
12. Liu H, Yao X, Zhang R, Liu M, Hu Z, Fan B. Accurate quantitative structure-property relationship model to predict the solubility of C60 in various solvents based on a novel approach using a least-squares support vector machine. Journal of Physical Chemistry B. 2005; 109(43):20565–71.
13. Toropov AA, Leszczynski J. A new approach to the characterization of nanomaterials: Predicting Young's modulus by correlation weighting of nanomaterials codes. Chemical Physics Letters. 2006; 433(1–3):125–129.
14. Toropov AA, Toropova AP, Benfenati E, Leszczynska D, Leszczynski J. Additive InChI-based optimal descriptors: QSPR modeling of fullerene C-60 solubility in organic solvents. Journal of Mathematical Chemistry. 2009; 46(4):1232–1251.
15. Sayes C, Ivanov I. Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis. 2010; 30(11):1723–1734. [PubMed: 20561263]
16. Fourches D, Pu D, Tassa C, Weissleder R, Shaw SY, Mumper RJ, Tropsha A. Quantitative nanostructure-activity relationship modeling. ACS Nano. 2010; 4(10):5703–12. [PubMed: 20857979]
17. Zheng W, Tropsha A. Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. Journal of Chem Information and Computer Science. 2000; 40(1):185–94.
18. Weissleder R, Kelly K, Sun EY, Shtatland T, Josephson L. Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. Nature Biotechnology. 2005; 23(11):1418–23.
19. Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. 1. Cambridge University Press; 2000.
20. Shaw SY, Westly EC, Pittet MJ, Subramanian A, Schreiber SL, Weissleder R. Perturbational profiling of nanomaterial biologic activity. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(21):7387–7392. [PubMed: 18492802]
21. Rabolli V, Thomassen LC, Princen C, Napierska D, Gonzalez L, Kirsch-Volders M, Hoet PH, Huaux F, Kirschhock CE, Martens JA, Lison D. Influence of size, surface area and microporosity on the in vitro cytotoxic activity of amorphous silica nanoparticles in different cell types. Nanotoxicology. 2010; 4(3):307–18. [PubMed: 20795912]
22. Waters KM, Masiello LM, Zangar RC, Tarasevich BJ, Karin NJ, Quesenberry RD, Bandyopadhyay S, Teeguarden JG, Pounds JG, Thrall BD. Macrophage responses to silica nanoparticles are highly conserved across particle sizes. Toxicological Sciences. 2009; 107(2): 553–69. [PubMed: 19073995]
23. Okuda-Shimazaki J, Takaku S, Kanehira K, Sonezaki S, Taniguchi A. Effects of Titanium Dioxide Nanoparticle Aggregate Size on Gene Expression. International Journal of Molecular Sciences. 2010; 11(6):2383–2392. [PubMed: 20640159]
24. Poater A, Saliner AG, Carbo-Dorca R, Poater J, Sola M, Cavallo L, Worth AP. Modeling the structure-property relationships of nanoneedles: A journey toward nanomedicine. Journal of Computational Chemistry. 2009; 30(2):275–84. [PubMed: 18615420]
25. Poater A, Saliner AG, Sola M, Cavallo L, Worth AP. Computational methods to predict the reactivity of nanoparticles through structure-property relationships. Expert Opinion on Drug Delivery. 2010; 7(3):295–305. [PubMed: 20201736]
26. Qin LJ, Zhang YH, Huang SP, Tian HP, Wang P. Atomic-scale structure of Co-Pt bimetallic nanoparticles: Monte Carlo simulations. Physical Review B. 2010; 82(7):075413.
27. Tian P. Molecular dynamics simulations of nanoparticles. Annual Reports on the Progress of Chemistry, Section C: Physical Chemistry. 2008; (104):142–164.
28. Guidance document on the validation of (Quantitative) structure-activity relationships [(Q)SAR] models. OECD Environment Health and Safety Publications Series on Testing and Assessment, Environment Directorate Organisation for Economic CoOperation and Development; Paris: No. 69ENV/JM/MONO(2007)2)
29. George S, Pokhrel S, Xia T, Gilbert B, Ji ZX, Schowalter M, Rosenauer A, Damoiseaux R, Bradley KA, Madler L, Nel AE. Use of a Rapid Cytotoxicity Screening Approach To Engineer a

