Chapter 14
Quantitative Nanostructure–Activity Relationships: Methods, Case Studies, and Perspectives

Denis Fourches and Ryan Lougee

Abstract In this chapter, we discuss the development and application of molecular modeling methods to analyze and forecast the experimental properties of nanomaterials. We mainly focus on Quantitative Nanostructure—Activity Relationships (QNAR) to evaluate the extent of biological activities potentially induced by various types of nanomaterials. First, we present the basic principles of QNAR modeling that uses machine-learning techniques to establish quantified links between the biological endpoint of interest (e.g., cytotoxicity, cell death, ROS production) and nanomaterials' characteristics. Second, we briefly review recently published studies reporting on the QNAR modeling of the largest and most significant datasets of nanomaterials available in the public domain. Third, we discuss some perspectives for the use of molecular modeling on nanomaterials. Overall, we show how molecular modeling can represent a key element for enabling the rational design of nanomaterials with the desired activity and safety profile.

Keywords Molecular modeling · Cheminformatics · QNAR · Machine learning · Virtual screening

14.1 Introduction

Nanotechnology [1] represents the final frontier for advanced material manufacturing at the atomic resolution. Manufactured nanoparticles (MNPs) are materials with at least one dimension varying from 1 to 100 nm. At that scale, such materials are characterized by unique optical, thermal, electrical, magnetic, and biological properties [2]. As a consequence, this is no surprise that the research on novel MNPs has led to a wide interest in many areas of research and industrial applications. Thus, nanotechnology is now seen as a global, multi-purpose technology [3].

D. Fourches (✉) · R. Lougee
Department of Chemistry, Bioinformatics Research Center,
North Carolina State University, 1 Lampe Dr, Raleigh, NC 27695, USA
e-mail: dfourch@ncsu.edu

B. Yan et al. (eds.), *Bioactivity of Engineered Nanoparticles*,
Nanomedicine and Nanotoxicology, DOI 10.1007/978-981-10-5864-6_14

Its worldwide impact is expected to be as big as that of plastics with a global market reaching three trillion dollars as early as 2020 [4, 5]. As of mid-2013, more than 1800 consumer products from 622 companies and 32 countries have already been inventoried [6].

One area of great interest for the use of nanotechnology is nanomedicine. As potential medical devices, MNPs are already capable of being used on various surfaces as antimicrobials, water purifying agents, or as electrochemical biosensors [7, 8]. Other nanomedicine-relevant applications have also been explored in which MNPs are directly interacting with biological systems. For instance, we can highlight the use of MNPs for achieving fluorescent labeling, drug delivery, detection of pathogens, detection of proteins, probing of DNA structure, tissue engineering, tumor destruction, purification of biomolecules and cells, or MRI contrast enhancement [9]. Furthermore, nature actually relies on complex nanoparticles in many organisms: e.g., cephalopods fabricate reflective protein platelet nanostructures [10], fireflies generate nanostructured cuticles on their abdomen, which enhance the emission of bioluminescent light [11]. We could also underline *Geobacter sulfurreducens,* a sulfur-reducing proteobacterium, which utilizes protein-based nanowires to transfer electron in the extracellular environment [12]. Thus, the potential of nanomaterials for medical applications with therapeutically relevant outcomes is tremendous. However, this objective is only valid as long as the MNP-based nanodevices have controlled and safe bioprofiles.

One particularly well-studied class of nanomedicine-relevant MNP is carbon nanotube (CNT). These cylindrical structures entirely composed of carbon are considered the quintessential nanomaterial. A varied list of properties and morphologies makes CNTs useful in many applications, especially when it comes to their mechanical strength, high thermal conductivity, optical properties, and outstanding field emission properties. As CNTs are increasingly considered in next-generation microelectronics (e.g., CNTs used as key components for the upcoming generation of 3-D microprocessors), they are also seen as a promising platform for carrying and delivering drugs in the human body [13–15]. However, up to this date, the lack of biocompatible CNTs has dramatically slowed down their development as devices for nanomedicine.

With MNPs being used in medicine, cosmetics, clothing, food, and even goods for children, it is of high importance to study and understand whether and how exposure to these highly diverse nanoparticles could impact their environment as well as human health. Are the unique properties of MNPs a potential source of short-term and/or long-term toxicity [6] for living organisms? Indeed, one important drawback of MNPs (and in particular CNTs) is their known toxicity potential due to their complex (and mostly unknown) bioprofiles as shown in various assays, cell lines, and organisms [16, 17]. For instance, even though CNTs are entirely made of carbon, these MNPs are not inert. In fact, they resist biological degradation and can potentially accumulate and induce toxicity in organs like lungs [18]. Moreover, not only humans are directly concerned but the whole environment including aquatic ecosystems due to industrial waste waters [19]. Therefore, eco- and human toxicological assessments are

increasingly needed as the number and diversity of consumer-oriented MNPs continue to follow a sustained rise.

When considering the outcome of federal screening efforts [20–22] such as Toxcast and Tox21, it is rather accurate to hypothesize that testing all the MNP-including products currently available on the market following similar experimental toxicological protocols would literally and optimistically take decades and cost several billion dollars. With the skyrocketing rate at which MNPs are being generated and incorporated into everyday products, there is a strong rationale for mainly relying on computational chemistry techniques to speed up the assessment of MNPs characteristics.

