Artificial Intelligence and Machine Learning in Computational Nanotoxicology: Unlocking and Empowering Nanomedicine

Ajay Vikram Singh,* Mohammad Hasan Dad Ansari, Daniel Rosenkranz, Romi Singh Maharjan, Fabian L. Kriegel, Kaustubh Gandhi, Anurag Kanase, Rishabh Singh, Peter Laux, and Andreas Luch

Advances in nanomedicine, coupled with novel methods of creating advanced materials at the nanoscale, have opened new perspectives for the development of healthcare and medical products. Special attention must be paid toward safe design approaches for nanomaterial-based products.

Recently, artificial intelligence (AI) and machine learning (ML) gifted the computational tool for enhancing and improving the simulation and modeling process for nanotoxicology and nanotherapeutics. In particular, the correlation of in vitro generated pharmacokinetics and pharmacodynamics to in vivo application scenarios is an important step toward the development of safe nanomedicinal products. This review portrays how in vitro and in vivo datasets are used in in silico models to unlock and empower nanomedicine. Physiologically based pharmacokinetic (PBPK) modeling and absorption, distribution, metabolism, and excretion (ADME)-based in silico methods along with dosimetry models as a focus area for nanomedicine are mainly described. The computational OMICS, colloidal particle determination, and algorithms to establish dosimetry for inhalation toxicology, and quantitative structure–activity relationships at nanoscale (nano-QSAR) are revisited. The challenges and opportunities facing the blind spots in nanotoxicology in this computationally dominated era are highlighted as the future to accelerate nanomedicine clinical translation.

1. Introduction

Computational toxicology aims to decode the factors which are responsible for toxic interactions using a sophisticated model. The model itself has to take all potential interactions of the substance into account to produce meaningful results. The evolution of artificial intelligence (AI) and machine learning (ML) led to the development of novel approaches for toxicity testing (Figure 1 and Table 1). Computational modeling in the field of nanomaterials is used to establish a correlation between their bioinformatics, the dynamics of the biological response, and the fate once the substance has reached the target organ of interest. Especially biology based mathematical models are well suited for the modeling of the fate of nanotherapeutics and engineered nanomaterials (ENMs) within experimental systems. ML was already successfully applied to simulate the bioinformatics and interaction of nanomaterials in a varying environment.[1] The most studied approaches for the assessment of nanomaterial induced toxicity are structure based mathematical models (e.g., Bayesian methods and Markov Chain Monte Carlo simulation) and in particular quantitative structure–activity relationships (QSAR) at nanoscale (nano-QSAR).[2] The QSAR computational approach endeavor to predict the biological function of a compound depending on the physicochemical properties and theoretical descriptors of molecules. The QSAR are most studied models owning the highest...
Physicochemical properties of the nanomaterial are a key requisite for the QSAR modeling. One approach for the determination of QSAR is based exclusively on a theoretical model as first principle. For example, chemical reactivity based on the energies of the molecular orbitals constituting the substance of interest could be used to derive the QSAR of the substance. Another modeling approach for QSAR is entirely statistical, i.e., essentially using pattern recognition in order to correlate descriptors that are potentially useful with the effect one is trying to predict. The hybrid QSAR model is a combination of mechanistic reasoning and statistical fit which use theoretical considerations to identify the descriptors that are likely to be predictive. The statistical models are effective to calibrate the parameters surrounding the molecular descriptors. There are some useful physicochemical properties (e.g., surface charge, corona, aggregation, solubility, etc.) for predicting biological activities of nanomaterials. Open-source ML tools can provide quantitative structure–property relationship (QSPR) workflow in this context to predict a variety of nanomaterial properties that could have cross-platform feasibility to integrate into existing nanoinformatics workflows.

These models can, of course, be used on their own right but they can also be employed to provide input parameters to the more complex physiologically based pharmacokinetic (PBPK) models. Through information on QSAR, simulations can be subsequently carried out, which are able to predict biological activities and potential interaction partners, which may induce a toxic response. A variety of different computational approaches was developed in the past years in order to improve simulations and produce reliable results.

2. Models Predicting Colloidal Properties as Determinant to Nanotoxicity

All QSAR models, which aim to predict the toxicity of a nanomaterial, rely on extrinsic material properties, which determine the bio-nanointeractions in the context of reactivity and nanotoxicity. These physicochemical descriptors, which describe the fate of nanomaterials within the body or in the environment, are incorporated into QSAR models. One of those descriptors is the point of zero zeta potential (PZZP) which is being used to predict the agglomeration/aggregation behavior of nanomaterials in computational models. The PZZPs are overlaid to simulate the potentials of ENMs in fluids with different pH levels to elucidate agglomeration of the ENMs within a particular tissue. Certain ENMs are able to exchange electrons with the cellular environment; a situation which may disrupt the cell redox states. In consequence, this leads to oxidative stress since the cell machinery has only limited capacity to compensate for radical reactions. Enhanced cytotoxic reactions, inflammation, and additional adverse effects in the body may arise.

Molecular modeling based SAR models for 1D ENMs such as nanoneedle or carbon nanotubes (CNTs) were shown suitable to predict their reactivity. The study demonstrated that the length versus the diameter ratio \((L/D)\) of a nanoneedle, shown as a
function of carbon atoms, has a vital influence on the reactivity. The frontier molecular orbital theory is another input parameter for the prediction of QSAR. A longer gap between the lowest unoccupied molecular orbital (LUMO) energy ($\epsilon_L$) and the highest occupied molecular orbital (HOMO) energy ($\epsilon_H$), reduces the stability and enables higher conductivity of the ENM, leading to an enhanced reactivity.$^{[16]}$ 3D Nano-QSAR model have also been developed via matching the low energy conformations, which were docked into ADME models to design novel nanotherapeutics as shown in Figure 2.

