ACCOUNTS
of chemical research

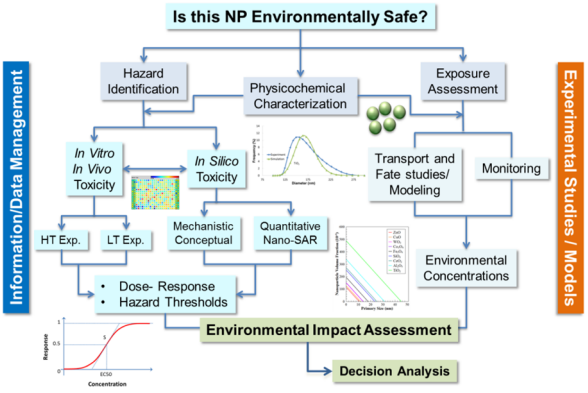
In Silico Analysis of Nanomaterials Hazard and Risk

YORAM COHEN,*,† ROBERT RALLO,‡ RONG LIU,† AND HAOYANG HAVEN LIU†

†*Chemical and Biomolecular Engineering Department and Center for Environmental Implications of Nanotechnology, University of California, Los Angeles, Los Angeles, California 90064, United States, and* ‡*Departament d'Enginyeria Informatica i Matematiques, Universitat Rovira i Virgili, Av. Paisos Catalans 26, 43007 Tarragona, Catalunya, Spain*

RECEIVED ON MARCH 3, 2012

CONSPECTUS



Because a variety of human-related activities, engineered nanoparticles (ENMs) may be released to various environmental media and may cross environmental boundaries, and thus will be found in most media. Therefore, the potential environmental impacts of ENMs must be assessed from a multimedia perspective and with an integrated risk management approach that considers rapid developments and increasing use of new nanomaterials.

Accordingly, this Account presents a rational process for the integration of *in silico* ENM toxicity and fate and transport analyses for environmental impact assessment. This approach requires knowledge of ENM toxicity and environmental exposure concentrations. Considering the large number of current different types of ENMs and that those numbers are likely to increase, there is an urgent need to accelerate the evaluation of their toxicity and the assessment of their potential distribution in the environment.

Developments in high throughput screening (HTS) are now enabling the rapid generation of large data sets for ENM toxicity assessment. However, these analyses require the establishment of reliable toxicity metrics, especially when HTS includes data from multiple assays, cell lines, or organisms. Establishing toxicity metrics with HTS data requires advanced data processing techniques in order to clearly identify significant biological effects associated with exposure to ENMs.

HTS data can form the basis for developing and validating *in silico* toxicity models (e.g., quantitative structure–activity relationships) and for generating data-driven hypotheses to aid in establishing and/or validating possible toxicity mechanisms. To correlate the toxicity of ENMs with their physicochemical properties, researchers will need to develop quantitative structure–activity relationships for nanomaterials (i.e., nano-SARs). However, as nano-SARs are applied in regulatory applications, researchers must consider their applicability and the acceptance level of false positive relative to false negative predictions and the reliability of toxicity data.

To establish the environmental impact of ENMs identified as toxic, researchers will need to estimate the potential level of environmental exposure concentration of ENMs in the various media such as air, water, soil, and vegetation. When environmental monitoring data are not available, models of ENMs fate and transport (at various levels of complexity) serve as alternative approaches for estimating exposure concentrations. Risk management decisions regarding the manufacturing, use, and environmental regulations of ENMs would clearly benefit from both the assessment of potential ENMs exposure concentrations and suitable toxicity metrics. The decision process should consider the totality of available information: quantitative and qualitative data and the analysis of nanomaterials toxicity, and fate and transport behavior in the environment.

Effective decision-making to address the potential impacts of nanomaterials will require considerations of the relevant environmental, ecological, technological, economic, and sociopolitical factors affecting the complete lifecycle of nanomaterials, while accounting for data and modeling uncertainties. Accordingly, researchers will need to establish standardized data management and analysis tools through nanoinformatics as a basis for the development of rational decision tools.

Published on the Web 11/08/2012 www.pubs.acs.org/accounts
10.1021/ar300049e

1. Introduction

Engineered nanomaterials (ENMs) which may be released to the environment as the result of a variety of human-related activities (air emissions and/or direct discharge to surface water, etc.) move across environmental boundaries and are therefore likely to be found in most media.[1] In order to appropriately assess the potential environmental impact of ENMs, it is imperative to evaluate existing and potential releases of ENMs to various environmental media. Such information is essential for estimating the expected concentration levels of ENMs in the environment and thus a possible level of exposures of ecological receptors to ENMs via multiple exposure pathways.

In order to evaluate the potential environmental impact of ENMs, one must then assess the expected levels of environmental concentrations at the exposure locations.[2] Such information can be provided via field monitoring or modeling of the fate of transport of ENMs in the environment. Environmental field monitoring of the concentrations of ENMs would clearly be valuable; however, this is a daunting and costly endeavor that would be impractical for the increasing number of ENMs. In this regard, estimation methods such as those based on life-cycle analysis of ENMs[3] can inform decision makers as to the potential environmental releases of ENMs during their manufacturing, use, and product disposal. Estimates of the potential environmental concentrations of ENMs as a result of various release scenarios could then be evaluated based on suitable fate and transport models that consider the environmental distribution of ENMs in the various environmental media (e.g., air, soil, water, sediment, and vegetation). Such information can serve to evaluate the possible exposures of ecological receptors to ENMs via multiple exposure pathways.

Risk characterization requires quantitative dose–response relationships for the target receptors and/or acceptable extrapolation from the test species.[4] In the absence of such information, alternative conservative approaches can be undertaken where acceptable dose (or dose below which there is no observed hazardous effect) can be established.[4] Such metrics can also be used, as in the case of quantitative risk assessment, to arrive at protective measures to establish allowable environmental concentrations. In some cases, it may only be possible to assert whether or not a given ENM may be potentially hazardous based on certain toxicity metrics, but without reliable concentration threshold. Although the quality of such information is below that of quantitative dose–response relations, it can be useful for making more informed decisions regarding the safe design, use, and disposal of ENMs.

