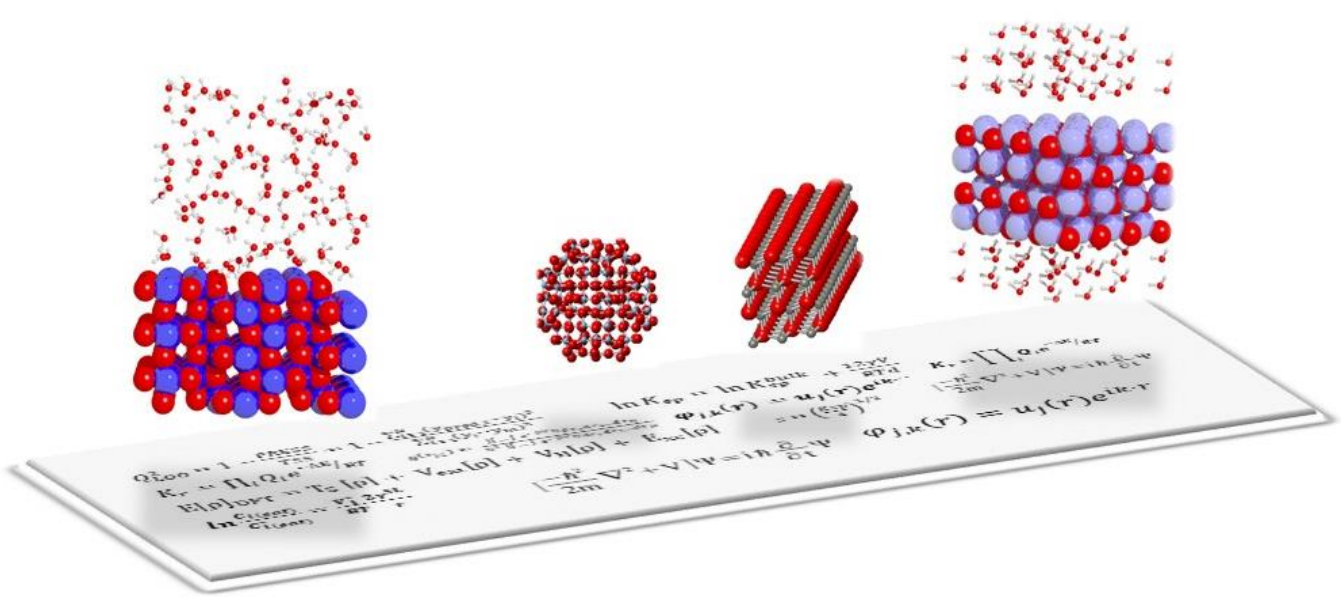
COMPUTATIONAL CHARACTERISATION OF METAL OXIDE NANOPARTICLES FOR HAZARD SCREENING AND RISK ASSESSMENT

Laura Escorihuela Martí

UNIVERSITAT ROVIRA i VIRGILI

Computational characterisation of metal oxide nanoparticles for hazard screening and risk assessment

Laura Escorihuela Martí



DOCTORAL THESIS
2019

Laura Escorihuela Martí

Computational characterisation of metal oxide nanoparticles for hazard screening and risk assessment

Tesi doctoral

Dirigida pel Dr. Alberto Fernández Sabater i Dr. Benjamí Martorell Masip

Departament d'Enginyeria Química

UNIVERSITAT ROVIRA i VIRGILI

Tarragona, 2019.

UNIVERSITAT ROVIRA I VIRGILI

Escola Tècnica Superior d'Enginyeria Química
Departament d'Enginyeria Química

Avinguda dels Països Catalans, 26
43007 Tarragona (Spain)
Tel. 977559603/04
Fax: 977559621

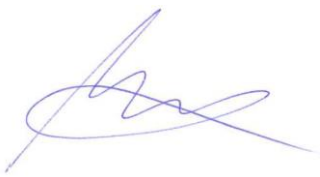
FAIG CONSTAR que aquest treball, titulat "Caracterització computacional de nanopartícules d'òxids metàl·lics per a la detecció i avaluació de riscos", que presenta Laura Escorihuela Martí per a l'obtenció del títol de Doctor amb menció internacional, ha estat realitzat sota la meva direcció al Departament d'Enginyeria Química d'aquesta universitat.

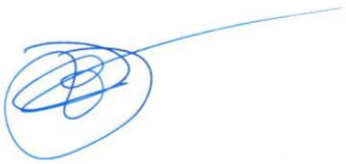
HAGO CONSTAR que el presente trabajo, titulado "Caracterización computacional de nanopartículas de óxidos metálicos para la detección y evaluación de riesgos", que presenta Laura Escorihuela Martí para la obtención del título de Doctor con mención internacional, ha sido realizado bajo mi dirección en el Departamento de Ingeniería Química de esta universidad.

I STATE that the present study, entitled "Computational characterisation of metal oxide nanoparticles for hazard screening and risk assessment", presented by Laura Escorihuela Martí for the award of the degree of Doctor with international distinction, has been carried out under my supervision at the Department of Chemical Engineering of this university.

Tarragona, 28/08/2019

El/s director/s de la tesi doctoral
El/los director/es de la tesis doctoral
Doctoral Thesis Supervisor/s





Alberto Fernández Sabater

Benjamí Martorell Masip

List of Publications

Articles:

- Laura Escorihuela, Alberto Fernández, Robert Rallo, Benjamí Martorell, " Molecular dynamics simulations of zinc oxide solubility: from bulk down to nanoparticle", Food and Chemical Toxicology, 2018, *112*, 518-525. https://doi.org/10.1016/j.fct.2017.07.038
- L. Escorihuela, B. Martorell, R. Rallo and A. Fernández, "Toward computational and experimental characterisation for risk assessment of metal oxide nanoparticles", Environmental Science-Nano, 2018, *5*, 2241-2251. DOI: 10.1039/C8EN00389K
- Laura Escorihuela and Benjamí Martorell," On the Use of GPUs in Density Functional Theory Atomistic Simulations: A Case of Acceleration Success", International Journal of Earth & Environmental Sciences, 2017, *2*, 133. https://doi.org/10.15344/2456-351X/2017/133

Book Chapters:

- Modelling the Toxicity of Nanoparticles, Ed. Springer. Editors: Tran, Lang, Bañares, Miguel A., Rallo, Robert. ISBN 978-3-319-47754-1. Book chapter title: An Integrated Data-Driven Strategy for Safe-by-Design Nanoparticles: The FP7 MODERN Project. Chapter authors: Martin Brehm, Alexander Kafka, Markus Bamler, Ralph Kühne, Gerrit Schüürmann, Lauri Sikk, Jaanus Burk, Peeter Burk, Tarmo Tamm, Kaido Tämm, Suman Pokhrel, Lutz Mädler, Anne Kahru, Villem Arouja, Mariliis Sihtmäe, Janeck Scott-Fordsmann, Peter B. Sorensen, Laura Escorihuela, Alberto Fernández, Francesc Giralt and Roberto Rallo.