In this chapter, we discuss the development and application of cheminformatics methods to analyze and assess the experimental properties of nanomaterials. We focus on a family of techniques entitled Quantitative Nanostructure—Activity Relationships (QNAR or nano-QSAR) for evaluating the biological events induced by nanomaterials based on their chemical, physical, and structural characteristics. In Sect. 14.2, we present the basic principles of QNAR modeling that employs machine-learning techniques (e.g., logistic regression, support vector machines, artificial neural network) to establish quantified links between the biological endpoint of interest (e.g., cytotoxicity, cell death, ROS production) and a selected pool of experimentally measured physical chemical properties and/or nanomaterials' characteristics computed from their structures. In Sect. 14.3, we briefly review a couple of recently published studies reporting on the QNAR modeling of datasets of nanomaterials, especially for carbon nanotubes. In Sect. 14.4, we discuss some perspectives for the use of molecular modeling on nanomaterials, especially the use of predictive molecular docking and molecular dynamics simulations in the complement of QNAR models.

14.2 Quantitative Nanostructure—Activity Relationships (QNAR)

14.2.1 Definitions and General Principles

Nanotechnology has been defined by the National Nanotechnology Initiative as "*the ability to control and restructure the matter at the atomic and molecular levels in the range of approximately 1–100 nm, and exploiting the distinct properties and phenomena at that scale as compared to those associated with single atoms or molecules or bulk behavior*" [23]. MNPs are nanoparticles that have been designed and manufactured through either top-down or bottom-up approaches, i.e., top-down approaches are processes where NPs are created from bulk materials through processes such as milling, repeated quenching, or photolithography [24, 25]; bottom-up approaches are based on molecular-sized components as starting materials and complex clusters are created through chemical reactions, nucleation,

and growth process [26]. Importantly, functionalization is the process of adding surface modifications to a material. In the case of nanomaterials, functionalization is used to modulate the bioprofiles of MNPs by decorating their surface with small molecules (see Sect. 14.3). Nanotoxicology is the field that seeks to understand how the unique characteristics and properties of nanomaterials can induce, modulate, and/or impact a potential detrimental and toxic effect in the human body and/or the environment.

The main cheminformatics method used to model MNPs is actually based on Quantitative Structure–Activity Relationships (QSAR). Based on the general principle that similar compounds should induce similar biological effects, Puzyn et al. [27] introduced the term "nano-QSAR" (which is equivalent to the term QNAR) referring to the use of QSAR models for nanoparticles. Puzyn et al. [27] proposed that "nano-QSAR" models would be capable of establishing key links between MNPs features and their biological properties. QSAR primarily relies on machine-learning algorithms (e.g., random forests, support vector machines, artificial neural networks) to generate prediction models using subsets of parameters (called *descriptors*) describing MNPs' chemical, physical, constitutional, and structural characteristics. QSAR methods have a long history of providing valuable and robust models used to identify and help designing drugs [28, 29]. For more technical details about the exact nature of a QSAR model, how to train and validate it, how to assess the domain of applicability, and how to screen a set of molecules using a QSAR model, we highly recommend the recent state-of-the-art review by Cherkasov et al. [30].

Again, Quantitative Nanostructure–Activity Relationships (QNAR) are based on the same principle as QSAR, i.e., nanomaterials with similar chemical, physical, and structural characteristics are likely to induce similar biological effects. Thus, QNAR models [29] involve the use of molecular descriptors that characterize the structures and other chemical physical properties of MNPs. Here, we should underline the fact that these descriptors can be either computed using dedicated software taking as input the chemical structures of the nanomaterials, or experimentally measured (e.g., zeta potential, size distribution) according to the same protocols and conditions. Some of the latter measured descriptors are sometimes referred as biological descriptors when MNPs are tested in a range of in vitro biochemical assays. Each QNAR model establishes quantified relationships between nanomaterials' descriptors and a particular endpoint (e.g., cytotoxicity). To do so, modelers use the exact same machine-learning techniques they employ when building more traditional QSAR models for small organic molecules [29].

14.2.2 Data Sources

Publicly available sources and repositories for MNP datasets suitable for QNAR modeling are slowly emerging. In this paragraph, we cite some of the most well-known repositories for MNP-related data:

- The DaNa$^{2.0}$ database (*Data and Knowledge on Nanomaterials*—accessible at www.nanopartikel.info/en/) incorporates key information on MNPs, exposure, uptake, and behavior in both the human body and environment. The information is organized first by the field of application of a given product, then by the type of nanoparticle.
- The Nanowerk nanomaterials database (accessible at www.nanowerk.com) currently contains almost 4000 unique nanomaterials. The information contained in this database is primarily originating from supplier information. Key physical characteristics (e.g., purity, size distribution) are available for the MNPs included in the dataset.
- The Nanodatabase (accessible at nanodb.dk) is a search engine containing a collection of 2340 different consumer nanomaterials. The collection contains information on the product category, year, type of nanomaterial, country of origin, country of production, manufacturer, waste products, and potential exposure pathways.
- Nano is a searchable database (accessible at https://nano.nature.com) of nanoscience data created by Springer. The database offers more than 200,000 curated profiles on nanomaterials and nanodevices. Each entry comes from high impact journals and patents, which are all evaluated by nanotechnology experts. Precise search tools help to categorize the structures, size, composition, properties, characterization methods, toxicity, other biological effect, synthesis methods, applications, and patent claims of these nanomaterials.
- The Nanomaterial Registry (accessible via https://nanohub.org/groups/nanomaterialregistry) contains information from many publicly available sources. This dataset contains several thousands of records, the most populated entry being for silver-based MNPs (ca. 200 records). Particles' size, size distribution, zeta potential, aggregation properties, and purity values are generally available for the records. The data quality control, MNPs' naming and description ontology, and storing protocols are state of the art by following the Nano-Tab recommendations [31, 32].
- The Nanomaterial–Biological Interactions Knowledgebase (accessible at nbi.oregonstate.edu) compiles experimental data on various types of MNPs and their effects on biological systems. The now famous weighted EZ metric scores are calculated from a panel of assays to characterize the bioprofiles of each MNP and can represent valuable biological descriptors to train nano-QSAR models [33].

