Using the Isalos platform to develop a (Q)SAR model that predicts metal oxide toxicity utilizing facet-based electronic, image analysis-based, and periodic table derived properties as descriptors

Thwala, M. M.; Afantitis, A.; Papadiamantis, A. G.; Tsoumanis, A.; Melagraki, G.; Dlamini, L. N.; Ouma, C. N.M.; Ramasami, P.; Harris, R.; Puzyn, T.; Sanabria, N.; Lynch, I.; Gulumian, M.

Link to publication on Research at Birmingham portal

ORIGINAL RESEARCH

Using the Isalos platform to develop a (Q)SAR model that predicts metal oxide toxicity utilizing facet-based electronic, image analysis-based, and periodic table derived properties as descriptors

M. M. Thwala[1,2] · A. Afantitis[3] · A. G. Papadiamantis[3,4] · A. Tsoumanis[3] · G. Melagraki[5] · L. N. Dlamini[1] · C. N. M. Ouma[6] · P. Ramasami[7] · R. Harris[8] · T. Puzyn[9,10] · N. Sanabria[2] · I. Lynch[4] · M. Gulumian[2,11,12]

Abstract

Engineered nanoparticles (NPs) are being studied for their potential to harm humans and the environment. Biological activity, toxicity, physicochemical properties, fate, and transport of NPs must all be evaluated and/or predicted. In this work, we explored the influence of metal oxide nanoparticle facets on their toxicity towards bronchial epithelial (BEAS-2B), Murine myeloid (RAW 264.7), and *E. coli* cell lines. To estimate the toxicity of metal oxide nanoparticles grown to a low facet index, a quantitative structure–activity relationship ((Q)SAR) approach was used. The novel model employs theoretical (density functional theory calculations) and experimental studies (transmission electron microscopy images from which several particle descriptors are extracted and toxicity data extracted from the literature) to investigate the properties of faceted metal oxides, which are then utilized to construct a toxicity model. The classification mode of the k-nearest neighbour algorithm (EnaloskNN, Enalos Chem/Nanoinformatics) was used to create the presented model for metal oxide cytotoxicity. Four descriptors were identified as significant: core size, chemical potential, enthalpy of formation, and electronegativity count of metal oxides. The relationship between these descriptors and metal oxide facets is discussed to provide insights into the relative toxicities of the nanoparticle. The model and the underpinning dataset are freely available on the NanoSolveIT project cloud platform and the NanoPharos database, respectively.

Keywords Cytotoxicity · Nano-(Q)SAR · Facets · Descriptors · Nanotopography

✉ M. Gulumian
MaryG@nioh.ac.za

1. Department of Chemical Sciences, University of Johannesburg, Doornfontein, Johannesburg 2028, South Africa

2. Department of Toxicology and Biochemistry, National Institute for Occupational Health, Constitutional Hill, Hospital Street, Johannesburg 2000, South Africa

3. NovaMechanics Ltd, Cheminformatics & Nanoinformatics Excellence, P.O Box 26014, Nicosia 1666, Cyprus

4. School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK

5. Division of Physical Sciences and Applications, Hellenic Military Academy, Vari, Greece

6. Faculty of Engineering, HySA Infrastructure Center of Competence, North West University, P. Bag X6001, Potchefstroom 2520, South Africa

7. Computational Chemistry Group, Department of Chemistry, Faculty of Science, University of Mauritius, Reduit 80837, Mauritius

8. Natural and Agricultural Sciences, University of Free State, PO Box 339, Bloemfontein 9300, South Africa

9. QSAR Laboratory Limited, Aleja Grunwaldzka 190/102, Gdansk 80-266, Poland

10. Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, Gdansk 80-308, Poland

11. Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa

12. Water Research Group, Unit for Environmental Sciences and Management, North West University, Private Bag X6001, Potchefstroom 2520, South Africa

Introduction

Metal oxide nanomaterials (NMs) have demonstrated exceptional structural, electronic, and chromic properties, and as such they find applications in various fields of technology. These include applications in cosmetics, sunscreens, textile, devices, and water treatment systems [1–3]. The vast application of metal oxides in various nanotechnology-related fields has drawn concerns with regard to their safety and toxicity, as some of these nanoparticles (NPs) have been reported to be toxic to some organisms [4, 5]. In this regard, there is a high possibility that the sunscreens, which are intended to protect individuals from UV radiation, may cause more harm than the UV radiation itself [6]. Furthermore, metal oxide–based water remediation systems may pose different risks to the environment than the conventional remediation methods [7]. This provides grounds for the need to evaluate metal oxide toxicity in parallel to development of applications, as part of a responsible innovation approach. Several experimental studies have been conducted to assess the toxicity of metal oxides, examining various parameters that can cause toxicity [4, 8, 9].

The rate at which metal oxides are applied in different fields of technology far exceeds the rate at which the toxicity of the materials is studied. Furthermore, conventional (i.e. experimental) risk assessment techniques are often time intensive and insufficient for ensuring the safety by design of newly produced materials in rapidly expanding application areas. To keep up with the rapid rates of metal oxide production, while also lowering the number of tests and the amount of consumable reagents utilized, a method that can predict the toxicity of metal oxides is required. The application of quantitative structure activity relationship ((Q)SAR) is one approach that has been used to evaluate the toxicity of metal oxides, but to a lesser extent compared to chemical compounds [10]. As a result, extensive studies on the use of (Q)SAR models in predicting toxicity of metal oxide NPs are required to support the validation and regulatory acceptance of NMs (Q)SARs in regulatory risk assessment..

Metal oxide NPs (Q)SAR is a relatively new type of (Q)SAR that is characterized as a mathematical relationship between the properties (descriptors) of metal oxides and their biological activity. This type of model is typically referred to as "nano-(Q)SAR," and the accompanying descriptors are referred to as "nano descriptors." It is possible to compute the activity of additional NPs (where toxicity tests have not yet been carried out) by using an appropriate mathematical model, i.e. once the physico-chemical parameters of metal oxides and their toxicity towards living cell lines have been calculated and the model has been validated using a test set of NMs. Predictions for new NMs can then be made once the new NMs properties fall with the domain of applicability of the (Q) SAR model, where reliable predictions can be made [11].

Classical NP properties, such as hydrodynamic size, core size, and surface charge based on 21 NPs, were used to develop a (Q)SAR model for predicting their toxicity towards human keratinocyte cells (HaCat) and bronchial epithelium transformed with Ad12-SV40 2B (BEAS-2B) cells [12]. Puzyn et al. also reported on a (Q)SAR model used to predict toxicity of 17 metal oxides, using enthalpy of formation as a descriptor for the model [13]. This model was able to successfully predict the cytotoxicity of metal oxides towards *E. coli* cells. Meanwhile, another model, based on seven metal oxides from different experimental conditions, has been reported [14]. On the other hand, Toporova et al. proposed (Q)SAR models based on simplified molecular input-line entry system (e.g. Simplified Molecular Input Line Entry System, "SMILES") [15–17].

