UNIVERSITYOF BIRMINGHAM

Computational enrichment of physicochemical data for the development of a ζ-potential read-across predictive model with Isalos Analytics Platform

Papadiamantis, Anastasios G.; Afantitis, Antreas; Tsoumanis, Andreas; Valsami-Jones, Eugenia; Lynch, Iseult; Melagraki, Georgia

Document Version
Publisher's PDF, also known as Version of record

[Link to publication on Research at Birmingham portal](#)

Research paper

Computational enrichment of physicochemical data for the development of a ζ-potential read-across predictive model with Isalos Analytics Platform

Anastasios G. Papadiamantis [a,b], Antreas Afantitis [a], Andreas Tsoumanis [a], Eugenia Valsami-Jones [b], Iseult Lynch [b,*], Georgia Melagraki [a,*]

[a] *NovaMechanics Ltd, 1065 Nicosia, Cyprus*
[b] *School of Geography, Earth and Environmental Sciences, University of Birmingham, B15 2TT Birmingham, United Kingdom*

ABSTRACT

The physicochemical characterisation data from a library of 69 engineered nanomaterials (ENMs) has been exploited in silico following enrichment with a set of molecular descriptors that can be easily acquired or calculated using atomic periodicity and other fundamental atomic parameters. Based on the extended set of twenty descriptors, a robust and validated nanoinformatics model has been proposed to predict the ENM ζ-potential. The five critical parameters selected as the most significant for the model development included the ENM size and coating as well as three molecular descriptors, metal ionic radius (r_{ion}), the sum of metal electronegativity divided by the number of oxygen atoms present in a particular metal oxide ($\Sigma\chi/n_O$) and the absolute electronegativity (χ_{abs}), each of which is thoroughly discussed to interpret their influence on ζ-potential values. The model was developed using the Isalos Analytics Platform and is available to the community as a web service through the Horizon 2020 (H2020) NanoCommons Transnational Access services and the H2020 NanoSoveIT Integrated Approach to Testing and Assessment (IATA).

1. Introduction

Engineered nanomaterials (ENMs) have become part of everyday life, as they are being used in a wide number of consumer and commercial products. The unique properties of ENM, their versatility, complexity and tuneable properties have put them at the centre of innovation in various fields (e.g. nanomedicine, photonics, energy production) (Kraegeloh et al., 2018). However, the innovation potential of ENMs is inhibited by concerns regarding their potential adverse effects. These can include toxic effects e.g. cytotoxicity, cell apoptosis, oxidative stress, genotoxicity, ecotoxicity etc. following accumulation in different organs, Trojan horse effects and more (Valsami-Jones and Lynch, 2015; Yan et al., 2019; Lin et al., 2018; Lead et al., 2018; Saarimäki et al., 2020). These concerns have led regulatory agencies, including the European Chemicals Agency (ECHA, 2020) and the U.S.A. Environmental Protection Agency (EPA, 2017) to publish guidelines on the testing and safety assessment of ENM. These regulatory frameworks (REACH, TSCA) are based on datasets generated in accordance with the OECD test guidelines for the physicochemical characterisation of ENMs, which currently have 15 listed endpoints (OECD, 2016), and for toxicity/ ecotoxicity assessment using model species and modified test guidelines (Rasmussen et al., 2019) including for dispersion of ENMs (Au-Kaur et al., 2017). Furthermore, the OECD and regulatory bodies are promoting the development of Alternative Testing Strategies (ATS) for the evaluation of ENMs and the development and application of safer by design (SbD) strategies (Hjorth et al., 2017; Fritsche et al., 2017; Bas-ketter et al., 2019). These aim to reduce the need for animal testing and the cost of detailed ENM testing via long-term animal studies, through the design of safer ENMs prior to production. A number of reviews deal with the application of SbD strategies in various ENM application areas (Kraegeloh et al., 2018; Yan et al., 2019; Lin et al., 2018).

The application of in silico approaches (Afantitis, 2020) for the prediction of specific ENM properties is emerging as an important step towards complete in silico hazard and risk assessment. Computational approaches can take advantage of existing data to develop predictive nanoinformatics models, which can be used to either design ENMs with specific properties (Melagraki and Afantitis, 2014; Varsou et al., 2019; Ling et al., 2018; Martin et al., 2019) or to predict their behaviour and effects (Afantitis et al., 2018; Toropova et al., 2016; Basei et al., 2019). One of the limitations ll to the widespread application of in silico

---

approaches is the lack of large quantities of high-quality data, or of data with adequate metadata that will allow dataset interoperability and their combination to create larger datasets. The effects of a lack of sufficient data or metadata on result reliability have been demonstrated in several meta-analysis studies (Bilal et al., 2019; Gernand and Casman, 2014; Oh et al., 2016; Ha et al., 2018; Shin et al., 2018; Labouta et al., 2019). In fact, a number of studies (Bilal et al., 2019; Oh et al., 2016; Ha et al., 2018; Labouta et al., 2019) found a significant correlation between the assay type and cytotoxic effects, with another (Shin et al., 2018) demonstrating the correlation between the ENM dispersion protocols and cytotoxicity. Furthermore, Labouta et al. (2019) demonstrated the advantages of data combination and computational approaches (decision trees) using the meta-analysis of 93 peer-reviewed papers, corresponding to over 3000 data points, on the cytotoxicity of organic and inorganic ENMs, to uncover hidden relationships. Thus, as more data is generated, and data management practices improve (Papadiamantis et al., 2020a), the power of in silico approaches will increase exponentially.

The role of metadata in facilitating dataset combination and interoperability becomes even more prominent when dealing with parameters which are dependent on the ENMs surrounding environment and can affect its behaviour, the so-called extrinsic parameters (Lynch et al., 2014; Casals et al., 2017). One example of an extrinsic, or context-dependent ENM parameter is ζ-potential, which corresponds to the electrostatic potential of the ENM at the slipping plane; i.e., at the surface where the ENM is considered to interact with its surrounding environment (Fig. 1). The size of this surface depends on both the properties of the ENM, e.g. size, coating, and the ionic strength of the surrounding solution (Lowry et al., 2016). In practice, it consists of an ionic liquid layer (Stern layer), which is strongly bound and moves with the particle as it flows inside the solution (Brownian motion) and an outer loosely bound layer (the diffusion layer) (Lowry et al., 2016). While in uncoated ENMs (Fig. 1, left) the Stern layer, and subsequently

ζ-potential, is outside the core of the ENM, this is not the case for polymeric coated ENMs (Fig. 1, middle and right). The ζ-potential of polymeric coated ENMs will be affected by both the bare ENM's ζ-potential and the charge of the respective coating, as the polymeric coating will act as a form of cation-exchange membrane separating the Stern from the diffusion layer (Batley and Apte, 2005). The charge of the polymeric coating is called the Donnan potential (Batley and Apte, 2005; Ohshima, 2006). Assuming a negatively charged ENM, covered with a negatively-charged polymeric coating (Fig. 1, middle and right), cationic species will be transported through the coating, driven by the negative charge, until the Stern layer is formed (Batley and Apte, 2005). If the coating is sufficiently thick (Fig. 1, middle), it will mask the bare particle ζ-potential and the apparent potential will be almost equal to the Donnan potential (Lowry et al., 2016; Ohshima, 2006). For thinner layers (Fig. 1, right), the apparent ζ-potential will be affected by both the bare particle ζ-potential and the coating charge and for a positively charged bare particle ζ-potential and negatively charged coating, it will be less than the Donnan potential (Lowry et al., 2016).

ζ-potential is considered one of the key ENM physicochemical parameters which need to be reported in regulatory dossiers, and several standards (e.g. ISO 13099) exist regarding its measurement. ζ-potential has been linked with ENM stability and its behaviour in solution, as well as with its toxicity. Furthermore, ζ-potential has substantial applications in nanoscience research broadly, as it can regulate the behaviour and functionality of ENMs when used in fields like nanomedicine (Honary and Zahir, 2013a; Honary and Zahir, 2013b), energy production (Sen et al., 2017; Caimi et al., 2018), fuel and combustion (Saxena et al., 2017; Kumar et al., 2020), construction (Ogunsona et al., 2020), consumer products (Ogunsona et al., 2020), water treatment (Ogunsona et al., 2020) and remediation of environmental heavy metal pollution (Yang et al., 2013).