- Computational Nanotoxicology: Challenges and Perspectives. Ed. Pan Stanford Publishing. *Waiting for ISBN*. Book chapter title: Physico-Chemical Properties of Nanomaterials from *in silico* simulations: An Introduction to Density Functional Theory and Beyond. Chapter authors: Laura Escorihuela, Alberto Fernández, Robert Rallo, Benjamí Martorell.

INDEX

Summary

Due to their intrinsic properties, nanomaterials (NMs) are the cornerstone of a wide range of technologically advanced applications, with metal oxide nanoparticles (MeO) NPs being the most used in areas such as electronics, optics, opto-electronics, pharmacy, medicine, cosmetics and textiles. However, there is still an important knowledge gap regarding how size influences their physicochemical properties as well as the risk to human health. Recent studies provided more insight on the size dependence of nanoparticle[1] properties and reactivity, revealing that small sized nanoparticles (NPs) have a more variable behaviour in terms of their properties than larger size NPs, which have a more constant behaviour. Therefore, nanomaterials (NMs) need a specific regulation to assess their toxicity. The generation of in *vitro* and in *vivo* toxicity characterization data is essential for risk assessment and establishment of safe use of engineered nanomaterials (ENMs). However, this is a formidable task given the expected growth in number and diversity of ENMs.

Nonetheless, toxicity assessment of NMs is a daunting task that involves multiple testing conditions and endpoints, and testing of different NP configurations. Computed based methods, *in silico* methods, based on theoretical and statistical domain, evaluate and determine and predict processes or even substance properties, this methodology is involved in very different disciplines given huge challenges. Apart from the legislation urgency for risk assessment exits a vacuum in literature, given that the data for the environmental risk assessment found in literature is uncertain and present knowledge gaps, though is not useful for the risk assessment for nanoparticles. This nanosafety data needs to be provided by standardised methods. One option to simplified this nanosafety data generation is the in silico methodology. *In silico* testing methods constitutes a cost-effective approach to fill the existing gaps in nanosafety data for being an effective tool for the safe-by-design of ENM. It exists an urgent necessity to develop nanosafety data for toxicity assessment in NPs, in particular for MeO NPs, and generate valuable QNAR (Quantitative Nano-Structure Activity Relationship) models for their risk assessment legislation using *in silico* methods to optimize time and resources to this purpose.

The most popular *in silico* method based on quantum mechanics for chemistry is Density Functional Theory (DFT), which is postulated on approximations to the exact exchange–correlation functional (e.g. LDA, GGA, hybrid GGA, meta-GGA) that are relatively computationally efficient; it is competitive in accuracy

for many interesting chemical phenomena, and it is computationally much less expensive than higher-level alternatives and useful for modelling chemical reaction pathways or for descriptor data in QSAR development. [2] DFT is the base of the huge part of the methodology implemented in this thesis.

In this thesis, we performed a strict and deep study of the best methods to evaluate the band gap and the solubility of MeO NP from a computational point of view. The use of periodical-DFT methods has allowed us to optimise structures computing the electronic ground state energy for nanostructures as the nanotubes or spherical nanoparticles. To get more reliability for band gap determination, the exchange-correlation functional has been improved using the DFT+U methodology. This type of functional gives us a reasonable computational cost and accuracy in band gap calculations or geometry optimisation, without affecting the predictive capability on the influence of experimental environment. After that, to reach large systems up to 10^3 atoms in order to simulate more realistic biological systems, it has been used the DFTB development for band gaps determination in large nanoparticles. Furthermore, the coupling of DFTB and molecular dynamics simulations has allowed the description of water-NP interaction, giving and extra value to this work.

In a work previous to this thesis, Liu *et al.* [3] determined that the descriptor E_C (conduction band energy) in the range between -4.84 and -4.12 eV of standard redox potential, can lead to the generation of cellular oxidative stress between NPs and cells developing then toxicity.[3,4,5] Furthermore, the ionic index of metal cations (Z) in MeO is correlated with their hydration energy, which is a measure of the affinity of the metal ion for water molecules. Therefore, how soluble is the MeO in the biology environment is also an important descriptor correlated with the toxicity found in other works. [6] As a consequence, this is the first descriptor chosen for this thesis. Given these two important descriptors, Liu *et al.* [4] postulated a nano-SAR (Structure Activity Relationship) developed from the toxicity data of twenty two MeO NPs from 10 to 70 nm of size.

The second descriptor chosen was the band gap between the HOMO and LUMO band energy for each size of NP, because it is totally linked in the E_c mentioned above. This is a property easy to measure via DFT or DFTB for small NPs, but it needs a large amount of computing resources for NPs with sizes greater than 2 nm. Therefore, in order to obtain band gaps for bigger NPs, the strategy followed was to create a prediction model of band gap MeO and compare the results obtained with the experimental values found in bibliography.

The computational results obtained with the methodology developed in this thesis for the ZnO case have been promising and, in order to make more robust the method employed, it has been tested for TiO_2 too, showing an excellent efficiency in results.

Great part of the work has been spent in the solubility evaluation of the MeO NP, because there is a few quantity of bibliography about that (both experimentally and computationally), what indeed is reasonable because the MeO NP has a very low solubility and it is quite difficult its evaluation either experimentally or computationally. In my opinion, the results obtained with the implementation of the use of thermodynamical theory as the Ostwald–Freundlich approach in the Molecular dynamics framework are fixing a new starting point for the solubility evaluation of NP in the near future, and it is a proof for computational engagement of legislation authorities for toxicity risk assessment.

Finally, the data obtained from the prediction models of band gap as well as the solubility models have been used to create nano-QSAR models, This is the reason why we used a nano-QSAR model published for these two MeO NPs and we added our computed descriptors in the work of Papa *et al.*,[7] in order to improve or equal the accuracy of the method. The data that may be used in these models is not enough to avoid statistical biases, but in this case it has been used Matlab software to implement statistical methods such as cross validation method, leave-on-out technique or the calculation of the root mean square error.

Finally, I would like to highlight that the work of this thesis has been carried out with the managing of different techniques, software and disciplines. So, this is an additional proof that the perfect combination of different disciplines can produce a great work.

Bibliography

1 L. K. Adams, D. Y. Lyon and P. J. J. Alvarez, *Water Res.*, 2006, **40**, 3527–3532.

2 P. G. Tratnyek, E. J. Bylaska and E. J. Weber, *Environ. Sci. Process. Impacts*, 2017, **19**, 188–202.

3 H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y.-P. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Maedler, Y. Cohen, J. I. Zink and A. E. Nel, *ACS Nano*, 2012, **6**, 4349–4368.