The risk assessment process may contain various explicit and implicit assumptions given the quality and availability of data for various components of such analysis.[5] Clearly, it is desired to have convincing experimental data to establish if specific ENMs may be hazardous and if so to establish quantitative dose–response relationships (Figure 1). However, it must be recognized that, given the rapid developments in nanotechnology[6] and thus additions of many

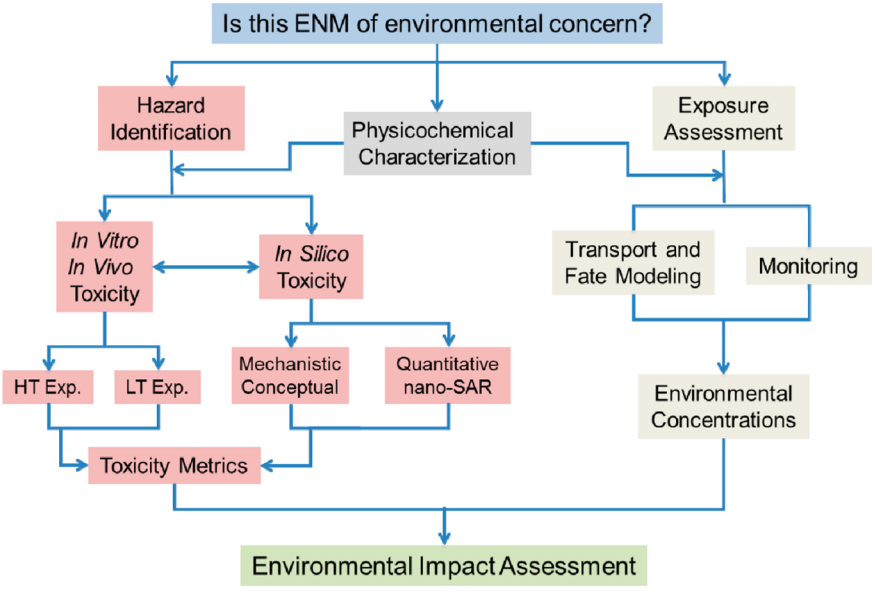


FIGURE 1. Schematic illustration of the components for environmental impact assessment associated with exposure to nanomaterials. (HT Exp.: High-Throughput Experiment; LT Exp.: Low-Throughput Experiment; nano-SAR: nanostructure–activity relationship).

different types of ENMs, there is also a need for hazard identification, establishing and correlating ENMs potential toxicity metrics with ENM properties and estimating the range of possible ENM exposure concentrations (Figure 1). In this regard, when experimental and field data are unavailable, *in silico* toxicity predictions[7] and models of ENMs fate and transport[8] are alternative approaches to arriving at the needed information for environmental impact assessment, the collective approach requiring ENM data managements and model building which are the focus of the emerging field of nanoinformatics.

2. Nanoparticle Hazard Identification

Characterization of the environmental hazard related to exposure to ENMs requires evaluation of adverse effects at multiple scales, ranging from molecular level (e.g., gene and protein expression) to ecosystem impacts that need to be assessed via *in vitro*, *in vivo*, and mesocosm studies for comprehensive hazard identification. At present, however, data scarcity and uncertainty are impediments to unambiguous ENMs hazard identification.[9] At the same time, given the rapid growth of nanotechnology in terms of the volume and different types of ENMs, there is a need for accelerated approaches to toxicity screening. In this regard, the combination of technologies for assay miniaturization (e.g., 384/1536 well plates) using single cell lines and simple organisms (e.g., zebrafish[10]), laboratory automation and robotic equipment enables high throughput screening (HTS[11]) testing of nanoparticles toxicity in biological systems. Indeed, in recent years, multiparametric assays involving multiple concentrations, exposure times, target cell lines, and toxicity end points have emerged as suitable platforms for rapid nanotoxicity screening via HTS toxicity studies.[12] HTS data can be invaluable for developing and validating *in silico* data-driven toxicity models (e.g., quantitative structure−activity relations) and for generating data-driven hypothesis regarding possible toxicity mechanisms. However, the use of HTS assays for hazard identification requires advanced data processing techniques in order to clearly identify significant biological effects associated with exposure to ENMs.

2.1. Processing of High-Throughput Nanotoxicity Data Sets. Data-driven techniques for knowledge extraction and toxicity model development require high quality data. However, rapid screening techniques such as HTS assays are vulnerable to systematic and random errors.[13] For instance, variations in liquid dispensing, cell growth variability, or fluctuations in nanoparticle concentration due to solvent evaporation can introduce significant levels of noise in HTS data.[14] The use of replicate measurements and procedural quality controls to reduce assay variability can, partially, compensate for random errors.[15] Systematic errors induced by across-plate and within-plate row and column biases are much more difficult to manage and require within plate reference controls.[15] Indeed, the use of control wells (e.g., cells not exposed to nanoparticles) is essential for assessing plate-to-plate variability in multiplate assays and for establishing proper assay background levels.[15] In this regard, normalization of raw HTS data, using control information, is required to remove systematic plate-to-plate variability and define toxicity metrics of the same statistically meaning so as to compare experimental measurements across plates. Also, data outliers must be identified and removed before HTS data normalization using acceptable statistical techniques such as the boxplot method or more sophisticated analysis.[16] HTS data normalization should, whenever feasible, be accomplished with clear interpretation of statistical parameters such as has been recently shown employing the strictly standardized mean difference (SSMD[15]). The SSMD can be used to quantify the statistical significance of the differences in the replicates of responses of exposed cells relative to the unexposed cell population. The SSMD, unlike other commonly used statistical measures such as Student's *t* test, accounts for intrinsic data variability without underestimating the likelihood of population similarity (i.e., lower *p*-values) with increasing sample size. A detailed discussion of additional statistical methods for HTS data preprocessing can be found elsewhere.[17]

Following HTS data normalization, hit-identification methods[17] can be applied to detect significant biological responses induced by nanomaterials. In such analysis, one must carefully control false-positive and false-negative identifications as it is crucial to ensure the data quality, which in turn is fundamental for the development of reliable models. False negative identifications (i.e., nanoparticles that induce adverse effects identified as no-effect nanoparticles) have significant implications in nanotoxicology and regulatory decision-making. A common practice to minimize false identifications is to apply predefined activity threshold levels to discriminate between nanoparticles that that lead to active (hits) and nonactive (non-hits) biological response. This approach tends to detect highly active nanoparticles as hits whereas lower activity nanoparticles, close to the threshold level, can then have a higher probability of being missed due to measurement errors.[13] Different methods have been proposed in the literature to address this critical

issue including the use of local thresholds based on similarity clustering.[18] It is noted that the use of SSMD for data normalization facilitates control of the false negative level during the hit-identification process and thus allows the definition of statistically significant thresholds.[19]

Screening of nanoparticle toxicity can be carried out over a broad domain of nanoparticle characteristics (e.g., size, surface charge, shape, and aggregation state), environmental parameters (pH, temperature, salinity, solution chemistry), and assay conditions (concentrations, exposure time, cell lines, measured response). However, due to the lack of a priori quantitative models of nanotoxicity, the challenge is to determine if the generated data sets span an adequate range (magnitude) of the selected nanoparticles types and properties, as well as environmental conditions. In this regard, feature selection[20] algorithms help reduce the required dimension of the nanoparticle data sets by keeping only information that is truly relevant for hazard identification. Feature selection analysis can assist in determining if experimental parameters span a range that is too narrow for extracting useful information.[20] A major benefit of using feature selection in HTS data analysis is in data and model complexity reduction.