The reliability of a (Q)SAR model depends on the set of data that was used to develop it [10]. There is no standardized dataset for the formulation of (Q)SAR models, as data used for model development are obtained from different sources and for varying experimental conditions. Density functional theory (DFT) calculations can play a vital role in the calculation of metal oxide descriptors as they have been successfully used for the prediction of metal oxide properties in various studies [18–21]. Properties obtained via DFT calculations can be used as descriptors for (Q)SAR model development. Venigalla and co-workers [22] reported on the development of a (Q)SAR model using 17 metal oxides, the descriptors of which were calculated using DFT methods. Meanwhile, Yunsong [23] explained the significance of using DFT and empirical-based descriptors to derive toxicity reaction mechanisms. The significance of in silico methods is better understanding of the features that influence MeONP potencies, as well as predict toxic responses and effect thresholds.

In the current study, a (Q)SAR model based on an extended database of the cytotoxicity of 26 metal oxide NPs to human BEAS-2B, Murine myeloid (RAW 264.7) cell lines, and *E. coli* was developed and validated. The confluence of the dataset is aimed to achieve a model with a broader spectrum of application and where the effects of facets towards toxicity have been taken into account. The bacteria were specifically selected for cytotoxicity assessment, since they are considered a good ecological indicator for assessing the persistence and impact of chemicals on the environment and human health. In addition, uncontrolled release of toxic substances to the bacterial environment may disturb their natural balance, resulting in unwanted effects on the environment [24]. Meanwhile, the

pulmonary epithelial BEAS-2B (non-tumorigenic human lung epithelial cells) and macrophage RAW 264.7 cell lines are good models for mimicking human cells during inhalation exposures. The catalytic properties of NPs are determined by the nature of the NP surface [25]. The goal of this study was to provide a method for generating metal oxide nano-descriptors that are related to exposed facets (i.e. one of the indicators specifically related to the surface nanotopography), which may be used to characterize the observed activity of NPs towards biological cell lines.

To the best of our knowledge, descriptors on metal oxide NPs that are facet-specific have not previously been reported in the literature and are reported herein for the first time. An interpretive nano-(Q)SAR model for an extended dataset of metal oxide NPs toxicity, based on various cell lines, relies on the combined theoretical and experimental work.

Materials and methods

Data

Toxicity data used for the model development was obtained from the literature [13], which was comprised of 26 metal oxides. (Q)SAR models have also been documented with significantly less datasets [26]. DFT simulations for low-indexed facets were used to calculate descriptors based on the surface energy and electronic properties of metal oxides. Descriptors were also generated using atomic periodicity, experimental data, and TEM data. Recent developments have focused on growing NPs with specific facets for specific applications and improved efficiency. These give a rationale for looking into the impact of exposed facets towards toxicity of NPs from a computational point of view. Various other publications have used these data, looking into different descriptors and models [27, 28]. There were no data gaps in the data set, which included 26 distinct metal oxide NPs. Types of metal oxides used in the model were of the form M_3O_4 M_2O_3, MO_2, and MO, where M (M = Al, Fe, Cr, Bi, La, Y, V, Sb, In, Ti, Sn, Si, Zr, Co, Cu, Ni, Zn, Co, Gd, Hf, Mn, W, and Yb) is the metal atom and O is the oxygen atom. The descriptors were classified into six categories: physicochemical, structural, image, periodic table, experimental, and molecular. Molecular descriptors included binding energy, Fermi energy, HOMO (highest occupied molecular orbitals), LUMO (lowest unoccupied energy orbitals), band gap, hardness, chemical potential, enthalpy of formation, and electronegativity. It is worth noting that molecular descriptors were derived from DFT calculations and were based on low index facet(s). The full list of descriptors collected from literature, calculated, and utilized in the model is given in the supporting information (Table S1).

Calculation of molecular descriptors

The molecular descriptors are numerical representations of the properties of metal oxides, which were based on their electronic properties. A calculated pool of molecular descriptors is presented in the supporting information (Table S1). These include surface energy, binding energy, HOMO, LUMO, band gap, electronegativity, Fermi level energy, chemical potential, and hardness. In spite of its shortcomings in estimating lattice constants and electronic band gaps, DFT is credited with giving insights to experimentally observed phenomena such as catalysis, photonics, and defects; hence, the reason for its use in this study to generate descriptors for (Q)SAR [29, 30]. Additionally, DFT calculations have been used to generate datasets for high throughput screening of materials, compounds and alloys, rational design of catalysts, and for training machine learning models [31, 32]. Metal oxide chemical space was limited to four types of metal oxides (M_3O_4, M_2O_3, MO_2, and MO).

It is worth mentioning that in doing DFT calculations, the primitive cell (lattice) and not the unit cell were used for the calculations. All the crystal structures were obtained from https://materialsproject.org. DFT calculations have been documented to over (under) estimate the lattice constants by almost 20% depending on whether generalized gradient approximation (GGA) pseudopotentials are used to account for core and valence electrons [21, 33]. However, this has not negated DFT from being used to offer insights on experimentally observed phenomena [19, 22, 34]. It is for this reason that DFT does not reproduce the same lattice constants as experiments. DFT calculations are ab initio in nature and hence do not account for temperature in the calculations. All metal oxides considered in this work were chosen with respect to their lowest convex hull energy.

All ab initio calculations, used to generate the descriptors in this study, were performed within DFT formalism as implemented in DMol [3] code within Material Studio (Accelrys, San Diego, CA) [35]. The GGA with the Perdew–Burke–Ernzerhof (PBE) functional was used to describe exchange–correlation effects [36]. The geometry optimization convergence tolerances were set at 10^{-5} Ha (1 Ha = 27.21 eV) for energy, 0.002 Ha for maximum force, and 0.005 Å for maximum displacement. The tolerance to electronic self consistent field (SCF) was set at 10^{-6} Ha, while a smearing of 0.02 Ha to orbital occupation was applied. Low index surfaces ((100), (110), and (111)) were constructed from the optimized bulk structures, using aslab with a vacuum of 30 Å along *c*-axis was used between the periodical slabs to eliminate spurious interactions.

The full, computationally enriched, dataset was made publicly available from the NanoPharos database (https://db.nanopharos.eu/), developed via the NanoSolveIT project and continues to be maintained by NovaMechanics Ltd.

Model development

Model development was performed using the Isalos Analytics Platform, powered by the Enalos + Tools [37, 38]. The dataset was checked for completeness, and no gaps were identified. The first step in the process was to feed the dataset into a low-variance filter in order to remove those descriptors that did not present significant variance and could not contribute to the model's predictive capacity [39]. In this way, the workload and time needed for the computational workflow to complete is reduced as well. The cut-off threshold for the model was set to 0.3, which meant the descriptors with 30% or similarity of values to another descriptor were removed. The next step was to account for the different numerical ranges of the included descriptors in the filtered dataset, for which Z-score normalisation was utilized. This ensured that the numerical range of all descriptors followed a Gaussian distribution with mean values and standard deviation of 0.0 and 1.0, respectively [40]. The model was then developed following random partitioning of the dataset into training and test sets, using a 75%:25% ratio. The training set was used for model development and training as well as to evaluate the model's performance and fine tune its parameters through cross-validation [41, 42]. The descriptors that were used to develop the model, (i.e. those presenting the highest statistical significance with respect to the dataset's variance), were identified using the Correlation based Feature Selection (CfsSubset) algorithm combined with the BestFirst evaluator [43].

