



HAL
open science

A computational view on nanomaterial intrinsic and extrinsic features for nanosafety and sustainability

Giulia Mancardi, Alicja Mikolajczyk, Vigneshwari K Annapoorani, Aileen Bahl, Kostas Blekos, Jaanus Burk, Yarkin A Çetin, Konstantinos Chairetakis, Sutapa Dutta, Laura Escorihuela, et al.

► To cite this version:

Giulia Mancardi, Alicja Mikolajczyk, Vigneshwari K Annapoorani, Aileen Bahl, Kostas Blekos, et al.. A computational view on nanomaterial intrinsic and extrinsic features for nanosafety and sustainability. *Materials Today*, 2023, 67, pp.344-370. 10.1016/j.mattod.2023.05.029 . hal-04227136

HAL Id: hal-04227136

<https://hal.science/hal-04227136v1>

Submitted on 3 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A computational view on nanomaterial intrinsic and extrinsic features for nanosafety assessment

Giulia Mancardi^{a,1}, Alicja Mikolajczyk^{b,2}, Vigneshwari K. Annapoorani^c,
Aileen Bahl^d, Kostas Blekos^e, Jaanus Burk^f, Yarkin A. Çetin^g,
Konstantinos Chairetakis^e, Sutapa Dutta^c, Laura Escorihuela^g, Karolina
Jagiello^b, Teboho Mokhena^h, Ankush Singhalⁱ, Rianne van der Polⁱ, Miguel
A. Bañares^j, Nicolae-Viorel Buchete^c, Monica Calatayud^k, Veronica I.
Dumit^d, Davide Gardini^l, Nina Jeliaskova^m, Andrea Haase^d, Regina
Maphangaⁿ, Effie Marcoulaki^e, Benjamí Martorell^g, Tomasz Puzyn^b, G.J.
Agur Sevinkⁱ, Felice C. Simeone^l, Kaido Tamm^f, Eliodoro Chiavazzo^{a,3}

^a*Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Italy*

^b*Laboratory of Environmental Chemometrics, Institute for Environmental and Human Health Protection, Faculty of Chemistry, University of Gdansk, 80-308 Gdansk, Poland*

^c*School of Physics, University College Dublin, Belfield, Dublin 4, Ireland*

^d*German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany*

^e*Institute of Nuclear & Radiological Sciences & Technology, Energy & Safety, National Centre for Scientific Research «Demokritos», Athens, Greece*

^f*Institute of Chemistry, University of Tartu, Ravila 14a, 50411, Tartu, Estonia*

^g*Departament d'Enginyeria Química, Universitat Rovira i Virgili, Avinguda dels Països Catalans, 26, 43007 Tarragona, Spain*

^h*Advanced Materials Division, Nanotechnology Innovation Center (NIC), Randburg 2194, South Africa*

ⁱ*Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands*

^j*Institute for Catalysis, ICP-CSIC, Marie Curie 2, E-28049, Madrid, Spain*

^k*Laboratoire de Chimie Théorique, CNRS, Sorbonne Université, 4 Place Jussieu, 75005, Paris, France*

^l*ISTEC-CNR, Institute of Science and Technology for Ceramics - National Research Council of Italy, Faenza, Italy*

^m*Ideaconsult Ltd, Sofia, Bulgaria*

ⁿ*Next Generation Enterprises and Institutions, Council for Science and Industrial Research, P.O. Box 395, Pretoria 0001, South Africa*

¹Corresponding Author email: giulia.mancardi@polito.it

²Corresponding Author email: alicja.mikolajczyk@ug.edu.pl

³Corresponding Author email: eliodoro.chiavazzo@polito.it

Abstract

In recent years, an increasing number of diverse Engineered Nano-Materials (ENMs), such as nanoparticles and nanotubes, have been included in many technological applications and consumer products. However, the desirable and unique properties of ENMs are accompanied by potential hazards whose impact is not always (well) evaluated or even known. Alongside established methods for experimental and computational characterisation, physics-based modelling tools like molecular dynamics are increasingly considered for safe and sustainable by design strategies that put user safety and sustainability at the centre of the design and development of new products. Hence, the further development of such tools is stimulated by numerous regulatory agencies.

This paper stems from a community effort and present the outcome of a four-year long discussion on the benefits, capabilities and limitations of adopting physics-based modelling for computing suitable features of nanomaterials that can be used for toxicity assessment of nanomaterials in combination with data-based models and experimental assessment of toxicity endpoints. We review how modern multi-scale physics-based models can generate advanced system-dependent (intrinsic) or time-dependent (extrinsic) descriptors/features of ENMs, with the former being related to the bare nanoparticle and the latter to the dynamic fingerprint upon entering biological media. The focus is on: (i) effectively representing all nanoparticle attributes for multicomponent nanomaterials, (ii) generation and inclusion of intrinsic nanoform properties, (iii) inclusion of selected extrinsic properties, (iv) the necessity of considering a distribution of structural advanced features rather than an average, (v) the implementation of adverse outcome pathways methodology into nanoinformatics. Such a review enables us to identify and highlight a number of key challenges associated with ENMs data generation, curation, representation and use within machine learning or other data based models. Finally, the set up of dedicated databases as well as the development of grouping and read-across strategies based on the mode of action of ENMs using omics methods are identified as emerging methodologies for safety assessment and reduction of animal testing.

Keywords: nanoinformatics, nanosafety, engineered nanomaterials, physicochemical descriptors, safe and sustainable by design, materials modeling, machine learning, grouping approaches, multiscale modeling

Contents

1	Introduction	6
2	Safe-and-sustainable-by-Design (SSbD) strategy: From nanoform description to safety modeling	9
2.1	The safety and sustainability assessment	9
2.2	Regulatory actions to drive SSbD chemicals and materials . .	10
2.3	Hazards of ENM	11
2.4	The role of materials modeling in NM hazard assessment . . .	13
3	Data-based models for for linking ENM features to safety	15
3.1	The critical role of a general representation of ENMs	15
3.2	Data-based models for safety assessment	16
3.2.1	System-independent intrinsic features	17
3.2.2	System-dependent extrinsic features	18
3.3	Modern machine-learning methods	21
3.3.1	Supervised learning	23
3.3.2	Unsupervised learning	25
4	Physics-based models for nanostructure characterization	31
4.1	Intrinsic advanced descriptors: Electronic level	32
4.2	Intrinsic advanced descriptors: Atomistic and Molecular level .	34
4.3	Moving from intrinsic to extrinsic descriptors	36
4.4	Extrinsic advanced descriptors: Mesoscopic level	38
5	Discussion: Challenges and perspectives	49
5.1	Databases: The eNanoMapper case study	51
5.2	Grouping approaches	52
6	Concluding remarks	57
7	Acknowledgements	58
8	Data availability	58
9	Conflicts of Interest	58

**Appendix A List of physics-based descriptors that can be
calculated through materials modelling techniques 86**

Nomenclature

AIMD	ab initio Molecular Dynamics	LUMO	Lowest Unoccupied Molecular Orbital
AO(P)	Adverse Outcome (Pathway)	MD	Molecular Dynamics
API	Application Programming Interface	MDReaxFF	Reactive Force Field Molecular Dynamics
BD	Brownian Dynamics	MIE	Molecular Initiating Events
CG(MD)	Coarse-Grained (Molecular Dynamics)	ML	Machine Learning
CSS	Chemicals Strategy for Sustainability	MoA	Mode of Action
DFT	Density Functional Theory	MODA	MOdelling DAta fiche
DFTB	Density Functional Tight-Binding	NAM	New Approach Methodologies
DLVO	Derjaguin-Landau-Verwey-Overbeek	PMF	Potential of Mean Force
ENM	Engineered NanoMaterials	QM	Quantum Mechanics
FAIR	Findable, Accessible, Interoperable and Reusable	QNAR	Quantitative Nanostructure-Activity Relationship
GA	Genetic Algorithm	QSAR	Quantitative Structure-Activity Relationship
GAN	Generative Adversarial Network	QSPR	Quantitative Structure-Property Relationship
HOMO	Highest Occupied Molecular Orbital	RNN	Recurrent Neural Network
hPF	hybrid Particle-Field	S(S)bD	Safe (and Sustainable) by Design
KE	Key Events	SASA	Solvent Accessible Surface Area
LCA	Life Cycle Assessment	SCFT	Self-Consistent Field Theory
LDM	Liquid Drop Model	VAE	Variational Autoencoder

1. Introduction

ENMs perform dedicated tasks in catalysis [1], medicine [2, 3], agriculture [4], food [5] and energy [6, 7] among many others, because of their exceptional characteristics. ENMs underpin new products and devices requiring control of matter at the nanometer scale, and the possibility to fine-tune their synthesis protocols results in countless variants with different physicochemical properties. The exceptional properties of ENMs - stemming from the manipulation of matter at the atomic scale - can also come with yet little-known risks to human health and the environment. In fact, the biological activity of ENMs is thought to be closely related to their physicochemical characteristics, which may be altered by the biological medium itself during the lifetime of ENMs (e.g., protein corona formation and structural modifications). Both the intrinsic and extrinsic physicochemical features of ENMs are key for recognizing and predicting hazards as well as assessing safety characteristics. The former are related to chemical composition, crystal structure, size, shape and surface structure, whereas the latter pertain to the non-trivial interaction with the environment. As such, a complete set of those features (here referred to as *advanced descriptors*) are critical in nanoinformatics for data-based model development, such as the popular quantitative nanostructure-activity/property models (QSAR/QSPR). In this respect, predictive models offer unprecedented opportunities for knowledge-based optimization and development of new ENMs improving their functionality and - at the same time - minimizing unexpected health and/or environmental risks. Preliminary *in silico* screening of possible versions of new ENMs can thus lead to optimal nanostructures with reduced hazardous characteristics even before the production stage. However, the pace of further progress in nanotechnology will critically depend on a synergistic knowledge integration of experimental evidence with data from reliable theoretical and computational models.

The combined study of materials modeling (nanostructure characterization) and predictive models for the design of safe nanostructures with desired properties (safe and sustainable by design perspective, SSbD) brings along new opportunities both in the academic and the industrial context. Nonetheless, significant challenges stemming from both the extremely demanding computational models (physics-based and data-based) and the practically unlimited number of possible combinations of different substances and nanoforms are ahead. Even for a single material, there is a plethora of possible bulk and surface structures, defects and terminations, each delivering

unique properties. This general challenge can only be tackled by an integrated approach. Here, we critically analyze the process associated with the *in silico* evaluation of the ENMs descriptors, with the aim of highlighting challenges and opportunities when using physics-based modelling for generating them.

Our discussion takes place under the broad perspective of the European and US legislation and their strategies related to risk assessment of nanomaterials (e.g., the EU Materials Modelling Council, the EU REACH regulation, US-EU nanoEHS initiative, the EU US Nanoinformatics Roadmap 2030[8]). Despite the well-accepted relationship between physicochemical properties (i.e. descriptors) and the (eco)-toxicological endpoints, a comprehensive computational description of ENMs (and understanding of the basic mechanisms behind their interaction with biological media) is still very challenging.

First, in this work we focus on some of the (physics-based) computational approaches for estimating advanced descriptors at time and space scales relevant to nanosafety. In this respect, we provide an overview on the approaches for investigating ENMs electronic and atomistic structure (thus focusing on intrinsic features) up to the mesoscopic description of interactions with biological matter, such as proteins or cell membranes (hence moving towards more extrinsic features).

Second, more recent data-based models and approaches for nanosafety assessment are analyzed. As shown in the pictorial representation reported in the top part of Figure 1, computations can be performed at different space/time scales by solving appropriate physical model equations with advanced descriptors collected from each simulator. Advanced descriptors can be passed across different time and space scales, realizing a chain of multiscale simulations. [9] This vision is certainly fascinating, however, as discussed below, it comes with formidable challenges: as opposed to the high accuracy data extracted by the physics-based models, the inherent computational cost is often prohibitively high. This leads to a low data variance that renders the subsequent translation of such data into input for QSAR/QSPR (or other data-based) models extremely difficult if not impossible. Details on obstacles at every relevant scale are discussed, and the current status and challenges with the European regulatory context on materials modeling and *in silico* estimate of descriptors for nanosafety purposes are reported below.[10]

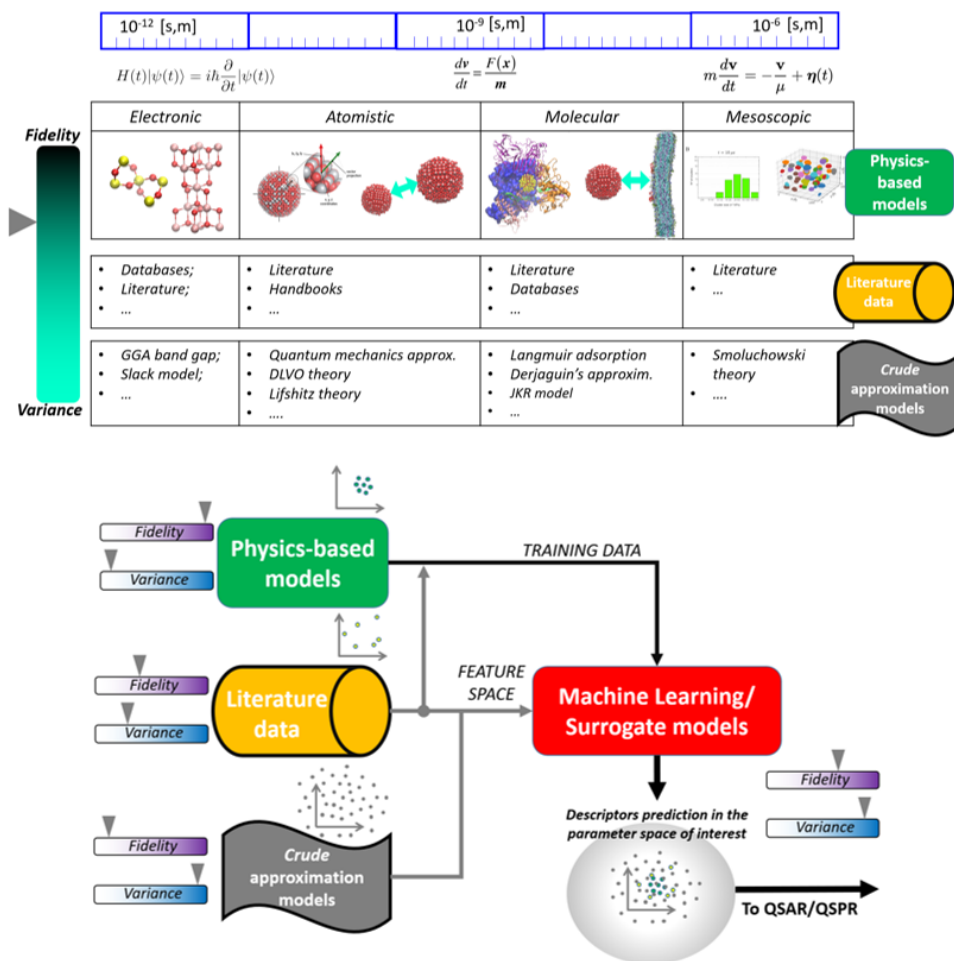


Figure 1: TOP: Physics-based models ensure high fidelity at a high computational cost, making it possible to investigate only a restricted region of the ENM design space. Hence, the necessary variance is needed to describe the large variety of parameters of interest (e.g., particle size, coating, etc.). Thus, predicting advanced descriptors for nanosafety has to stem from different sources, such as literature data and crude approximation models characterized by a lower fidelity level. BOTTOM: Discrepancy in fidelity and variance is to be properly orchestrated, possibly using ML models.[11]

2. Safe-and-sustainable-by-Design (SSbD) strategy: From nanoform description to safety modeling

Within the European Green Deal, the Chemicals Strategy for Sustainability (CSS)[12] identified several actions to reduce negative impacts on human health and the environment associated with chemicals, materials, products, and services commercialized or introduced onto the EU market. In particular, the ambition of the CSS is to phase out the most harmful substances and substitute, as far as possible, all other substances of concern and otherwise minimize their use and track them. This objective requires novel approaches to analyze and compare all life cycle stages, effects, releases, and emissions for specific chemicals, materials, products, and services, and moving towards zero pollution for air, water, soil, and biota. The SSbD framework aims to support the design and development of safe and sustainable chemicals and materials with research and innovation (R&I) activities.

Although the safety of nanomaterials has been of concern to the scientific community for more than two decades, there is still a limited number of validated and regulatory accepted alternative nano-specific approaches for assessing their human safety. In fact, the current knowledge about various adverse effects induced by nanomaterials after exposure does not yet enable a broad development of the SSbD strategy.

2.1. *The safety and sustainability assessment*

Reliable and efficient methods allowing, in a timely manner, to assess exposure risks should be first developed. In this context, the development of new alternative methods combining *in vitro*, chemical analysis as well as *in silico* models to predict the potential adverse impact of chemicals, including nanomaterials, on human health is highly needed.[13] These tools are expected to be useful for regulators and policymakers; therefore, they need to consider biologically plausible and regulatory-relevant events essential for the occurrence of possible adverse outcome.[14] In line with this point of view is the strategy that integrates nanoinformatics models with the Adverse Outcome Pathways (AOPs).[15, 16] The AOP is a framework that describes a sequence of biological events following stressor exposure and leading to various Adverse Outcomes (AO). This concept links Molecular Initiating Events (MIE) to the series of key cellular- or tissue-level changes (so-called Key Events, KE) that culminate in the manifestation of AO. By supporting the identification of AOP-relevant events and information that might be applied

to the weight of evidence safety decisions, AOP can serve as a framework for developing nanoinformatics models useful for regulatory actions. This would mean that events recognized as crucial to manifest nano-specific adverse effects should be considered for modeling. As an example, recently, Jagiello et al.[15] proposed a model that linked the structural properties of multiwalled carbon nanotubes with the recruitment of pro-inflammatory mediators into the lungs, that finally leads to lung fibrosis according to AOP173. Applying the AOP framework in this model enables to understand events triggers following nanotube exposure and occurring in different biological organizations (molecular, cellular, tissue, organ, individual). To sum up, with the development of the AOP-anchored nanoinformatics models (models for key biologically plausible and regulatory-relevant events), the usage and acceptance of those approaches in making regulatory decisions about the safety of nanomaterials exposure can increase.

Additionally, the JRC (2022)[17] reviewed previous decision frameworks for materials safety to investigate how sustainability (broadly including social and economic aspects) was considered, and define a set of criteria to integrate safety and sustainability. Sustainability assessment can use conventional sustainability metrics adapted for nanotechnology products and processes. In this context, Stieberova et al.[18] used techno-economic and life cycle environmental criteria for sustainability assessment, to compare alternative nanoparticle production technologies; García-Quintero and Palencia[19] analyzed conventional quantitative sustainability metrics and proposed Life Cycle Assessment (LCA) as a suitable metric to compare, optimize and quantify bio-based nanobiotechnology and nanosynthesis protocols.

2.2. Regulatory actions to drive SSbD chemicals and materials

Predictive *in silico* modeling, accessible and searchable databases and quantitative tools for risk assessment and prevention are critical for the successful implementation of SbD/SSbD strategies in the nanosafety context and to develop relevant protocols, reference materials, realistic *in vitro* and computational models, as well as grouping and read-across methods.[20] Additional challenges stem from the need to assess the fate and reactivity of next generations of nanomaterials[21] including smart nanomaterials and nano-enabled products,[22]. New Approach Methodologies (NAMs), comprise a wide range of such models and tools[23] to conduct robust and reliable chemical safety assessments and reduce animal testing.

With substantial funding, the European Commission has supported NAMs for nanomaterials Refs.[24, 25], acknowledging their potential in regulatory

decision-making and innovation. NAMs could increase regulatory preparedness through fit-for-purpose data and standardized tests specific to nanomaterials,[26] and their potential would be better exploited within tiered regulatory schemes.[27] Consideration of available standards (as the OECD standard for QSARs[28] and recommendations of the European Materials Modelling Council for validating *in silico* tools is essential to promote their adoption for regulatory use. On the technical side, this also enables model integration into complete computational IATA workflows.[29]

There is already a quite rich literature on the development of tools for designing ENMs with defined properties. As a representative example, preliminary works considered a conditional deep convolutional Generative Adversarial Network (GAN) using competitive learning to suggest nanophotonic structures with desired optical properties,[30] a GAN to generate crystalline porous materials, in particular pure silica zeolite structures.[31] However, since there is still much to do for understanding the behavior of ENMs, several projects have been funded by the EU in the last years[32, 33, 34, 35] to address safety and risk governance issues.

With this in mind researchers in NanoInformaTIX[36] project have developed an optimization methodology to guide the search for safer ENMs and to support the application of safe(r)-by-design (SbD) approaches. The implemented tool can be trained using available datasets and can enable the assessment of various design options generated during the optimal search. The assessment and screening processes use efficient representations coupled with quantitative tools for similarity assessment to investigate morphology-based behavior.[37]

2.3. Hazards of ENM

As mentioned above, physicochemical properties of nanomaterials may significantly differ from their bulk counterparts. Despite the beneficial technological consequences associated with this, ENMs might exhibit hazardous effects for the human health or the environment. The experience gathered from the employment of asbestos fibers decades ago, serves as cautionary tale for the potential hazard of nanomaterials. Asbestos were extensively used in various products with a wide range of applications because of the advantageous properties, like heat resistance/isolation and durability. However, after many years of ubiquitous cotidian presence, the insidious harmful effects on health became evident. In this regard, it was discovered that after a long latency period asbestos were able to cause lung cancer and mesothelioma [38].

This effect was linked with its high-aspect ratio fiber structure and frustrated phagocytosis. Gathering materials by common features leading to similar adverse effects is a first step to group and read-across. Therefore, investigating the potential hazards of nanomaterials is crucial to guarantee their safety enabling their responsible usage. Understanding their potential hazard is important to engineer relevant modifications and undertake the necessary measurements to minimize exposure, while still profiting from their various beneficial properties. Understanding their potential hazard requires describing surface structure, properties and reactivity, which determines how these engineered nanomaterials interact with each other and with the environment. The latter determines trend to aggregate, which affects fate and exposure; moreover it shapes their surface properties, thus altering interaction among particles (fate, exposure) and modifying its reactivity. Although there are several exposure scenario for nanomaterial toxicity, inhalation is considered the most critical uptake route since it allows the particles to reach the sensitive tissues deep within the lungs [39]. There, the particles could be taken up by lung cells or interact with the immune system, leading to harmful effects. Asbestos is not the only 1-dimensional material potentially harmful. Other materials, with very different chemistry and composition, share common features. Studies on specific types of carbon nanotubes suggested that they may show similar toxicity as asbestos fibers at lung level [40, 41]. Similarly, research on silver nanoparticles showed genotoxicity and pointed out that smaller-sized particles exhibited higher toxicity in in vitro settings [42]. It is important to emphasize that not all nanomaterials are associated with risks for human health. Their toxicological potential depends on different factors such as size, shape, functionalization, besides its chemical composition. Ultimately, toxic effect is a phenomenon triggered at the surface of nanomaterials, and how it interacts with its environment. Bulk properties do not typically correlate with surface properties or structure; therefore, no direct translation of bulk characteristics can be made to surface reactivity. Experimental characterization may be hampered by this, which must be taken in consideration when defining characterization strategies. From a more methodological perspective, material characterization should include detailed info on the surface and its defects, how the latter and the former depend on the underlying bulk structure and defects and how those are affected and, ultimately, shaped by the interaction with the environment or biological media. The combination of these properties results in countless possible nanomaterials that can enter the market, which can rapidly overwhelm the current risk assessment

procedure.