A high value (empirically $\geq \pm 30$ mV) is considered to provide sufficient electrostatic repulsion between ENMs to prevent agglomeration
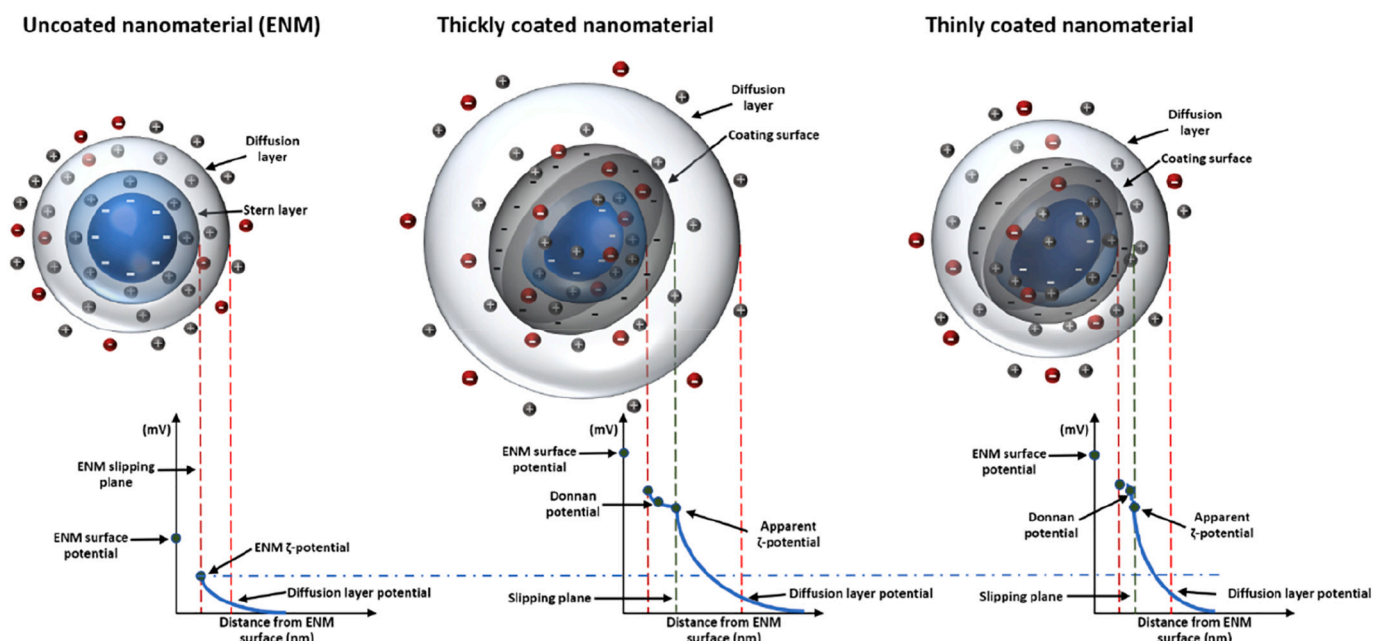
Fig. 1. Schematic representation of a negatively charged uncoated (left), a negatively charged thickly polymer-coated (middle) and negatively charged thinly polymer-coated (right) ENM. In the uncoated ENM the Stern layer consists of firmly bound positively charged ions, which move with the ENM. Positive and negative charges outside the slipping plane and within the ENM's diffusion layer are loosely bound to the ENM and remain with the bulk fluid. In this case, the ζ-potential corresponds to the electric charge at the slipping plane. For a negatively charged and thickly coated ENM the Donnan potential falls between the metal core and the coating surface and is approximately equal to the apparent ζ-potential. For a similar thinly coated ENM, the apparent ζ-potential is less than the Donnan potential and is affected by the bare ENM ζ-potential, the coating charge and the surrounding solution. In the image, white and black dashes correspond to the charge of the metal core and the coating respectively. Grey and red spheres correspond to positive and negative free radicals (counterions) inside the solution. Image adapted from Lowry et al. (2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(in the absence of steric stabilisation resulting from coatings) (Fatehah et al., 2014), while the diffusion layer (regulated by the ζ-potential) plays a vital role in ENM dissolution (Borm et al., 2006). Increasing medium ionic strength, changes in pH, and subsequent alterations in ζ-potential can affect the way in which ENM react with the cell membrane, altering the degree of ENM uptake by organisms or cells. This will influence the extent of reactive oxygen species (ROS) production, which is the most common adverse effect following exposure to ENM (Abdal Dayem et al., 2017). As a result, being able to predict an ENM's ζ-potential based on specific characteristics such as its elemental composition, will allow the development of more stable, functional and safer ENMs. Previous ζ-potential related predictive studies (Varsou et al., 2020a; Mikolajczyk et al., 2015; Toropov et al., 2018; Wyrzykowska et al., 2016; Toropov et al., 2016) have mainly focussed on metal oxides, whereas recent work by Varsou et al. (Varsou et al., 2020a) has developed a fully validated model for the prediction of ζ-potential of both metal oxide and metal ENMs. These previous studies developed ζ-potential predictive models through the analysis of TEM images of existing ENMs (Varsou et al., 2020a; Mikolajczyk et al., 2015), predicted the ζ-potential of ENM in different media based on known aqueous ζ-potential values (Wyrzykowska et al., 2016), or did not take into account different surface characteristics (coating, functionalisation) (Toropov et al., 2018; Toropov et al., 2016). Most studies used literature curated data (Mikolajczyk et al., 2015; Toropov et al., 2018; Wyrzykowska et al., 2016; Toropov et al., 2016) that could result in discrepancies regarding the experimental protocols used for data generation. Protocol variation has been shown to have a significant effect on the acquired results (Bilal et al., 2019; Gernand and Casman, 2014; Oh et al., 2016; Ha et al., 2018; Shin et al., 2018; Labouta et al., 2019; Langevin et al., 2018) and should be taken into account (potentially as an extra variable) while preparing datasets for in silico exploitation. In our case, the ENM physicochemical characterisation data were all generated within one project (the Framework Program 7 (FP7) project NanoMILE, 2013), which minimises the risk of protocol variation.

The aim of this study was to enrich this library containing the full physicochemical characterisation of 69 ENMs, with molecular descriptors to increase the value of the available information. The enriched dataset was used to develop an in silico workflow to predict ENM ζ-potential based on a number of descriptors that can be used as part of SbD approaches for design and production of safer and more functional ENMs. The read across predictive model has been made publicly and freely available as a webservice through the Horizon 2020 (H2020) NanoCommons project (http://enaloscloud.novamechanics.com/nanocommons/mszeta/) and the H2020 NanoSolveIT Cloud Platform (https://mszeta.cloud.nanosolveit.eu/) to ensure accessibility to the community and interested stakeholders.

2. Experimental section

2.1. Experimental data

The experimental dataset used for the development of the nanoinformatics predictive model was generated during the FP7 NanoMILE project (2013). The dataset, published by Joossens et al. (2019), contains the physicochemical characterisation in water of 69 ENM and is available through the NanoPharos database (NovaMechanics Ltd., 2020) developed under the H2020 NanoSolveIT (Afantitis et al., 2020a) and (NanoCommons (2018)) www.nanocommons.eu projects. The dataset consists of pristine Ag ($n = 3$), AgO ($n = 1$), Ag_2S ($n = 1$), AlOOH ($n = 1$), Au ($n = 5$), $BaTiO_3$ ($n = 1$), CeO_2 ($n = 13$), Zr-doped CeO_2 ($n = 5$) and their ($n = 5$) aged equivalents, $CePO_4$ ($n = 1$), CuO ($n = 1$), Fe_2O_3 ($n = 1$), hydroxylapatite ($n = 1$), SiO_2 ($n = 14$), TiO_2 ($n = 8$), ZnO ($n = 7$) and ZrO_2 ($n = 1$) ENMs. The dataset contains information on the core chemistry, where applicable coating type and coating charge (positive, negative, neutral), morphology (identified using TEM) and whether the particles were aged or not (no ageing protocol was provided).

Furthermore, the physicochemical characterisation of ENM in MilliQ water includes: core size (measured using TEM or STEM), hydrodynamic diameter, ζ-potential (although the data producers have not provided information on the concentrations used during the ζ-potential measurements but it was performed in parallel to the dynamic light scattering measurements and as such was typically ~10 mg/L), specific and geometric surface area (calculated using primary particle size and morphology) and energy band gap (calculated from UV–Vis measurements using Tauc plots (Tauc et al., 1966)). The characterisation methodologies and protocols used for the ENM characterisation can be found in (Joossens et al., 2019; Römer et al., 2019; Briffa et al., 2019), while Fig. 2 presents an overview of the metals present in the ENM library along with basic atomic parameters and atomic periodicity trends. All information was handled as data points (descriptors), either nominal (core chemistry, coating type, coating charge, morphology, ageing) or continuous (core size, diameter, ζ-potential, specific and geometric surface area, energy band gap).

2.2. Data pre-processing and enrichment

The NanoMILE dataset was initially evaluated in terms of completeness. To increase the model robustness and reliability all descriptors with data gaps that could not be filled either experimentally or using suppliers' information were excluded. Hence, 7 descriptors were included in the model development. Core chemistry, coating, coating charge, ageing, primary particle size, morphology and geometric surface area were used as the independent variables and ζ-potential was used as the dependent variable.

Taking into account that ζ-potential is regulated by the intrinsic molecular properties of the ENM, the dataset was enriched using a number of molecular descriptors, which have been used in past studies for model development and computational workflows (Kar et al., 2014; Zhang et al., 2012). The descriptors used are metal atomic radius (r_{at}), metal ionic radius (r_{ion}), metal electronegativity (χ), sum of metal electronegativity ($\Sigma\chi$), sum of metal electronegativity divided by the number of oxygen atoms present in a particular metal oxide ($\Sigma\chi/n_O$), number of metal atoms (N_{Metal}), number of oxygen atoms (N_{Oxygen}), the charge of the metal cation corresponding to a given oxide (χ_{ox}), the molecular weight (MW) and the period and group of the metal in the periodic table. Furthermore, the absolute electronegativity (Mulliken electronegativity, χ_{abs}), which is a measure of the tendency of an atom to attract electrons, and energy band gap (*EBG*) were calculated using the equations reported by Portier et al. (2001a, 2001b). In short, χ_{abs} was calculated using:

$$\chi_{abs} = 0.45\chi(eV) + 3.36 \tag{1}$$

and *EBG* was calculated using:

$$EBG = Ae^{0.34E_{\Delta H}}, \tag{2}$$

where *EBG* is the energy band gap in eV, $E_{\Delta H}$ is the standard enthalpy of formation of the ENM and A is a pre-exponential constant, which varies from 0.5–1.7 depending on the metal. $E_{\Delta H}$ values were taken from Portier et al. (2001a, 2001b) or other literature or database sources. The A value was taken from Portier et al. (2001b) where possible. For the rest, the average A value reported by Portier et al. (2001a, 2001b), depending on the position of the metal in the periodic table was used:

- For s- and f-block elements the average A value is 0.80
- For d-block elements the average A value is 1.00
- For p-block elements the average A value is 1.35