Safer Zinc Oxide Nanoparticle through Iron Doping. ACS Nano. 2010; 4(1):15–29. [PubMed: 20043640]

30. Zhang XD. A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. Journal of Biomolecular Screening. 2007; 12(5):645–55. [PubMed: 17517904]
31. Zhang XD. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. Genomics. 2007; 89(4):552–561. [PubMed: 17276655]
32. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, Smith Q, Beijersbergen RL, Ghazal P, Shamu CE. Statistical methods for analysis of high-throughput RNA interference screens. Nature Methods. 2009; 6(8): 569–75. [PubMed: 19644458]
33. Xia T, Kovochich M, Liong M, Madler L, Gilbert B, Shi HB, Yeh JI, Zink JI, Nel AE. Comparison of the Mechanism of Toxicity of Zinc Oxide and Cerium Oxide Nanoparticles Based on Dissolution and Oxidative Stress Properties. ACS Nano. 2008; 2(10):2121–2134. [PubMed: 19206459]
34. Karlsson HL, Cronholm P, Gustafsson J, Moller L. Copper oxide nanoparticles are highly toxic: A comparison between metal oxide nanoparticles and carbon nanotubes. Chemical Research in Toxicology. 2008; 21(9):1726–1732. [PubMed: 18710264]
35. Meng H, Chen Z, Xing GM, Yuan H, Chen CY, Zhao F, Zhang CC, Zhao YL. Ultrahigh reactivity provokes nanotoxicity: Explanation of oral toxicity of nano-copper particles. Toxicology Letters. 2007; 175(1–3):102–110. [PubMed: 18024012]
36. Park EJ, Park K. Oxidative stress and pro-inflammatory responses induced by silica nanoparticles in vivo and in vitro. Toxicology Letters. 2009; 184(1):18–25. [PubMed: 19022359]
37. Li Y, Pan D, Liu J, Kern PS, Gerberick GF, Hopfinger AJ, Tseng YJ. Categorical QSAR Models for skin sensitization based upon local lymph node assay classification measures part 2: 4D-fingerprint three-state and two-2-state logistic regression models. Toxicological Sciences. 2007; 99(2):532–44. [PubMed: 17675333]
38. Worth AP, Cronin MTD. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. Journal of Molecular Structure-Theochem. 2003; 622(1–2):97–111.
39. Grassian VH, Adamcakova-Dodd A, Pettibone JM, O'Shaughnessy PT, Thorne PS. Inflammatory response of mice to manufactured titanium dioxide nanoparticles: Comparison of size effects through different exposure routes. Nanotoxicology. 2007; 1(3):211–226.
40. Chithrani BD, Chan WCW. Elucidating the mechanism of cellular uptake and removal of protein-coated gold nanoparticles of different sizes and shapes. Nano Letters. 2007; 7(6):1542–1550. [PubMed: 17465586]
41. Jiang JK, Oberdorster G, Biswas P. Characterization of size, surface charge, and agglomeration state of nanoparticle dispersions for toxicological studies. Journal of Nanoparticle Research. 2009; 11(1):77–89.
42. Petosa AR, Jaisi DP, Quevedo IR, Elimelech M, Tufenkji N. Aggregation and Deposition of Engineered Nanomaterials in Aquatic Environments: Role of Physicochemical Interactions. Environmental Science & Technology. 2010; 44(17):6532–6549. [PubMed: 20687602]
43. Rivera Gil P, Oberdorster G, Elder A, Puntes V, Parak WJ. Correlating Physico-Chemical with Toxicological Properties of Nanoparticles: The Present and the Future. ACS Nano. 2010; 4(10): 5527–5531. [PubMed: 20973573]
44. Burello E, Worth AP. A theoretical framework for predicting the oxidative stress potential of oxide nanoparticles. Nanotoxicology. 2010 Early Online 1–8.
45. Vold, RD.; Vold, MJ. Colloid and Interface Chemistry. 1. Addison-Wesley; 1983.
46. Lide, DR., editor. Handbook of Chemistry and Physics. Vol. 90. CRC Press; 2009.
47. Ji ZX, Jin X, George S, Xia TA, Meng HA, Wang XA, Suarez E, Zhang HY, Hoek EMV, Godwin H, Nel AE, Zink JI. Dispersion and Stability Optimization of TiO_2 Nanoparticles in Cell Culture Media. Environmental Science & Technology. 2010; 44(19):7309–7314. [PubMed: 20536146]
48. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. Nature Biotechnol. 2006; 24(2):167–75. [PubMed: 16465162]