Clearly, the definition of meaningful nanotoxicity metrics based on HTS processed data is of paramount importance for *in silico* hazard identification. For example, the definition of metrics that are based on averaged cell responses across multiple assay conditions must consider possible cancellation effects due to the coexistence of positive (e.g., up-regulated) and negative (e.g., down-regulated) values. In a similar way, cytotoxicity end points should be defined on the basis of irreversible biological effects (e.g., cell membrane damage). Accordingly, use of the proper toxicity metrics (based on suitable end point definition) will be fundamental for the interpretability of subsequent nanotoxicity models.

2.2. Knowledge Extraction from High-Throughput Screening Data. Data mining techniques[21] can be applied during exploratory analysis of HTS data to extract information for hypothesis formulation of possible toxicity mechanisms and relationship among different cell responses, and relevance of environmental conditions and ENMs properties. At the basic level of analysis, heat maps combined with hierarchical clustering are most commonly used for high-throughput data exploration. Heat maps provide, via row and/or column clustering, ordered representations of data that facilitate identification of similarity patterns. For instance, hierarchical clustering analysis applied to *in vitro* HTS data identified nanomaterials with similar patterns of biologic activity across a broad sampling of cellular contexts.[22] More comprehensive visualizations can be obtained via topology preserving mapping techniques such as the self-organizing map (SOM[23]). Briefly, SOM provides an ordered 2D projection of data vectors, which for the case of HTS analysis contain the cell response information. These 2D projections form cell lattices such that distances among cells are preserved in relation to the similarity of elements of the original HTS data set.[23] Such topology preservation allows, for example, clustering of nanoparticles of similar behavior on neighboring regions of the 2D map and facilitates the identification and visualization of groups of nanoparticles that trigger similar biological responses. Layers of SOM maps then form component planes that contain additional ordered representations of the detailed information (e.g., specific cell responses) over the SOM space (i.e., mapped unit cells).

Clusters of similar nanoparticles identified either using heatmaps or SOM need to be validated to ensure their significance. Cluster validation is challenging for HTS data since (a) lack of prior knowledge about the expected cluster structure, and (b) clustering of a data set of high dimensionality is sensitive to data quality (e.g., both in terms of data set size and data uncertainty). Techniques based on statistical sampling such as consensus clustering[24] should be used to quantify the quality (e.g., validity and stability) of the clusters. The use of validated SOM analysis for knowledge extraction from HTS was recently demonstrated for cell signaling pathways related to nanotoxicity.[25] SOM analysis identified groups of nanoparticles that induced similar effects on the regulation of signaling pathways in RAW264.7 macrophage cells. For example, the analysis revealed that exposure to high concentrations (>20 $mg \cdot L^{-1}$) of ZnO and Pt nanoparticles induced a significant up-regulation of pathways related to DNA damage.

SOM analysis can also be applied to integrated analysis of nanoparticle biological activity profiles encompassing multiple assay conditions, cell types, and cell responses. Figure 2 demonstrates the application of SOM analysis to an HTS data set that includes activity of 10 cell signaling pathways and 4 cytotoxicity responses for RAW264.7 macrophages and BEAS-2B epithelial cells exposed to 6 metal and metal oxide nanoparticles.[12,25] A complete description of the HTS assays including details of the experimental protocols and data preprocessing can be found elsewhere.[12] Briefly, the SOM provides a compact data visualization and facilitates identification and interpretation of similarities of biological response profiles (i.e., clusters). Cluster I identifies similarities

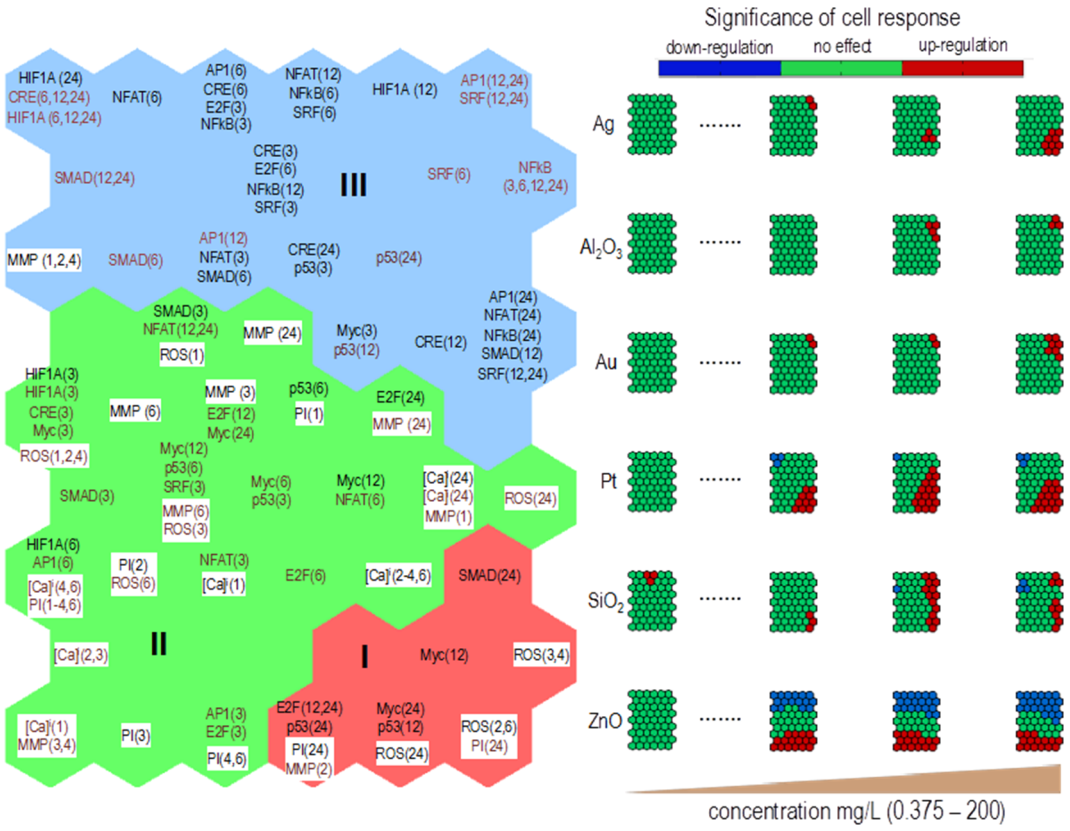


FIGURE 2. SOM clusters of biological responses of RAW264.7 (black labels) and BEAS-2B (red labels) cells exposed to six metal and metal oxide nanoparticles at different doses in the range of 0.375−200 ppm and after different exposure times (1−24 h). (Left) Clusters of signaling pathway activity and cytotoxicity effects (white background). (Right) SOM projection of activity profiles at different concentrations for each nanoparticle. Blue color indicates down-regulation/low activity, whereas red color indicates up-regulation/high activity.