For this reason, SSbD strategies are a particularly important as they have the potential to facilitate and expedite risk assessment, necessary to minimize the potential risks associated with the use of nanomaterials throughout their lifecycle. Moreover, SSbD strategies are crucial to ensure the responsible development of nanomaterials and their applications.

2.4. *The role of materials modeling in NM hazard assessment*

It is worth stressing that accurate physics based modelling of materials are typically not meant to directly simulate the basic mechanisms underpinning ENMs toxicity. On the contrary, such tools and methods can be used to compute properties that can better capture the complexity of ENMs composition and the influence of external conditions on toxicity, as opposed to difficult, time consuming and expensive experiments (when experiments are possible). Despite the benefits of using materials modelling to calculate nano-descriptors, even until a few years ago, the scale and complexity of system simulations were challenging, and there was shortage of models to predict important properties, such as the NM dissolution rate.[8]

One may thus conclude that it is highly desirable to base nanosafety assessment upon accurate intrinsic and extrinsic properties, i.e. *features* that represent a particular structure and chemical nature of the ENMs of interest and how these properties affect or are affected by their (biological) environment. In the following, such features or *advanced descriptors* and their evaluation by means of physics-based models is discussed in more details. Interpolation- or extrapolation-based methods like QSAR and ML will be served by a deterministic calculation of such advanced descriptors in a part of the parameter space that is expensive or hard to assess experimentally, or when experimental data are unclear or incomplete. On the other hand, physics-based models may also provide *direct* insight into the relation between observed ensemble properties (phenomena) and system parameters (simple and advanced descriptors) that stems from a systematic computational screening for one or more well-defined design parameters. Examples of phenomena considered in this review are protein absorption, NP aggregation and membrane binding. An instance of direct insight gained by computational means is the finding that small nanoparticles ($\ll d$, with d the membrane thickness, usually 4-5 nm) can simply permeate through the membrane, similar to small molecules, while large nanoparticles ($\gg d$) will be fully engulfed or wrapped by the membrane upon binding. The binding

characteristics for nanoparticle sizes comparable to the membrane thickness is still unclear.[43]

A separation of descriptors into intrinsic and extrinsic provides a straightforward basis for the selection of the most appropriate physics-based model. Section 3 discusses the current knowledge on ENMs and the (additional, advanced) features that need to be considered, as well as the progress on nanomaterial representations and on data-based models, like nano-QSARs, that can receive and process the calculated nano-descriptors.

3. Data-based models for linking ENM features to safety

3.1. The critical role of a general representation of ENMs

In the attempt of gaining a deep rationale from data linking ENMs and their observed safety properties, an essential prerequisite is the ability of representing complex materials in well structured and machine readable format. Unfortunately, to date, the lack of a standardized semantic characterization of the structural ENMs features and environmental variables makes it difficult to aggregate, curate and evaluate data from different sources and to use them for simulations or for training new (data-driven) models; thus making the meaningful integration of datasets a very demanding task. Web databases and repositories for chemical substances (e.g., ChEMBL, PDBe, ZINC15, Pubmed etc.) use standardized linear notations for substance identification (SMILES, SYBYL Line Notation, or InChI).[44, 45] These notations are also employed in deep learning tools to guide the generation of feasible molecular structures with desirable properties (e.g., Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAEs), Recurrent Neural Networks (RNNs), etc.). The extension of these notations to polymers, mixtures, reactions, etc. has also been proposed, but ENMs entail additional challenges compared to conventional chemicals. The key properties of ENMs have a strong dependence on the physical and structural features. ENMs characteristics such as the spatial relationship between components, their relative sizes, etc., all play an important role in their inherent and emergent properties. What is more, many of these properties are environment-dependent, as they are affected by various external factors such as temperature and concentration through non-linear dependencies, making ENMs descriptions even more subtle and complicated.

A complete ENM representation, therefore, would require the inclusion of information on many more aspects than what current linear notations provide. Extending the current representations to include such advanced information content is not a straightforward task. Lynch et al.[46] have initiated an extensive discussion among various stakeholders and proposed a framework for an InChI standard applied to ENMs as well as a roadmap for its development. They aimed to address the variety of complex nanostructures, using a hierarchical approach that introduces new layers on the InChI notation for the size, shape, crystal structure, and ligand binding of the ENM, and, possibly, extrinsic and surface properties. Recently, Blekos et al. proposed principles for a more accurate, complete, flexible and incre-

mental representation approach. The principles were demonstrated through the development of extensions to nano-InChI to encode morphology/mixture properties and statistical distributions of properties, and to store metadata and enable their reuse.

The proposed representation framework could also provide the theoretical background to gradually capture the real particle dynamics under specific environmental conditions. [37] These are currently under consideration in the Nanomaterials InChI Working Group (<https://www.inchi-trust.org/nanomaterials/>) and their proposal for a new InChI standard. A hope for the near future is that an increasing number of advanced descriptors - as those discussed below in Section 4 will be gradually incorporated into such standard notations.

3.2. Data-based models for safety assessment

In nanoinformatics, popular data-based models include Quantitative Structure Activity Relationship modeling (so-called nano-QSAR/-QSPR) that utilizes Machine Learning (ML) and artificial intelligence to predict the desired response (e.g., biological activity, toxicity endpoints or any physicochemical property of interest). Those approaches are based on a set of computational and/or experimentally developed descriptors representing nanoparticle structure. As a representative example, Wyrzykowska, Mikolajczyk, et al.[47] proposed a concept in which a nanostructure is characterized by a triad that describes: (i) molecular structure; (ii) molecular descriptors; (iii) molecular properties that correspond to chemical composition and chemical structures of its components. The proposed triad covers the characterization of the intrinsic properties of nanoparticle structure (so-called system-independent or intrinsic (nano-)descriptors (S-descriptors in Figure 2, left panel).[48]

Recent advancements in nano-QSAR modeling have introduced hybrid models that enhance predictive capabilities by integrating multiple modeling approaches. These hybrid models often combine molecular dynamics simulations data with machine learning techniques to better understand and predict the complex behavior of nanomaterials in biological systems and their interactions with the environment.[49, 50] Although QSAR/QSPR can be certainly regarded as valuable tools to complement experimental studies on chemicals and nanomaterials, they come with multiple challenges that should be carefully analyzed. In particular, there is an absence of comprehensive methods to characterize nanoparticles, which are essential for a standard QSAR procedure or a dependable computational approach that accurately represents the unique features of nanostructures – in short, a lack of trustworthy

nanodescriptors.

Furthermore, depending on the surrounding environment, ENMs features may change; thus, the data-based models should be developed based not only on the characterization of ENMs chemical composition/chemical structures of its components but also on the influence of the environment (including experimental conditions, or biological medium). In other words, the so-called system-dependent (extrinsic) nano-descriptors or the environment (E-descriptors) should also be considered to describe nanostructure as a whole (Figure 2, right panel).[47, 48] In the following two subsections, we provide a brief critical review of intrinsic and extrinsic descriptors in preparation to the most recent and advanced computational approaches to estimate them, as discussed in detail in Section 4.

3.2.1. System-independent *intrinsic* features

System-independent (intrinsic) nano-descriptors refer to the composition, components' structures, and properties that may be measured or calculated under a well-defined, unchanging set of conditions. A first example of advanced descriptors that were successfully applied for data-based modeling of nanoparticles was developed based on quantum mechanical calculations (so-called QM Descriptors).[51] The QM Descriptors that describe the core chemistry of the structure were proposed in 2011 by Puzyn et al.[51] QM descriptors reflect the electronic form of a chemical compound. They are obtained by applying quantum mechanics to the appropriate molecular model of an ENMs structure. During the last 10 years, different research groups have been working to develop more sophisticated types of descriptors that are not related to detailed atomistic simulations. For example, in 2012, Toropov et al.[52] proposed SMILES-based optimal descriptors based on the encoded one-, two- and three-element SMILES attributes of a compound and can be calculated with the CORAL software.[53] This idea was then extended to the simplex representation of molecular structure (SiRMS). Here, another type of descriptor based on the Liquid Drop Model (LDM descriptors) was proposed by Sizochenko et al. in 2014.[54] The methodology of LDM descriptors calculation assumes that an ENM can be represented as a spherical drop in which elementary molecules are tightly packed. At the same time, the density of clusters is equal to the particle mass density.[55] The proposed methodology is based on the thermodynamically most stable unit cell of the considered crystal structure, that is replicated in three dimensions. Afterwards, a spherical ENM is created by removing all atoms outside the indi-

cated diameter. This is a clear simplification: In fact, bulk-cut surfaces will restructure, and even a spherical particle may exhibit regular regions as well as a variety of defects. [56]. Moreover, surfactants may reorder the surface differently.[57] Clearly, during the last decade, computational scientists and nanoinformaticians active within the European safety community have made a considerable effort to integrate knowledge from existing EU completed and ongoing projects within EU FP7 and HORIZON 2020 to develop a more comprehensive approach for nanostructure characterization.[58] However, we should also recognize that several key challenges related to the appropriate representation and description of ENMs structure are still ahead. Among others, those aspects have been highlighted by results provided by Mikolajczyk et al., Wyrzykowska, Mikolajczyk, et al.[59, 48]. As far as intrinsic properties of ENMs are concerned, one of the main challenges is related to the description of the complexity of nanomaterials composition, thus clearly requiring the development and use of more advanced computational tools capable to evaluate/estimate descriptors that are difficult or even impossible to access by current experiments.

While such aspects are discussed in detail below within Section 4, an additional critical point, partly related to the latter computational assessment of intrinsic ENMs features, is related to the lack of detailed material characterization in the published literature. Too often, indeed, only nominal composition, shape and size of nanomaterial are reported, with their possible coating only vaguely (if at all) defined. Clearly, those uncertainties add even more complexity and make benchmarking of computational material modeling particularly difficult.

3.2.2. *System-dependent extrinsic features*

While using QSAR/QSPR or other data-based models, a second challenge concerns representing the influence of external conditions (surrounding environment).[47] A recently published study[48] indicates that the system-dependent (extrinsic) nano-descriptors, also referred to as *environment descriptors* (E-descriptors), are much more critical for controlling and managing out the properties of ENMs than nanostructure characteristics themselves. Thus, in addition to standard characterization, the experimentalist should provide information about changes in the structure of the nanoparticles depending on the environment (the surrounding conditions). As a result, next to the core and coating, surface properties such as protein corona formation (so-called “biomolecular corona”) play an essential role in characterizing

ENMs' behavior and may be considered its fingerprint in a biological medium see also Section 4.4 below for more details.

The system-dependent (extrinsic) nano-descriptors are crucial in describing physicochemical properties such as electrophoretic mobility or zeta potential value under specified conditions, reflecting the hydrophobicity, biomolecular corona, dissolution rate, sorption, surface reactivity, degree of aggregation/agglomeration, or ENM persistence.

Developing the system-dependent (extrinsic) nano-descriptors is challenging because the nanostructure may change during its lifetime due to its transport through different environments. In fact, nano-bio interactions are in principle driven by the nanomaterials fate in biological and environmental compartments, meaning how materials translocate, act as carriers of further toxicants and finally come in touch with biological targets. Under this perspective, the surface charge and wettability are considered the key determinants of the fate and behavior of nanomaterials dispersed in the exposure media.

Furthermore, the electrostatic interactions that keep particles dispersed, preventing or promoting contact with cell membranes, depend on the surface potential shown at the slipping plane (Zeta potential), as well as other important properties that drive nano-bio reactivity such as hydrophilicity or the surface interaction with biomolecules solubilized in the media. In this respect, the identification of the pH at which the Zeta potential is equal to zero (isoelectric point) allows making hypothesis on the type of acid/base behavior of surfaces and on the presence of charged molecules specifically adsorbed, as well as on the colloidal stability of nano dispersed phases and of the occurring of hetero-aggregation phenomena.[59, 60, 61, 62] In addition, the Zeta potential is very useful also for the design optimization of surface functionalization strategies applied to the control of nanoparticles reactivity both for nanosafety and nanomedicine purposes, because it is predictive of the amount and type of coating that masks surface sites, providing a new biological identity to the dispersed phases.[59, 63] Few studies report *in silico* models for the prediction of Zeta potential based on physicochemical intrinsic properties.[59, 64]

One of the most promising approaches that could cope with the challenge of extrinsic descriptors is an application of atomistic simulations complemented with coarse-grained models of ENM and biomolecules (see the dedicated sections below). In such multiscale approach, the coarse-grained models (nano and microscale) are parameterized and possibly validated by

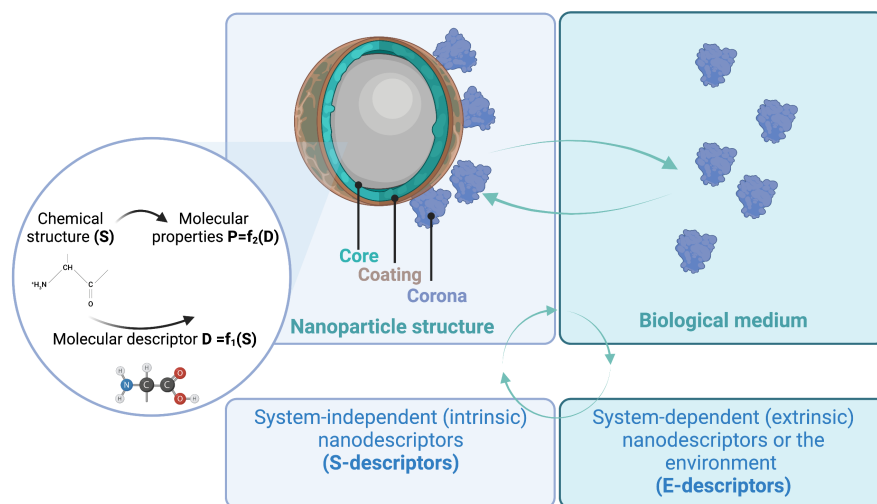


Figure 2: Molecular and time-dependent descriptors: system-independent descriptors refer to the composition, components' structures, and properties that may be measured or calculated under a well-defined, unchanging set of conditions; system-dependent descriptors represent the influence of the surrounding environment.

the data obtained from more detailed models at smaller scales (atomistic and quantum chemical). In this context, novel descriptors can also be derived from the statistics of adsorbed molecules after an analysis of the biomolecular corona.[65, 66]

3.3. *Modern machine-learning methods*

Machine learning (ML) is a branch of artificial intelligence (AI) that involves training computer programs to make predictions or take actions based on data. In ML, algorithms are designed to learn from data, instead of being explicitly programmed to perform specific tasks. The idea is to provide the computer with a large amount of data, and then use this data to teach the computer how to identify patterns, make predictions, or classify new data. At its core, ML includes four basic notions: algorithms, models, data, and training. ML algorithms are designed to learn from data and make predictions or take actions based on that data. The algorithms generate models, which are representations of the patterns and relationships in the data. The quality of the model depends on the quality and quantity of the data used to train it, as well as the algorithm's ability to learn from that data. Therefore, the process of training an ML model involves feeding it with labeled data, measuring its performance, and refining the model until it achieves satisfactory accuracy on new, unseen data.

The rationale for using ML approaches for nanosafety assessment is based on the need to efficiently process, analyze, and extract meaningful information from possibly vast amounts of data generated from both computational and experiments. ML algorithms can learn complex relationships and patterns from those data sets, enabling researchers to make predictions, optimize material properties, and identify novel materials with desired (e.g. less toxic) characteristics.

The advantages of using ML - as opposed to more traditional approaches - include:

- Accelerated materials discovery and optimization: ML algorithms can quickly process large amounts of data and identify potential new material candidates or modifications for further investigation.
- Reduced experimental and computational costs: By predicting material properties and hidden pattern identification, ML can help reduce the number of experiments and simulations required in the development process.

- Enhanced understanding of complex material systems: ML can capture non-linear relationships and intricate patterns in data, leading to better insights into the underlying physics and chemistry of materials.

However, there are also challenges and limitations associated with such ML approaches:

- Data quality and availability: ML algorithms rely on high-quality and abundant data, which can be a limiting factor in materials science, where data may be scarce, noisy or heterogeneous.
- Interpretability and explainability: ML models can be complex and difficult to interpret, making it challenging to understand the underlying reasons for their predictions and build trust in their outcomes.
- Overfitting and generalization: ML models may be prone to overfitting, namely they perform well on the training data but poorly on unseen data, thus reducing predicting accuracy.

ML techniques have been applied to a wide range of applications and fields including nanoinformatics [67]. In such a context, ML methods have the potential to significantly impact the design, characterization, and safety assessment of ENMs. However, the availability of high-quality and abundant data is still an important aspect to address.

Among other challenges, a particularly important aspect associated to data-based modelling in nanoinformatics is the proper handling of highly imbalanced datasets. Imbalanced datasets are characterized by a skewed class distribution (e.g., over 1:100 observations in the minority class compared to the majority class), where usually the minority (underrepresented) class is the most interesting one to predict. For example, in an unbalanced dataset, there could be a majority class of nanomaterials that are composed of a single element (e.g. gold nanoparticles) and minority classes of nanomaterials that are composed of multiple other elements. **As a result, when** training a ML to classify the nanomaterials the model is likely to be biased towards the majority class and may not perform well on the minority class of high interest.

To address the issue of class imbalance, various data-level and algorithm-level approaches have been proposed over the last decades [68, 69, 70]. Data level approaches are addressing class imbalance via resampling (undersampling and oversampling), as well as evolutionary algorithms for sampling,

active learning for selecting the most appropriate data points or more recently adversarial learning algorithms for new points generation and meta-learning.[71] Random undersampling (i.e. removal of observations from the majority class), Near Miss, Tomek links (i.e. removal of boundary observations) are examples of undersampling (or downsampling) algorithms. The disadvantage of the latter methods is the loss of useful information as well as possible increase of the data bias.

Oversampling approaches include random oversampling (multiplication of observations from the minority class), SMOTE (finding nearest neighbors of the minority class observations and adding points on the line joining the point and the nearest neighbor).[72] Review of SMOTE variants can be found in Ref.[73] SMOTE extensions to handle multiclass and multilabel (MLSMOTE) classification and regression (SMOTER) have been proposed,[74] and, more recently, DeepSMOTE[75] and GraphSMOTE.[76] Oversampling has the advantage of retaining all the information and usually performs better than undersampling. However, oversampling may increase the probability of overfitting.

Another important aspect to be considered is that accuracy is not an appropriate metric for assessing model performance on imbalanced datasets. Algorithm-level approaches aim at modifying the loss metric, assigning different costs to penalize errors in each class (cost-sensitive training) or introducing new algorithms which can inherently deal with imbalanced data. Instead of accuracy, the recommended metrics are confusion matrix, precision and recall, F1 score, Kappa, Area under curve (AUC) (see also Ref.[73]). Ensemble algorithms using bagging and boosting (e.g., tree ensembles as Random Forest) are known to be more robust in imbalanced settings. Recent literature addresses handling imbalanced data (also known as *long tail learning*) by deep learning methods.[77, 78] Imbalanced datasets are typical in high throughput screening[79, 80, 81] and chemogenomics.[82]

ML algorithms can be broadly categorized into supervised, and unsupervised learning. We begin with a discussion of supervised learning in the following subsection.

3.3.1. Supervised learning

Supervised learning works on labelled data, with the goal of approximating a function that maps input to labelled data. A prototypical example would be the ability of linking a set of relevant intrinsic and extrinsic advanced descriptors for a number of ENMs to their observed (eco-)toxicological

endpoints (e.g. the survival and/or reproduction rate of a chosen organisms). Supervised learning techniques include a large variety of algorithms and methods like Decision Trees, Random Forest, Support Vector Machines, various classifiers like k-Nearest Neighbors, Neural Networks and Instance-Based Learning methods (see Figure 3, left panel).[83, 84]

In one of the earliest applications of ML methods in manufactured nanoparticles, Fourches *et al.* used Support Vector Machine-based classification and kNN-based regression to generate Quantitative Nanostructure-Activity Relationship (QNAR) models to predict biological activity profiles of novel nanomaterials.[85] Later, Puzyn *et al.*[51] presented a method to quickly test the potential toxicity of engineered nanoparticles. They applied a multiple regression method combined with a Genetic Algorithm (GA-MLR) to create a model that described the cytotoxicity of 17 different types of metal oxide nanoparticles to bacteria *Escherichia coli*. [51] Gernand and Casman performed a regression-tree-based meta-analysis on rodent pulmonary toxicity exposed to uncoated, non-functionalized carbon nanotubes. They reported the application of Regression Tree models, Random Forest models, and a random-forest-based dose-response model.[86] Winkler *et al.* used novel sparse ML methods to model the biological effects of nanoparticles with various compositions, including iron oxide nanoparticles and gold nanoparticles. They employed Bayesian neural networks using both linear and nonlinear ML methods.[87]

Evolutionary approaches have also been explored. Le and Winkler,[88] for example, reviewed the use of artificial evolutionary methods for the identification and optimization of novel materials. They report uses of genetic algorithms to investigate properties of bimetallic core-shell and titanium dioxide nanoparticles.[89, 90] Martinez *et al.* presented decision tree models based on evolutionary algorithms that classified nanoparticle aggregates into morphological classes.[91] kNN algorithms have also been used to model the toxicological properties of nanomaterials. Wang *et al.* used a kNN algorithm to develop QNAR models for biological activity profiles like cellular uptake in various human cells and the ability to induce oxidative stress.[92] Kovalishyn *et al.* used kNN, random forest, and neural network methods to generate models for the analysis of eco/toxicological and physicochemical properties for metal and metal oxide nanoparticles.[93]

Various Neural Network architectures have also been investigated in relation to predictive nanoinformatics. Gomez-Bombarelli *et al.* trained a Deep Neural Network to automatically generate novel chemical structures

and demonstrated their method on small drug-like molecules.[94] Hataminia et al. used a neural network to model the cytotoxicity of iron oxide nanoparticles in relation to kidney cells.[95] Lazerovits et al. trained a Deep Neural Network to predict nanoparticle biological fate immediately after intravenous injection and tested it by predicting nanoparticle spleen and liver accumulation.[96] Very recently, Balraadsing et al. investigated the performance of various supervised ML algorithms in the context of acute *Daphnia Magna* nanotoxicity prediction. Here, Authors created classification models based on Random Forest, Neural Networks and kNN algorithms.[97]

Li et al. used multi-target Random Forest Regression to predict the performance of sunscreen based on the type of [titanium dioxide](#) nanoparticle additives. Based on those models, they demonstrated the use of inverse design models that identify nanoparticle configurations based on desired sunscreen properties.[98]

While the literature provides numerous examples of supervised learning techniques applied to nanoinformatics, the choice of an appropriate method is not always straightforward and often depends on several factors. For example, the choice between classification and regression usually depends on the type of problem being addressed: classification is used for categorical output variables, while regression is used for continuous output variables. Various factors, such as data size, dimensionality, complexity, and desired model interpretability, influence the choice of a specific algorithm.