49. Sanderson, RT. Inorganic Chemistry. Van Nostrand Reinhold; 1967.
50. Koelling DD, Boring AM, Wood JH. The Electronic-Structure of CeO2 and PrO2. Solid State Communications. 1983; 47(4):227–232.
51. Zuin, S.; Pojana, G.; Marcomini, A. Effect-Oriented Physicochemical Characterization of Nanomaterials. In: Monteiro-Riviere, NA.; Tran, CL., editors. Nanotoxicology: characterization, dosing and health effects. Informa Healthcare; 2007.
52. Harper G, Pickett SD. Methods for mining HTS data. Drug Discovery Today. 2006; 11(15–16): 694–699. [PubMed: 16846796]
53. Han, J.; Kamber, M. Data Mining: Concepts and Techniques. 1. Morgan Kaufmann; 2000.
54. Wasserman, L. Springer texts in statistics. New York: Springer; 2004. All of statistics : a concise course in statistical inference.
55. Bishop, CM. Information Science and Statistics. 2. Springer; Pattern Recognition and Machine Learning.
56. LeCessie S, vanHouwelingen JC. Ridge Estimators in Logistic Regression. Applied Statistics. 1992; 41(1):191–201.
57. Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on Knowledge and Data Engineering. 2005; 17(4):491–502.
58. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross Validation. In: Liu, L.; Özsu, MT., editors. Encyclopedia of Database Systems. Springer; 2009.