Another mathematical model which predicts toxicity of the nanoparticulate and long aspect ratio nanoneedle, and CNT, has only limited success for the crystalline materials.$^{[17]}$ The close proximity of atoms in a crystal structure increases the probability of an overlap of orbital energies and a subsequently split. This results in a valence band and a conduction band, separated by an energy gap. Models which are extended with this input parameter were able to estimate redox active nanomaterials more reliably. Moreover, it can also be predicted if chemicals in the cellular environments are able to donate or receive the electrons to the conduction band.$^{[18]}$ Overlapping conduction bands indicating redox potentials in the cellular environment of ENMs predicted the cytotoxicity and other disruptive effects for several metal oxide ENMs.$^{[19]}$ However, the common QSAR methods used for
Many of the assessed nano-QSAR models are also able to predict the biopersistance of the ENMs which support a safer regulation.\(^{[42]}\) Based on analyses drawn from the nanocluster growth rate and the physicochemical properties of the ENM a read across of the bulk chemical properties is used.\(^{[43]}\) The physicochemical properties of ENMs are employed to provide input parameters to the more complex PBPK models. The sophisticated Nano-QSAR models are employed to develop safer ENMs for nanomedicinal approaches as described below.

### 2.1. Nano-QSAR Cytotoxicity Models

A variety of different models are available to overcome the limits of standard QSAR approaches for nanomaterials. The computational hybrid nano-QSAR model for the nanocytotoxicity adopts dual descriptors: enthalpy of a formation (related to the bandgap energy), and the electronegativity (related to stability). The simplistic nano-QSAR approach successfully computed the cytocompatibility of metal oxide ENMs for a number of cell lines.\(^{[40]}\) In order to foster the development of computational tools for the assessment of toxicity induced by ENMs, the European Commission has launched different projects under the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) regulation.\(^{[41]}\) Many of the assessed nano-QSAR models are also able to predict the biopersistance of the ENMs which supports a safe regulation.\(^{[42]}\)

However, the value of these models is rather different when it comes to grouping of nanomaterials based on their similarity in terms of physicochemical properties and biological activities.\(^{[44]}\) Based on analyses drawn from the nanocluster growth rate and the physicochemical properties of the ENM a read across of the toxicological properties ENMs based on governance, risk, and compliance (GRC) transformation can be performed.\(^{[45]}\) Such QSAR based models can be used in their own right but they can also be employed to provide input parameters to the more complex PBPK models. The sophisticated Nano-QSAR models are employed to develop safer ENMs for nanomedicinal approaches as described below.

### 2.2. Metal Containing NPs

Fourches et al.\(^{[46]}\) considered four experimentally determined descriptors of 44 different nanoparticles (NPs)\(^{[47]}\) from size, zeta potential arising from the intensity of charge on the surface, R1 relaxivity, and R2 relaxivity. The last two terms are related to the magnetic properties of the NPs, which determine their ability to influence the relaxation rates of proton spins in the surrounding water molecules. They also analyzed lipophilicity dataset to evaluate cellular uptake. The authors investigated the effectiveness of support vector machines (SVM) and QSAR methodology in predicting nanotoxicology, and in designing safer NPs.\(^{[48]}\) Another study uses nano-SAR model\(^{[49]}\) to build and integrate the naïve Bayesian classifier on the same dataset.\(^{[49]}\) This hybrid model classify the 44 NPs into bioactive or inactive groups based on similar four descriptors—size, zeta potential, spin–lattice relaxivity, and spin–spin relaxivity.\(^{[47]}\) Yet another study investigated the toxicity of 82 metallic, metal–oxide, dendrimer, and...
polymer NPs against zebrafish embryos using six descriptors—concentration, shell composition, functional groups, purity, core structure, and surface charge.[50] Kleandrova et al. developed a nano-QSAR model to test the ecotoxicity of NPs taking into account six descriptors—molar volume, polarizability, size, electronegativity, hydrophobicity, and polar surface area.[51] Applying different cut-off values for several endpoints like CC_{50}, EC_{50}, IC_{50}, TC_{50}, and LC_{50}, the NPs were divided into either toxic or nontoxic. Further, ecotoxicity of three nickel based NPs was predicted which was in good agreement with the experimental
evidence. Researchers have also reported an accuracy of more than 90% in their cytotoxicity predictions in NPs. Luan et al. developed a QSAR-perturbation model to predict the cytotoxicity of NPs against mammalian cell lines considering three descriptors—molar volume, size, and polarizability of NPs; and reported an accuracy of 93%. Concu et al. developed a unified in silico machine learning model based on artificial neural networks to predict general toxicity profiles of NPs. Applying this model to 260 NPs using two families of descriptors—physicochemical and 2D topology, showed an accuracy of more than 97%. Boukhvalov and Yoon developed descriptors based on the results of first principle calculations for metal NPs. They considered two reactions: Ion extraction from the surface of a NP to aqueous media and water dissociation on the crystal planes with different miller indices, such as (001) and (111), nanorods, and two cubic nanoparticles of 0.6–0.3 nm size of different metals—Al, Fe, Cu, Ag, Au, Pt. The dependence of chemical activity on the shape and size of NPs was investigated. Another study using QSAR reports the toxicity tests for three types of cluster NPs—monometallic (Au-Pd) clusters, core–shell particles, and bimetallic clusters (Au/Pd) on *Escherichia coli* and CHO-K1 cells based on size and specific surface area of NPs. The cytotoxicity of bimetallic clusters (Au-TiO2, Pd-TiO2) and bimetallic clusters (Au/Pd-TiO2) was found to be enhanced compared to that of pure TiO2. The proposed mechanism of cytotoxicity of NPs was shown as the release of ions and reactive oxygen species (ROS) from the TiO2 surface.