14.2.3 QNAR Modeling

QNAR modeling workflow is strictly similar to the classical predictive QSAR workflow [30, 34]. Therefore, we refer the readers to these papers for more details regarding the exact procedures for training, validating, and selecting the best models using a particular set of molecules, a machine-learning technique, and one

(or several) endpoint(s) to assess. Importantly, QNAR models need to account for the unique properties arising with MNPs.

There are many features and other specificities that contribute to the diversity and uniqueness of MNPs [35]. For instance, we can underline the molecular shape (e.g., cubes, cylinders, platelets, hollow spheres), the dispersion medium (e.g., liquids, gels, solid matrix), or the surface modifications (e.g., pristine, polymer grafting, biomolecules, surface coatings). These aforementioned features should thus be characterized either by the computational descriptors or the experimentally measured properties [27]. Therefore, modelers should consider various types and combinations of descriptors (such as chemical composition, size distribution, zeta potential, agglomeration state, porosity, overall molecular shape, surface chemistry) in order to build their nano-QSAR models.

This structural diversity of MNPs makes the choice of descriptors very challenging and so critical when it comes to prediction performances and the interpretability of the QNAR models built with these descriptors [27]. For instance, the chemical descriptors used for quantum dots will likely not be the same as the ones used for functionalized multiwalled carbon nanotubes. Therefore, we proposed to distinguish two categories of nanomaterials: (1) those with different cores and surface chemistry and (2) those possessing the same core but different surface functionalization. In order to afford high-performing QNAR models, the descriptors used to characterize MNPs need to be well-chosen. In case of MNPs with different cores, descriptors need to describe the whole MNP taking into account both the cores and the surface modifiers (if any). In the case of a mono-core with different functionalization, the descriptors can either describe the whole MNP or simply the surface modifiers. In the latter case, the nano-QSAR model is simply a traditional QSAR model of the surface decorators.

The chemical data curation workflow we published [36, 37] is also critical for MNPs. While each dataset of MNPs should theoretically undergo its own customized curation procedure, curation should always involve the removal of certain records (e.g., mixtures), structural cleaning (e.g., neutralization, removal of counterions), normalization of specific chemotypes, treatment of tautomeric forms, analysis and removal of duplicates, and a final manual inspection of the curated dataset of structures.

Once the chemical datasets has been compiled and curated, a set of chemical descriptors needs to be obtained for each MNP. Among the computed descriptors, we can mention: 0D/1D descriptors are a single number parameter usually referring to a global property of the MNP (e.g., presence of carbon atoms, average molecular weight, number of oxygen atoms). 2-D descriptors are computed from the 2-D molecular representations of the MNPs or its surface modifier. 2-D descriptors traditionally encompass molecular fragments, topological indices, and other graph-derived parameters [30]. 3-D descriptors refer to parameters and indices computed from the three-dimensional structure of the MNP and/or its surface decorators [30]. Quantum descriptors computed from semiempirical or ab initio quantum chemistry software are also very useful for characterizing the distinct properties of MNPs [30].

Similarly to classical QSAR models, QNAR can be either continuous or classification models. Classification QNAR models relate the predictor variables to a categorical value (binary or multi-class) of the response variable, while regression models have continuous response variables. QNAR models can not only be utilized to forecast the properties of MNPs, but also to understand which functional groups, physical characteristics, or quantum parameters have significant effects in the modulation of those experimental properties. This is crucial when it comes to the rational design of new MNPs.

14.3 Recent Case Studies of Nano-QSAR Modeling

Gajewicz et al. [38] developed nano-QSAR models for a set of 18 metal oxide nanoparticles to evaluate their toxicity on cells. Quantum descriptors used in this study included ΔH_f^c (related to the band gap width) and X^c (related to Fermi level of the oxide). The authors showed these computer-selected descriptors could help the understanding of the mechanism of action for these MNPs when tested against HaCat keratinocyte human cells and *Escherichia coli* cell lines.

Interestingly, this paper echoes another study [39] published in Nanotoxicology regarding the nano-QSAR models based on a set of 70 oxide nanoparticles and their oxidative stress potential. The authors showed those MNPs were capable of inducing oxidative stress in vitro because of their specific band energy characteristics similar to redox potentials of antioxidants or radical formation reactions.

Recently, Puzyn and coworkers [40, 41] further demonstrated the usefulness of characterizing the physicochemical features of metal oxide nanoparticles, especially for predicting the zeta potential of such MNPs, a determining factor of their aggregation properties and ultimate behavior once released in the environment.

As illustrated by the three aforementioned examples, the compendium of studies regarding the development of QNAR and nano-QSAR models is growing fast. We recommend reading the excellent reviews by Kar et al. [42], Puzyn et al. [27, 43], or Tantra et al. [44] One should also note the growing interest in the modeling of gold-based nanomaterials [45].