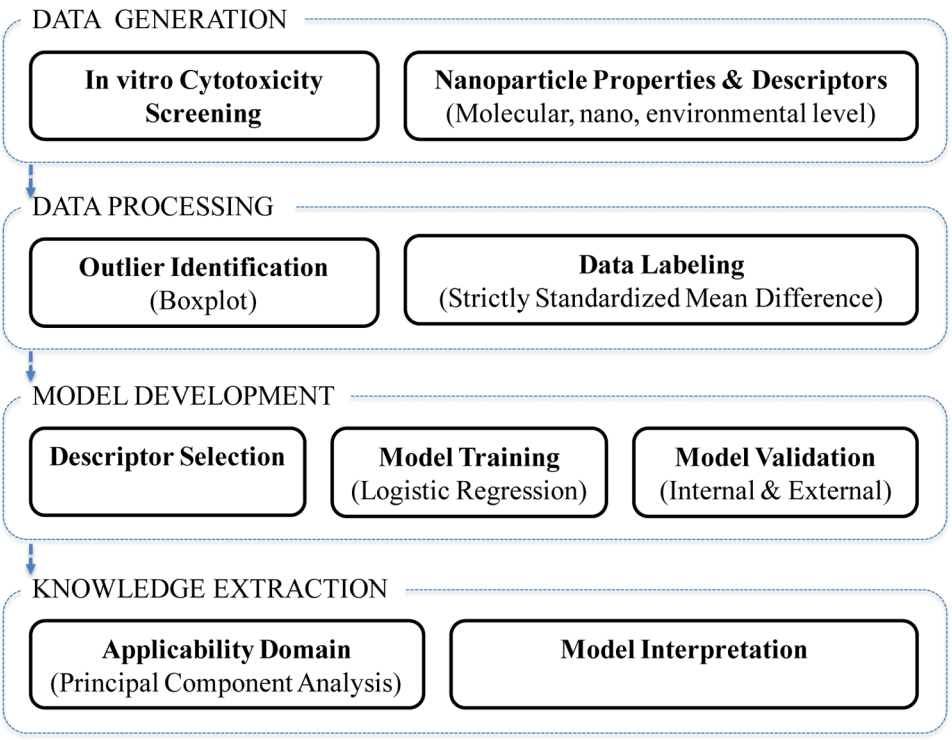

Figure 1.
Workflow for nano-SAR development, validation and interpretation

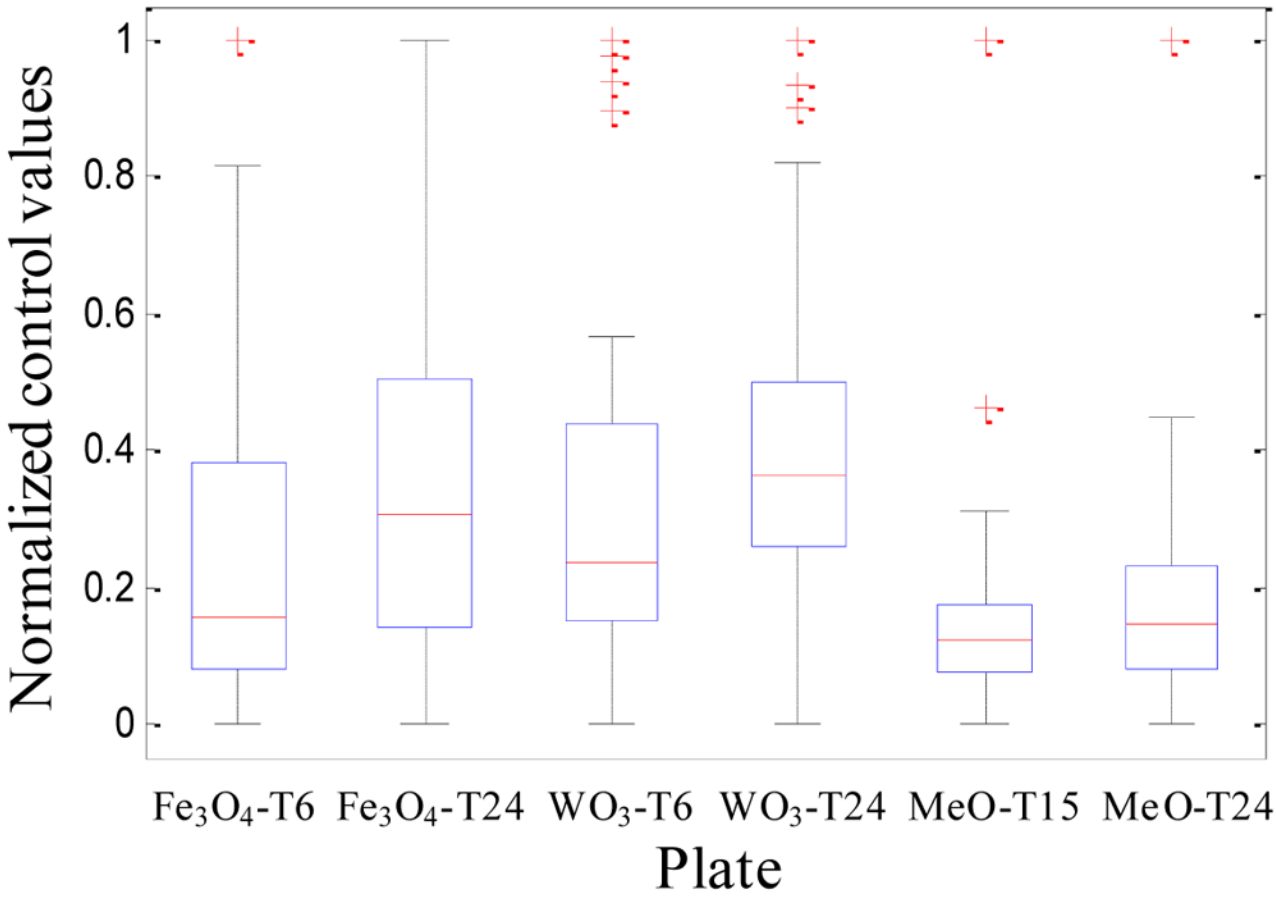

Figure 2.

