Environmental Science Nano

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Check for updates

Cite this: *Environ. Sci.: Nano*, 2023, 10, 325

Development of a nano-QSAR model for predicting the toxicity of nano-metal oxide mixtures to *Aliivibrio fischeri*†

Minju Na,[ac] Sang Hwan Nam,[b] Kyonghwan Moon[c] and Jongwoon Kim *[a]

Received 15th July 2022,
Accepted 30th November 2022

DOI: 10.1039/d2en00672c

rsc.li/es-nano

Metal oxide nanoparticles (MONPs) have various applications, including cosmetics, detergents, and antibacterial agents, owing to their unique physicochemical properties. MONPs are often mixed and used in products and exist together through various exposure routes in the environment. Toxicity can occur between chemicals at no observed effect concentrations (NOECs) by the cocktail effect (*e.g.*, addition, synergism, potentiation), but a definitive toxic assessment for the nanoparticles is still lacking. There have been several studies on nano quantitative structure–activity relationships (nano-QSAR), but the calculations of the descriptors (<1000 atoms) for the engineered size of the nanoparticles (NPs) based on density functional theory (DFT) are unclear. In this study, we conducted both mixture toxicity assays and molecular dynamics (MD)-based molecular descriptor calculations to develop a nano-mixture QSAR model. A toxicity assay was performed for a mixture of SiO_2, TiO_2, and ZnO NPs of various sizes (8–140 nm), targeting a marine bioluminescent bacterium called *Aliivibrio fischeri*, a decomposer in aquatic ecosystems. Theoretical molecular descriptors were calculated based on molecular dynamics (MDs) to reflect the characteristics of NPs of different sizes. Two different types of descriptors (total descriptors and the calculated descriptors) were used to develop the models. In this study, four machine-learning algorithms (random forest (RF), support vector machine (SVM), Bayesian regularized neural network (BRNN), and multilinear regression (MLR)) were applied to develop a nano-mixture QSAR model. The proposed model based on MD shows potential for use in the selection of a safer MONPs combination design.

Environmental significance

Metal oxide nanoparticles (MONPs) widely used in different industrial applications are frequently mixed in products, and encounter each other through various environmental exposure routes. The mixture toxicity of MONPs presenting at no observed effect concentrations can be caused by the cocktail effect. As one of the first studies on developing a nano-mixture quantitative structure–activity relationship model for *Aliivibrio fischeri* using molecular dynamics-based descriptors for MONP mixtures (SiO_2, TiO_2, and ZnO groups) in a wide range of sizes (8–140 nm), this study also provides experimental data for the first time on the mixture toxicity of the MONPs to *A. fischeri*. A random forest based nano-QSAR model was successfully developed to predict the EC_{10} and EC_{50} of nano-MONP without the dose–response curves of the components.

1. Introduction

Nanoparticles (NPs) have unique physical and chemical properties, including small and quantum size effects.[1,2] Owing to their unique properties, NPs have a wide range of applications in electronics, consumer products (*e.g.*, paints and fabrics), personal care products (*e.g.*, sunblock), and biomedicine. Although various nanomaterials are typically used, metal oxide nanoparticles (MONPs) constitute approximately 80% of the market volume (Abercade Research Company, 2009). As they are released into the external environment after use, MONPs encounter various chemicals, and even at environmentally safe concentrations, combinations of MONPs can be toxic to aquatic organisms. Upon release, MONPs can interact with the environment, leading to a toxic mixture that is exposed to the organisms and causes mixture toxicity.[3] When chemicals interact, synergistic or antagonistic effects are considered. A synergistic effect occurs when the mixture toxicity is greater

[a] *Chemical Safety Research Center, Korea Research Institute of Chemical Technology (KRICT), Daejeon, Republic of Korea. E-mail: jkim@krict.re.kr; Tel: +82 (0)42 860 7482*
[b] *Laboratory for Advanced Molecular Probing, Korea Research Institute of Chemical Technology (KRICT), Daejeon, Republic of Korea*
[c] *Department of Public Health, College of Health Science, Korea University, Seoul, Republic of Korea*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2en00672c

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

than the sum of the toxicity of individual components, whereas the antagonistic effect has the opposite result.[4–6] However, effective experimental studies may be unfeasible because a series of concentrations of all components of various combinations of nano-mixtures must be established to test the toxic effects of the different mixtures. Additionally, the complete experimental characterization of toxicity for all varying preparations is laborious, but predictions of the relationship between the structure of NPs and their biological activity are in high demand. The quantitative structure–activity relationship (QSAR) approach is one of the most time-efficient computational methods, a molecular modeling approach that evaluates the relationship between structure and activity using machine-learning methods and mathematical statistics. The properties of various NPs have been modeled and predicted over the last decade using QSAR approaches.[4]

The term "nano-QSAR" refers specifically to the search for quantitative links between NP features and target activity. Structural features are typically represented by various numeric parameters (also known as "descriptors").[7,8] Tomasz Puzyn *et al.* (2011)[9] developed models describing the cytotoxicity of 16 metal oxide NPs to *Escherichia coli*. Quantum chemical calculations were performed using the PM6 semi-empirical method. Cytotoxicity predictions based on the same dataset of 17 metal oxide NPs were further investigated.[10] Density functional theory (DFT)-based descriptors (hardness, softness, energy gap, and electrophilicity index) indicated a high correlation between the experimental and predicted values. Current nano-QSAR models concerning the toxicity of nano-mixtures are mostly focused on predicting the cytotoxicity of TiO_2-based nano-mixtures on Chinese hamster ovary cell lines[4,7,8] and human kidney two cell lines.[11] Mikolajczyk *et al.* (2018)[8] developed a cytotoxicity prediction model for 34 TiO_2 samples modified with different types and amounts of noble metals (mixtures of Ag, Au, and Pt), and quantum-mechanical (QM) descriptor calculations were performed at the DFT level with metal clusters (5 × 5 × 5 Å). Several nano-QSAR models, based on metallic and metal oxide nanoparticles, have been developed for predicting cytotoxicity.[4,7,8,11–13] Furthermore, Trinh *et al.* (2022)[14] developed QSAR models to predict the mixture toxicity of TiO_2-based nano-mixtures to *Daphnia magna*, and the descriptors were calculated using the semi-empirical parametric method seven (PM7) based on the TiO_2 anatase nanocluster (0.75 × 0.75 × 1.35 nm^3). Although the descriptors mentioned above are widely used in nano-QSAR studies, there is a limitation regarding their applicable size. DFTs generally offer higher material fidelity, but it is limited to small systems, which are usually less than 1000 atoms. Therefore, it is necessary to develop QSAR models using a different descriptor-calculation approach. The force field (also called an interatomic potential)-based method is the only remaining possible option, and large-scale atomic/molecular massively parallel simulator (LAMMPS)[15] software (a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales) was chosen for calculating descriptors in this study. Molecular dynamics (MD) is an essential modeling technique, where Newton's second law of motion for the atoms/molecules in a system is applied to obtain their physical properties, and describes the thermodynamic behaviors of the system.

In an aquatic environment, several mixtures are present, including various sizes of MONPs. Therefore, models that can accommodate size-dependent MONP mixtures are necessary for the risk assessment of nano-mixtures. Since the cytotoxicity of TiO_2-based nano-mixtures[4,7,8,11] and toxicity of *D. magna*[14] could be predicted using the structures of the mixtures, we hypothesized that the mixture toxicity of MONPs to *A. fischeri* might also be related to the structures of the mixtures. This marine bacterium has widely been used as a test organism to investigate the toxicity of pollutants, including heavy metals[16] and nZnO.[17,18] Since bacteria act as decomposers in an ecosystem, the effect of MONPs on bacteria can disrupt the entire ecosystem.

In this study, we aimed to minimize the experimental cost of the environmental risk assessment of nano-mixtures and also overcome the size limitations. We developed nano-mixture QSAR models using MD descriptors to predict the toxicity of MONPs to *A. fischeri*. The nano-mixture QSAR models were developed using mixture descriptors, which were calculated by combining the molecular descriptors of the components of the mixtures in a component-based approach. The QSAR models developed in this study can be used in web-based applications (**https://mjna-nano.shinyapps.io/monps**) to assess the aquatic toxicity of the MONP mixtures, where users can input individual MONP information (*i.e.*, type, size, and mole fraction) and predict the concentration of the mixture toxicity value (EC_{50} and EC_{10}).