4 R. Liu, H. Y. Zhang, Z. X. Ji, R. Rallo, T. Xia, C. H. Chang, A. Nel and Y. Cohen, *Nanoscale*, 2013, **5**, 5644–5653.

5 E. Burello and A. P. Worth, *Nanotoxicology*, 2011, **5**, 228–235.

6 H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y. P. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Mädler, Y. Cohen, J. I. Zink and A. E. Nel, *ACS Nano*, 2012, **6**, 4349–4368.

7 E. Papa, J. P. Doucet and A. Doucet-Panaye, *SAR QSAR Environ. Res.*, 2015, 1–19.

Chapter 1. Introduction

In Europe, nanotechnology is considered as a key enabling technology (KET) that provides the basis for new advances and innovations in many fields of science and technology. In terms of economic impact, the global market of nano-enabled products was valued at $26 billion in 2014 and is expected to reach about $64.2 billion by 2019.[1] Due to their intrinsic properties, nanomaterials (NMs) are the cornerstone of a wide range of technologically advanced applications, with metal oxide (MeO) nanoparticles (NPs) being the most used ones in areas such as electronics, optics, opto-electronics, pharmacy, medicine, cosmetics and textiles. However, there is still an important knowledge gap regarding how size influences their physicochemical properties, and their toxic (cytotoxic, mutagenic or carcinogenic) effects on human health are still not well established.

Particles with one or more of their dimensions in the range of a few nm up to tenths of µm have different properties, effects and behaviour relative to their microscale counterparts. Recent studies provided more

insight on the size dependence of nanoparticle [2] properties and reactivity, revealing that small sized NPs have a more variable behaviour in terms of their properties than larger size NPs, which have a more constant behaviour. For example, size dependent changes in NPs below 5 nm have more influence than changes in NPs in the range of 15 to 90 nm due to the quantum size and macro-quantum tunnelling effects. [3] Another relevant effect of the smaller NPs is the direct exposure in an organism via the mechanism of entering directly inside the body and dissolving and delivering toxic metals, known as the Trojan effect. This effect is specific for nanoscale particles given the inadvertent recognition by cell receptors.[4]

Therefore, NMs need a specific regulation to assess their toxicity. The generation of *in vitro* and *in vivo* toxicity characterization data is essential for risk assessment and the establishment of safe use practices for engineered nanomaterials (ENMs). However, this is a formidable task given the expected growth in number and diversity of ENMs. In the European Union (EU), the REACH[5] (Registration, Evaluation, Authorisation and Restriction of Chemicals) agency directive is the current regulatory framework for chemical risk assessment and management. NMs are considered as independent "chemical substances" and therefore their registration and labelling are also regulated. Furthermore, REACH considers a "chemical element obtained by any manufacturing process, including any impurity deriving from the process used". Any chemical element can be classified within different levels of impurities if it has hazardous properties. Therefore, REACH forces NP producers and importers to provide toxicological data and environmental impact assessments (e.g., environmental exposure) when the NP concentrations are lower than 0.1% in weight.

The EU acknowledges that the application of REACH may cause administrative burden, affect time to market and increase marginal costs of nano-enabled products and technologies. In the United States, the Environmental Protection Agency (EPA) has a special regulation for NMs, the Toxic Substances Control Act (TSCA). NMs are referred to in TSCA as chemicals at the nanoscale. Due to their increased use in a huge range of products, in 2015 the TSCA regulation was extended to include chemical substances manufactured or processed as nanoscale materials (https://www.epa.gov/reviewing-new-chemicals-under-toxic-substances-control-act-tsca/control-nanoscale-materials-under).

Nonetheless, toxicity assessment of NMs is a daunting task that involves multiple testing conditions and endpoints, and testing of different NP configurations (i.e., different combinations of core, shell and functionalization layers). Computer based methods, also known as *in silico* methods, can be

used to evaluate, determine and predict processes or even substance properties, this methodology is involved in very different disciplines given huge challenging solution Regarding the use of theoretical testing methods, REACH promotes the use of computational methods to implement 3R (replacement, reduction and refinement) approaches aimed at reducing and ultimately avoiding animal testing.

In addition to the legislation urgency for risk assessment, exists a vacuum in literature given that the data for environmental risk assessment found in literature is uncertain and presents knowledge gaps and, therefore, it is not useful for the risk assessment of nanoparticles. This nanosafety data needs to be provided by standardised methods. One option to simplify this nanosafety data generation is the *in silico* methodology.

In silico testing methods, specifically the establishment of quantitative (nano)structure–activity relationships (QNARs), nano quantitative structure–property relationships (nano-QSPRs) or quantitative structure–toxicity relationships (QSTRs), constitute a cost-effective approach to fill the existing gaps in nanosafety data since they are an effective tool for the safe-by-design development of ENMs. Specifically, the development of QNAR models has been recognised as a key objective by the EU Nanosafety Cluster in its Strategic Research Agenda for 2015-2025.[1] The establishment of nano-QSPRs and QNARs requires (i) a detailed physicochemical and biological characterization of NMs, and (ii) the development of computational nano-descriptors suitable to represent the electronic, atomic and molecular structures of NMs. The development and validation of standard protocols for the experimental and theoretical characterization of NPs is fundamental to the generation of the high-quality data required to develop reliable nano-QSPRs and QNARs. For models being acceptable for regulatory decision making, they need a clear applicability domain and clear definition of the endpoint. Several reference descriptions of experimental and theoretical research protocols have been published by the Organisation for the Economic Cooperation and Development (OECD). [6] In addition, the Nanosafety Cluster, [7] promoted by the EU commission, helps to monitor and harmonize the European activities related to the risk assessment of NMs. The majority of published nano-SAR studies[8–12] have focused on MeO NPs[8,10,11] that have a high commercial production volume.[13]

Recent results have described mathematical models linking NM structure descriptors with toxicity effects. These descriptors include physical and chemical properties such as electronic band gap, or surface properties such as surface formation energy or reactive sites.[8,9,14] Regarding this relationship, for example, a band gap descriptor can be used to estimate the oxidative stress of MeO NPs.[10] Recent studies separate the surface modifiers

of NPs from the core of a MeO for predicting cellular uptake.[16] In particular, QSAR models show that diverse combinations of NP properties can be used to classify different levels of biological response for ZnO and TiO_2 NPs.[12]

As we mentioned above, the majority of the toxicity studies are based on MeO NPs, given that they are widely used in many technological applications, such as semiconductors, capacitors, coatings, solar cells, etc. because of their highly efficient properties due to their limited size and high density of corner or edge surfaces that result in unique optical, chemical sensing, and semiconducting properties. However, the high diversity of MeO and sizes of NPs in use and the lack of standardized measurement protocols makes complicated the use of physicochemical parameters for risk assessment. Bulk oxides are stable in a well-defined solid crystallographic structure (or a few structures, in some cases) under standard conditions. However, for smaller sizes (e.g. microscopic scale), the lattice stress must be taken into account because it can affect the structural properties up to the total disappearance of the crystallographic structure at the NP limit. Accordingly, phases with low stability in bulk form can be found at the nanoscale. This structural phenomenon has been reported for TiO_2, VO_x, Al_2O_3 and MoO_x oxides.[15] Nanoparticle size also influences other important features of electronic and physicochemical properties such as electrical conductivity and colour. At the nanoscale, semiconducting materials become metallic and non-magnetic particles become magnetic due to quantum-size and macro-quantum tunnelling effects. From the point of view of solid-state physics, both the superposition of bulk states and the increase in the material strength may affect electronic properties such as the band gap.[4]