Some general guidelines that can be considered when selecting an algorithm include the following: Decision Trees and Random Forest are well-suited for problems with mixed data types (numerical and categorical), and they provide easily interpretable results. Support Vector Machines (SVM) are appropriate for high-dimensional data and complex decision boundaries; however, they might be computationally intensive for large datasets. k-Nearest Neighbors (kNN) is a simple, instance-based method that performs well with small datasets but can be computationally intensive and sensitive to noise for larger datasets. Lastly, Neural Networks are effective for modeling complex patterns and relationships in high-dimensional data, but they may require more extensive computational resources and may not provide easily interpretable results.

3.3.2. *Unsupervised learning*

Unsupervised learning mostly deals with unlabelled data. This is particularly advantageous when data labelling is a resource-intensive task. The

goal of unsupervised learning is to learn or discover the structure and patterns of the input data, and is often based on exploring the similarity among the input variables. In the case of nanomaterials, [unsupervised learning is thus useful to explore similarities based advanced descriptors](#). Clustering algorithms are the most representative algorithms of unsupervised learning. Popular examples include k-means, Principal Component Analysis, various Neural Networks like Self-Organizing Maps and Hierarchical Clustering (see Figure 3, right panel).[99, 83, 84]

Unsupervised learning techniques have been used for the classification of nanomaterials or nanomaterial properties and quality assessment. They have also been used in combination with supervised or other statistical methods to assist, for example, the development of QSAR and neural network models. Wang et al. used Principal Component Analysis to analyze the structure toxicity relationship for various nanoparticles and identified the physicochemical properties of the nanoparticles that are risk factors for cytotoxicity.[100] Jha et al. modeled the toxicity of nanomaterials using a multivariate statistical analysis approach: through a multivariate Principal Component Analysis, they selected descriptors that optimally separated toxic from non-toxic nanomaterials.[101] Sizochenko et al. used a Self-Organizing Map in their hybrid approach in order to identify hidden patterns of toxicity among nanoparticles and to determine the underlying factors responsible for the toxicity.[102] Sizochenko et al., in order to evaluate the genotoxicity of metal oxide nanoparticles, developed another hybrid supervised and unsupervised ML approach where they used a Self-Organizing Map to estimate relative distances between nanoparticles.[84] Kotzabasaki et al. also used a hybrid approach in developing a predictive model for the prediction of the genotoxicity of Multi-Walled Carbon Nanotubes. The latter Authors used the information derived from the experimental characterization of CNTs and a combination of Principal Component Analysis and supervised classification techniques to improve the accuracy of the analysis in their parameters.[83]

[Similarly to the supervised learning, selecting an appropriate method is often contingent on the specific goals and data structure. Unsupervised learning aims to identify underlying patterns, structures, or relationships within the data without relying on labeled outcomes. Several factors, including data size, dimensionality, complexity, and the nature of the problem, influence the choice of a specific algorithm. Some general guidelines to consider when selecting an unsupervised learning algorithm are as follows:](#)

[Clustering methods, such as K-means, DBSCAN, and hierarchical clus-](#)

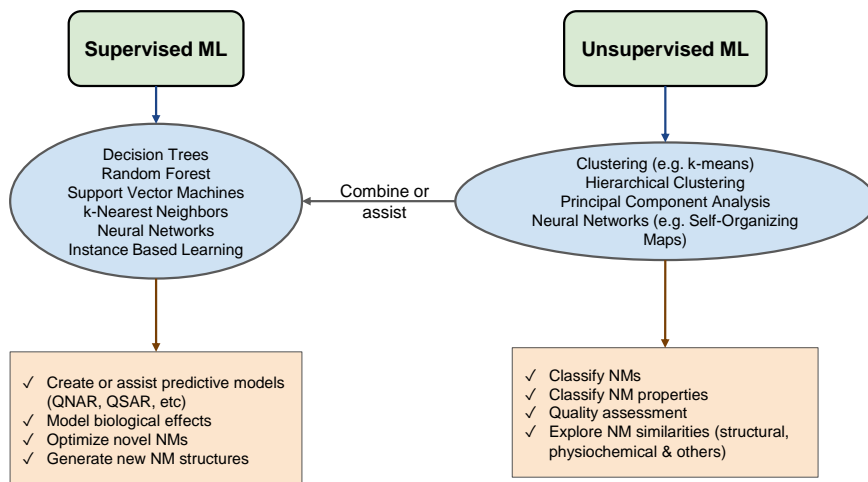


Figure 3: Use of Supervised and Unsupervised Machine ML in predictive nanoinformatics: Supervised ML methods (Decision Trees, Random Forest, Neural Networks etc.) are often used to create or improve predictive models while Unsupervised ML methods (Clustering, Self-Organizing Maps, etc.) are more often used to group and explore nanomaterials properties or in combination with Supervised ML methods.

tering, are suitable for partitioning data into groups based on similarity or distance metrics. They are particularly useful when exploring the intrinsic structure of the data or identifying previously unknown subgroups.

Dimensionality reduction techniques, including Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), are employed to project high-dimensional data into lower-dimensional spaces. These methods can aid in data visualization, noise reduction, and improving the performance of other machine learning algorithms.

Autoencoders, a type of neural network, are effective for unsupervised feature learning and representation. They can be used to reduce the dimensionality of data, denoise data, or learn more complex and abstract features.

For the sake of clarity, in the Table 1 below, we summarize what we believe are the most important open issues in data-based modeling assessment of ENMs safety. Concurrently, we provide recommendations on possible effective actions that could help addressing such challenges in the near future.

Identified challenge	Recommendations for future research
Representation and description of intrinsic properties of ENMs is much more complex and challenging as compared to chemicals	Effort should be devoted in adding additional and more comprehensive information layers to the nano-InChI notation following the ongoing effort in [46, 103]. Additional intrinsic descriptors should be developed on the basis of experimental data but also advanced material modeling tools, such as quantum and atomistic simulations as discussed in Section 4 below. Attention should be paid to the use of computational techniques as their robustness and the energetic needs associated with massive calculations might be a constrain in the next future.
Detrimental lack of well characterized ENMs data: go beyond nominal data	An effort should be made to design experimental and theoretical characterization including interfacial regions, key in the ENMs reactivity. In particular, the accurate description of the electronic structure at the surface is needed to capture specific interactions and mechanisms.
The influence of external conditions may play an even more critical role as compared to intrinsic ENMs properties	Experimental testing and characterization should focus on approaches (and report data) capable to track changes in the nanoparticle structure as a function of environmental conditions.
Unlike intrinsic features, extrinsic ones are time dependent and may change during ENMs lifetime while transported through different environments	As discussed in detail in Section 4, some of the advanced material modeling approaches can be used both for estimating the value of extrinsic advanced descriptors under disparate conditions and for gaining further understanding on the basic mechanisms underpinning their change in time.

Training datasets for ML based models are often imbalanced with the most interesting material class being underrepresented	Use of undersampling and/or oversampling techniques trying to minimize loss of information and overfitting. Explore data-level and algorithm-level approaches. Develop new methods or adapt existing techniques, such as SMOTE and its variants. More focus on evaluating model performance using metrics suitable for imbalanced datasets, such as precision, recall, F1 score, Kappa, and AUC.
Addressing class imbalance issues in supervised learning for nanoinformatics.	Utilize data augmentation techniques, resampling methods (e.g., undersampling and oversampling), cost-sensitive training, and ensemble approaches to mitigate imbalance and improve prediction accuracy.
Ensuring adequate data quality and quantity for effective supervised learning.	Investigate ensemble methods such as bagging and boosting, feature selection techniques, and regularization methods to enhance model performance and stability.
Ensuring generalizability of ML models to new data.	Evaluate model performance on external datasets to ensure generalizability and robustness. Utilize cross-validation techniques to assess model performance on different subsets of the data. Pay attention to model overfitting, which occurs when the model is too complex and performs well on the training data but poorly on new, unseen data. Regularization methods such as L1 and L2 regularization can be used to reduce overfitting and improve generalizability.

<p>Selecting appropriate ML algorithms for specific data structures and goals.</p>	<p>Consider various factors, such as data size, dimensionality, and complexity when selecting specific ML algorithms. For example, Decision Trees and Random Forests are good for mixed data types and provide interpretable results, while Support Vector Machines (SVM) are suitable for high-dimensional data and complex decision boundaries. k-Nearest Neighbors (kNN) is a simple, instance-based method for small datasets, while Neural Networks are effective for modeling complex patterns and relationships in high-dimensional data but may require more computational resources.</p>
<p>Integration of different ML models for a comprehensive understanding of ENM safety.</p>	<p>The literature provides numerous examples of ML models applied to nanoinformatics. However, the majority of the published studies focus on a single ML algorithm or approach. A comprehensive understanding of ENM safety requires the integration of multiple sources of data, including experimental and theoretical, and the use of different ML models to capture the complexity and variability of the system. The integration of different ML models can enable a more accurate and robust prediction of ENM safety by incorporating different types of information and reducing the uncertainty associated with each individual model. However, integrating different models poses several challenges, including data compatibility, model complexity, and the need for appropriate validation methods.</p>

Table 1: Summary of the main challenges associated to data-based modelling in the ENMs safety context

4. Physics-based models for nanostructure characterization

In general, physics-based modelling techniques can be grouped according to the dominating length and time scales (or resolution), and equivalent representations at different resolutions can be related to each other via a coarsening or fine-graining procedure, in which degrees of freedom (electrons, atoms) are averaged out (forward-mapping) or introduced (back-mapping). For instance, the familiar quantum methods at the *electronic* level explicitly consider electrons and atoms, enabling them to calculate several intrinsic properties of the material with great precision, including the electronic bandgap and surface reactivity. At the *atomistic* level considered by classical molecular dynamics, explicit electronic structures like the bonds are only implicitly represented in a force field. The resulting enhanced sampling rates can be exploited to study larger system, for instance, the long-term conformational dynamics of one protein or the interaction between a nanoparticle and several biomolecules.

Moving to even larger scales, interactions between ENMs and their surroundings, reflected by extrinsic ENM properties, often involve processes that exhibit a considerable disparity in length and time scales. From a computational perspective, they therefore require yet another - *mesoscopic* - level to be tractable *in silico*: a level that is and will not be genuinely within reach of atomistic methods for some time despite the continuing advances in computer power. For instance, even when investigating the uptake of tiny particles such as fullerenes (<1 nm) by human cells (10-100 μm), which first cluster to larger aggregates in solution before binding to a membrane due to their hydrophobicity, the spatial scales that have to be adequately represented span many orders of magnitude.

This challenge is even more significant for the (competitive) binding of one or more molecules like proteins onto a nanoparticle, since they generally experience conformational changes on various length scales upon absorption. The same holds for time, as diffusion or translocation processes in dense molecular environments like membranes and intracellular spaces take place at timescales that are orders of magnitude greater than the fastest vibrations inside a molecule - molecular bond stretching - that set the elementary time scale. At the most basic level, this challenge can be faced by experiment, via trial and error for individual setups. Yet, as the experimental resolution is also essentially limited at the bottom where many relevant molecular mechanisms of interest take place, one may formulate theoretical or computational

answers for more well-posed questions.

In the following, we will concentrate on three extrinsic properties for which classical molecular dynamics significantly falls too short: nanoparticle clustering, the formation of a protein-corona, and nanoparticle uptake by a lung membrane.

A subset of advanced descriptors also stemming from such calculations is reported in the Appendix, where more detailed info are reported. We note that, especially at the mesoscopic level, this overview of progress in the field of nanoparticle simulations is not exhaustive. It should also be noted that much effort has been put into ensuring *equivalence* at the highest level of resolution, e.g. by proper parameterization of tight-binding methods at the quantum level and determination of proper force fields for atomistic molecular dynamics. At the mesoscopic level, many developments are quite recent and the issue of equivalence is more complicated to satisfy, which hampers the application range of genuine multi-scale methodology based on systematic coarse graining. For this reason, many mesoscopic investigations have focussed on evaluating mechanisms for generic setups, which cannot be directly related to specific systems that are needed for advanced descriptors. For details of such activities, we refer the reader to published reviews.[104]

4.1. *Intrinsic advanced descriptors: Electronic level*

The quantum-mechanical computations of nanoparticle properties relevant to nanotoxicology is still a challenge as it typically involves large size and time scales. The key question is to reveal the electronic structure nature in the complex toxicity mechanisms. A valid strategy involves the accurate determination of descriptors (heat of formation, lattice energy, enthalpies of cation detachment) and band structure (bandgap, HOMO and LUMO levels) that could be later related to toxicity endpoints by data-based modelling. This approach was successfully proved in Nano-QSAR models [51].

The computational cost of *ab initio* calculations, typically density functional theory DFT, remains a strong limitation to the deployment of quantum-chemical descriptors, mainly due the size of realistic nanoparticles (from few nanometers to hundreds of nanometers) and the complexity of the surface region (often unknown in experimental data). A detailed quantum chemical description of those systems severely increases the time to acquire data.

Faster computational schemes use the so-called *quasi-ab initio* methods based on DFT applying expansions of the atomic electronic structure using the Tight Binding approximation (DFTB) [105]. Such techniques, 3,000

to 30,000 times faster than regular DFT, allow considering more realistic sizes and conditions such as solvation or molecule adsorption. To compute dynamic effects, *ab initio* molecular dynamics (AIMD) is a promising technique, yet at a **too** high computational cost.

In the last years, some of the aforementioned techniques have been successfully applied to investigate nanosized titania with sizes 2.2-4.4 nm. An example of a spherical 3 nm titanium dioxide nanoparticle is shown in Figure 4, panels A and B. The model contains more than 1200 atoms, 6000 electrons and 50000 cpu hours [58] for routine DFT characterization. In panel A, the titanium dioxide nanoparticle is shown (left), with an inset highlighting some surface atoms (middle); it can be seen that they differ from the bulk geometry. A scheme of DOS (Density of States) is shown (right) as an example of electronic descriptor associated with chemical reactivity. In panel B, the decrease in energy during a DFTB simulation in water is shown (left, middle), and the corresponding RDF (radial distribution function) for pairs of atoms (right), accounting for the surface structure in the proximity of the solvent molecules.

Other properties recently reported in the literature for titanium dioxide nanoparticles concern crystallinity [106] and hydration [107, 108]. Interestingly, DFTB tools have been recently applied to understand biological processes such as the inactivation of SARS-CoV-2 virus [109], or the coating by biological molecules [110, 111]. It is however advised to carefully check the applicability of DFTB methods for electronic structure characterization as their accuracy critically depend on the parameterization; for instance the oxygen vacancy energy for different titanium dioxide termination is not always well captured by DFTB [112].

The extension of quantum-based methods to the field of (nano)toxicology should be naturally observed in the next years, provided that a fundamental understanding of the physicochemical behavior behind toxicological effects is achieved. This involves several aspects.

First, the identification of new relevant descriptors based on a detailed knowledge of the electronic structures of realistic complex interfaces. The quality of the descriptors is crucial for the precision of the results and should be carefully addressed. Experiments focused on revealing the role of the surface in the toxicology mechanisms should be designed to properly interplay with electronic structure calculations.

Second, **we envision interesting progress in the near future due to** the development of faster and more accurate techniques to account for the complex

structure and reactivity, based on machine and deep learning [113, 114, 115, 116] or automatized structure screening tools (like Grand Canonical [117], evolutionary [118] or clustering [119] algorithms). However, they should be still tested on realistic nanotoxicity models and could possibly become routine methods accounting for composition, coating and biological media.

4.2. *Intrinsic advanced descriptors: Atomistic and Molecular level*

In atomistic (also called classical or all-atom) MD simulations, atoms are represented by particles interacting according to Newton's second law. In such an approach, electron dynamics is neglected, dramatically speeding up the simulations (as compared to DFT calculations), and system with larger dimension become affordable (with the typical computational domain having up to dozens of nanometers edge length). Given the size range of nanoparticles, MD is a powerful tool to study such systems in a detailed way that is often far from the experimental capabilities.

With respect to the interaction of nanoparticles with biological moieties, MD was recently used to determine the binding energies for nanoparticles made from three bare materials (silver, silica and titanium dioxide) of three different sizes (1, 3 and 5 nm diameter) via MD-based Potential of Mean Forces (PMFs) using umbrella sampling.[120] To cover the diversity of responses that are possible for nanoparticle binding of a real human lung membrane, which is facilitated by a mixture of different lipids with their own phase behavior, a membrane with an equivalent lipid composition was considered. Calculating binding free energies via the PMF along a normal reaction coordinate (typically the nanoparticle-membrane distance) for real materials provides a new and useful advanced descriptor for QSAR, especially when the particle size effect can be explored within the domain where the membrane response is sensitive. For non-spherical nanoparticles, however, the PMF calculation becomes less trivial and involves several reaction coordinates. Moreover, the nanoparticle sizes within reach of MD typically only cover part of the domain of interest, and are significantly below the experimental sizes.

While current MD results are exciting and make a first step towards the *in silico* assessment of advanced nano-descriptors for nanosafety, they also illustrate the essential challenge associated with MD. In the first place, experimental nanoparticles are usually at least 50 nm in size, meaning that the maximum sizes considered in current studies,[104] i.e. in the order of the

membrane thickness, are still far from most of the real applications, especially given that many advanced descriptors are size-dependent and that it is unknown if and how one may extrapolate. The most important drawback, however, is the cost of individual simulations given the immense nanoparticle design space, which comprises size, shape, elasticity, charge, composition/hydrophobicity, and surface modification,[104] if we leave dynamic surface modifications, such as the formation of a protein corona, by the surrounding medium out of the picture. The majority of existing MD studies of nanoparticle-nanoparticle and nanoparticle-membrane interactions focus on the molecular understanding that can be gained for specific experimental setups. A very promising application of MD is the calculation of binding energies between vitamins and specific nanoparticle interfaces, represented as an infinite slab, using metadynamics, for the purpose of validating QSAR predictions and nano-descriptors.[121]

Theoretical descriptors for small organic molecules[122] can be readily calculated with various levels of theory to represent most molecular features. In the case of nanoparticles, the size is the obvious limiting factor for the calculation of whole particle nano-descriptors where all atoms are considered. To address those issues, full particle molecular nano-descriptors developed by Tämm et al.[123, 124] were calculated directly and solely from the structure of the considered nanoparticles. Such calculations can be performed with LAMMPS program[125] together with Buckingham potential[126] and Wolf summation.[127] The developed set of atomistic nano-descriptors is based on the chemical composition, potential energy, lattice energy, topology, size, and force vectors. Reported studies based on the latter approach can cover different properties of the nanoparticles including also differentiating such properties in core and shell regions of the nanoparticle.[123, 124, 128] The core region usually captures similar properties to bulk material, while the shell region is expected to account for the special nanoparticle properties. While there are different methods for the *in silico* generation of nanoparticles, there is still a need to define and study the property differences in core and shell regions of the nanoparticle. Therefore, a software tool has been recently developed to defines the shell depth for all nanoparticles (<https://nanogen.me/shell-depth>). The tool requires the *xyz* type files as input and calculates the optimal shell depth for nanoparticles together with the average coordination numbers for different atom types in the nanoparticle.

As far as reactive phenomena are concerned, attempts to go beyond classical force field simulations could be based on reactive force fields (ReaxFF).[129,

130, 131] Such an approach unlocks the possibility of studying larger computational systems at a fraction of the cost as compared to quantum-level simulations (i.e. DFT and DFTB). However, it requires a delicate tuning of several dozens of parameters against energetics from first principle simulations at lower scales. This fine-tuning often leads to a lack of generality of force field parameters and may need a case-by-case optimization.

Nanomaterial dimension and shape are important to determine the corresponding toxicological endpoints: the formation of aggregates from small nanoparticles modulates the amount of material entering the cells, regulating its interaction with the DNA. Aggregates are too large to study using Classical Molecular Dynamics: this kind of system can be simulated at the mesoscopic level, as discussed below.