This study presents two novel aspects: 1) single and mixture toxicity assessment for *A. fischeri* and toxicity data production. 2) Nano-mixture QSAR model development based on MD descriptors to characterize the MONPs properties of various sizes and reflect the exact engineered size. Four machine learning algorithms (random forest [RF], support vector machine [SVM], multiple linear regression [MLR], and Bayesian-regularized neural network [BRNN]) were applied to develop nano-mixture QSAR models based on the calculated descriptors using the MD tool, including properties of the various sizes of MONPs.

2. Methods

2.1 Target MONPs

A total of 12 targets in four different types of MONPs of various sizes (8–140 nm) were purchased in nano powder form from US Research Nanomaterials (Houston, TX, USA). We attempted to expand the applicability domain of MONPs of various size ranges. Accordingly, a range of commercially available MONP sizes were investigated. Of these, SiO_2, TiO_2, and ZnO were selected as the target materials in this study.

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 1 Target metal oxide nanoparticles (MONPs)

No.	Particle[a]	Engineered nanoparticles (NPs) diameter [nm]	Molecular weight [g mol^{-1}]	Purity [%]
1	SiO_2	8	60.08	99.0
2	SiO_2 (porous)	15–20	60.08	99.5
3	SiO_2 (spherical)	15–20	60.08	99.5
4	SiO_2	20–30	60.08	99.0
5	SiO_2	60–70	60.08	98.0
6	TiO_2 (rutile)	30	79.87	99.9
7	TiO_2 (anatase)	15	79.87	99.5
8	TiO_2 (anatase)	18	79.87	99.9
9	TiO_2 (anatase)	30	79.87	99.9
10	ZnO	10–30	81.38	99.0
11	ZnO	35–45	81.38	99.0
12	ZnO	80–200	81.38	99.9

[a] Denomination of particles.

Additionally, these fall in the top five in terms of general usage. Further details on the target MONPs are presented in Table 1.

2.2 Characterization of MONPs

Different MONP suspensions were prepared in a 2% NaCl solution, and the solutions were subjected to 30 min sonication with an ultrasonic probe (15 °C, 60 W) to decrease the agglomeration. To determine the surface charge and the actual size of the MONPs in suspension, it was necessary to measure the zeta potential and hydrodynamic size. Both single and mixed samples of MONPs were characterized by measuring the hydrodynamic size (*Z*-average) and zeta potential using a Zeta Nano-ZS with a 532 nm laser (Malvern Instruments, UK). The surface charge of the MONPs was characterized *via* their zeta potential. The measurement sample was a 2% NaCl solution (salinity of seawater) for marine bacteria, and the particle size and surface charge were measured. All measurements were conducted in triplicate at 25 °C, and the average values were established.

2.3 Toxicity test

The toxicity of MONPs was experimentally determined using the bioluminescent bacterium, *A. fischeri*. The sample effective concentration (EC) was determined using the International Organization of Standardization (ISO) 21338,[19] which stipulates the kinetic inhibitory effects of the sediment, other solids, and color samples on the light emission of *A. fischeri*. Our target samples, MONPs, have low solubility in water. Suspensions of the three different MONPs prepared with the 2% NaCl solution were serially diluted at a 1 : 1 (100 μL) ratio in a white 96-well plate. *A. fischeri* bioluminescence was recorded using a CentroXS LB 960 high-sensitivity microplate luminometer (Berthold Technologies, France). Triplicates were considered for each concentration and control, and the exposure time of the tested MONPs to *A. fischeri* was set at 30 min. For quality assurance of the bacteria, 100 ppm zinc sulfate solution (Sigma-Aldrich) was measured for each case.

2.4 Mixture toxicity test

Mixture toxicity was tested for all the binary mixtures prepared from the 12 target MONPs. The binary mixtures were prepared, including two equitoxic mixtures at a 50% effective concentration for each MONP as a high effective concentration ratio mixture (EC_{50} ratio mixture) and at 10% effective concentration ratio mixture (EC_{10} ratio mixture) as a low effective concentration ratio mixture. Additionally, a combination of EC_{50} and EC_{10} ratio mixtures was used to establish a medium effective concentration. Binary mixture toxicity tests were conducted using the same method as for the analysis of individual toxicity tests. A total of 93 binary mixtures were tested in equitoxic doses at high and low effective concentrations (*e.g.*, EC_{50} and EC_{10}, respectively). However, the total doses of the mixtures systematically differed. Further details on the mixture design are presented in Table S3.†

2.5 Statistical analysis of the mixture toxicity

Sigmoidal regression equation parameters were estimated using Sigma Plot® (Ver. 14.0, Systat Software, Chicago, IL, USA). A best-fit approach was used to select the model with the smallest sum of absolute residuals and the highest coefficient of determination (R^2) among the different sigmoidal functions. Using three-parameter sigmoidal equations, the best-fit models were finally determined and applied to describe the experimental data of the single and mixture MONPs tested in this study. The EC*x* (*e.g.*, EC_{10} and EC_{50}) values of the test chemicals were derived from the regression models. One of the widely used classical models for mixture toxicity prediction is the concentration addition (CA) model. CA is expressed mathematically as shown in eqn (1):

$$EC x_{mix}\left(\sum_{i=1}^{n}\frac{p_i}{EC x_i}\right)^{-1} \quad (1)$$

where $EC x_{mix}$ is the predicted total concentration of the mixture, which provokes the *x*% effect, and $EC x_i$ and p_i are the

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

individual effective concentration and the fraction of the component within the mixture, respectively.

The model deviation ratio (MDR) values were used to quantify the interaction between the mixture components, and determine the mixture toxicity type of the interactions. MDR is defined as:

$$\mathrm{MDR} = \frac{\text{Predicted EC}x\text{ of mixture}}{\text{Observed EC}x\text{ of mixture}} \quad (2)$$

where the predicted ECx indicates the effective concentration of a mixture based on the predictive model (*i.e.*, the CA model), and the observed ECx is the effective mixture concentration obtained from the experimental toxicity test. This study used the CA model, which is recommended as a default approximation for mixtures, to predict the mixture toxicity. Based on the MDR value, the combined effects were divided into three groups in this study: synergistic (MDR $>$ 2), additive ($0.5 \leq$ MDR ≤ 2), and antagonistic (MDR < 0.5).

2.6 Nano-QSAR model development

2.6.1 Mixture descriptor calculations

Experimental descriptors. In the present study, the five experimental descriptors, excluding molecular dynamic (MD) calculations, were defined as follows: 1) core diameter, 2) molecular weight, 3) specific surface area, 4) concentration, and 5) hydrodynamic size. The information on the core diameter, molecular weight, and specific surface area given by the material safety data sheet (MSDS) from US Research Nanomaterials (Huston, HX USA) was used. Hydrodynamic size information was obtained by measuring the *Z*-average size using DLS.

Descriptor calculations using MD. Molecular descriptors calculation is the most critical step in the development of nano-QSAR models. MD-based optimal descriptors were employed to describe the engineered MONPs structure information by characterizing each engineered size of the MONP structure and calculating the properties of the MONPs. The unit cells of the MONPs were replicated in all three dimensions using the Open Visualization Tool (OVITO) to create MONPs of engineered size. The resulting spherical MONPs were generated by removing all atoms outside the set radius of the produced NPs, and maintaining the electroneutrality of the final MONPs. The potential energies of the atoms were calculated based on the third-generation charge-optimized many-body (COMB3).[20] These calculations were performed under periodic boundary conditions in all three Cartesian directions using the LAMMPS software. The length of the simulation box in each direction was much larger than the diameter of the MONP, such that all atoms and interatomic interactions of the MONPs were contained within the box. During the LAMMPS simulation, structural optimization was performed through the energy minimization process, and the energy and characteristics of the structure were calculated. The interactions and potential energies between the atoms are defined by a molecular mechanical force field.

Mixture descriptor calculations. The mixture descriptor (D_{mix}) is a hypothetical descriptor that measures the contribution of a component to the overall activity of a mixture.[6] Based on theoretical considerations, the descriptors of the components were obtained by combining them numerically. Previous studies[4,7,8,11] used the arithmetic mean (eqn (3)) to calculate the mixture descriptors of the nano-mixtures:

$$D_{\mathrm{mix}} = x_1 D_1 + x_2 D_2 \quad (3)$$

where x_1 and x_2 are the mole fractions, and D_1 and D_2 are the descriptors of the individual components in the mixture. For user convenience of the developed model, the mole fraction and core diameter of every single substance constituting the mixture were also included as descriptors and used without using D_{mix}. A total of four descriptors (NP1 mole fraction, NP2 mole fraction, NP1 core diameter, and NP2 core diameter) were used as a single descriptor to identify the effect of individual NPs on the mixture toxicity.

