Quantitative Nanostructure—Activity Relationship Modeling

Denis Fourches,[†] Dongqiuye Pu,[†] Carlos Tassa,[‡] Ralph Weissleder,[‡] Stanley Y. Shaw,[‡] Russell J. Mumper,[§] and Alexander Tropsha[†,*]

[†]Laboratory of Molecular Modeling, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, [‡]Center for Systems Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, and [§]Center for Nanotechnology in Drug Delivery, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599

ABSTRACT Evaluation of biological effects, both desired and undesired, caused by manufactured nanoparticles (MNPs) is of critical importance for nanotechnology. Experimental studies, especially toxicological, are time-consuming, costly, and often impractical, calling for the development of efficient computational approaches capable of predicting biological effects of MNPs. To this end, we have investigated the potential of cheminformatics methods such as quantitative structure—activity relationship (QSAR) modeling to establish statistically significant relationships between measured biological activity profiles of MNPs and their physical, chemical, and geometrical properties, either measured experimentally or computed from the structure of MNPs. To reflect the context of the study, we termed our approach quantitative nanostructure—activity relationship (QNAR) modeling. We have employed two representative sets of MNPs studied recently using *in vitro* cell-based assays: (i) 51 various MNPs with diverse metal cores (*Proc. Natl. Acad. Sci.* 2008, *105*, 7387—7392) and (ii) 109 MNPs with similar core but diverse surface modifiers (*Nat. Biotechnol.* 2005, *23*, 1418—1423). We have generated QNAR models using machine learning approaches such as support vector machine (SVM)-based classification and *k* nearest neighbors (*k*NN)-based regression; their external prediction power was shown to be as high as 73% for classification modeling and having an R^2 of 0.72 for regression modeling. Our results suggest that QNAR models can be employed for: (i) predicting biological activity profiles of novel nanomaterials, and (ii) prioritizing the design and manufacturing of nanomaterials toward better and safer products.

KEYWORDS: nanoparticles · QSAR · cheminformatics · nanotoxicity · modeling

More than 1000 manufacturer-identified, nanotechnology-based consumer products are currently available on the market (The Woodrow Wilson International Center for Scholars, 2010). A growing fraction of them represent green products intended to achieve efficient and less polluting energy sources.[1] However, some manufactured nanoparticles (MNPs) intended for industrial applications may cause toxic effects in humans,[2−4] and public concern about the safety of MNPs is increasing.[5] Induced biological effects could result from exposure and subsequent absorption of ultrafine MNPs *via* different routes[6] and lead to their potentially detrimental delivery to critical organs.[7] Once MNPs gain entry into the systemic circulation, they can immediately interact with blood cells and can then be either distributed throughout the body or quickly captured by macrophages of the reticuloendothelial system. Thus, understanding the biological effects of exposure to MNPs is of paramount importance.

Experimental nanotoxicology is a very young field.[8−14] There remain significant scientific gaps in our understanding of the toxicology of nanobased materials that, (i) are already contained in commercial products not intended for human exposure, (ii) could contaminate the environment while also not intended for human exposure, and (iii) are intended for biomedical applications such as drug delivery, imaging, and sensing. Thus, it is imperative to develop a comprehensive, and ideally, predictive knowledge of the effects of MNPs on the environment as well as animals and humans. Recently, Mumper and colleagues published a comprehensive study on the hemocompatibility of lipid NPs for drug delivery.[15] There are several reports on the deleterious effects of MNPs on humans and wildlife. For example, Radomski *et al.*[16] reported that both multiwalled and single-walled carbon nanotubes caused platelet aggregation and vascular thrombosis acceleration. Harhaji *et al.*[17] showed that even at the "high dose" of 1 μg/mL, the C60 fullerenes caused reactive-oxygen, species-mediated, necrotic cell damage[18] and proposed C60 fullerenes as an anticancer agent. Kane *et al.*[19] found that silica MNPs directly interacted with plasma and lysosomal membranes leading to Ca^{2+} influx, ATP depletion, and cell death. Kang *et al.*[20] observed that nano-TiO_2 caused ROS stress and DNA damage in lymphocytes. Leonard *et al.*[21] showed that $PbCrO_4$ particles resulted in ROS generation and up-regulation of NF-kappaB and AP-1 in RAW 264.7 cells. Pulskamp *et al.*[22] reported that several

*Address correspondence to alex_tropsha@unc.edu.

Received for review June 15, 2010 and accepted September 08, 2010.

Published online September 21, 2010.
10.1021/nn1013484

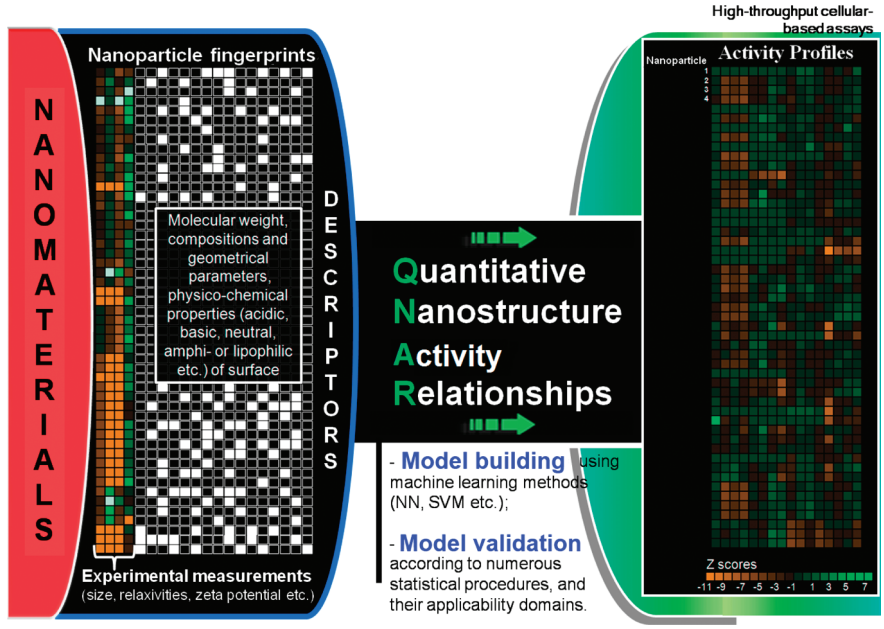

Figure 1. Study design for quantitative nanostructure−activity relationship (QNAR) modeling using both calculated as well as experimentally measured properties of manufactured nanoparticles as descriptors.

carbon MNPs (multiwalled, single-walled, carbon black, quartz) increased ROS and decreased mitochondrial membrane potential in a dose- and time-dependent manner in rat macrophages and human A549 lung cells. An important review on the subject of nanotoxicity was recently published[7] that describes examples of known toxic effects of MNPs.