between the activity of p53, Myc, E2F, and SMAD signaling pathways with cell membrane damage (PI) and oxidative stress responses (ROS) after long time exposure (12−24 h) of macrophage cells to nanomaterials. The projection of HTS data onto the SOM (component planes) identifies ZnO, Pt, and SiO_2 at high concentrations as the nanoparticles responsible for triggering the above cellular activity. Further analysis of component planes shows that ZnO produces a significant down-regulation of multiple signaling pathways in both RAW264.7 and BEAS-2B cell lines (cluster III). The results presented in Figure 2 highlight the use of SOM as a computational and visualization tool for knowledge extraction from high-dimensional HTS data sets and as a tool for *in silico* hazard screening and identification.

2.3. *In Silico* Nanotoxicity Models and Nanotoxicity Metrics. The development of structure−activity relationships for nanomaterials (nano-SARs) is regarded as being essential for the implementation of ENM hazard ranking and regulatory decision-making.[26] Nano-SAR development follows a data-driven approach that requires sufficiently large data repositories of reasonable diversity (e.g., with respect to the heterogeneity of nanoparticles and biological receptors) and suitable nanoparticle descriptors (i.e., physicochemical and structural properties of nanoparticles, environmental conditions such as particle concentrations and solution properties). Ideally, the development of nano-SARs should be integrated with the experimental design of ENMs toxicity studies to ensure that generated data span the desired application domain as discussed in sections 2.1 and 2.2. Also, the generation of detailed structure and chemical descriptors of nanoparticles may demand computational modeling (e.g., quantum mechanics, molecular dynamics, and Monte Carlo methods) at a molecular level.[27] Such an endeavor could require extensive computational effort and resources depending on the complexity of the ENM structure. However, the preferred practical approach[28] is to utilize a small set of fundamental nanoparticle descriptors, especially given the yet-limited nanoparticle characterization and toxicity databases[5a,28b] relative to the chemical world. Moreover, the development of nano-SARs must ensure high data quality when characterizing biological end points, and

TABLE 1. Summary of Recently Published Nano-SARs

ref	materials	size (nm)	receptor	input variable	end point	model	performance	data
ref 31	TiO_2, ZnO	30−125, 50−1500	co-cultures of immortalized rat L2 lung epithelial cells and rat lung alveolar macrophages.	primary size, aggregate sizes (in different mediums), concentration, zeta potential	LDH[a]	linear regression LDA[b] classifier	$r^2 = 0.77$ 100% resubstitution accuracy	N/A
ref 32	44 various ENMs	20−74	four cell lines[c]	primary size, magnetic properties, zeta potential	four assays[d]	SVM[f] classifier	73% 5-fold cross-validation	N/A
ref 33	17 metal oxides	15−90	E. coli cells	ΔH_{Me+}[e]	log(1/EC_{50})	linear regression	0.77 cross-validated regression coefficient	N/A
ref 28b	9 metal oxides	8−19	transformed bronchial epithelial cells	primary size, particle volume fraction, period of particle metal, and atomization energy of metal oxide(d)	PI uptake[g]	logistic regression (classifier)	100% classification accuracy (leave-one-out validation and external validation)	available

[a]LDH: lactate dehydrogenase release. [b]LDA: linear discriminate analysis. [c]Aorta endothelial, vascular smooth muscle, hepatocyte, and monocyte/macrophage. [d]Apoptosis, mitochondrial potential, reducing equivalents, and ATP content. [e]The enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure. [f]SVM: support vector machine. [g]PI: propidium iodide uptake.

concepts such as reliability, variability, and uncertainty management must be considered.[29] Overall, nano-SAR development should follow acceptable guidelines such as those previously established for chemical structure−activity relationships.[28a]

To date, the few nano-SARs have been published (Table 1), based on relatively small data sets have focused on primarily metal and metal oxide nanoparticles. The majority of models included information related to the size (e.g., primary and aggregate sizes) and fundamental nanoparticle properties (e.g., zeta potential, magnetic properties) and fundamental molecular descriptors. This is consistent with recent work[30] that recognized the importance of including particle size information in nano-SAR models to distinguish between nanoscale effects and bulk properties.

The definition of toxicological end points plays a key role in nano-SARs development. Given the uncertainties and variability in high-throughput assays, it is difficult and often impractical to develop quantitative toxicity end points from HTS data that are both meaningful (for regulatory use) and accurate. Therefore, toxicological end points (or metrics) that can be correlated with ENM descriptors using classifier based nano-SARs are alternative screening tools for identification of active nanoparticles (with respect to toxicity) for subsequent detailed toxicity and/or in support of regulatory actions. Following this approach, a nano-SAR classifier[28b] for cytotoxicity of metal oxide nanoparticles was recently developed for transformed bronchial epithelial cells (BEAS-2B), making use of a logistic regression model. This simple nano-SAR demonstrated that the metal-oxide energy of atomization, period of the metal, and the nanoparticle primary size and volume fraction were suitable parameters for identifying the propensity of the nanoparticles to damage the cell membrane.

When data uncertainty levels are significant, it is important to control the nano-SAR misclassification of ENMs as being either toxic when they are not (false positive) or as nontoxic when they are actually toxic (false negative). In the regulatory context, false negatives should be avoided to ensure that correct decisions are reached with respect to environmental and human health protection. In order to address this issue, penalty (or cost) functions can be introduced (in the nano-SAR development stage) to weight nano-SAR predictions according to predefined acceptance level of false negatives relative to false positives. This approach is demonstrated in Figure 3 using available literature data[33] for EC_{50} toxicity to *E. coli* for a set of 16 metal oxide nanoparticles with primary sizes in the range of 15−90 nm.