### 2.3. Multiwall Carbon Nanotubes (MWCNTs)

Fiber-shape and ability to generate ROS are considered indicators of high toxicity of a material. While all CNTs are fiber shaped, some also produce ROS. To study the toxicity of MWCNTs, quasi-QSAR models were developed based on the representation of conditions—concentration, presence of S9 mix, type of MWCNT (surface area), and usage of preincubation in a quasi-simplified molecular input-line entry system (SMILES) form whose descriptor correlation weights were calculated with the Monte Carlo method. In another study, the genotoxicity was modeled as a function of five parameters—particle type (MWCNT or fullerene), illumination, concentration, metabolic activation, and preincubation producing satisfactory statistical parameters. In these examples, the endpoint of cytotoxicity measurement is the reverse mutation test (TA100). In another study, nano-QSAR models were developed to predict the toxicity of 20 MWCNT types in human lung cells. Quasi-SMILES was used to represent the physicochemical properties and experimental conditions: diameter, length, surface area, in vitro toxicity assay, cell line, exposure time, and dose. The quasi-SMILES-based nano-QSAR model calculation using Monte Carlo method, provided sufficient statistical parameters. In this case, the endpoint of cytotoxicity measurement is the cell viability (%).

### 2.4. Fullerenes

Taking the cytotoxicity data of C60 NPs toward *Salmonella typhimurium* from ref., a mathematical model as a function of dose, S9 mix, and illumination was constructed using quasi-SMILES descriptors obtained with the Monte Carlo method resulting in statistical parameters: $R^2 = 0.755, q^2 = 0.571$. Continuing the study, taking two datasets from ref., mathematical models as a function of dose, S9 mix, and illumination were constructed using quasi-SMILES optimal descriptors obtained with the Monte Carlo method. In this study several splits were made into the training, calibration, and validation datasets as opposed to one split in the earlier study. In both of these studies, the endpoint of cytotoxicity measurement was the reverse mutation test, either TA100 or WP2 uvrA/pKM101.

### 2.5. Silica Nanomaterials

Taking a similar approach as described above, a predictive model for calculating the cytotoxicity of 20 and 50 nm silica NPs was built as a mathematical function of size, concentration, and exposure time using quasi-SMILES descriptors. The dataset was split into three random sets—training, calibration, and validation. The toxicity was estimated by 3-[4,5-dimethylthiazole-2-yl]-2,5-diphenyltetrazolium bromide (MTT) assay as cell viability of cultured human embryonic kidney cells exposed to different concentrations of silica NPs. In another similar study, taking numerical cell viability data for silica NPs from literature and splitting it randomly into three datasets—sub-training, calibration, and validation, high determination coefficients (0.83–0.89) and $q^2$ parameter values (0.71–0.82) are reported for the subtraining set. The cell viability is also shown as a function of size, concentration, and exposure time. The endpoint of cytotoxicity measurements in both of these examples is the cell viability. Quasi-QSAR method (using quasi-SMILES optimal descriptors) was compared with the random forest (RF) approach in a set of cytotoxicity experimental data containing 19 data points for silica NPs. While for RF, aspect ratio and zeta potential were found to be the most important variables, for quasi-QSAR no such conclusion could be drawn. It was also demonstrated that the RF approach is applicable to modeling the cytotoxicity of silica. Later, better nano-QSAR models based on quasi-SMILES were built using CORAL software with high determination coefficients (0.8–0.95).

### 2.6. The ADMET In Silico Modeling Accelerating the Safer ENMs

Traditionally, nanotherapeutics was developed by testing them against several biological screens in time-consuming multi-step processes. The target compounds which showed good potential were then investigated for their pharmacokinetic properties, metabolism, and potential toxicity, during which adverse effects most commonly have been discovered. As can be imagined, this entire process was extremely expensive, tedious, labor-intensive, and time consuming, and most often resulted in a nonusable drug; requiring starting the process all-over again with a new drug. With the increase in biological screening and chemical synthesis, the demands for useful early information on absorption, distribution, metabolism, excretion and toxicity data (together called ADMET data) has also dramatically increased. This has led to the development and use of various high-throughput in
vitro ADMET screens. There is also an increasing need for reliable and robust tools for predicting these properties to serve two key aims. First, prediction at the design stage of new compound libraries so as to reduce the risk of late-stage attrition; and second, to optimize the screening and testing by looking at only the most promising compounds.

It has been reported that the quest for early, fast, and relevant ADMET data can be tackled in three ways. First, a variety of in vitro assays are automated using robotics and miniaturization as high throughput (HTP) advances. Second, in silico models are used to assist in the selection of both appropriate assays, as well as in the selection of subsets of compounds to go through these screens. Third, predictive models have been developed that might ultimately become sophisticated enough to replace in vitro assays and/or in vivo experiments. The need for ADMET information starts with the design of new compounds. At such an early stage, computational approaches can be used to make predictions and decide a lead, even though they may not be perfect at this point. This may also help in deciding the synthesis route of the compound. More robust tools can then be used to optimize and predict the mechanism of a lead compound into a feasible clinical product.