Descriptor selection. Two types of models were developed (descriptor-based and calculation descriptor-based) with a component-based approach to predict the $\log \mathrm{EC}_{50}$ and $\log \mathrm{EC}_{10}$ of the MONPs mixtures. In addition, all single descriptors (independent variables) correlated with the luminescence inhibition (dependent variable). Descriptors, which indicate theoretical properties, were derived from force field calculations (MD calculations), corresponding to the arithmetic means of the potential energies for specific atom types and locations in the MONP.

The main advantages of these novel nano descriptors, compared to previously published descriptors for MONPs, can be summarized as follows:

1. These descriptors are theoretically based solely on the unit cell structure and COMB3 potential parameters, which are directly available for many compounds or derived from quantum chemical calculations.

2. The descriptors indicate inherent particle size-dependent properties.

3. Overall, the inter-correlation among the descriptors is low.

These descriptors were used selectively to build a model. In model development, having too many descriptors increases the model complexity, as well as the risk of overfitting. Therefore, dispensable descriptors were removed, and only vital descriptors were used to build the model. Generally, a correlation coefficient of $(r) > \lfloor 0.9 \rfloor$ is interpreted as having a strong association. Therefore, a correlation analysis was performed to select one of these descriptors when the correlation coefficient exceeded 0.9 between several descriptors. The descriptor selection, correlation analysis, and model development were performed using R software. Based on the correlation analysis results, mixture descriptors were calculated using D_{mix} for the remaining single

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

descriptor. Subsequently, the recursive feature elimination (RFE) algorithm was applied to define only the essential and relevant features of the dataset. RFE is one of the most preferred feature selection algorithms for analyzing datasets and achieving the best model performance.[21,22] The RFE algorithm is a method wherein characteristics with higher feature importance corresponding to the desired number of features become the final feature selection result. With a range of descriptors from 1 to 20, the RFE algorithm was applied to determine the most appropriate number and type of descriptors.

2.6.2 Data splitting for model development. For the training and evaluation of the model, the MONPs experimental data were split into two sets. The training set (to develop a nano-QSAR model) and test set (to measure the generalization ability of the model) guarantee a balanced distribution of MONPs covering the entire range of both training and test sets (Fig. 1). If data points from the training set were used in the test set, the evaluation would always result in high performance as the model already knows the data point. As a result, the two sets had to be strictly separated. Therefore, the external validation was assessed by randomly splitting the data set into a training set and a test set at 80% and 20% of the data set, respectively. Details of the separation results are provided in the supplementary data. For $\log EC_{50}$, 44 data points were split into groups of 37 and 7 for the training and test sets, respectively. Likewise, for $\log EC_{10}$, 85 data points were divided into 68 and 17 in the training and test sets, respectively.

2.6.3 Applied machine learning algorithm. Four machine-learning algorithms (RF, SVM, MLR, and BRNN) were applied to develop a nano-mixture QSAR model with associated R packages ('RandomForest', 'e1071', 'statas', and 'brnn',

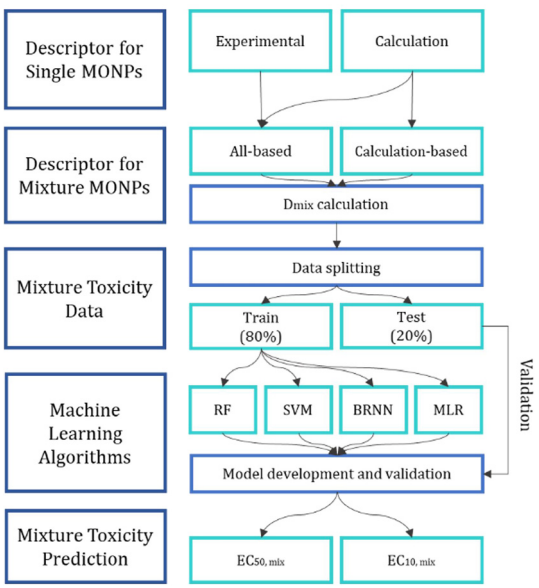

Fig. 1 A scheme for developing the nano-quantitative structure–activity relationships (QSAR) model.

respectively). RF is one of the most powerful algorithms for nonlinear models, and is a classifier ensemble composed of numerous decision trees. The hyper-parameter 'mtry' was used in the RF, which defines the number of variables randomly sampled as candidates at each split, set in the range of 2 to 7. The eps-regression type and polynomial kernel were used in the SVM algorithm. As hyperparameters, degree = 2, scale = 0.1, and C = 0.25, were selected as the optimal values. The neuron range was set from 1 to 3 for the BRNN algorithm. The root mean square deviation (RMSD) was used to select the optimal model using the smallest value. We used R (R Core Team, 2021)[23] and RStudio (RStudio Team, 2021)[24] to develop the predictive models.

2.6.4 Model validation. Model validation is critical for guaranteeing that a fitted model can effectively predict responses for future or unknown subjects. In addition, any QSAR model must comply with the Organization for Economic Cooperation and Development (OECD) recommendations. The goodness-of-fit was evaluated by calculating the squared correlation coefficient (R^2), and adjusted correlation coefficients between the observed and predicted values (AdjR^2).[25] Furthermore, the robustness and internal predictivity of the models were verified using a 10-fold cross-validation algorithm. The Y-randomization method[26] based on the Z-score was used to ensure the reliability and robustness of the developed model. $R_{train}{}^2$ values of models built from the shuffled data were used to calculate the Z-score for validation of the final models:

$$Z_{score} = \frac{R_{train(val)}{}^2 - \overline{R_{train(val)}{}^2}}{\sigma_{rand}} \quad (4)$$

where $R_{train(val)}{}^2$ is the $R_{train}{}^2$ value of the validated model, and σ_{rand} are average and standard deviation of $R_{train}{}^2$ obtained from ten randomly shuffled datasets. A Z_{score} greater than three indicates that the validated model is unique, statistically significant and robust when compared with models developed with random data. External validation is essential in determining both the generalizability and external predictivity of the nano-mixture QSAR models for the MONPs not involved in model training. In the present study, the external validation was assessed by randomly split test sets, 20% of the data set, ten times, as described in section 2.6.2.

2.6.5 Evaluation of the predictive models. Statistical quality of the activity prediction was defined by using the number of variables that enter a QSAR model using the FIT Kubinyi function (eqn (5)).[27] A criterion closely related to the F-value was proven to be valid:

$$FIT = R^2(n - k - 1)/(n + k^2)(1 - R^2) \quad (5)$$

where n is the number of data points, and k is the number of variables in the QSAR equation. The FIT criterion has a low sensitivity toward changes in k values if they are small, and there is a substantial increase in sensitivity for large k values.

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 2 Particle characteristics of single MONPs

No.	Particle[a]	MONPs diameter [nm]	Specific surface area [m^2 g^{-1}]	Hydrodynamic size in 2% NaCl[b] [nm]	Zeta-potential [mV]
1	SiO_2	8	185	614.1	−5.79
2	SiO_2 (porous)	15–20	640	784.8	−4.65
3	SiO_2 (spherical)	15–20	170–200	509.7	−5.52
4	SiO_2	20–30	180–600	358.9	−3.56
5	SiO_2	60–70	160–600	1079	−14.80
6	TiO_2 (rutile)	30	35–60	1875	−4.97
7	TiO_2 (anatase)	15	60	1571	0.33
8	TiO_2 (anatase)	18	200–240	1509	−0.17
9	TiO_2 (anatase)	30	50	2274	1.73
10	ZnO	10–30	35	1434	−6.45
11	ZnO	35–45	85	1954	−15.50
12	ZnO	80–200	20–60	1171	−10.12

[a] Denomination of particles. [b] All DLS measurements were performed in triplicate.

The best model is one that possesses a high value for this function. Therefore, this was used as a criterion to confirm the robustness of the QSAR models developed in this study.