Another point that shows the importance of size is the recent development of size-dependent nano-descriptors using *in silico* methods.[1,16,17] The descriptors used required only one easy experimental part, just to know the crystallographic phase of the unit cell of the metal oxide it belongs to. These descriptors are size-dependent, what it is important to take in account given that in a recent study it has been shown that some QSARs developed for MeO NPs using the enthalpy of formation of gaseous metal cation or the SMILE code, they are not capable to predict toxicity properly well because they are not size-dependent.[1]

The most popular *in silico* method based on quantum mechanics for the calculation of chemical properties from electronic structure is Density Functional Theory (DFT), which is postulated on approximations to the exact exchange–correlation functional (e.g. LDA, GGA, hybrid GGA, meta-GGA) that are relatively computationally efficient. It is competitive in accuracy for many interesting chemical phenomena, and it is computationally much less

expensive than higher-level alternatives and useful for modelling chemical reaction pathways or for descriptor data in QSAR development.[18] DFT is the base of the huge part of the methodology implemented in this thesis and is explained in the chapter of methods.

In a work previous to this thesis, Liu *et al.*[13] determined that the descriptor E_C (conduction band energy), in the range between -4.84 and -4.12 eV of standard redox potential, can lead to the generation of cellular oxidative stress between NPs and cells developing then toxicity.[10,13,19] Moreover, the ionic index of metal cations (Z) in MeO is correlated with their hydration energy, which is a measure of the affinity of the metal ion for water molecules. Therefore, how soluble is the MeO in the biology environment is also an important descriptor correlated with the toxicity found in other works. [10] Given these two important descriptors, Liu *et al.*[13] postulated a nano-SAR developed from the toxicity data of twenty two MeO NPs from 10 to 70 nm of size. Predictions regarding if NPs are toxic or not, they quantify the probability of NP *x* as being toxic (i.e., P(T|*x*)) from the intrinsic probability function used in a classification model. The results are plotted in Figure 1:

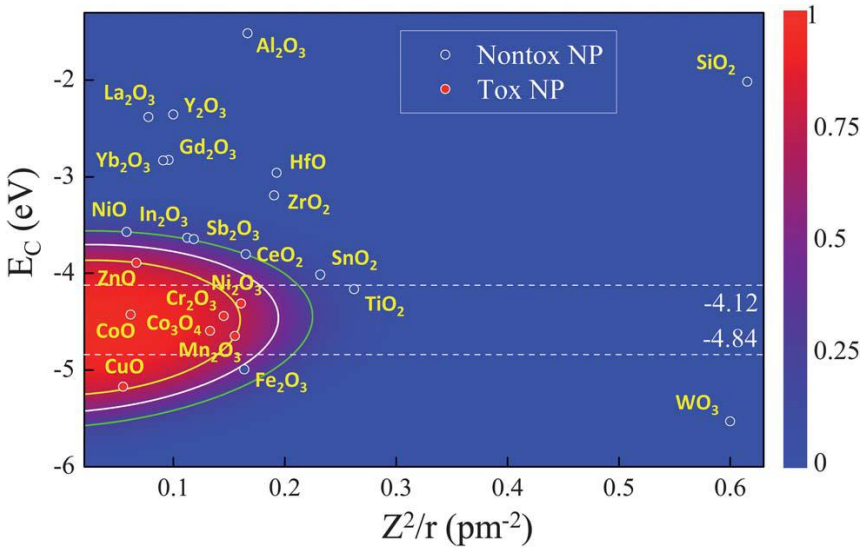


Figure 1. Toxicity probability of NP (***x***) belonging to the toxic class given by the SVM based nano-SAR. The posterior toxicity probability *P*(*T*|***x***) is depicted by the color scale in the descriptor space. The contour in the middle (*i.e.*, *P*(*T*|***x***) = 0.5) defines the nano-SAR classification boundary, while the inner (*P*(*T*|***x***) = 0.73) and outer contours (*P*(*T*|***x***) = 0.27) correspond to the decision boundaries for penalty ratios of false negative relative to false positive predictions.[13]

The nano-SAR classification boundary represented in Figure 1 spans the above suggested redox potential range. However, two (ZnO and CuO) of

the seven toxic NPs fall outside, while the non-toxic TiO_2 is located inside the redox potential range. On the other hand, for the nano-SAR that includes both E_C and Z^2/r, all of the toxic and non-toxic NPs are correctly classified.

As the ionic index decreases, the probability of a MeO becoming toxic increases. This behaviour is consistent with studies where it is shown that the metal ion toxicity increases transport across cell membranes for metal ions of lower hydration energy. [20] In this context, it is important to highlight that ZnO and CuO are classified as toxic despite their E_C being outside the suggested redox potential range. Overall, the nano-SAR (Figure 1) suggests that the probability of a MeO NP being classified as toxic would increase if its E_C is within or close to the redox potential range and its metal ionic index decreases.[13,21]

In addition to this work, in our results the ZnO was identified in the category of soluble and the most toxic together with the CuO, in total agreement with the work of Liu *et al.*. According to the experimental results, ZnO does not dissolve in biological medium; this dissolution starts in tissue or environmental culture and goes through cells or organs. On the other hand, TiO_2 was in the category of non-toxic effects at concentrations below 100 mg/l. TiO_2 is identified as an insoluble NP but its exposure to UV light can cause that the electrons are excited to conduction band creating a hole in the valence band and this effect could interact with H_2O and O_2 and generate reactive oxygen species (ROS).

The results of the works above commented are the seed for this thesis. Because it exists an urgent necessity to develop nanosafety data for toxicity assessment in NPs, in particular for MeO NPs, and generate valuable QNAR models for their risk assessment legislation using *in silico* methods to optimize time and resources to this purpose.

For both MeO NPs studied, ZnO and TiO_2, two key descriptors to identify the toxicity were evaluated. The first descriptor used is the solubility of MeO NPs, because it is linked with the enthalpy of hydration and the ionic release facility mentioned in the nano-SAR work of Liu *et al*. The study of NPs dissolution in water has been highlighted as an important physicochemical property to assess their environmental impact, as well as the study of their behavior in biological media, because their toxicity mechanism is not well known. Some studies attribute the toxicity to Zn^{2+} ions dissolved, nanoscale properties, and generation of ROS. [22,23] After obtaining the key parameters for solubility evaluation of the MeO, one model of solubility prediction on MeO NP versus size has been created for the ZnO and another one for the TiO_2, these two models are shown in the statistical chapter of this thesis. The

solubility study of NPs it has been a big challenge, because it was not still implemented at this level of theory for this purpose in literature, and for this reason all the work done has a special chapter in this thesis, the solubility chapter.