The presented metal oxide cytotoxicity model was developed using the Enalos implementation of the classification mode of the *k*-nearest neighbour algorithm (Enalos*k*NN), Enalos Chem/Nanoinformatics) [44]. Enalos*k*NN is an instance-based (lazy) method that uses the distance of the predicted endpoint from its k ($k = 1, 2, 3, …$) nearest neighbours in the feature space R^n created by the "n" identified significant descriptors, which are used to make the prediction, with "k" being defined based on the model's best performance. The prediction is achieved based on the Euclidean distance, a similarity measure, of the target variable from its "k" closest neighbours [45]. The prediction is performed via the weighted average of the independent variable values of these neighbours, with the inverse of the Euclidean distance being used as the weighing factor [45, 46]. In the case of 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 [47].

The Euclidean distance calculated though model development and usage can be used as a metric to not only predict a specific endpoint, but to also identify groups of neighbouring nanomaterials (NMs). The identification and analysis of these groups can lead to mapping the prediction space into specific NMs groups, which can then be used for the development of read-across strategies [48]. Therefore, by taking advantage of the Enalol*k*NN ability to provide the Euclidean distances, we can visualize and study the entire predictive space R^n. As a result, the kNN algorithm can be used according to ECHA's read across framework [49] for NMs only if the following criteria are satisfied:

- Gathering of required descriptors for each NM.
- Construction of data matrix including properties and endpoints.
- Development of a correlation between end points to reactivity properties.
- Assessment of the applicability of the approach.
- Ensurance that there is no missing data.
- Assessment of the robustness of the grouping.
- Justification of the method.

Model validation

The model's validity and robustness were tested using its sensitivity (Sn), specificity (Sp), and accuracy (Ac). These metrics describe the proportion of the correct predictions of toxic NMs, the proportion of the NMs that were correctly classified as non-toxic and the model's overall success rate, respectively [50]. Cohen's κ, which measures the model's reliability, while taking into account any successful predictions to be based on chance correlation [51] was calculated as well.

$$Sn = \frac{TP}{TP + FN} \tag{1}$$

$$Sp = \frac{TN}{TN + FP} \tag{2}$$

$$Ac = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

where, TP are true positives, TN are true negatives, FP are false positives, and FN are false negatives.

The model was furthermore evaluated using the Matthews correlation coefficient (MCC) [52], which is used as a quality measure for the development of predictive classification workflows. The MCC takes into account the true and false positive and negative outcomes of the developed model, and it is considered as a good quality metric of the model, even

in cases of unbalanced datasets [53], The MCC results range between -1 and $+1$. An MCC value of -1 corresponds to total disagreement between observed (experimental) and predicted results, while a value of $+1$ corresponds to total agreement. An MCC value of 0 corresponds to a random prediction [54]. The MCC is calculated using Eq. (4):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{4}$$

where TP are true positives, TN are true negatives, FP are false positives, and FN are false negatives.

Y-randomisation ($n = 10$) was used to guarantee that the created model was not the result of chance correlation and to test its statistical significance and robustness. Random shuffling of the prediction using all original descriptors yielded different data sets. The model acceptance criteria given above were calculated for each iteration. The revised criteria were expected to be lower than the original model for the model to be valid [55–57].

Applicability domain

In order to ensure accessibility within the scientific community and to interested stakeholders as well as to ensure its validity and reliable usage and applicability to external datasets, the applicability domain (APD) of the (Q)SAR model was calculated using the Euclidean distances calculated and retrieved via the Enalso$k$NN. The following equation was used:

$$APD = <d> + Z\sigma \tag{5}$$

where $<d>$ and σ are the average and standard deviation of all Euclidean distances in the training set, respectively, and Z is the empirical cut off value, which is usually 0.5. Any predictions made outside these defined limits are regarded as unreliable [58].

An analytical summary of the produced model, along with the full demonstration that the produced model meets the OECD criteria for the validation of (Q)SAR models for regulatory purposes is demonstrated via the completed (Q) SAR Model Reporting Format (QMRF) template which is included in the supplementary information (S1).

Webservice development

The fully documented model and relevant tutorials have been made publicly available, through the NanoSolveIT cloud platform (https://cloud.nanosolveit.eu/), as a user-friendly webservice (https://facetcytotoxicity.cloud.nanosolveit.eu/) to ensure accessibility within the scientific community and to interested stakeholders. The model has been complemented with a REST API (Fig. 1; https://facetcytotoxicity.cloud.nanosolveit.eu/swagger-ui/) to make it easily accessible and usable programmatically and to enable to implementation into a computational workflow, e.g. as KNIME node. The API has been implemented using the POST Request Method to be able to transfer and handle large amounts of data that are necessary to run the model. Following analysis, the results are returned in JSON format.
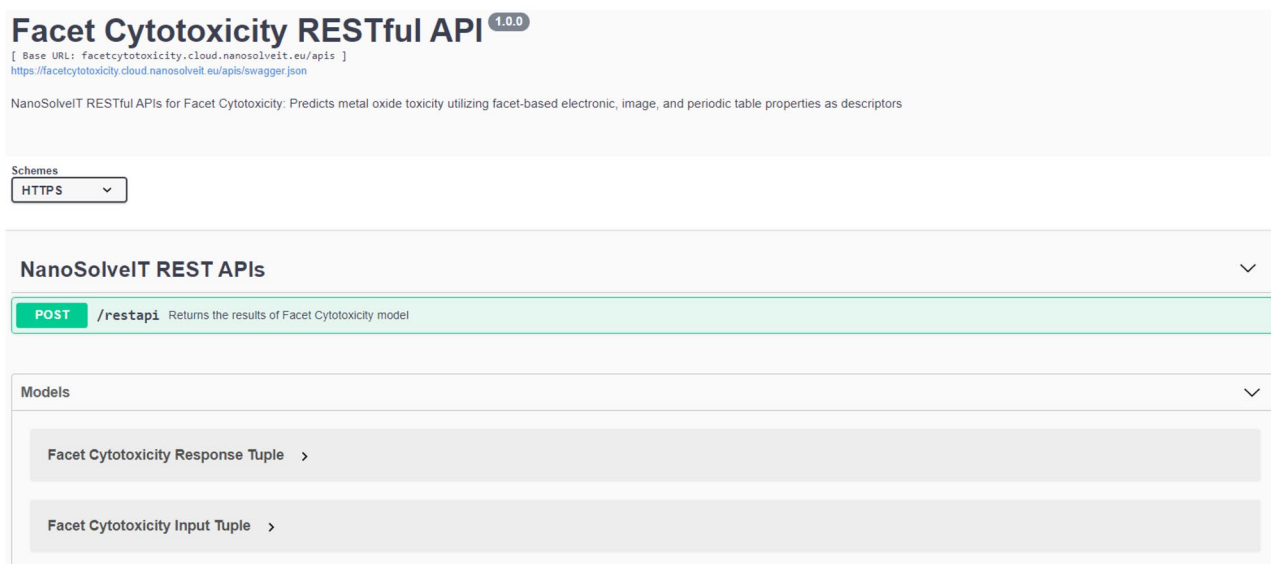

Fig. 1 The REST API environment for the facet-driven cytotoxicity model presented in this study. Through the API users can implement the model into their own computational workflows

Results and discussion

The goal of this study was to show that cytotoxicity of NPs may be predicted using a combination of physicochemical, molecular, and periodic table-based descriptors derived for low index facets of metal oxide NMs. The dataset included 26 metal oxides in replicates of varied experimental settings, as well as a total of 32 descriptors, i.e. 8 molecular descriptors, 9 periodic table-based descriptors, 10 TEM image-based descriptors, and 5 experimental/physicochemical descriptors (Table S1). Although the data set is modest statistically, it yields a solid predictive model. Merging datasets was carried out to ensure that they were interoperable in order to identify probable reasons for variability. This also emphasized the importance of including sufficient metadata in public datasets to improve their quality, FAIRness score, and consequently reproducibility, reusability, and scientific transparency in general [59, 60].