In this mini-chapter, we specially focus on the papers reporting on QNAR models for carbon nanotubes (CNT) as the current knowledge is still limited regarding CNTs' in vivo bioprofiles and potential induced toxicity. Indeed, a significant range of negative effects caused by pristine CNTs has been reported in the literature for various assays, cell lines, and organisms [16, 17]. Interestingly, recent studies have proven these detrimental effects can be noticeably reduced by functionalizing CNTs' surface with organic molecules [46]. Therefore, the rational design of CNTs' surface chemistry could lead to safe and controlled bioprofiles for industrial CNTs, and obviously a more appropriate biocompatibility for CNTs potentially relevant for medical applications.

To allow a safe and optimal use of CNTs in medical applications, CNTs can be functionalized to optimize blood circulation and biocompatibility [47].

On a practical point of view, the nanotube surface is drastically modified by adding biocompatible organic compounds. These surface modifications can lead to improved interactions with biological components, altered homeostasis, and improved permeability with plasma membranes [48].

In 2006, Dumortier et al. [49] studied the effects of functionalized carbon nanotubes (f-CNTs) on immune system cells. T lymphocytes, B lymphocytes, and macrophages were shown to take up four different types of f-CNTs and the authors did not observe any detrimental effects on cell viability.

In another study, Liu et al. [50] studied the biodistribution of radio-labeled f-CNTs in mice by in vivo positron emission tomography and Raman spectroscopy. They found that CNT functionalized with poly ethylene-glycol phospholipids were stable in vivo. These f-CNTs exhibited relatively long blood circulation times with a half-life of $\sim$2 h for longer phospholipid chains, and $\sim$0.5 h for the shortest chains. Also, the authors showed these f-CNTs exhibited surprisingly low uptake into the liver and kidneys. Additionally, these f-CNTs efficiently showed significant uptakes in various tumor types giving a strong indication of their potential abilities in cancer nanomedicine. In a follow-up publication, Liu et al. [51] increased blood circulation to 24 h and showed complete clearance of CNTs from major organ systems in about two months. Since f-CNTs were detected in feces, kidneys, and bladder, the study of Liu et al. [51] demonstrated that clearance is possible via the biliary and renal pathways.

In the context of structure–toxicity relationships, Sayes et al. [52] analyzed the effects of the density of functionalization for f-CNTs. The authors showed that the type and density of functionalization were correlated with CNTs' cytotoxicity observed in cultures of human dermal fibroblasts. Interestingly, the authors found that when the degree of sidewall functionalization increases, the cytotoxicity induced by the f-CNTs decreases. These results illustrate how CNTs can be rendered less toxic to cells. A complementary study by Chen et al. [53] showed that polymer-coated CNTs could be better interfaced with living cells.

As shown by the aforementioned studies, the experimental testing of pristine and functionalized CNTs is well underway and has now led to more knowledge on the actual bioprofiles of some compounds. However, are these preliminary results sufficient to start building predictive models to assess the induced effects of new CNTs? Or help in designing CNTs with the desired biocompatibility?

There are many challenges in developing and using computational chemistry methods to evaluate the biological effects induced by CNTs. The ultimate objective is to develop techniques that are effective and accurate in identifying biological effects (harmful or beneficial) for the various forms of CNTs being synthesized. The current experimental methods to characterize CNTs experimentally are expensive and time-consuming. Therefore, high-throughput computational methods with the ability to assess biological outcomes for hundreds of thousands of virtual f-CNTs in a time-effective manner would dramatically reduce overhead cost and allow for the exploration of the chemical space of f-CNTs.

Modern cheminformatics methods such as QNAR and nano-QSAR models utilize biological and chemical data including physical and geometric properties in order to

create statistically significant and externally predictive models capable of accurately forecasting the adverse and therapeutic biological effects of f-CNTs [29]. Although interest in the predictive modeling of MNPs is increasing in the computational chemistry community, literature specifically related to the cheminformatics modeling of f-CNTs for the purpose of understanding and forecasting their biological activity is still scarce. Below are recapitulated some of the most recent studies:

Monajjemi and Mollaamin [54] conducted molecular dynamics simulations of f-CNTs in different solvents. The functionalization was based on the anti-cancer drug cisplatin. The authors explored the potential of those f-CNTs for drug delivery.

Puzyn et al. [43] developed a robust nano-QSAR model based on ensemble learning regression methods in order to predict the biological effects of diverse nanomaterials including CNTs. Based on CDK molecular descriptors, these models were tested against *in vitro* assays to determine their reliability and underwent fivefold cross validation. The prediction performances of the nano-QSAR model for f-CNTs were as high as $R^2 = 0.922$ for the full set.

Fourches et al. [29, 55, 56] developed a series of QNAR models based on a dataset of 83 f-CNTs tested in in vitro toxicological assays. Four protein-binding assays (bovine serum albumin, carbonic anhydrase, chymotrypsin, and haemoglobin) were conducted as well as acute and immune toxicity assays. External prediction accuracy of the QNAR models based on 2-D descriptors and support vector machines were shown to be as high as 74% ($n = 73$, sensitivity = 79%, specificity = 69%) for the cytotoxicity models. Protein-binding classification QNAR models afforded 77% external prediction accuracy. Importantly, these models were used to screen a large library of 240,000 potential surface modifiers. The modifiers predicted to lead to f-CNTs with low toxicity and low protein affinity were identified and recommended for experimental synthesis. Ten putatively active and 10 putatively inactive CNTs were synthesized and tested. We found that all 10 putatively inactive and 7 of 10 (*6 of 10*) putatively active CNTs were confirmed in the protein-binding (*cytotoxicity*) assay. These results suggested that QNAR models can be employed for predicting biological activity profiles of novel nanomaterials, and prioritizing the design and manufacturing of nanomaterials toward better and safer products.