Understanding the relationship between important ADMET parameters and chemical structures helps in developing in silico models for estimating ADMET properties. For example, dose size and dose frequency (Figure 3) can be estimated from oral absorption, bioavailability, tissue penetration, clearance, and the distribution volume. Yet, despite its importance, lack of published data makes it difficult to predict such properties from chemical structures alone. A good understanding of physicochemical properties, coupled with their measurement and prediction, are crucial for a successful nanomedicine project.

Another example of an ADMET property that can be predicted is drug absorption. Simulation programs, such as GastroPlus and Idea, are available which help in lead optimization and compound selection. Such programs use computer simulation models to predict the rate and extent of absorption using in vitro data inputs.

There are two different kinds of approaches for predicting potential nanotoxicology issues in in silico models. The first approach uses models derived on the basis of extracting and organizing human knowledge and scientific literature. The second method approaches chemical structure descriptors and performs systematic analysis of the relationships between these descriptors and the nanotoxicological endpoints. Automated tools and new methodologies are required to ease the metadata extraction, semantic representation and analysis of published content to complement the first approach. Open source and freeware offer a forum for automatically extracting, enriching, and characterizing various structural and semantic aspects of scientific data published. Open forums help to evaluate the published data by depending on the scientific text mining system established within the system of the open sources. However, most of the systems did not analyze whether the system promotes fundamental task of classification of rhetorical data and sum-up the extractive text.

Major cons of the mathematical models arguably mimicking in vivo physiology and pharmacokinetics is of model animals, which largely encompass rodents, canine and rabbits which...
morphological differences differ from target humans. Further, morphological differences into internal organ systems in different age groups of humans (adult vs children) are generally not considered while physiological fitting parameters (i.e., breathing rate and route). Until recently, more advanced dosimetry models are introduced giving an opportunity to choose in accordance with relevance and biological similarities (e.g., swine or human model itself) which reflect accurate prediction in context with physiologically regional tissue deposition and distribution.

The second approach is hampered by the limited nanotoxicology data available in the literature. It has been reported that current software includes models for the irritation, sensitization, immunotoxicology, and neurotoxicity.[70,75,77] While these are important endpoints, there are several others for which in silico modeling can be faster, cheaper, and can fill gaps that could not be accessible. Understanding of the processes and product design is immensely helpful to extrapolate the modeling result to occupational exposure assessment. However, the models are still only at early stage as the data they are based on are limited.[70]

2.7. The Mechanistic Dosimetry Models for Inhalation Nanotoxicology

The Paracelsus adage “sola dosis facit venenum” that means dose makes poisons, makes “dosimetry” a strong force in nanotoxicology.[83] Why do we need a computational tool for establishing the prediction between dosimetry and risk assessment? First, it helps in the assessment of external exposure and biological response. The early dosimetry modeling complements the correlation of complication of those in vitro and inhalation studies.[84] When we do an experiment to decipher toxicity of a particulate matter, all we get is the results but we do not really gain any insight into mechanistic aspects of toxicology.[85] Being able to model the mechanisms empowers us to gain better understanding of the processes and the product design. Further, the modeling can be faster, cheaper, and can fill gaps that could not be addressed through experimental studies.[86] The experimental planning and execution is expensive and takes a long time, minimum 4–12 weeks for in vivo test. Therefore, dosimetry modeling could be a great help during the handling of preliminary data and information. However, subsequently this information can be extrapolated for the prediction to humans’ and animals’ responses. Here, we describe ML tools and AI advances applied for the development of key models in inhalation nanotoxicology and nanomedicine.

The multiple path particle dosimetry is a quantitative computational tool to predict the deposition of inhaled nanoparticulate in the upper and lower respiratory tract of a lung in a human and several other mammalian species.[87] Dosimetry is part of the ADMET model and describes the administered amount of substance in a given uptake pathway, e.g., the uptake of nanoparticles by inhalation. An experimental determination of the available dose for uptake into the body, whether human or animal, is very difficult. A direct determination of the actual number of particles in the different lung areas is not possible or only possible to a limited extent. Another parameter of interest is the variability of the species under investigation in terms of age and sex, also for PBPK and genome scale metabolic network (GSMN) models. Mathematical models developed by the national and international commission on radiological protection (NCRP, ICRP) or the multi path particle dosimetry model (MPPD) are able to address the issues without animal or human testing. The real objectives of computational dosimetry are to develop mechanistic models based on the physical and physiological parameters to govern the transport within the respiratory tract. As a result, the local and regional deposition of particulate matter in the lung can be predicted.[88] With the MPPD, the local and regional depositions and clearance of particulate matter in the lung are predicted. It is practically impossible to design an entire lung models since we do not know every detail of all the elements, about ventilation physiology and physics in context with aerodynamics to assign the boundary conditions.[89] An alternative is to develop a dose-limiting model from the entire respiratory tract by making simplified assumptions regarding the geometry, lung ventilation and particle transport. The MPPD engine models picture the entire breathing process complementing inhalation and exhalation for deposition, and clearance as respiratory cycles, applicable to inhalation toxicology and workplace exposure assessment (Figure 4).[90]

If the exposure of an engineered, accidental or manmade nanomaterial is known and the subsequent inhalation test is to be performed in an animal model, the MPPD user interface makes it extremely easy to run the model quickly to speed up the test, significantly reducing the time.[91,92] The morphometry for different age groups of humans (adults vs children) and experimental animal variations (guinea pigs, rhesus monkeys, rats, mice, etc.) help to precisely decide input characteristics.[93] The morphometric data such as mass number, particle characteristics such as concentration, diameter, size distribution, and physiological fitting parameters (e.g., breathing frequency, total volume, breathing routes: nasal vs oral) give key information about regional tissue deposition and distribution.[93] The addition of more advanced features in inhalation methods immensely helps to extrapolate the modeling results to occupational exposure limits (OELs).[84] These computational models also offer a useful platform for interfacing with other physiological and mechanistic models such as PBPK/ADMET, etc.