Development of the model Following a random division of the dataset into training and test sets using a ratio of 75%:25%, respectively (Fig. 1), the predictive model was created. Using the CFsSubset algorithm and the BestFit evaluator, the descriptors that contributed the most to the model's variance and subsequently used to perform the classifications were determined. As a result, four descriptors were identified to be significant, i.e. core size, chemical potential, metal electronegativity count, and enthalpy of formation of metal the oxide. These results contain a mixture of descriptors, which includes physicochemical, periodic table and molecular-based descriptors calculated with low index facets in consideration. They are also in good agreement with a similar study by Papadiamantis et al. [44] on metal oxide toxicity to BEAS-2B and RAW 264.7 cell lines, where the core size was also identified as significant. Papadiamantis et al. identified the energy of the conduction band as significant, which is directly correlated with the metal electronegativity, and subsequently the metal electronegativity count presented herein (see also Eqs. 10 and 11 below) and the enthalpy of formation [61]. Furthermore, the chemical potential, which is linked to the arrangement of the atoms on the NPs surface, is correlated to the identified average coordination number of metal atoms in the surface region of the NPs and the average length of the surface normal component of force vector of atoms in the surface region of the NP, which describes the potential energy (stability and activity) of the atoms on the surface of NPs [61]. This parameter is unique for each exposed facet (e.g. {100} or {111}) on metal oxide MNs and hence explains their varied toxicity behaviour. The produced model had high predictivity, having an Ac value of 0.929, Sn of 0.889, and Sp of 1.000. Cohen's κ was calculated to be 0.851, and the APD is 1.951. The MCC value of the produced model was 0.861, denoting a good prediction for both classes. The predictions were performed for $k = 4$ closest neighbours (Fig. 2).
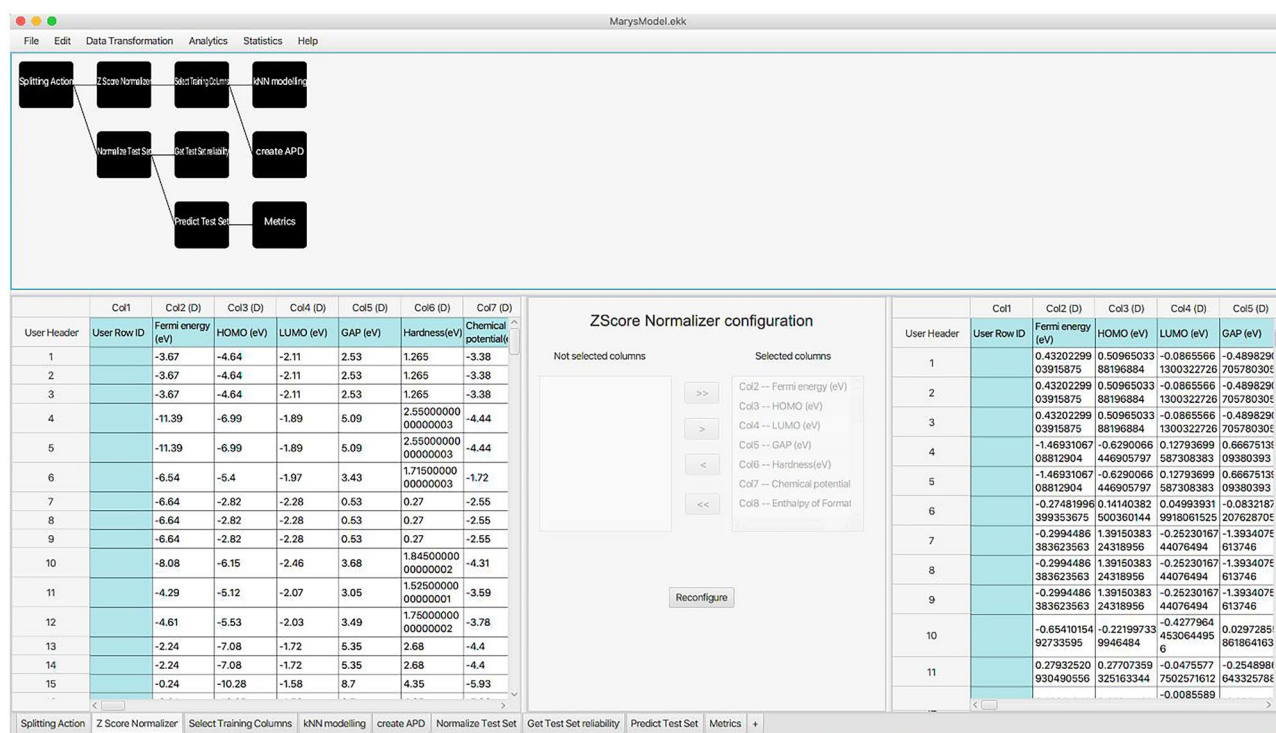

Fig. 2 The Isalos Analytics Platform environment, with the developed workflow and the respective dataset

Justification of the descriptors implemented

There are four main mechanisms for NP toxicity: (i) the release of chemical constituents from nanomaterials; (ii) the size and shape of particle, which produce steric hindrances or interference with important binding sites of macromolecules; (iii) the surface properties of the material such as photochemical and redox potentials; and (iv) the capacity of nanomaterials to act as vectors for the transport of other toxic chemicals to sensitive tissues [13].

Core size was found to be one of the significant descriptors of the model presented herein, and this is in agreement with mechanism (ii). The effect of core size towards toxicity has been reported in the past, where a trend of decreasing core size has been connected to greater toxicity [62–64]. Earlier studies by Karlsson et al. compared metal oxide NPs of various sizes when they were exposed to adenocarcinomic human alveolar basal epithelial A549 cells [65]. Their findings showed that toxicity of NPs cannot be generalized solely on the basis of their core size.. The toxicity of CuO NPs increased with the observed decrease in core size, whereas the toxicity of TiO_2, Fe_2O_3, and Fe_3O_4 was unaffected by core size, regardless of having the same chemical composition. Furthermore, Warheit and co-workers reported that the toxicity of TiO_2 particles was determined by surface properties, rather than size and surface area [66]. Similarly, Ivask and colleagues discovered that while human colorectal adenocarcinoma (Caco2) cells ingested TiO_2 particles, no cytotoxicity was induced [4].