All data are available through the NanoPharos database (https://db.nanopharos.eu/Queries/Datasets.zul) (NovaMechanics Ltd., 2020) and as a supplementary file.
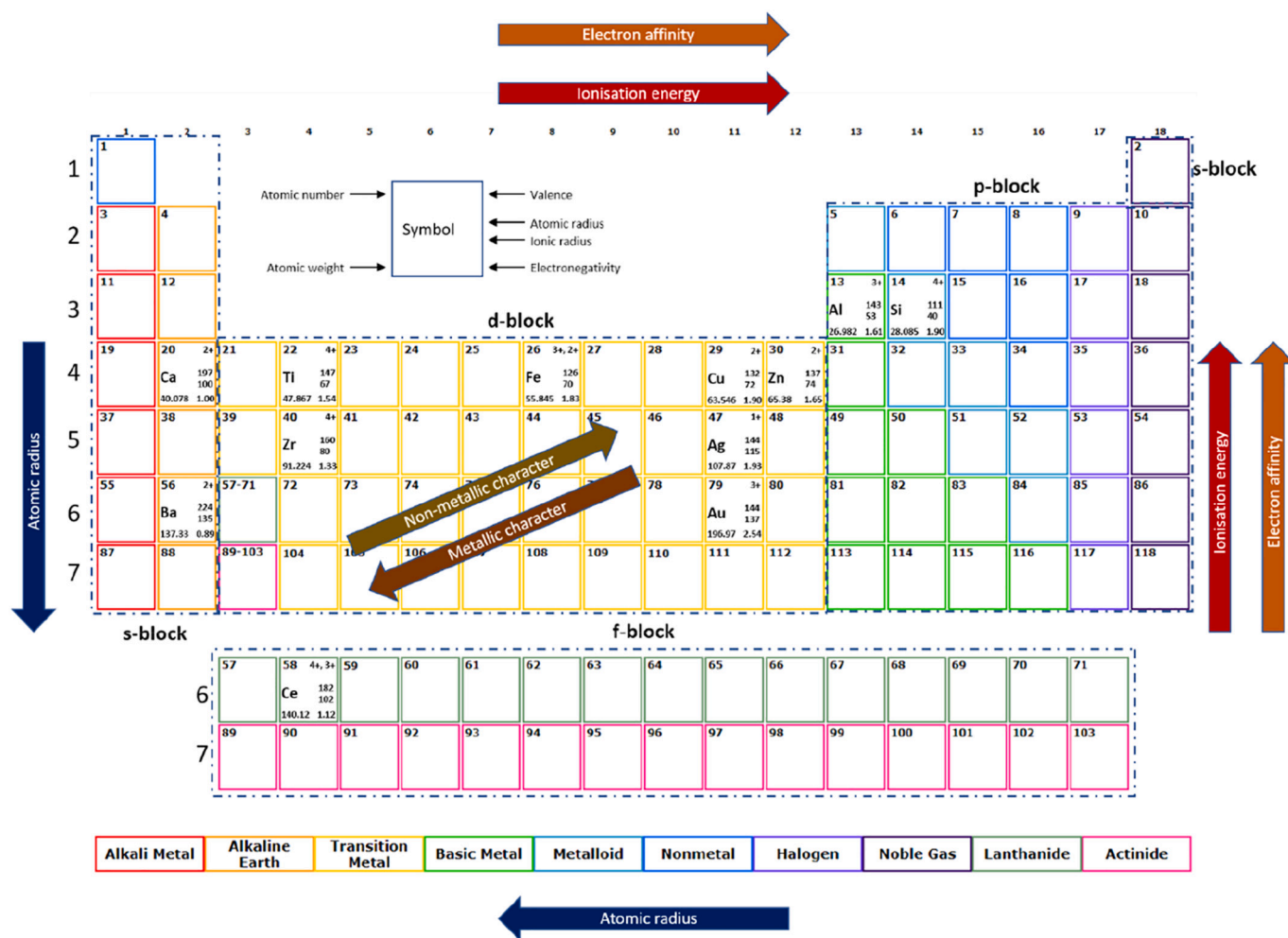
Fig. 2. Schematic overview of the metals included in the NanoMILE ENM library along with their atomic characteristics (atomic number, atomic weight, valence, atomic and ionic radii, electronegativity). The schematic presents the recurring atomic trends (i.e. atomic periodicity) to assist with the visual understanding of the chemical similarities and differences between the different metals.

2.3. Model development and read across

Model development was performed using the Isalos Analytics Platform, powered by the Enalos+ Tools (Varsou et al., 2018; Afantitis et al., 2020b). The dataset was first passed through a low-variance filter to remove descriptors that did not present significant variance and thus did not contribute to the discrimination power of the model (Ojha and Roy, 2011). The upper variance bound was set to 0.2, thus all descriptors having 20% or more equal values to another descriptor were excluded. Due to the varying numerical ranges of the used descriptors, *Z*-score normalisation was then applied to the dataset to ensure that the values of all descriptors followed Gaussian distribution with a mean value of 0.0 and a standard deviation of 1.0 (Leach and Gillet, 2007). Random partitioning of the data into training and test sets took place using a ratio of 75%: 25% respectively. The training set was used to train the model and to blindly evaluate its performance, following parameter tuning through cross-validation (Bishop, 2007; Xu and Goodacre, 2018). The most significant descriptors, those contributing the most to dataset variance, were then identified using the Correlation based Feature Selection (CfsSubset) algorithm combined with the BestFirst evaluator (Hall et al., 2009; Witten et al., 2016).

The ζ-potential predictive model was developed using the Enalos implementation of the *k*-nearest neighbours (Enalos*k*NN, Enalos Chem/Nanoinformatics tools) regression methodology (Papadiamantis et al., 2020b). This is an instance-based (lazy) method used for data

classification and regression. Enalos*k*NN predicts the unknown endpoint based on the k (k = 1, 2, 3, …) nearest neighbours in the features space R^n where *n* is the total number of descriptors used for the prediction. ζ-potential prediction was achieved based on the Euclidian distance (similarity measure) of the target variable from its closest neighbours (Witten et al., 2016). The prediction was then performed using the weighted average of the ζ-potential values of these neighbours, with the weighing factor for each neighbour being the inverse of its Euclidean distance (Varsou et al., 2019; Witten et al., 2016). For nominal descriptors, the individual values are compared and if the values are the same the Euclidean distance is set equal to 0, otherwise it is set to 1 (Larose and Larose, 2014). The optimal number of neighbours (k) was identified based on the best model performance.

The *k*NN algorithm can be used within a read-across strategy that can also be employed for ENMs. According to ECHA's read across framework (ECHA, 2017) a computational workflow needs need to include the following steps:

1. Gathering of the required descriptors (physicochemical, molecular) for each ENM.
2. Construction of a data matrix including properties and endpoints.
3. Development of an initial grouping hypothesis that correlates an endpoint to different behaviour and reactivity properties. Assignment of the samples to groups.

4. Assessment of the applicability of the approach using computational techniques, and data gaps filling. If no regular pattern can be identified, an alternative grouping hypothesis must be proposed.
5. In the case that the grouping hypothesis is robust, but adequate data are not available, additional testing should be considered.
6. Justification of the method.

Using the Euclidian distance as a metric, it is possible to predict ζ-potential and identify the respective groups of neighbouring ENMs. By identifying and studying the resulting groupings it is possible to map the prediction space into specific groups that can be translated into read across strategies, as per the requirements of ECHA's read across framework. The Enalos*k*NN node has the benefit of providing not only the predictive results, but also the specific neighbours along with their Euclidian distances, as well allowing the visualization of the entire predictive space.

2.4. Model validation

Internal and external model validation followed the OECD principles for validation, for regulatory purposes, of (Q)SAR models (OECD, 2004). Goodness-of-fit, robustness and predictivity were used to validate the model internally and externally (Zhang et al., 2006; Puzyn et al., 2018). The statistical criteria used to evaluate the model's performance were the coefficient of determination between experimental values and model predictions (R^2), validation through an external test set, the leave-many-out cross validation procedure and Quality of Fit and Predictive Ability of a continuous predictive model according to Tropsha's tests (Tropsha, 2010). This was achieved using the Enalos Model Acceptability Criteria, where the following equations (to calculate Tropsha's tests) have been implemented:

$$R^2_{cvtext} = 1 - \frac{\sum_{i=1}^{ntest} (y_i - \widetilde{y}_i)^2}{\sum_{i=1}^{ntest} \left(y_i - \overline{y}_{tr}\right)^2} \tag{3}$$

$$k = \frac{\sum_{i=1}^{ntest} y_i \widetilde{y}_i}{\sum_{i=1}^{ntest} \widetilde{y}_i^2} \tag{4}$$

$$R^2_o = 1 - \frac{\sum_{i=1}^{ntest} \left(\widetilde{y}_i - \widetilde{y}_i^{ro}\right)^2}{\sum_{i=1}^{ntest} \left(\widetilde{y}_i - \overline{\overline{y}}\right)^2}, \text{where } \widetilde{y}_i^{ro} = ky_i, i = 1, 2, \ldots, ntest, \tag{5}$$

where *ntest* is the number of ENM in the test set, $\overline{y}_{tr}$ is the average ζ-potential for the training set; $y_i, \widetilde{y}_i, i = 1, 2, \ldots,$ ntest are the experimentally measured and the predicted ζ-potential values for the validation set, respectively; and $\overline{\overline{y}}$ is the average predicted ζ-potential over all of the model predictions $\widetilde{y}_i, i = 1, 2, \ldots,$ ntest.

Furthermore, according to Tropsha (2010), a predictive model is considered predictive if all of the below conditions are satisfied:

$$R^2_{cvtext} > 0.5 \tag{6}$$

$$R^2_{pred} > 0.6 \tag{7}$$

$$\frac{R^2_{pred} - R^2_o}{R^2_{pred}} < 0.1 \tag{8}$$

$$0.85 \leq k \leq 1.15 \tag{9}$$

The statistical significance and robustness of the produced model was also evaluated using Y-randomisation. During Y-randomisation the

modelling calculations were repeated several times using all original descriptors and shuffling the ζ-potential predictions to produce 10 different datasets. The model acceptability criteria, described above, were calculated in each case and were expected to be reduced compared to those of the original model. This would demonstrate that the developed model was not due to chance correlation, as in this case we would also get similar acceptability criteria. If this was the case (i.e., if the model was based on chance correlation), it would not be possible to obtain a valid model using the specific partitioning (Tropsha et al., 2003).

2.5. Domain of applicability

To ensure the applicability of the produced model to external datasets it is important to define the limits within which future predictions will be considered reliable; i.e., the applicability domain (APD). If the model is applied to ENMs that are too different from those used to train the model, the predictions will be outside the defined limits, these will be flagged as unreliable and filtered out (Melagraki and Afantitis, 2014). The APD can be defined through the calculation of the Euclidian distances of all ENM in the training set. The APD was calculated using:

$$APD = <d> + Z\sigma \tag{10}$$

where $<d>$ and σ are the average and standard deviation of all Euclidian distances in the training set. *Z* is an empirical cut-off value and was set to 0.5 (Zhang et al., 2006). All calculations were performed using the Isalos Analytics Platform (Papadiamantis et al., 2020b; Afantitis et al., 2020b) and the full demonstration that the produced model meets the OECD criteria as listed above is demonstrated via the completed QSAR Model Reporting Format (QMRF) template which is included in the supplementary information (SI).