Distribution of the control values (normalized over the range of [0, 1]) in the HTS plates. The upper and lower bound of a box identify the first quartile (Q_1) and the third quartile (Q_3) of the data, the line inside a box represent the median of the data; ends of the whiskers are the smallest observation and the largest observation within the inter-quartile range (defined by [Q_1-1.5(Q_3-Q_1), Q_3+1.5(Q_3-Q_1)]), and outliers (i.e., the data outside the inter-quartile range) are marked by "+". MeO-T15 and MeO-T24identify the plates that contained Al_2O_3, CeO_2, Co_3O_4, TiO_2, ZnO, CuO, and SiO_2 after 15 h and 24 h of exposure, respectively. Fe_3O_4-T6 and Fe_3O_4-T24 identify the plates containing Fe_3O_4 after exposure of 6 h. and 24 h, respectively. WO_3-T6 and WO_3-T24 identify the plates containing WO_3 after 6 h and 24 h, respectively.

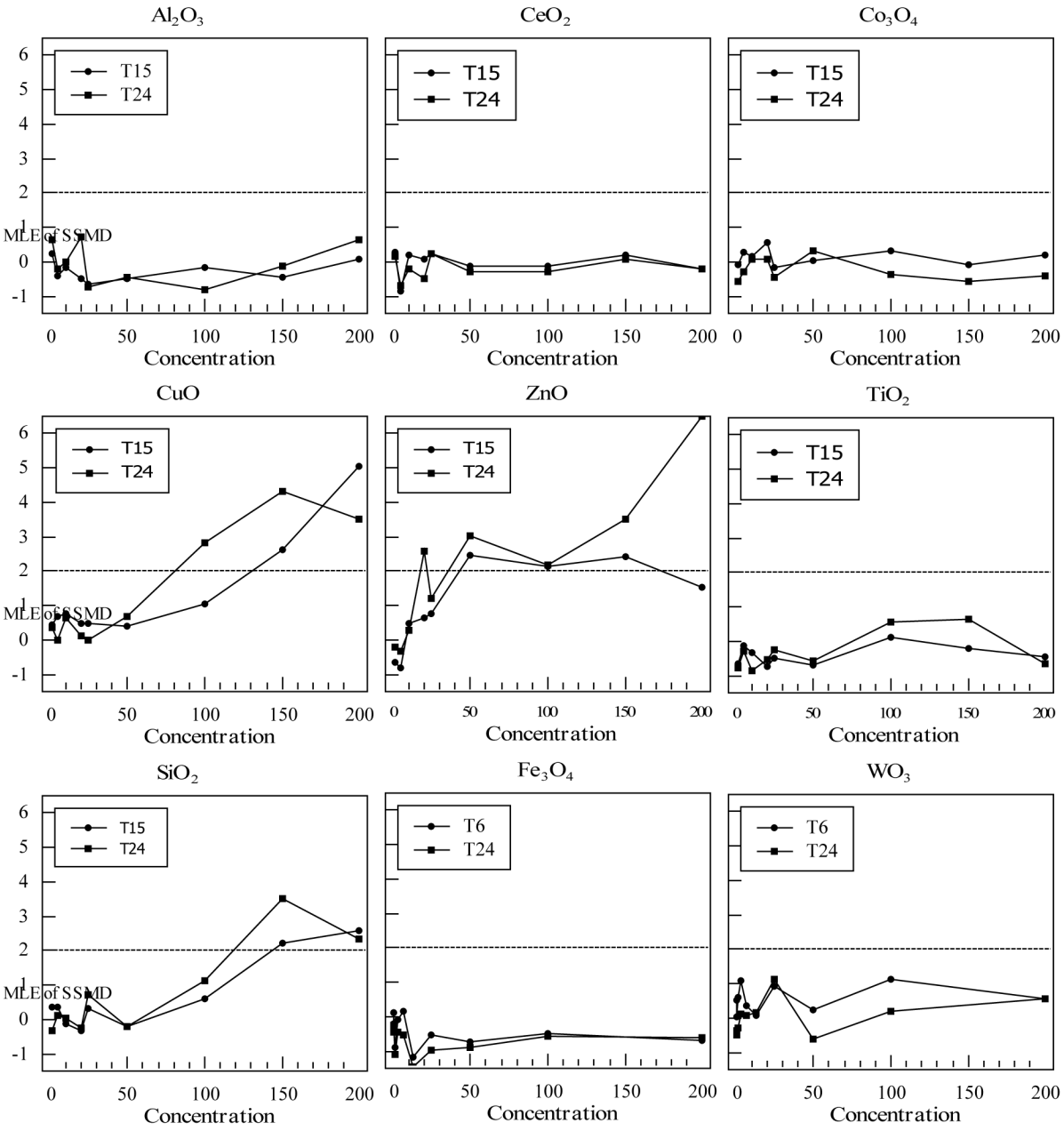

Figure 3.

The Maximum Likelihood Estimates (MLE) of SSMD for nanoparticles of different concentrations (mg $\cdot L^{-1}$) and exposure times (labeled as T6, T15, and T24 for 6h, 15h, and 24, respectively) (The red dash line corresponds to the threshold of MLE of SSMD=2)