Facets are among the surface properties that have a significant influence on the toxicity of metal oxides. Liu and colleagues observed that faceted TiO_2 metal oxides (e.g. {001} facet) are more toxic than spherical metal oxides, due to their preferentially exposed crystallographic facets with large densities of unsaturated bonds [67]. The precise property-activity relationship was used to evaluate the toxicity of faceted metal oxides, using TiO_2 bipyramids with varying percentages of exposed {001} and {101} facets on the surface. The {001} facet was found to elicit severe toxicitys compared to the {101} facet, and this was attributed to the high production rate of hydroxyl radicles on the {001} surface. Similarly, differently exposed facets ({100} and {111}) of Cu_2O nanocrystals were reported to have varied cytotoxic effects on RAW 264.7 cells [68]. Plausible mechanistic explanations were attributed to the formation of hydroxyl radicals on the facet surfaces of Cu_2O for short term exposure, and release of Cu^+ ions for long-term exposure. The {100} facet was considered to release a higher concentration of Cu ions than the {111} facet. This was attributed to the alternating stacking of Cu^+ and O [2] ion on the {100} facet, while the Cu^+ ion in the {111} facet are packed between the O^{2-} ions making release of Cu^{2+} ions more challenging [69].

Chemical stability, which is connected to particle dissolution, catalytic properties, and redox alteration on the surface, is the most important regulating parameter for metal oxide toxicity [70]. This is consistent with the aforementioned mechanisms (i) and (iii). Ions can be released by breaking chemical bonds in the metal oxide's lattice structure. Such reactions are widespread near the material's surface and are influenced by the metal oxide's exposed facet, as indicated in for Cu_2O NMs [71]. Furthermore, the growth of metal oxide NMs towards certain facets is strongly linked to a specific lattice energy, which defines the dissolution of NMs without oxidation or reduction. As a result, the lattice energy of metal oxides varies depending on their facets, and hence, their stability varies.

Negative values for lattice energy increase with increasing cation charge (n). Similarly, the increase in positive value of enthalpy of formation is associated with an increase in cation charge. Consequently, the release of cation Me^{n+} having smaller charge is more energetically favoured than the release of ions with larger cation charges. This explains why the toxicity of studied metal oxides decreases in the following order: $Me^{2+} > Me^{3+} > Me^{4+}$. Furthermore, because creation of Me^{n+} cations necessitates sublimation followed by ionization processes, the enthalpy of formation (ΔH_f) is linked to the sum of ionization potentials of a specific metal, and thus, can be calculated as

$$\Delta H_f = \Delta H_s + \sum_{i=1}^{n} IP_i \tag{6}$$

where ΔH_s is the enthalpy of sublimation and IP_i is the nth ionization potentials of the metals.

Enthalpy of formation is not connected to metal oxide size NM, which supports our previous assertion that size cannot be the only contributing factor to metal oxide NM toxicity. Furthermore, as much as enthalpy of formation influences the release of metal ions, it also serves as an indicator for average, metal–oxygen bond strength [72]. The metal–oxygen bond strengths vary within the same crystal structure, and this is most noticeable with different facets. This explains why the ion release varies for different facets of the same metal oxide crystallographic structure.

Curvature is a fundamental variable that modulates the forces and controls the size and shape of NMs. The shape index is defined as follows:

$$S = \frac{2}{\pi}\arctan\left(\frac{K_2 + K_1}{K_2 - K_1}\right), \left(K_1 \geq K_2, K_1 = 0 \text{ only if } K_2 \neq 0\right) \tag{7}$$

where K_2 and K_2 are the maximum and minimum principal curvature, respectively. It represents a visual interpretation of curvature with a singular range of values ($-1 \geq S \geq 1$). A

positive value depicts a concave surface, a negative value depicts a convex surface and a zero value represents a near flat surface. Furthermore, the surface curvature radius of an NP determines the chemical potential (μ) of atoms on an NP surface and is given by the Young–Laplace equation [73]:

$$\Delta\mu = 2\gamma\frac{\Omega}{R} \tag{8}$$

where R denotes the radius of a spherical NP, Ω denotes the volume of the particles, and γ is the surface energy.

Chemical potential determines the crystallographic structure, ionicity of metal–oxygen bonds, and electrostatic potential. The latter is essential for explaining experimentally observed interactions between NPs and the cellular membrane, where the positively charged surface of the NP tends to be attracted to the negatively charged cell membrane surfaces. Meanwhile, induced stress onto the crystallographic structure tends to alter the unit cell parameters and, thus, causes structural changes such as size [74]. Ionicity is inversely proportional to size, but has a direct influence on properties such as chemical reactivity [75]. However, the observed behaviour of metal oxides is linked to the atomic arrangement of atoms on the exposed surface (facet). As a result, a relationship is established between the facets of metal oxides and their chemical potential. As a consequence, using chemical potential as a descriptor aids in model interpretation and demonstrates how facets influence metal oxide toxicity.

The electronegativity count of metal is defined as follows:

$$\varepsilon_{\text{metal}} = -\alpha_{\text{metal}} + \left(0.3 * Z_{\text{metal}}^v\right) \tag{9}$$

where α_{metal} is defined by $\alpha_{\text{metal}} = \lambda * \gamma$; λ is defined by $\lambda = {(Z_{\text{metal}} - Z_{\text{metal}}^v)}/{Z_{\text{metal}}^v}$; $\gamma = {1}/{\text{PN}_{\text{metal}}}$; Z_{metal}^v is the valence electron of metal; Z_{metal} is the atomic number of metal.

Electronegativity was also used as a descriptor in the model. The electronegativity (χ) value for a given metal oxide is strongly related to the electronegativity of the corresponding cation (χ^+). The cation electronegativity depends on the ionic radius and formal charge of the cation; i.e. higher cation electronegativity values characterize cations with a broad charge distribution over a narrow atomic radius., It is evident that increasing cationic electronegativity will enhance the catalytic property (because electronegativity is a measure of an atom's tendency to attract a bonding pair of electrons) and, hence, the toxicity of metal oxide NPs.

Electronegativity scale also gives the potential of a metal oxide to transfer an electron towards a chemical reaction. It is an electronic based property that relates to HOMO and LUMO energies as follows:

$$E_{\text{HOMO}} = (\chi_m{}^a \chi_O{}^b)^{1/n} - E_o + 0.5E_g \tag{10}$$

and

$$E_{\text{LUMO}} = E_{\text{HOMO}} - E_g \tag{11}$$

where E_{HOMO} and E_{LUMO} denote the highest occupied molecular orbitals and lowest unoccupied molecular orbitals' potentials respectively, χ_m and χ_O denote the absolute electronegativity of metal and oxygen atoms respectively [76], a and b are number of metal of oxygen atoms in chemical formula, n is the total number of atoms in the chemical formula, E_o is the standard electrode potential that assumes the value of 4.5 eV on the hydrogen scale, and E_g is the estimated band gap.