3. Decoding Nanotoxicology Related OMICS Profile with PBPK Models

The OMICS (proteomics, genomics, and metabolomics) datasets are often complex and heterogeneous.[94] Extracting meaningful data from this vast amount of information is by far the most important challenge for the machine learning experts and bioinformaticians. In this context, there is an increasing interest in the potential of deep learning (DL) and artificial neural network methods to create predictive trends and to identify complex patterns from these large datasets. The PBPK models are the mathematical model that aims to integrate the known information from existing data about the physiological processes for the particular species of interest. It further enables the possibility to add any other known attributes about a specific target compound in order to predict or simulate formal kinetic properties in vivo.[95] The PBPK models are quite similar to classic top-down two compartment models and structure. However, it involves organs as the
compartments and the blood flow rates decide basically the drug transport from one compartment to the other. In PBPK models, we have organs arrow as our compartments they can be used to simulate concentration time profiles in any tissue of interest (Figure 5). The PBPK models have great extrapolation abilities from one species to another and one population to another.[96] For the prediction model, the common parameters can be divided into two main groups: physiological parameters which are the species specific which would include the organ volume surface area tissue compositions blood flow rates protein abundance. On the other side, the nanotherapeutics specific parameters, one need mainly the distribution related parameters and disposition related parameters. Of course, for the extra vascular absorption and the permeability calculations, one also needs metabolism, elimination and absorption rates to predict accurately.

The PBPK based toxicity models are a suitable tool to identify the toxicity mechanism associated with different classes of nanomaterial. However, if the target genes are located in the toxicity related network pathway, it does not give us more information about the causal mechanisms of regulation.[97,98] Therefore, we cannot predict if a particular gene regulates the other genes or any other working mechanism inside. Hence, we need to apply other types of ML approaches including deep learning, random forests, k-nearest neighbors, and support vector machines to identify the relevant effector pathways.

In OMICS screening of nanomaterials, we start with a nanomaterial with known surface physicochemistry and we identify which proteins/sites get perturbed or targeted by the specific nanoparticle interaction.[99] The PBPK model enables to screen the transcription factors involved in the signaling pathway, which regulate the expression of the target genes related with the nanotoxicity (e.g., cell death, proliferation, oxidative stress, and inflammation).[97] The versatility in machine learning algorithm helps to predict absorption, distribution, metabolism and excretion of different classes of nanomaterials via PBPK modeling in different organs.[100] There are a variety of modeling approaches that aim at identifying nanotoxicity pathways. Table 1 examines the most commonly utilized machine learning models for the prediction of nanotoxicity along with their pros and cons.
other organisms with an ensemble of interaction networks be-
with the toxicity-associated genes.\[105\] Chemical biology will fur-
protein–protein interactions that connect the toxicity phenotype
tree approaches to identify the pathways to help identifying the
analyzethedatawiththeMonteCarloalgorithmsuchasSteiner
erapeutics. Another example is the implementation of these to
understand the crystal structure information relevant to nanoth-
ment, will help us to cope with a situation.\[106\] This core network could be fed
into the PBPK model algorithm to further expand the scope of
This is of utmost importance for nanotoxicology response
analysis. The curiosity model must be incorporated one way or
another in modeling the NP-associated metabolism. The curios-
ity models or algorithms implement the curiosity reward to the
agent learning process and curiosity is framed as the interest of
an agent to learn unknown regularities. The idea is to reward
the system if it finds something unexpected.\[107\]. A curiosity re-
ward is added to the agent learning process, proportional to the
difference between the expected result from a given input and the
actual output. Such approaches have been successfully ap-
plicated in cancer nanomedicine to address the metabolomics\[108\]
and could be very helpful in perspective we describe herein. Dur-
ing the last decade, there has been a lot of work done for the
development of GSMN in the context of nanotoxicology.\[97,109\]
These are being used to model human metabolism at different
levels, particularly at the cellular level for tissues and organs de-
pending on the models. The GSMN models are being used to
predict the human metabolomics integrating organ- and tissue-
specific genome-scale metabolic network. Since vast informa-
tion is present at genome level, the implementation of GSMN
models might deepen our mechanistic understanding toward hu-
mnan metabolism. A few of the models are publicly available and
are represented by a set of differential equations.\[110\] The large-
scale models imply hundreds of equations to estimate the human
metabolic reactions in the nanotoxicology perspective. Many pub-
lc available models also give information about gene expres-
sion network, protein confirmation and co-crystallization in the
toxic environment as implemented in contemporary protein data
banks (PDB) (Figure 7).

PBPK models help in the development of formulation
for nanocarriers, predicting dose–effect relationships and risk
assessments toward nanopharmaceuticals.\[99,111\] The draft is
adopted for the eco guidance on the grouping of nanomaterials
for the assessment of these substances according to REACH, the
chemicals regulation of the European Union. The guideline helps
to distinguish between different kinds of properties that essen-
tially describe the chemical and physical identity of the substance.