New studies are under way to build QNAR models for even larger sets of f-CNTs tested in more diverse assays. The progress of machine-learning techniques (e.g., deep learning) will enable the prediction performances to afford higher levels of accuracy and allow the rational design of f-CNTs with perfectly controlled bioprofiles. This is the only way to achieve a fast and robust screening of millions of hypothetical f-CNTs and to prioritize the experimental testing to the most interesting compounds.

14.4 Perspectives

As modelers, we should recognize that the current QNAR modeling technology is still in its infancy. In that regard, there are multiple ways QNAR models will evolve in the coming years. Not too surprisingly, these evolutions will mostly follow the

directions taken by the more traditional QSAR models for small organic molecules, and we describe those potential evolutions in the section below. Not only the actual prediction performances of these models will improve, but more generally the overall robustness and interpretability of cheminformatics methods we rely on when it comes to the modeling of nanomaterials.

However, it is important to emphasize the fact that the QNAR modeling field has completely different types of challenges comparing to the QSAR modeling field. The latter deals with very large sets of well-defined and characterized molecules tested against multiple and diverse of biochemical and cell-based assays. There is no lack of small molecule datasets to be modeled by QSAR models due to the efforts of the research community to deposit and maintain experimental data in freely accessible online repositories (such as PubChem, ChEMBL, or Chemspider). On the contrary, the field of QNAR modeling faces a severe paucity of nanomaterial datasets available in the public domain, and this lack of experimental data limits the type, quality, and applicability domain of the current generation of QNAR models.

Below, we underline several approaches regarding the future evolution of QNAR and nano-QSAR modeling:

- **Consensus QNAR models**: the vast majority of QNAR and nano-QSAR modeling studies rely on the use of one single type of machine-learning technique and one type of chemical descriptors per study. However, the benefits of using a collection of independent models based on various learning algorithms and chemical descriptors have been shown and established in several key community benchmarks [57, 58]. Therefore, it is likely that future QNAR and nano-QSAR models will, in fact, be consensus models, i.e., *an* ensemble of individual models averaging their predictions to assess the bioactivity of a given compound. The averaging procedure can be complex with a weighting scheme based on models' individual characteristics and/or performances. The two main advantages of using a consensus QNAR models are the gain of prediction performances and an assessment of the models' concordance allowing a better estimation of the prediction reliability for a particular compound. Importantly, methods like read-across [59, 60] can also be considered for taking part in such types of consensus models.
- **Use of biological descriptors**: hybrid QSAR models involving both computed molecular descriptors and experimentally measured biological properties have been shown to afford higher prediction performances [30, 61–63]. For instance, concentration-response curves can be used as descriptors in a QSAR model [64]. Datasets including hundreds or thousands of chemicals fully tested against tens or hundreds of biological assays are still rare in the public domain, but one could note several recent examples of such screening efforts [21, 65]. Due to the critical lack of MNP-related data in the public domain, it is way more difficult to obtain and use biological descriptors for training a QNAR model. The largest datasets of MNPs in the public domain contain ca. 150 compounds and are generally tested in one single biological assay. Therefore, the use of biological

descriptors for QNAR models will certainly intensify as soon as more experimental screens are conducted for larger sets of MNPs.

- **Mixtures of nanomaterials:** Slurries are stable suspensions of abrasive nanomaterials (e.g., alumina, silica, and/or ceria) dispersed in water with other chemicals. They are notably employed for polishing tasks in the manufacturing of semi-conductors (e.g., chemical mechanical planarization process). These types of suspensions represent a real challenge for molecular modelers as they involve complex mixtures of different nanomaterials and various organic molecules in solution. As their bioprofiles are difficult to assess, it is likely more cheminformatics studies will be done to create simpler "*model systems*" for attempting the modeling of such types of mixtures of nanomaterials.
- **Quantum mechanics**: QM represents one of the cornerstones of computational chemistry. As shown by the work of Puzyn and coworkers [40], QM-based calculations are essential in enabling the characterization of the electronic, physical, and chemical properties of nanomaterials. With the development of new DFT functionals adapted for subtypes of nanomaterials, the use of QM calculations for computing MNPs' descriptors in order to build QNAR models will skyrocket in the coming years. This is especially true for challenging series of analogous MNPs with subtle structural variations, for which QM-based descriptors will help in discriminating.
- **Molecular docking**: Three-dimensional molecular docking [66] is a popular technique used for screening large libraries of molecules in drug discovery [67, 68]. Docking allows modelers to forecast the binding mode of small molecule ligands in the active site of a biological target (e.g., protein, DNA). Obviously, the 3-D structure of the target is needed. Molecular docking not only predicts the binding mode of the ligand but also scores the actual molecular interactions to estimate the free energy of binding. These docking scores can be used to rank ligands in virtual screening studies [69, 70]. These methods can be applied to estimate the binding modes of carbon nanotubes with small proteins. Since the scoring functions used by molecular docking programs have not been designed and trained for that purpose, the docking of f-CNTs is still very prototypical and not ready for reliable virtual screening.
- **Molecular dynamics**: Another computational methodology utilized in cheminformatics is molecular dynamic simulations (MDS) [71]. MDS allows modelers to simulate the dynamic motions of molecules by solving Newton's equations of motions for every single atom in the system. A force field [72] is used to compute all intra- and intermolecular forces so that the full-atom system with the explicit solvent can evolve over several hundreds of nanoseconds of biological time in a "realistic" manner. There are several examples of MDS for nanomaterials and CNTs [73–76]. These simulations can give clear insight into the intermolecular forces of f-CNTs interacting with biological targets in an explicit solvent as well as the effects of different functionalization. In the future, these MDS trajectories are likely to be used in complement to nano-QSAR models.