4.3. *Moving from intrinsic to extrinsic descriptors*

For the physics-based determination of intrinsic and extrinsic descriptors, DFT or classical MD are the first methods of choice. In contrast to mesoscopic methods, which are based on averaging, they incorporate the most direct and material-specific reaction, chemical, conformational, and interaction detail possible. Yet, for all these methods, the total computational effort invested for the determination of material properties is proportional to the number of degrees of freedom (basis functions, atoms or groups of atoms) that have to be taken into account multiplied by the number of discrete (time) steps needed for numerically stable minimization or equilibration of the system at hand. Even on exascale high-performance computing environments, where the original calculation/simulation can be divided into parts and distributed over multiple processors, nowadays even up to a million, this proportionality represents a serious computational limitation.[132]

Intrinsic properties, which are determined for (solvated) isolated molecular systems, generally fall inside the reach of the most detailed DFT or MD modelling, because the system size and equilibration times usually remain manageable. Yet, although quantum methods have a clear advantage over classical MD, their applicability is restricted to small systems compared to most experimental ENMs, see section 4.1, meaning that one should extrapolate or concentrate on surface properties. Classical MD pushes this boundary up somewhat, and is capable of simulating, for instance, the formation of a stable ligand coating around a nanoparticle of relevant size, or the exploration of the protein folding funnel on a second scale in explicit solvent. The

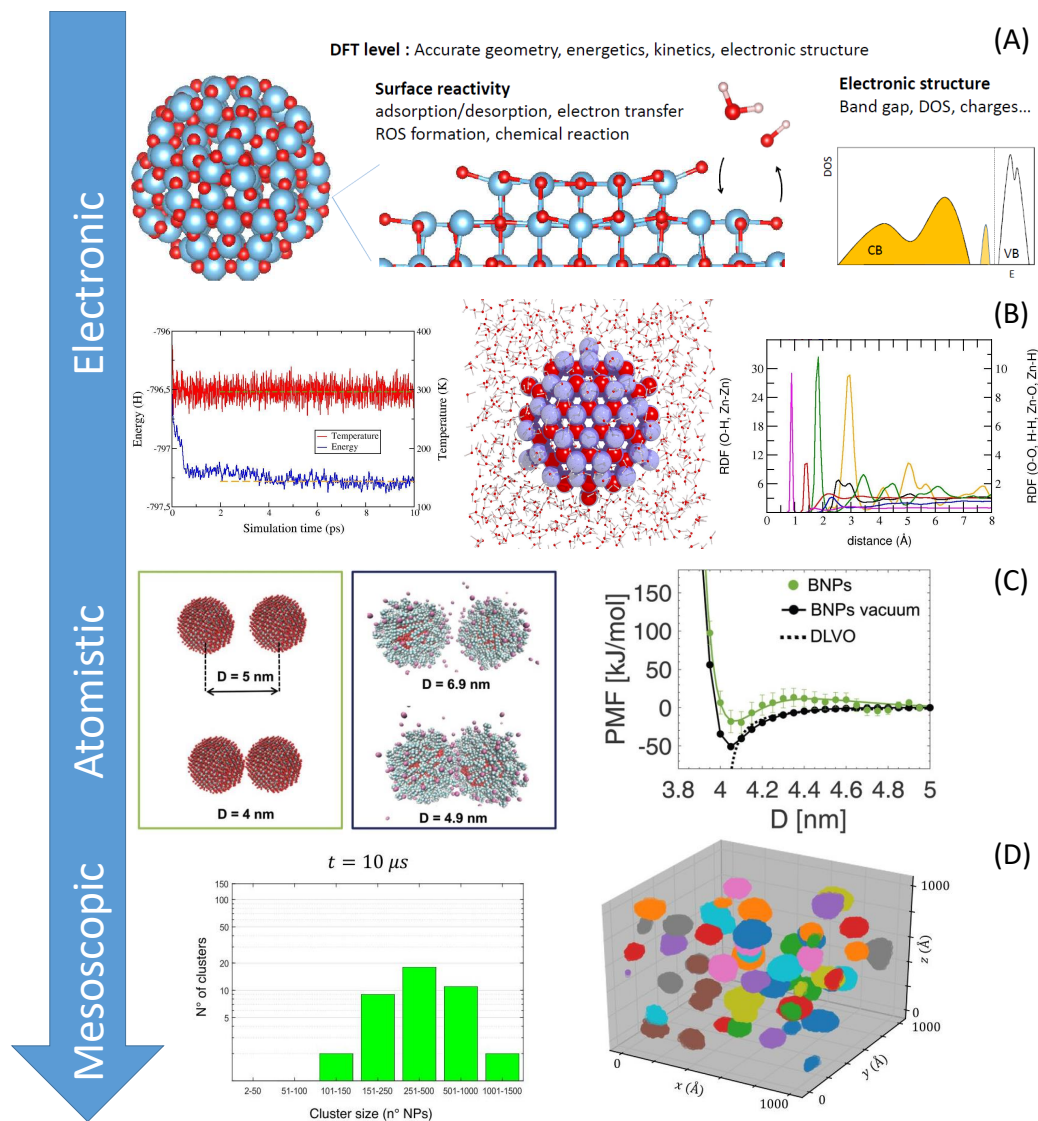


Figure 4: Multi Scale Materials Modelling: from the electronic level to the atomistic description up to the mesoscopic level. Panel A: DFT level allows accounting explicitly for electrons, providing accurate descriptors for geometry, energetics and electronic structure; Panel B: (left) Energy and temperature stabilization of ZnO nanoparticle of 2 nm in water medium at 300 K for 10 ps simulation (centre) ZnO nanoparticle of 2 nm in diameter in water (right) Radial Distribution Function of ZnO nanoparticle of 2n in water medium at 300 K. Panel C: bare and coated nanoparticles on the left and the potential of mean force for the approaching of two identical nanoparticles in a vacuum and in water and the comparison with the original DLVO theory (Reproduced from Ref. [9] with permission from the Royal Society of Chemistry); Panel D: Brownian Dynamics simulations of nanoparticle clustering, cluster size distribution on the left and snapshot of the simulated system on the right (Reproduced with permission from Mancardi et al., MDPI Manomaterials 2022.[58])

most detailed methods are thus particularly useful for providing implicit descriptors.

This situation changes when dealing with extrinsic descriptors. Nanoparticle aggregation and uptake, as well as biomolecule absorption, are at play upon the release of nanoparticles into an biological environment. The length and time scales involved in these processes are many orders of magnitude greater than the elementary Angstrom and femtosecond scales of classical MD. In particular, when computationally evaluating the nanoparticle interaction with a surrounding bio-matrix of structured lipid envelopes and unstructured mixtures of shorter and longer bio-molecules, care should be taken in selecting proper system sizes. After all, these systems should sufficiently represent the key elements in the larger open system, and be large enough to avoid computational artifacts due to boundary conditions.

As such, multi-scale approaches, in which the system is evaluated at a coarse time and length scale at some stage, become compulsory.[132] In this context, fine grained electronic and atomistic methodology still serves a distinct role as a reference for equivalence, via systematic mapping, or by providing structural input for realistic evaluation at a coarser level, e.g. relevant nanoparticle-nanoparticle interaction potentials or (static) protein conformations for docking to a nanoparticle.

A general issue in assessing ecotoxicity by multiscale physics-based modelling is that, historically, mesoscopic modelling primarily aims at providing fundamental insight into phenomena at a structural or kinetic level rather than determining accurate extrinsic descriptors, since it puts stringent constraints on the way atomistic or molecular detail is absorbed into a coarser description. Also after coarsening it is wise to build up complexity step-by-step, testing the averaging procedure at each increment. Therefore, in membrane binding, the current focus is on calculating advanced extrinsic descriptors like binding energies for passive rather than spontaneous (protein-induced) binding. In the foreseeable future, the addition of more players, including the many lipid types that render a cellular membranes, membrane-bound and adsorbed proteins, and the cytoskeleton, will become an option, and the particular role of these actors can be studied *in-silico*.

4.4. *Extrinsic advanced descriptors: Mesoscopic level*

The simulation of aggregation phenomena involving thousands of nanoparticles is simply too demanding to be carried out by all-atom MD and coarse-graining procedures are mandatory to make such simulations feasible. An

effective strategy is to employ Brownian Dynamics (BD) simulations, see Figure 4, panel D.[9, 58] Here, each nanoparticle is represented by a spherical bead characterized by the nanoparticle’s diameter and interacting with other beads according to mathematical equations describing the interaction potential (e.g. calculated by means of classical Molecular Dynamics for a nanoparticle pair, see Figure 4, panel C). Applying this coarse-graining procedure makes it possible to calculate new molecular descriptors ruling particle aggregation that could be fed into QSAR models.[58] Additional challenges faced by developing general computational and theoretical modelling for understanding the nanosafety of ENMs stem from the role of the environment in determining the observed toxicity. Once the ENM enters a living organism, it gets in **contact** with the biological molecules, in particular proteins and lipids.

Nanoparticle-proteins and nanoparticle-lipids interaction could in theory be investigated using Brownian Dynamics simulations, which can run even on a 16 core workstation, provided that all pair interactions are known; in practice, this has not yet been done because the calculation of the free energy profiles for each pair by all-atom MD is too computationally demanding. Anyway, this could be an interesting attempt to bridge the gap between the molecular simulations scale and the experimental scale.

The oldest approach for computationally determining material properties, i.e. the continuum mechanics pioneered in the 19th century by Cauchy, is in fact most suited for screening purposes, since it combines modest computational costs for realistic system sizes with a few effective screening parameters. This screening idea is at the basis of continuum Self-Consistent Field Theory (SCFT), which was developed to describe phase behavior and phase separation dynamics in block copolymers (represented as flexible chains) based on an implicit molecular representation [133].

Rigid objects like nanoparticles have also been incorporated into SCFT, primarily for the purpose of modelling polymer nanocomposites,[134] and we refer to early papers for details about the different approaches.[135, 136, 137] While these field-based methods possess a clear advantage of efficiency over particle-based methods like AAMD and CGMD, which stems from the choice to deal with ensembles rather than individual chains, and SCFT interactions are of the desired many-body type by definition, this is offset by the serious disadvantage of not being able to represent specific interactions at the molecular level and having no access to conformational detail. In addition, the commonly used excluded volume interactions in SCFT do not allow for

phase transitions that can play a role in membrane binding processes. Only recently, hybrid particle-field (hPF) approaches such as hPF-MD[138] and single chain in mean field (SCMF)[139] have introduced the ability to combine particle-based (atomistic or segmental) molecular detail with the efficiency and multi-body nature of Hamiltonians from continuum theory like SCFT. Phase transitions and/or coexistence has also been added recently.[140] Until these hybrid methods offer a validated solution for the need to combine efficiency with specificity and molecular detail, we conclude that SCFT is useful for the investigation of general phenomena, but not suited for the extraction of extrinsic material descriptors.

A popular and very efficient representation of lipid membranes is that of a thin elastic sheet without any molecular detail. Shape, dynamics and responses to deformation are dictated by a continuum Helfrich free energy that depends on a few collective properties like bending rigidity and lateral membrane tension, which can be directly related to specific membrane compositions via particle-based simulation. An extended Helfrich model developed later provided straightforward conditions for nanoparticle uptake, *i.e.* the balance of energy needed to stretch and bend the membrane around the nanoparticle and the energetic gain of nanoparticle binding. The latter is provided by the adhesion energy density of (coated) nanoparticles in the contact region and, as was later found, also in near-contact regions. Extracting adhesion energy density for real materials from more detailed descriptions is unfortunately a far from simple task.[141] Early on, Deserno et al. used this continuum model to show that a tensionless membrane can only adopt two states: an unbound state where the membrane is flat, or a state where the nanoparticle is fully wrapped by the membrane.[142, 143]. Accounting for factors that were missing in the original Helfrich-based analysis, such as the membrane thickness and interaction range, Raatz et al. and Spangler et al. found that also partially wrapped cases could be stable.[144, 145]. While providing important energetic insight, and therefore being of potential use for restricted but quick screening, the lack of lipid detail already seriously hampers the use of such methods for extracting extrinsic material descriptors. Another disadvantage is that the nature of the membrane deformation upon nanoparticle binding is not an outcome, but required *a priori*, introducing a risk of overlooking alternative binding mechanisms. The historical solution to this issue is to employ highly coarse-grained particle-based models with implicit solvent for studying generic membrane dynamics and nanoparticle-membrane interactions. Using such a model, the wrapping characteristics

for nanoparticles up to 40 nm was considered, *i.e.* the entire range for the mechanism is expected to switch, and a discontinuous transition from partial to full wrapping was predicted around 10 – 15 nm nanoparticles.[145, 146] Since these models lack solvent, which is known to modulate the (free) energy landscape, and they also lack the resolution to distinguish between different lipids, they do not represent a decisive step forward in the search for accurate descriptors.

Summarizing, one may conclude that these efficient molecule-based mesoscopic methods are useful for extracting information about general balances and mechanisms for systems of relevant size, but they were never meant to provide extrinsic nanoparticle descriptors. Just this, balancing (chemical) information and efficiency with the aim to retain the necessary detail, is the main purpose of recent development in systematic coarse-grained methodology. Although there are several ways to perform systematic coarse graining, depending on the characteristics of the reference atomistic system that one wants to reproduce, they are all based on lumping groups of atoms into CG particles or beads. The most popular method, CG Martini, combines 2-4 heavy atoms in a single bead. Generating a description in terms of CG beads does not only reduce the computational load, but it also softens the interactions. As a result, also the system evolution is significantly accelerated. Methods based on such types of coarse graining have been applied to study lipid partitioning in general, the binding of elastic nanoshells,[147, 148] the adhesion of anisotropic nanoparticles[149, 150] and of functionalized nanoparticles.[151, 152] Also the role of bending and adhesion in the distribution of multiple nanoparticles inside the membrane has been investigated, both in terms of a generic representations in a highly CG method[153] and for more chemically resolved CG representations.[154, 155, 156] Yet, while these CGMD studies have either been designed to properly represent a particular experimental nanoparticle or to obtain insight into more general binding mechanisms, very few have focused on the challenge of developing a transferable representation or map from the atomistic to the coarse-grained domain. Yet, determining such a map that is valid for all nanoparticles of the same material is a prerequisite for the extraction of advanced descriptors and trends that enable extrapolation.

One very recent example of such a new development is the special CG nanoparticle representation within the familiar Martini CG approach that is required for studying binding and translocation pathways of realistic silver nanoparticles across solvated lipid barriers in the lungs, see Figure 5.

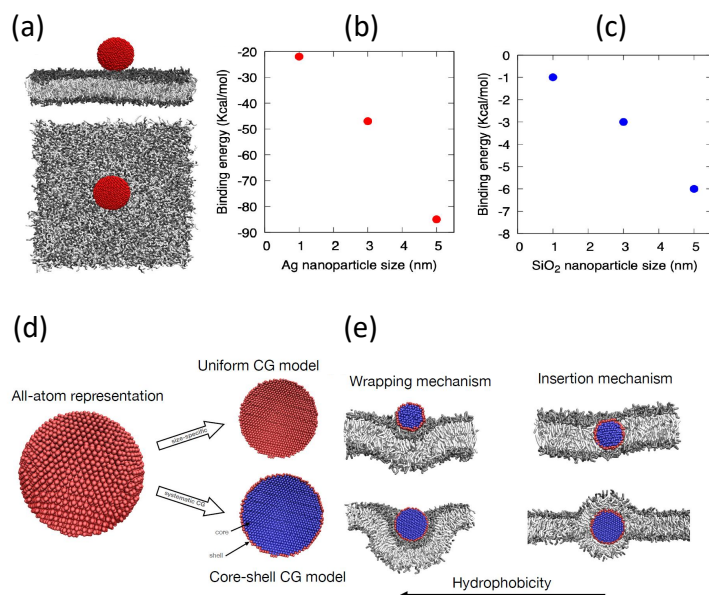


Figure 5: (a) All-atom MD of nanoparticle/membrane/water system, (b) all-atom MD binding free energy for three different sizes of Ag nanoparticles, (c) all-atom MD binding free energy for silica for three nanoparticle sizes (graphs reproduced from Ref. [120] with permission from the Royal Society of Chemistry). (d) Standard uniform CG model and the new core-shell CG model, reproduced with permission from Singhal et al., MDPI Nanomaterials, 2022[158]. (e) With increasing hydrophobicity, the mechanism of direct insertion into the model lung membrane switches to a wrapping mechanism.

Whereas the modelling community has thus far generally approached the fundamentals of such large-scale phenomena via implicit-solvent continuum[142, 143] or highly coarse-grained descriptions[145, 146, 157], a core-shell CG representation was developed that is transferable with respect to size and enables the simulation of relevant nanoparticle sizes including solvation effects, in the size range where interesting switching in binding behavior is expected.[158] The development of this transferable map was based on matching potentials of mean force (PMFs) for silver nanoparticles obtained using all-atom molecular dynamics.[120] The systematic development of transferable CG representations for the other materials, such as the silica and titanium dioxide that were also considered in atomistic studies, is a future desire. As the determination of binding free energies for CGMD NP by standard methods like umbrella sampling gets prohibitive with increasing size, string methods could be employed as an alternative.[159]

A key requirement in nanosafety assessment is how to include systematically a detailed molecular description of ENM-protein interactions. In biological environments, proteins organize on ENM surfaces forming the so-called nanoparticle protein corona (NPC) structures which play a central role in biological interactions and nanotoxicity.[160, 161, 162] The NPC formation around a variety of nanoparticles was evidenced and characterized in terms of its biochemical composition by several experimental studies,[163, 164]. However, its effects on biological interactions and implications to nanosafety considerations remain largely unknown.[163, 165, 166, 167, 168, 169] Similarly as above, a main research challenge is how to develop efficient yet accurate computational methods and tools that can bridge the gap between a detailed, molecular-level description of ENMs interacting with solvents and biomolecules such as proteins at an atomistic level, and the much larger scale (i.e., tens or hundreds of nanometers to micrometers) corresponding to biological structures such as NPCs or cellular membranes.[170] Coarse-grained computational methods of protein-covered nanoparticles are often limited to modeling entire proteins as single particles. Such models are successful in showing how nanomaterials type, size, and shape can lead to diverse protein composition of the NPC.[171, 172] However, detailed atomistic aspects of modeling protein interactions are required to calculate other key experimentally-relevant mesoscopic descriptors such as the hydrophobic fraction of the solvent accessible surface area ($SASA_H$) (see Figure 6 (a-b-c)). Recent developments in atomistic and multiscale computational methods allow unique opportunities to probe the detailed molecular mechanisms that modulate interactions at bio-nano interfaces.[173, 174, 29] This approach can be extended to the calculation of mesoscopic biophysical descriptors for an NPC and relies on simplified models allowing further computational studies of protein interactions with ENMs (see Figure 6 (a-b-c)).[163]. This has the potential of (i) unveiling the role of specific proteins in NPC's stability and biophysical properties (e.g., hydrophobic surface area, charged patches), and (ii) quantifying the way in which nanoparticle corona properties and protein-protein interactions in the corona are modulated for different nanoparticle types. First, all-atom molecular dynamics (MD) simulations of key plasma proteins (e.g., human serum albumin, fibrinogen, immunoglobulin gamma-1 chain-C, complement C3, and apolipoprotein A1) can be used to study adsorption on typical nanoparticle surfaces (e.g., titanium dioxide or silica). For binary protein-protein interactions (e.g., only two interacting proteins) it is possible to perform exhaustive atomistic MD simulations, both in the vicinity

of nanoparticle surfaces and in bulk to compare directly the results and infer the influence of the presence of the specific nanoparticles on the dynamic and thermodynamics aspects of protein-protein interactions. Figure 6 (c) illustrates the possibility to identify residues crucial to protein-nanoparticle interactions in a specific system (human serum albumin-titanium dioxide). In the second stage, the molecular mechanisms of protein-nanoparticle interactions are probed by looking at the dynamic and structural proteins of several proteins (and possibly lipids) in the crowded environment of nanoparticle coronas, using also molecular docking simulations and, depending on systems size, coarse-grained simulations of mixtures of multiple proteins[175, 176] that can investigate the formation of the protein layer on the nanoparticle surface, as illustrated in Figure 6 (d-e-f). Preliminary studies on multi-protein docking on nanoparticles, suggest that knowledge of protein composition and conformations (e.g., refined from MD simulations) can be used to estimate the overall biophysical properties of NPCs, such as the hydrophobic fraction of their solvent-accessible surface area, and surface charge distributions.[177] Outstanding challenges in modeling the interactions of biological molecules in contact with ENMs are to extend the capability of docking programs to include more than just a few tens of proteins,[178, 179, 180, 181] and to include detailed information on the specific corona molecular composition that is seldom available.[170] Additionally, besides proteins, it is expected that future studies will also include (i) other components such as lipids[182] and glycans[183] which play pivotal roles in ENMs uptake and could also be key for the modeled ENM systems, as well as (ii) an accurate description of the corresponding surface functionalization.[184] Finally, metallic nanoparticles deserve a special mention as they are routinely employed in cancer therapy, where they need to be selectively delivered to the tumor tissues. Among all metals, gold nanoparticles are widely used as radiosensitizing agents because of their biocompatibility and simplicity of synthesis.[185] Coated metallic nanoparticles are used in catalysis, self-assembly, imaging, drug delivery, and sensing applications. Metallic nanoparticles are very sensitive to the local environment because of a phenomenon called “localized surface plasmon resonance” deriving from the collective oscillation of surface electrons.[186] When coated with a monolayer ligand, the metallic nanoparticles’ properties such as metal reduction and colloidal stability can be adjusted for the desired application.[187] Atomistic and mesoscale simulations allowed an understanding of the atypical distribution of multiple ligands on gold and silver nanoparticles observed in the experiment,[188], as well as the adsorption

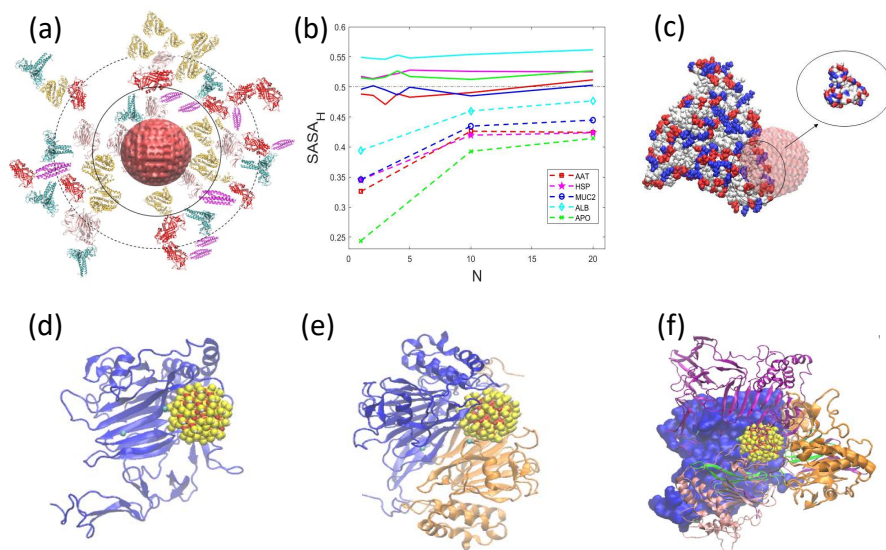


Figure 6: (a) Schematic atomistic model of an NPC. A “soft corona” layer (dashed line) of loosely bound proteins surrounds a “hard” corona layer (continuous black line), proximal to the ENM’s surface. Even for spherical nanoparticles, the overall shape and biophysical properties of the NPC surface will depend on its composition. (b) Comparing values of the $SASA_H$ for various coronas around a 4 nm spherical silica nanoparticle (dashed) with the values calculated for similar protein aggregates without including a nanoparticle (continuous). (c) Atomistic corona models allow the identification of protein residues that may play significant roles at the nanoparticle-protein interfaces. (d-e-f) Building an atomistic model of an NPC by sequential docking of protein structures (mucin, pre-equilibrated using MD) on a spherical silica nanoparticle. Mesoscopic descriptors such as $SASA_H$ can be estimated as statistical averages over results from docking multiple representative structures.