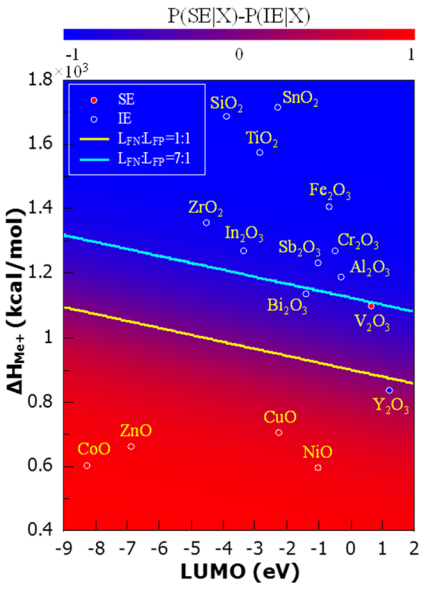


FIGURE 3. Probability map of nanoparticle (*X*) having significant effect (*P*(SE|*X*)) versus having insignificant effect (*P*(IE|*X*)) (i.e., *P*(SE|*X*) versus P(IE|*X*)). The log ratio of the two probabilities is correlated with two ENM descriptors via a simple nano-SAR: $\ln(P(SE|X)/P(IE|X)) = -0.1882\text{LUMO} - 0.0087\Delta H_{Me+} + 7.834$. The curves illustrate two decision boundaries corresponding to different acceptance levels of false negative relative to false positive (expressed via the ratio L_{FN}:L_{FP}).

The analysis proceeds by first dividing the nanoparticles into two categories (significant effect (SE) and insignificant effect (IE)) according to a predefined EC_{50} threshold (e.g., $\log(1/EC_{50}) \geq 3$ identifies ENM having a significant effect on *E. coli*). A nano-SAR, based on a logistic regression model, was then developed for classifying the metal oxides nanoparticles as having either significant effect (SE) or insignificant effect (IE) using two quantum chemical energy descriptors (energy of the lowest unoccupied molecular orbital (LUMO) and enthalpy of formation of a gaseous cation (ΔH_{Me+})).[33] The results as shown in Figure 3 depict a decision boundary (in yellow) that separates the two categories (SE and IE) (with equal acceptance level of false negative and false positive). In this example, the nano-SAR model yields a false positive classification of Y_2O_3 and misclassifies V_2O_3 as having insignificant effect. Upon introducing a penalty function or an acceptance level of 7:1 (i.e., false negatives have a penalty (L_{FN}) of classification 7 times greater than for false positives (L_{FP})) produces a more conservative nano-SAR with zero false negatives at the expenses of adding Bi_2O_3 as another false positive. The above example illustrates that the use of nano-SARs for regulatory applications could benefit from considerations of the level of acceptance of false positives together with the reliability of toxicity data.

3. Environmental Impact Assessment

Environmental impact assessment requires identification and acceptance criteria of potential risk or hazard ranking. The analysis should include the various factors discussed in sections 1 and 2, such as production volume, emission rates and modes of release, likely concentrations in the various environmental media, exposure pathways, toxicity data (e.g., dose−response), ENMs' physicochemical properties, possible environmental transformations, as well as the multimedia distribution of ENMs that govern exposures (Figure 4).

Prospects for quantitative risk assessment of nanomaterials are plagued by a general lack of environmental concentrations, exposure, and toxicity data.[34] At the same time, the number of commercially produced ENMs could increase from the current 10^3 different nanomaterials to an order of 10^5 within a decade.[34] Clearly, applying conventional risk assessment techniques used for chemicals to nanomaterials would be a formidable task given the lack of environmental monitoring data. Therefore, risk assessment for nanomaterials requires carefully crafted strategies that optimally use the available information to guide the decision-making process.[35] For example, cause and effect relationships involving multiple interdependent ranking criteria can be modeled using Bayesian networks.[36] The use of a Bayesian network is particularly useful since this network encodes, as a joint probability distribution, the domain knowledge (either given explicitly by an expert or extracted from data) of interdependency relationships between variables. In the decision-making problem one strives to recommend the alternative that maximizes the expected objective given the observation of a set of external factors and preferences of the decision maker.

Irrespective of the complexity of establishing toxicity metrics and exposure assessment, one is likely to be confronted with significant fuzzy information (i.e., qualitative or quantitative but of various levels of uncertainty). Therefore, uncertainties should be considered in various paths of the analysis process. The premise of such an approach is that a given ENMs would be of environmental concern if it is hazardous and there is exposure to the receptors of concern at concentration levels that may induce an adverse effect. Accordingly, in order to determine if a given ENM should be of environmental concern, one can proceed with an initial screening to first evaluate exposure likelihood and subsequently (or in parallel) the potential hazard associated with the ENM, followed by detailed environmental impact analysis as may be suggested by the initial screening.

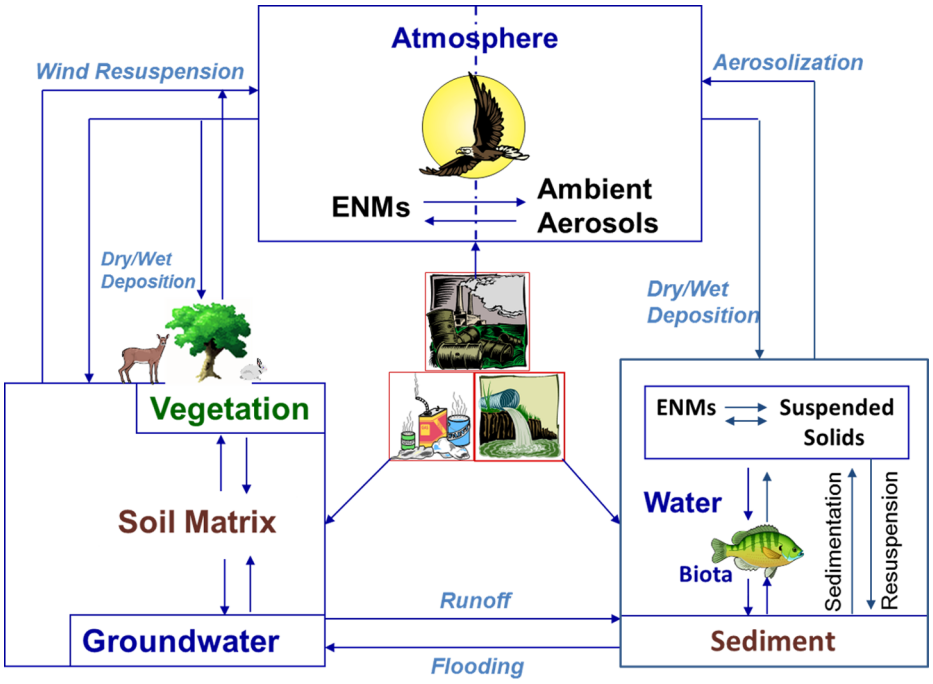


FIGURE 4. Major intermedia transport processes in a multimedia environment.