The second descriptor chosen it was the band gap between the HOMO and LUMO band energy for each size of NP, because it is totally linked in the E_c mentioned above. This is a property easy to measure via DFT for small NPs, but it needs a large amount of computing resources for NPs with sizes greater than 2 nm. Therefore, in order to obtain band gaps for bigger NPs, the strategy followed it was to create a prediction model of band gap MeO and compare the results obtained with the experimental values found in bibliography. These models are shown in the statistical chapter of the present thesis. The results and the methods used for band gap calculation for all the sizes studied are explained in the chapters of each MeO.

Finally, the data obtained from the prediction models of band gap as well as the solubility models have been used to create nano-QSAR models, in order to show if it is possible the goal proposed, to avoid some of the highly time consuming experimental part for the toxicity risk assessment. One of the main problems found in the course of this thesis, it was the experimental data scarcity for the two descriptors of interest (band gap and solubility) in the range of sizes of interest for comparing with our computed values. Because of the lack in standardisation of experimental measurements and details in the publications about the methodology used, it was not possible to get homogenised data. This is the reason why we used a nano-QSAR model published for these two MeO NPs and we added our computed descriptors in the work of E. Papa et al.,[24] in order to improve or equal the accuracy of the method. All this work is explained in the chapter of statistical analysis.

2. Objectives of this thesis

After the comprehension of the work explained, we fixed the objectives as it follows:

- Calculation of MeO NP physicochemical properties from their atomic structure using *in silico* methods based on DFT.

- Calculation of chemical and physical properties, such as band gap, solubility, surfaces energy, total energies, of MeO NPs at bigger sizes to predict solid behaviour in realistic sizes of nanoparticles.

- Development of a normalised method to calculate the solubility for MeO NPs in aqueous media.

- Generation of new data for realistic sizes of the descriptors band gap and solubility used for the modelling of toxicity prediction.

- Compilation of the data generated with experimental data available in literature, to create a new nano-QSAR for toxicity prediction of MeO NPs.

3. Bibliography

1 L. Tran, Bañares, Miguel A. and R. Rallo, *Modelling the toxicity of nanoparticles*, Cham, Switzerland : Springer, 2017.

2 L. K. Adams, D. Y. Lyon and P. J. J. Alvarez, *Water Res.*, 2006, **40**, 3527–3532.

3 J. A. Rodríguez and M. Fernández Garcia, *Synthesis, properties, and applications of oxide nanomaterials*, Wiley-Interscience, Hoboken, New Jersey, 2007.

4 L. Escorihuela, B. Martorell, R. Rallo and A. Fernández, *Environ. Sci. Nano*, 2018, **5**, 2241-2251.

5 European Comission, *Off. J. Eur. Union*, 2006, 396–849.

6 A. Hunt and U. Kingdom, Unclassified Working Party on Global and Structural Policies, *Most*, 2007, **33**.

7 M. Riediker, Compendium of Projects in the European NanoSafety Cluster 2012 Edition February. NanoSafety Cluster, 2013, 287.

8 A. E. Nel, W. J. Parak, W. C. W. Chan, T. Xia, M. C. Hersam, C. J. Brinker, J. I. Zink, K. E. Pinkerton, D. R. Baer and P. S. Weiss, *ACS Nano*, 2015, **9**, 5627–5630.

9 A. Gajewicz, N. Schaeublin, B. Rasulev, S. Hussain, D. Leszczynska, T. Puzyn and J. Leszczynski, *Nanotoxicology*, 2015, **9**, 313–325.

10 H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y. P. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Mädler, Y. Cohen, J. I. Zink and A. E. Nel, *ACS Nano*, 2012, **6**, 4349–4368.

11 D. Fourches, D. Pu, C. Tassa, R. Weissleder, S. Y. Shaw, R. J. Mumper and A. Tropsha, *ACS Nano*, 2010, **4**, 5703–5712.

12 C. Sayes and I. Ivanov, *Risk Anal.*, 2010, **30**, 1723–1734.

13 R. Liu, H. Y. Zhang, Z. X. Ji, R. Rallo, T. Xia, C. H. Chang, A. Nel and Y. Cohen, *Nanoscale*, 2013, **5**, 5644–5653.

14 T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska and J. Leszczynski, *Nat. Nanotechnol.*, 2011, **6**, 175–8.

15 A. Nel, T. Xia, L. Madler and N. Li, *Science (80-.).*, 2006, **311**, 622–627.

16 J. Ying, T. Zhang and M. Tang, *Nanomaterials*, 2015, **5**, 1620–1637.

17 A. Gajewicz, T. Puzyn, B. Rasulev, D. Leszczynska and J. Leszczynski, *Nanosci. Nanotechnologt-Asia*, 2011, **1**, 53–58.

18 P. G. Tratnyek, E. J. Bylaska and E. J. Weber, *Environ. Sci. Process. Impacts*, 2017, **19**, 188–202.

19 E. Burello and A. P. Worth, *Nanotoxicology*, 2011, **5**, 228–235.

20 F. Thévenod, *BioMetals*, 2010, **23**, 857–875.

21 H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y.-P. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Maedler, Y. Cohen, J. I. Zink and A. E. Nel, *ACS Nano*, 2012, **6**, 4349–4368.

22 I. a Mudunkotuwa, T. Rupasinghe, C.-M. Wu and V. H. Grassian, *Langmuir*, 2012, **28**, 396–403.

23 C. A. David, J. Galceran, C. Rey-Castro, J. Puy, E. Companys, J. Salvador, J. Monné, R. Wallace and A. Vakourov, *J. Phys. Chem. C*, 2012, **116**, 11758–11767.