3. Results and discussion

3.1 Characteristics of MONPs

Previous studies have observed the aggregation of MONPs. Our study also revealed the same trend, as displayed in Table 2 (for single) and Table S3† (for the mixture). The hydrodynamic size of the MONPs in the solution increased considerably compared with the diameter of the engineered MONPs. The behavior of NPs entering the marine environment could change owing to their intrinsic chemical signature, and the physical and chemical properties of the receiving water bodies (*i.e.*, temperature and pH).[28] As such, the hydrodynamic diameters, which describe sizes more precisely in an aqueous environment, were much larger than their advertised sizes, reflecting considerable aggregation of the NPs.

Additionally, the MONPs aggregation tendency was ascribed to their relatively low zeta potential (<|30 mV|). The ultrasonic probe disperses MONPs with the help of acoustic ultrasonic forces, which generate sound waves to help break down the agglomeration. Adams *et al.* (2006)[29] reported that the aggregation of particles in water led to their actual size in suspension differing widely from that of the dry powders. Comparing the anatase and rutile structures of TiO_2 30 nm, anatase has a lower absolute zeta-potential and larger hydrodynamic size, indicating a higher aggregation tendency for anatase than rutile. Even though they have the same atomic compositions and sizes, the differences in structure and shape induce different physicochemical properties. In addition, the physicochemical properties, fate, and toxicity can vary depending on the structural properties of the same material.

The hydrodynamic size ranges were TiO_2 (anatase) (1509–2274 nm) > ZnO (1171–1954 nm) > TiO_2 (rutile) (1875 nm) > SiO_2 (509.7–1079 nm). As the zeta potential of ZnO is higher than that of other MONPs, it has a relatively strong negative charge on the surface. NPs with a positive surface charge tend to exert higher toxicity on bacteria.[30] However, ZnO is negatively charged and exhibited the highest toxicity in several studies, indicating that the surface charge is not a critical factor for NP toxicity.[31] In the present study, 10 MONPs, except for two (TiO_2 anatase, 15 nm, and 30 nm), had a positive surface charge.

3.2 *A. fischeri* toxicity test

3.2.1 Single MONPs dose–response curves (DRCs). The toxicity of the 12 MONPs was evaluated. As shown in Fig. 2, significant dose-dependent inhibition of bioluminescence was observed with an increase in NP concentration, suggesting that the MONP toxicity is dose-dependent. Among the NPs, ZnO exhibited the highest toxicity (0.012–0.034 mM), whereas SiO_2 (7.464–63.131 mM) and TiO_2 (9.841–58.042 mM) were relatively less toxic. According to previous studies, ZnO shows relatively higher toxicity than other NPs.[29,32] All 12 single MONPs had EC_{10} values, with a total of six MONPs obtained up to EC50 values. For EC_{10} and ZnO in all three size ranges, SiO_2 showed 54.183 mM at 60–70 nm, TiO_2 (rutile) showed 318.642 mM at 30 nm, and TiO_2 (anatase) showed 23.573 mM at 18 nm. Table S2†

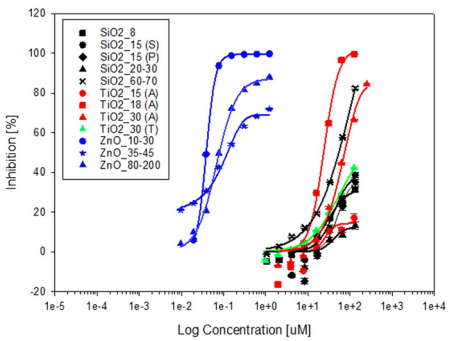

Fig. 2 Dose–response curves (DRCs) for *A. fischeri* growth inhibition of single metal oxide nanoparticles (MONPs). ZnO (blue lines), TiO_2 anatase (red lines), TiO_2 rutile (green lines), and SiO_2 (black lines).

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

summarizes the parameter values of all best-fitting regression models for the DRCs of single MONPs.

3.2.2 Binary nano-mixtures DRCs. Binary mixture DRCs (Table S4†) were experimentally evaluated with two different equitoxic levels (EC_{50} and EC_{10} ratio mixtures) and a EC_{10} + EC_{50} mixture based on a single toxicity result. Mixture toxicity analysis was performed for 93 binary mixtures, but only 85 mixtures exhibited levels above EC_{10}. As shown in Table S4,† the best-fitting curves for the 85 mixtures had high regression coefficients (R^2), ranging from 0.749 to 0.999. More than 10% effect concentrations could not be obtained under the testing conditions, which could be due to the limited water solubility or nontoxicity. As a result, 44 mixtures reached up to EC_{50}, and 39 mixtures reached up to EC_{10}. All substances showing EC_{50} values in the mixture and six MONPs showing EC_{50} values at a single level were used as constituents of the mixture.

Table S4† illustrates DRC information for the observed and predicted bioluminescence inhibition of *A. fischeri* by the CA model for the binary mixtures based on ratios at 50% and 10% effective inhibition concentrations. We calculated MDR values to quantify the toxicity interactions between MONPs. Based on the MDR value, we divided the combined effects into three types:

For EC_{50}, 41% (n = 18) showed synergism, 52% (n = 23) showed additive, and 7% (n = 3) showed antagonistic interactions among the mixtures. For EC_{10}, 61% (n = 52) showed synergism, 31% (n = 26) showed additive, and 7% (n = 7) showed antagonistic interactions. In particular, at the EC_{10} level, the synergism was found to be as much as 61%, and there was a mixture that increased up to 65.486 times from the concentration addition predictions. These results show that synergism due to mixture toxicity is more likely to appear at the EC_{10} level, which has a low effect concentration value. This indicates the need for the regulation of nanomaterials and the consideration of mixed toxicity, not just the level of a single substance. For more details about the MDR accumulative frequency and the pie chart of the toxicity interaction, please refer to Fig. 3.

3.3 Nano-mixture QSAR model development

3.3.1 Experimental and calculated descriptors. A total of 26 single descriptors were derived from the size (five descriptors), experimental conditions (seven descriptors), and nano-based characteristics (14 descriptors) used in the model development stage.

The proposed methodology for computing the additive descriptors for mixture MONPs were applied, and we calculated a set of mixture descriptors reflecting the properties of the $\log EC_{10}$ (n = 85) and $\log EC_{50}$ (n = 44) values of the MONP mixtures. The RFE algorithm was applied for the descriptor selection process. As a result, nine descriptors were selected for all descriptor-based models (including experimental properties), and seven descriptors were selected for the calculation-based model (Fig. 4). By combining the experimental and theoretical studies (*i.e.*, toxicity data obtained for the *A. fischeri* and selected additive structural descriptors for the nano-metal oxide mixtures), we have developed two types (all descriptor-based and calculated descriptor-based) of statistically significant nano-mixture QSAR models that reliably predict the mixture toxicity.

From two types of descriptor sets, using four algorithms (RF, SVM, BRNN, and MLR), we obtained eight models for predicting two endpoints, each $\log EC_{50}$ and $\log EC_{10}$ of the MONP mixtures.

3.3.2 Mixture toxicity prediction and validation. Following the OECD QSAR validation recommendations, all nano-

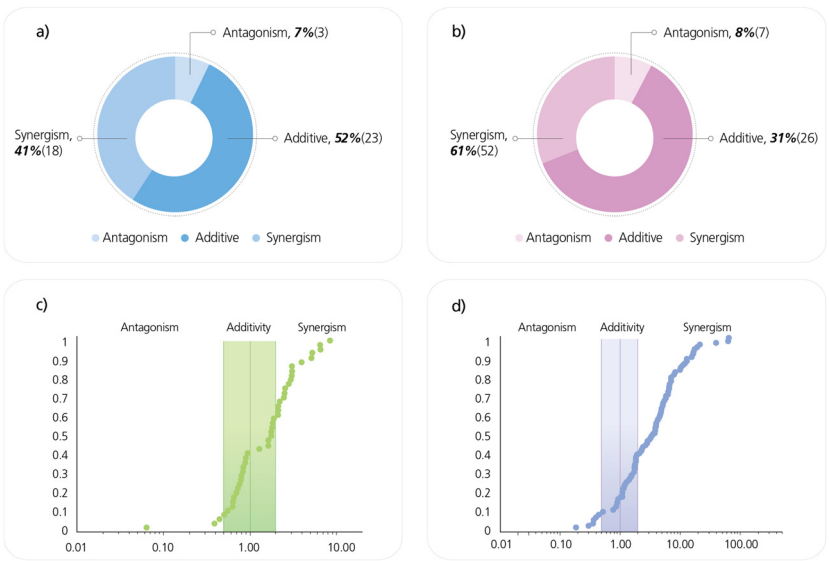

Fig. 3 Model deviation ratio (MDR) value chart of the binary mixture interaction between the metal oxide nanoparticles (MONPs) EC_{50} ratio (Fig. 3a) and EC_{10} ratio (Fig. 3b). Cumulated frequency of MDRs at EC_{50} ratio (Fig. 3c) and EC_{10} ratio (Fig. 3d).