3. Results

The main purpose of this study was the development of a robust and predictive model for the prediction of ζ-potential that can be used in SbD approaches for the design and development of novel ENM. Using the Isalos Analytics Platform (Papadiamantis et al., 2020b)(Afantitis et al., 2020b) and the respective Enalos+ nodes (Varsou et al., 2018) we worked with a dataset (Joossens et al., 2019) of 69 ENM, which was further enriched with a number of molecular descriptors. The benefits of using this dataset is that the physicochemical characterisation was performed using harmonised protocols, which reduced the risk for data variability due to experimental practices which has been identified as a major obstacle for data interoperability and reusability. This eliminates the need to identify protocols used and include them as potential descriptors, the significance of which has been demonstrated in previous meta-analysis studies where the type of assay was identified as a key parameter in ENM toxicity studies (Bilal et al., 2019; Gernand and Casman, 2014; Oh et al., 2016; Ha et al., 2018; Shin et al., 2018; Labouta et al., 2019). Following exclusion of incomplete descriptors and enrichment with calculated descriptors, the used dataset includes a total of 20 descriptors - 7 physicochemical and 13 molecular descriptors (Table 1).

Model development took place by randomly dividing the dataset into training and test sets using a ratio of 75%: 25%. The descriptors of the training set were normalised using *Z*-score normalisation and the applied normalisation parameters were then used to normalise the test set. The most significant parameters, those contributing the most to dataset variability, were identified using the CfsSubset algorithm combined with the BestFirst evaluator (Hall et al., 2009; Witten et al., 2016). Out of the 20 descriptors imported into the computations, as none of the descriptors presented variance lower than 20%, 5 were identified as the most significant. Two of those were physicochemical; i.e. CT and SZ, and 3 were molecular; i.e. r_{ion}, $\Sigma\chi/n_O$ and χ_{abs}.

Table 1

Physicochemical and molecular descriptors used for model development.

Physicochemical descriptors	Molecular descriptors
Chemical formula (CF)	Metal atomic radius (r_{at})
Coating/ functionalisation (CT)	Metal ionic radius (r_{ion})
Coating charge (CTC)	Metal electronegativity (χ)
Ageing (AGE)	Sum of metal electronegativity ($\Sigma\chi$)
Core size (SZ)	Sum of metal electronegativity divided by the number of oxygen atoms present in a particular metal oxide ($\Sigma\chi/n_O$)[a]
Morphology (MP)	Number of metal atoms (N_{Metal})
Geometric surface area (GSA)	Number of oxygen atoms (N_{Oxygen})
	Metal charge (χ_{ox})
	Molecular weight (MW)
	Metal periodic table group
	Metal periodic table period
	Absolute electronegativity (χ_{abs})
	Energy band gap (EBG)

[a] For metallic ENM, where n0 = 0, $\Sigma\chi/n_O$ was assumed to be 0 to assist with computations.

The developed workflow provided the opportunity to test a number of different predictive algorithms. This enabled us to pick kNN as the best performing algorithm for the specific dataset. kNN was applied to the training dataset with k = 7, which was the value that presented the best model performance (Table 2), based on the cross validated R^2_{cvext} performnce. The R^2 (R^2_{Pred}) value; i.e., the coefficient of determination between experimental values and model prediction on the test set, also presented its maximum value ($R^2 = 0.900$, Table 2) for k = 7. This denotes a substantial correlation between the values measured experimentally and the predicted values (presented graphically in Fig. S1 of the ESI).

The model was validated as described in the Experimental Section as per the OECD's guidelines. The produced model successfully passed Tropsha's tests (Tropsha, 2010) (Table 3) demonstrating the robustness and predictivity of the model. The calculation of the presented parameters was performed using the regression between the experimentally measured ζ-potential vs. the predicted values and vice versa. Similarly, Y-randomisation demonstrated the model's robustness and validity. Finally, the APD to test potential model limitations was calculated to provide future users with an indication of the reliability of their predictions. The APD limit was calculated, in this case, to be 1.947 with all predictions being deemed reliable (test set range: 0–1.605).

The results acquired from the EnaloskNN node were used to study the neighbouring space relative to the Euclidian distance between the ENM descriptors. This allowed us to perform an initial grouping of the ENM based on their distances (Fig. 3). The results demonstrate certain patterns with respect to grouping. In most cases, grouping has been performed with chemically similar ENM when enough neighbours are available. SiO_2 ENM are grouped with other types of SiO_2 ENM only. CeO_2 ENM are grouped with other CeO_2 ENM types, while Zr-doped CeO_2 ENM are grouped with other similarly doped ENM and ZrO_2. Ag

Table 2

k value selection based on model performance (R^2).

k value	R_{cvext}	R^2	k value	R_{cvext}	R^2
1	0.450	0.584	8	0.908	0.896
2	0.764	0.765	9	0.903	0.890
3	0.787	0.771	10	0.908	0.896
4	0.865	0.850	11	0.901	0.887
5	0.888	0.874	12	0.903	0.890
6	0.896	0.881	13	0.901	0.887
7	**0.913**	**0.900**	14	0.701	0.888

The bold values highliight the optimal number of neighbours for the *k*NN model to have the best performance, as both the R^2_{cvext} and the R^2_{Pred} present maximum values.

Table 3

Model validation criteria and acquired results.

Criterion	Result	Assessment
$R^2 > 0.6$	0.9	Pass
$R_{cvext} > 0.5$	0.913	Pass
$\frac{R^2 - R_0^2}{R^2} < 0.1$	0.005	Pass
$\frac{R^2 - R_0'^2}{R^2} < 0.1$	0.001	Pass
$|R^2 - R_0'^2| < 0.3$	0.004	Pass
$0.85 < k < 1.15$	1.053	Pass
$0.85 < k' < 1.15$	0.869	Pass

ENM are generally grouped with other types of Ag ENM (metallic Ag, AgO, Ag_2S) Au ENM are clustered with other metals from the same group of the periodic table and with hydroxyapatite ENM with which they have similar χ_{abs} and close r_{ion}. AgO is grouped with a number of metallic (Ag), and metal oxide (ZnO, CuO and Fe_2O_3) ENM. ZnO ENM are grouped with either ZnO ENM or ENM belonging to the same periodic table period (TiO_2, CuO, Fe_2O_3). $BaTiO_3$ is grouped with CeO_2 and Zr-doped CeO_2 ENM. Barium is in the same period as Ce^{4+} and Ti^{4+} and in the same group as Zr^{4+}. Finally, TiO_2 ENM are grouped with other TiO_2 ENM, ENM from the same period (Fe_2O_3, ZnO), CeO_2 and AlOOH ENM.

The ζ-potential predictive model has been made available through the Enalos Cloud Platform (Varsou et al., 2020b) to facilitate further use within the wider nano-community and interested stakeholders. The corresponding webservice can be found at:

http://enaloscloud.novamechanics.com/nanocommons/mszeta/

The service is complemented with a user-friendly interface (Fig. 4). The user needs to input the required values (ENM size, coating, metal r_{ion}, $\Sigma\chi/n_O$ and χ_{ab}) that were selected as the most significant descriptors during model development. In the case of the coating type, the user needs to pick from the available coatings in the dropdown menu, as, currently, only these coatings are supported in the model as they constituted the training set. A note has been added to the tool (Fig. 4) warning the users that the prediction assumes specific dispersion conditions, i.e., in water at pH 7 and with a theoretical ionic strength of $10^{-6.998}$ mol/L. The linked tutorial provides a short library of the molecular properties (r_{ion}, $\Sigma\chi/n_O$ and χ_{ab}) used during model development that are available for re-use, since these are ENM size/coating independent. Upon submission of the values, the results page appears (Fig. 5) indicating the value of the predicted ζ-potential along with the 7 closest neighbours that were used to perform the prediction. The Euclidean distances of each neighbour from the tested ENM are provided as well. Furthermore, the prediction's reliability is stated to indicate whether the prediction falls within the calculated APD or not, and the reliability value is next to the domain value calculated for the specific ENM, based on its distances from the identified closest neighbours. All results appear on screen, and can be downloaded as a .CSV file.

4. Discussion

ζ-potential is a key parameter from a regulatory context, as it can directly affect ENMs behaviour in different media and their interaction with biological organisms. However, ζ-potential's correlation with medium properties like ionic concentration and chemistry makes it difficult to measure and interpret (Kirby and Hasselbrink Jr., 2004) and highly context dependent. Additionally, the calculations for ζ-potential assume that measurements are performed on spherical and reasonably monodisperse particles, which is often not the case for ENMs. This results in substantial variability in the ζ-potential measurements reported for the same ENM by different laboratories, on top of variability due to different experimental protocols used.

An in silico workflow for the development of a robust and validated ζ-potential kNN predictive model is presented. The dataset published by
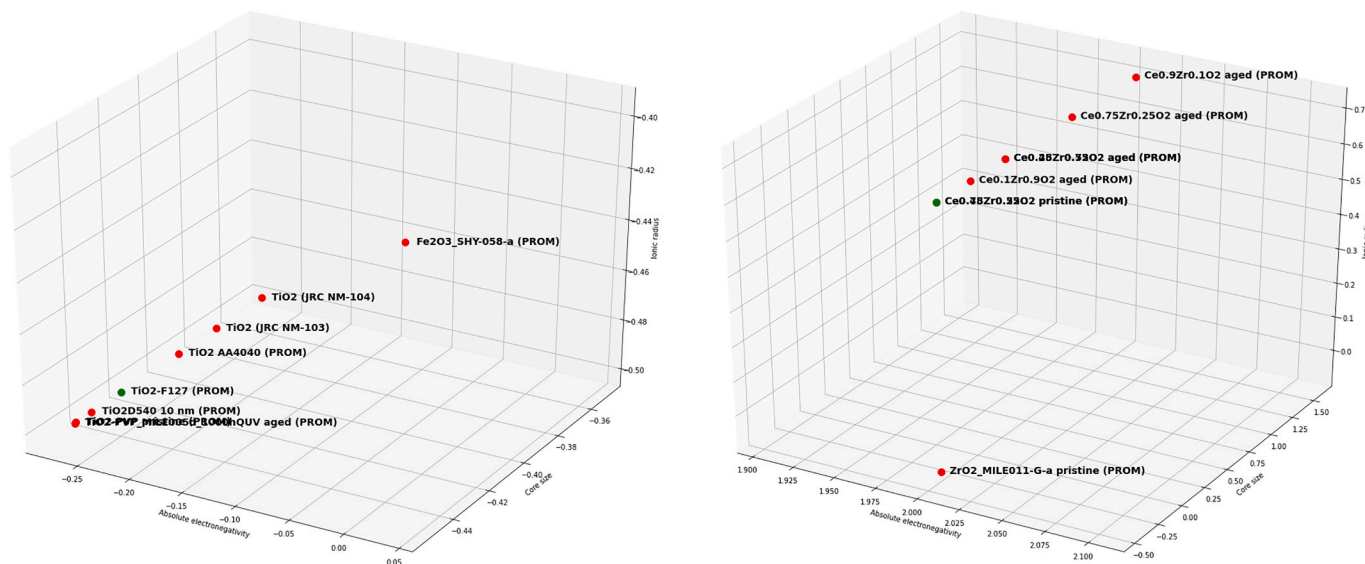
Fig. 3. Indicative results for TiO$_2$ (left) and Zr-doped CeO$_2$ (right) ENMs from the normalised kNN space of the produced ζ-potential predictive model. The ENMs studied are placed based on their Euclidian distances. Some overlap is observed due to small Euclidian distances. Red and green spheres correspond to ENMs from the training and test sets respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. MS^{3}bD (MScubed: Molecular, Size and Surface based Safe by Design) Zeta Potential predictive model webservice for the prediction of ENM ζ-potential. The user inputs the required parameters. For the coating type, the user needs to pick from those available in the dropdown menu, while for the rest of the parameters the user needs to manually input them. The calculation is then performed automatically.