Modeling MNPs and their biological effects is challenging. First, because of the high structural complexity and diversity of MNPs, it is difficult to develop quantitative parameters capable of characterizing the structural and chemical properties of MNPs. Second, systematic physicochemical, geometrical, structural, and biological studies of MNPs are nearly absent in the public domain, making the development of statistically significant computational models and their validation difficult as these procedures require relatively large amounts of data. For instance, most papers cited in the previous paragraph reported experimental studies on one or a few MNPs. Not surprisingly, the reports on computational modeling of MNPs, especially in the area of nanotoxicology have been scarce.[23] Liu *et al.*[24] demonstrated the utility of molecular dynamics simulations for (i) revealing the overall changes in the structure of cellular membranes caused by the insertion of carbon nanotubes as well as (ii) estimating the affinity of drug-like molecules for carbon nanotubes in an aqueous environment.[25] In another recent study by Shaw *et al.*,[26] as many as 51 MNPs were thoroughly tested *in vitro* against four cell lines in different assays to study their induced biological effects. Different statistical techniques were applied to find the correlations between the biological activity profiles of MNPs and to discover hidden structure−property relationships. Recently, Puzyn *et al.*[27] suggested that quantitative structure−activity relationship (QSAR) modeling can be employed in computational nanotoxicology studies. The authors appropriately concluded that no universal "nano-QSAR" model can accurately assess the toxicity of all possible MNPs. They also reported several QSAR models largely developed for carbon nanotubes and fullerenes to assess their solubility and lipophilicity. However, these models were built using very small data sets, usually less than 20 MNPs, and insufficient validation procedures according to common QSAR modeling practices (such as OECD principles[28]).

The main objective of this study is to develop predictive quantitative nanostructure−activity relationship (QNAR) models following the established principles of conventional QSAR modeling workflows.[29] Similar to general QSAR modeling strategies, the overall objective of QNAR models is to relate a set of descriptors characterizing MNPs with their measured biological effects, for example, cell viability, or cellular uptake (Figure 1). Such models can then be applied to newly designed or commercially available MNPs in order to quickly and efficiently assess their potential biological effects. As a proof-of-concept, we describe case studies for two relatively large series of MNPs that have been tested for their effects in different *in vitro* cellular-based assays. The first series[26] comprises 51 diverse MNPs with different metal cores and surface modifications (Case Study 1) that were tested in different cell-based assays, whereas the second series[30] includes 109 NPs with the same core but different surface modifiers (Case Study 2) that were tested for their cell uptake activity. We have applied conventional cheminformatics techniques such as (i) cluster analysis to examine if MNPs with similar biological activities are also structurally similar, and (ii) QNAR modeling to establish quanti-

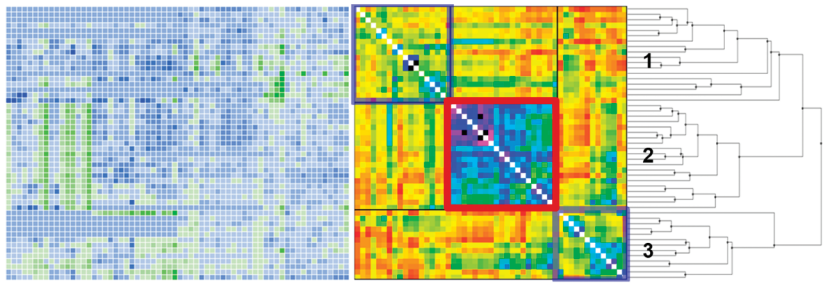

Figure 2. Hierarchical clustering analysis of 51 MNPs using their biological activity profiles. The clustered distance matrix reveals three distinct clusters of MNPs based on their biological activity profiles (on the distance matrix, blue colors = high similarity between nanoparticles, red/green/yellow colors = low/medium similarity between nanoparticles).

tative links between available MNP descriptors (that characterize their structure) and their biological activity. In Case Study 1, the structure of MNPs was characterized by their experimental properties treated as molecular descriptors. Conversely, Case Study 2 could be regarded as a conventional QSAR investigation since 109 MNPs with the same metal core (analogous to common chemical scaffold for organic molecules) were characterized by conventional chemical descriptors of surface-modifying organic molecules. In both case studies, QNAR calculations led to statistically validated and externally predictive models; these models quantitatively relate the chemical, physical, and geometrical properties of MNPs with their biological effects measured *in vitro* in different cell-based assays. We believe that this report, which to the best of our knowledge is the first example of QNAR analysis of relatively large data sets of MNPs, successfully demonstrates the high potential of cheminformatics approaches for improving the experimental design and prioritizing the biological testing of novel MNPs.

RESULTS

Case Study 1—Modeling of Cellular Effects Induced by Diverse MNPs. Using the Heat Map Viewer (HMV) software developed internally, we first visualized the biological activity profile of the entire data set comprising 51 MNPs tested *in vitro* using four doses, four different cell lines, and four different assays of cellular physiology, forming an activity matrix of 64 biological parameters for each MNP (using data reported by Shaw *et al.*[26]); this array represents a 64-feature biological activity profile for each MNP (Supporting Information, Figure S1), which we subsequently normalized to the unit (variation between 0 and 1) and clustered using ISIDA/Cluster[31] (hierarchical algorithm, Euclidean distance between compounds, complete linkage between clusters). It should be pointed out that no obvious clusters of MNPs appeared on this initial heat map.

We identified three clusters (Figure 2) using this hierarchical clustering procedure. Five types of MNPs were represented in the data set; each MNP had a metal core, that is, Fe_2O_3-predominant, Fe_3O_4-predominant, Cd−Se, or Fe^{III}; an organic coat, either acidic, basic, amphiphilic, or lipophilic; and various surface modifiers for some of the MNPs. First, we confirmed some of the results of Shaw *et al.*[26] that all MION nanoparticles were found in cluster 2, whereas all three quantum dot-based MNPs were found in cluster 1. Importantly, further analysis (see Table 1) revealed that all MNPs included in cluster 2 featured the same metal core (Fe_3O_4-predominant) independently of their surface modifiers and the type of MNPs; for instance, cluster 2 contains 13 CLIO, 2 PNP, and 4 MION each with different coatings, such as cross-linked dextran, arabino-galactan, or carboxymethyl dextran, and different additional organic surface modifiers (see Supporting Information, Table 1 for the complete description of MNPs). All of these MNPs displayed relatively similar biological activity profiles (Figure 2), and data visualization with the HMV program confirmed that MNPs within cluster 2 were very similar in terms of their biological properties, whereas those in clusters 1 and 3 were more chemically diverse. These similarity discrepancies were also observed in the distance matrix, where intracluster pairwise similarities were significantly higher for cluster 2 compared to the other two clusters. These results demonstrate that, at least in some cases, MNPs with similar biological activity can be also recognized as similar by their structural descriptors (*e.g.*, type of metal core). This initial ob-