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

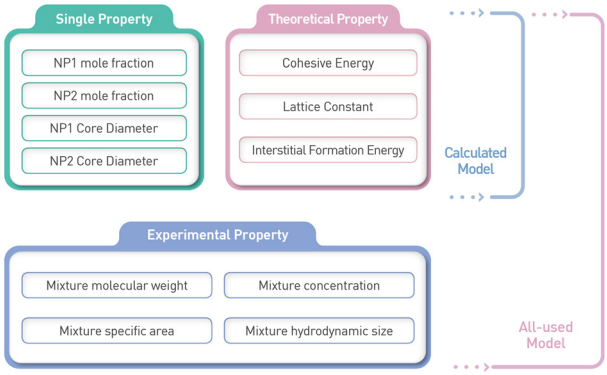

Fig. 4 Descriptors for developing the nano-QSAR model.

mixture QSAR models were validated (internally and externally). All developed models were employed to predict the test set for external validation. As shown in Fig. S4,† a comparison between the predicted and observed $\log EC_{50}$ and $\log EC_{10}$ values illustrates the goodness-of-fit and high predictive ability of the four models, with $R_{train}{}^2$ showing the average result of 10-fold cross-validation. Among the four algorithms, the RF-based models showed good performance in predicting the toxicity of $\log EC_{50}$ and $\log EC_{10}$ for both types of descriptors. For predicting $\log EC_{50}$, the all-descriptor based model showed $R^2 = 0.924$ and $adjR^2 = 0.818$, whose R^2 value is highest among the eight models. The calculation-based model yielded $R^2 = 0.866$ and $adjR^2 = 0.705$. For predicting $\log EC_{10}$, the all-descriptor-based model showed $R^2 = 0.901$ and $adjR^2 = 0.789$, whereas the calculation-based model showed $R^2 = 0.887$ and $adjR^2 = 0.766$. All RF-based models showed high correlations ($adjR^2 > 0.7$) between the experimental and predicted values. Besides the RF algorithm, the other three algorithms showed lower $adjR^2$ than R^2, which indicated that the selected descriptors may have relatively less importance in the three algorithms. In addition, the results suggest that the presented RF-based nano-mixture QSAR models have satisfactory prediction and generalization performances (similar values) (Fig. 5). The *Y*-randomization test yielded Z_{score} values of 15.66 and 6.31 for the RF-based model (calculated descriptor-based) built from $\log EC_{50}$ and $\log EC_{10}$, respectively (Tables S5 and S6†). Therefore, the calculated descriptor-based RF model is statistically significant and unique when compared with models developed with random data. While most previous studies[4,7,8,11] have indicated that the cytotoxicity of the heterogeneous metallic TiO_2 NPs was linearly dependent on the descriptors, the results of our study suggest that the toxicity of the MONP mixtures is nonlinearly dependent on the mixture descriptors.

3.3.3 Model evaluation. Firstly, from the perspective of model performance in predicting the mixture toxicity of the MONPs, nano-mixture QSAR models showed good predictive

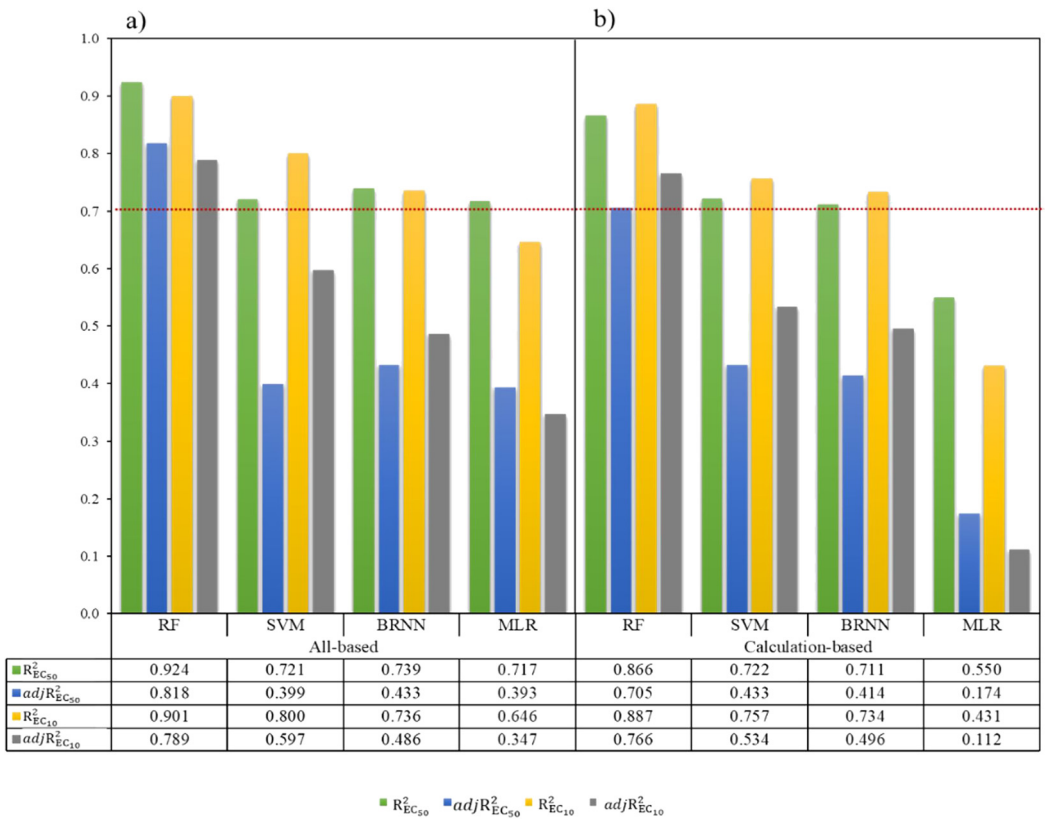

Fig. 5 Model performances for predicting $\log EC_{50}$ and $\log EC_{10}$ of the metal oxide nanoparticles (MONPs). These are descriptor-based models (Fig. 5a) and calculated descriptor-based models (Fig. 5b).

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

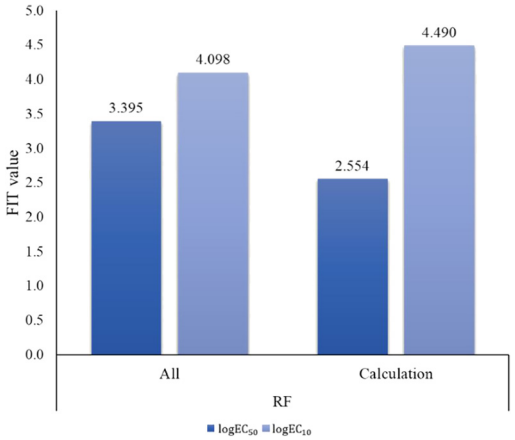

Fig. 6 Kubinyi function (FIT) of the RF-based model predictions for log EC_{50} and log EC_{10} of MONPs.

power for RF-based models (Fig. 5 and 6). The calculation-based RF model for predicting log EC_{10} showed the highest FIT value (FIT = 4.490) among the 16 models. To predict the log EC_{50}, the all-based RF model showed the highest FIT value (FIT = 4.409) and the second-highest value. These two nano-mixture QSAR models constructed with the RF algorithm were recommended to predict the mixture toxicity for MONPs with unknown experimental values. The resultant model can be described as a calculation-based model that is as effective as an experiment-based model. This means that we can predict the toxicity of metal-oxide NPs by calculating only the properties of the individual components, without experimentation or sample measurement.

Secondly, from the perspective of the model application coverage, the nano-mixture QSAR model can predict the SiO_2, TiO_2 (anatase), TiO_2 (rutile), and ZnO mixture toxicity. Although the nano-mixture QSAR model can predict the mixture toxicity of MONPs, its application coverage needs to be extended for various MONPs.