Overall, QNAR modeling represents a reliable and potentially disrupting way we assess the properties of nanomaterials. This is obviously critical from a chemical risk assessment standpoint and thus for regulators that need to fully evaluate the potentially detrimental effects induced by a given MNP in a particular organism. But QNAR modeling is also essential for enabling the rational design of new MNPs with a defined list of characteristics and controlled bioprofiles. In fact, developing reliable QNAR models to help identifying those highly valuable MNPs is critical for the future of nanotechnology. Nanomedicine-oriented MNPs are requiring enormous amounts of costly and time-consuming rounds of structural optimization to make them efficient and safe *by design*. Thus, we posit that any new computational technique enabling or facilitating that MNP design process is relevant and worth investigating as part of establishing the future toolbox of next-generation chemists.

References

1. Drexler KE (2004) Nanotechnology: from feynman to funding. Bull Sci Technol Soc 24:21–27
2. Gogotsi Y, Presser V (2014) Carbon nanomaterials, 2nd edn. Taylor & Francis Group LLC, Abingdon
3. Helpman E (1998) General purpose technologies and economic growth. MIT Press, Cambridge
4. Berube D (2006) Nano-hype: the truth behind the nanotechnology buzz. Prometheus Books, Amherst, N.Y. ISBN 978-1-59102-351-7
5. Roco MC (2011) The long view of nanotechnology development: the national nanotechnology initiative at 10 Years. In: Nanotechnology research directions for societal needs in 2020. Science Policy Reports, vol 1. Springer, Dordrecht. ISBN 978-94-007-1167-9
6. Vance ME, Kuiken T, Vejerano EP, McGinnis SP, Hochella MF, Rejeski D, Hull MS (2015) Nanotechnology in the real world: redeveloping the nanomaterial consumer products inventory. Beilstein J. Nanotechnol 6:1769–1780
7. Li Q, Mahendra S, Lyon DY, Brunet L, Liga MV, Li D, Alvarez PJJ (2008) Antimicrobial nanomaterials for water disinfection and microbial control: potential applications and implications. Water Res 42:4591–4602
8. Wang J (2005) Carbon-nanotube based electrochemical biosensors: a review. Electroanalysis 17:7–14
9. Salata O (2004) Applications of nanoparticles in biology and medicine. J Nanobiotechnol 2:3
10. Crookes WJ, Ding LL, Huang QL, Kimbell JR, Horwitz J, McFall-Ngai MJ (2004) Reflectins: the unusual proteins of squid reflective tissues. Science 303(5655):235–238
11. Kim JJ, Lee Y, Kim HG, Choi KJ, Kweon HS, Park S, Jeong KH (2012) Biologically inspired LED lens from cuticular nanostructures of firefly lantern. Proc Natl Acad Sci USA 109:18674–18678
12. Reguera G, McCarthy KD, Mehta T, Nicoll JS, Tuominen MT, Lovley DR (2005) Extracellular electron transfer via microbial nanowires. Nature 435:1098–1101
13. Franklin AD (2013) Electronics: the road to carbon nanotube transistors. Nature 498:443–444
14. Singh RK, Patel KD, Kim JJ, Kim TH, Kim JH, Shin US, Lee EJ, Knowles JC, Kim HW (2014) Multifunctional hybrid nanocarrier: magnetic CNTs ensheathed with mesoporous silica for drug delivery and imaging system. ACS Appl Mater Interfaces 6:2201–2208

