Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: Nano-read-across

Journal:	*Nanotechnology*
Manuscript ID:	Draft
Manuscript Type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Gajewicz, Agnieszka; University of Gdansk, Faculty of Chemistry Cronin, Mark; School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Rasulev, Bakhtiyor; Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, Leszczynski, Jerzy; Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, Puzyn, Tomasz; University of Gdansk, Faculty of Chemistry
Article Keywords:	nanoparticles, nano-QSAR, nano-read-across, cytotoxicity
Abstract:	Creating suitable chemical categories and developing read-across methods, supported by quantum-mechanical calculations, can be an effective solution solving key problems related to current scarcity of data on the toxicity of various nanoparticles. This study has demonstrated that by applying a nano-read-across, the cytotoxicity of nano-sized metal oxides could be estimated with a similar level of accuracy as provided by nano-QSAR model(s). The method presented is a suitable computational tool for the preliminary hazard assessment of nanomaterials. It also could be used for the identification of nanomaterials that may pose potential negative impact to the human health and the environment. Such approaches are especially necessary, when there is paucity of relevant and reliable data points to develop and validate nano-QSAR models.

Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: Nano-read-across

*Agnieszka Gajewicz[1], Mark T.D Cronin[2], Bakhtiyor Rasulev[3], Jerzy Leszczynski[3] and Tomasz Puzyn[1]**

[1] *Laboratory of Environmental Chemometrics, Institute for Environmental and Human Health Protection, Faculty of Chemistry, University of Gdansk, Gdansk, Poland*

[2] *School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK*

[3] *Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, Jackson MS, USA;*

Correspondence: T. Puzyn, Laboratory of Environmental Chemometrics, Department of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland,

phone: (+48 58) 523 52 48; fax: (+48 58) 523 50 12; e-mail: t.puzyn@qsar.eu.org

Abstract

Creating suitable chemical categories and developing read-across methods, supported by quantum-mechanical calculations, can be an effective solution solving key problems related to current scarcity of data on the toxicity of various nanoparticles. This study has demonstrated that by applying a nano-read-across, the cytotoxicity of nano-sized metal oxides could be estimated with a similar level of accuracy as provided by nano-QSAR model(s). The method presented is a suitable computational tool for the preliminary hazard assessment of nanomaterials. It also could be used for the identification of nanomaterials that may pose potential negative impact to the human health and the environment. Such approaches are especially necessary, when there is paucity of relevant and reliable data points to develop and validate nano-QSAR models.

Keywords: nanoparticles, nano-QSAR, nano-read-across, cytotoxicity

1. Introduction

There is a large number of engineered nanomaterials of varying structure as well as those of the same chemical formula that differs only in terms of their physico-chemical properties. The differences in the physico-chemical properties of nanoparticles often determine the changes in their hazardous properties [1]. Unfortunately, it is still largely unknown, which properties determine and/or influence the toxicity of particular nanoparticles [2-4]. Therefore, effort should be placed into defining and developing methods for characterization, exposure, engineering control, potential toxicity, fate and transport, and life cycle assessment of engineered nanoparticles [5]. This leads to an increase in testing, and thus animal use, when traditional human health and the environment hazard assessments are conducted for all different nanomaterials. According to the recent report of the European Commission Joint Research Centre approximately 3.9 million additional test animals could potentially be used as a consequence of the introduction of REACH (Registration, Evaluation, Authorisation and Restriction of CHemicals) program [6] if the use of alternative methods would not be available. However, if these techniques would be applied in a maximal extend, a reduction in animal use could be obtained with potential savings of 1.9 million test animals [7]. Therefore, with regard to the ethical aspects, and following the recommendations by the REACH system and new European Directive 2010/63/EU [8], testing should be performed by means other than vertebrate animal tests, whenever possible. This requirement specifically encourages the use of alternative approaches, for example, *in vitro* methods or qualitative or quantitative structure-activity and structure-property relationship models

(QSAR/QSPR), or from information arising from the grouping of structurally related substances (chemical category formation leading to read-across) [6].

The concept of using (Q)SAR methodologies in the risk assessment of engineered nanomaterials has been discussed extensively for at least five years by many national and international bodies (e.g., World Health Organization (WHO), Organization for Economic Co-operation and Development (OECD), European Commission (REACH, NanoSafety Cluster), European Food Safety Authority (EFSA), Scientific Committee on Consumer Products (SCCP), Environmental Protection Agency (EPA), etc.). (Q)SAR/QSPR approaches are based on defining mathematical dependencies between the variation in molecular structures (encoded by so-called molecular descriptors) and a given physico-chemical or biological property (so-called endpoint). It is usually done for a series of often related compounds - the dataset. These approaches have been applied successfully to predict the toxicity and selected physico-chemical properties of a number of different types of nanoparticles [9-15]. However, since (Q)SAR/QSPR algorithms are based on various statistical/probabilistic approaches, sufficiently large data set must be obtained for those methods to be functional [16]. On the other hand, engineered nanoparticles represent very structurally diverse groups of chemicals (organic, inorganic, metal, carbon-based nanoparticles etc.). Thus, it is difficult to build a significant dataset of structurally related nanoparticles. In addition, since nanoparticles are not structurally homogenous, a common mechanism of toxicity (which is a fundamental requirement for modeling) cannot be expected for all of them. As a consequence, toxicity and other properties should be assessed within specific applicability domains, i.e. groups of

sufficiently similar nanoparticles that can be expected to have similar properties. Examples of such groups include metal oxides, various modifications of a single metal oxide, carbon nanostructures etc. Furthermore, systematic data available in the literature for specific groups of nanomaterials are usually of limited use in the context of (Q)SAR/QSPR modeling and risk assessment purposes. Very often, even if the quality of the measurements is very high, the number of data points (i.e., given property measured for a series of differing nanoparticles) is insufficient to allow for the development of robust models [17].

In the absence of relevant, reliable and sufficient data to build an appropriately validated (Q)SAR/QSPR model, the read-across approach is an attractive and pragmatic technique to fill data gaps [6]. The read-across has been successfully applied to predict many properties and toxicity (i.e., teratogenicity, mutagenicity, skin and respiratory tract sensitization effects, fish acute toxicity, etc.) of compounds, such as carbonyl compounds, phospho-organic pesticides and polar organic compounds [18-20]. However, so far this computational technique has not been used to fill data gaps for nanoparticles.

The focus of this work is to provide an efficient methodology that will eliminate shortcomings of the dataset scarcity for nanomaterials. Our approach provides a capable tool allowing for predictions of various properties (physico-chemical as well as toxicological) of unknown nanomaterials based on information extracted from very few known species. It opens new opportunities to evaluate their properties without necessity of performing expensive experimental studies on large pool of nanomaterials.