TABLE 1. Case Study 1. Cluster Membership According to the MNP Types and Their Metal Core Types

	cluster 1	cluster 2	cluster 3	total
MNP Type				
CLIO	7	13	3	23
PNP	7	2	10	19
MION	0	4	0	4
Qt-dot	3	0	0	3
Feridex	0	1	0	1
Ferrum Haussmann	1	0	0	1
	18	20	13	51
MNP Metal Core				
Fe_2O_3	5	0	9	14
Fe_3O_4	9	20	4	33
Cd−Se	3	0	0	3
Fe(III)	1	0	0	1
	18	20	13	51

TABLE 2. Case Study 1—QNAR Modeling of the Biological Effects for 44 MNPs

	modeling set				external set			
fold	no. MNPs	no. models	% accuracy internal 5-fold CV	% accuracy	no. MNPs	% accuracy	% sensitivity	% specificity
1	35	11	51.4−60.0	71.4−82.9	9	78	67	100
2	35	13	51.4−60.0	71.4−77.1	9	78	50	100
3	35	16	57.1−62.9	74.3−82.9	9	78	80	75
4	35	11	60.0−62.9	77.1−88.6	9	56	50	60
5	36	4	66.7	83.3−86.1	8	88	33	100
					44	73	60	86

servation is important to demonstrate the applicability of cheminformatics approaches to the analysis of nanostructure−activity relationships.

To further demonstrate the overall feasibility of QNAR modeling, we used experimentally measured physical parameters (descriptors) of MNPs to build binary classification models (*i.e.*, models capable of assigning MNPs to one of two distinct classes defined by their biological activity). Four such structural descriptors were available for 44 of the 51 MNPs: nanoparticle size, ranging from 20 to 74 nm, *R*1 and *R*2 relaxivities representing their magnetic properties, and zeta potential representing the intensity of charge on their surface. On the other hand, the entire biological activity profile included 64 features, that is, a total number of all possible combinations of four doses, four cell lines, and four assays. To enable a binary classification study, we transformed the 64 features into 1 by calculating their arithmetic mean (Z_{mean}; *cf.* Table S1 in Supporting Information). It should be noted that when Shaw *et al.*[26] expressed the biological activity of MNPs as a 64-feature vector (4 cell lines × 4 assays × 4 doses), the correlation coefficient between vectors associated with the independent replicates for the same nanoparticle was as high 0.93; furthermore, these independent replicates for the same nanoparticle were more similar to each other, than to any other nanoparticle (*cf.* Figure 3 in Shaw *et al.*[26]). We then defined two binary classes using an arbitrary threshold at $Z_{mean} = -0.40$, which allowed us to split the set into two groups each containing the same number of MNPs. As a result, 22 MNPs belonged to class 1 ($Z_{mean} \geq -0.40$), and the remaining 22 were put in class 0 ($Z_{mean} < -0.40$).

To derive QNAR models, we used the WinSVM program developed in-house. WinSVM implements an external 5-fold cross-validation procedure: the program splits the entire data set five times into a modeling set including 80% of the nanoparticle data set, and the external validation set, comprising the remaining 20% of the nanoparticle data set. Only the modeling set (which is divided additionally into multiple training and test sets) was used to build and validate models, and models with appreciable training and test set prediction accuracies were selected for predicting class membership of the external set. Each MNP was included into a validation set only once, allowing us to calculate the overall *external* prediction accuracy for the whole set (see Table 2). The data indicate that SVM models had relatively high external prediction accuracies of 56−88% for the five independent external validation sets, with the mean external accuracy as high as 73%. To assess model significance, we also applied a *Y*-randomization procedure and found no statistically significant model according to CCR acceptance thresholds (see Materials and Methods); this result indicates that models developed with the original data are statistically robust.

In terms of applicability domain, the high similarity of biological profiles for particles of cluster 2 could be expected to yield better prediction performances within this cluster. To evaluate this hypothesis, we recalculated all statistical parameters per cluster for 5-fold external cross-validation results: cluster 1 ($n = 13$, CCR = 0.65, sensitivity = 0.5, specificity = 0.8), cluster 2 ($n = 18$, CCR = 0.78, sensitivity = 0.78, specificity = 0.78), and cluster 3 ($n = 13$, CCR = 0.7, sensitivity = 0.4, specificity = 1). These results confirm that prediction performances of our model were indeed better for MNPs comprised in cluster 2.

Then, we investigated the dose-dependency of biological effects induced by MNPs. The activity heat map representing biological activity induced by MNPs at four different concentrations is shown in Supporting Information, Figure S2. We also plotted the *Z* score variations for all 51 MNPs tested against aorta endothelial cells in the ATP content assay at four different concentrations (Figure 3). Overall, the higher was the dosage, the stronger were the NP-induced effects; however, we observed some interesting cases where this rule was not clearly followed. Although the vast majority of MNPs are characterized by small linear variations of *Z* scores corresponding to increases in their concentrations, some MNPs induced significantly higher *Z* scores at higher concentrations: NP_36 (PNP-Fe_2O_3−PVA, PEG), NP_29 (PNP-Fe_3O_4−PVA, protamine, rhodamine), NP_28 (PNP-Fe_3O_4−PVA, ethylenediamine), NP_37 (PNP-Fe_2O_3−PVA), and NP_20 (CLIO-SIA-FITC-Fe_3O_4-succinimidyl iodoacetate). Of all these MNP with "outlier" dose dependencies, our binary classification model assigns the class correctly for all but NP_20 (see Supporting Information, Table S1); NP_20 features a unique combination of molecular coating and surface modifi-

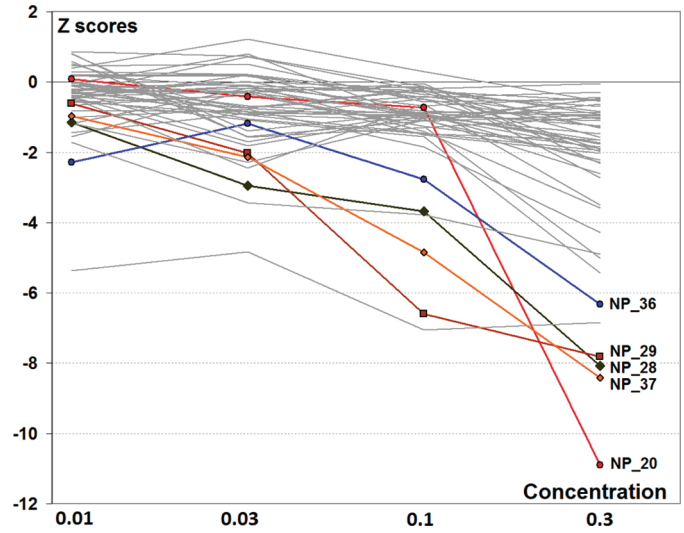

Figure 3. Analysis of *Z* score variations for all 51 nanoparticles tested against AO aorta endothelial cells in the ATP content assay at four different concentrations (0.01, 0.03, 0.1, and 0.3 Fe (mg/mL) for iron-based nanoparticles (NP_1−48) respectively; for the three quantum dot-based nanoparticles (NP_49−51), concentrations were equal to 1, 3, 30, and 100 nM). The labeled MNPs show the most dramatic dose-dependence of their biological effects, particularly at high MNP concentrations.

ers, and this nanoparticle displayed the highest variation in biological activity at a high concentration. Such examples highlight the complexity involved in modeling these chemical systems where minor changes such as a small variation of nanoparticle concentrations or surface modifiers may dramatically affect their biological activity profile. Although the modeling of such cases remains very challenging, we believe that this proof-of-concept study illustrates the ability of QNAR models to establish predictive relationships between structural attributes and biological activity of MNPs.