Joossens et al. (2019), which contained the physicochemical characterisation in water of 69 ENM, was exploited. Eight descriptors were included in the model development: core chemistry, coating, coating charge, ageing, primary particle size, morphology, geometric surface area and ζ-potential. The dataset was evaluated regarding its completeness and enriched using 12 molecular descriptors: metal atomic radius (r_{at}), metal ionic radius (r_{ion}), metal electronegativity (χ), sum of metal electronegativity ($\Sigma\chi$), sum of metal electronegativity divided by the number of oxygen atoms present in a particular metal oxide ($\Sigma\chi/n_O$),

number of metal atoms (N_{Metal}), number of oxygen atoms (N_{Oxygen}), the charge of the metal cation corresponding to a given oxide (χ_{ox}), the molecular weight (MW), the period and group of the metal in the periodic table and the absolute electronegativity (Mulliken electronegativity, χ_{abs}).

The resulting dataset was passed through a low-variance filter to remove any descriptors that did not present significant variance and thus did not contribute to the discrimination power of the model (Ojha and Roy, 2011) and the dataset was then divided into training and test

MS³bD Zeta Potential Predictive Model Results

Download files

ID	Zeta potential prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5	Closest NN6	Distance from NN6	Closest NN7	Distance from NN7	Domain	APD	Prediction
Ag	-1.931	Ag JRC NM-300K	0.016	Fe2O3_SHY-058-a (PROM)	1.290	AgPURE_15nm	1.705	Ag2S 36 nm PVP	1.707	CeO2 JRC NM-212	2.123	CeO2-PVP40 (UoB)	2.123	CeO2-PVP10 (UoB)	2.123	0.016969	1.570952	reliable
Ag2S	-42.904	AgPURE_15nm	0.034	Ag2S 36 nm PVP	1.001	CeO2 (PROM)	1.179	CeO2 JRC NM-212	1.181	CeO2-PVP40 (UoB)	1.181	CeO2-PVP10 (UoB)	1.182	CeO2-IKR1_ (UoB)	1.182	0.034071	1.570952	reliable
SiO2	-22.771	SiO2-COOH (JRC) 50-60 nm	0.0	SiO2-uncoated (JRC) 50-60 nm	0.0	SiO2-COOH (JRC) 25-30 nm	0.100	SiO2-uncoated (JRC) 25-30 nm	0.100	SiO2-COOH (JRC) < 20 nm	0.188	SiO2-COOH (JRC) 100 nm	0.225	SiO2-NH2 (JRC) < 20 nm	0.227	0.000000	1.570952	reliable
CeO2	5.743	CeO2 (PROM)	0.0	CeO2 JRC NM-212	1.001	CeO2-PVP40 (UoB)	1.001	CeO2-PVP10 (UoB)	1.002	CeO2-IKR1_ (UoB)	1.002	CeO2-IK56_ (UoB)	1.003	CeO2-IKC3 (UoB)	1.004	0.000000	1.570952	reliable
Au	-22.435	Au 12 nm PEG-Ome (JRC-IO166D)	0.022	Au 12 nm PEG-OH/PEG-Ome (50:50) (JRC-IO166C)	0.022	AlOOH_SHY-226-B-a (PROM)	1.000	AgPURE_15nm	1.399	Ag2S 36 nm PVP	1.402	CeO2 JRC NM-212	1.573	CeO2-PVP40 (UoB)	1.573	0.022332	1.570952	reliable
CeZrO2	43.881	Ce0.92Zr0.1O2 aged (PROM)	0.002	Ce0.1Zr0.9O2 pristine (PROM)	0.325	Ce0.25Zr0.75O2 pristine (PROM)	0.666	Ce0.48Zr0.52O2 pristine (PROM)	0.666	Ce0.1Zr0.9O2 aged (PROM)	0.755	Ce0.75Zr0.25O2 pristine (PROM)	1.349	Ce0.9Zr0.1O2 pristine (PROM)	2.031	0.002932	1.570952	reliable
TiO2	-9.455	TiO2-PVP pristine (PROM)	0.0	TiO2 uncoated (PROM)	6.141	TiO2 JRC NM-104	1.000	TiO2-F127 (PROM)	1.001	TiO2 AA4040 (PROM)	1.002	TiO2 JRC NM-103	1.004	SiO2 20 nm Dodecyl (JRC-IO 168C Silica)	1.032	0.000000	1.570952	reliable
TiO2	16.992	TiO2-F127 (PROM)	1.878	TiO2 uncoated (PROM)	1.000	TiO2 JRC NM-104	1.001	TiO2 JRC NM-103	1.001	TiO2 uncoated (PROM)	1.001	TiO2-PVP pristine (PROM)	1.001	SiO2 20 nm Dodecyl (JRC-IO 168C Silica)	1.032	0.000019	1.570952	reliable
TiO2	-42.789	TiO2 JRC NM-104	0.0	TiO2 uncoated (PROM)	1.000	TiO2-PVP pristine (PROM)	1.000	TiO2-F127 (PROM)	1.001	TiO2 AA4040 (PROM)	1.002	TiO2 JRC NM-103	1.004	SiO2 20 nm Dodecyl (JRC-IO 168C Silica)	1.032	0.000000	1.570952	reliable
ZnO	-2.719	Au 12 nm PEG-COOH (JRC-IO166E)	1.0	TiO2D540 (PROM)	1.015	TiO2 JRC NM-103	1.096	ZnO-TECS (JRC NM-111) Nanogenotox protocol	1.099	TiO2 AA4040 (PROM)	1.105	TiO2-F127 (PROM)	1.112	TiO2 JRC NM-104	1.129	0.000000	1.570952	reliable
Ca5(PO4)3OH	-8.911	Fe2O3_SHY-058-a (PROM)	4.734	Ag JRC NM-300K	1.290	Ag2S 36 nm PVP	1.957	AgPURE_15nm	1.968	CeO2 JRC NM-212	2.056	CeO2-IKR1_ (UoB)	2.062	TiO2D540 (PROM)	2.209	0.000047	1.570952	reliable
AlOOH	-26.033	ZnO-TECS (JRC NM111) 2939b_1000hQUV aged	1.481	TiO2D540 (PROM)	1.636	TiO2 JRC NM-103	1.753	TiO2 AA4040 (PROM)	1.766	Fe2O3_SHY-058-a (PROM)	1.793	Au 12 nm PEG-COOH (JRC-IO166E)	1.857	TiO2-F127 (PROM)	2.038	1.481366	1.570952	reliable
Fe2O3	63.728	CuO-PVP360 (UoB)	0.0	TiO2 AA4040 (PROM)	0.341	TiO2 JRC NM-103	0.348	TiO2D540 (PROM)	0.558	SiO2 20 nm Dodecyl (JRC-IO 168C Silica)	1.018	TiO2-PVP pristine (PROM)	1.053	TiO2 uncoated (PROM)	1.053	0.000000	1.570952	reliable

Fig. 5. The results provided from the MS³bD Zeta potential webservice include the ζ-potential prediction, the identities and distances of the closest neighbours on which the prediction was based. and whether the prediction falls within the model's APD (reliable prediction) or not (unreliable prediction), along with the calculated domain value and the model's APD value.

sets, using a ratio of 75%: 25% respectively. The training set was used to identify the descriptors contributing the most to the dataset variance. Five descriptors were deemed as the most statistically significant, two of which were physico-chemical, i.e., CT and SZ, and 3 of which were molecular, i.e., r_{ion}, $\Sigma\chi/n_O$ and χ_{abs}. These descriptors were used to train and validate the model using the training data set. To do this, the Enalos implementation of the kNN algorithm was used, the accuracy of which has been cross-checked using the established and widely used Weka kNN implementation in the Konstanz Information Miner (KNIME) open-source data analytics platform. The Enalos kNN algorithm offers the benefit of providing the calculated Euclidean distances to each of the nearest neighbours allowing their relative influence to be understood and visualised.

Based on the identified variables, the kNN model was trained, its parameters fine-tuned and validated internally and externally using the OECD principles for validation of (Q)SAR models for regulatory purposes (OECD, 2004). Goodness-of-fit, robustness and predictivity were used to validate the model internally and externally (Zhang et al., 2006; Puzyn et al., 2018), while the coefficient of determination between experimental values and model predictions (R^2), validation through an external test set, the leave-many-out cross validation procedure and Quality of Fit and Predictive Ability of a continuous predictive model according to Tropsha's tests (Tropsha, 2010) were used to evaluate the model's performance (Table 3). Furthermore, the model's robustness and statistical significance were tested using Y-randomisation. Finally, the APD of the model was calculated using the Euclidean distances of all ENM in the training set, to be able to evaluate whether a future prediction is reliable or not. The APD limit was found to be 1.947 such that all predictions resulting in an APD value lower than this are considered reliable.

The descriptors identified as significant, and used for model development can be easily calculated using periodic table or bibliographic information (χ_{abs}, r_{ion}, $\Sigma\chi/n_o$,), or can be imported from the manufacturer's description of the ENM (ENM core size, coating) without the need for any experimental measurements. Thus, the developed model can be used as a SbD tool for the development of stable and safer ENMs, as ζ-potential can affect ENM behaviour by influencing agglomeration (Fatehah et al., 2014) or dissolution (Borm et al., 2006). ζ-potential has also been linked with ENM toxicity, in terms of reactivity with cell

membranes and ENM uptake from organisms (Schwegmann et al., 2010; Deryabin et al., 2015; Cho et al., 2012), especially for those ENM that undergo dissolution. Given that suspension medium ionic strength and pH changes affect ζ-potential and ENM dissolution, subsequent models can incorporate such information enabling prediction of dissolution and toxicity as part of a complete in silico workflow and IATA.