15. Song W, Zheng Z, Tang W, Wang X (2007) A facile approach to covalently functionalized carbon nanotubes with biocompatible polymer. Polymer (Guildf) 48:3658–3663
16. Kumar A, Kumar P, Anandan A, Fernandes TF, Ayoko GA, Biskos G (2014) Engineered nanomaterials: knowledge gaps in fate, exposure, toxicity, and future directions. J Nanomater 2014:1–16
17. Qu G, Bai Y, Zhang Y, Jia Q, Zhang W, Yan B (2009) The effect of multiwalled carbon nanotube agglomeration on their accumulation in and damage to organs in mice. Carbon 48:2060–2069
18. Ryman-Rasmussen JP, Cesta MF, Brody AR, Shipley-Phillips JK, Everitt JI, Tewksbury EW, Moss OR, Wong BA, Dodd DE, Andersen ME, Bonner JC (2009) Inhaled carbon nanotubes reach the subpleural tissue in mice. Nat Nanotechnol 4:747–751
19. Moore MN (2006) Do nanoparticles present ecotoxicological risks for the health of the aquatic environment? Environ Int 32:967–976
20. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95:5–12
21. Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM, Houck KA, Dix DJ, Kavlock RJ, Knudsen TB (2013) Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. Chem Res Toxicol 26:878–895
22. Thomas RS, Black MB, Li L, Healy E, Chu TM, Bao W, Andersen ME, Wolfinger RD (2012) A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. Toxicol Sci 128:398–417
23. Roco M, Chen H (2009) Mapping nanotechnology innovations and knowledge. Springer, Boston
24. Niemeyer CM (2001) Nanoparticles, proteins, and nucleic acids: biotechnology meets materials science. Angew Chemie Int Ed 40:4128–4158
25. Ju-Nam Y, Lead JR (2008) Manufactured nanoparticles: an overview of their chemistry, interactions and potential environmental implications. Sci Total Environ 400:396–414
26. Rotello V (2004) Nanoparticles: building blocks for nanotechnology. Nanostructure Science and Technology. Springer US. ISBN 978-1-4613-4770-5
27. Puzyn T, Leszczynska D, Leszczynski J (2009) Toward the development of "nano-QSARs": advances and challenges. Small 5:2494–2509
28. Verhaar HJM, van Leeuwen CJ, Hermens JLM (1992) Classifying environmental pollutants. Chemosphere 25:471–491
29. Fourches D, Pu D, Tropsha A (2011) Exploring quantitative nanostructure-activity relationships (QNAR) modeling as a tool for predicting biological effects of manufactured nanoparticles. Comb Chem High Throughput Screen 14:217–225
30. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57:4977–5010
31. Thomas DG, Pappu RV, Baker NA (2011) NanoParticle Ontology for cancer nanotechnology research. J Biomed Inform 44:59–74
32. Thomas DG, Klaessig F, Harper SL, Fritts M, Hoover MD, Gaheen S, Stokes TH, Reznik-Zellen R, Freund ET, Klemm JD, Paik DS, Baker NA (2011) Informatics and standards for nanomedicine technology. Wiley Interdiscip Rev Nanomed Nanobiotechnol 3(5):511
33. Harper B, Thomas D, Chikkagoudar S, Baker N, Tang K, Heredia-Langner A, Lins R, Harper S (2015) Comparative hazard analysis and toxicological modeling of diverse nanomaterials using the embryonic zebrafish (EZ) metric of toxicity. J Nanopart Res 17:250
34. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inform 29:476–488
35. Nagarajan R (2008) Nanoparticles: building blocks for nanotechnology. ACS Symp Ser 996:2–14

36. Fourches D, Muratov E, Tropsha A (2010) Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J Chem Inf Model 50:1189–1204
37. Fourches D, Muratov E, Tropsha A (2015) Curation of chemogenomics data. Nat Chem Biol 11:535
38. Gajewicz A, Schaeublin N, Rasulev B, Hussain S, Leszczynska D, Puzyn T, Leszczynski J (2014) Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: hints from nano-QSAR studies. Nanotoxicology 1–13
39. Burello E, Worth AP (2011) A theoretical framework for predicting the oxidative stress potential of oxide nanoparticles. Nanotoxicology 5:228–235
40. Mikolajczyk A, Gajewicz A, Rasulev B, Schaeublin N, Maurer-Gardner E, Hussain S, Leszczynski J, Puzyn T (2015) Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. Chem Mater 27:2400–2407
41. Wyrzykowska E, Mikolajczyk A, Sikorska C, Puzyn T (2016) Development of a novel in silico model of zeta potential for metal oxide nanoparticles: a nano-QSPR approach. Nanotechnology 27:445702
42. Kar S, Gajewicz A, Puzyn T, Roy K (2014) Nano-quantitative structure-activity relationship modeling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells. Toxicol In Vitro 28:600–606
43. Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari TP, Michalkova A, Hwang HM, Toropov A, Leszczynska D, Leszczynski J (2011) Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nat Nanotechnol 6:175–178
44. Tantra R, Oksel C, Puzyn T, Wang J, Robinson KN, Wang XZ, Ma CY, Wilkins T (2015) Nano(Q)SAR: challenges, pitfalls and perspectives. Nanotoxicology. 9:636–642
45. Bigdeli A, Hormozi-Nezhad MR, Parastar H, Rother J, Sunnick E, Rosman C, Pierrat S, Sönnichsen C, Wegener J, Janshoff A, Sperling R, Schmid G, Simon U, Parak WJ, Semmler-Behnke M (2015) Using nano-QSAR to determine the most responsible factor(s) in gold nanoparticle exocytosis. RSC Adv 5:57030–57037
46. Zhou H, Mu Q, Gao N, Liu A, Xing Y, Gao S, Zhang Q, Qu G, Chen Y, Liu G, Zhang B, Yan B (2008) A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. Nano Lett 8:859–865
47. Singh R, Pantarotto D, Lacerda L, Pastorin G, Klumpp C, Prato M, Bianco A, Kostarelos K (2006) Tissue biodistribution and blood clearance rates of intravenously administered carbon nanotube radiotracers. Proc Natl Acad Sci USA 103:3357–3362
48. Vardharajula S, Ali SZ, Tiwari PM, Eroğlu E, Vig K, Dennis VA, Singh SR (2012) Functionalized carbon nanotubes: biomedical applications. Int J Nanomed 7:5361–5374
49. Dumortier H, Lacotte S, Pastorin G, Marega R, Wu W, Bonifazi D, Briand JP, Prato M, Muller S, Bianco A (2006) Functionalized carbon nanotubes are non-cytotoxic and preserve the functionality of primary immune cells. Nano Lett 6:1522–1528
50. Liu Z, Cai W, He L, Nakayama N, Chen K, Sun X, Chen X, Dai H (2007) In vivo biodistribution and highly efficient tumour targeting of carbon nanotubes in mice. Nat Nanotechnol 2:47–52
51. Liu Z, Davis C, Cai W, He L, Chen X, Dai H (2008) Circulation and long-term fate of functionalized, biocompatible single-walled carbon nanotubes in mice probed by Raman spectroscopy. Proc Natl Acad Sci USA 105:1410–1415
52. Sayes CM, Liang F, Hudson JL, Mendez J, Guo W, Beach JM, Moore VC, Doyle CD, West JL, Billups WE, Ausman KD, Colvin VL (2006) Functionalization density dependence of single-walled carbon nanotubes cytotoxicity in vitro. Toxicol Lett 161:135–142
53. Chen X, Tam UC, Czlapinski JL, Lee GS, Rabuka D, Zettl A, Bertozzi CR (2006) Interfacing carbon nanotubes with living cells. J Am Chem Soc 128:6292–6293
54. Monajjemi M, Mollaamin F (2011) Molecular modeling study of drug-DNA combined to single walled carbon nanotube. J Clust Sci 23:259–272