2. Read-across - general approach

The read-across concept has received much attention in recent years since is a non-testing approach that can be used for data gap filling. In principle, read-across can be used to predict endpoint information for one, or more, chemical(s) (the so-called "target chemical(s)") by using data from the same endpoint from another substance(s) (the "source chemical(s)"), which are considered to be 'similar' in some way. Similarity is usually based on structural and/or physico-chemical properties of new molecules [21]. The similarities may be based on common functional groups, common constituents or the likelihood of common precursors and/or breakdown products [21]. In other words, the known activity of one substance could be used to estimate the unknown value, for the same activity, for another substance if they have a similar chemical structure. According to OECD Guidance [21] read-across may be performed in a qualitative or quantitative manner, depending on the data used (discrete or numerical). If the presence (or absence) of a property/activity for the source chemical(s) is expressed on a discrete scale, then qualitative read-across is used to obtain a 'yes/no' answer for the presence (or absence) of the same property/activity for one or more target chemical(s). Whereas, quantitative read-across is used to predict the unknown value of the property/activity for the target chemical(s) based on the known value(s) of the same property/activity for source chemical(s) expressed on a numerical scale. Furthermore, as already mentioned, read-across may be performed between two chemicals or for a group of chemicals in one of four ways shown schematically in Figure 1.

[INSERT FIGURE 1]

Regardless of whether the property/activity is expressed in a quantitative or qualitative manner, the procedure of read-across involves two following steps: the identification of a structural feature that is common to two substances or a group of substances (which are considered to be similar) and the assumption that the known value of a physico-chemical property, toxicological effect or environmental fate property for one chemical can be used to assess the unknown value of the same property or activity for another chemical. However, as already mentioned, read-across approach has not been yet applied to fill data gaps for nanoparticles.

3. Materials and methods

The methodology applied in this study involves the following steps: (i) exploration of the multidimensional space of calculated molecular descriptors in order to obtain the preliminary information on structural similarities and dissimilarities on the studied datasets; (ii) grouping of nanoparticles with pattern recognition techniques based on structural features; (iii) performing nano-read-across analysis in a variant of the many-to-many approach.

Metal oxide nanoparticles were selected as the target group for the study. These compounds are of the highest priority due to the fact that metal oxide nanoparticles are commonly used in nanotechnology. Cytotoxicity data (the concentration of metal oxide nanoparticles that reduces bacteria viability of 50%,

EC_{50}) for 17 nano-sized metal oxides to the bacterium *Escherichia coli* were taken from our previous study [9]. Whereas the toxicity data (the concentration of metal oxide nanoparticles that caused a 50% reduction of the cells after 24h exposure, LC_{50}) for 18 nano-metal oxides to a human keratinocyte (HaCaT) cell line were taken from [15]. To make the current study directly comparable with the results obtained from the previously described nano-QSAR models [9, 15], we utilized the same method to split the data into training and validation sets (for more details, please refer to the Electronic Supplementary Material, Table S3 and Table S9). The training set was later used to identify of MeOx or a group of MeOx(s) (which are considered to be similar) and an external validation set to evaluate the predictive ability of the nano-read-across models. Finally, the models were applied to predict the cytotoxicity towards bacteria *E. coli* and human keratinocyte (HaCaT) cell line respectively for untested metal oxide nanoparticles from the selected groups.

The structural characteristics that quantitatively describe the variation of the nanoparticles' structure (structural descriptors) also were taken from our previous studies [9, 15]. Detailed lists of the descriptors calculated on the basis of small, stoichiometric clusters, reflecting all the characteristics of fragments of the crystal structures of particular oxides are provided in the Electronic Supplementary Material (Table S1 and Table S7). To ensure that the influence of each variable was equivalent (i.e. the same scale and range of all variables), the descriptors were standardized, which means that the average value of the descriptor for the set of compounds considered was substracted from the descriptors and the resultant values divided by the standard deviation, according to the formula:

$$z_i = \frac{x_i - \bar{x}_j}{s_j} \quad (1)$$

where:

z_i - is the transformed value of a given variable,

x_i - is the original value of a given variable,

$\bar{x}_j$ - is the mean value of a given variable calculated across a group of all studied compounds,

s_j - is the standard deviation of a given variable calculated across a group of all studied compounds.

Two-dimensional hierarchical cluster analysis (t-HCA) was employed to search for similarities between the nanoparticles in the feature space, assuming that two NPs are similar when located close (i.e., in terms of a measure of distance) to each other [22, 23]. The similarities (and dissimilarities) of the MeOx nanoparticles studied were explored using the multidimensional space of the calculated molecular descriptors and experimentally measured cytotoxicity to bacteria *E. coli* and human keratinocyte (HaCaT) cell line, respectively. Two or more objects may form clusters indicating groups of NPs having similar property and, thus, similar biological activity/toxicity. An important advantage of applying such methods is that it enables simultaneously analyze similarities of compounds and variables, being able to identify categories (classes) of NPs. These classes may be used to predict biological activity/toxicity of other NPs that could be assigned to the same family, based on its structural similarity to other members. Two-way hierarchical cluster analysis was conducted using Euclidean distance as similarity measure and Ward's method of

linkage. The Euclidean distance, defined as the geometric distance in the multidimensional space, was computed according to the formula:

$$d_{(ij)}=\sqrt{\sum_{k=1}^{p}(x_{(i)k}-x_{(j)k})^2} \quad (2)$$

where $x_{(i)k}$, $x_{(j)k}$ are *k*-coordinate values for *i* and *j* object respectively.

Ward's method of clustering (also known as the Ward's minimum variance) is based on the inner squared distance of clusters, so that at each stage these two clusters are grouped together, for which the minimum increase in the total within a group error sums of squares is observed [24]. The final number of clusters was defined based on Sneath's criterion of distance measure relation, i.e. $1/2D_{max}$, where D_{max} is the maximal distance in the similarity matrix [25]. The Sneath index of cluster significance implies that only clusters remaining compact after breaking the linkage at $1/2D_{max}$ are considered significant and should be interpreted.

On the basis of these clusters, the MeOx nanoparticles from the training set were divided into groups with similar cytotoxicity profiles. After the identification of the chemicals considered to be analogues, the information on cytotoxicity of MeOx available for members of each class was utilized to estimate the cytotoxicity for the compounds from the validation set. Therefore, the data from validation set were rescaled, using the mean and standard deviation values from the predefined similarity classes, in order to incorporate them into the classes.

Finally, it was hypothesized that the cytotoxicity of metal oxide nanoparticles to bacteria *E. coli* and human keratinocyte (HaCaT) cell line obtained from the nano-read-across techniques should not be substantially different than the values obtained experimentally, as well as those predicted from the nano-QSAR models. To verify whether the hypothesis and conclusions can be extended to other properties and groups of nanoparticles, a comparison of predictive power was performed for both techniques. Since the method proposed in this study to fill data gaps is qualitative rather than quantitative, the non-parametric Spearman rank correlation test [26] was applied as a measure of the strength of the relationship between the two variables. Spearman's rho (ρ_S), based on rank orderings, describes whether two variables are correlated and is defined as (eq. 3):

$$\rho_S = 1 - \frac{6\Sigma d_i^2}{N(N^2 - 1)} \quad (3)$$

where:

ρ_S – is Spearman's rank correlation coefficient;

d^2 - is the square of the difference between ranks;

N - is the number of data pairs.