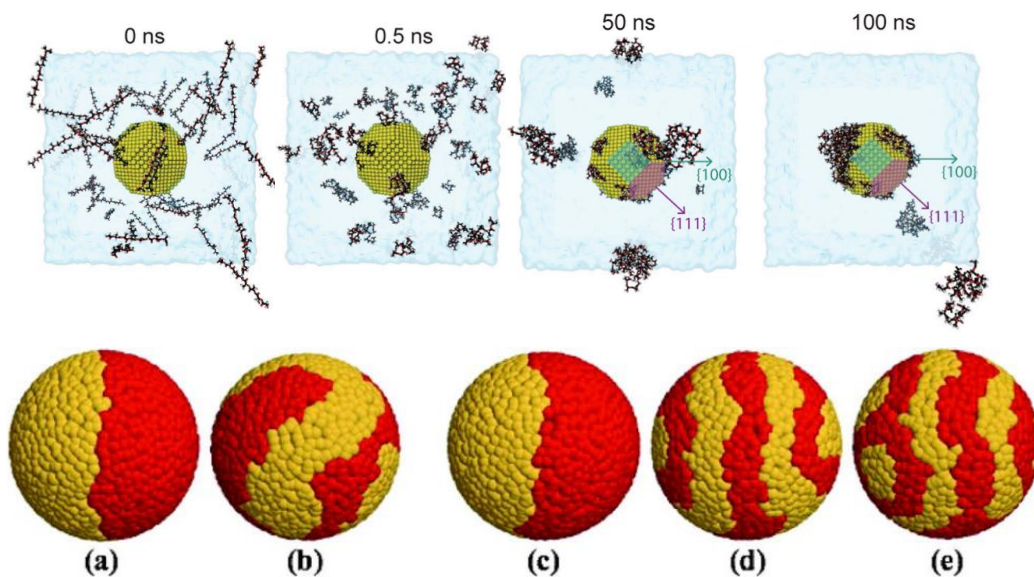


Figure 7: Upper panel: MD snapshots of self-assembly simulations of 60 PLGAs and the Au nanoparticle in aqueous solution at 0, 0.5, 50, and 100 ns, reproduced with permission from Cappabianca et al., ACS Omega, 2022[190]; Bottom panel: Equilibrium structures obtained by mesoscale simulations of self-assembly of binary mixtures of surfactants with varying length difference or bulkiness difference on a spherical nanoparticle. Dark (red) beads and light (yellow) beads represent head groups of the two species of surfactants, from Ref.[188]. [See text for more details.](#)

of biomolecules on gold nanoparticles of different sizes,[189]. We report in Figure 7 an example of [how molecular modelling simulations](#) (here all-atom molecular dynamics simulations) can be used to investigate the behavior of metallic nanoparticles, in particular, to capture adsorption phenomena of polymers. Interestingly in [190], it is clearly shown that the design of shape and topology of the surface in metallic nanoparticles is a rather effective strategy to control the preferential polymer coating in some particle regions as compared to others. Furthermore, using similar simulations tools, other works [188] have investigated precise patterning of coadsorbed surfactants on silver and gold nanoparticles. Hence, by targeting more effective control on the crystallographic features of metallic nanoparticles, we can envision an improved control of ENMs coating thus also significantly affecting their toxicological properties.

Modelling technique	Recommendations for future research
Quantum computations (DFT and DFTB)	Those are among the most time-consuming methods for extracting advanced descriptors. The most recent methodologies based on machine learning algorithms are promising to expand the set of systems, but attention should be paid to their robustness and the energetic needs associated with massive calculations. Pluridisciplinary physico-chemical approaches, such as those widely used in other technological fields like catalysis, should be used on a regular basis, to capture toxicity mechanisms on the molecular level. This may lead to new descriptors and increase the predictive power.
Reactive atomistic simulations	Suitable reactive force fields are still to be developed for accurately predicting the development of ENMs surface electrical charges. In this respect, reaxFF [191] or the more recent Machine Learning based potentials [192] are expected to likely play an important role for coping with sufficiently large particles beyond capacity of standard DFT and DFTB computations.
Classical All-Atom molecular dynamics (MD)	Calculate the adsorption affinity of small metabolites on nanoparticles using enhanced sampling techniques such as metadynamics. Molecular nano-descriptors in [124] have proved very computationally effective to cope with large size particles, nonetheless they are limited to non-metallic particles: Additional effort should be spent to relax this constraint. MD is also needed for molecular docking of proteins to ENM surfaces (see below). To speed up future calculations, an MD database of typical conformations and their corresponding thermodynamic weights could be pre-built for proteins that play a major role in interacting with ENMs (i.e., albumin, mucin, etc.) in order to increase the accuracy of docking-based calculations of nano-descriptors (e.g., $SASA_H$).

magentaProtein docking	Current docking software is optimised for protein-drug and protein-protein interactions (PPIs). Future research would benefit from optimization for sampling efficiently and accurately protein-ENM surfaces (i.e., inorganic materials). Future docking programs should be able to handle multiple molecules, including larger proteins with diverse conformations and in a much larger number than currently possible (i.e., from tens to hundreds or even thousands). One possible approach to bridge the gap between the limitations of docking programs and the complex and large-scale nature of PPIs and protein-ENM interactions is the development of advanced machine learning approaches.
Coarse-Grained Molecular Dynamics (CGMD)	Estimating absorption energy densities by AAMD for pre-screening by continuum Helfrich methods. Calculating CG PMFs for all possible core-shell combinations to generate a matrix for mapping material-specific atomistic PMFs. When addressing extrinsic descriptors for ENMs-cell membrane binding, overcoming the current passive configurations, by adding many lipid types rendering the cellular membranes as well as membrane proteins and the cytoskeleton.
Brownian Dynamics	Simulating nanoparticles in contact with biological molecules to bridge the gap between the molecular modelling scale and the experimental scale. More effective approaches are needed to compute and use PMFs in case of non-spherical or anisotropic particles.

Table 2: Summary of recommendations and possible future research avenues in the context of using material modeling techniques for computing advanced descriptors.

5. Discussion: Challenges and perspectives

On one hand, from the above overview, it clearly emerges that the accuracy in predicting possible hazard of ENMs critically relies upon the ability of using features beyond what can be accessed in typical experimental test and characterization. Importantly, those features depend on intrinsic and extrinsic properties aiming at describing both materials and biological environments.

Nonetheless, when it comes to the direct link of ENMs to their expected toxicity endpoints, it is fair to say that the current status of development of the hardware and algorithms does not allow a brute-force assessment of nanosafety. A more viable approach is expected to be the calculation of intrinsic and extrinsic advanced descriptors using a plethora of methodologies mostly developed and used in other fields (e.g. materials modeling and biochemistry) to extract input features for data based or statistical models (e.g. QSAR, ML algorithms) to finally link them to the toxicity endpoints.

However, even the latter approach comes with formidable challenges, mostly associated to the size and complexity of ENMs of practical use. On one hand, ENMs of experimental interest may have dimensions orders of magnitude larger than their computationally affordable counterparts. On the other hand, precise compositions of particles and their coating are often known with little detail level: This calls for a crucial effort of the relevant scientific community in future experimental works where, in addition to the valuable measurement of toxicological endpoints, a more comprehensive characterization beyond nominal values of ENMs is requested.

Furthermore, currently, an interesting (and perhaps necessary) approach seems to be *hybridization* of pure physics-based models with disparate data sources. In particular, due to a practically unlimited number of different ENMs with great chemical and geometrical variety, a truly extended adoption of advanced descriptors in data-based models for nanoinformatics looks inconceivable without leveraging high-fidelity multiscale modeling data from both literature and crude or analytical (yet computationally efficient) approximations models. As a representative example, biased classical molecular dynamics simulations can certainly be used to accurately compute the Potential of Mean Force between nanoparticle pairs. At the same time, classical approaches such as the theory of Derjaguin-Landau-Verwey-Overbeek (DLVO) cannot be discarded and efforts should be devoted to finding new approaches capable of orchestrating and integrating such multi-fidelity and

multi-source data. Another example of a similar synergy has been described above in the manuscript and it has to do with CGMD based simulations of cell membranes and the corresponding continuum Helfrich free energy models.

One possibility would be the adoption of descriptors from crude approximation models, characterized by a lower fidelity level and available literature data as a subset of features for ML models where physics-based model results are used as training sets. Such an approach has been proven successful in significantly improving ML model predictions in the presence of a small or incomplete training dataset.[11, 193] We thus envision a more comprehensive and multi-layered approach where detailed physics-based models (first layer), literature/experiment data (second layer), and crude approximation models (third layer) are synergetically managed for the estimate of relevant descriptors by means of surrogate (statistical) and ML models (see Figure 1, bottom panel). In this respect, further research is requested to investigate to what extent the small variance of high-fidelity data from physics-based model predictions combined with the high-variance (and low-fidelity) of other sources can deliver descriptors predictions with high/medium fidelity and variance that are suitable for QSAR/QSPR models.

It is also worth stressing that, beyond the mere usefulness of advanced descriptors for data based models targeting nanosafety assessment, further improved physics-based models describing phenomena from the electronic up to the mesoscopic level can offer the unique opportunity of gaining insights at the nano-level, which are hardly accessible by experimental techniques and therefore remain critical for: i) unveiling basic mechanisms behind the possible hazardous character of ENMs; ii) possibly complement information from less detailed and less demanding models.

For all the above reasons, we expect and hope that the progress of high-performance computing power combined with advanced tools for acceleration of atomistic simulations [194] can soon make the calculation of advanced descriptors more competitive as compared to experiments in terms of both the amount and quality of generated data.

Another interesting aspect to highlight is related to the possible exploitation of the large body of knowledge available in the context of modern computational approaches to investigate nanoparticles in other technological fields such as catalysis and materials science [195, 196, 197, 76, 198, 199, 200]. Specifically, we refer to:

1. High-throughput screening of nanoparticles in drug delivery [201]
2. Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics,[202].
3. Computational high-throughput screening of alloy nanoclusters for electrocatalytic hydrogen evolution [203]

It would be desirable that the safe and sustainable design strategies in toxicology could take advantage of this knowledge and tools, combining efficiently biology and medicine with physics and chemistry.

Furthermore, it is likely that an even larger amount of data on nanosafety of ENMs will be generated in the near future. It is therefore of increasing importance to comply with the FAIR principles, so that metadata and data can be efficiently reused in data-based models for predicting the hazard of ENMs. Specifically, in this respect, the role of databases for collecting and storing a large amount of curated and well structured data is likely to play a major role in the near future. As such, we discuss one prototypical case study in the following subsection.

Finally, as a long-term goal, we expect that nanoinformatics models based on advanced descriptors could be integrated with AOPs, in order to better assess the potential exposure throughout the entire life of the nanoparticles. In this respect, modern grouping strategies taking into account the mode of actions and developed based on ML techniques processing data from omics studies appear particularly promising and therefore it will be specifically discussed below in a dedicated subsection.

5.1. Databases: The eNanoMapper case study

Storage and organization as well as organization of curated data from disparate sources within databases appears critical for nanosafety assessment. To this end, below we review a prototypical case study. The eNanoMapper database is an open-source chemical substance data management solution,[204] adopted by more than 20 European projects and facilitating the Findable, Accessible, Interoperable and Reusable (FAIR) data collection and reuse of the nanosafety community. To provide aggregated findability, accessibility, and interoperability across project-specific databases, the Nanosafety Data Interface (<https://search.data.enanomapper.net>) was created, and currently represents one of the largest searchable nanosafety data collections.[205] The eNanoMapper is based on data and software originally developed to represent industrial chemicals and related experimental or calculated data. It

was one of the first cheminformatics platforms to offer open REST Application Programming Interface (API) supporting integrated services such as data, descriptor calculations, and ML.[206, 207] A visual representation of the eNanoMapper data model is reported in Figure 8, where substances are characterized by names and IDs, which can be multiple, the composition refers to the components of the material (core, coating, chemical structure), each of them having different properties; the same material can have different compositions. A protocol consists in measurements of a specific endpoint in given conditions, related protocols form an investigation entity; finally, different substances can be grouped into an assay entity when the same protocol applies, giving an extremely flexible structure.[208] With the explosive growth of material databases, ML frameworks, and their success in material modeling, it is critical to explore the link between the estimated material properties and experimentally measured safety or functional properties. In NanoInformaTIX,[36] both experimental data from selected use cases and also calculated descriptors are stored into the eNanoMapper database and an effort is devoted towards providing open source libraries to facilitate integration with data analysis frameworks and developing exploratory data analysis methods. The validation of computational models relies on high-quality experimental data; such data may not always be complete, and it is necessary to identify data gaps and, eventually, to generate additional data based on experimental and theoretical chemistry and on biology. The goal of this computational and theoretical endeavor is the realization of safe nanoparticles and nanomaterials. To fulfil such expectation, theoretical descriptors must be translated into measurable parameters, and this challenge entails a deep knowledge of the mode of actions of nanoparticles that lead to harmful outcomes. On another level, identification and estimation of theoretical drivers of toxicity can facilitate the prioritization of experimental tests, which, due to the uncontrollable inhomogeneity of any ensemble of nanoparticles, is always needed for a robust risk assessment of nano-enabled technology.

5.2. Grouping approaches

Regulatory processes, which often rely on *in vivo* testing, are outpaced by the increasing number of ENMs on the market. To cope with this situation, the lack of data and to ensure the safety of new materials, grouping approaches emerge as an interesting method. Those approaches are accepted within the overarching EU chemicals regulation REACH (EC 1907/2006)

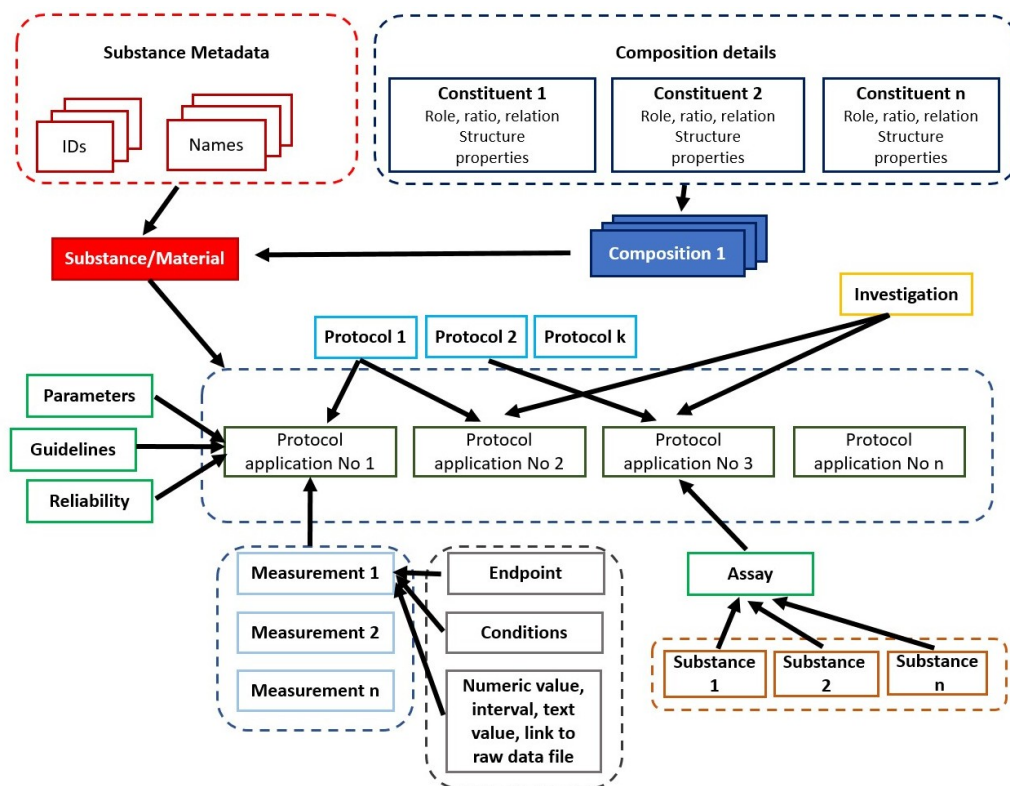


Figure 8: eNanoMapper/Ambit data model adapted from Ref.[208]. Substances are characterized by their composition and identified by name and ID, which can be multiple. Complex relations between the substance components can be specified.

and are commonly used to consider more than one chemical at the same time.[209, 210, 211] Within an established group, data gaps can be filled by read-across. Here, existing data on a particular (eco)toxicological endpoint linked to one or several source chemicals can be employed to estimate the same property of one or more target chemical(s).

Chemicals can be grouped together on the basis of well-defined physicochemical similarities, like common functional groups, precursors, and/or breakdown products. However, ENMs pose an additional challenge compared to chemicals, since there is a very limited understanding on how individual physicochemical parameters influence cellular uptake and toxicity. In addition, the properties of ENMs can change depending on the surrounding medium and over time. To establish grouping approaches for ENMs, it is essential to understand how the individual physicochemical properties are linked to toxicity. To support a grouping justification, additional information on a common Mode of Action (MoA) or toxicity mechanism is advantageous.[209] The MoA of a substance describes the functional or physiological changes it causes to a living organism or cell. One of the drawbacks is that toxicity mechanisms are only partially understood, and for plenty of ENMs variants the precise MoA remains elusive. In this regard, the potential of systems biology to contribute to the development of reliable grouping approaches for ENMs should not be obliterated. Modern omics-based approaches (transcriptomics, proteomics, metabolomics), in combination with sophisticated bioinformatics and data analysis tools, are very important to characterize toxicity pathways and unravel relationships between individual physicochemical properties and cellular responses. This knowledge can then be applied in the context of grouping and categorization. Moreover, omics approaches are of relevance for the development of AOPs and to establish reliable, comprehensive testing strategies building on the known MoA.[212] AOPs are a conceptual construct that integrates known information from various sources in a sequential chain of causally linked key events that cover different levels of biological organization (i.e. cellular, organ level) starting with a molecular initiating event leading to the final adverse outcome.[213] The knowledge gained from grouping approaches can then be directly used for SSbD of ENMs. Currently, there are several ENM grouping frameworks with different approaches.[214, 215, 216]. However, there are only very few case studies for which the frameworks have been applied to ENMs.

Recently, several publications have taken advantage of different techniques, including ML approaches. [217, 218, 219] Additionally, the incorpora-

tion of omics datasets to support the development of more accurate grouping strategies of ENMs take into account the mode of action [220, 221, 222, 223].

Bioinformatics and ML techniques are thus essential when approaching cellular effects in a comprehensive manner, particularly in combination with high-throughput techniques like omics. [Figure 9 depicts the ML random forest approach used for the grouping of ENMs, with this strategy, the biological activity of ENMs can be predicted provided that physicochemical properties are known.](#) Omics studies are a massive source of data sets, which comprehensively describe the cellular alterations caused by any treatment, importantly in this case, by ENMs. Thus, omics methods are highly useful to identify the MoA of ENMs to be employed in the grouping approaches as additional biological descriptors. So far, most of the efforts to understand the MoA of ENMs have been undertaken in the field of transcriptomics.[224, 225] However, proteomics can be even more informative since it depicts the cellular alterations much closer to the phenotype than transcriptomics. The challenge here is the standardization of the methods, particularly for data analysis and interpretation.[226, 227]

Meta-analysis of publicly available proteome data targeted to specific organ alterations is being carried out to investigate the MoA of ENMs within the organ, based on proteome alteration evidence. Relevant datasets from publicly accessible proteomics databases such as PRIDE are identified in the first step. These data do not necessarily involve only ENM treatments, but also other alterations like disease, cancer, and chemical treatments. A recently developed workflow by the BfR for standardized data analysis was applied to the data sets. These results are then integrated into proteomics data *de novo* generated from studies evaluating the effect of ENMs in vitro. Correlations between nanomaterials effects and organ-specific alterations can thus be detected.[228]

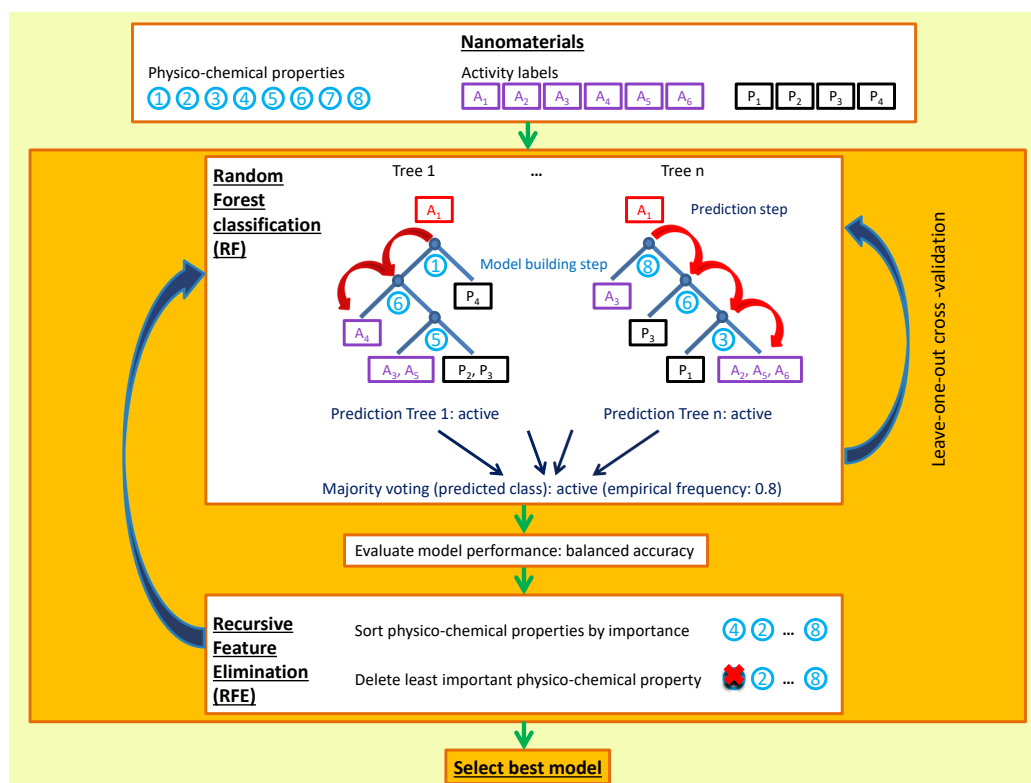


Figure 9: Schematic representation of the random forest approach used for grouping of ENMs. ENMs are described by a set of different physicochemical properties, and an activity label is assigned to each ENM (based on the outcome of biological assays). ENMs are then subjected to a random forest model, in which ENMs are divided into active and passive materials, based on splits made on their physicochemical properties in each tree. Recursive feature elimination is used to select only the most important physicochemical properties for achieving maximum accuracy. To address and avoid a large overfitting bias, the model is validated in a leave-one-out approach. This approach can be used to predict the biological activity of new ENMs, for which the physicochemical properties are known.