Case Study 2—Modeling of MNPs Uptake in PaCa2 Cancer Cells. Unlike the MNP set employed in Case Study 1, all MNPs included in the second set possessed exactly the same metal core. The structure of organic small molecule conjugated to the MNP surface was the only difference from one MNP to another. As a result, each MNP was represented by a unique set of descriptors determined by the conjugated small molecule. There were 150 MOE descriptors calculated for all 109 organic compounds. We expressed cellular uptake as the decadic logarithm of the concentration (pM) of MNP per cell, which varied from 2.23 to 4.44 (see Supporting Information, Table S2). Next, we performed a QSAR investigation and descriptor analysis to uncover major structural attributes responsible for cellular uptake of MNPs. An external 5-fold cross validation exercise was carried out in the same manner as in Case Study 1 employing the *k* nearest neighbors (*k*NN) modeling approach. Results showed that prediction accuracies expressed as coefficients of correlation R_{abs}^2 ranged from 0.65 to 0.80 for external sets (see Table 3). These results were slightly improved to 0.67−0.90 by taking into account the applicability domain of the models and removing compounds found to be outside the domain. We also performed *Y*-randomization, and no statistically significant models were retrieved, proving the robustness of QNAR models built on this data set.

To enable model interpretation, we identified descriptors that occurred most frequently in *k*NN models with the highest prediction accuracy. We calculated average values of these descriptors for MNPs with the highest (top 20) and the lowest (bottom 20) cellular uptakes (Figure 4a) and found that these values were significantly different in several cases. The top-10 most frequently selected descriptors in each individual fold and the averaged frequency across five folds are listed in Supporting Information (SM_Tables S3 and S4). It is of notice that several descriptors such as SlogP_VSA1, SlogP_VSA2, and SlogP_VSA5 represent different as-

TABLE 3. Case Study 2−QNAR Modeling of PaCa2 Cell Uptake for 109 MNPs with Different Surface Attachment

				no applicability domain		with applicability domain		
fold	modeling set	external set	no. models	R_{abs}^2	MAE	R_{abs}^2	MAE	coverage (%)
1	87	22	371	0.65	0.18	0.67	0.18	86
2	87	22	282	0.67	0.14	0.73	0.13	91
3	87	22	266	0.72	0.22	0.75	0.21	82
4	87	22	183	0.75	0.19	0.90	0.14	64
5	88	21	145	0.80	0.16	0.78	0.17	76
cumulative external sets (109 MNPs)				0.72	0.18	0.77	0.17	80

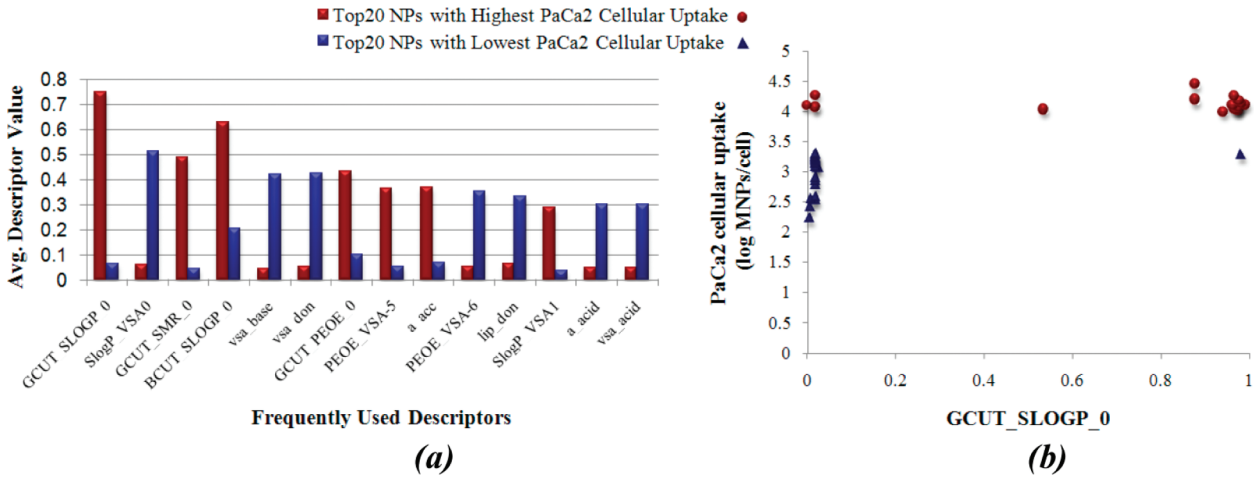

Figure 4. Analysis of descriptors used most frequently in *k*NN-QSAR models of 109 MNPs. (a) Average descriptor values in MNPs with highest and lowest PaCa2 cellular uptake. (b) Example of a lipophilicity related descriptor (GCUT_SLOGP_0), which significantly discriminates between nanoparticles with highest and lowest PaCa2 cellular uptake.

pects of van der Waals surface area's contribution to compound lipophilicity; another relatively frequent descriptor is b_double (representing the number of double bonds in a molecule).

Thus, lipophilicity was found to be the most discriminating factor; it is quantified by several descriptors such as GCUT_SLOGP_0, SlogP_VSA0, BCUT_SLOGP_0, and SlogP_VSA1. Consistent with this observation, MNPs with the highest PaCa2 cellular uptake are highly enriched for lipophilic surface compounds (high values of GCUT_SLOP_0); conversely, MNPs with the lowest PaCa2 uptake are highly enriched for low values of GCUT_SLOP_0 (Figure 4b). However, this phenomenon was only found in Paca2 cell lines. In the other cell lines tested by Weissleder *et al.*,[30] cellular uptake measured for the same series of MNPs revealed no significant variations correlating with the structural properties of MNPs. Other descriptors like molecular refractivity (GCUT_SMR_0), specific van der Waals surface area (basic vsa_base, acidic vsa_acid, and donor vsa_don), and electrostatic descriptors also reasonably discriminated between MNPs possessing high or low Paca2 cellular uptake. These findings imply that the cellular behavior of a nanoparticle library based on a common core can be predicted using QNAR analysis of the surface modifying ligands, and thus that rational design of organic compounds attached to the surface of MNPs is possible using QNAR models and descriptor analysis.