ζ-potential as a parameter is distinct from the surface charge of an ENM (Fig. 1). When suspended in an ionic solution, a negatively charged uncoated ENM (Fig. 1a) attracts ions from the surrounding solution. As a result, the electrical double layer (EDL) is formed between the surface of the ENM and the bulk of the solution (Schmickler, 2014). The innermost layer, called the Stern layer, is firmly bound to the surface of the particle and consists of positively charged ions, which are attracted to the particle via the Coulomb force and move around in the liquid with the ENM. The outermost layer, consisting of ions that are loosely bound to the rest of the particle and the Stern layer, due to electrostatic attraction, is called the diffusion layer (Schmickler, 2014). In fact, the EDL thickness can be calculated through the Debye length (λ_D, eq. 11), which expresses the distance over which the electrostatic effect of the ENM charge persists (Russel et al., 1989). For a colloidal or ionic solution, the Debye length is given by the equation:

$$\lambda_D = \sqrt{\frac{\varepsilon_r \varepsilon_0 k_B T}{e^2 \sum_i n_i z_i^2}} \tag{11}$$

where ε_r is the solution's dielectric constant, ε_0 is the free space permittivity, k_B is Boltzmann's constant, T the temperature in K, e is the electron charge, n_i is the number of ions in the solution and z_i is the valence of the ith ion (Russel et al., 1989).

ζ-potential is calculated at the slipping plane, where the EDL (and thus the Debye length) ends, using Henry's equation (Eq. (12)) for electrophoretic mobility (u_e) (Delgado et al., 2005):

$$u_e = \frac{2\varepsilon_{rs} \varepsilon_0}{3\eta} \zeta f_1(\kappa\alpha) \tag{12}$$

where ε_{rs} is the relative permittivity of the electrolyte solution, ε_0 is the electric permittivity of vacuum, η is the dynamic viscosity of the solution, ζ is the ζ-potential, κ is the inverse Debye length, α is the ENM size

and f_1 is Henry's function, the value of which varies from 1 to 1.5 based on the ratio of the ENM size to the Debye length ($\kappa\alpha$). Hence, electrophoretic mobility is directly correlated to the ENM size, confirming its presence as one of the key parameters for the prediction of ζ-potential. Interestingly, it was the ENM core size that was the key parameter rather than the experimentally measured hydrodynamic diameter that was found to be most predictive.

Furthermore, the presence of the Debye length in the ζ-potential formula (Eq. (12)); i.e., the thickness of the EDL, means that the measured ζ-potential, depends on the ability of an ENM to attract and redistribute ions. Based on studies on solid-solid electrification, which are dominated by electron transfers, Wang and Wang (2019) suggested, extending to solid-liquid interface, that the EDL is formed in two stages. Firstly, the solution molecules interact with the atoms on the surface of the solid and form strong overlaps of electron clouds, with electron transfers taking place to ionise the atoms on the solid surface to form the first electrostatic charges. These are followed by ion exchange, which is the dominant process thereafter, between the charged solid atoms and free H^+ and OH^- radicals in the solution, forming the EDL (Wang and Wang, 2019). These results were further supported by Lin et al. (2020), who demonstrated that both electron and ion transfers co-exist at the solid-liquid interface. In practice, this explains the identification of χ_{abs} as a significant parameter for the prediction of ζ-potential, as it is the average (Eq. (13)) of the ionisation energy (E_i) and the electron affinity (E_{ea}) and, according to Mulliken (Mulliken, 1934; Mulliken, 1935), expresses the tendency of an atom to attract electrons and thus create the EDL.

$$\chi_{abs} = \frac{E_i - E_{ea}}{2} \qquad (13)$$

This also explains why the ionic radius (r_{ion}) of metals has been identified as a significant parameter, also reported by Wyrzykowska et al. (2016), for the prediction of ζ-potential, which regulates the bond strength of the atom's valence electrons, and thus the energy needed to detach them and commence the electron exchange needed to form the EDL. The higher r_{ion}, the further the valence electrons are from the nucleus and, thus, the easier it is to remove an electron from its orbital. Similarly, metal electronegativity plays an important role in determining ζ-potential since its increase corresponds to a higher number of protons in the atom's nucleus, increasing its electron attracting potential. Thus, the summation of total metal electronegativity ($\Sigma\chi$) would provide an indication of the ENMs' total metal reactivity and ability to attract electrons. On the other hand, weighing $\Sigma\chi$ with the total number of oxygens present (n_O) would provide an indication of the total ENM electronegativity, and thus reactivity, and is based on Sanderson's electronegativity equalisation principle (Sanderson, 1951). Sanderson's principle states that when two atoms with different electronegativities combine, their electronegativities equalise in the formed molecule, with molecular electronegativity being the geometric mean of the electronegativities of the isolated atoms (Parr and Bartolotti, 1982). Hence, $\Sigma\chi/n_O$ can be considered as a measure of the reactivity of the entire ENM, with ions and electrons, explaining why it has been identified as a significant parameter in our model.

In the case of coated ENMs (Fig. 1b and c), ζ-potential measurement is more complicated and depends on both the ENM's EDL and the coating's charge and thickness (Lowry et al., 2016). For an ENM with a thick polymeric coating layer (Fig. 1b), which is thicker than the EDL, the coating completely masks the core ENM ζ-potential pushing the slipping plane to the surface of the coating (Ohshima, 2006). As a result, the ζ-potential measured in this case will correspond to the Donnan potential (Batley and Apte, 2005; Ohshima, 2006). This is the result of the Gibbs-Donnan effect of two ionic solutions separated by a semi-permeable membrane, the role of which is played by the coating, which are not distributed evenly (due to certain charged elements being unable to cross the membrane) resulting in a potential difference measured as the ζ-potential. For thinly polymer coated ENM (Fig. 1c),

the apparent ζ-potential will be influenced by both the coating charge (neutral, positive, negative) and the ζ-potential of the core ENM.

Having studied the kNN ENM grouping, certain trends could be observed that could provide a meaningful basis for ENM grouping and read-across. Elements, in general, group with other similar elements in terms of core chemistry, in most cases where enough neighbouring ($n = 7$) ENM species exist. This is the case for SiO_2 and CeO_2 ENM, which present the same molecular properties with their neighbours, but differ in terms of size and coating. Results demonstrate that, for same core chemistry, the closest neighbours are those with similar sizes, followed by those with different surface properties. These are followed by changes in size initially and then coating demonstrating the ranking of the descriptors significance. While the molecular properties will initially define the ENM properties and reactivity, the larger the ENM core size the lower the effect of the coating on EDL and ζ-potential. Similarly, Zr-doped CeO_2 ENM are grouped with other similar ENM with different doping ratios and ZrO_2 ENM, spanning the entire range from low to complete Ce^{4+} substitution by Zr^{4+} that can be considered a form of "grading" the behaviour of ENM based on varying doping ratios.

In the case where not enough similar neighbours exist, certain patterns of grouping can also be observed. Metallic Ag ENMs are grouped initially with other Ag ENMs (metallic Ag, AgO, Ag_2S) and are then grouped with Au ENM as they are in the same group of the periodic table, and thus present chemical similarity. AgO ENM are grouped with metallic ENM and ZnO, CuO and Fe_2O_3 ENM. Ag and Cu^{2+} are in the same periodic table group, and thus present chemical similarity, while Fe^{3+} and Zn^{2+} have electronegativities (Zn: 1.65, Fe: 1.83) close to those of Ag (1.93). The fact that AgO is grouped with other metal oxides and not with mostly metallic Ag or Au points out the significance of both metal electronegativity and oxygen presence and how these interact based on Sanderson's principle of electronegativity discussed above. The grouping of $BaTiO_2$ with CeO_2 and Zr-doped CeO_2 can be explained, initially, by the chemical similarity between Ti^{2+} and Zr^{4+}, which belong to the same periodic table group. Barium (Ba^{2+}) and Ce^{4+}, on the other hand, are in the same periodic table period (row) having the same number of electron shells and a difference of two protons (Ba: 56, Ce: 58) in their nuclei and close electronegativities (Ba: 0.9, Ce: 1.12). Period (row) grouping is also taking place in the case of ZnO ENM. These are grouped not only with other ZnO ENMs, but with TiO_2, CuO, Fe_2O_3 as well. This can be explained by the trends existing for electronegativity, atomic structure, electron affinity and ionisation energy for elements in the same period row (as shown in Fig. 2). Similar grouping is also observed for TiO_2 ENM with other TiO_2 species, ENM from the same row (Fe_2O_3, ZnO), CeO_2, which is in the subsequent periodic table group, and AlOOH ENM with which they present similar χ and the closest $\Sigma\chi/n_O$ value in the dataset.

Although the scope of the produced model is to be used for safe by design strategies, i.e. to design new safer and more functional ENM, it can also be used to test existing ENMs as long as there is experimental information on the ENM core size. As a result, it would be of benefit to test how the current model compares with other ζ-potential models, like the "Enalos Nanoinformatics Model for Zeta Potential Prediction" (), which uses the pH of the suspension used for ζ-potential measurement and the ENM's main elongation as extracted through TEM image analysis ()(Varsou et al., 2020a). Using ENMs characterised in MiliQ water ($n = 8$) that were not used in the current model's training or test sets, the results suggest that the current model has better predictive ability ($R^2 = 0.921$) when compared to the "Enalos Nanoinformatics Model for Zeta Potential Prediction" ($R^2 = 0.906$). Furthermore, our model's predictions were all deemed reliable, compared to two unreliable measurements produced with the Enalos model.