24 E. Papa, J. P. Doucet and A. Doucet-Panaye, *SAR QSAR Environ. Res.*, 2015, 1–19.

Chapter 2. Methods

1. Density Functional Theory

A formal proof of the notion of Density Functional Theory (DFT) came in the 1960s, when Hohenberg and Kohn[1,2] published their two well-known theorems, where they reported that the total energy and properties of an electronic structure are a unique functional of the electron density (ρ) of a non-degenerate system, and that the electron density that minimises the total energy is the exact ground state density. [2]

In 1965, W. Kohn and L. Sham introduced the atomic orbitals in the DFT theory and proposed the Kohn-Sham equations. [3] This is a set of N/2 equations (for closed shell systems with no spin polarization) describing the wave function of each pair of the N-electrons as a non-interacting particle. This

formulation developed the application of DFT in computational fields. In these Kohn-Sham equations (1.1) - (1.3):

$$\frac{dE[\rho]}{d\rho(\vec{r})} = \frac{dT_S[\rho]}{d\rho(\vec{r})} + \int \frac{\rho(\vec{r_2})}{r_{12}} d\vec{r}_2 + V_{ext}(\vec{r}) + \frac{dE_{xc}}{d\rho(\vec{r})} \quad (1.1)$$

$$h_{\widehat{KS}\varphi_i} = \varepsilon\varphi_i \quad (1.2)$$

$$h_{KS} = \frac{1}{2}\nabla^2 + V_{eff}(\vec{r}) \quad (1.3)$$

The electron is subjected to a modified effective potential from the nuclei and electrons, much easier to solve than a wave function of the whole three dimensional system, having this non-interacting system the same ground–state density as the real system.[4,5] In fact, since the electron-electron interactions are always the same, it is the external potential due to ions that uniquely determines the Hamiltonian. Therefore, the total energy of the non-interacting system is given in equation 1.4:

$$\mathrm{E}[\rho]_{\mathrm{DFT}} = \mathrm{T_S}\,[\rho] + \mathrm{V_{ext}}[\rho] + \mathrm{V_H}[\rho] + \mathrm{E_{xc}}[\rho] \quad (1.4)$$

where T_S is the kinetic energy of the electrons, V_{ext} is the external potential created by the nuclei, V_H is the electrons Coulombic repulsion, and E_{XC} is the exchange-correlation (XC) term of the electronic energy, that represents the difference of kinetic energies expression and the energy of Coulombic repulsion. In this non-interacting system, the kinetic energy for the non-interacting particles is expressed in two parts: the first part is the kinetic energy exactly calculated for the non-interacting electrons, and the second part considers the interaction between electrons included in the Exc[ρ] terms. Therefore the expression of the total energy can be expanded as:

$$\mathrm{E}[\rho]_{\mathrm{DFT}} = -\frac{1}{2}\sum_{\mathrm{i=1}}^{\mathrm{n}} \int \varphi_{\mathrm{i}}^{*}\,(\mathrm{r}_1)\nabla_{\mathrm{i}}^2\,\varphi_{\mathrm{i}}(\mathrm{r}_1)\mathrm{dr}_1 - \sum_{\mathrm{x=1}}^{\mathrm{N}} \int \frac{\mathrm{Z_x}}{\mathrm{r_{xi}}}\,\rho(\mathrm{r}_1)\mathrm{dr}_1 + \frac{1}{2}\iint \frac{\rho(\mathrm{r}_1)\rho(\mathrm{r}_2)}{\mathrm{r}_{12}}\mathrm{dr}_1\mathrm{dr}_2 + \mathrm{E_{xc}}[\rho] \quad (1.5)$$

In equation 1.5, φ_i (i= 1,2…) are the Kohn–Sham orbitals, the first term in the right hand side is the kinetic energy for non-interacting electrons, the second term is the nuclear-electron interactions, the third part is the Coulombic interactions corresponding at the energy between the interacting electrons, and at the end there is the exchange-correlation term, which contains the correction of the kinetic part for the interacting electrons and the

classic correction for the electron-electron repulsion, the sum of the error made in using a non-interacting kinetic energy and the error made in treating the electron-electron interaction classically.

The correspondence of the charge density and energy of the many-body and the non-interacting system is only exact if the exact functional is known. For any particular system we could, in principle, solve the Schrödinger equation exactly and determine the energy functional and its associated potential. This, of course, involves a greater effort than a direct solution for the energy. Nevertheless, the ability to determine exact properties of the universal functional in a number of systems allows excellent approximations to the functional to be developed and used in unbiased and thus predictive studies of a wide range of materials – a property usually associated with an ab initio theory. For this reason the approximations to DFT discussed below are often referred to as ab initio or first principles methods.

Depending on the treatment of the exchange-correlation term, the accuracy of the DFT calculations varies. Classically, these approximations have been classified depending on their complexity. This gave rise to the concept of DFT Jacob's ladder, where each level of approximation adds more complexity in the treatment of the E_{XC} term.[6] The approximations are, in terms of increasing complexity, the Local Density Approximation (LDA), the Generalised Gradient Approximation (GGA), the meta-GGA, the Hybrid Functionals, and finally the exact electron density functional. In this thesis, we have made use of the GGA method. In order to explain how it works, we will start with the simplest model LDA, and the pass to the more sophisticated GGA. The use of higher levels of theory have been disregarded in this thesis because they do not substantially increase the quality of results (meta-GGA) or their computational cost is too high in periodic systems to take them into account (hybrid functionals).[7,8]

1.1. Local Density Approximation

The Local Density Approximation (LDA) was the first approach (and the simplest one used nowadays) to evaluate the exchange-correlation energy. In this approximation, the exchange-correlation energy is evaluated as a function of a uniform gas of electrons.[4,9] The exchange part for the model of a homogeneous, constant density gas is:

$$\mathrm{E}_{\mathrm{LDA}}^{\mathrm{X}}[\rho] = -\mathrm{C}_{\mathrm{x}} \int \rho^{\frac{4}{3}}(\mathrm{r})\mathrm{dr} \quad ; \quad C_x = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \tag{1.6}$$

In LDA, the XC energy density at position r depends only on the particle density at that point, ρ(r). An important property of any density functional is the exchange–correlation hole, the region around any particle in which the probability of finding another identical particle is reduced. In other words, an electron never interacts with itself, thus, the exact exchange-correlation energy provides a self-interaction correction to the Hartree electrostatic energy. This hole arises from Pauli's exclusion principle, and the total reduction of the density of the other particles should equal to −1. The LDA satisfies this exclusion principle exactly, meaning the shape of the LDA exchange-correlation hole is incorrect, although it has the correct average shape. [10]

As one can expect, LDA presents some limitations due to its simplicity, especially for the ionisation potentials and binding energies for molecules. However, it works well in electronic homogeneous systems as general solids, it improves the quality of the results with the size of the system, and it is a relatively computationally cheap methodology.[11]

1.2. Generalised Gradient Approximation (GGA) and meta-GGA

Although LDA gave some good results for systems with homogeneous electronic density structures, it also failed to describe many materials and properties. This is because one cannot consider most materials as a homogenous gas. The energy is not only dependent on the local value of the electronic density, but it is also dependent on the gradient of the density (∇ρ(r)). This leads to the formulation of the Generalised Gradient Approximation, or GGA. The typical expression of the energy for GGA is:

$$E_{GGA}^{XC}[\rho] \approx \int \rho(r)\varepsilon_{xc}(\rho, \nabla\rho)dr \qquad (1.7)$$

The GGA functional ensures the normalization condition of the spherical hole and a negative factor in the exchange correlation leading to a dependence in the density as in the gradient based on the weak of any perturbation on the homogenous electron gas. In addition, it gives an analytical improvement in the atomization energies and corrects the overbonding of H-bonding solids.[8] Several versions for the GGA functional have been reported depending on the form of the functional: PW91,[12] BLYP,[13] PBE,[14] etc. In this work PBE has been chosen as work functional.