DISCUSSION

Although QSAR methodology is well-known and is extensively applied in the areas of drug discovery[29] and chemical toxicity modeling,[32] its application to model the biological effects of MNPs presents a real challenge for several reasons: (i) MNPs are complex assemblies of inorganic and/or organic elements, sometimes mixed or coated with diverse organic compounds where the exact stoichiometry may vary from one MNP to another, making classical molecular descriptors not appropriate for this type of study; (ii) the exact composition of a given MNP is not known in most cases; (iii) three-dimensional nanostructures that include thousands of atoms are highly complex. Many computational approaches, like *ab initio* quantum chemistry methods, are inadequate for such large, complex systems. Systematic physicochemical, geometrical, structural, and biological studies of MNPs are rare. Therefore, computational modeling of MNPs is only beginning to emerge. Most likely, a comprehensive computational nanotechnology and nanotoxicology effort would require the integration of several computational techniques, such as quantum mechanics, molecular dynamics simulations,[24,25,33] and cheminformatics.[27,34] The success will certainly depend on close collaboration with experimental scientists, as well as the application of high-throughput assay technologies to test MNPs, resulting in a sufficiently large body of data to enable large-scale modeling. Such strategies are entirely within a vision for toxicity testing outlined in a recent *Science* paper[35] and implemented in a joint project between EPA, NIEHS, and the NIH Chemical Genomics Center.

The overall goal of this study was to demonstrate the potential benefits of using cheminformatics approaches such as QSAR (or QNAR) modeling to obtain predictive knowledge for MNPs that affect human cells and utilize this knowledge to improve the experimental design of MNPs and enable their prioritization for *in vivo* testing (*e.g.*, to evaluate MNPs for therapeutic efficacy or toxicity). There were three fundamental hypotheses that drove this study: (i) the effects of MNPs on different types of human cells depend on the physical/chemical/geometrical properties of MNPs; (ii) high-throughput cellular-based assays can provide useful and predictive information about pleiotropic biological properties of MNPs; (iii) it is feasible to develop predictive quantitative nanostructure−activity/toxicity (QNAR/QNTR) models using physical/chemical charac-

terization and toxicological screens for an ensemble of MNPs.

Obviously, in our study we have attempted to tackle a very challenging problem of establishing predictive relationships between the structure of MNPs and their biological activity. It is undoubtedly true that biological effects of nanomaterials are strongly dependent on a large variety of factors. Should this consideration alone prevent any attempt to model MNPs? It may appear so; however, we should note that the problem of accurate prediction of biological activity (let alone, toxicity) of organic molecules, that is, traditional objective of QSAR modeling, is no less challenging when one thinks about dozens if not hundreds of interlinked mediatory processes in living cells (let alone whole organisms) that are responsible for the observed biological response phenotypes. Nevertheless, the QSAR approach has been successfully applied for many years to model very complex biological end points. Thus, we have been motivated to examine the applicability of QSAR approaches to modeling nanoparticles empirically. We should also emphasize that a critical strength of the data sets analyzed in our paper is that a large number of nanomaterials were simultaneously tested in the same laboratory under the identical culture conditions, thus enabling direct comparisons across nanomaterials.

Our computational approach addresses a significant near- and long-term problem that relates to the complexity, time, and cost associated with performing subchronic and chronic studies of novel nanomaterials in animals.[36] Because these types of comprehensive studies are impossible for all available MNPs, high-throughput cellular-based assays are needed that provide critical and predictive data in just a few hours. Shaw and colleagues[26] provided examples where these *in vitro* biologic activity profiles were correlated with *in vivo* MNP effects. As current efforts to correlate *in vitro* cellular activity with *in vivo* behavior (including toxicity) improve, QNAR models such as those presented here could help predict toxicity of newly designed nanomaterials and bias the design and manufacturing toward safer products.

To demonstrate the validity of the QNAR modeling approach, we have applied it in two case studies comprising two series of diverse MNPs. In case study 1, we studied a data set of 51 MNPs that Shaw *et al.*[26] tested extensively against four cell lines in four different assays. Our studies revealed three clusters of MNPs based on their induced biological activity and established specific nanostructure−activity relationships using cheminformatics approaches relying on multiple molecular descriptors of MNPs. We demonstrated the feasibility of deriving robust QNAR models using the following four experimental descriptors: size, relaxivities, and zeta potential. We should note that in an attempt to capture MNPs activity across a broader swath of biology, the Shaw group measured the effects of nanomaterials in four diverse cell lines, at four doses, and using four assays that interrogate different aspects of cellular physiology. Thus, the biological effects of each nanomaterial can be described as a 64-feature vector (4 cell lines × 4 assays × 4 doses). An analogy or inspiration for this approach may be found in the field of cancer genomics, where describing cancer cell lines using a common multidimensional vector (in this case composed of the expression levels of many different genes) has enabled many powerful computational analyses.

Theoretically, one could develop 64 independent QNAR models, with each model attempting to reproduce the biological response induced by 44 nanoparticles for a given assay in a particular cell line at a given dose. Actually, we have attempted to obtain such highly specific models but with no success for most cases (data not shown): despite pretty high fitting accuracy (~85%), the external predictivity (assessed by 5-fold external cross-validation) of these models were dramatically low (*ca.* 40−50%), not significantly different than the predictivity of models built with randomly shuffled activity of the training set (*i.e.*, using the standard *Y*-randomization test). Meanwhile, the combination of the entire 64 dimensional vector for each nanoparticle into one single averaged response apparently helped with the detection of the overall biological signal from noise; this was achieved by defining two different classes of particles (the threshold of averaged *Z* score have been put to −0.4 to balance nanoparticle distribution between the two classes). As we demonstrate in the manuscript, this data transformation allowed us to succeed in obtaining models characterized by both good internal fitness and external predictive power.

In the course of additional studies, we are pursuing different approaches for characterizing biological activity profiles, defining thresholds between profiles, and reducing the number of biological dimensions by compressing multiple measurements. For instance, we could define profiles based on clusters resulting from the previous analysis, or we could compare biological activity profiles using different general profile similarity metrics such as the Tanimoto coefficient, which is widely used to compare chemical structures. We anticipate that the quest to identify and predict rigorous relationships between MNP structures and their biological activity will require empiric exploration of several different approaches.

In case study 2, we investigated 109 MNPs with the same core structure but diverse organic molecules attached to their surfaces that were tested for cellular uptake against different cell lines.[30] The PaCa2 cell line was selected for in-depth QSAR study because of the significant variance of cellular uptakes among all tested MNPs. Each individual MNP was represented by the structure of the organic molecule attached to its surface. Statistically robust *k*NN QSAR models, linking

chemical descriptors and MNP cellular uptakes, were developed and validated using 5-fold external validation procedure. Their external prediction power was shown to be as high as R^2 of 0.72. Additional investigations are in progress to map chemical features responsible for differential uptake of MNPs onto chemical structures of surface modifiers and to detect the key structural fragments that mostly influence the cellular uptake. Overall, models assessing the potential cellular uptakes for particular cell lines are likely to be important tools to design novel cell-targeting particles that deliver drugs to those specific cells. We aim to develop an ensemble of models for use as efficient filters for computer-aided MNP design.