The significance of the presented model is further enhanced due to the importance of ζ-potential and thus the applicability of the model to different fields in which ENM are applied, including nanomedicine (Honary and Zahir, 2013a; Honary and Zahir, 2013b), energy production (Sen et al., 2017; Caimi et al., 2018), fuel and combustion (Saxena

et al., 2017; Kumar et al., 2020), construction (Ogunsona et al., 2020), consumer products (Ogunsona et al., 2020), water treatment (Ogunsona et al., 2020) and remediation of heavy metal pollution (Yang et al., 2013). In most cases, ζ-potential plays a key role in keeping the ENMs dispersed inhibiting agglomeration to maximise their functional potential. In the case of construction and consumer products (Ogunsona et al., 2020), high ENM dispersibility assisted by high ζ-potential values can enhance their potential as antimicrobial agents in substrates (e.g. coatings), alternative food packaging, antifouling coatings and more. Furthermore, ζ-potential can assist with maximising the efficiency of redox flow batteries, as well as dispersed non-agglomerating suspensions of electrochemically active ENMs to reduce system viscosity such that the suspensions can undergo electrochemical charge and discharge (Sen et al., 2017) and thus increase the battery's conductivity. Similarly, the inclusion of stable ENM suspensions in diesel and biodiesel fuels for combustion engines has led to substantial enhancement in thermophysical and chemical properties and decrease in environmental pollutants (Saxena et al., 2017; Kumar et al., 2020). In nanomedicine, an extensive review by Honary and Zahir (2013a, 2013b) demonstrates the significance of ζ-potential with respect to blood circulation by inhibiting agglomeration, targeting specific sites, absorption through bodily membranes and controlled drug release. Finally, ζ-potential can substantially help with decreasing environmental pollution through water treatment (Ogunsona et al., 2020) and binding of heavy metal pollutants from soil and groundwater (Yang et al., 2013). In these cases, ENMs with a sufficiently negative charge can attract and bind positively charged metals, like Pb or As, inhibiting their release into the environment.

The ζ-potential predictive model has been made available as a webservice and as such it is now easily accessible and fully exploitable by interested users. It requires the submission of the values of the identified significant parameters (ENM core size, coating, r_{ion}, $\Sigma\chi/n_O$ and χ_{abs}). The predicted ζ-potential values, as well as the domain of applicability warnings, are returned to the user within seconds. It is important to note that the presented ζ-potential predictive model has been developed based on experimental measurements performed with water as the dispersion medium. As such, users must assume that the acquired predictions are at pH 7 and that the theoretical ionic strength is $10^{-6.998}$ mol/L. Furthermore, model usage is restricted to the coatings offered in the webservice's dropdown menu currently. Direct usage for different conditions or coatings would result in unreliable measurements. Another point to take into account when using the tool, is that a concentration-related effect has been reported when measuring the ζ-potential of ENM (Kaszuba et al., 2010; Tantra et al., 2010; Wang et al., 2013). These changes were observed when the ENM concentration dropped below a certain threshold (Tantra et al., 2010; Wang et al., 2013), which Tantra et al. (2010) have reported to be between 10^{-2} and 10^{-4} wt% depending on the tested material. This change was attributed to either ambient CO_2 interacting with the ENM surface functional groups to form $\vdots MOCO_2^-$ (Wang et al., 2013) or to extraneous particulate matter (Tantra et al., 2010) the contribution of which became significant as the ENM concentration decreased. While in most cases, we would not expect this to affect model performance, users need to take care when predicting ζ-potential values and applying the results to low-concentration systems. The webservice is offered as a service through the NanoCommons research infrastructure and is fully compatible with, and will be integrated into, the NanoSolveIT IATA that is currently under development within the NanoSolveIT project.

5. Conclusions

In silico approaches can substantially reduce the effort and time needed for the design, production and evaluation (e.g. preparation of regulatory dossiers) of stable and safe ENM. This study presents a robust, validated and applicable ζ-potential predictive model, developed using the Isalos Analytics Platform and the Enalos+ nodes (Papadiamantis et al., 2020b; Afantitis et al., 2020b). The model was developed via a

library of 69 ENMs, using 7 physico-chemical descriptors, enriched with 13 molecular descriptors, which can be easily identified or calculated using periodic table information. The produced model, which predicts ζ-potential ($R^2 = 0.9$) can be applied in SbD workflows for the development of novel ENM, and is fully independent of experimental input parameters, requiring just descriptive information and molecular descriptors calculated from the periodic table. The parameters identified as significant, i.e., ENM core size, coating, absolute electronegativity (χ_{abs}), metal ionic radius (r_{ion}) and sum of metal electronegativity divided by the number of oxygen atoms present ($\Sigma\chi/n_o$) play a clear role in regulating ζ-potential in water. The benefits of this group of parameters are that they can either be calculated or acquired through the periodic table or existing libraries of calculated descriptors (χ_{abs}, r_{ion}, $\Sigma\chi/n_o$), or can be imported based on ENM physicochemical descriptions typically provided by manufacturers (core size, coating type and charge), and thus is straightforward to use to design specific ENM as only limited and easily acquired input data is required. Furthermore, the used descriptors allow ENM read-across, based on chemical similarity and specific trends observed within the groups (columns) and periods (rows) of the periodic table, followed by ENM core size and the coating used. In any case, no ENM synthesis or physicochemical characterisation is needed to successfully use the model.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.impact.2021.100308.

References

Abdal Dayem, A., et al., 2017. The role of reactive oxygen species (ROS) in the biological activities of metallic nanoparticles. Int. J. Mol. Sci. 18 (1), 120.

Afantitis, A., 2020. Nanoinformatics: artificial intelligence and nanotechnology in the new decade. Comb. Chem. High Throughput Screen. 23 (1), 4–5.

Afantitis, A., et al., 2018. A nanoinformatics decision support tool for the virtual screening of gold nanoparticle cellular association using protein corona fingerprints. Nanotoxicology 12 (10), 1148–1165.

Afantitis, A., et al., 2020a. NanoSolveIT project: driving nanoinformatics research to develop innovative and integrated tools for in silico nanosafety assessment. Comput. Struct. Biotechnol. J. 18, 583–602.

Afantitis, A., Tsoumanis, A., Melagraki, G., 2020b. Enalos suite of tools: enhance cheminformatics and nanoinformatics through KNIME. Curr. Med. Chem. 27, 1–13.

Au-Kaur, I., et al., 2017. Dispersion of nanomaterials in aqueous media: towards protocol optimization. JoVE 130, e56074.

Basei, G., et al., 2019. Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of in silico tools: a critical review. NanoImpact 13, 76–99.

Basketter, D., et al., 2019. Applying non-animal strategies for assessing skin sensitisation report from an EPAA/cefic-LRI/IFRA Europe cross sector workshop, ECHA helsinki, February 7th and 8th 2019. Regul. Toxicol. Pharmacol. 109, 104477.

Batley, G.E., Apte, S.C., 2005. ELEMENTAL SPECIATION | waters, sediments, and soils. In: Worsfold, P., Townshend, A., Poole, C. (Eds.), Encyclopedia of Analytical Science, , Second edition481-488. Oxford, Elsevier.

Bilal, M., et al., 2019. Bayesian network resource for meta-analysis: cellular toxicity of quantum dots. Small 15 (34), 1900510.

Bishop, C.M., 2007. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer New York, NY, USA.

Borm, P., et al., 2006. Research strategies for safety evaluation of nanomaterials, part V: role of dissolution in biological fate and effects of nanoscale particles. Toxicol. Sci. 90 (1), 23–32.

Briffa, S.M., et al., 2019. Physical and chemical transformations of zirconium doped ceria nanoparticles in the presence of phosphate: increasing realism in environmental fate and behaviour experiments. Environ. Pollut. 252, 974–981.

Caimi, S., et al., 2018. Effect of SiO2 nanoparticles on the performance of PVdF-HFP/ ionic liquid separator for lithium-ion batteries. Nanomaterials 8 (11), 926.

Casals, E., et al., 2017. Intrinsic and extrinsic properties affecting innate immune responses to nanoparticles: the case of cerium oxide. Front. Immunol. 8, 970.

Cho, W.-S., et al., 2012. Zeta potential and solubility to toxic ions as mechanisms of lung inflammation caused by metal/metal oxide nanoparticles. Toxicol. Sci. 126 (2), 469–477.

Delgado, A.V., et al., 2005. Measurement and interpretation of electrokinetic phenomena (IUPAC technical report). Pure Appl. Chem. **77** (10), 1753.

Deryabin, D.G., et al., 2015. A zeta potential value determines the aggregate's size of penta-substituted [60]fullerene derivatives in aqueous suspension whereas positive charge is required for toxicity against bacterial cells. J. Nanobiotechnol. 13 (1), 50.

ECHA, 2017. Read-Across Assessment Framework (RAAF). https://echa.europa.eu/docu ments/10162/13628/raaf_en.pdf. Accessed 8 November 2019.

ECHA, 2020. Overview of REACH Information Requirements and Available Methods. Available from: https://euon.echa.europa.eu/reach-test-methods-for-nanomaterials.

Engineered Nanomaterial Mechanisms of Interactions with Living Systems and the Environment: A Universal Framework for Safe Nanotechnology (NanoMILE) [cited 13 November 2020; Available from: https://cordis.europa.eu/project/id/310451.

EPA, 2017. Federal Register Notice: Chemical Substances When Manufactured or Processed as Nanoscale Materials; Notice of Availability and Request for Comment. Available from: https://www.epa.gov/reviewing-new-chemicals-under-toxic-subs tances-control-act-tsca/federal-register-notice-chemical.

Fatehah, M.O., Aziz, H.A., Stoll, S., 2014. Stability of ZnO nanoparticles in solution. Influence of pH, dissolution, aggregation and disaggregation effects. J. Colloid Sci. Biotechnol. 3 (1), 75–84.

Fritsche, E., et al., 2017. OECD/EFSA workshop on developmental neurotoxicity (DNT): the use of non-animal test methods for regulatory purposes. Altex 34 (2), 311–315.

Gernand, J.M., Casman, E.A., 2014. The Toxicity of carbon nanotubes. Risk Anal. 34(3).

Ha, M.K., et al., 2018. Toxicity classification of oxide nanomaterials: effects of data gap filling and PChem score-based screening approaches. Sci. Rep. 8 (1), 3141.

Hall, M., et al., 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11 (1), 10–18.

Hjorth, R., et al., 2017. The role of alternative testing strategies in environmental risk assessment of engineered nanomaterials. Environ. Sci. Nano 4 (2), 292–301.

Honary, S., Zahir, F., 2013a. Effect of zeta potential on the properties of nano-drug delivery systems-a review (part 1). Trop. J. Pharm. Res. 12 (2), 255–264.

Honary, S., Zahir, F., 2013b. Effect of zeta potential on the properties of nano-drug delivery systems-a review (part 2). Trop. J. Pharm. Res. 12 (2), 265–273.

Joossens, E., et al., 2019. A high throughput imaging database of toxicological effects of nanomaterials tested on HepaRG cells. Sci. Data 6 (1), 46.

Kar, S., et al., 2014. Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: a mechanistic QSTR approach. Ecotoxicol. Environ. Saf. 107, 162–169.

Kaszuba, M., et al., 2010. High-concentration zeta potential measurements using light-scattering techniques. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 368 (1927), 4439–4451.