Therefore, to verify whether the predictions from the nano-read-across technique differ significantly from the experimental values of $\log(EC_{50})^{-1}$ as well as those predicted from the nano-QSAR modeling, the Spearman's rank correlation coefficient was calculated as an alternative to Pearson's correlation coefficient. All calculations within this study were carried out with the Statistics Toolbox for MATLAB v. 7.6.0.324 [27].

4. Results

4.1. Case study 1

Since the general concept of read-across methods is based on the assumption that nanoparticles with similarities in their chemical structure will give similar toxic responses, we decided to identify those structural features that can be related to the toxicity of metal oxide nanoparticles to bacteria *E. coli*.

The starting point for estimating the mutual similarity of MeOx in terms of the toxicity of metal oxide nanoparticles to bacteria *E. coli* was a matrix of 11 structural descriptors (Table S2). Selection of independent variable(s) that define the similarity of MeOx was conducted based on the value of the Perason's correlation coefficient calculated between the matrix of all descriptors (**X**) and the vector of the dependent variable (**y**) for training set. The size of the training set for the identification of structure-based categories (i.e. developing a model) was n=10. In contrast, the size of the test set for carrying out the external validation and estimating the error of classification for novel NPs (not previously utilized for training the model) was k=7. To eliminate redundancy in the structural data, we utilized for further modeling only those descriptors which had been found by us to contribute significantly to understanding the mechanism of toxicity and have a Pearson correlation coefficient with the endpoint (i.e. cytotoxicity) with an absolute value greater than 0.8. It should be mentioned, however, that sometimes a model can include several descriptors which may have a low individual correlation with activity, but in combination provide a good model. Nevertheless, we generally assume that the proposed nano-read-across

approach should be maximally simplified and based on minimal number of utilized descriptor(s). Finally, we selected only one molecular descriptor: the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure (ΔH_{Me+}). The correlation coefficient for the ΔH_{Me+} (describing ionization enthalpy of the (detached) metal atoms) with the toxicity was -0.92. This means the descriptor ΔH_{Me+} can explain approximately 85% of the variability of the toxicity data for the oxides (R^2=0.85).

The selected descriptor i.e., enthalpy of formation of the metal cation in the gas phase, has been used to define groups of similar chemical species for the read-across of *E. coli* toxicity. The Hierarchical Cluster Analysis (t-HCA) was performed using the data matrix, where the first column corresponded to the experimentally determined cytotoxicity of MeOx to *E. coli,* and the second represented the theoretically calculated values of enthalpy of formation of the metal cation in the gas phase, for all individual metal oxides from the training set. As a result, a dendrogram - a tree-like graphic, shown graphically in Figure 2, was derived. This dendrogram displays the linkages between the clustered objects with respect to their similarity.

[INSERT FIGURE 2]

Based on the structural similarities derived from the selected quantum-mechanical descriptor of the compounds studied, three classes (groups) of metal nano-oxides were identified. The classes consist of compounds with similar values of

cytotoxicity. The first class contained the following oxides: ZnO, CuO and Y_2O_3, second: Bi_2O_3, In_2O_3, Al_2O_3 and Fe_2O_3, while the third group includes: SiO_2, SnO_2 and TiO_2. By analyzing the results obtained, it can be seen that compounds from class I seem to have the highest toxicity, whereas the majority of compounds in class II and all compounds from class III qualify as being non-toxic. The observed trends are in accordance with the results of previous studies by other groups. For instance, Heinlaan et al. [28] observed that the toxicity rank order for ZnO, TiO_2, and CuO to bacteria *Vibrio fisheri* and to crustaceans *D. magna* and *Thamnocephalus platyurus* was as follows: ZnO > CuO > TiO_2.

In the next step the predictive ability of the nano-read-across model has been evaluated. The set of seven metal oxides (i.e., CoO, Cr_2O_3, La_2O_3, NiO, Sb_2O_3, V_2O_3, ZrO_2) from the validation set, which have not been previously applied for classification was used for this task. The ΔH_{Me+} values for the compounds in the validation set were rescaled in order to incorporate them into the previously identified classes, which contained MeOx with similar values of cytotoxicity (Table S4). Based on information on the cytotoxicity of MeOx available for training set members of each class, as well as on the rescaled values of the selected descriptor, the cytotoxicity of each metal oxide in the validation set was estimated. NiO and CoO were assigned to the first class, while the rest of the metal oxides (i.e., Cr_2O_3, La_2O_3, Sb_2O_3, V_2O_3 and ZrO_2) were included in the second class, as determined by the limit values of $\log(EC_{50})^{-1}$, which were respectively 2.29 and 2.82 log unit [mol/dm^3] (Table S4).

The final stage of the evaluation of the accuracy of the data estimation was to perform two-dimensional hierarchical cluster analysis for all 17 metal oxide

nanoparticles from the training and validation sets. The result obtained is the dendrogram, which is shown graphically in Figure 3.

[INSERT FIGURE 3]

The dendrogram confirms that the distribution of the validation set of oxides into the toxicity classes, based on the rescaled descriptor values, is highly consistent with the class assignation we obtained by performing the t-HCA analysis on the experimental data for the training and validation sets. It should be also highlighted that based on the rescaled value of ΔH_{Me+} relevant predictions were made even for CoO - the metal oxide with toxicity ($\log(EC_{50})^{-1} = 3.51$) slightly exceeding the cytotoxicity range covered by the training compounds ($3.45 > \log(EC_{50})^{-1} > 1.74$). Thus the validation stage confirmed the significance of our nano-read-across model. After positive validation, the nano-read-across model was applied to estimate the cytotoxicity to *E. coli* for 19 untested metal oxide nanoparticles. The predicted results (Figure 4) suggest that toxicity of 19 metal oxides increases in the following order: $GeO_2 < Ga_2O_3 < Tl_2O_3 < Au_2O_3 < Yb_2O_3 \approx Mn_2O_3 < Er_2O_3 < Ho_2O_3 < Eu_2O_3 < Tb_2O_3 < Gd_2O_3 < Sm_2O_3 < Nd_2O_3 < Ag_2O_3 < FeO < AuO < MnO < MgO < PbO$.

[INSERT FIGURE 4]

In summary, the results obtained (for compounds from training and validation sets) that are presented in Figure 4 indicate a very good agreement with both the

experimental data and previously predicted with nano-QSAR model results.