Lastly, from the data availability perspective, the nano-mixture QSAR model has a notable characteristic advantage in that it does not require the mode of action (MoA) information tailored to the target organism, unlike conventional models. Conventional models require toxicity endpoints (*i.e.*, the CA model) and MoA information to predict the mixture toxicity. Since the calculation-based model indicated good performance, the nano-mixture QSAR model could be used without any experiments.

This section addresses the advantages and disadvantages of the nano-mixture QSAR model from three perspectives: model performance, application coverage, and data availability.

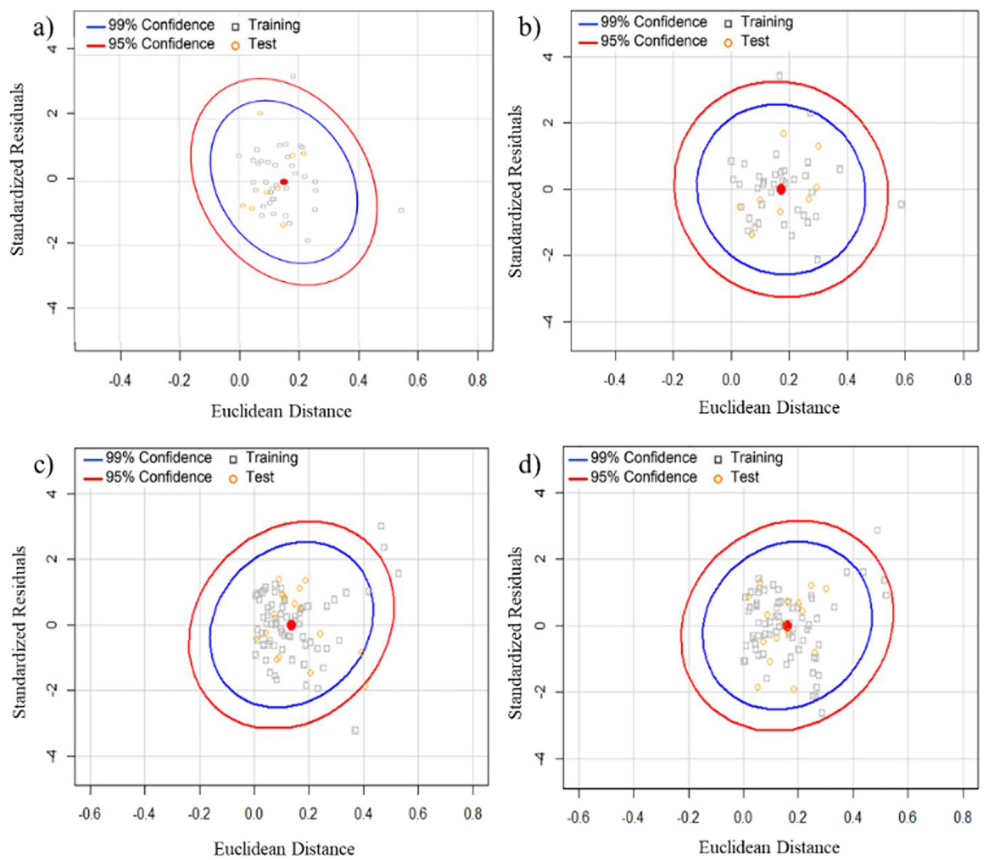

Fig. 7 Applicability domains of the random forest (RF)-based models predicting log EC_{50} (Fig. 7a and b) and log EC_{10} (Fig. 7c and d). All descriptor-based models (Fig. 7a and c) and calculated descriptor-based models (Fig. 7b and d).

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

3.3.4 Applicability domain analysis. According to OECD guidelines,[33] the OECD principle states that "a (Q)SAR should be associated with a defined domain of applicability". The applicability domain (AD) represents the response and chemical structure pace, in which models provide predictions with a given reliability. The ADs in the present study were obtained by applying the Student's *t*-distribution on Euclidean distances (structural domain) and standardized residuals (response domain) of a training dataset to define the space in which reliable predictions are obtained at a certain confidence.[2] The red elliptical boundary corresponded to a confidence interval of 95%, and the toxicity prediction for the data present inside the AD designed by the red ellipse was reliable (Fig. 7). For the models predicting $\log EC_{50}$, all descriptor-based and calculated descriptor-based models, two data points fell outside the confidence area of 99%. Regarding both models, one is too far from the structural domain (Euclidean distances are greater than 0.4), and the other has a high standardized residual value (Fig. 7a and b). For the models predicting $\log EC_{10}$, all descriptor-based and calculated descriptor-based models, five data points fell outside the confidence area of 99%. For all descriptor-based models, three data points were too far from the structural domain, and two data points had high standardized residual values (Fig. 7c). For the calculated descriptor-based models, four data points were far from the structural domain, and the two data points had high standardized residual values (Fig. 7d). Nano-mixture QSAR models in previous studies[4,7,8,11,14] were limited to only TiO_2-based nano-mixtures and limited in nano-size. The QSAR models in the present study have exhibited an extended applicability domain in which toxicity is based on TiO_2, SiO_2, and ZnO of various sizes up to 140 nm. As various types of NPs can be released into the environment, our models were limited to these four types of MONPs. Additional data on other metal oxide-based nano-mixtures should be collected in the future to develop predictive models with a larger applicability domain.

3.3.5 Descriptor importance and mechanistic interpretation. The relative importance of the descriptors in the models predicting $\log EC_{10}$ and $\log EC_{50}$ is shown in Fig. 8. The importance was measured by the weight of descriptors in the RF-based model, with higher weights indicating more relevance to the accurate prediction endpoint. For the descriptor-based model to predict $\log EC_{50}$, the specific surface area, mixture concentration, and mixture molecular weight were the three most important descriptors (Fig. 8a). For all descriptor-based models used to predict $\log EC_{10}$, the mixture concentration, hydrodynamic size, and mixture molecular weight were the three most important descriptors (Fig. 8c). The specific surface area is the most important factor, especially for a high concentration ratio ($\log EC_{50}$). This result is consistent with the fact that NPs are characterized by a large specific surface area, which determines their high reaction capacity and activity. Both models agree that the mixture concentration and mixture molecular weight are important and relevant for predicting the mixture toxicity. In addition, the hydrodynamic size is important and relevant, especially for low concentration ratios ($\log EC_{10}$). This result is in agreement with a previous study, since all powders resulted in similarly sized particles in suspension, regardless of the advertised powder size, and

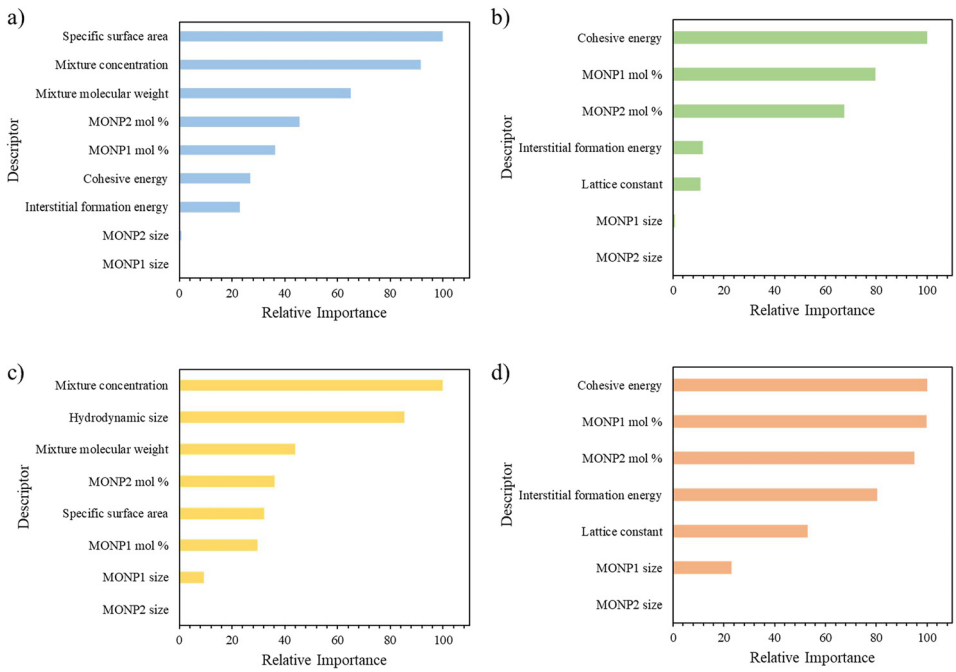

Fig. 8 Relative importance of descriptors in random forest (RF)-based models predicting $\log EC_{50}$ (Fig. 8a and b) and $\log EC_{10}$ (Fig. 8c and d). All descriptor-based models (Fig. 8a and c) and calculated descriptor-based models (Fig. 8b and d).