Kirby, B.J., Hasselbrink Jr., E.F., 2004. Zeta potential of microfluidic substrates: 1. Theory, experimental techniques, and effects on separations. Electrophoresis 25 (2), 187–202.

Kraegeloh, A., et al., 2018. Implementation of safe-by-design for nanomaterial development and safe innovation: why we need a comprehensive approach. Nanomaterials 8 (4), 239.

Kumar, S., et al., 2020. Combined effect of oxygenated liquid and metal oxide nanoparticle fuel additives on the combustion characteristics of a biodiesel engine operated with higher blend percentages. Energy 197, 117194.

Labouta, H.I., et al., 2019. Meta-analysis of nanoparticle cytotoxicity via data-mining the literature. ACS Nano 13 (2), 1583–1594.

Langevin, D., et al., 2018. Inter-laboratory comparison of nanoparticle size measurements using dynamic light scattering and differential centrifugal sedimentation. NanoImpact 10, 97–107.

Larose, D.T., Larose, C.D., 2014. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons.

Leach, A.R., Gillet, V.J., 2007. An Introduction to Chemoinformatics. Springer Science & Business Media.

Lead, J.R., et al., 2018. Nanomaterials in the environment: behavior, fate, bioavailability, and effects—an updated review. Environ. Toxicol. Chem. 37 (8), 2029–2063.

Lin, S., et al., 2018. Nanomaterials safer-by-design: an environmental safety perspective. Adv. Mater. 30 (17), 1705691.

Lin, S., et al., 2020. Quantifying electron-transfer in liquid-solid contact electrification and the formation of electric double-layer. Nat. Commun. 11 (1), 399.

Ling, S., et al., 2018. Combining in silico design and biomimetic assembly: a new approach for developing high-performance dynamic responsive bio-nanomaterials. Adv. Mater. 30 (43), 1802306.

Lowry, G.V., et al., 2016. Guidance to improve the scientific value of zeta-potential measurements in nanoEHS. Environ. Sci. Nano 3 (5), 953–965.

NovaMechanics Ltd., N., 20 July 2020. NanoPharos Database. Available from: www. nanopharos.eu.

Lynch, I., Weiss, C., Valsami-Jones, E., 2014. A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. Nano Today 9 (3), 266–270.

Martin, P., et al., 2019. Simulations of morphological transformation in silver nanoparticles as a tool for assessing their reactivity and potential toxicity. NanoImpact 14, 100147.

Melagraki, G., Afantitis, A., 2014. Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles. RSC Adv. 4 (92), 50713–50725.

Mikolajczyk, A., et al., 2015. Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure–property relationship approach. Chem. Mater. 27 (7), 2400–2407.

Mulliken, R.S., 1934. A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. J. Chem. Phys. 2 (11), 782–793.

Mulliken, R.S., 1935. Electronic structures of molecules XI. Electroaffinity, molecular orbitals and dipole moments. J. Chem. Phys. 3 (9), 573–585.

OECD, 2016. Physical Chemical Parameters: Measurements and methods relevant for the regulation of nanomaterials (ENV/JM/MONO(2016)2). OECD Workshop Report: Series on the Safety of Manufactured Nanomaterials No. 63.

OECD, 2004. Validation of (Q)SAR Models. https://www.oecd.org/chemicalsafety/ris k-assessment/validationofqsarmodels.htm. Accessed 8 November 2019.

Ogunsona, E.O., et al., 2020. Engineered nanomaterials for antimicrobial applications: a review. Appl. Mater. Today 18, 100473.

Oh, E., et al., 2016. Meta-analysis of cellular toxicity for cadmium-containing quantum dots. Nat. Nanotechnol. 11 (5), 479–486.

Ohshima, Hiroyuki, 2006. Chapter 2. Potential distribution around a soft particle. In: Theory of Colloid and Interfacial Electric Phenomena, Volume 12, 1st Edition. https ://www.elsevier.com/books/theory-of-colloid-and-interfacial-electric-phenomena /ohshima/978-0-12-370642-3.

Ojha, P.K., Roy, K., 2011. Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. Chemom. Intell. Lab. Syst. 109 (2), 146–161.

Papadiamantis, A.G., et al., 2020a. Metadata stewardship in nanosafety research: community-driven organisation of metadata schemas to support FAIR nanoscience data. Nanomaterials 10 (10), 2033.

Papadiamantis, A.G., et al., 2020b. Predicting cytotoxicity of metal oxide nanoparticles using Isalos Analytics Platform. Nanomaterials 10 (10), 2017.

Parr, R.G., Bartolotti, L.J., 1982. On the geometric mean principle for electronegativity equalization. J. Am. Chem. Soc. 104 (14), 3801–3803.

Portier, J., et al., 2001a. Degenerate semiconductors in the light of electronegativity and chemical hardness. Int. J. Inorg. Mater. 3 (7), 1039–1043.

Portier, J., et al., 2001b. Relationships between optical band gap and thermodynamic properties of binary oxides. Int. J. Inorg. Mater. 3 (7), 1091–1094.

Puzyn, T., et al., 2018. Perspectives from the NanoSafety modelling cluster on the validation criteria for (Q)SAR models used in nanotechnology. Food Chem. Toxicol. 112, 478–494.

Rasmussen, K., et al., 2019. Developing OECD test guidelines for regulatory testing of nanomaterials to ensure mutual acceptance of test data. Regul. Toxicol. Pharmacol. 104, 74–83.

Römer, I., et al., 2019. Impact of particle size, oxidation state and capping agent of different cerium dioxide nanoparticles on the phosphate-induced transformations at different pH and concentration. PLoS One 14 (6), e0217483.

Russel, W.B., Saville, D.A., Schowalter, W.R., 1989. Colloidal Dispersions. Cambridge University Press.

Saarimäki, L.A., et al., 2020. Toxicogenomics analysis of dynamic dose-response in macrophages highlights molecular alterations relevant for multi-walled carbon nanotube-induced lung fibrosis. NanoImpact 100274.

Sanderson, R., 1951. An interpretation of bond lengths and a classification of bonds. Science 114 (2973), 670–672.

Saxena, V., Kumar, N., Saxena, V.K., 2017. A comprehensive review on combustion and stability aspects of metal nanoparticles and its additive effect on diesel and biodiesel fuelled C.I. engine. Renew. Sust. Energ. Rev. 70, 563–588.

Schmickler, W., 2014. Electrochemical Theory: Double Layer, in Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Elsevier.

Schwegmann, H., Feitz, A.J., Frimmel, F.H., 2010. Influence of the zeta potential on the sorption and toxicity of iron oxide nanoparticles on *S. cerevisiae* and *E. coli*. J. Colloid Interf. Sci. 347 (1), 43–48.

Sen, S., et al., 2017. Electroactive nanofluids with high solid loading and low viscosity for rechargeable redox flow batteries. J. Appl. Electrochem. 47 (5), 593–605.

Shin, H.K., et al., 2018. Meta-analysis of Daphnia magna nanotoxicity experiments in accordance with test guidelines. Environ. Sci. Nano 5 (3), 765–775.

Tantra, R., Schulze, P., Quincey, P., 2010. Effect of nanoparticle concentration on zeta-potential measurement results and reproducibility. Particuology 8 (3), 279–285.

Tauc, J., Grigorovici, R., Vancu, A., 1966. Optical properties and electronic structure of amorphous germanium. Phys. Status Solidi B **15** (2), 627–637.

Toropov, A.A., Achary, P.G.R., Toropova, A.P., 2016. Quasi-SMILES and nano-QFPR: the predictive model for zeta potentials of metal oxide nanoparticles. Chem. Phys. Lett. 660, 107–110.

Toropov, A.A., et al., 2018. Towards the development of global nano-quantitative structure–property relationship models: zeta potentials of metal oxide nanoparticles. Nanomaterials 8 (4), 243.

Toropova, A.P., et al., 2016. Quasi-SMILES as a tool to utilize eclectic data for predicting the behavior of nanomaterials. NanoImpact 1, 60–64.

Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Mol. Inform. 29 (6–7), 476–488.

Tropsha, A., Gramatica, P., Gombar, V.K., 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR & Comb. Sci. 22 (1), 69–77.

Valsami-Jones, E., Lynch, I., 2015. How safe are nanomaterials? Science 350 (6259), 388–389.

Varsou, D.-D., et al., 2018. Enalos+ KNIME nodes: new cheminformatics tools for drug discovery. In: Mavromoustakos, T., Kellici, T.F. (Eds.), Rational Drug Design: Methods and Protocols. Springer New York, New York, NY, pp. 113–138.

Varsou, D.-D., et al., 2019. A safe-by-design tool for functionalised nanomaterials through the Enalos nanoinformatics cloud platform. Nanoscale Adv. 1 (2), 706–718.

Varsou, D.-D., et al., 2020a. Zeta-potential read-across model utilizing nanodescriptors extracted via the nanoxtract image analysis tool available on the enalos nanoinformatics cloud platform. Small 16 (21), 1906588. **n/a**(n/a): p. 1906588.

Varsou, D.-D., et al., 2020b. Enalos cloud platform: nanoinformatics and cheminformatics tools. In: Roy, K. (Ed.), Ecotoxicological QSARs. Springer US, New York, NY, pp. 789–800.

Wang, Z.L., Wang, A.C., 2019. On the origin of contact-electrification. Mater. Today 30, 34–51.

Wang, N., et al., 2013. Influence of metal oxide nanoparticles concentration on their zeta potential. J. Colloid Interf. Sci. 407, 22–28.

Witten, I.H., et al., 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Wyrzykowska, E., et al., 2016. Development of a novelin silicomodel of zeta potential for metal oxide nanoparticles: a nano-QSPR approach. Nanotechnology 27 (44), 445702.

Xu, Y., Goodacre, R., 2018. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. J. Anal. Test. 2 (3), 249–262.

Yan, L., et al., 2019. A safe-by-design strategy towards safer nanomaterials in nanomedicines. Adv. Mater. 31 (45), 1805391.

Yang, X., et al., 2013. Silica nanoparticles capture atmospheric lead: implications in the treatment of environmental heavy metal pollution. Chemosphere 90 (2), 653–656.

Zhang, S., et al., 2006. A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. J. Chem. Inf. Model. 46 (5), 1984–1995.

Zhang, H., et al., 2012. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. ACS Nano 6 (5), 4349–4368.

NanoCommons (2018). The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators (NanoCommons). (Accessed 23 March 2021) www.nanocommons.eu.