The quality of all QNAR models derived in this study was rigorously estimated according to their external prediction abilities assessed by a 5-fold external cross-validation procedure. Unlike many QSAR (or similar multidimensional data modeling) studies, we did not evaluate the power of our models based on their "too optimistic" fitting performances but on external predictions only (models are built and selected using modeling set only, not the external set). *Y*-randomization technique also assessing the chance correlation likelihood was used in both case studies to confirm the predictivity of generated models.

Before embarking on the huge task of predictive, computational nanotoxicology, it is necessary to demonstrate that statistical and data-mining techniques could indeed uncover the nonspurious nanostructure−activity correlations using experimental or computed properties of MNPs as structural descriptors. Our preliminary analysis of these two data sets provides a clear indication that this approach could indeed bear fruit. We also believe the two case studies reported in this paper represent the first attempts to build robust and validated QNAR models using either MNPs as a whole (case study 1) or particle-specific organic compounds representing the whole structure (case study 2). The two types of data sets studied in this report are representative of many similar data sets that hopefully will emerge in the published scientific literature and that could be subject to similar computational analysis. All too often the results of even large-scale experimental projects remain confined to individual laboratories or are published in unstructured format making it difficult, if not impossible, to access these data. We hope that with time and more data available in the public domain we will be able to establish an Integrated Nanotoxicology Web-Portal to enable the scientific community free access to both data and computational models. As part of these efforts, all data sets used in this study can be downloaded from the ChemBench portal (http://chembench.mml.unc.edu/) developed in our laboratory and are also provided as Supporting Information in Tables S1 and S2.

In summary, the trends in experimental nanotechnology and nanotoxicology require not only exploration and rationalization of experimental nanostructure−activity relationships, but most importantly, development of models that will help in designing environmentally benign nanomaterials, and prioritizing existing and novel MNPs for *in vivo* testing. Integrated data obtained from the characterization of the MNPs and systematically acquired *in vitro* data could enable the development of predictive QNAR models to correlate descriptors of MNPs with clinically important *in vivo* end points.

MATERIALS AND METHODS

Data Sets. ***Case Study 1: MNPs with Diverse Core Structures.*** Recently Shaw *et al.*[26] published a unique and comprehensive study in which 51 diverse MNPs were tested in various cell-based assays. Among these MNPs, 23 were cross-linked iron oxide (CLIO) derivates; 19 were pseudocaged nanoparticle (PNP) based; 4 were monocrystalline iron oxide nanoparticle (MION) based; 3 were quantum dot-based MNPs with a CdSe core, a ZnS shell, and a polymer coating; and 2 other were iron-based MNPs: Feridex IV (approved for *in vivo* imaging) and Ferrum Hausmann (approved for iron supplementation). All these MNPs were tested *in vitro* against four cell lines in four different assays at four different concentrations resulting in a 51 * 64 data matrix of experimental results. Each cell of this matrix (Supporting Information, Figure S1) reports the biological activity profile induced by a given MNP at a certain concentration in a particular assay for a given cell line. The four cell lines included monocytes, hepatocytes, and two types of vascular cells, namely, endothelial and smooth muscle. The four assays measured (i) ATP content, (ii) reducing equivalents, (iii) caspase-mediated apoptosis, and (iv) mitochondrial membrane potential. Biological activity profiles were recorded for the following concentrations of MNPs: 0.01, 0.03, 0.1, and 0.3 mg/mL for all iron-based MNPs; and 3, 10, 30, and 100 nM for the three quantum dot-based MNPs. Assay response values were expressed in units of standard deviations of the distribution obtained when control cells were treated with PBS (phosphate buffered saline) alone: $Z_{NP} = (\mu_{NP} - \mu_{PBS})/\sigma_{PBS}$, where μ_{PBS} is the mean of control tests with PBS, and σ_{PBS} is their standard deviation. The authors also reported four experimentally measured descriptors for 44 out of 51 tested MNPs: size, relaxivities, and zeta potential.

Case Study 2: MNPs with Common Core but Diverse Surface Modifiers. Weissleder *et al.*[30] recently synthesized a library composed of 109 MNPs in which a superparamagnetic nanoparticle (cross-linked iron oxide with amine groups, CLIO-NH_2) was decorated with different synthetic small molecules. NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate) molecules on their surfaces to enable measurement of cellular uptake. Then, NPs were screened against different cell lines, including PaCa2 human pancreatic cancer cells, U937 macrophage cell lines, resting and activated primary human macrophages, and HUVEC human umbilical vein endothelial cells. Unlike the other cell lines, the uptake of the NPs in PaCa2 pancreatic cancer cells was diverse and highly dependent on surface modification, enabling the application of QSAR modeling approach to this data.

QSAR Modeling. QSAR models establish quantitative relationships between chemical structures characterized by chemical descriptors and a target property, *e.g.,* biological activity of chemicals in specific biological assays. Validated and externally predictive models[29] can be applied to screen virtual chemical li-

braries to retrieve compounds with desired properties.[32,37,38] QSAR modeling employs complex machine learning algorithms such as support vector machines (SVM) or the *k* nearest neighbors (*k*NN) that take the descriptor matrix of compounds as inputs and output a predicted value for the modeled property.

The QSAR modeling workflow can be divided into three major steps: (i) data preparation/analysis[39] (selection of compounds and descriptors), (ii) model building, and (iii) model validation/selection (including the evaluation of its applicability domain, AD). A set of compounds with known experimental activity is randomly split into several training and test sets. Models are built using compounds of each training set and then applied to test set compounds to assess their properties. After application of rigorous tests (such as leave one out, *n*-fold cross-validation, and *Y*-randomization) and calculation of model accuracy metrics described below, certain models are selected if and only if they can reasonably predict both the training set as assessed by cross-validation procedures and the test set.[40] Models obtained for the modeling set with randomized activities (*Y*-randomization) should have significantly lower predictive capabilities than models built using the modeling set with real activities. Finally, the selected models are applied to the external validation set compounds.

Chemical structures are represented by molecular descriptors.[41] In Case Study 2, we used the following two-dimensional MOE descriptors (commercial software distributed by Chemical Computing Group): physical properties, surface areas, atom and bond counts, Kier & Hall connectivity indices, kappa shape indices, adjacency and distance matrix descriptors, pharmacophore feature descriptors, and molecular charges.

The clustering of a chemical data set consists of merging compounds into independent clusters that include chemically similar molecules[42] based on any similarity metrics (*e.g.*, compounds can be clustered based on their biological activity profiles). In this study, we have employed the ISIDA/Cluster program[31] implementing the sequential agglomerative hierarchical nonoverlapping (SAHN) method. The parent−child relationships between clusters result in a hierarchical data representation, or dendrogram. In particular, we used ISIDA/Cluster to obtain the heat map of the proximity matrix and the dynamic dendrogram (Figure 2).