4. Challenges Ahead with ML and AI in
Nanotoxicology Modeling

Like there is a need for a discussion to establish ethical guide-
lines for the operation of robots (roboethics), there must be dis-
cussions regarding AI and ML too, including for nanotoxicology
applications.\[112\] The guidelines must be established to provide
indications and suggest principles, which, at the opportune mo-
ment, will help us to cope with a situation.

Recent developments in modeling ligand–protein interac-
tions are opening new avenues for superior nanomaterials design,
which needs ML approaches to optimize the nanomedicine ap-
plications. This could be particularly promising for the per-
sonalized cancer nanomedicine since cancer cells have recently
been known to secret protein/peptides which can make advance
Figure 6. Linking PBPK models interfacing relevant physiological organs and tissues including the drug-specific ADME processes with the genome-scale metabolic network. The latter describes the cellular biochemistry in the interstitial (extracellular) and intracellular (cytosol, mitochondria, peroxisome, etc.) space. Reproduced with permission. [97] Copyright 2018, Nature.

Materials of biological origin, One of the biggest challenges in nanomaterial library design is the scarcity of computational tools to predict the ligand–metal interactions (number of reducing amphiphilic amino acid ligand and gold ion binding site), which could be helpful in predicting the final shape and size of the nanostructure formed. The machine learning based hierarchical approach termed as Iterative Threading ASSEmblY Refinement (I-TASSER) is able to predict the gold ion ligand binding sites into self-assembled short amino acid sequences (15–20 amino acid long). The i-TASSER approach could be tamed to design thin platelets nanoflower like structures with superior plasmonic wave guidance for photothermal applications [114,115] (Figure 8). Predictive efficacy of any mathematical model depends on simplifying assumptions and at least a compromise
Table 2. Algorithms and modeling for nanotoxicology and nanomedicine.

| Models          | Nanomaterial                                      | Algorithm                                           | In vitro/in vivo/in silico | Reference |
|-----------------|---------------------------------------------------|-----------------------------------------------------|-----------------------------|-----------|
| PBPK            | 5 and 30 nm CeO₂ NPs                              | Gear algorithm or Berkeley Madonna with linear regression | In vivo rat model           | [25]      |
|                 | Nanocarrier                                       |                                                    |                             |           |
|                 | TiO₂                                              |                                                    |                             |           |
|                 | Carbon nanotube                                   |                                                    |                             |           |
|                 | Microspheres                                     |                                                    |                             |           |
|                 | Functionalized iron Oxide NPs                    |                                                    |                             |           |
|                 | Metal oxide nanoparticles                         |                                                    |                             |           |
|                 | C₆₀ NPs                                           |                                                    |                             |           |
|                 | Small molecules and NPs                           |                                                    |                             |           |
|                 | ADMET                                             |                                                    |                             |           |
|                 | i-TASSER                                           |                                                    |                             |           |
|                 | ZnO, CuO, Co₃O₄, and TiO₂                          | Focus on activity calculations                      | In silico                   |           |

between physiology and physics of input labels. In dosimetry modeling for inhalation toxicology, multiscale airway dimensions and complicated airway morphology (straight tube vs dichotomous branching) of the lower respiratory tract (LRT) make conditions challenging. The LRT is basically tube-within-tube and implementing those features is a herculean task in dosimetry and allied regenerative nanomedicine models. Further, physics and aerodynamics of physiological lung ventilation is not fully understood, which makes it difficult to implement this in dosimetry. For example, parabolic flow velocity assuming low Reynolds number, convective versus resistive forces for air transport assuming independent events are only some of the fundamental assumptions that need more concerted evidence. To assume nanoparticulate versus microparticulates transport depositions, averaging cross sections as 1D models may compromise predictive power, and need further improvement. High data requirements and tuning the code could be complicated for a use to application variability in nanomedicine. The current algorithms have specific challenges in implementing each kind of modeling (Figure 9). Many parameters, which are correlated, have been implemented under different ML toolboxes in each kind of modeling. There must be adoptable common standard operating procedures (SOPs) to use the best predictive ML toolbox across all models.

5. Future and Outlook: Green Algorithms, Ethics, and Prospects of Remote Monitoring of Nanomaterials for Public Safety

With respect to the application of the 3R principles (replacement, refinement, and reduction of animal experiments) during implementation of the EU REACH regulation, recent algorithms have shown higher accuracy than animal test prediction. They have proved a promising potential and perspective to diminish animal testing during the development of future nanomedicine. The algorithms in focus can generate a chemical map that contains the putative toxicity ingredients of thousands of compounds.
Figure 7. Hierarchical integration of genome-scale metabolic networks at multiple orders of biological organization into whole body PBPK model. Top panel: Preparations and input of pharmacokinetics (PK) and OMICS data on the organism level. A PBPK model is developed and validated with human PK data. A human organ-specific genome-scale metabolic network (GSMN) model is reconstructed with the help of OMICS data. Middle panel: A hybrid PBPK–GSMN model is developed to estimate the in vivo organ-specific drug metabolism as time-resolved reaction rates including the absorption, distribution, metabolism, excretion processes. Lower panel: Prediction of cellular responses with the combined multiscale PBPK–GSMN model to predict organ-specific drug-induced metabolic perturbations, resulting in altered intracellular and extracellular reactions rates. The combined model allows the explicit consideration of specific dosing schemes, patient physiology, and genetic characteristics (abbreviations: dynamic version of the minimization of metabolic adjustment algorithm (dMOMA). Reproduced with permission.[97] Copyright 2018, Nature.