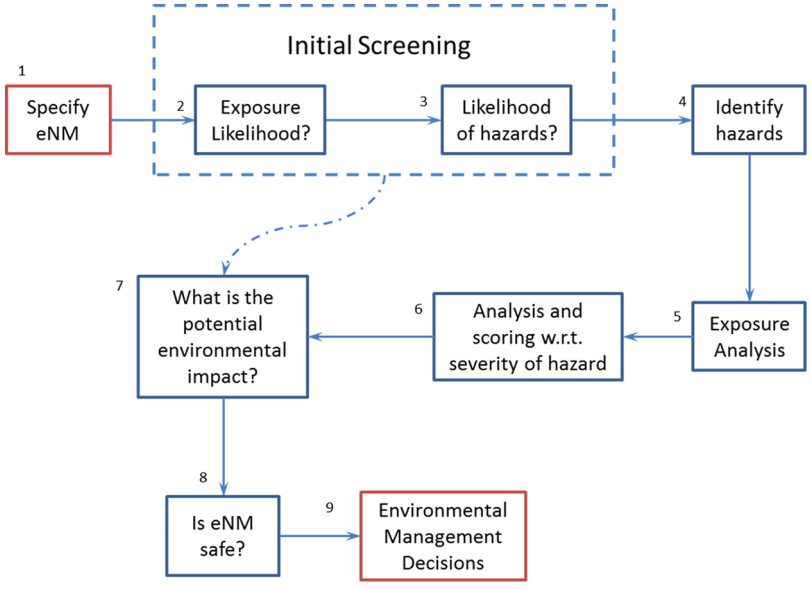


FIGURE 5. Example of sequential steps in assessing the environmental impact of nanomaterials.

An example of a possible decision analysis process is depicted in Figure 5. The question to be asked is whether or not there is likelihood for exposure (step 2) to the specified ENM and if it is hazardous (step 3). Screening assessment should consider the environmentally relevant exposure period and potential environmental ENM release during production, use or disposal. If the ENM is also deemed to be potentially hazardous at the screening stage (e.g., based on available information), then decision makers could opt to assert their authority to make early judgment regarding the potential safety of such a material (proceed from step 3 to step 8 in Figure 5) or request analysis that follows steps 4–8. Hazard identification (step 4) may include experimental toxicity studies (*in vivo*/*in vitro*) and/or toxicity modeling (*in silico*). Subsequently, step 5 serves to determine if there is indeed ENM exposure for the expected exposure scenarios. Subsequently, the severity of the identified hazards can be ranked (step 6), based on the collection of quantitative and qualitative information, while considering the uncertainties in the various analysis steps (step 7). The outcome of the analysis

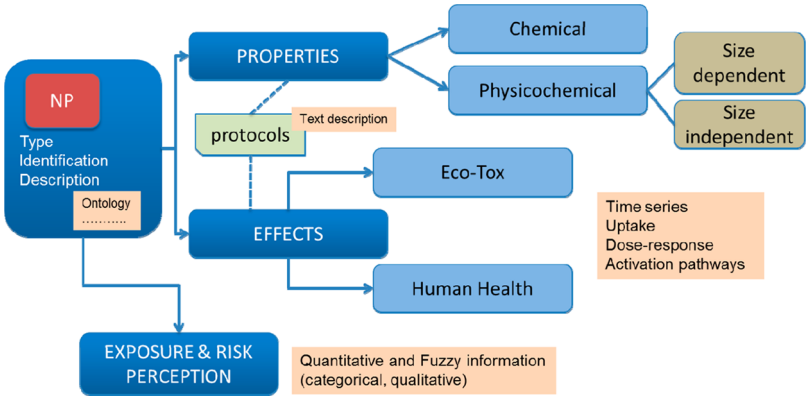


FIGURE 6. Data required for nanomaterial hazard and risk assessment.

(whether in the form of relative ranking of exposures among analyzed ENMs, risk of adverse outcomes, or level of exposures above an acceptable threshold) can then be used by regulators to address the question "Is this ENM safe?" (step 8). It is acknowledged that decisions regarding environmental risk management (step 9) may need to be made even when faced with partial information. Therefore, the final risk management decision (step 9) should consider uncertainties of exposure and risk (steps 6 and 7). The outcome may include ranking following established criteria based, for example, on methods such as multicriteria decision analysis (MCDA[37]). MCDA methods utilize a decision matrix of criteria and performance scores to provide a systematic analytical approach, which enables evaluation and ranking of alternatives. To generate rankings, each criterion requires weights that in many practical applications may be difficult to assign. In this regard, the application of MCDA would allow for uncertainty quantification,[38] enabling statements about the likelihood of miss-rankings.

4. Nanoinformatics

Given the need for integrated information sources and *In Silico* tools (e.g., nano-SARs and ENM transport and fate models) for the analysis of nanomaterials hazard and risk, the nanoinformatics field has emerged over the past few years as "*The science and practice of determining which information is relevant to the nanoscale science and engineering community, and then developing and implementing effective mechanisms for collecting, validating, storing, sharing, analyzing, modeling, and applying that information.*"[39] A key challenge in nanoinformatics is in establishing interoperability of data repositories containing heterogeneous data sets (Figure 6), common vocabulary (i.e., ontology) to unambiguously describe nanoparticles,[40] standard formats for data exchange (e.g., ISA-TAB Nano specification;[41] NCBO;[42] NBI[43]), definition of the minimum set nanomaterial characterization parameters (e.g., MINChar[44]), and ENM databases and related information portals (e.g., EU nanoHUB;[45] NCI caNanoLab;[46] CEIN NDR[47]).

5. Concluding Remarks

Effective decision-making to address the potential impact of nanomaterials will require an integrated analysis platform that considers the relevant environmental, ecological, technological, economic, and sociopolitical factors affecting the complete lifecycle of nanomaterials. Generating the necessary scientific data and knowledge pertaining to the toxicity of nanomaterials and their transport and fate behavior in the environment is paramount to conducting environmental impact assessments. However, it must be recognized that heterogeneous information with respect to human aspirations and technical applications of ENMS demand a systematic and coherent framework to organize people, data, processes, and tools for making structured and defensible decisions.

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 authors 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. Support was also provided by CICYT (Project CTQ2009-14627), Generalitat de Catalunya (2009SGR-01529) and the EU Commission (OSIRIS, Contract No. 037017).

BIOGRAPHICAL INFORMATION

Yoram Cohen is a Professor of Chemical and Biomolecular Engineering at UCLA. He received his B.A.Sc (1975) and M.A.Sc (1977) from the University of Toronto and Ph.D. (1981) from

the University of Delaware. He is also a Faculty member of the the UCLA Institute of the Environment and Sustainability and an Adjunct Professor at Ben-Gurion University. He was a Visiting Professor at the Technion (1987−1988), Universitat Rovira i Virgili (1994), and a Distinguished Visiting Professor at Victoria University (2006).