Compared to LDA functionals, GGAs are more accurate for calculating total energies and atomization energies, because the method is more based

on the principles of the quantum mechanics scaling the densities, and then achieving correction in the upper and lower part of the scale of densities. Several approximations on the gradient have been implemented, and all of them give better results than LDA for geometries, vibrational frequencies and charge densities, being its computational cost higher.[15,16] However, it is still not enough accurate for some chemical properties of the molecules, such as the Van de Waals interactions, ionization potentials or electron affinities.

Although not being used in this thesis, it is noteworthy to mention that beyond the complexity of GGA functionals, the next step in Jacob's ladder are the meta-GGA functionals, [17,6,18] where the semi-local information of the electronic density is introduced in the Laplacian in terms of spin density or kinetic energy density, which includes some derivation of the occupied Kohn-Sham orbitals. The typical form for this functional is:

$$\mathrm{E}^{\mathrm{XC}}_{\mathrm{m-GGA}}[\rho] \approx \int \rho(\mathrm{r})\varepsilon_{\mathrm{xc}}(\rho, \nabla\rho, \nabla^2\rho, \tau)\mathrm{dr} \qquad (1.8)$$

where $\nabla^2\rho$ is the kinetic energy. With this approximation, some properties such as atomization energies are slightly improved at a much larger computational cost respect to GGA. Among those functionals one can find derivations, such as TPSS [19] or the recent SCAN. [15]

1.3. Pseudopotentials, PAW method.

For solid state chemistry, the goal in using pseudopotentials theory is to improve the convergence of the plane wave expansion. Bloch's theorem uses the periodicity of a crystal to reduce the infinite number of one-electron wavefunctions to be calculated to simplify the number of electrons in the unit cell of the crystal. The wavefunction of an electron ($\varphi_{j,k}(r)$) is written as the product of a cell periodic part and a wavelike part:

$$\varphi_{j,k}(r) = u_j(r)e^{ik\cdot r} \qquad (1.9)$$

Where $u_j(r)$ is the lattice periodic part, $e^{ik\cdot r}$ the wave part, *j* is the band index and *k* is an arbitrari vector that is confined to the first Brillouin zone of the reciprocal lattice.

Since $u_j(r)$ has the same periodicity as the direct lattice, it can be expressed in terms of a discrete plane-wave basis set with the vectors G, that are reciprocal lattice vectors of the crystal :

$$u_j(r) = \sum_G C_j G e^{iGr} \qquad (1.10)$$

The above results show that the electron wavefunctions can be expanded in terms of a linear combination of plane-waves:

$$\varphi_{j,k}(r) = \sum_G C_{j,k+} G e^{i(k+G)r^{ik\cdot r}} \quad (1.11)$$

Plane-waves are a simple way of representing electron wavefunctions. They offer a complete basis set that is independent of the type of crystal and treats all areas of space equally. To obtain its coefficients (k+G), it is need to insert this wavefunction into the Kohn-Sham equation. The Kohn-Sham operator must, therefore, be diagonalized in the space of all plane waves of momentum k + G for any vector k chosen in the first Brillouin zone. Indeed, the periodicity of k in the reciprocal space implies that it is sufficient to choose the arbitrary vector k in the first Brillouin zone

By the use of Bloch's theorem, the problem of the infinite number of electrons has now been mapped onto the problem of expressing the wavefunction in terms of an infinite number of reciprocal space vectors within the first Brillouin zone of the periodic cell, k . This problem is dealt with by sampling the Brillouin zone at special sets of *k*-points

Instead of applying Blöch theorem, and cutting kinetic energy of plane wave at huge values to have a good modelling of the core electrons of the atoms, the strategy is to use pseudopotentials to describe the core electrons. [1]This strategy reduces the time demanding process of an all-electron calculation.

The electrons not used for bonding between atoms (non-valence electrons) are “frozen” in the called frozen –core electrons approximation. They are calculated as a reference configuration but kept constant in the rest of calculation. The core electrons are represented by pseudowave functions, which reproduce the energy levels obtained by an all-electron calculation. These pseudo-wave-functions are different from the all-electron wave-functions because in the inner zone, near the nucleus, are designed not to have any node. This makes decrease in a substantial way the number of plane-waves required. [20]

The pseudopotentials used in this thesis are PAW,[21] because they have proven to give more accurate results in solid state chemistry, despite being more time-demanding than other pseudopotentials.

The biggest difference between PAW pseudopotentials, non-conserving pseudopotentials [21] or ultrasoft pseudopotentials [22] (US) and the

other ones is that PAW tries to reproduce the nodal structure of the upper core states in addition to valence states in the self-consistent iterations, so PAW is a frozen-core method, but it tries to introduce the advantages of all-electron calculations. PAW describes core electrons with a frozen nodal structure and valence electrons by an all-electron wave-function. Thus, the quality of results improves substantially specially for first row atoms.

1.4. Beyond classic DFT

DFT methods have very favourable characteristics and a huge number of scientists of different areas employ them to compute and predict properties in numerous systems (in materials science, nanotechnology and nano-engineering, chemistry, physics, Earth & environmental sciences, etc).

Although DFT is a very powerful tool, as explained in the previous section, it still has some limitations in its use. Therefore, new developments based on DFT have been developed to compute and predict better a wider range of properties and materials. Nevertheless, the main limitation in DFT methods from my point of view is computational: at present, systems with more than 1000 atoms are difficult to compute, even in large supercomputers. This is due to the cubic increase in the computational requirements respect to the number of atoms.

There are many fields where the current functionals are known to perform poorly. Current DFT methods describe poorly weak interactions due to dispersed forces such as Van der Waals interactions, and even stronger interactions such as hydrogen bonds present many difficulties for DFT functionals due to the electrostatic long-range interactions. [20]

In the next sections it is given an overview of computational methods beyond classic DFT, used in this work to improve results in the electronic structure properties calculations and to study large nanoparticles and their solubility in water.

1.5. DFT+U

In the Hartree-Fock (HF) method,[23] the Coulombic interactions are correctly described with the presence of the exact exchange term, cancelling the self-interaction of the electron. However, due to the fittings of the functionals, in DFT that exchange is only approximated, and as a consequence, the self-interaction term is not correctly cancelled. This is known as the self-interaction error. This error causes the poor description of electrons

interaction in strongly correlated systems. The Coulomb interaction between electrons is important and dominant for some transition metals, rare earth or insulating metal oxides, and the idea of a simple correlated electron gas is not enough to properly describe their properties. Because the local Coulomb repulsion of the *d* orbitals is underestimated by the classic LDA or GGA, an extra repulsion term needs to be added to correct this underestimation (U repulsion term, the Hubbard term). This correction is known as DFT+U (LDA+U or GGA+U, depending on the initial functional used) methodology.[24] Simply said, DFT+U describes the "strongly correlated" electronic states of a system (typically, localized *d* or *f* orbitals) using the Hubbard model, whereas the rest of electrons are treated at the level of classic DFT functionals. This is a methodology that is similar to hybrid functionals, in the sense that it tries to correct the exchange term in the E_{XC}, but in this case without mixing the exact correlation of the HF method.