The *k*NN QSAR method[43,44] is based on the idea that the activity of a given compound can be predicted by averaging the activities of *k* compounds from the modeling set, which are most chemically similar to this compound. Briefly, our algorithm employs the *k*NN classification principle and variable selection procedure (simulated annealing with the Metropolis-like acceptance criteria): it generates both an optimum *k* value, typically from one to five, and an optimal *nvar* subset of descriptors that maximize the QSAR model's training set accuracy as estimated by the Q_{abs}^2 statistical parameter. The Euclidean distance between compounds is used as a metric that characterizes compounds' dissimilarity in multidimensional descriptor space. Additional details of the method can be found elsewhere.[29] For SVM classification, we used the WinSVM program (version 1.1.8)[37] developed in our group at UNC, which implements the open-source libsvm package (*http://www.csie.ntu.edu.tw/~cjlin/libsvm/*).

The applicability domain (AD) of a model is defined in order to determine if a given model is capable of predicting the activity of a query compound[29,38] within a reasonable error. In this study, we defined the AD as a threshold distance D_T between a query compound and its nearest neighbors in the training set, calculated as follows: $D_T = \bar{y} + Z\sigma$ where $\bar{y}$ is the average Euclidean distance between each compound and its *k* nearest neighbors in the training set, σ is the standard deviation of the Euclidean distances, and *Z* is an arbitrary parameter to control the significance level; *k* is the parameter optimized in the course of QSAR modeling. We set the default value of *Z* at 0.5, which formally places the allowed distance threshold at the mean plus one-half of the standard deviation. If the distance of the test compound from any of its *k* nearest neighbors in the training set exceeds the threshold, the prediction is considered unreliable. In this study, we used this same approach for both case studies 1 and 2.

We used different statistical parameters to evaluate the performance of models. For binary classification problems (like case study 1), these are defined as: accuracy = (TP + TN)/(NA + NI); sensitivity = TP/NA; specificity = TN/NI; CCR = 0.5 (sensitivity + specificity), where NA is the total number of actives (or class 1), NI is the total number of inactives (or class 0), TP is the number of true positives (experimentally actives predicted as actives), TN is the number of true negatives (experimentally inactives predicted as inactives), and CCR is the correct classification rate.

When activities were represented by a range of values (case study 2), we used squared correlation coefficient (R_{abs}^2) for test set compounds, squared leave-one-out cross-validation correlation coefficient (Q_{abs}^2) for training set compounds, and mean absolute error (MAE) for the linear correlation between predicted (Y_{pred}) and experimental (Y_{exp}) data. For this study, *Y* is the Paca2 cellular uptake. These parameters are defined as follows:

$$R_{abs}^2 = \frac{1 - \sum_Y (Y_{expt} - Y_{pred})^2}{\sum_Y (Y_{expt} - \langle Y \rangle_{expt})^2}$$

$$Q_{abs}^2 = \frac{1 - \sum_Y (Y_{expt} - Y_{LOO})^2}{\sum_Y (Y_{expt} - \langle Y \rangle_{expt})^2}$$

$$\mathrm{MAE} = \frac{\sum_Y |Y - Y_{pred}|}{n}$$

In case study 1, the classification models were considered acceptable if $CCR_{CV} \geq 0.6$ and $CCR_{test} \geq 0.6$, whereas the regression models were considered acceptable in case study 2 if $Q_{abs}^2 > 0.6$ and $R_{abs}^2 > 0.6$.

Acknowledgment. We gratefully acknowledge the support from the Semiconductor Research Corporation (SRC) to A.T. and R.J.M., NIH Grant R01-GM66940 and EPA (Grant RD832720) to A.T., and NIH Grant U01-HL80731 to R.W. and S.Y.S.

Supporting Information Available: Sets of nanoparticles and supporting figures. This material is available free of charge *via* the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

1. Schmidt, K. Green Nanotechnology: It's Easier Than You Think; Project on Emerging Nanotechnologies, April 26, 2007, Washington, DC, 2007; pp 1−36..
2. Service, R. F. Nanotoxicology. Nanotechnology Grows Up. *Science* **2004**, *304*, 1732–1734.
3. Heinemann, M.; Schafer, H. G. Guidance for Handling and Use of Nanomaterials at the Workplace. *Hum. Exp. Toxicol.* **2009**, *28*, 407–411.
4. Kipen, H. M.; Laskin, D. L. Smaller Is Not Always Better: Nanotechnology Yields Nanotoxicology. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2005**, *289*, 696–697.
5. Hart, P. *Nanotechnology, Synthetic Biology, & Public Opinion* ; Project on Emerging Nanotechnologies, September 29, 2009; Washington, DC, 2009; p 117.
6. Oberdorster, G.; Oberdorster, E.; Oberdorster, J. Nanotoxicology: An Emerging Discipline Evolving From Studies of Ultrafine Particles. *Environ. Health Perspect.* **2005**, *113*, 823–839.
7. Bystrzejewska-Piotrowska, G.; Golimowski, J.; Urban, P. L. Nanoparticles: Their Potential Toxicity, Waste and Environmental Management. *Waste Manage.* **2009**, *29*, 2587–2595.
8. Oberdorster, G. Safety Assessment for Nanotechnology and Nanomedicine: Concepts of Nanotoxicology. *J. Intern. Med.* **2010**, *267*, 89–105.