from databases with superior predictability. The algorithm can compare and replace an unknown chemical moiety within the map from thousands of databases available in nanochemistry, predicting the potential side effects and toxicity. The algorithm, named after its discoverer Hartung,[131] is an improvement on the read across methodology in toxicology and hailed as a software which could replace animal tests.[132] Both, EU and USA make it mandatory to perform health-based risk assessments prior to the sale of consumer products containing nanomaterials. Unlike most of Asian countries and other Schengen states including Norway and Israel where animal testing of cosmetic and allied products with nanomaterials is feasible, the EU has set a high priority on the development of alternatives to animal tests. The read across algorithm and recent ML tools such as the Hartung algorithm will immensely support big data analysis from animal alternatives, organs-on-chip, and 3D culture models in vitro.[133] These methods are getting an ethical voice as insightful alternative methods for animal tests, supporting the 3R principle. Europe’s REACH regulation, an ambitious project to implement the safety of chemicals, has largely benefited from the support of ML tools.[134] Further, the point that ML/AI helps to accomplish more in less time, will shorten the duration of experimental task and analysis lags in nanomedicine. The functional aspect and biological responses of every individual against particular toxicogenic stimuli is different. The AI gives an edge to crack down the differences since “one size fits all” is not applied to read the output of modern modeling tools. Despite these ML advances, a long road ahead awaits to completely replace animal tests with sole
Figure 8. A,B) Mathematical simulations of temperature distribution in electron microscopy images using the 3D Finite difference time domain (FDTD) algorithm in nanoflower petals after excitation with $\lambda = 780$ nm of two different anisotropic nanoflowers shown with SEM image in (C,D). E,F) Infrared thermal imaging to read rise in temperature. G–R) Photothermal effect based killing of untreated bacteria biofilms and their comparison with biofilms supplemented with K−N) nanoflowers and O−R) spherical gold NPs. S) iterative threading assembly refinement (i-TASSER) algorithm for the secondary structure predictions of a short peptide in the presence of gold ions (green dots), docking as ligand into the peptide to produce nanoclusters. Reproduced with permission.[114] Copyright 2018, American Chemical Society.
computational intelligence, since it will be scientifically a hard pill to swallow. However, riding on advances in ML toolbox and AI software which are more accurate compared to animal testing in predicting nanotoxicity, emphasis is now shifting toward a sustainable nanomedicine expansion as “green nanotoxicology.”[135]

Like telemedicine, we demonstrate that the remote monitoring of nanomaterial containing consumer products will further win public confidence and bring transparency in biosafety measures. AI and machine learning will be at heart of such technology complemented with advanced in nano-biosensors, which may enable the ability to institute the “cradle-to-grave” monitoring of nano and other advanced material-enabled products. With a unique universal material ID (UID) and potentially also the incorporation of trustless, block chain-based ledger system, future nanomaterials could be tracked from their production, deployment into the consumer or industrial market and eventual end-of-life recycling and safe disposal; all information about the safety and handling requirements for each enhanced material could be easily available through a simple scan on a handheld device. Recent developments in breathing/air quality monitoring sensors, which measure gases/volatile compounds as input and laser sensors to measure particulate in surrounding, can be used to scan the consumer products (e.g., textiles, toys, food/food contact and education/recreational kits containing nanomaterials). In case of complexity, the information collected after scan can be wirelessly sent to regulatory experts in local area for interpretation and feedback (Figure 10).[136] The wireless communication between ‘fitness monitoring’ wearable e-gadgets and smartphone with sophisticated analysis app developed with AI/ML, could be followed for nanomedicine and monitoring biosafety of nanomaterials. This will also fill the communication gap between regulatory bodies and public domains.[137]

As shown in Figure 11, expanding the AI toolbox will help to connect and benefit the interdisciplinary research in

---

**Figure 9.** A futuristic hybrid multimodal system linking in vitro models. The formation of nanomaterials (NM) is based on DFT calculations including I) the material synthesis ligand associated, II) structural information like size, shape and composition, III) the interaction and adsorption of molecules or proteins at the surface of created NM. Relevant physical–chemical and biological properties (like solubility, surface charge, partition coefficient, etc.) are predicted by nano quantitative structure–activity relationship (nano-QSAR). The synthesized NM is investigated with respect to administration, adsorption, distribution, metabolism, excretion, and toxicity (ADMET) into human body over material or human lifespan by multiple administration. Physiologically base pharmacokinetic models (PBPK) are used to predict organ specific and respectively whole body adsorption and distribution. By combination of PBPK models to metabolic networks the NM metabolism can be investigated for all organs implemented in the human body model. By using optimized models, the excretion of the administrated nanomaterials can be calculated if necessary. Additionally, outputs of the multiscale PBPK-metabolic networking modeling route are the toxicity and estimated input parameters for NM administration. All models investigated with multiple input parameters are supported by machine learning (ML) approach for faster and reliable prediction.
nanotoxicology. For examples, the nanotheranostic (nanocarrier and nanosensors) and food nanotechnology is rather less explored in context with new materials introduced and their possible health risk associated with exposure, particularly leaching of nanomaterials from coating. The Machine Learning algorithm can have better applications in predicting the different unknown nodes of material properties and composition (grey and cyan box in Figure 11) to connect with the surrounding behavior media (teal box) to predict the toxic response (purple). The rapidly computed empirical and ab initio models (e.g., i-TASSER, ab initio modeling of proteins for molecular replacement (AMPLE) as iterative-and-truncated approaches) could be key to solve the physicochemical property conundrum of hidden layers of toxicology.
Acknowledgements

The authors thank Anne Rosenkranz for her constructive suggestions in drawing graphics. This work was supported by BFR SFP 1322-725; BFR SFP 1322-733 (A.V.S.), BFR SFP 1322-642 (F.L.K.) and BFR SFP 1322-724 (D.R.). Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

AI, machine learning, nanomedicines, nanotoxicology, physiologically based pharmacokinetic modeling

Received: December 27, 2019
Revised: April 17, 2020
Published online: July 6, 2020
Polacchini, F. Rotondo, M. Vaccari, M. Villani, R. Serra, A. Colacci, Microchem. J. 2018, 136, 143.