6. Concluding remarks

Gaining a clear and deep rationale behind the safety properties of [nanomaterials](#) is a multifaceted issue still posing formidable challenges: Nanomaterials are far more complex than their macroscopic counterparts, having high surface reactivity and the ability to enter living cells, potentially causing damage to the organisms.

In this work, mostly taking a computational perspective, we made an effort in analyzing state of the art physics-based models for computing both intrinsic and extrinsic ENMs properties that are crucial for setting up reliable data-based models for nanosafety assessment. In this spirit, one major aim of this work was the identification of the most critical roadblocks towards computer-aided support nanosafety assessment. Importantly, we have identified and discussed opportunities in advancing the field and open research points, as conveniently summarized in Tables 1 and 2. We have extensively illustrated the methods for computing advanced descriptors and the current associated challenges mainly leading to low data variance despite the expected higher fidelity. Hence, possible suggestions on hybridization strategies for moving towards models with higher data variance and fidelity, as well as the inclusion of Adverse Outcome Pathways (AOPs) are envisioned.

We stress that the reported analysis and suggested guidelines for future reflect our current best understanding of the field after several years of discussion among experts in disparate related fields. As such, our main hope is that this work can serve as a stimulus for future research, and could thus help triggering next breakthroughs in nanosafety assessment.

7. Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 814426 (NanoInformaTIX project).

8. Data availability

This is a review article and data availability does not apply

9. Conflicts of Interest

Authors have no conflict of interest to declare

References

- [1] D. Astruc, Introduction: nanoparticles in catalysis (2020).
- [2] A. Gizzatov, J. Key, S. Aryal, J. Ananta, A. Cervadoro, A. L. Palange, M. Fasano, C. Stigliano, M. Zhong, D. Di Mascolo, et al., Hierarchically structured magnetic nanoconstructs with enhanced relaxivity and co-operative tumor accumulation, *Advanced functional materials* 24 (29) (2014) 4584–4594.
- [3] A. Cardellini, M. Fasano, E. Chiavazzo, P. Asinari, Interfacial water thickness at inorganic nanoconstructs and biomolecules: Size matters, *Phys. Lett. Sect. A Gen. At. Solid State Phys.* 380 (20) (2016) 1735–1740.
URL <http://dx.doi.org/10.1016/j.physleta.2016.03.015>
- [4] S. S. Mukhopadhyay, Nanotechnology in agriculture: prospects and constraints, *Nanotechnology, science and applications* (2014) 63–71.
- [5] B. S. Sekhon, Food nanotechnology—an overview, *Nanotechnology, science and applications* (2010) 1–15.
- [6] E. Chiavazzo, P. Asinari, Enhancing surface heat transfer by carbon nanofins: towards an alternative to nanofluids?, *Nanoscale research letters* 6 (1) (2011) 1–13.

- [7] E. Chiavazzo, P. Asinari, Reconstruction and modeling of 3d percolation networks of carbon fillers in a polymer matrix, *International journal of thermal sciences* 49 (12) (2010) 2272–2281.
- [8] A. Haase, F. Klaessig, EU US Roadmap Nanoinformatics 2030 (2018) 1–126.
- [9] A. Cardellini, M. Alberghini, A. Govind Rajan, R. P. Misra, D. Blankschtein, P. Asinari, Multi-scale approach for modeling stability, aggregation, and network formation of nanoparticles suspended in aqueous solutions, *Nanoscale* 11 (9) (2019) 3925–3932.
- [10] Making nano data FAIR enough, *Nat. Nanotechnol.* 16 (6) (2021) 607.
- [11] Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science, *npj Comput. Mater.* 4 (1) (2018) 28–33.
- [12] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Chemicals Strategy for Sustainability Towards a Toxic-Free Environment, in: COM/2020/667 Final, 2020.
- [13] S. F. Bezerra, B. dos Santos Rodrigues, A. C. G. da Silva, R. I. de Ávila, H. R. G. Brito, E. R. Cintra, D. F. M. C. Veloso, E. M. Lima, M. C. Valadares, Application of the adverse outcome pathway framework for investigating skin sensitization potential of nanomaterials using new approach methods, *Contact Dermatitis* 84 (2) (2021) 67–74.
- [14] K. Jagiello, K. Ciura, In vitro to in vivo extrapolation to support the development of the next generation risk assessment (NGRA) strategy for nanomaterials, *Nanoscale* 14 (18) (2022) 6735–6742.
- [15] K. Jagiello, S. Halappanavar, A. Rybińska-Fryca, A. Williams, U. Vogel, T. Puzyn, Transcriptomics-Based and AOP-Informed Structure–Activity Relationships to Predict Pulmonary Pathology Induced by Multiwalled Carbon Nanotubes, *Small* 17 (15) (2021).
- [16] K. Jagiello, B. Judzinska, A. Sosnowska, I. Lynch, S. Halappanavar, T. Puzyn, Using AOP-Wiki to support the ecotoxicological risk assessment of nanomaterials: first steps in the development of novel adverse outcome pathways, *Environ. Sci. Nano* 9 (5) (2022) 1675–1684.

- [17] C. Caldeira, R. Farcal, C. Moretti, L. Mancini, H. Rauscher, J. Riego Sintes, S. Sala, K. Rasmussen, Safe and sustainable by design chemicals and materials : review of safety and sustainability dimensions, aspects, methods, indicators, and tools, Publications Office of the European Union, 2022.
- [18] B. Stieberova, M. Zilka, M. Ticha, F. Freiberg, P. Caramazana-González, J. McKechnie, E. Lester, Sustainability assessment of continuous-flow hydrothermal synthesis of nanomaterials in the context of other production technologies, *J. Clean. Prod.* 241 (2019).
- [19] A. García-Quintero, M. Palencia, A critical analysis of environmental sustainability metrics applied to green synthesis of nanomaterials and the assessment of environmental risks associated with the nanotechnology, *Sci. Total Environ.* 793 (2021) 148524.
- [20] E. Marcoulaki, J. M. López de Ipiña, S. Vercauteren, J. Bouillard, M. Himly, I. Lynch, H. Witters, N. Shandilya, B. van Duuren-Stuurman, V. Kunz, W. E. Unger, V. D. Hodoroaba, D. Bard, G. Evans, K. A. Jensen, M. Pilou, A. K. Viitanen, A. Bochon, A. Duschl, M. Geppert, K. Persson, I. Cotgreave, P. Niga, M. Gini, K. Eleftheriadis, S. Scalbi, B. Caillard, A. Arevalillo, E. Frejafon, O. Aguerre-Chariol, V. Dulio, Blueprint for a self-sustained European Centre for service provision in safe and sustainable innovation for nanotechnology, *NanoImpact* 23 (2021) 100337.
- [21] L. J. Johnston, N. Gonzalez-Rojano, K. J. Wilkinson, B. Xing, Key challenges for evaluation of the safety of engineered nanomaterials, *NanoImpact* 18 (2020) 100219.
- [22] S. Gottardo, A. Mech, J. Drbohlavová, A. Małyska, S. Bøwadt, J. Riego Sintes, H. Rauscher, Towards safe and sustainable innovation in nanotechnology: State-of-play for smart nanomaterials, *NanoImpact* 21 (2021) 100297.
- [23] P. Nymark, M. Bakker, S. Dekkers, R. Franken, W. Fransman, A. García-Bilbao, D. Greco, M. Gulumian, N. Hadrup, S. Halappanavar, V. Hongisto, K. S. Hougaard, K. A. Jensen, P. Kohonen, A. J.

- Koivisto, M. Dal Maso, T. Oosterwijk, M. Poikkimäki, I. Rodriguez-Llopis, R. Stierum, J. B. Sørli, R. Grafström, Toward Rigorous Materials Production: New Approach Methodologies Have Extensive Potential to Improve Current Safety Assessment Practices, *Small* 16 (6) (2020).
- [24] S. H. Doak, M. J. Clift, A. Costa, C. Delmaar, I. Gosens, S. Halappanavar, S. Kelly, W. J. Peijnenburg, B. Rothen-Rutishauser, R. P. Schins, V. Stone, L. Tran, M. G. Vijver, U. Vogel, W. Wohlleben, F. R. Cassee, The Road to Achieving the European Commission’s Chemicals Strategy for Nanomaterial Sustainability—A PATROLS Perspective on New Approach Methodologies, *Small* 18 (17) (2022).
- [25] L. Bajard, O. Adamovsky, K. Audouze, K. Baken, R. Barouki, J. B. Beltman, A. Beronius, E. C. Bonfeld-Jørgensen, G. Cano-Sancho, M. L. de Baat, F. Di Tillio, M. F. Fernández, R. E. FitzGerald, C. Gundacker, A. F. Hernández, K. Hilscherova, S. Karakitsios, E. Kuchovska, M. Long, M. Luijten, S. Majid, P. Marx-Stoelting, V. Mustieles, C. K. Negi, D. Sarigiannis, S. Scholtz, I. Sovadinova, R. Stierum, S. Tanabe, K. E. Tollefsen, A. D. van den Brand, C. Vogs, M. Wielsøe, C. Wittwehr, L. Blaha, Application of AOPs to assist regulatory assessment of chemical risks – Case studies, needs and recommendations, *Environ. Res.* (2022) 114650.
- [26] S. I. Gomes, J. J. Scott-Fordsmand, M. J. Amorim, Alternative test methods for (nano)materials hazards assessment: Challenges and recommendations for regulatory preparedness, *Nano Today* 40 (2021) 101242.
- [27] C. Westmoreland, H. J. Bender, J. E. Doe, M. N. Jacobs, G. E. Kass, F. Madia, C. Mahony, I. Manou, G. Maxwell, P. Prieto, R. Roggeband, T. Sobanski, K. Schütte, A. P. Worth, Z. Zvonar, M. T. Cronin, Use of New Approach Methodologies (NAMs) in regulatory decisions for chemical safety: Report from an EPAA Deep Dive Workshop, *Regul. Toxicol. Pharmacol.* 135 (2022) 105261.
- [28] OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, 2014.

- [29] A. Afantitis, G. Melagraki, P. Isigonis, A. Tsoumanis, D. Danai, E. Valsami-jones, A. Papadiamantis, L.-j. A. Ellis, H. Sarimveis, P. Doganis, P. Karatzas, P. Tsiros, I. Liampa, V. Lobaskin, D. Greco, A. Serra, P. Anneli, S. Kinaret, L. Aliisa, R. Grafström, P. Kohonen, P. Nymark, E. Willighagen, T. Puzyn, A. Rybinska-fryca, A. Lyubartsev, K. Alstrup, J. Gerit, S. Lofts, C. Svendsen, S. Harrison, D. Maier, K. Tamm, J. Jänes, L. Sikk, M. Dusinska, E. Longhin, E. Rundénpran, E. Mariussen, N. El, W. Unger, J. Radnik, A. Tropsha, Y. Cohen, J. Leszczynski, C. Ogilvie, M. Wiesner, D. Winkler, N. Suzuki, T. Hyun, J.-s. Choi, N. Sanabria, M. Gulumian, I. Lynch, NanoSolveIT Project : Driving nanoinformatics research to develop innovative and integrated tools for in silico nanosafety assessment, *Comput. Struct. Biotechnol. J.* 18 (2020) 583–602.
- [30] S. So, J. Rho, Designing nanophotonic structures using conditional deep convolutional generative adversarial networks, *Nanophotonics* 8 (7) (2019) 1255–1261.
- [31] I. Kim, K. Viswanathan, G. Kasi, K. Sadeghi, S. Thanakkasaranee, J. Seo, Preparation and characterization of positively surface charged zinc oxide nanoparticles against bacterial pathogens, *Microb. Pathog.* 149 (2020) 104290.
- [32] NanoSolveIT, NanoSolveIT Horizon 2020 project, <https://cordis.europa.eu/project/id/814572>.
URL <https://cordis.europa.eu/project/id/814572>
- [33] Gov4Nano, Gov4Nano Horizon 2020 project, <https://cordis.europa.eu/project/id/814401>.
- [34] NANORIGO, NANORIGO Horizon 2020 project, <https://cordis.europa.eu/project/id/814530>.
- [35] RiskGONE, RiskGONE Horizon 2020 project, <https://cordis.europa.eu/project/id/814425>.
- [36] NanoInformaTIX, NanoInformaTIX Horizon 2020 project, <https://cordis.europa.eu/project/id/814426>.

- [37] K. Blekos, K. Chairetakis, I. Lynch, E. Markoulaki, Principles and requirements for nanomaterial representations to facilitate machine processing and cooperation with nanoinformatics tools, *Journal of Cheminformatics* (2022).
- [38] H. Nagai, S. Toyokuni, Biopersistent fiber-induced inflammation and carcinogenesis: lessons learned from asbestos toward safety of fibrous nanomaterials, *Archives of Biochemistry and Biophysics* 502 (1) (2010) 1–7.
- [39] F. S. Bierkandt, L. Leibrock, S. Wagener, P. Laux, A. Luch, The impact of nanomaterial characteristics on inhalation toxicity, *Toxicology research* 7 (3) (2018) 321–346.
- [40] K. Donaldson, F. A. Murphy, R. Duffin, C. A. Poland, Asbestos, carbon nanotubes and the pleural mesothelium: a review of the hypothesis regarding the role of long fibre retention in the parietal pleura, inflammation and mesothelioma, *Particle and fibre toxicology* 7 (1) (2010) 5.
- [41] A. A. Shvedova, E. R. Kisin, R. Mercer, A. R. Murray, V. J. Johnson, A. I. Potapovich, Y. Y. Tyurina, O. Gorelik, S. Arepalli, D. Schwegler-Berry, et al., Unusual inflammatory and fibrogenic pulmonary responses to single-walled carbon nanotubes in mice, *American Journal of Physiology-Lung Cellular and Molecular Physiology* 289 (5) (2005) L698–L708.
- [42] M. V. Park, A. M. Neigh, J. P. Vermeulen, L. J. de la Fonteyne, H. W. Verharen, J. J. Briedé, H. van Loveren, W. H. de Jong, The effect of particle size on the cytotoxicity, inflammation, developmental toxicity and genotoxicity of silver nanoparticles, *Biomaterials* 32 (36) (2011) 9810–9817.
- [43] D. J. Smith, L. G. Leal, S. Mitragotri, M. S. Shell, Nanoparticle transport across model cellular membranes: When do solubility-diffusion models break down?, *J. Phys. D. Appl. Phys.* 51 (29) (2018).
- [44] R. W. Homer, J. Swanson, R. J. Jilek, T. Hurst, R. D. Clark, SYBYL line notation (SLN): A single notation to represent chemical structures,

- queries, reactions, and virtual libraries, *J. Chem. Inf. Model.* 48 (12) (2008) 2294–2307.
- [45] S. J. Coles, N. E. Day, P. Murray-Rust, H. S. Rzepa, Y. Zhang, Enhancement of the chemical semantic web through the use of InChI identifiers, *Org. Biomol. Chem.* 3 (10) (2005) 1832–1834.
- [46] I. Lynch, A. Afantitis, T. Exner, M. Himly, V. Lobaskin, P. Doganis, D. Maier, N. Sanabria, A. G. Papadiamantis, A. Rybinska-fryca, M. Gromelski, T. Puzyn, E. Willighagen, B. D. Johnston, M. Gulumian, M. Matzke, A. G. Etxabe, N. Bossa, A. Serra, I. Liampa, S. Harper, K. Tämm, A. C. Jensen, P. Kohonen, L. Slater, A. Tsoumanis, D. Greco, D. A. Winkler, H. Sarimveis, G. Melagraki, Can an inchi for nano address the need for a simplified representation of complex nanomaterials across experimental and nanoinformatics studies?, *Nanomaterials* 10 (12) (2020) 1–44.
- [47] E. Wyrzykowska, A. Mikolajczyk, I. Lynch, N. Jeliaskova, N. Kochev, H. Sarimveis, P. Doganis, P. Karatzas, A. Afantitis, G. Melagraki, A. Serra, D. Greco, J. Subbotina, V. Lobaskin, M. A. Bañares, E. Valsami-jones, K. Jagiello, T. Puzyn, Representing and describing nanomaterials in predictive nanoinformatics, *Nat. Nanotechnol.* 17 (2022) 924–932.
- [48] M. Swirog, A. Mikolajczyk, K. Jagiello, J. Jänes, K. Tämm, T. Puzyn, Predicting electrophoretic mobility of TiO₂, ZnO, and CeO₂ nanoparticles in natural waters: The importance of environment descriptors in nanoinformatics models, *Sci. Total Environ.* 840 (2022) 1–7.
- [49] M. Fronzi, R. D. Amos, R. Kobayashi, N. Matsumura, K. Watanabe, R. K. Morizawa, Evaluation of machine learning interatomic potentials for the properties of gold nanoparticles, *Nanomaterials* 12 (21) (2022). doi:10.3390/nano12213891.
- [50] H. Li, H. Tian, Y. Chen, S. Xiao, X. Zhao, Y. Gao, L. Zhang, Analyzing and predicting the viscosity of polymer nanocomposites in the conditions of temperature, shear rate, and nanoparticle loading with molecular dynamics simulations and machine learning, *The Journal of Physical Chemistry B* 0 (0) (0) null, PMID: 37020173. arXiv:<https://doi.org/10.1021/acs.jpcc.3c01697>, doi:10.1021/acs.jpcc.3c01697.

- [51] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H. M. Hwang, A. Toropov, D. Leszczynska, J. Leszczynski, Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles, *Nat. Nanotechnol.* 6 (3) (2011) 175–178.
- [52] A. A. Toropov, A. P. Toropova, E. Benfenati, G. Gini, T. Puzyn, D. Leszczynska, J. Leszczynski, Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli*, *Chemosphere* 89 (9) (2012) 1098–1102.
- [53] A. A. Toropov, A. P. Toropova, G. Raitano, E. Benfenati, CORAL: Building up QSAR models for the chromosome aberration test, *Saudi J. Biol. Sci.* 26 (6) (2019) 1101–1106.
- [54] N. Sizochenko, B. Rasulev, A. Gajewicz, V. Kuz'min, T. Puzyn, J. Leszczynski, From basic physics to mechanisms of toxicity: the “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles, *Nanoscale* 6 (22) (2014) 13986–13993.
- [55] N. Sizochenko, K. Jagiello, J. Leszczynski, T. Puzyn, How the “Liquid Drop” Approach Could Be Efficiently Applied for Quantitative Structure–Property Relationship Modeling of Nanofluids, *J. Phys. Chem. C* 119 (45) (2015) 25542–25547.
- [56] R. Slapikas, I. Dabo, S. B. Sinnott, Surface reconstruction of oxidized platinum nanoparticles using classical molecular dynamics simulations, *Comput. Mater. Sci.* 209 (2022) 111364.
- [57] D. L. Liao, B. Q. Liao, Shape, size and photocatalytic activity control of TiO₂ nanoparticles with surfactants, *J. Photochem. Photobiol. A Chem.* 187 (2-3) (2007) 363–369.
- [58] G. Mancardi, M. Alberghini, N. Aguilera-Porta, M. Calatayud, P. Asinari, E. Chiavazzo, A multi-scale modelling of aggregation of TiO₂ nanoparticle suspensions in water, *MDPI Nanomater.* 12 (217) (2022) 1–18.
- [59] A. Mikolajczyk, A. Gajewicz, B. Rasulev, N. Schaeublin, E. Maurer-Gardner, S. Hussain, J. Leszczynski, T. Puzyn, Zeta potential for

- metal oxide nanoparticles: A predictive model developed by a nano-quantitative structure-property relationship approach, *Chem. Mater.* 27 (7) (2015) 2400–2407.
- [60] S. Ortelli, A. L. Costa, M. Blosi, A. Brunelli, E. Badetti, A. Bonetto, D. Hristozov, A. Marcomini, Colloidal characterization of CuO nanoparticles in biological and environmental media, *Environ. Sci. Nano* 4 (2017) 1264–1272.
- [61] S. Ortelli, G. Malucelli, F. Cuttica, M. Blosi, I. Zanoni, A. Luisa, Coatings made of proteins adsorbed on TiO₂ nanoparticles : a new flame retardant approach for cotton fabrics, *Cellulose* (2018).
- [62] S. Ortelli, G. Malucelli, M. Blosi, I. Zanoni, A. Luisa, Journal of Colloid and Interface Science NanoTiO₂@DNA complex: a novel eco , durable , fire retardant design strategy for cotton textiles, *J. Colloid Interface Sci.* 546 (2019) 174–183.
- [63] S. Ortelli, A. L. Costa, I. Zanoni, M. Blosi, O. Geiss, I. Bianchi, D. Mehn, F. Fumagalli, G. Ceccone, G. Guerrini, L. Calzolari, Colloids and Surfaces B : Biointerfaces TiO₂ @ BSA nano-composites investigated through orthogonal multi-techniques characterization platform, *Colloids Surfaces B Biointerfaces* 207 (2021) 112037.
- [64] N. Sizochenko, A. Mikolajczyk, M. Syzochenko, T. Puzyn, J. Leszczynski, Zeta potentials (ζ) of metal oxide nanoparticles: A meta-analysis of experimental data and a predictive neural networks modeling, *NanoImpact* 22 (December 2020) (2021) 100317.
- [65] C. D. Walkey, J. B. Olsen, F. Song, R. Liu, H. Guo, D. W. H. Olsen, Y. Cohen, A. Emili, W. C. Chan, Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles, *ACS Nano* 8 (3) (2014) 2439–2455.
- [66] I. Rouse, D. Power, E. G. Brandt, M. Schneemilch, K. Kotsis, N. Quirke, A. P. Lyubartsev, V. Lobaskin, Advanced in silico characterization of nanomaterials for nanoparticle toxicology (2020). [arXiv: 2007.04017](https://arxiv.org/abs/2007.04017).
- [67] S. I. Gomes, M. J. Amorim, S. Pokhrel, L. Mädler, M. Fasano, E. Chiavazzo, P. Asinari, J. Jänes, K. Tämm, J. Burk, et al., Machine learning

and materials modelling interpretation of in vivo toxicological response to tio 2 nanoparticles library (uv and non-uv exposure), *Nanoscale* 13 (35) (2021) 14666–14678.