9. Nyland, J. F.; Silbergeld, E. K. A Nanobiological Approach to Nanotoxicology. *Hum. Exp. Toxicol.* **2009**, *28*, 393–400.
10. Donaldson, K.; Stone, V.; Tran, C. L.; Kreyling, W.; Borm, P. J. Nanotoxicology. *Occup. Environ. Med.* **2004**, *61*, 727–728.
11. Xia, T.; Li, N.; Nel, A. E. Potential Health Impact of Nanoparticles. *Annu. Rev. Public Health* **2009**, *30*, 137–150.
12. Marquis, B. J.; Love, S. A.; Braun, K. L.; Haynes, C. L. Analytical Methods to Assess Nanoparticle Toxicity. *Analyst* **2009**, *134*, 425–439.
13. Suh, W. H.; Suslick, K. S.; Stucky, G. D.; Suh, Y. H. Nanotechnology, Nanotoxicology, and Neuroscience. *Prog. Neurobiol.* **2009**, *87*, 133–170.
14. Shvedova, A. A.; Kagan, V. E. The Role of Nanotoxicology in Realizing the "Helping Without Harm" Paradigm of Nanomedicine: Lessons From Studies of Pulmonary Effects of Single-Walled Carbon Nanotubes. *J. Intern. Med.* **2010**, *267*, 106–118.
15. Koziara, J. M.; Oh, J. J.; Akers, W. S.; Ferraris, S. P.; Mumper, R. J. Blood Compatibility of Cetyl Alcohol/Polysorbate-Based Nanoparticles. *Pharm. Res.* **2005**, *22*, 1821–1828.
16. Radomski, A.; Jurasz, P.; onso-Escolano, D.; Drews, M.; Morandi, M.; Malinski, T.; Radomski, M. W. Nanoparticle-Induced Platelet Aggregation and Vascular Thrombosis. *Br. J. Pharmacol.* **2005**, *146*, 882–893.
17. Harhaji, L.; Isakovic, A.; Raicevic, N.; Markovic, Z.; Todorovic-Markovic, B.; Nikolic, N.; Vranjes-Djuric, S.; Markovic, I.; Trajkovic, V. Multiple Mechanisms Underlying the Anticancer Action of Nanocrystalline Fullerene. *Eur. J. Pharmacol.* **2007**, *568*, 89–98.
18. Harhaji, L.; Isakovic, A.; Vucicevic, L.; Janjetovic, K.; Misirkic, M.; Markovic, Z.; Todorovic-Markovic, B.; Nikolic, N.; Vranjes-Djuric, S.; Nikolic, Z.; *et al.* Modulation of Tumor Necrosis Factor-Mediated Cell Death by Fullerenes. *Pharm. Res.* **2008**, *25*, 1365–1376.
19. Kane, A. B.; Petrovich, D. R.; Stern, R. O.; Farber, J. L. ATP Depletion and Loss of Cell Integrity in Anoxic Hepatocytes and Silica-Treated P388D1Macrophages. *Am. J. Physiol.* **1985**, *249*, 256–266.
20. Kang, S. J.; Kim, B. M.; Lee, Y. J.; Chung, H. W. Titanium Dioxide Nanoparticles Trigger P53-Mediated Damage Response in Peripheral Blood Lymphocytes. *Environ. Mol. Mutagen.* **2008**, *49*, 399–405.
21. Leonard, S. S.; Roberts, J. R.; Antonini, J. M.; Castranova, V.; Shi, X. $PbCrO_4$ Mediates Cellular Responses *via* Reactive Oxygen Species. *Mol. Cell. Biochem.* **2004**, *255*, 171–179.
22. Pulskamp, K.; Diabate, S.; Krug, H. F. Carbon Nanotubes Show No Sign of Acute Toxicity but Induce Intracellular Reactive Oxygen Species in Dependence on Contaminants. *Toxicol. Lett.* **2007**, *168*, 58–74.
23. Meng, H.; Xia, T.; George, S.; Nel, A. A Predictive Toxicological Paradigm for the Safety Assessment of Nanomaterials. *ACS Nano* **2009**, *3*, 1620–1627.
24. Liu, J.; Hopfinger, A. J. Identification of Possible Sources of Nanotoxicity From Carbon Nanotubes Inserted into Membrane Bilayers Using Membrane Interaction Quantitative Structure−Activity Relationship Analysis. *Chem. Res. Toxicol.* **2008**, *21*, 459–466.
25. Liu, J.; Yang, L.; Hopfinger, A. J. Affinity of Drugs and Small Biologically Active Molecules to Carbon Nanotubes: A Pharmacodynamics and Nanotoxicity Factor. *Mol. Pharm.* **2009**, *6*, 873–882.
26. 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.
27. Puzyn, T.; Leszczynska, D.; Leszczynski, J. Toward the Development of "Nano-QSARs": Advances and Challenges. *Small* **2009**, *5*, 2494–2509.
28. QSAR Expert Group. The Report From the Expert Group on (Quantitative) Structure−Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs. Series on Testing and Assessment, No. 49; Organisation For Economic Co-operation and Development 2004, 206.
29. Tropsha, A.; Golbraikh, A. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Curr. Pharm. Des.* **2007**, *13*, 3494–3504.
30. Weissleder, R.; Kelly, K.; Sun, E. Y.; Shtatland, T.; Josephson, L. Cell-Specific Targeting of Nanoparticles by Multivalent Attachment of Small Molecules. *Nat. Biotechnol.* **2005**, *23*, 1418–1423.
31. Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, V.; Hoonakker, F.; Tetko, I. V.; Marcou, G. ISIDA−Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 191–198.
32. Zhu, H.; Tropsha, A.; Fourches, D.; Varnek, A.; Papa, E.; Gramatica, P.; Oberg, T.; Dao, P.; Cherkasov, A.; Tetko, I. V. Combinatorial QSAR Modeling of Chemical Toxicants Tested against Tetrahymena Pyriformis. *J. Chem. Inf. Model.* **2008**, *48*, 766–784.
33. Song, Y.; Luo, M.; Dai, L. L. Understanding Nanoparticle Diffusion and Exploring Interfacial Nanorheology Using Molecular Dynamics Simulations. *Langmuir* **2010**, *26*, 5–9.
34. Martin, D.; Maran, U.; Sild, S.; Karelson, M. QSPR Modeling of Solubility of Polyaromatic Hydrocarbons and Fullerene in 1-Octanol and *n*-Heptane. *J. Phys. Chem. B* **2007**, *111*, 9853–9857.
35. Collins, F. S.; Gray, G. M.; Bucher, J. R. Toxicology. Transforming Environmental Health Protection. *Science* **2008**, *319*, 906–907.
36. Donaldson, K.; Borm, P. J.; Castranova, V.; Gulumian, M. The Limits of Testing Particle-Mediated Oxidative Stress *in Vitro* in Predicting Diverse Pathologies; Relevance for Testing of Nanoparticles. *Part. Fibre Toxicol.* **2009**, *6*, 13–21.
37. Fourches, D.; Barnes, J. C.; Day, N. C.; Bradley, P.; Reed, J. Z.; Tropsha, A. Cheminformatics Analysis of Assertions Mined From Literature That Describe Drug-Induced Liver Injury in Different Species. *Chem. Res. Toxicol.* **2010**, *23*, 171–183.
38. Tetko, I. V.; Sushko, I.; Pandey, A. K.; Zhu, H.; Tropsha, A.; Papa, E.; Oberg, T.; Todeschini, R.; Fourches, D.; Varnek, A. Critical Assessment of QSAR Models of Environmental Toxicity Against Tetrahymena Pyriformis: Focusing on Applicability Domain and Overfitting by Variable Selection. *J. Chem. Inf. Model.* **2008**, *48*, 1733–1746.
39. Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50*, 1189–1204.
40. Golbraikh, A.; Tropsha, A. Beware of Q2. *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
41. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000; pp 1−667.
42. Downs, G.; Barnard, J. Clustering Methods and Their Uses in Computational Chemistry. *Rev. Comput. Chem.* **2002**, *18*, 1–40.
43. Zheng, W.; Tropsha, A. Novel Variable Selection Quantitative Structure−Property Relationship Approach Based on the *k*-Nearest-Neighbor Principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185–194.
44. Shen, M.; Xiao, Y.; Golbraikh, A.; Gombar, V. K.; Tropsha, A. Development and Validation of *k*-Nearest-Neighbor QSPR Models of Metabolic Stability of Drug Candidates. *J. Med. Chem.* **2003**, *46*, 3013–3020.