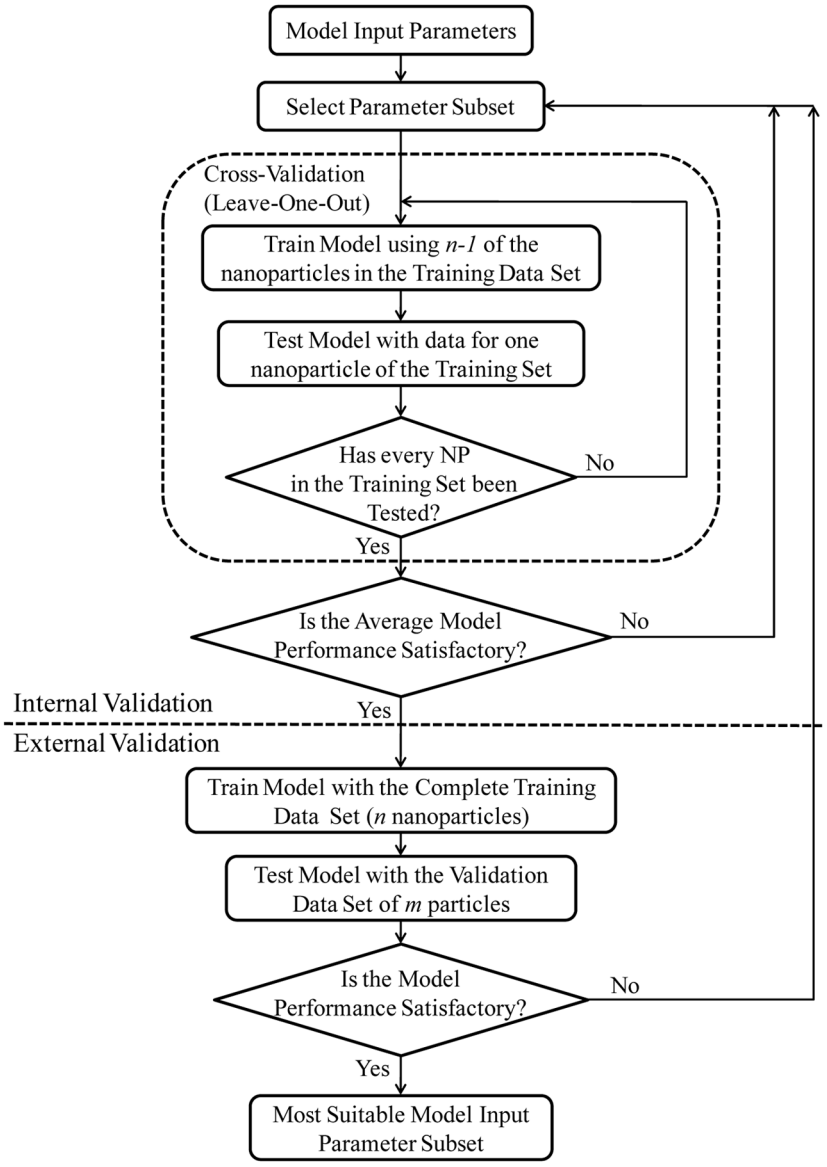

Figure 4.
The integrated scheme for nano-SAR model development

Table 1

Nanoparticle Descriptors

NP [a]	N_{Me}	N_O	m_{Me}	m_{MeO}	G_{Me}	P_{Me}	E_{MeO}	d	Z_w	IEP
Al_2O_3	2	3	26.98	101.96	13	3	122	12	7.75	7.60
CeO_2	1	2	140.12	172.12	3	6	116	10	12.74	7.90
Co_3O_4	3	4	58.93	240.79	9	4	93	10	34.58	9.60
TiO_2	1	2	47.87	79.87	4	4	148	18	−21.19	6.90
ZnO	1	1	65.38	81.38	12	4	87	10	29.26	9.60
CuO	1	1	63.55	79.55	11	4	89	10	−9.29	7.30
SiO_2	1	2	28.09	60.09	14	3	108	19	−31.48	1.00
Fe_3O_4	3	4	55.85	231.55	8	4	100	8	−31.60	4.14
WO_3	1	3	183.84	231.84	6	6	96	10	−61.77	0.23

[a] (N_M and N_O: number of metal and Oxygen atoms, m_{Me} (g·mol^{-1}): atomic mass of the nanoparticle metal, m_{MeO} (g·mol^{-1}): molecular weight of the metal oxide, G_{Me} and P_{Me}: group and period of the nanoparticle metal, E_{MeO} (kcal·eqv^{-1}): atomization energy of the metal oxide, d (nm): nanoparticle primary size, Z_w (mV): zeta potential (in water at pH=7.4), IEP: isoelectric point.)

Table 2

Toxicity Labels of Nanoparticle at given Concentration

C_m [a]	1	5	10	20	25	50	100	150	200
Al_2O_3	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox
CeO_2	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox
Co_3O_4	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox
TiO_2	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox
ZnO	Nontox	Nontox	Nontox	Nontox	Nontox	Tox	Tox	Tox	Tox