A more detailed comparison between the classes provides interesting information. Figure 4 shows how the cytotoxicity of MeOx to *E. coli* varies with the oxidation state of the metal for each class. Metal cations are formed by the initial loss of electron(s) and many metals can form cations in several oxidation states. Since cations of the metals in lower oxidation states (+2) are much less chemically stable than the corresponding cation of the metal in the +4 oxidation state, thus the release of metal cations having a smaller charge (*n*) is energetically more favorable than the release of cations with larger *n*. This explains why the metal oxides containing a cation with lower oxidation numbers (i.e., +2) exhibit the highest cytotoxicity to *E. coli*, whereas the nano-sized metal oxides in the +4 oxidation state of the metal - not. The above observation sheds light on the mechanism of cytotoxicity to bacteria *E. coli*. However, since to the identification of classes of similar nanoparticles based on the structural features used the same quantum-mechanical descriptor that has been utilized to develop the final nano-QSAR model (i.e., enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure), its interpretation is consistent with the mechanism of cytotoxicity previously discussed by Puzyn et al. [9]. As already demonstrated, the cytotoxicity of metal oxides nanoparticles to *E. coli* is associated with the process of metal cations release from the particle surface and decreases in the following order: $Me^{2+} > Me^{3+} > Me^{4+}$. For more details, please refer to Puzyn et al. [9].

Predictions are always less accurate than the experimental data. However, the results obtained from: (1) nano-QSAR modeling and (2) nano-read-across techniques

estimation that are summarized in Figure 4, provide mechanistic interpretation of the investigated toxicity and indicate an almost identical order of assignment of the individual metal oxides (from validation and prediction sets) to particular classes of similar toxicity profile. Moreover, we found that the Spearman's rank correlation coefficient calculated between experimentally measured and estimated with nano-read-across approach values of $\log(EC_{50})^{-1}$ is $\rho_S = 0.955$ (Table 1). Since the absolute value of ρ_S was greater than the critical $\rho_{S(\alpha\ 0.05)}$, and the p-value was less than the significance level of 5% (p=0.0001), thus the null hypothesis has been rejected and may be concluded that there is a rank order relationship between the variables.

[INSERT TABLE 1]

Additionally, the relationship between the cytotoxicity of metal oxide nanoparticles to *E. coli* obtained from the nano-QSAR model and estimated with nano-read-across technique was found to be statistically significant. Since our calculated value of $\rho_S = 0.961$ exceeds the critical value and the p-value was 0.0001, we concluded that the correlation is considered to be statistically significant at the 95% probability level (Table 1). For more details, please refer to Electronic Supplementary Material (Table S5 and Table S6). The results obtained support the research hypothesis that there is a rank order relationship between the predictions from nano-read-across technique and experimentally measured (or predicted from the nano-QSAR model) values of $\log(EC_{50})^{-1}$.

4.2. Case study 2

The goal of this study was to identify similar groups of metal oxide nanoparticles, to find relationships between the feature/property of NPs, to detect factors responsible for their cytotoxicity to the human keratinocyte cell line as well as to reveal discriminating parameters which determine the classification of the MeOx in different groups of similarity (or dissimilarity). To explore the data set studied and to examine the similarities of the MeOx nanoparticles, the hierarchical clustering method was used. The results presented below are based on the Euclidean distance as the similarity measure, Ward's linkage algorithm, z-transformation of data and check of the cluster significance by the Sneath index. The data set studied presents 26 of structural descriptors (including 15 descriptors derived from quantum-mechanical calculations and 11 descriptors derived from TEM image analysis) and toxicity data to a human keratinocyte (HaCaT) cell line in term of LC_{50} for 18 different metal oxides nanoparticles. The calculated/measured parameters are listed in Table S8 and Table S9.

To define structure-based categories of MeOx the Mulliken's electronegativity of the cluster (χ^C) has been selected by employing Pearson correlation coefficient. The correlation coefficient for the χ^C with the toxicity was 0.81. This means χ^C can explain approximately 66% of the variability of the toxicity the metal oxide nanoparticles.

The dendrogram formed for the selected variable (Figure 5) reveals three main classes of metal nano-oxides that consist of compounds with similar values of cytotoxicity. The first class includes: ZnO, CoO and In_2O_3, second class contains: Bi_2O_3, Mn_2O_3 and Sb_2O_3, whereas the third class is: ZrO_2, TiO_2, SiO_2, V_2O_3. By

analyzing the results obtained, we found that the cytotoxicity of MeOx systematically decreases when moving from class I to class III.

[INSERT FIGURE 5]

In this way, we obtained a classification model capable of assigning novel compounds to an appropriate class, based on the values of selected descriptor. Since according to OECD recommendations mechanistic interpretation is vital for validation, in the next step the predictive ability of the developed nano-read-across model has been evaluated with the set of eight metal oxides (i.e., Al_2O_3, Cr_2O_3, Fe_2O_3, La_2O_3, NiO, SnO_2, WO_3, Y_2O_3) from the validation set. Based on information on the cytotoxicity of MeOx available for training set members of each class, as well as on the rescaled values of the selected descriptor, we have estimated cytotoxicity of each metal oxide in the validation set. We found that La_2O_3 and WO_3 have been assigned to the first class, Cr_2O_3, NiO and SnO_2 to the second class, while the rest of the metal oxides (i.e., Y_2O_3, Al_2O_3 and Fe_2O) were included in the third class. We obtained relevant estimation of the majority of MeOx that was confirmed by performing two-dimensional hierarchical cluster analysis for all metal oxide nanoparticles from both: training and validation sets (Figure 6).

[INSERT FIGURE 6]

Interestingly, there were false negative predictions for three oxides. The nano-

read-across model failed to predict the high toxic effect to human keratinocytes cell line for the two oxides, namely: SnO_2 and Mn_2O_3. Additionally V_2O_3 was wrongly assigned to the non-toxic class instead of to the weak-toxic class. One possible explanation is that all of these oxides are extremely close to the lower limit of the appropriate toxicity class (i. e. weak-toxic and toxic class respectively for (i) V_2O_3 and (ii) MnO_2 and SnO_2). Thus this might be a reason why they were not properly assigned to the adequate class of toxicity. Therefore in consequence MnO_2 and SnO_2 were introduced to the weak-toxicity class instead of toxic class whereas V_2O_3 was assigned to non-toxic class in spite it being classified previously as a weak-toxic MeOx based on experimentally measured $\log(LC_{50})^{-1}$ values.

Finally, we applied the nano-read-across model to estimate the toxicity to the human keratinocyte cell line for seven - so far experimentally untested - metal oxide nanoparticles. Based only on structural similarity (i.e. rescaled values of selected descriptor), we assigned FeO, PbO, PbO_2 and Gd_2O_3 to the first class determined by the limit values of $\log(LC_{50})^{-1}$, higher than 2.51 log unit [mol/dm^3] (Table S10). CuO was introduced to the second class determined by the limit values of $\log(LC_{50})^{-1}$=2.21 - 2.50. Whereas, MnO and MoO_3 were included to the third class with the upper experimentally measured limit value of $\log(LC_{50})^{-1}$ equal 2.21 log unit [mol/dm^3] (Table S10).