Changes in the wave function due to quantum confinement of electrons results in a change of electronic properties for metal oxides. Hence, the electronegativity of the same kind of an atom in different systems is not the same. Moreover, for different facets of the same material, the electron wave function is not the same, thus its electronegativity also. It follows that varied facets of a metal oxide will yield different oxidative and reductive reactions as a result of varied electron properties (Fig. 3). This result was also observed experimentally and reported in the literature [8].

Metal electronegativity count was identified as a descriptor significant for predicting toxicity of metal oxides and is related to the band gap energy (E_g), the LUMO and the HOMO according to Eqs. 10 and 11. Thus, while the model descriptor is the electronegativity with respect to the exposed facet, the band edges can be used to describe the toxicological potential of metal oxides in cellular environments where the aqueous redox potential ranges from -4.12 to -4.84 eV [77]. In our study, HOMO energies for La_2O_3, Sb_2O_3, In_2O_3, TiO_2, and NiO metal oxides fall within the range of aqueous redox potential. These findings show that the apparent toxicity of metal oxides is due to a distinct mechanism involving the ability to transport electrons between the surface of NPs and intracellular redox couples. When a metal oxide is irradiated under UV light, electrons are extracted from the valence band into the conduction band, leaving unoccupied electron states (holes). The holes (h^+) are capable of transferring between biological media and metal oxide, reacting with OH^- and/or H_2O to produce hydroxide radicals ($\bullet OH$). The free electrons may react with O_2 to form superoxide radical anions ($O\bullet{}^-_2$). These reactive oxygen species (ROS) are capable of causing membrane disruptions that can lead to cell death [78–80].
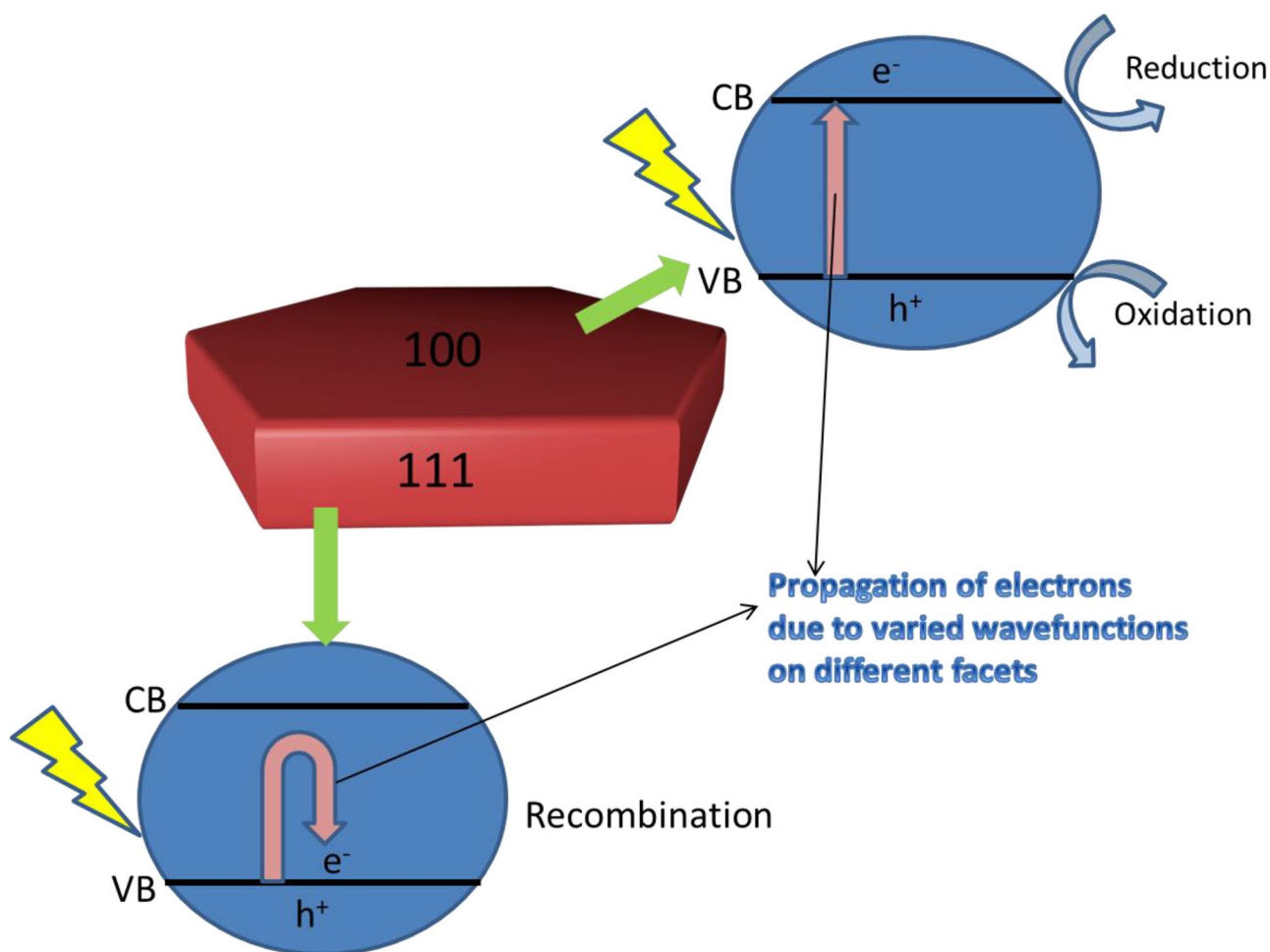
Fig. 3 A faceted metal oxide demonstration with specific behaviours observed on its various facets (surfaces). Some facets are inactive and thus non-responsive, while others are highly reactive. The vari-able electron wavefunction experienced on each facet can be used to explain this occurrence

Conclusion

The study used toxicity data from the literature to develop a (Q)SAR model for the prediction of the toxicity of metal oxide NPs using a combination of physicochemical, molecular, and periodic table-based descriptors. The additional descriptors used in this study were calculated with the aid of DFT on the basis of up to 3 different crystal facets per NM composition (i.e. {100}, {110} and {111}), TEM images of the NMs (extracted from the original publications and processed via the NanoXtract tool). The materials' core size, chemical potential, enthalpy of formation, and electronegativity count of metal oxides were found to be the most significant descriptors of the model. All DFT calculations for the metal oxides were based on low index facets. Metal oxide chemical space was limited to four types of metal oxides (M_3O_4, M_2O_3, MO_2, and MO). A model that is both reliable and basic for theoretical assessment of the toxicity of untested metal oxides, was successfully developed and validated using the OECD guidelines for the validation of (Q)SAR models for regulatory purposes. The model provides an insight into how facet tuning could lead to different degrees of toxicity. Finally, the types of mechanistic pathways that can lead to toxicity have been explained based on electronic properties of the metal oxide NMs. Defined structure–activity relationships in the study could play a vital role in the design of safer nanomaterials.