55. Mu Q, Jiang G, Chen L, Zhou H, Fourches D, Tropsha A, Yan B (2014) Chemical basis of interactions between engineered nanoparticles and biological systems. Chem Rev 114 (15):7740–7781
56. Fourches D, Pu D, Tassa C, Weissleder R, Shaw SY, Mumper RJ, Tropsha A (2010) Quantitative nanostructure-activity relationship modeling. ACS Nano 4:5703–5712
57. Zhu H, Tropsha A, Fourches D, Varnek A, Papa E, Gramatica P, Oberg T, Dao P, Cherkasov A, Tetko IV (2008) Combinatorial QSAR modeling of chemical toxicants tested against *Tetrahymena pyriformis*. J Chem Inf Model 48:766–784
58. Sushko I, Novotarskyi S, Körner R, Pandey AK, Cherkasov A, Li J, Gramatica P, Hansen K, Schroeter T, Müller KR, Xi L, Liu H, Yao X, Öberg T, Hormozdiari F, Dao P, Sahinalp C, Todeschini R, Polishchuk P, Artemenko A, Kuz'min V, Martin TM, Young DM, Fourches D, Muratov E, Tropsha A, Baskin I, Horvath D, Marcou G, Muller C, Varnek A, Prokopenko VV, Tetko IV (2010) Applicability domains for classification problems: benchmarking of distance to models for Ames mutagenicity set. J Chem Inf Model 50:2094–2111
59. Gajewicz A, Cronin MTD, Rasulev B, Leszczynski J, Puzyn T (2015) Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: nano-read-across. Nanotechnology. 26:15701
60. Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, Tropsha A (2013) Integrative chemical-biological read-across approach for chemical hazard classification. Chem Res Toxicol 26:1199–1208
61. Rusyn I, Sedykh A, Low Y, Guyton KZ, Tropsha A (2012) Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. Toxicol Sci 127:1–9
62. Low Y, Uehara T, Minowa Y, Yamada H, Ohno Y, Urushidani T, Sedykh A, Muratov E, Kuz'min V, Fourches D, Zhu H, Rusyn I, Tropsha A (2011) Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. Chem Res Toxicol 24:1251–1262
63. Zhu XW, Sedykh A, Liu SS (2013) Hybrid in silico models for drug-induced liver injury using chemical descriptors and in vitro cell-imaging information. J Appl Toxicol 34(3): 281–288
64. Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A (2011) Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect 119:364–370
65. Elkins JM, Fedele V, Szklarz M, Abdul Azeez KR, Salah E, Mikolajczyk J, Romanov S, Sepetov N, Huang XP, Roth BL, Al Haj Zen A, Fourches D, Muratov E, Tropsha A, Morris J, Teicher BA, Kunkel M, Polley E, Lackey KE, Atkinson FL, Overington JP, Bamborough P, Müller S, Price DJ, Willson TM, Drewry DH, Knapp S, Zuercher WJ (2015) Comprehensive characterization of the published kinase inhibitor set. Nat Biotechnol 34:95–103
66. Sousa SF, Fernandes PA, Ramos MJ (2006) Protein-ligand docking: current status and future challenges. Proteins. 65:15–26
67. Shoichet BK (2004) Virtual screening of chemical libraries. Nature 432:862–865
68. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949
69. Damm-Ganamet KL, Smith RD, Dunbar JB, Stuckey JA, Carlson HA (2013) CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series. J Chem Inf Model 53:1853–1870
70. Fourches D, Muratov E, Ding F, Dokholyan NV, Tropsha A (2013) Predicting binding affinity of CSAR ligands using both structure-based and ligand-based approaches. J Chem Inf Model 53:1915–1922
71. Mortier J, Rakers C, Bermudez M, Murgueitio MS, Riniker S, Wolber G (2015) The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. Drug Discov Today 20:686–702

72. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput 6:1509–1519
73. Oliveira OV (2009) Atomistic molecular dynamics simulation of the CeO_2 nanoparticle aggregation. International conference on advanced materials, Rio de Janeiro, Brazil
74. Argyris D, Ho T, Cole DR, Striolo A (2011) Molecular dynamics studies of interfacial water at the alumina surface. J Phys Chem C 115:2038–2046
75. Sayle TXT, Molinari M, Das S, Bhatta UM, Möbus G, Parker SC, Seal S, Sayle DC (2013) Environment-mediated structure, surface redox activity and reactivity of ceria nanoparticles. Nanoscale 5:6063–6073
76. Proctor EA, Yin S, Tropsha A, Dokholyan NV (2012) Discrete molecular dynamics distinguishes nativelike binding poses from decoys in difficult targets. Biophys J 102:144–151