Robert Rallo is an associate professor in the Departament d'Enginyeria Informatica i Matematiques at Universitat Rovira i Virgili (URV) and was a visiting scholar at UCLA (2007, 2009−2011). He received his B.S (1985) and M.A (1987) from Universitat de Barcelona and his Ph.D (2007) from URV.

Rong Liu is a postdoctoral research associate in the Department of Chemical and Biomolecular Engineering at UCLA. He received his B.S. (1999) from Northwestern Polytechnical University (China) and M.S. (2005) and Ph.D. (2008) from Graduate University of Chinese Academy of Sciences.

Haoyang Haven Liu is a Ph.D. candidate at the Department of Chemical and Biomolecular Engineering at UCLA. He received his B.S. (2007) from Washington University in Saint Louis.

FOOTNOTES

*To whom correspondence should be addressed. E-mail: yoram@ucla.edu. Telephone: (310) 825-8766.

The authors declare no competing financial interest.

REFERENCES

1 Boxall, A. B.; Tiede, K.; Chaudhry, Q. Engineered nanomaterials in soils and water: how do they behave and could they pose a risk to human health? *Nanomedicine* **2007**, *2*, 919–927.

2 Cowan, C. E. *The multi-media fate model: a vital tool for predicting the fate of chemicals*; SETAC Press: Pensacola, FL, 1995.

3 (a) Mueller, N. C.; Nowack, B. Exposure modeling of engineered nanoparticles in the environment. *Environ. Sci. Technol.* **2008**, *42*, 4447–4453. (b) Linkov, I.; Seager, T. P. Coupling multi-criteria decision analysis, life-cycle assessment, and risk assessment for emerging threats. *Environ. Sci. Technol.* **2011**, *45*, 5068–5074.

4 Paustenbach, D. J. *Human and ecological risk assessment: theory and practice*; J. Wiley: New York, 2002.

5 (a) Tervonen, T.; Linkov, I.; Figueira, J.; Steevens, J.; Chappell, M.; Merad, M. Risk-based classification system of nanomaterials. *J. Nanopart. Res.* **2009**, *11*, 757–766. (b) Linkov, I.; Satterstrom, F. K. Nanomaterial Risk Assessment and Risk Management. In *Real-Time and Deliberative Decision Making*; Linkov, I.; Ferguson, E.; Magar, V. S. Eds; Springer: Netherlands, 2009; pp 129−157. (c) Grieger, K. D.; Linkov, I.; Hansen, S. F.; Baun, A. Environmental risk analysis for nanomaterials: review and evaluation of frameworks. *Nanotoxicology* **2012**, *6*, 196–212.

6 (a) Maynard, A. D. Commentary: Oversight of Engineered Nanomaterials in the Workplace. *J. Law, Med. Ethics* **2009**, *37*, 651–658. (b) Klaine, S. J.; Koelmans, A. A.; Horne, N.; Carley, S.; Handy, R. D.; Kapustka, L.; Nowack, B.; von der Kammer, F. Paradigms to assess the environmental impact of manufactured nanomaterials. *Environ. Toxicol. Chem.* **2012**, *31*, 3–14.

7 Cronin, M.; Madden, J. *In Silico Toxicology: Principles and Applications*; Royal Society of Chemistry: Cambridge, UK, 2010.

8 Cohen, Y.. National Center for Intermedia Transport Research (U.S.); University of California Los Angeles. Engineering and Systems Analysis for the Control of Toxics Program. *Pollutants in a multimedia environment*; Plenum Press: New York, 1986.

9 Zbilut, J. P.; Giuliani, A. Biological uncertainty. *Theory Biosci.* **2008**, *127*, 223–227.

10 Lin, S. J.; Zhao, Y.; Xia, T.; Meng, H.; Ji, Z. X.; Liu, R.; George, S.; Xiong, S. J.; Wang, X.; Zhang, H. Y.; Pokhrel, S.; Madler, L.; Damoiseaux, R.; Lin, S.; Nel, A. E. High Content Screening in Zebrafish Speeds up Hazard Ranking of Transition Metal Oxide Nanoparticles. *ACS Nano* **2011**, *5*, 7284–7295.

11 Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discovery* **2011**, *10*, 188–195.

12 George, S.; Xia, T. A.; Rallo, R.; Zhao, Y.; Ji, Z. X.; Lin, S. J.; Wang, X.; Zhang, H. Y.; France, B.; Schoenfeld, D.; Damoiseaux, R.; Liu, R.; Lin, S.; Bradley, K. A.; Cohen, Y.; Nal, A. E. Use of a High-Throughput Screening Approach Coupled with *In Vivo* Zebrafish Embryo Screening To Develop Hazard Ranking for Engineered Nanomaterials. *ACS Nano* **2011**, *5*, 1805–1817.

13 Malo, N.; Hanley, J. A.; Cerquozzi, S.; Pelletier, J.; Nadon, R. Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.* **2006**, *24*, 167–175.

14 Wawer, M.; Bajorath, J. Extraction of Structure-Activity Relationship Information from High-Throughput Screening Data. *Curr. Med. Chem.* **2009**, *16*, 4049–4057.

15 Zhang, X. D. *Optimal High-Throughput Screening: Practical Experimental Design and Data Analysis for Genome-Scale RNAi Research*; Cambridge University Press: Cambridge, UK, 2011.

16 (a) Shieh, A. D.; Hung, Y. S. Detecting Outlier Samples in Microarray Data. *Stat. Appl. Genet. Mol. Biol.* **2009**, *8*, Article 13. DOI: 10.2202/1544-6115.1426. (b) Botella, C.; Ferre, J.; Boque, R. Outlier detection and ambiguity detection for microarray data in probabilistic discriminant partial least squares regression. *J. Chemom.* **2010**, *24*, 434–443.

17 Birmingham, A.; Selfors, L. M.; Forster, T.; Wrobel, D.; Kennedy, C. J.; Shanks, E.; Santoyo-Lopez, J.; Dunican, D. J.; Long, A.; Kelleher, D.; Smith, Q.; Beijersbergen, R. L.; Ghazal, P.; Shamu, C. E. Statistical methods for analysis of high-throughput RNA interference screens. *Nat. Methods* **2009**, *6*, 569–575.

18 (a) Sukuru, S. C. K.; Jenkins, J. L.; Beckwith, R. E. J.; Scheiber, J.; Bender, A.; Mikhailov, D.; Davies, J. W.; Glick, M. Plate-Based Diversity Selection Based on Empirical HTS Data to Enhance the Number of Hits and Their Chemical Diversity. *J. Biomol. Screening* **2009**, *14*, 690–699. (b) Posner, B. A.; Xi, H. L.; Mills, J. E. J. Enhanced HTS Hit Selection via a Local Hit Rate Analysis. *J. Chem. Inf. Model.* **2009**, *49*, 2202–2210.