- [68] N. Chawla, Data Mining for Imbalanced Datasets: An Overview, in: O. Maimon, L. Rokach (Eds.), *Data Min. Knowl. Discov. Handb.*, Springer US, 2010, pp. 875–886.
- [69] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera, *Learning from Imbalanced Data Sets*, Springer Cham, 2018.
- [70] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, G. Bing, Learning from class-imbalanced data: Review of methods and applications, *Expert Syst. Appl.* 73 (2017) 220–239.
- [71] H. B. Lee, H. Lee, D. Na, S. Kim, M. Park, E. Yang, S. J. Hwang, Learning to Balance: Bayesian Meta-Learning for Imbalanced and Out-of-distribution Tasks, in: *ICLR 2020*, 2020.
- [72] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.
- [73] P. Branco, L. Torgo, R. P. Ribeiro, A Survey of Predictive Modelling under Imbalanced Distributions (2015). [arXiv:1505.01658](https://arxiv.org/abs/1505.01658).
- [74] A. Fernández, S. García, F. Herrera, N. V. Chawla, SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary, *J. Artif. Intell. Res.* 61 (2018) 863–905.
- [75] D. Dablain, B. Krawczyk, N. V. Chawla, DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data, *IEEE Trans. Neural Networks Learn. Syst.* (2021).
- [76] T. Zhao, X. Zhang, S. Wang, GraphSMOTE: Imbalanced Node Classification on Graphs with Graph Neural Networks, *WSDM 2021 - Proc. 14th ACM Int. Conf. Web Search Data Min.* (2021) 833–841.
- [77] M. Buda, A. Maki, M. A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, *Neural Networks* 106 (2018) 249–259.

- [78] J. M. Johnson, T. M. Khoshgoftaar, Survey on deep learning with class imbalance, *J. Big Data* 6 (1) (2019) 1–54.
- [79] A. V. Zakharov, M. L. Peach, M. Sitzmann, M. C. Nicklaus, QSAR modeling of imbalanced high-throughput screening data in PubChem, *J. Chem. Inf. Model.* 54 (3) (2014) 705–712.
- [80] J. Liu, K. Mansouri, R. S. Judson, M. T. Martin, H. Hong, M. Chen, X. Xu, R. S. Thomas, I. Shah, Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure, *Chem. Res. Toxicol.* 28 (4) (2015) 738–751.
- [81] K. Klimenko, S. A. Rosenberg, M. Dybdahl, E. B. Wedebye, N. G. Nikolov, QSAR modelling of a large imbalanced aryl hydrocarbon activation dataset by rational and random sampling and screening of 80,086 REACH pre-registered and/or registered substances, *PLoS One* 14 (3) (2019) e0213848.
- [82] N. Sturm, A. Mayr, T. Le Van, V. Chupakhin, H. Ceulemans, J. Wegner, J. F. Golib-Dzib, N. Jeliazkova, Y. Vandriessche, S. Böhm, V. Cima, J. Martinovic, N. Greene, T. Vander Aa, T. J. Ashby, S. Hochreiter, O. Engkvist, G. Klambauer, H. Chen, Industry-scale application and evaluation of deep learning for drug target prediction, *J. Cheminform.* 12 (1) (2020) 1–13.
- [83] M. Kotzabasaki, I. Sotiropoulos, C. Charitidis, H. Sarimveis, Machine Learning Methods for Multi-Walled Carbon Nanotubes (MWCNT) Genotoxicity Prediction, *Nanoscale Adv.* 3 (11) (2021) 3167–3176.
- [84] N. Sizochenko, M. Syzochenko, N. Fjodorova, B. Rasulev, J. Leszczynski, Evaluating Genotoxicity of Metal Oxide Nanoparticles: Application of Advanced Supervised and Unsupervised Machine Learning Techniques, *Ecotoxicol. Environ. Saf.* 185 (2019) 109733.
- [85] D. Fourches, D. Pu, C. Tassa, R. Weissleder, S. Y. Shaw, R. J. Mumper, A. Tropsha, Quantitative Nanostructure-Activity Relationship Modeling, *ACS Nano* 4 (10) (2010) 5703–5712.
- [86] J. M. Gernand, E. A. Casman, A Meta-Analysis of Carbon Nanotube Pulmonary Toxicity Studies-How Physical Dimensions and Impurities

Affect the Toxicity of Carbon Nanotubes, *Risk Anal.* 34 (3) (2014) 583–597.

- [87] D. A. Winkler, F. R. Burden, B. Yan, R. Weissleder, C. Tassa, S. Shaw, V. C. Epa, Modelling and Predicting the Biological Effects of Nanomaterials, *SAR QSAR Environ. Res.* 25 (2) (2014) 161–172.
- [88] T. C. Le, D. A. Winkler, Discovery and Optimization of Materials Using Evolutionary Approaches, *Chem. Rev.* 116 (10) (2016) 6107–6132.
- [89] N. S. Froemming, G. Henkelman, Optimizing Core-Shell Nanoparticle Catalysts with a Genetic Algorithm, *J. Chem. Phys.* 131 (23) (2009) 234103.
- [90] S. Kim, K.-S. Sohn, M. Pyo, Genetic Algorithm-Assisted Optimization of Nanoporous TiO₂ for Low-Temperature Processable Photoanodes of Dye-Sensitized Solar Cells, *ACS Comb. Sci.* 13 (2) (2011) 101–106.
- [91] R. Fernandez Martinez, A. Okariz, J. Ibarretxe, M. Iturrondobeitia, T. Guraya, Use of Decision Tree Models Based on Evolutionary Algorithms for the Morphological Classification of Reinforcing Nanoparticle Aggregates, *Comput. Mater. Sci.* 92 (2014) 102–113.
- [92] W. Wang, A. Sedykh, H. Sun, L. Zhao, D. P. Russo, H. Zhou, B. Yan, H. Zhu, Predicting Nano-Bio Interactions by Integrating Nanoparticle Libraries and Quantitative Nanostructure Activity Relationship Modeling, *ACS Nano* 11 (12) (2017) 12641–12649.
- [93] V. Kovalishyn, N. Abramenko, I. Kopernyk, L. Charochkina, L. Metelytsia, I. V. Tetko, W. Peijnenburg, L. Kustov, Modelling the Toxicity of a Large Set of Metal and Metal Oxide Nanoparticles Using the OCHEM Platform, *Food Chem. Toxicol.* 112 (2018) 507–517.
- [94] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules, *ACS Cent. Sci.* 4 (2) (2018) 268–276.

- [95] F. Hataminia, Z. Noroozi, H. Mobaleghol Eslam, Investigation of Iron Oxide Nanoparticle Cytotoxicity in Relation to Kidney Cells: A Mathematical Modeling of Data Mining, *Toxicol. Vitro.* 59 (2019) 197–203.
- [96] J. Lazarovits, S. Sindhwani, A. J. Tavares, Y. Zhang, F. Song, J. Audet, J. R. Krieger, A. M. Syed, B. Stordy, W. C. W. Chan, Supervised Learning and Mass Spectrometry Predicts the in Vivo Fate of Nanomaterials, *ACS Nano* 13 (7) (2019) 8023–8034.
- [97] S. Balraadsing, W. J. G. M. Peijnenburg, M. G. Vijver, Exploring the Potential of in Silico Machine Learning Tools for the Prediction of Acute Daphnia Magna Nanotoxicity, *Chemosphere* 307 (2022) 135930.
- [98] S. Li, A. S. Barnard, Safety-by-Design Using Forward and Inverse Multi-Target Machine Learning, *Chemosphere* 303 (2022) 135033.
- [99] L. E. Vivanco-Benavides, C. L. Martínez-González, C. Mercado-Zúñiga, C. Torres-Torres, Machine Learning and Materials Informatics Approaches in the Analysis of Physical Properties of Carbon Nanotubes: A Review, *Comput. Mater. Sci.* 201 (2022) 110939.
- [100] X. Z. Wang, Y. Yang, R. Li, C. McGuinness, J. Adamson, I. L. Megson, K. Donaldson, Principal Component and Causal Analysis of Structural and Acute in Vitro Toxicity Data for Nanoparticles, *Nanotoxicology* 8 (5) (2014) 465–476.
- [101] S. K. Jha, T. H. Yoon, Z. Pan, Multivariate Statistical Analysis for Selecting Optimal Descriptors in the Toxicity Modeling of Nanomaterials, *Comput. Biol. Med.* 99 (2018) 161–172.
- [102] N. Sizochenko, A. Mikolajczyk, K. Jagiello, T. Puzyn, J. Leszczynski, B. Rasulev, How the Toxicity of Nanomaterials towards Different Species Could Be Simultaneously Evaluated: A Novel Multi-Nano-Read-across Approach, *Nanoscale* 10 (2) (2018) 582–591.
- [103] K. Blekos, K. Chairetakis, I. Lynch, E. Marcoulaki, Principles and requirements for nanomaterial representations to facilitate machine processing and cooperation with nanoinformatics tools, *Journal of Cheminformatics* 15 (1) (2023) 1–17.

- [104] X. Zhang, G. Ma, W. Wei, Simulation of nanoparticles interacting with a cell membrane: probing the structural basis and potential biomedical application, *NPG Asia Mater.* 13 (1) (2021).
- [105] F. Spiegelman, N. Tarrat, J. Cuny, L. Dontot, E. Posenitskiy, C. Martí, A. Simon, M. Rapacioli, Density-functional tight-binding: basic concepts and applications to molecules and clusters, *Adv. Phys.* X 5 (1) (2020).
- [106] O. Lamiel-Garcia, A. Cuko, M. Calatayud, F. Illas, S. T. Bromley, Predicting size-dependent emergence of crystallinity in nanomaterials: titania nanoclusters versus nanocrystals, *Nanoscale* 9 (3) (2017) 1049–1058.
- [107] A. Cuko, A. M. Escatllar, M. Calatayud, S. T. Bromley, Properties of hydrated tio 2 and sio 2 nanoclusters: dependence on size, temperature and water vapour pressure, *Nanoscale* 10 (45) (2018) 21518–21532.
- [108] F. A. Soria, C. Di Valentin, Reactive molecular dynamics simulations of hydration shells surrounding spherical TiO₂nanoparticles: Implications for proton-Transfer reactions, *Nanoscale* 13 (7) (2021) 4151–4166.
- [109] M. Kohantorabi, M. Wagstaffe, M. Creutzburg, A. Ugolotti, S. Kulkarni, A. Jeromin, T. Krekeler, M. Feuerherd, A. Herrmann, G. Ebert, U. Protzer, G. Guñdez, C. Löffler, R. Thuenauer, C. Schlueter, A. Gloskovskii, T. F. Keller, C. Di Valentin, A. Stierle, H. Noei, Adsorption and inactivation of sars-cov-2 on the surface of anatase tio₂(101), *ACS Applied Materials & Interfaces* 15 (6) (2023) 8770–8782. doi:10.1021/acscami.2c22078.
URL <https://doi.org/10.1021/acscami.2c22078>
- [110] F. A. Soria, C. Daldossi, C. Di Valentin, Tuning the electron injection mechanism by changing the adsorption mode: the case study of alizarin on tio₂, *Materials Today Energy* 28 (2022) 101085.
URL <https://www.sciencedirect.com/science/article/pii/S2468606922001435>
- [111] P. Siani, G. Frigerio, E. Donadoni, C. Di Valentin, Molecular dynamics simulations of crgd-conjugated pegylated tio₂ nanoparticles for targeted photodynamic therapy, *Journal of*

Colloid and Interface Science 627 (2022) 126–141. doi:<https://doi.org/10.1016/j.jcis.2022.07.045>.
URL <https://www.sciencedirect.com/science/article/pii/S0021979722012218>

- [112] Y. A. Çetin, B. Martorell, F. Serratosa, N. Aguilera-Porta, M. Calatayud, Analyzing the tio₂ surface reactivity based on oxygen vacancies computed by dft and dftb methods, *Journal of Physics: Condensed Matter* 34 (31) (2022) 314004.
- [113] R. Iftimie, P. Minary, M. E. Tuckerman, Ab initio molecular dynamics: Concepts, recent developments, and future trends, *Proc. Natl. Acad. Sci. U. S. A.* 102 (19) (2005) 6654–6659.
- [114] P. J. Ollitrault, A. Miessen, I. Tavernelli, Molecular Quantum Dynamics: A Quantum Computing Perspective, *Acc. Chem. Res.* 54 (23) (2021) 4229–4238.
- [115] L. Shen, W. Yang, Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks, *J. Chem. Theory Comput.* 14 (3) (2018) 1442–1455.
- [116] S. Chmiela, H. E. Sauceda, K. R. Müller, A. Tkatchenko, Towards exact molecular dynamics simulations with machine-learned force fields, *Nat. Commun.* 9 (1) (2018).
- [117] M. K. Bisbo, B. Hammer, Efficient global structure optimization with a machine-learned surrogate model, *Phys. Rev. Lett.* 124 (2020) 086102. doi:[10.1103/PhysRevLett.124.086102](https://doi.org/10.1103/PhysRevLett.124.086102).
URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.086102>
- [118] X. Liu, H. Niu, A. R. Oganov, Copex: co-evolutionary crystal structure prediction algorithm for complex systems, *npj Computational Materials* 7 (1) (2021) 199. doi:[10.1038/s41524-021-00668-5](https://doi.org/10.1038/s41524-021-00668-5).
URL <https://doi.org/10.1038/s41524-021-00668-5>
- [119] K. H. Sørensen, M. S. Jørgensen, A. Bruix, B. Hammer, Accelerating atomic structure search with cluster regularization, *The Journal of Chemical Physics* 148 (24) (2018) 241734. arXiv:<https://doi.org/>

10.1063/1.5023671, doi:10.1063/1.5023671.
URL <https://doi.org/10.1063/1.5023671>

- [120] A. Singhal, G. J. Agur Sevink, The role of size and nature in nanoparticle binding to a model lung membrane: an atomistic study, *Nanoscale Adv.* 3 (2021) 6635–6648.
- [121] B. W. Brinkmann, A. Singhal, G. J. Sevink, L. Neeft, M. G. Vijver, W. J. Peijnenburg, Predicted Adsorption Affinity for Enteric Microbial Metabolites to Metal and Carbon Nanomaterials, *J. Chem. Inf. Model.* 62 (15) (2022) 3589–3603.
- [122] M. Karelson, *Molecular descriptors in QSAR/QSPR*, John Wiley & Sons, New York, 2000.
- [123] J. Burk, L. Sikk, P. Burk, B. B. Manshian, S. J. Soenen, J. J. Scott-Fordsmand, T. Tamm, K. Tämm, Fe-Doped ZnO nanoparticle toxicity: Assessment by a new generation of nanodescriptors, *Nanoscale* 10 (46) (2018) 21985–21993.
- [124] K. Tämm, L. Sikk, J. Burk, R. Rallo, S. Pokhrel, L. Mädler, J. J. Scott-Fordsmand, P. Burk, T. Tamm, Parametrization of nanoparticles: Development of full-particle nanodescriptors, *Nanoscale* 8 (36) (2016) 16243–16250.
- [125] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, S. J. Plimpton, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, *Comput. Phys. Commun.* 271 (2022) 108171.
- [126] R. Buckingham, The classical equation of state of gaseous helium, neon and argon, *Proc. R. Soc. Lond. A* 168 (1938) 264–283.
- [127] D. Wolf, P. Keblinski, S. R. Phillpot, J. Eggebrecht, Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r-1 summation, *J. Chem. Phys.* 110 (17) (1999) 8254–8282.
- [128] B. B. Manshian, S. Pokhrel, U. Himmelreich, K. Tämm, L. Sikk, A. Fernández, R. Rallo, T. Tamm, L. Mädler, S. J. Soenen, In Silico Design

- of Optimal Dissolution Kinetics of Fe-Doped ZnO Nanoparticles Results in Cancer-Specific Toxicity in a Preclinical Rodent Model, *Adv. Healthc. Mater.* 6 (9) (2017) 1–11.
- [129] A. C. Van Duin, S. Dasgupta, F. Lorant, W. A. Goddard, ReaxFF: A reactive force field for hydrocarbons, *J. Phys. Chem. A* 105 (41) (2001) 9396–9409.
- [130] K. Chenoweth, A. C. Van Duin, W. A. Goddard, ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation, *J. Phys. Chem. A* 112 (5) (2008) 1040–1053.
- [131] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama, A. C. Van Duin, The ReaxFF reactive force-field: Development, applications and future directions, *npj Comput. Mater.* 2 (2016) 15011.
- [132] A. Hoekstra, B. Chopard, D. Coster, S. Portugies Zwart, P. Coveney, Multiscale computing for science and engineering in the era of exascale performance, *Phil. Trans. R. Soc. A* 377 (2018) 0180144.
- [133] M. Müller, K. Katsov, M. Schick, Biological and synthetic membranes: What can be learned from a coarse-grained description?, *Phys. Rep.* 434 (5-6) (2006) 113–176.
- [134] K. M. Langner, G. J. Sevink, Mesoscale modeling of block copolymer nanocomposites, *Soft Matter* 8 (19) (2012) 5102–5118.
- [135] G. J. Sevink, A. V. Zvelindovsky, B. A. Van Vlimmeren, N. M. Maurits, J. G. Fraaije, Dynamics of surface directed mesophase formation in block copolymer melts, *J. Chem. Phys.* 110 (4) (1999) 2250–2256.
- [136] A. C. Balazs, V. V. Ginzburg, F. Qiu, G. Peng, D. Jasnow, Multi-scale model for binary mixtures containing nanoscopic particles, *J. Phys. Chem. B* 104 (15) (2000) 3411–3422.
- [137] S. W. Sides, B. J. Kim, E. J. Kramer, G. H. Fredrickson, Hybrid particle-field simulations of polymer nanocomposites, *Phys. Rev. Lett.* 96 (25) (2006) 1–4.

- [138] G. Milano, T. Kawakatsu, Hybrid particle-field molecular dynamics simulations for dense polymer systems, *J. Chem. Phys.* 130 (21) (2009).
- [139] K. C. Daoulas, M. Müller, J. J. De Pablo, P. F. Nealey, G. D. Smith, Morphology of multi-component polymer systems: Single chain in mean field simulation studies, *Soft Matter* 2 (7) (2006) 573–583.
- [140] G. J. Sevink, E. M. Blokhuis, X. Li, G. Milano, Efficient and realistic simulation of phase coexistence, *J. Chem. Phys.* 153 (24) (2020).
- [141] M. Schneemilch, N. Quirke, Predicting nanoparticle uptake by biological membranes: theory and simulation, *Molecular Simulation* 48 (2022) 150–167.
- [142] M. Deserno, W. M. Gelbart, Adhesion and wrapping in colloid-vesicle complexes, *J. Phys. Chem. B* 106 (21) (2002) 5543–5552.
- [143] M. Deserno, T. Bickel, Wrapping of a spherical colloid by a fluid membrane, *Europhys. Lett.* 62 (5) (2003) 767–773.
- [144] M. Raatz, R. Lipowsky, T. R. Weikl, Cooperative wrapping of nanoparticles by membrane tubes, *Soft Matter* 10 (20) (2014) 3570–3577.
- [145] E. J. Spangler, S. Upreti, M. Laradji, Partial wrapping and spontaneous endocytosis of spherical nanoparticles by tensionless lipid membranes, *J. Chem. Phys.* 144 (4) (2016).
- [146] E. J. Spangler, M. Laradji, Discontinuous wrapping transition of spherical nanoparticles by tensionless lipid membranes, *J. Chem. Phys.* 152 (10) (2020).
- [147] X. Yi, X. Shi, H. Gao, Cellular uptake of elastic nanoparticles, *Phys. Rev. Lett.* 107 (9) (2011) 1–5.
- [148] X. Yi, H. Gao, Cell membrane wrapping of a spherical thin elastic shell, *Soft Matter* 11 (6) (2015) 1107–1115.
- [149] S. Dasgupta, T. Auth, G. Gompper, Wrapping of ellipsoidal nanoparticles by fluid membranes, *Soft Matter* 9 (22) (2013) 5473–5482. [arXiv:1303.5567](https://arxiv.org/abs/1303.5567).

- [150] S. Dasgupta, T. Auth, G. Gompper, Shape and orientation matter for the cellular uptake of nonspherical particles, *Nano Lett.* 14 (2) (2014) 687–693.
- [151] R. Vácha, F. J. Martinez-Veracoechea, D. Frenkel, Receptor-mediated endocytosis of nanoparticles of various shapes, *Nano Lett.* 11 (12) (2011) 5391–5395.
- [152] J. Huang, A. D. Mackerell, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data, *J. Comput. Chem.* 34 (25) (2013) 2135–2145.
- [153] A. Šarić, A. Cacciuto, Fluid membranes can drive linear aggregation of adsorbed spherical nanoparticles, *Phys. Rev. Lett.* 108 (11) (2012) 1–5.
- [154] P. Angelikopoulos, L. Sarkisov, Z. Cournia, P. Gkeka, Self-assembly of anionic, ligand-coated nanoparticles in lipid membranes, *Nanoscale* 9 (3) (2017) 1040–1048.
- [155] F. Simonelli, D. Bochicchio, R. Ferrando, G. Rossi, Monolayer-Protected Anionic Au Nanoparticles Walk into Lipid Membranes Step by Step, *J. Phys. Chem. Lett.* 6 (16) (2015) 3175–3179.
- [156] S. Salassi, F. Simonelli, D. Bochicchio, R. Ferrando, G. Rossi, Au Nanoparticles in Lipid Bilayers: A Comparison between Atomistic and Coarse-Grained Models, *J. Phys. Chem. C* 121 (20) (2017) 10927–10935.
- [157] T. Ruiz-Herrero, E. Velasco, M. F. Hagan, Mechanisms of budding of nanoscale particles through lipid bilayers, *J. Phys. Chem. B* 116 (32) (2012) 9595–9603.
- [158] A. Singhal, A. Sevink, A systematic core-shell approach for coarsening nanoparticle-membrane interactions: application to silver nanoparticles, *Nanomaterials* 12 (2022) 3859.
- [159] Y. Smirnova, H. Risselada, M. Müller, Thermodynamically reversible paths of the first fusion intermediate reveal an important role for membrane anchors of fusion proteins, *PNAS* 116 (2019) 2571–2576.