One more important note, the nano-read-across model obtained utilizes one of the two descriptors (i.e. Mulliken's electronegativity of the cluster) that previously were used for nano-QSAR model development [15]. As we have already demonstrated in [15] electronegativity (χ), corresponding to the Fermi level, falls within the middle

of the forbidden gap region and mainly depends on the formal charge of the cation and ionic radius (eqs. 4-6):

$$\chi(P.u.) \approx 0.274z - 0.15zr - 0.01r + 1 + \alpha \quad (4)$$

$$\chi_{cation}(eV) \approx \frac{(\chi_{cation}(P.u.) + 0.206)}{0.336} \quad (5)$$

$$\chi_{cation}(eV) \approx 0.45\chi_{cation}(eV) + 3.36 \quad (6)$$

where: z is the formal charge of the cation, r is the Shannon ionic radius and α is a correcting term specific for each cation.

Conversely, following Portier et al. [29] it should be highlighted that, even if the formal charge is large, but it is distributed over a sizeable cationic volume, no one should expect a high value of the cation electronegativity. Thus, in the context of the Haber-Weiss-Fenton cycle, mechanistic interpretation of the developed nano-read-across model was intuitive: the increase of the cation electronegativity should result in the increase of catalytic properties of metal cations and, consequently, increase the toxicity of the metal oxide nanoparticle. For more details please refer to Gajewicz et al. [15].

By employing the Spearman rank correlation test we confirmed that the values obtained from the nano-read-across technique did not differ significantly from those measured experimentally ($\rho_S = 0.732$), as well as those predicted from the nano-QSAR model ($\rho_S = 0.866$). Since the absolute values of ρ_S were greater than the critical $\rho_{S(\alpha\ 0.05)}$, and the p-values were less than the significance level of 5%, we concluded that

the present method is sufficiently accurate to fill data gaps (Table 1). For more details, please refer to Electronic Supplementary Material (Table S11 and Table S12).

It should be also highlighted that, from an economical point of view, both computational methods (i.e. nano-QSAR and nano-read-across techniques) are acceptable, since they require a relatively small number of experimental data. In fact, both are based on data from the literature, thus performing of any extensive empirical work was unnecessary. However, the use of the nano-read-across approach seems to be much more profitable, as it enables the prediction for those groups of NPs, for which the number of experimental data is insufficient to develop appropriate nano-QSAR model(s). For example, if the experimentally determined data for biological activity are available only for 8-10 NPs, there are too few to calibrate and validate a nano-QSAR model. However, it is sufficient to estimate the unknown value of endpoint information for one, or more, chemical(s) by employing the nano-read-across approach. This makes the proposed method a very useful computational tool, especially when the number of data points to develop and validate nano-QSAR model is inadequate. We hope that the proposed method will soon be verified by experimental and additional theoretical studies.

5. Conclusions

In conclusion, it can be stated that although two-way hierarchical clustering is usually applied at the first stage of data exploration, it can lead to many valuable observations and conclusions. Applied to the data reflecting properties and toxicity of nanoparticles, it allows for the classification of MeOx NPs according to their potential

toxic effects as well as for the identification of factors responsible for their toxicity. The proposed method is the first attempt to use a two-dimensional hierarchical cluster analysis to identify classes of similar nanoparticles based on the structural features and then use them to estimate the biological activity for empirically untested metal oxide nanomaterials. Based on structural similarity, we have estimated the cytotoxicity for - so far untested experimentally - metal oxides (for which only structural descriptors have been calculated).

The method proposed, although it does not provide quantitative information on the cytotoxicity of nanoparticles towards *E. coli* or to human keratinocytes cell line, nevertheless it can be used as an efficient tool for the preliminary hazard assessment of nanomaterials, as well as to identify nanomaterials that may pose a potential negative impact to the human health and the environment. One great advantage of grouping and nano-read-across technique is the fact that it does not require a large number of data to identify groups of similar compounds. Thereby, in the light of REACH regulations the proposed method has a significant practical aspect and it may be an alternative to the extremely time-consuming, costly and questionable from ethical point of view animal experiments.

Electronic Supplementary Material

Electronic Supplementary Material is available in the online version, at http://www.springer.com or from the author.

Acknowledgements

This material is based on research sponsored by the Polish National Science Center (grant no. UMO-2011/01/M/NZ7/01445). The authors also thank the National Science Foundation for support from the NSF CREST Interdisciplinary Nanotoxicity Center – grant - #HRD-0833178. A.G. thanks the European Social Fund, the State Budget and the Pomorskie Voivodeship Budget according to the Operational Programme Human

Capital, Priority VIII, Action 8.2, Under-action 8.2.2: 'Regional Innovation Strategy' for granting her with a fellowship in frame of the project "InnoDoktorant – Scholarships for PhD students, IVth edition".

References

1. Winkler D A, Mombelli E, Pietroiusti A, Tran L, Worth A, Fadeel B and McCall M J 2013 Applying quantitative structure-activity relationship approaches to nanotoxicology: current status and future potential. *Toxicology* **313** 15-23

2. Gordon T, Chen L C, Fine J M, Schlesinger R B, Su W Y, Kimmel T A and Amdur M O 1992 Pulmonary effects of inhaled zinc oxide in human subjects, guinea pigs, rats, and rabbits. *Am. Ind. Hyg. Assoc. J.* **53** 503-9

3. Rehn B, Seiler F, Rehn S, Bruch J and Maier M 2003 Investigations on the inflammatory and genotoxic lung effects of two types of titanium dioxide: untreated and surface treated. *Toxicol. Appl. Pharmacol*. **189** 84-95

4. Chen Y, Chen J, Dong J and Jin Y 2004 Comparing study of the effect of nanosized silicon dioxide and microsized silicon dioxide on fibrogenesis in rats. *Toxicol. Ind. Health*. **20**, 21-7

5. Savolainen K, Backman U, Brouwer D, Fadeel B, Fernandes T, Kuhlbusch T, Landsiedel R, Lynch I and Pylkkänen L 2013 Nanosafety in Europe 2015-2025: Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations. Finnish Institute of Occupational Health: http://www.ttl.fi/en/publications/Electronic_publications/Nanosafety_in_europe_2015-2025/Documents/nanosafety_2015-2025.pdf

6. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union, L 396/1 of 30.12.2006. Office for Official Publications of the European Communities (OPOCE)

7. van der Jagt K, Munn S, Tørsløv J and de Bruijn J 2004 Alternative Approaches can Reduce the Use of Test Animals Under REACH. *JRC Report EUR 21405*

8. Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes. OfficialJournaloftheEuropeanUnion L276/33.

9. Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari T P, Michalkova A, Hwang H M, Toropov A, Leszczynska D and Leszczynski J 2011 Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat. Nanotechnol.* **6** 175-8

10. Fourches D, Pu D, Tassa C, Weissleder R, Shaw S Y, Mumper R J and Tropsha A 2010 Quantitative Nanostructure - Activity Relationship Modeling. *ACS Nano* **4** 5703-12

11. Fourches D, Pu D and Tropsha A 2011 Exploring quantitative nanostructure–activity relationships (QNAR) modelling as a tool for predicting biological effects of manufactured nanoparticles. *Comb. Chem. High Throughput Screen*. **14** 217-25

12. Liu R, Rallo R, George S, Ji Z, Nair S, Nel A E and Cohen Y 2011 Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles. *Small* **7**, 1118-26

13. Toropov A A, Toropova A P, Benfenati E, Gini G, Puzyn T, Leszczynska D and Leszczynski J 2012 Novel application of the CORAL software to model cytotoxicity of

metal oxide nanoparticles to bacteria Escherichia coli. *Chemosphere* **89** 1098-102

14. Zhang H, Ji Z, Xia T, Meng H, Low-Kam C, Liu R, Pokhrel S, Lin S, Wang X, Liao Y P, Wang M, Li L, Rallo R, Damoiseaux R, Telesca D, Mädler L, Cohen Y, Zink Y I and Nel A E 2012 Use of Metal Oxide Nanoparticle Band Gap To Develop a Predictive Paradigm for Oxidative Stress and Acute Pulmonary Inflammation. *ACS Nano* **6** 4349-68

15. Gajewicz A, Schaeublin N, Rasulev B, Maurer E I, Hussain S, Puzyn T and Leszczynski J 2014 Towards Understanding Mechanisms Governing Cytotoxicity of Metal Oxides Nanoparticles: Hints from Nano-QSAR Studies. *Nanotoxicology* DOI:10.3109/17435390.2014.930195

16. Schultz T W, Cronin M T D, Walker J D and Aptula A O 2003 Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. *J. Mol. Struc-Theochem.* **622** 1-22

17. Lubinski L, Urbaszek P, Gajewicz A, Cronin M T D, Enoch S J, Madden J C, Leszczynska D, Leszczynski J and Puzyn T 2013 Evaluation criteria for the quality of published experimental data on nanomaterials and their usefulness for QSAR modelling. *SAR QSAR Environ. Res.* **24** 995-1008

18. Enoch S J, Cronin M T D, Schultz T W and Madden J C 2008 Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem. Res. Toxicol*. **21** 513-20

19. Enoch S J, Hewitt M, Cronin M T D, Azam S and Madden J C 2008 Classification of chemicals according to mechanism of aquatic toxicity: an evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere* **73** 243-8

20. Enoch S J, Roberts D W and Cronin M T D 2009 Formation of structural categories to allow for read-across for teratogenicity. *QSAR Comb. Sci.* **28** 696-708

21. Organisation of Economic Cooperation and Development 2007 Guidance of gruping of chemicals, ENV/JM/MONO(2007)28

22. Massart D L and Kaufman L 2008 *The interpretation of analytical data by use of cluster analysis*, Wiley, New York, USA

23. Vogt W, Nagel D and Sator H 1987 *Cluster analysis in clinical chemistry: A Model*, Wiley, New York, USA

24. Ward Jr. J H 1963 Hierarchical Grouping to Optimize an Objective Function. *J. Am. Satat. Assoc.* **58** 236-44

25. Sneath P H A and Sokal R R 1973 *Numerical taxonomy. The principles and practice of numerical classification*, Freeman, San Francisco, USA

26. Spearman C E 1910 Correlation calculated from faulty data. *Brit. J. Psychol*. **3** 271-95

27. MATLAB 7.6 R2008, MATLAB® The Language of Technical Computing v. R2008. The MathWorks Inc. http://www.mathworks.com

28. Heinlaan M, Ivask A, Blinova I, Dubourguier H C and Kahru A 2008 Toxicity of nanosized and bulk ZnO, CuO and TiO_2 to bacteria Vibrio fischeri and crustaceans Daphnia magna and Thamnocephalus platyurus *Chemosphere* **71** 1308-16

29. Portier J, Campet G, Poquet A, Marcel C and Subramanian M A 2001 Degenerate semiconductors in the light of electronegativity and chemical hardness. *Int. J. Inorg. Mater.* **3** 1039-43

Table 1. Results of the Spearman's rank correlation test between results obtained from nano-read-across technique and experiment/nano-QSAR respectively.

Statistics	nano-read-across vs.	
	Experiment	nano-QSAR
Case study 1		
ρ_S	0.955	0.961
critical $\rho_{S\ (\alpha\ 0.05)}$	0.414	0.279
p	0.0001	0.0001
Case study 2		
ρ_S	0.732	0.866
critical $\rho_{S\ (\alpha\ 0.05)}$	0.401	0.337
p	0.0026	0.0001

Figure Captions

Fig. 1. Idea of read-across approach.

Fig. 2. Two-dimensional cluster analysis. Three identified natural clusters (classes) in the data are presented. Colors represent the auto-scaled values of the selected descriptor and cytotoxicity to bacteria *E. coli*: green color, the mean value of a given descriptor/endpoint (zero value on the scale); yellow and red colors, the values higher than the mean value of a given descriptor/endpoint (higher up to 1.5 standard deviations); light and dark blue colors, the values lower than the mean values of a given descriptor/endpoint (lower up to -1.5 standard deviations). When moving down the figure from Class I to Class III, log $(EC_{50})^{-1}$ systematically decreases, and $\left(\Delta H_{Me^+}\right)$ increases.

Fig. 3. Two-dimensional cluster analysis for all 17 metal oxide nanoparticles from both: training and validation sets.

Fig. 4. Comparison of the predictive power of nano-QSAR and nano-read-across approaches with experimentally measured values of $\log(EC_{50})^{-1}$.

Fig. 5. Two-dimensional cluster analysis. Three natural clusters (classes) in the data were identified. Colors represent the auto-scaled values of the selected descriptor and cytotoxicity to human keratinocyte cell line: green color, the mean value of a given descriptor/endpoint (zero value on the scale); yellow and red colors, the values higher than the mean value of a given descriptor/endpoint (higher up to 1.5 standard deviations); light and dark blue colors, the values lower than the mean values of a given descriptor/endpoint (lower up to -1.5 standard deviations). When moving down the figure from Class I to Class III, log $(LC_{50})^{-1}$ systematically decreases, and (χ^c) increases.

Fig. 6. Two-dimensional cluster analysis for all 18 metal oxide nanoparticles from both: training and validation sets.

Figure 1.