Availability of data and material Model: https://facetcytotoxicity.cloud.nanosolveit.eu/swagger-ui/); Dataset: https://db.nanopharos.eu/

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Zheng H, Ou J, Strano M (2011) Nanostructured tungsten oxide — properties, synthesis, and applications. Adv Funct Mater 21:2175–2196

2. Marschall R, Wang L (2014) Non-Metal doping of transition metal oxides for visible-light photocatalysis. Catal Today 225:111–135

3. Hua M, Zhang S, Pan B (2012) Heavy metal removal from water/wastewater by nanosized metal oxides: A review. J Hazard Mater 212:317–331

4. IIvask A, Titma T, Visnapuu M, Vija H, Kakinen A, Sihtmae M, Pokhrel S, Madler L, Heinlaan M, Kisand V, Shimmo R (2015) Toxicity of 11 metal oxide nanoparticles to three mammalian cell types in vitro. Curr Top Med Chem 15 1914–1929

5. Chaudhary RG, Bhusari S, Tiple A, Rai A, Somkuvar S, Potbhare A, Lambat T, Ingle P, Abdala A (2019) Metal/metal oxide nanoparticles: toxicity, applications, and future prospects. Curr Pharm Des 25:4013–4029

6. Egambaram OP, Pillai SK, Ray SS (2020) Materials science challenges in skin UV protection : A Review. Photochem Photobiol 96:779–797

7. Ghadimi M, Zangenehtabar S, Homaeigohar S (2020) An overview of the water remediation potential of nanomaterials and their ecotoxicological impacts. Water 12:1150–1173

8. Chang Y, Li K, Feng Y, Liu N, Cheng Y, Sun X, Feng Y, Li X, Wu Z, Zhang H (2016) Crystallographic facet-dependent stress responses by polyhedral lead sulfide nanocrystals and the potential "safe-by-design" approach. Nano Res 9:3812–3827

9. Strigul N, Koutsospyros A, Christodoulatos C (2010) Tungsten speciation and toxicity: acute toxicity of mono- and poly-tungstates to fish. Ecotoxicol Environ Saf 73:164–171

10. Burello E (2017) Review of (Q)SAR models for regulatory assessment of nanomaterials risks. NanoImpact 8:48–58

11. Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O (2016) Applicability domain for QSAR models 1:45–63

12. Choi JS, Trinh TX, Yoon TH, Kim J, Byun HG (2019) Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. Chemosphere 217:243–249

13. Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari T, Michalkova A, Hwang H, Toropov A, Leszczynska D, Leszczynski J (2011) Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nat Nanotechnol 6:175–178

14. Choi JS, Ha MK, Trinh TX, Yoon TH, Byun HG (2018) Towards a generalized toxicity prediction model for oxide nanomaterials using integrated data from different sources. Sci Rep 8:1–10

15. Toropov AA, Toropova AP, Benfenati E, Gini G, Puzyn T, Leszczynska D, Leszczynski J (2012) Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli. Chemosphere 89:1098–1102

16. Toropov AA, Toropov AP, Benfenati E, Korenstein R, Leszczynska D, Leszczynski J (2015) Optimal nano-descriptors as translators of eclectic data into prediction of the cell membrane damage by means of nano metal-oxides. Environ Sci Pollut Res 22:745–757

17. Manganelli S, Leone C, Toropov AA, Toropova AP, Benfenati E (2016) QSAR model for predicting cell viability of human embryonic kidney cells exposed to SiO2 nanoparticles. Chemosphere 144:995–1001

18. Mulwa WM, Ouma CNM, Onani MO, Dejene FB (2016) Energetic, electronic and optical properties of lanthanide doped TiO2: an ab initio LDA+U study. J Solid State Chem 237:129–137

19. Hsieh PL, Lee AT, Chen LJ, Huang MH (2018) Germanium wafers possessing facet-dependent electrical conductivity properties. Angew Chemie - Int Ed 57:16162–16165

20. Di Valentin C, Wang F, Pacchioni G (2013) Tungsten oxide in catalysis and photocatalysis: hints from DFT. Top Catal 56:1404–1419

21. Thwala MM, Dlamini LN (2019) Improved photocatalytic activity of Y-doped WO3 in reduction of Cu(II) in industrial effluent. Int J Environ Sci Technol 16:4887–4898

22. Venigalla S, Dhail S, Ranjan P (2014) Computational study about cytotoxicity of metal oxide nanoparticles invoking nano-Qsar 23:3999

23. Mu Y, Wu F, Zhao Q, Ji R, Qie Y, Zhou Y, Hu Y, Pang C, Hristozov D, Giesy J, Xing B (2016) Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. Nanotoxicology 10:1207–1214

24. Neal AL (2008) What can be inferred from bacterium–nanoparticle interactions about the potential consequences of environmental exposure to nanoparticles? Ecotoxicology 17:362–371

25. Lee K, Kim M, Kim H (2010) Catalytic nanoparticles being facet-controlled. J Mater Chem 20:3791–3798

26. Gajewicz A, Schaeublin N, Rasulev B, Hussain S, Leszczynska D, Puzyn T, Leszczynski J (2015) Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. Nanotoxicology 9:313–325

27. Caixeta MB, Araújo PS, Gonçalves BB, Silva LD, Grano-Maldonado MI, Rocha TL (2020) Toxicity of engineered nanomaterials to aquatic and land snails: a scientometric and systematic review. Chemosphere 260:127654

28. Kaweeteerawat C, Ivask A, Liu R, Zhang H, Chang C, Low-Kam C, Fischer H, Ji Z, Pokhrel S, Cohen Y, Telesca D, Zink J, Mädler L, Holden P, Nel A, Godwin H (2015) Toxicity of metal oxide nanoparticles in Escherichia coli correlates with conduction band and hydration energies. Environ Sci Technol 49:1105–1112

29. Chen ZW, Chen LX, Wen Z, Jiang Q (2019) Understanding electrocatalysis by using density functional theory. 23782–23802

30. Lee Y (2021) Density Functional Theory (DFT) Calculations and catalysis. Catalysis 10–12

31. Ghamali M, Chtita S, Aouidate A, Ghaleb A, Bouachrine M, Lakhlifi T (2017) Combining DFT and QSAR computation to predict the interaction of flavonoids with the GABA (A) receptor using electronic and topological descriptors. Integr Med Res 11:422–433

32. Adegoke RO, Oyebamiji AK, Semire B (2020) Dataset on the DFT-QSAR , and docking approaches for anticancer activities of 1 , 2 , 3-triazole-pyrimidine derivatives against human esophageal carcinoma (EC-109). Data Br 31:105963

33. Dal Corso A, Pasquarello A, Baldereschi A, Car R (2015) Generalized-gradient approximations to density-functional theory: a comparative study for atoms and solids A comparative study for atoms and solids

34. Baseden KA, Tye JW (2014) Introduction to density functional theory: calculations by hand on the helium atom. J Chem Educ 91:2116–2123

35. Delley B (2002) Hardness conserving semilocal pseudopotentials. Phys Rev B 66:155125

36. Perdew JP, Burke K, Ernzerhof M (1997) Generalized gradient approximation made simple. Phys Rev Lett 78:1396

37. Varsou DD, Nikolakopoulos S, Melagraki G, Toumanis A, Afantitis A (2018) Enalos+ KNIME Nodes: New Cheminformatics Tools for Drug Discovery, in Rational Drug Design: Methods and Protocols. Springer, New York