19 Zhang, X. D. A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *J Biomol Screen* **2007**, *12*, 645–655.

20 Liu, R.; Rallo, R.; Cohen, Y. Unsupervised Feature Selection Using Incremental Least Squares. *Int. J. Inf. Tech. Decis.* **2011**, *10*, 967–987.

21 Masulli, F.; Mitra, S. Natural computing methods in bioinformatics: A survey. *Inform. Fusion* **2009**, *10*, 211–216.

22 Shaw, S. Y.; Westly, E. C.; Pittet, M. J.; Subramanian, A.; Schreiber, S. L.; Weissleder, R. Perturbational profiling of nanomaterial biologic activity. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7387–7392.

23 Kohonen, T. The Self-Organizing Map. *Proc. IEEE* **1990**, *78*, 1464–1480.

24 Monti, S.; Tamayo, P.; Mesirov, J.; Golub, T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **2003**, *52*, 91–118.

25 Rallo, R.; France, B.; Liu, R.; Nair, S.; George, S.; Damoiseaux, R.; Giralt, F.; Nel, A.; Bradley, K.; Cohen, Y. Self-Organizing Map Analysis of Toxicity-Related Cell Signaling Pathways for Metal and Metal Oxide Nanoparticles. *Environ. Sci. Technol.* **2011**, *45*, 1695–1702.

26 (a) Walker, N. J.; Bucher, J. R. A 21st Century Paradigm for Evaluating the Health Hazards of Nanoscale Materials? *Toxicol. Sci.* **2009**, *110*, 251–254. (b) Lai, D. Y. Toward toxicity testing of nanomaterials in the 21st century: a paradigm for moving forward. *Wiley Interdiscip. Wiley Interdiscip. Rev.-Nanomed. Nanobiotechnol.* **2012**, *4*, 1–15.

27 (a) Poater, A.; Saliner, A. G.; Carbo-Dorca, R.; Poater, J.; Sola, M.; Cavallo, L.; Worth, A. P. Modeling the Structure-Property Relationships of Nanoneedles: A Journey Toward Nanomedicine. *J. Comput. Chem.* **2009**, *30*, 275–284. (b) Poater, A.; Gallegos Saliner, A.; Sola, M.; Cavallo, L.; Worth, A. P. Computational methods to predict the reactivity of nanoparticles through structure-property relationships. *Expert Opin. Drug Delivery* **2010**, *7*, 295–305. (c) Barnard, A. S. Computational strategies for predicting the potential risks associated with nanotechnology. *Nanoscale* **2009**, *1*, 89–95.

28 (a) Puzyn, T.; Leszczynska, D.; Leszczynski, J. Toward the development of "nano-QSARs": advances and challenges. *Small* **2009**, *5*, 2494–509. (b) Liu, R.; Rallo, R.; George, S.; Ji, Z.; Nair, S.; Nel, A. E.; Cohen, Y. Classification NanoSAR Development for Cytotoxicity of Metal Oxide Nanoparticles. *Small* **2011**, *7*, 1118–1126.

29 Fernandez, A.; Rallo, R.; Giralt, F. Uncertainty Reduction in Environmental Data with Conflicting Information. *Environ. Sci. Technol.* **2009**, *43*, 5001–5006.

30 Carbo-Dorca, R.; Besalu, E. Construction of coherent nano quantitative structure-properties relationships (nano-QSPR) models and catastrophe theory. *SAR QSAR Environ. Res.* **2011**, *22*, 661–665.

31 Sayes, C.; Ivanov, I. Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Anal.* **2010**, *30*, 1723–1734.

32 Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S. Y.; Mumper, R. J.; Tropsha, A. Quantitative Nanostructure-Activity Relationship Modeling. *ACS Nano* **2010**, *4*, 5703–5712.

33 Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat. Nanotechnol.* **2011**, *6*, 175–178.

34 Morgan, K. Development of a preliminary framework for informing the risk analysis and risk management of nanoparticles. *Risk Anal.* **2005**, *25*, 1621–1635.

35 Linkov, I.; Satterstrom, F. K.; Steevens, J.; Ferguson, E.; Pleus, R. C. Multi-criteria decision analysis and environmental risk assessment for nanomaterials. *J. Nanopart. Res.* **2007**, *9*, 543–554.

36 Pearl, J. *Causality: Models, Reasoning, and Inference*; Cambridge University Press: Cambridge, UK, 2000.

37 Linkov, I.; Steevens, J.; Adlakha-Hutcheon, G.; Bennett, E.; Chappell, M.; Colvin, V.; Davis, J. M.; Davis, T.; Elder, A.; Hansen, S.; Hakkinen, P. B.; Hussain, S. M.; Karkan, D.; Korenstein, R.; Lynch, I.; Metcalfe, C.; Ramadan, A. B.; Satterstrom, F. K. Emerging methods and tools for environmental risk assessment, decision-making, and policy for nanomaterials: summary of NATO Advanced Research Workshop. *J. Nanopart. Res.* **2009**, *11*, 513–527.

38 Seager, T. P.; Linkov, I. Coupling multicriteria decision analysis and life cycle assessment for nanomaterials. *J. Ind. Ecol.* **2008**, *12*, 282–285.

39 Nanoinformatics 2020 Roadmap; http://eprints.internano.org/607/.

40 Thomas, D. G.; Pappu, R. V.; Baker, N. A. NanoParticle Ontology for cancer nanotechnology research. *J. Biomed. Inform.* **2011**, *44*, 59–74.

41 Thomas, D. G.; Klaessig, F.; Harper, S. L.; Fritts, M.; Hoover, M. D.; Gaheen, S.; Stokes, T. H.; Reznik-Zellen, R.; Freund, E. T.; Klemm, J. D.; Paik, D. S.; Baker, N. A. Informatics and standards for nanomedicine technology. *Wiley Interdiscip. Rev.-Nanomed. Nanobiotechnol.* **2011**, *3*, 511–532.

42 NCBO BioPortal; bioportal.bioontology.org.

43 NBI Knowledgebase; nbi.oregonstate.edu.

44 Auffan, M.; Rose, J.; Bottero, J. Y.; Lowry, G. V.; Jolivet, J. P.; Wiesner, M. R. Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective. *Nat. Nanotechnol.* **2009**, *4*, 634–641.

45 EU nanoHUB; www.napira.eu.

46 NCI caNanoLab; cananolab.nci.nih.gov.

47 CEIN Nanomaterials Data Repository; nanoinfo.cein.ucla.edu.