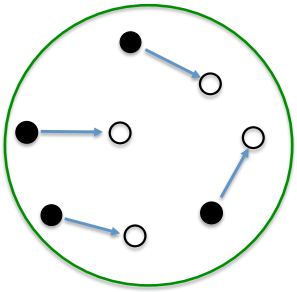
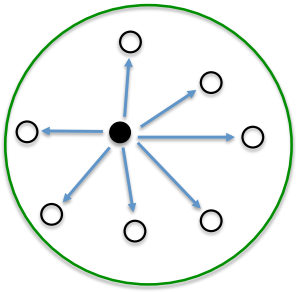
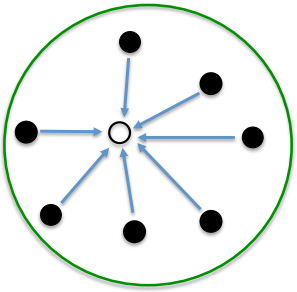
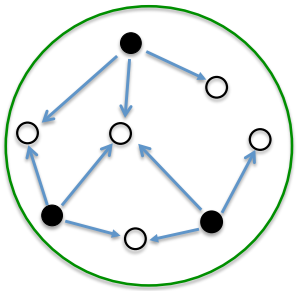
One-to-one	One-to-many
one analogue used to make an estimation for a single chemical	one analogue used to make estimations for two or more chemicals
	
Many-to-one	**Many-to-many**
two or more analogues used to make an estimation for a single chemical	two or more analogues used to make estimations for two or more chemicals
	

Figure 2.

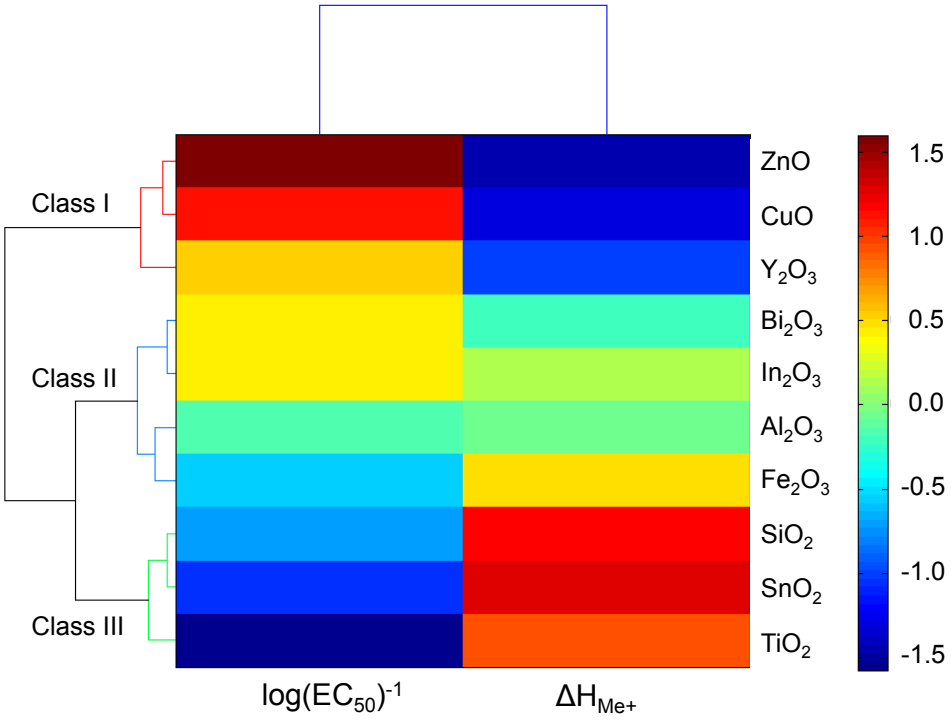

Class I
Class II
Class III
ZnO
CuO
Y_2O_3
Bi_2O_3
In_2O_3
Al_2O_3
Fe_2O_3
SiO_2
SnO_2
TiO_2
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
$\log(EC_{50})^{-1}$
ΔH_{Me+}

Figure 3.

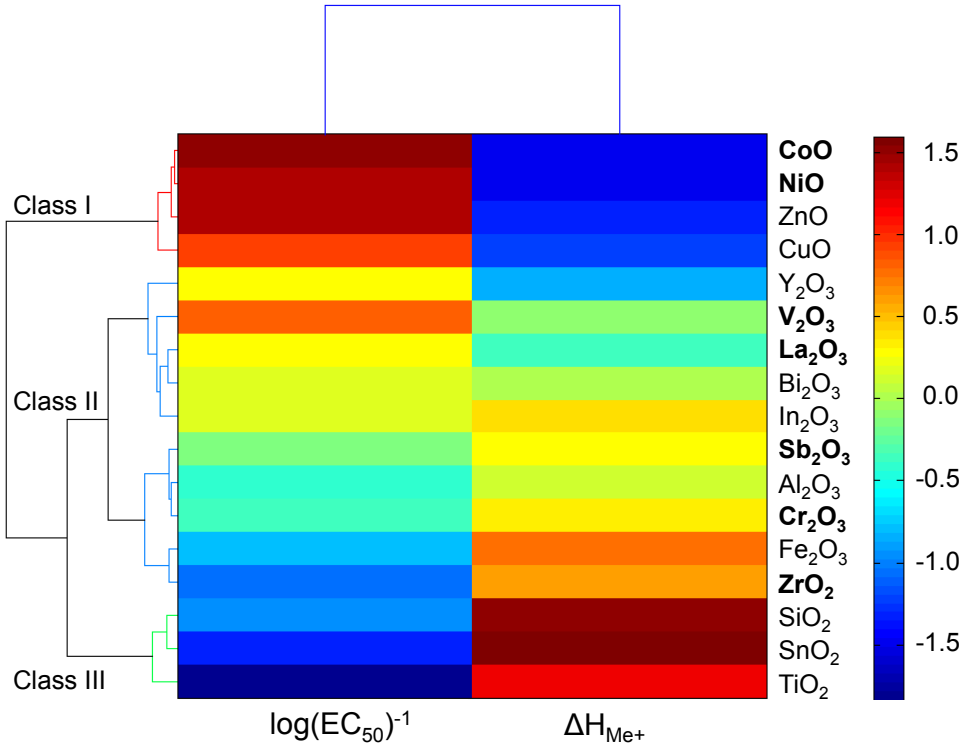

Class I
Class II
Class III
CoO
NiO
ZnO
CuO
Y2O3
V2O3
La2O3
Bi2O3
In2O3
Sb2O3
Al2O3
Cr2O3
Fe2O3
ZrO2
SiO2
SnO2
TiO2
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
log(EC50)-1
ΔHMe+

Figure 4.