CuO	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Tox	Tox
SiO_2	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Tox	Tox

C_m	0.375	0.75	1.6	3.12	6.25	12.5	25	50	100	200
Fe_3O_4	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox
WO_3	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox	Nontox

[a](C_m (mg·L^{-1}): nanoparticle mass concentration. Volume fraction (θ_V (10^{-6})), surface concentration (C_S (m^2·L^{-1})) and number concentration (C_n (10^{22}·L^{-1})) can be derived from nanoparticle mass concentration by $\theta_V=C_m/\rho$, $C_S=2\theta_V/d$ and $C_n=8\theta_V/d^3$. Here ρ (g·cm^{-3}) is the nanoparticle density and d (nm) is the primary size. The above four different concentration measures together with the ten nanopaticle descriptors (Table 1) comprise the initial set of fourteen input parameters (i.e., features) for nano-SAR development.)

Table 3

Nano-SAR input parameters and performance measures for selected models

Model	Descriptors	Internal validation		External validation	
		Acc [a]	FN [b]	Acc	FN
M1	$\boldsymbol{P_{Me}, E_{MeO}, d, C_v}$	**100%**	**0**	**100%**	**0**
M2	$P_{Me}, E_{MeO}, d, C_n, C_s$	100%	0	96%	0
M3	N_O, G_{Me}, IEP, C_m	98%	0	93%	2
M4	$N_{Me}, N_O, G_{Me}, IEP, C_m$	98%	0	93%	2
M5	$N_O, G_{Me}, P_{Me}, IEP, C_m$	96%	0	93%	2
M6	$N_O, m_{MeO}, P_{Me}, Z_w, C_m$	96%	0	93%	2
M7	$G_{Me}, P_{Me}, E_{MeO}, C_v$	96%	0	93%	0

[a] (Acc: classification accuracy)

[b] (FN: number of false negatives)