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

that advertised particle size did not affect the antibacterial activity.[1] For the calculated descriptor-based model, cohesive energy showed the highest relative importance in both the $\log EC_{50}$ and $\log EC_{10}$ prediction models, and identified the most influential properties of the MONPs mixture to *A. fischeri*.

The cohesive energy ($E_C(n)$) is defined as the difference between the average energy of the atoms in solid and isolated atoms, where n denotes the number of atoms in the material. Cohesive energy is considered one of the most fundamental thermal properties that describes the inner structural energy of nanoparticles, and shows how strongly atoms hold together. It is well known that cohesive energy is a parameter that describes the bond strength of a material. Decreasing the cohesive energy of nanoparticles reduces the strength of the corresponding metallic bond. Both experimentally and theoretically, it has been well established that the cohesive energy of nanoparticles decreases with decreasing particle size.[34] The mechanical–electrical properties change with the size and shape of the nanomaterials.[35] The size effect is prominent in nanostructures up to the size limit of approximately 30 nm. However, the effect of size becomes less significant when the size is more than 30 nm,[36] and our results showed the same trend. The cohesive energy decreases with a reduction in the size of the nanomaterials owing to an increase in the number of dangling bonds in the nanomaterials. The decrease in the cohesive energy of nanomaterials results in a decrease in the melting temperature and an increase in the energy bandgap, as is corroborated by previous studies.[35,37] The energy bandgap can be described as the variation between the valence and conduction energy bands of the Gibbs free energy.[35] Additionally, the cohesive energy ($E_C(n)$) determines the size dependence of several physicochemical properties of materials, such as the evaporation temperature, melting temperature, formation enthalpy, surface energy, diffusion activation energy, and bandgap energy.

Consequently, the cohesive energy investigation has become one of the most important topics in the thermodynamics of materials. This is directly related to the thermal stability of the nanoparticles. Therefore, it can be considered that the energy that induces changes in the interactions in oxides is an important property of toxicity. In addition, both models agree that the mole fraction (MONP1 mole fraction, MONP2 mole fraction of components in mixtures) is important and relevant for predicting the mixture toxicity. Based on the relative importance plot for all RF-based models (Fig. 8), it can be inferred that the MONP size had little effect on mixture toxicity.

3.4 Implementation of the RF-based QSAR model

An application with a graphical user interface was developed to enable users to use our model (Fig. 9). This application can be accessed at **https://mjna.shinyapps.io/monps/**. After accessing the website, users can view two subcategories: "log EC_{10}" and "log EC_{50}". The user might need to set parameters for predicting the mixture toxicity value (*e.g.*, MONPs type, core diameter, and mole ratio of NPs). Users do not need to input molecular dynamic descriptors because they have already been calculated between 1 and 100 nm on the web server in the DB format. The application uses user input

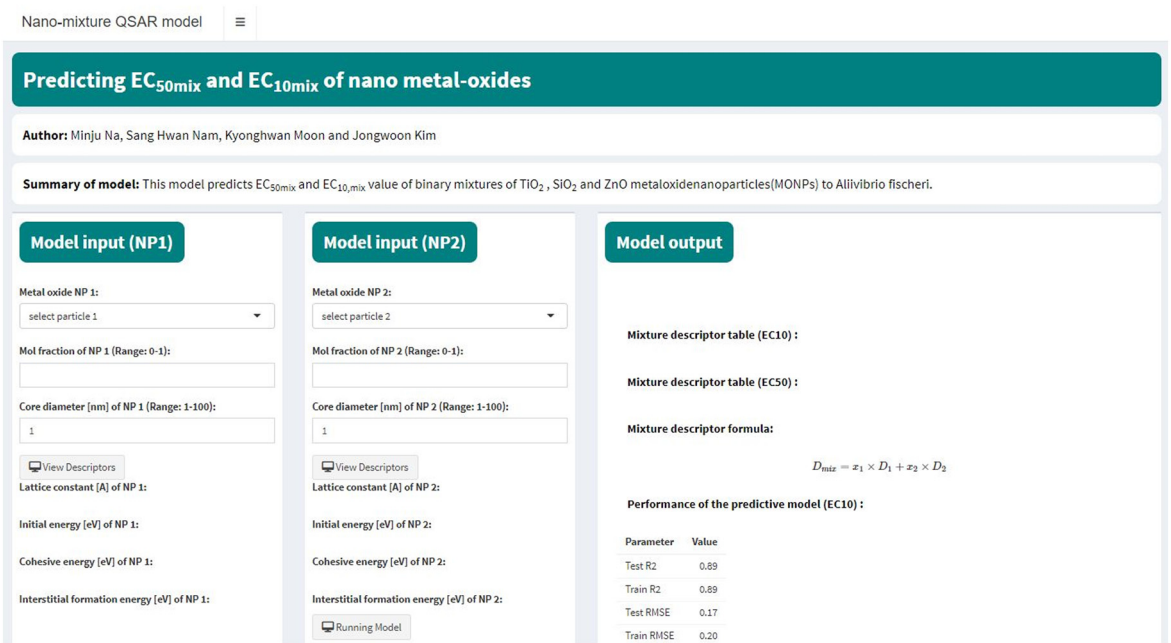

Fig. 9 Screenshot of the prototype of the nano-mixture quantitative structure–activity relationship (QSAR) for the metal oxide nanoparticles (MONPs).

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

data, the MD descriptor corresponding to the input data is called from the built-in DB, and the descriptor is calculated. Then, the D_{mix} descriptors are calculated, the prediction is run using pre-trained models in this study, and the predicted mixture toxicity value is shown ($\log EC_{10}$ and $\log EC_{50}$). The application is currently limited to mixtures of SiO_2, TiO_2, and ZnO MONPs. However, in the future, we might collect more mixture toxicity data and train models again to expand the applicability domain of the application. Furthermore, it will be possible to develop a model based on the toxicity data of MONPs on human cells, and expand it from environmental toxicity to human toxicity.

4. Conclusions

In this study, the toxicity of 12 MONPs (*e.g.*, four types of MONPs) and their binary mixtures (93 samples) were tested at high and low effective concentrations (EC_{50} and EC_{10} ratio mixtures) and EC_{50} + EC_{10} on the bioluminescent bacteria *A. fischeri*. Toxicological interactions were evaluated based on the MDRs between the observed and predicted toxicity values. In addition, the CA model predicted the mixture toxicities to evaluate their toxicological interactions (antagonistic, additive, and synergistic effects) based on the MDR values. Nano-mixture QSAR models were successfully developed in this study to predict the mixture toxicity of MONPs using experimental toxicity data. The proposed approach is the first model of molecular dynamic (MD) descriptors to characterize the structure of nano-metal oxides and develop a mixture toxicity prediction QSAR model. A significant advantage of this proposed approach is that it allows the description of the properties of large (up to 140 nm) MONPs. In addition, we present an easily applicable model for the prediction of the mixture toxicity of MONPs in EC_{50} and EC_{10} (which is similar to the NOEC value). The main findings are summarized as follows: (1) to describe the structure and properties of various sized MONPs, and (2) act as an efficient tool to predict the mixture toxicity of MONPs of various sizes and compositions. Compared to previous nano-mixture QSAR models, the newly developed models in the present study extended the applicability domain to a broader domain, where the toxicity of various sizes of metal oxide-based nano-mixtures.

Author contributions

M. N. and J. K. conceptualized the study. S. H. N. supported sample characterization. M. N. designed and performed the experimental research, performed molecular dynamic calculations, and developed and validated the nano-mixture QSAR models. M. N., K. M., and J. K. discussed the results and compiled the manuscript. All authors approved the final version of the manuscript.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

This study was funded by the Korea Research Institute of Chemical Technology (KRICT) through the development of chemical safety platform technologies (Project No. KK2152-10 and KK2252-10 (NTIS Grant No. 1711134765)). J. Kim acknowledges support from the European Union's Horizon 2020 Research and Innovation Program (SABYDOMA Project under Grant Agreement No. 862296).