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

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

40. Leach AR (2007) An introduction to chemoinformatics. 2007: Springer Science & Business Media. Springer Netherlands

41. Bishop CM (2007) Pattern recognition and machine learning (information science and statistics). Springer New York, NY, USA

42. 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 2:249–262

43. Hall M, Eibe Frank G, Holmes B, Pfahringer P, Reutemann IW (2009) The WEKA data mining software: an update. SIGKDD Explor Newsl 11(1):10–18

44. Papadiamantis AG, Jänes J, Voyiatzis E, Sikk L, Burk J, Burk P, Tsoumanis A, Kieu Ha M, Hyun Yoon T, Valsami-Jones E, Melagraki G, Tämm K, Afantitis A (2020) Predicting cytotoxicity of metal oxide nanoparticles using isalos analytics platform. Nanomaterials 10:2017

45. Witten IH, Hall M, Frank E (2016) Data mining: practical machine learning tools and techniques. 2016: Morgan Kaufmann

46. Varsou D, Afantitis A, Tsoumanis A, Melagraki G, Sarimveis H, Valsami-Jones E, Lynch I (2019) A safe-by-design tool for functionalised nanomaterials through the Enalos Nanoinformatics Cloud platform. Nanoscale Adv 1:706–718

47. Larose DT, Larose C (2014) Discovering knowledge in data: an introduction to data mining. John Wiley & Sons

48. Papadiamantis AG, Tsoumanis A, Melagraki G, Valsami-Jones E, Lynch I (2021) Computational enrichment of physicochemical data for the development of a ζ -potential read-across predictive model with Isalos Analytics Platform. NanoImpact 22:100308

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

50. Directorate E (2007) Environment health and safety publications series on testing and assessment No. 69, Guidance document on the validation of (quantitative) structure-activity relationships [(Q) SAR] models

51. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partialcredit. Psychol Bull 70:220–231

52. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta - Protein Struct 405:442–451

53. Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE 12:1–17

54. Chicco D, Tötsch N, Jurman G (2021) The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Min 14:1–22

55. Afantitis A, Melagraki G, Tsoumanis A, Valsami-Jones E, Lynch I (2018) A nanoinformatics decision support tool for the virtual screening of gold nanoparticle cellular association using protein corona fingerprints. Nanotoxicology 12:1148–1165

56. Melagraki G, Afantitis A (2015) A risk assessment tool for the virtual screening of metal oxide nanoparticles through enalos insiliconano platform. Curr Top Med Chem 15:1827–1836

57. Varsou D, Afantitis A, Tsoumanis A, Melagraki G, Sarimveis H, Valsami-Jones E, Lynch I (2019) A safe-by-design tool for functionalised nanomaterials through the Enalos Nanoinformatics. Nanoscale Adv 1:706–718

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

59. Bonaretti S, Willighagen E (2019) Two real use cases of FAIR maturity indicators in the life sciences. BioRxiv 1:739334

60. Papadiamantis AG, Klaessig FC, Exner TE, Hofer S, Hofstaetter N, Himly M, Williams MA, Doganis P, Hoover MD, Afantitis A, Melagraki G (2020) Metadata stewardship in nanosafety research community-driven organisation of metadata schemas to support FAIR nanoscience data

61. Portier J, Hilal HS, Saadeddin I, Hwang SJ, Subramanian MA, Campet G (2004) Thermodynamic correlations and band gap calculations in metal oxides. Prog. Solid State Chem 32:207e217

62. Li Q, Hu H, Jiang L, Zou Y, Duan J, Sun Z (2016) Cytotoxicity and autophagy dysfunction induced by different sizes of silica particles in human bronchial epithelial BEAS-2B cells. Toxicol Res 5:1216–1228

63. Ali SA, Rizk MZ, Hamed MA, Aboul-Ela EI, El-Rigal NS, Aly HF, Abdel-Hamid A (2019) Assessment of titanium dioxide nanoparticles toxicity via oral exposure in mice: Effect of dose and particle size. Biomarkers 24:492–498

64. Loan TT, Do LT, Yoo H (2018) Platinum nanoparticles induce apoptosis on raw 264.7 macrophage cells. J Nanosci Nanotechnol 18:861–864

65. Karlsson HL, Gustafsson J, Cronholm P, Möller L (2009) Size-dependent toxicity of metal oxide particles — a comparison between nano- and micrometer size. 188:112–118

66. Warheit DB, Webb TR, Sayes CM, Colvin VL, Reed KL (2006) Pulmonary instillation studies with nanoscale TiO2 rods and dots

in rats: toxicity is not dependent upon particle size and surface area. Toxicol Sci 91(1):227–36

67. Liu N, Li K, Li X, Chang Y, Feng Y, Sun X, Cheng Y, Wu Z, Zhang H (2016) Crystallographic facet-induced toxicological responses by faceted titanium dioxide nanocrystals

68. Feng Y, Chang Y, Sun X, Liu N, Cheng Y, Feng Y, Zhang H, Li X (2017) Understanding the property–activity relationships of polyhedral cuprous oxide nanocrystals in terms of reactive crystallographic facets. 156:480–491

69. Schulz KH, Cox DF (1991) Photoemission and lowenergy-electron-diffraction study of clean and oxygen-dosed Cu$_2$O (111) and (100) surfaces. Phys Rev B 43:1610–1621

70. Brown Jr GE, Henrich V, Casey W, Clark D, Eggleston C, Andrew Felmy AF, Goodman (1999) Metal oxide surfaces and their interactions with aqueous solutions and microbial organisms. Chem Rev 99:77–174

71. Shang Y, Guo L (2015) Facet-controlled synthetic strategy of Cu$_2$O-based crystals for catalysis and sensing

72. Deml AM, Muhich CL, Musgrave CB (2014) Oxide enthalpy of formation and band gap energy formation energetics †. Environ Sci 7:1996–2004

73. Cao G (2004) Nanostructures & nanomaterials: synthesis, properties & applications. Imperial college press

74. Zhang H, Banfield JF (1998) Thermodynamic analysis of phase stability of nanocrystalline titania. J Mater Chem 8:2073–2076

75. McHale JM, Auroux A, Perrotta AJ, Navrotsky A (1997) Surface energies and thermodynamic phase stability in nanocrystalline aluminas. Science 277(80):788–791

76. Parr R, Pearson R (1983) Absolute hardness: companion parameter to absolute electronegativity. J Am Chem Soc 105:7512–7516

77. Zhang H, Ji Z, Xia T, Meng H, Low-Kam C, Liu R, Pokhrel S, Lin S, Wang X, Liao YP (2012) Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. ACS Nano 6:4349–4368

78. Khanna P, Ong C, Bay B, Baeg G (2015) Nanotoxicity: an interplay of oxidative stress, inflammation and cell death. Nanomaterials 5:1163–1180

79. Dayem AA, Hossain MK, Lee SB, Kim K, Saha SK, Yang GM, Choi HY, Cho SG (2017) The role of reactive oxygen species (ROS) in the biological activities of metallic nanoparticles. 1–21

80. Manke A, Wang L, Rojanasakul Y (2013) Mechanisms of nanoparticle-induced oxidative stress and toxicity