Calibration and validation				Class	Prediction			
MeOx	Experimental $log(EC_{50})^{-1}$	Predicted $log(EC_{50})^{-1}$ nano-QSAR	Predicted $log(EC_{50})^{-1}$ nano-read-across		MeOx	Experimental $log(EC_{50})^{-1}$	Predicted $log(EC_{50})^{-1}$ nano-QSAR	Predicted $log(EC_{50})^{-1}$ nano-read-across
CoO	3.51	3.38	CoO	Class I	PbO	N/A	3.51	PbO
ZnO	3.45	3.30	NiO		MgO	N/A	3.45	MgO
NiO	3.45	3.38	ZnO		MnO	N/A	3.44	MnO
CuO	3.20	3.24	CuO		AuO	N/A	3.23	AuO
					FeO	N/A	3.19	FeO
					Ag_2O_3	N/A	3.08	Ag_2O_3
					Nd_2O_3	N/A	2.91	Nd_2O_3
V_2O_3	3.14	2.74	Y_2O_3		Sm_2O_3	N/A	2.90	Sm_2O_3
Y_2O_3	2.87	3.08	V_2O_3		Gd_2O_3	N/A	2.88	Gd_2O_3
La_2O_3	2.87	2.85	La_2O_3		Tb_2O_3	N/A	2.87	Tb_2O_3
Bi_2O_3	2.82	2.69	Bi_2O_3		Eu_2O_3	N/A	2.86	Eu_2O_3
In_2O_3	2.81	2.52	In_2O_3		Er_2O_3	N/A	2.85	Er_2O_3
Sb_2O_3	2.64	2.57	Sb_2O_3	Class II	Ho_2O_3	N/A	2.85	Ho_2O_3
Cr_2O_3	2.51	2.52	Al_2O_3		Mn_2O_3	N/A	2.84	Yb_2O_3
Al_2O_3	2.49	2.63	Cr_2O_3		Yb_2O_3	N/A	2.82	Mn_2O_3
Fe_2O_3	2.29	2.35	Fe_2O_3		Au_2O_3	N/A	2.48	Au_2O_3
ZrO_2	2.15	2.41	ZrO_2		Tl_2O_3	N/A	2.43	Tl_2O_3
					Fe_2O_3	N/A	2.40	Fe_2O_3
					Ga_2O_3	N/A	2.38	Ga_2O_3
SiO_2	2.20	1.99	SiO_2					
SnO_2	2.01	1.95	SnO_2	Class III	GeO_2	N/A	1.77	GeO_2
TiO_2	1.74	2.13	TiO_2					

Increasing toxicity

N/A – Experimentally measured value of $log(EC_{50})^{-1}$ is not available.

Figure 5.

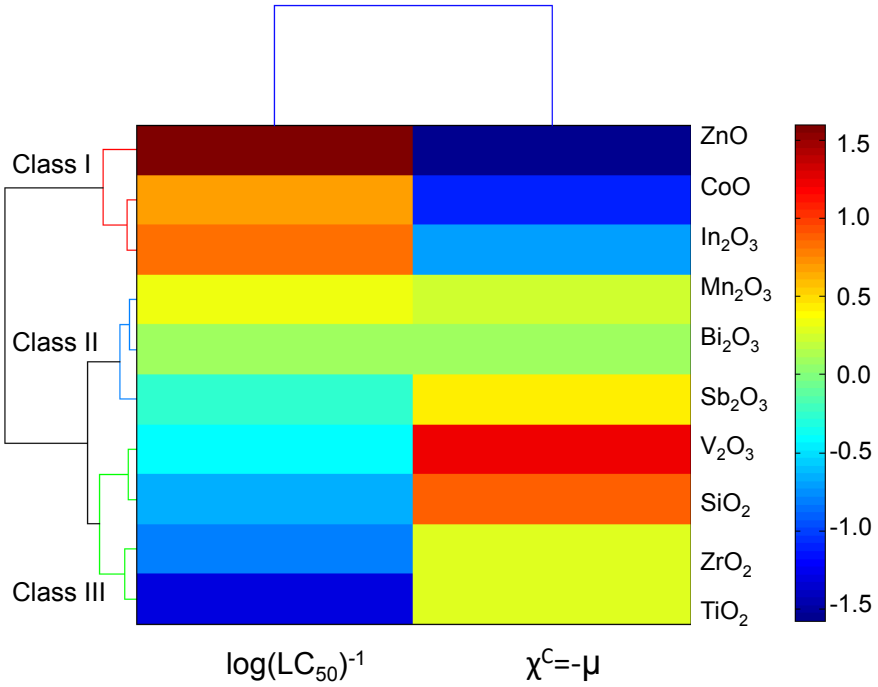
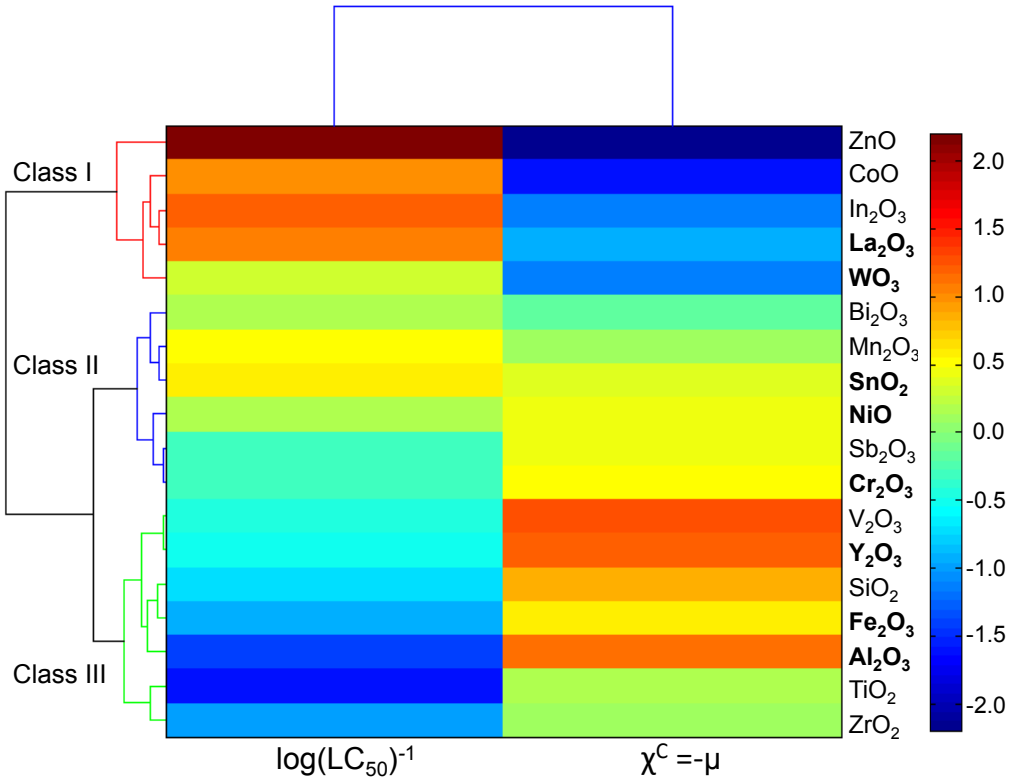

Class I
Class II
Class III
ZnO
CoO
In2O3
Mn2O3
Bi2O3
Sb2O3
V2O3
SiO2
ZrO2
TiO2
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
log(LC50)-1
χC=-μ

Figure 6.


Class I
Class II
Class III
ZnO
CoO
In_2O_3
La_2O_3
WO_3
Bi_2O_3
Mn_2O_3
SnO_2
NiO
Sb_2O_3
Cr_2O_3
V_2O_3
Y_2O_3
SiO_2
Fe_2O_3
Al_2O_3
TiO_2
ZrO_2
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
$log(LC_{50})^{-1}$
$\chi^C = -\mu$