Notes and references

1 J. Hou, L. Wang, C. Wang, S. Zhang, H. Liu, S. Li and X. Wang, Toxicity and mechanisms of action of titanium dioxide nanoparticles in living organisms, *J. Environ. Sci.*, 2019, **75**, 40–53.
2 Y. Teow, P. V. Asharani, M. P. Hande and S. Valiyaveettil, Health impact and safety of engineered nanomaterials, *Chem. Commun.*, 2011, **47**, 7025–7038.
3 S. Naasz, S. Weigel, O. Borovinskaya, A. Serva, C. Cascio, A. K. Undas, F. C. Simeone, H. J. P. Marvin and R. J. B. Peters, Multi-element analysis of single nanoparticles by ICP-MS using quadrupole and time-of-flight technologies, *J. Anal. At. Spectrom.*, 2018, **33**, 835–845.
4 A. Mikolajczyk, N. Sizochenko, E. Mulkiewicz, A. Malankowska, B. Rasulev and T. Puzyn, A chemoinformatics approach for the characterization of hybrid nanomaterials: safer and efficient design perspective, *Nanoscale*, 2019, **11**, 11808–11818.
5 J. Kim, S. Kim and G. E. Schaumann, Reliable predictive computational toxicology methods for mixture toxicity: toward the development of innovative integrated models for environmental risk assessment, *Rev. Environ. Sci. Bio/Technol.*, 2012, **12**, 235–256.
6 R. Altenburger, M. Nendza and G. Schuurmann, Mixture toxicity and its modeling by quantitative structure-activity relationships, *Environ. Toxicol. Chem.*, 2003, **22**, 1900–1915.
7 A. Mikolajczyk, A. Malankowska, G. Nowaczyk, A. Gajewicz, S. Hirano, S. Jurga, A. Zaleska-Medynska and T. Puzyn, Combined experimental and computational approach to developing efficient photocatalysts based on Au/Pd–TiO2nanoparticles, *Environ. Sci.: Nano*, 2016, **3**, 1425–1435.
8 A. Mikolajczyk, A. Gajewicz, E. Mulkiewicz, B. Rasulev, M. Marchelek, M. Diak, S. Hirano, A. Zaleska-Medynska and T. Puzyn, Nano-QSAR modeling for ecosafe design of heterogeneous TiO2-based nano-photocatalysts, *Environ. Sci.: Nano*, 2018, **5**, 1150–1160.
9 A. M.-S. Tomasz Puzyn, A. Gajewicz, M. Skrzyński and A. P. Worth, Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models, *Struct. Chem.*, 2011, **22**, 795–804.
10 K. Choudhary, T. Liang, A. Chernatynskiy, S. R. Phillpot and S. B. Sinnott, Charge optimized many-body (COMB) potential

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 22 2022. Downloaded on 2025/8/29 16:15:59.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

for Al2O3 materials, interfaces, and nanostructures, *J. Phys.: Condens. Matter*, 2015, **27**, 305004.
11 B. Yuan, P. Wang, L. Sang, J. Gong, Y. Pan and Y. Hu, QNAR modeling of cytotoxicity of mixing nano-TiO2 and heavy metals, *Ecotoxicol. Environ. Saf.*, 2021, **208**, 111634.
12 G.-X. Zhang, A. M. Reilly, A. Tkatchenko and M. Scheffler, Performance of various density-functional approximations for cohesive properties of 64 bulk solids, *New J. Phys.*, 2018, **20**, 063020.
13 J. Li, C. Wang, L. Yue, F. Chen, X. Cao and Z. Wang, Nano-QSAR modeling for predicting the cytotoxicity of metallic and metal oxide nanoparticles: A review, *Ecotoxicol. Environ. Saf.*, 2022, **243**, 113955.
14 T. X. Trinh, M. Seo, T. H. Yoon and J. Kim, Developing random forest based QSAR models for predicting the mixture toxicity of TiO2 based nano-mixtures to *Daphnia magna*, *NanoImpact*, 2022, **25**, 100383.
15 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 and S. J. Plimpton, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, *Comput. Phys. Commun.*, 2022, **271**, 108171.
16 V. Tsiridis, M. Petala, P. Samaras, S. Hadjispyrou, G. Sakellaropoulos and A. Kungolos, Interactive toxic effects of heavy metals and humic acids on *Vibrio fischeri*, *Ecotoxicol. Environ. Saf.*, 2006, **63**, 158–167.
17 G. Chen, M. G. Vijver, Y. Xiao and W. Peijnenburg, A Review of Recent Advances towards the Development of (Quantitative) Structure-Activity Relationships for Metallic Nanomaterials, *Materials*, 2017, **10**(9), 1013.
18 M. Heinlaan, A. Ivask, I. Blinova, H. C. Dubourguier and A. Kahru, Toxicity of nanosized and bulk ZnO, CuO and TiO2 to bacteria *Vibrio fischeri* and crustaceans *Daphnia magna* and Thamnocephalus platyurus, *Chemosphere*, 2008, **71**, 1308–1316.
19 DS/ISO 21338 – Water quality – Kinetic determination of the inhibitory effects of sediment, other solids and coloured samples on the light emission of *Vibrio fischeri* (kinetic luminescent bacteria test).
20 K. Choudhary, T. Liang, A. Chernatynskiy, S. R. Phillpot and S. B. Sinnott, Charge optimized many-body (COMB) potential for Al2O3 materials, interfaces, and nanostructures, *J. Phys.: Condens. Matter*, 2015, **27**, 305004.
21 B. Gregorutti, B. Michel and P. Saint-Pierre, Correlation and variable importance in random forests, *Stat. Comput.*, 2016, **27**, 659–678.
22 P. M. Granitto, C. Furlanello, F. Biasioli and F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, *Chemom. Intell. Lab. Syst.*, 2006, **83**, 83–90.
23 R Development Core Team Version 3.5.1, *R: a language and environment for statistical computing. Version 3.6.2*, R Foundation for Statistical Computing, Vienna, 2020.
24 RStudio Team, *RStudio: Integrated Development for R. RStudio*, PBC, Boston, MA, 2021, **https://www.rstudio.com/**.
25 P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR Comb. Sci.*, 2007, **26**, 694–701.
26 C. Rücker, G. Rücker and M. Meringer, y-Randomization and Its Variants in QSPR/QSAR, *J. Chem. Inf. Model.*, 2007, **47**, 2345–2357.
27 H. Kubinyi, Variable Selection in QSAR Studies. II. A Highly Efficient Combination of Systematic Search and Evolution, *QSAR Comb. Sci.*, 1994, **13**, 393–401.
28 I. Corsi, A. Bellingeri, M. C. Eliso, G. Grassi, G. Liberatori, C. Murano, L. Sturba, M. L. Vannuccini and E. Bergami, Eco-Interactions of Engineered Nanomaterials in the Marine Environment: Towards an Eco-Design Framework, *Nanomaterials*, 2021, **11**(8), 1903.
29 L. K. Adams, D. Y. Lyon and P. J. Alvarez, Comparative eco-toxicity of nanoscale TiO2, SiO2, and ZnO water suspensions, *Water Res.*, 2006, **40**, 3527–3532.
30 W. Jiang, K. Yang, R. W. Vachet and B. Xing, Interaction between oxide nanoparticles and biomolecules of the bacterial cell envelope as examined by infrared spectroscopy, *Langmuir*, 2010, **26**, 18071–18077.
31 D. Wang, Z. Lin, Z. Yao and H. Yu, Surfactants present complex joint effects on the toxicities of metal oxide nanoparticles, *Chemosphere*, 2014, **108**, 70–75.
32 Y. W. Baek and Y. J. An, Microbial toxicity of metal oxide nanoparticles (CuO, NiO, ZnO, and Sb2O3) to *Escherichia coli, Bacillus subtilis*, and *Streptococcus aureus*, *Sci. Total Environ.*, 2011, **409**, 1603–1608.
33 OECD, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, 2014.
34 X. Li, Modeling the size- and shape-dependent cohesive energy of nanomaterials and its applications in heterogeneous systems, *Nanotechnology*, 2014, **25**, 185702.
35 G. Guisbiers, Advances in thermodynamic modelling of nanoparticles, *Adv. Phys.: X*, 2019, **4**, 1668299.
36 K. P. Singh and S. Gupta, Nano-QSAR modeling for predicting biological activity of diverse nanomaterials, *RSC Adv.*, 2014, **4**, 13215–13230.
37 M. Singh, M. Goyal and K. Devlal, Size and shape effects on the band gap of semiconductor compound nanomaterials, *J. Taibah Univ. Sci.*, 2018, **12**, 470–475.