[101] a) S. M. Bryce, D. T. Bernacki, J. C. Bemis, R. A. Spellman, M. E. Engel, M. Schuler, E. Lorge, P. T. Heikkinen, U. Hemmann, V. Thymbra, S. Wilde, N. Queisser, A. Sutter, A. Zeller, M. Guérard, D. Kirkland, S. D. Dertinger, Environ. Mol. Mutagen. 2017, 58, 146; b) H. L. Karlsson, S. Di Bucchianico, A. R. Collins, M. Dusinska, Environ. Mol. Mutagen. 2015, 56, 82; c) Y. Zerdoumi, K. K. Raman, F. Soubigou, S. Adriouch, G. Bougeda, T. Frebourg, J. M. Flaman, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 2015, 789–790, 28.

[104] a) X. Wang, Y. Jin, BMC Genomics 2017, 18, 716; b) J. Dapkokas, K. Olechnovíc, Č. Venclová, Proteins: Struct., Funct., Bioinf. 2018, 86, 292; c) F. Zhang, H. Song, M. Zeng, Y. Li, L. Kurgan, M. Li, Proteomics 2019, 19, 1900019; d) M. Howarth, Nat. Struct. Mol. Biol. 2015, 22, 349.

[105] a) M. Bereta, J. Telecommun. Inf. Technol. 2017, 4, 71; b) Y. Sun, C. Ma, S. Halgamuge, BMC Bioinformatics 2017, 18, 551; c) M. Wakeham, A. Nguyen, Q. Zhang, FASEB J. 2018, 32, 1035.

[106] O. Basha, R. Shrpringer, C. M. Argov, E. Yeger-Lotem, Nucleic Acids Res. 2018, 46, D522.

[107] Y. Burda, H. Edwards, A. Storkey, O. Klimov, arXiv preprint arXiv:1810.12894 2018.

[108] A. Mao, presented at 2018 IEEE 23rd Int. Conf. on Digital Signal Processing (DSP), Shanghai, China, November 2018.

[109] a) E. M. Maldonado, V. Leoncikas, C. P. Fisher, J. B. Moore, N. J. Plant, A. M. Kierzek, CPT: Pharmacometrics Syst. Pharmacol. 2017, 6, 732; b) C. Maroteaux, E. L. Schymanski, S. Neumann, B. Merlet, R. M. Salek, F. Jourdan, O. Yanes, Metabolites 2018, 8, 51.

[110] a) J. E. Sager, J. Yu, I. Raguenneau-Majlessi, N. Isotier-Carrera, Drug Metab. Dispos. 2015, 43, 1823; b) V. Yellepeddi, J. R. Brown, J. Liu, S. Kumar, J. Rashid, C. M. T. Sherwin, Clin. Pharmacokinet. 2019, 58, 1.

[111] a) I. E. Templeton, N. J. Jones, L. Musib, AAPS J. 2018, 20, 31; b) X. Zhang, S.-M. Huang, J. Pharm. Sci. 2017, 106, 2204.

[112] P. Houdy, M. Lahmani, F. Marano, Nanoethics and Nanotoxicology, Springer Science & Business Media, Berlin 2011.

[113] a) A. V. Singh, M. Batuwangala, R. Mundra, K. Mehta, S. Patke, E. Falletta, R. Patil, W. N. Gade, ACS Appl. Mater. Interfaces 2014, 6, 14679; b) A. V. Singh, T. Jahnke, V. Kishore, B. W. Park, M. Batuwangala, J. Bill, M. Sitti, Acta Biomater. 2018, 71, 61.

[114] A. V. Singh, T. Jahnke, S. Wang, Y. Xiao, Y. Alapan, S. Krishnati, M. C. Onbasli, K. Kozieński, H. David, G. Richter, J. Bill, P. Laux, A. Luch, M. Sitti, ACS Appl. Nano Mater. 2018, 1, 6205.

[115] A. V. Singh, Y. Alapan, T. Jahnke, P. Laux, A. Luch, A. Aghakhanlou, S. Krishnati, M. C. Onbasli, J. Bill, M. Sitti, J. Mater. Chem. B 2018, 6, 7573.

[116] J. Basu, J. W. Ludlow, Trends Biotechnol. 2010, 28, 526.

[117] J. P. Butler, A. Tsuda, Compr. Physiol. 2011, 1, 1301.

[118] F. Jung, L. Nothnagel, F. Gao, M. Thurn, V. Vogel, M. G. Wacker, Eur. J. Pharm. Biopharm. 2018, 127, 462.

[119] G. Bachler, N. von Goetz, K. Hungerbuhler, Nanotoxicology 2015, 9, 373.

[120] C. Riebeling, A. Luch, M. E. Götz, Nanotoxicology 2016, 10, 343.

[121] A. V. Singh, Y. Alapan, T. Jahnke, P. Laux, A. Luch, M. Sitti, J. Mater. Chem. B 2018, 6, 7573.