- [160] T. Cedervall, I. Lynch, S. Lindman, T. Berggård, E. Thulin, H. Nilsson, K. A. Dawson, S. Linse, Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles, *Proc. Natl. Acad. Sci. U. S. A.* 104 (7) (2007) 2050–2055.
- [161] M. Lundqvist, J. Stigler, G. Elia, I. Lynch, T. Cedervall, K. A. Dawson, Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts, *Proc. Natl. Acad. Sci. U. S. A.* 105 (38) (2008) 14265–14270.
- [162] I. Lynch, K. A. Dawson, Protein-nanoparticle interactions, *nanotoday* 3 (2008) 40–47.
- [163] C. K. Payne, A protein corona primer for physical chemists, *J. Chem. Phys.* 151 (13) (2019).
- [164] T. Casalini, V. Limongelli, M. Schmutz, C. Som, O. Jordan, P. Wick, G. Borchard, G. Perale, Molecular modeling for nanomaterial–biology interactions: Opportunities, challenges, and perspectives, *Front. Bioeng. Biotechnol.* 7 (OCT) (2019) 1–14.
- [165] J. E. Vance, Phospholipid Synthesis and Transport in Mammalian Cells, *Traffic* 16 (1) (2015) 1–18.
- [166] I. Capjak, S. Š. Goreta, D. D. Jurašin, I. V. Vrček, How protein coronas determine the fate of engineered nanoparticles in biological environment, *Arh. Hig. Rada Toksikol.* 68 (4) (2017) 245–253.
- [167] R. García-álvarez, M. Vallet-Regí, Hard and soft protein corona of nanomaterials: Analysis and relevance, *Nanomaterials* 11 (4) (2021).
- [168] M. Ovais, S. Mukherjee, A. Pramanik, D. Das, A. Mukherjee, A. Raza, C. Chen, Designing Stimuli-Responsive Upconversion Nanoparticles that Exploit the Tumor Microenvironment, *Adv. Mater.* 32 (22) (2020) 1–19.
- [169] V. Srivastava, D. Gusain, Y. C. Sharma, Critical Review on the Toxicity of Some Widely Used Engineered Nanoparticles, Vol. 54, 2015.
- [170] I. Hasenkopf, R. Mills-Goodlet, L. Johnson, I. Rouse, M. Gepfert, A. Duschl, D. Maier, V. Lobaskin, I. Lynch, M. Himly,

Computational prediction and experimental analysis of the nanoparticle-protein corona: Showcasing an in vitro-in silico workflow providing fair data, *Nano Today* 46 (2022) 101561. doi:<https://doi.org/10.1016/j.nantod.2022.101561>. URL <https://www.sciencedirect.com/science/article/pii/S174801322200189X>

- [171] I. Rouse, D. Power, E. G. Brandt, M. Schneemilch, K. Kotsis, N. Quirke, A. P. Lyubartsev, V. Lobaskin, First principles characterisation of bio-nano interface, *Phys. Chem. Chem. Phys.* 23 (24) (2021) 13473–13482.
- [172] J. Subbotina, V. Lobaskin, Multiscale modeling of bio-nano interactions of zero-valent silver nanoparticles, *The Journal of Physical Chemistry B* 126 (6) (2022) 1301–1314, pMID: 35132861. doi:10.1021/acs.jpcc.1c09525.
- [173] D. Power, I. Rouse, S. Poggio, E. Brandt, H. Lopez, A. Lyubartsev, V. Lobaskin, A multiscale model of protein adsorption on a nanoparticle surface, *Model. Simul. Mater. Sci. Eng.* 27 (8) (2019) 84003.
- [174] S. A. Alsharif, D. Power, I. Rouse, V. Lobaskin, In silico prediction of protein adsorption energy on titanium dioxide and gold nanoparticles, *Nanomaterials* 10 (10) (2020) 1–21.
- [175] N.-v. Buchete, J. E. Straub, D. Thirumalai, Development of novel statistical potentials for protein fold recognition, *Curr. Opin. Struct. Biol.* 14 (2004) 225–232.
- [176] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. De Vries, The MARTINI force field: Coarse grained model for biomolecular simulations, *J. Phys. Chem. B* 111 (27) (2007) 7812–7824.
- [177] V. Karunakaran Annapoorani, E. al., Atomistic models of protein-protein interactions in nanoparticle coronas, *Prep.* (2022).
- [178] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, PatchDock and SymmDock: Servers for rigid and symmetric docking, *Nucleic Acids Res.* 33 (SUPPL. 2) (2005) 363–367.

- [179] S. Chibber, I. Ahmed, Molecular docking, a tool to determine interaction of CuO and TiO₂ nanoparticles with human serum albumin, *Biochem. Biophys. Reports* 6 (2016) 63–67.
- [180] B. G. Pierce, K. Wiehe, H. Hwang, B. H. Kim, T. Vreven, Z. Weng, ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers, *Bioinformatics* 30 (12) (2014) 1771–1773.
- [181] M. S. Ali, M. Altaf, H. A. Al-Lohedan, Green synthesis of biogenic silver nanoparticles using *Solanum tuberosum* extract and their interaction with human serum albumin: Evidence of “corona” formation through a multi-spectroscopic and molecular docking analysis, *J. Photochem. Photobiol. B Biol.* 173 (May) (2017) 108–119.
- [182] T. Lima, K. Bernfur, M. Vilanova, T. Cedervall, Understanding the Lipid and Protein Corona Formation on Different Sized Polymeric Nanoparticles, *Sci. Rep.* 10 (1) (2020) 1–9.
- [183] A. Pancaro, M. Szymonik, P. G. Georgiou, A. N. Baker, M. Walker, P. Adriaensens, J. Hendrix, M. I. Gibson, I. Nelissen, The polymeric glyco-linker controls the signal outputs for plasmonic gold nanorod biosensors due to biocorona formation, *Nanoscale* 13 (24) (2021) 10837–10848.
- [184] A. Kurtz-Chalot, C. Villiers, J. Pourchez, D. Boudard, M. Martini, P. N. Marche, M. Cottier, V. Forest, Impact of silica nanoparticle surface chemistry on protein corona formation and consequential interactions with biological cells, *Mater. Sci. Eng. C* 75 (2017) 16–24.
- [185] J. F. Hainfeld, F. A. Dilmanian, D. N. Slatkin, H. M. Smilowitz, Radiotherapy enhancement with gold nanoparticles, *J. Pharm. Pharmacol.* 60 (2008) 977–985.
- [186] O. Zeiri, Metallic-Nanoparticle-Based Sensing: Utilization of Mixed-Ligand Monolayers, *ACS Sensors* 5 (12) (2020) 3806–3820.
- [187] A. Heuer-Jungemann, N. Feliu, I. Bakaimi, M. Hamaly, A. Alkilany, I. Chakraborty, A. Masood, M. F. Casula, A. Kostopoulou, E. Oh, K. Susumu, M. H. Stewart, I. L. Medintz, E. Stratakis, W. J. Parak,

- A. G. Kanaras, The Role of Ligands in the Chemical Synthesis and Applications of Inorganic Nanoparticles, *Chem. Rev.* 119 (8) (2019) 4819–4880.
- [188] C. Singh, P. K. Ghorai, M. A. Horsch, A. M. Jackson, R. G. Larson, F. Stellacci, S. C. Glotzer, Entropy-mediated patterning of surfactant-coated nanoparticles and surfaces, *Phys. Rev. Lett.* 99 (22) (2007) 1–4.
- [189] M. Baranov, E. Nepomyashchaya, E. Velichko, Computer Simulation of Biomolecules Around Metallic Nanoparticle for Biomolecular Electronics, *Proc. 2021 Int. Conf. Electr. Eng. Photonics, EExPolytech 2021* (2021) 171–174.
- [190] R. Cappabianca, P. De Angelis, A. Cardellini, E. Chiavazzo, P. Asinari, Assembling Biocompatible Polymers on Gold Nanoparticles: Toward a Rational Design of Particle Shape by Molecular Dynamics, *ACS Omega* 7 (2022) 42292–42303.
- [191] A. C. T. van Duin, C. Zou, K. Joshi, V. Bryantsev, W. A. Goddard, CHAPTER 6. A Reaxff Reactive Force-field for Proton Transfer Reactions in Bulk Water and its Applications to Heterogeneous Catalysis (14) (2014) 223–243.
- [192] O. T. Unke, S. Chmiela, H. E. Saucedo, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, K.-R. Müller, Machine learning force fields, *Chemical Reviews* 121 (16) (2021) 10142–10186.
- [193] G. Trezza, L. Bergamasco, M. Fasano, E. Chiavazzo, Minimal crystallographic descriptors of sorption properties in hypothetical mofs and role in sequential learning optimization, *npj Computational Materials* 8 (1) (2022) 1–14.
- [194] E. Chiavazzo, R. Covino, R. R. Coifman, C. W. Gear, A. S. Georgiou, G. Hummer, I. G. Kevrekidis, Intrinsic map dynamics exploration for uncharted effective free-energy landscapes, *Proceedings of the National Academy of Sciences* 114 (28) (2017) E5494–E5503.
- [195] O. T. Unke, S. Chmiela, H. E. Saucedo, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, K.-R. Müller, Machine learning force fields, *Chemical Reviews* 121 (16) (2021) 10142–10186. doi:10.1021/

acs.chemrev.0c01111.

URL <https://doi.org/10.1021/acs.chemrev.0c01111>

- [196] S. Bag, M. Konrad, T. Schl  nder, P. Friederich, W. Wenzel, Fast generation of machine learning-based force fields for adsorption energies, *Journal of Chemical Theory and Computation* 17 (11) (2021) 7195–7202. doi:10.1021/acs.jctc.1c00506.
URL <https://doi.org/10.1021/acs.jctc.1c00506>
- [197] T. Seki, K. Takeda, A. Nakayama, Ab Initio Molecular Dynamics Simulations for Adsorption Structures of Picolinic Acid at the Water/CeO₂ Interface, *J. Phys. Chem. C* 126 (7) (2022) 3404–3410.
- [198] S. Wei, S. Zheng, C. Xie, J. Liang, Ab initio molecular dynamics study of wet H₂S adsorption and dissociation on Fe(100) surface, *J. Mol. Liq.* 319 (2020) 114135.
- [199] K. Leung, J. A. Greathouse, Ab initio molecular dynamics free energy study of enhanced copper (II) dimerization on mineral surfaces, *Commun. Chem.* 5 (1) (2022) 1–8.
- [200] D. Chen, C. Shang, Z.-P. Liu, Automated search for optimal surface phases (asops) in grand canonical ensemble powered by machine learning, *The Journal of Chemical Physics* 156 (9) (2022) 094104. arXiv: <https://doi.org/10.1063/5.0084545>, doi:10.1063/5.0084545.
URL <https://doi.org/10.1063/5.0084545>
- [201] I. Tom  , V. Francisco, H. Fernandes, L. Ferreira, High-throughput screening of nanoparticles in drug delivery, *APL Bioeng.* 5 (3) (2021) 1–12.
- [202] J. E. Dahlman, K. J. Kauffman, Y. Xing, T. E. Shaw, F. F. Mir, C. C. Dlott, R. Langer, D. G. Anderson, E. T. Wang, Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics, *Proc. Natl. Acad. Sci. U. S. A.* 114 (8) (2017) 2060–2065.
- [203] X. Mao, L. Wang, Y. Xu, P. Wang, Y. Li, J. Zhao, Computational high-throughput screening of alloy nanoclusters for electrocatalytic hydrogen evolution, *npj Comput. Mater.* 7 (1) (2021) 1–9.

- [204] N. Jeliaskova, C. Chomenidis, P. Doganis, B. Fadeel, R. Grafström, B. Hardy, J. Hastings, M. Hegi, V. Jeliaskov, N. Kochev, P. Kohonen, C. R. Munteanu, H. Sarimveis, B. Smeets, P. Sopasakis, G. Tsiliki, D. Vorgrimmler, E. Willighagen, The eNanoMapper database for nanomaterial safety information, *Beilstein J. Nanotechnol.* 6 (1) (2015) 1609–1634.
- [205] N. Jeliaskova, M. D. Apostolova, C. Andreoli, F. Barone, A. Barrick, C. Battistelli, C. Bossa, A. Botea-Petcu, A. Châtel, I. De Angelis, M. Dusinska, N. El Yamani, D. Gheorghe, A. Giusti, P. Gómez-Fernández, R. Grafström, M. Gromelski, N. R. Jacobsen, V. Jeliaskov, K. A. Jensen, N. Kochev, P. Kohonen, N. Manier, E. Mariussen, A. Mech, J. M. Navas, V. Paskaleva, A. Precupas, T. Puzyn, K. Rasmussen, P. Ritchie, I. R. Llopis, E. Rundén-Pran, R. Sandu, N. Shandilya, S. Tanasescu, A. Haase, P. Nymark, Towards FAIR nanosafety data, *Nat. Nanotechnol.* 16 (6) (2021) 644–654.
- [206] N. Jeliaskova, V. Jeliaskov, AMBIT RESTful web services: An implementation of the OpenTox application programming interface, *J. Cheminform.* 3 (1) (2011) 1–18.
- [207] N. Jeliaskova, Web tools for predictive toxicology model building, *Expert Opin. Drug Metab. Toxicol.* 8 (7) (2012) 791–801.
- [208] N. Kochev, N. Jeliaskova, V. Paskaleva, G. Tancheva, L. Iliev, P. Ritchie, V. Jeliaskov, Your spreadsheets can be fair: A tool and fairification workflow for the enanomapper database, *Nanomaterials* 10 (10) (2020) 1–23.
- [209] OECD, Guidance on Grouping of Chemicals, Second Edition, 2017.
- [210] ECHA, Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals, 2008.
- [211] A. Mech, K. Rasmussen, P. Jantunen, L. Aicher, M. Alessandrelli, U. Bernauer, E. A. Bleeker, J. Bouillard, P. Di Prospero Fanghella, R. Draisci, M. Dusinska, G. Encheva, G. Flament, A. Haase, Y. Handzhiyski, F. Herzberg, J. Huwyler, N. R. Jacobsen, V. Jeliaskov, N. Jeliaskova, P. Nymark, R. Grafström, A. G. Oomen, M. L. Polci, C. Riebeling, J. Sandström, B. Shivachev, S. Stateva, S. Tanasescu,

- R. Tsekovska, H. Wallin, M. F. Wilks, S. Zellmer, M. D. Apostolova, Insights into possibilities for grouping and read-across for nanomaterials in EU chemicals legislation, *Nanotoxicology* 13 (1) (2019) 119–141.
- [212] S. Halappanavar, S. Van Den Brule, P. Nymark, L. Gaté, C. Seidel, S. Valentino, V. Zhernovkov, P. Høgh Danielsen, A. De Vizcaya, H. Wolff, T. Stöger, A. Boyadziev, S. S. Poulsen, J. B. Sørli, U. Vogel, Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale, *Part. Fibre Toxicol.* 17 (1) (2020) 1–24.
- [213] G. T. Ankley, R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, J. A. Serrano, J. E. Tietge, D. L. Villeneuve, Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment, *Environ. Toxicol. Chem.* 29 (3) (2010) 730–741.
- [214] A. Giusti, R. Atluri, R. Tsekovska, A. Gajewicz, M. D. Apostolova, C. L. Battistelli, E. A. Bleeker, C. Bossa, J. Bouillard, M. Dusinska, P. Gómez-Fernández, R. Grafström, M. Gromelski, Y. Handzhiyski, N. R. Jacobsen, P. Jantunen, K. A. Jensen, A. Mech, J. M. Navas, P. Nymark, A. G. Oomen, T. Puzyn, K. Rasmussen, C. Riebeling, I. Rodriguez-Llopis, S. Sabella, J. R. Sintes, B. Suarez-Merino, S. Tanasescu, H. Wallin, A. Haase, Nanomaterial grouping: Existing approaches and future recommendations, *NanoImpact* 16 (2019) 100182.
- [215] L. Lamon, K. Aschberger, D. Asturiol, A. Richarz, A. Worth, Grouping of nanomaterials to read-across hazard endpoints: a review, *Nanotoxicology* 13 (1) (2019) 100–118.
- [216] A. G. Oomen, K. G. Steinhäuser, E. A. Bleeker, F. van Broekhuizen, A. Sips, S. Dekkers, S. W. Wijnhoven, P. G. Sayre, Risk assessment frameworks for nanomaterials: Scope, link to regulations, applicability, and outline for future directions in view of needed increase in efficiency, *NanoImpact* 9 (2018) 1–13.
- [217] A. Bahl, B. Hellack, M. Balas, A. Dinischiotu, M. Wiemann, J. Brinkmann, A. Luch, B. Y. Renard, A. Haase, Recursive fea-

ture elimination in random forest classification supports nanomaterial grouping, *NanoImpact* 15 (2019) 100179.

- [218] A. Bahl, B. Hellack, M. Wiemann, A. Giusti, K. Werle, A. Haase, W. Wohlleben, Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods, *NanoImpact* 19 (2020) 100234.
- [219] L. Lamon, D. Asturiol, A. Richarz, E. Joossens, R. Graepel, K. Aschberger, A. Worth, Grouping of nanomaterials to read-across hazard endpoints: From data collection to assessment of the grouping hypothesis by application of chemoinformatic techniques, Part. *Fibre Toxicol.* 15 (1) (2018) 1–17.
- [220] I. Karkossa, A. Bannuscher, B. Hellack, A. Bahl, S. Buhs, P. Nollau, A. Luch, K. Schubert, M. Von Bergen, A. Haase, An in-depth multi-omics analysis in RLE-6TN rat alveolar epithelial cells allows for nanomaterial categorization, Part. *Fibre Toxicol.* 16 (1) (2019) 1–19.
- [221] I. Karkossa, A. Bannuscher, B. Hellack, W. Wohlleben, J. Laloy, M. S. Stan, A. Dinischiotu, M. Wiemann, A. Luch, A. Haase, M. von Bergen, K. Schubert, Nanomaterials induce different levels of oxidative stress, depending on the used model system: Comparison of in vitro and in vivo effects, *Sci. Total Environ.* 801 (2021) 149538.
- [222] A. Bannuscher, I. Karkossa, S. Buhs, P. Nollau, K. Kettler, M. Balas, A. Dinischiotu, B. Hellack, M. Wiemann, A. Luch, M. von Bergen, A. Haase, K. Schubert, A multi-omics approach reveals mechanisms of nanomaterial toxicity and structure–activity relationships in alveolar macrophages, *Nanotoxicology* 14 (2) (2020) 181–195.
- [223] A. Bannuscher, B. Hellack, A. Bahl, J. Laloy, H. Herman, M. S. Stan, A. Dinischiotu, A. Giusti, B. C. Krause, J. Tentschert, M. Roşu, C. Balta, A. Hermenean, M. Wiemann, A. Luch, A. Haase, Metabolomics profiling to investigate nanomaterial toxicity in vitro and in vivo, *Nanotoxicology* 14 (6) (2020) 807–826.
- [224] L. A. Saarimäki, A. Federico, I. Lynch, A. G. Papadiamantis, A. Tsoumanis, G. Melagraki, A. Afantitis, A. Serra, D. Greco, Manually

curated transcriptomics data collection for toxicogenomic assessment of engineered nanomaterials, *Sci. Data* 8 (1) (2021) 1–10.

- [225] V. Fortino, P. A. S. Kinaret, M. Fratello, A. Serra, L. A. Saarimäki, A. Gallud, G. Gupta, G. Vales, M. Correia, O. Rasool, J. Ytterberg, M. Monopoli, T. Skoog, P. Ritchie, S. Moya, S. Vázquez-Campos, R. Handy, R. Grafström, L. Tran, R. Zubarev, R. Lahesmaa, K. Dawson, K. Loeschner, E. H. Larsen, F. Krombach, H. Norppa, J. Kere, K. Savolainen, H. Alenius, B. Fadeel, D. Greco, Biomarkers of nanomaterials hazard from multi-layer data, *Nat. Commun.* 13 (1) (2022).
- [226] OECD, Transcriptomic Reporting Framework (TRF), 2021.
- [227] OECD, Metabolomic Reporting Framework (MRF), 2021.
- [228] A. Bahl, C. Ibrahim, K. Plate, A. Haase, J. Dengjel, P. Nymark, V. I. Dumit, Proteomas: a workflow enabling harmonized proteomic meta-analysis and proteomic signature mapping, *Journal of Cheminformatics* 15 (1) (2023) 1–17.

Appendix A. List of physics-based descriptors that can be calculated through materials modelling techniques

Scale	Method	Advanced descriptor	Units	Typical values for a 3 nm TiO ₂ nanoparticle	Notes and limitations
Quantum - Space scale: 10 ⁻¹¹ -10 ⁻⁹ m Time scale: static	DFT & DFTB	Standard enthalpy of formation	eV	-9 to -8	Very accurate but ideal only for small nanoparticles up to 4 nm in diameter. Highly dependent on the basis set and the pseudopotential of choice. Absolute energy values are only useful for comparison with other structures.
		Total energy	eV	-90000 to -10000	
		Electronic energy	eV	≈ -	
		Energy of the Highest Occupied Molecular Orbital (HOMO)	eV	≈ -3.2	
		Energy of the Lowest Unoccupied Molecular Orbital (LUMO)	eV	≈ -3.1	
		HOMO-LUMO energy gap	eV	≈ 0-3	
		Valence band width	eV	7.3	
		Conduction band width	eV	2.6	
		Fermi level	eV	≈ -3.3	
		Hydration energy	eV	≈ -143	
		Vertical ionization potential	eV	≈ 3.4	
		Vertical electron affinity	eV	≈ -3	
		Oxygen vacancy formation energy	eV	3 to 4	

Scale	Method	Advanced descriptor	Units	Typical values for a 3 nm TiO ₂ nanoparticle	Notes and limitations
Molecular, atomistic - Space scale: 10 ⁻⁹ -10 ⁻⁸ m - Time scale: static	Quantum mechanics at DFT level	Chemical composition (x9)	-	N.A.	Fast to calculate because there is no need of expensive energy minimization. Not for metallic ENMs. Calculable also for larger nanoparticles up to 60 nm. Possible use in QSAR and ML models. Detailed description of those quantities can be found in Refs. [123, 124]
		Potential energy (x9)	eV	-75 to -18	
		Topology descriptors (x9)	-	2.5 to 6	
		Size descriptors (x3)	Å, Å ² , Å ³	N.A.	
		Lattice energy descriptors (x5)	eV, eV/Å, eV/Å ² , eV/Å ³	-110 to 0	
		Force field descriptors (x27)	-	N.A.	

Scale	Method	Advanced descriptor	Units	Typical values for a 3 nm TiO ₂ nanoparticle	Notes and limitations
	ReaxFF MD	Surface charge	C/m ²	-0.06	Need a reactive force field developed for a similar system
		Particle-Membrane binding free energy	eV	≈ 0	
		Aggregation free energy	KJ/mol	54	
		Solvent Accessible Surface Area (SASA)	Å ²	??? VIO	
	AAMD and docking	Hydrophobic SASA fraction	-	44.5-56.4	
		Polar SASA fraction	-	??? VIO	
		Negatively and positively charged SASA fraction	-	??? VIO	
		Deviation from Smoluchowski's aggregation kinetics theory	-	N.A.	Simulations are fast but require previous AAMD simulations to get the force field parameters
Mesoscopic - Space scale: 10 ⁻⁹ -10 ⁻⁶ m - Time scale: 10 ⁻⁹ -10 ⁻⁶ s	Brownian Dynamics Coarse Grained MD	Membrane bending rigidity and other membrane related descriptors	-	N.A.	AGUR's comments