In the DFT+U method, two new parameters are needed: U and J. The U term represents the strengths of the on-site Coulomb interaction, and the parameter J adjusts the strengths of the exchange interaction. These two parameters are combined into a new single parameter called effective U or U_{eff} = U-J. With this correction, the electron potential is locally reduced for the specified orbitals of the respective atoms or elements, and, as a consequence reducing the hybridation with the orbitals of other elements and their interaction. The limit of U_{eff} = 0 (eV) represents the classic DFT. [25]The value of the U_{eff} is fitted against experimental data for materials representative of the elements to be corrected with this methodology. Within DFT+U, the total energy is expressed as follows:

$$\mathrm{E_{DFT+U}}[\rho] = \mathrm{E_{DFT}}[\rho] + \mathrm{E}_{Hub}(\sigma_{nn}) - \mathrm{E}_{dc}(\sigma_{nn}) = \mathrm{E_{DFT}}[\rho] + \sum_n \frac{U_{eff}}{2} Tr(\sigma_n - \sigma_n\sigma_n) \quad (1.12)$$

where ρ is the electronic density, E_{Hub} is the Hubbard correction energy, E_{dc} is the energy that accounts for doubly counted electrons in the E_{DFT} and E_{Hub} terms, and σ is the electron density matrix. E_{DFT} represents the approximate DFT total energy functional being corrected and E_{Hub} is the term that contains the Hubbard Hamiltonian to model correlated states.[26]

The DFT+U method has helped to improve the results on the band gaps of some insulators and semi-conductors that were underestimated with classic DFT, as the case of ZnO. Nevertheless, they depend on an empirical correction that can overlocalize the d- or f- electrons of the metals. Besides, different parameters U depend on the chemical environment of the element, not being transferable, what is a problem in elements with different oxidation

states. Besides, due to the overlocalization, unpaired electrons interactions and magnetic structures can be wrongly predicted.[27] In this thesis, U_{eff} obtained from literature have been used for the metal elements.

1.6. Density Functional Tight Binding

As commented in the previous sections, DFT have some limitations, DFT cannot deal with large numbers of atoms (or electrons, to be fair) and has difficulties when performing long Molecular Dynamics simulations due to the expenses of the method.

A powerful alternative developed in recent years is the application of Tight Binding (TB) theory into DFT implementation. In classic DFT, the energy of a system of *M* electrons in the field of *N* nuclei at positions $\vec{R}$ is based on a functional of the electrons density $n(\vec{r})$ (or ρ in the section 1.1), following the Kohn-Sham equation. In Density Functional Tight Binding theory (DFTB), [28–30] the energy is expressed as the second-order expansion of a small charge density fluctuation $\delta n(\overrightarrow{r'})$ respect to the reference density $n_0(\overrightarrow{r'})$ (the zero-th order expansion corresponds to the TB method). The total DFTB energy can be expressed after this expansion as:

$$E_{DFTB} = \sum_{i=1}^{occ\ states}\langle\Psi_i|H_0|\Psi_i\rangle + E_{ii} - \frac{1}{2}\iint' \frac{n_0(\overrightarrow{r'})n_0(\vec{r})}{|\vec{r}-\overrightarrow{r'}|}d\overrightarrow{r'}d\vec{r} + E_{XC}[n_0(\vec{r})] - \int E_{XC}[n_0(\vec{r})]n_0(\vec{r})\,d\vec{r} + \frac{1}{2}\iint' \left(\frac{1}{|\vec{r}-\overrightarrow{r'}|} + \frac{\delta^2 E_{XC}}{\delta n(\vec{r})\delta n(\overrightarrow{r'})}\Big|_{n_0(\vec{r})}\right)\delta n(\vec{r})\delta n\left(\overrightarrow{r'}\right)d\overrightarrow{r'}d\vec{r} \quad (1.13)$$

where the first term is the Kohn-Sham energy of the occupied orbitals, the second one is the interatomic repulsion, the third one is the Coulombic repulsion, the fourth and fifth terms are the exchange-correlation energy derived terms, and the final term is related to the small fluctuation applied to the system $\delta n(\overrightarrow{r'})$.

The evaluation of E_{DFTB} implies computing several elements depending on the fluctuation of the electronic charge density, which is not simple. The first approximation consists in neglecting the last term in equation 1.13, which is the classic form of DFTB. However, in doing this, the charge density on atomic centers is kept constant in the energy evaluation, and therefore this can only be applied to covalent systems. More commonly, one develops the last term of this expression and truncates the multipole expansion after the

monopole term. This approximation is known as the Self-Consistent Charge DFTB (SCC-DFTB), and the energy is expressed as:

$$\mathrm{E_{DFTB}} = \sum_{i=1}^{occ\ states}\langle\Psi_i|H_0|\Psi_i\rangle + E_{rep} + \frac{1}{2}\sum_{\alpha,\beta}^{N}\Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \qquad (1.14)$$

where the second term encloses the short-range repulsive terms of the nuclei-nuclei interactions plus other exchange-correlation factors, and the third term accounts for the charge transfer between different atomic constituents. In more detail, the E_{rep} is defined as:

$$\mathrm{E_{rep}} = \mathrm{E_{ii}} - \frac{1}{2}\iint' \frac{\mathrm{n_0}(\vec{r'})\mathrm{n_0}(\vec{r})}{|\vec{r}-\vec{r'}|}\mathrm{d}\vec{r'}\mathrm{d}\vec{r} + \mathrm{E_{XC}}[\mathrm{n_0}(\vec{r})] - \int \mathrm{E_{XC}}[\mathrm{n_0}(\vec{r})]\mathrm{n_0}(\vec{r})\,\mathrm{d}\vec{r} \qquad (1.15)$$

However, the full evaluation of all terms in equation 1.15 requires a huge computational effort and, since it is based on the exchange-correlation functional, the final result will depend on this election. Therefore, the strategy typically used in DFTB is to represent this repulsion energy as a sum of pair-wise functions between atoms. These functions are obtained by fitting them to high-level theoretical calculations. [5]

The third term of equation 1.14 takes into account the charge fluctuations between atoms in the system and it is derived from the last term in equation 1.11. In this last term, for local XC functionals (see section 1.2), the second-order terms become a Dirac delta function, becoming this last term the electrostatic interaction shown in equation 1.14. [5] For equation 1.14, Δq_α represents the Mulliken charge change respect the neutral atom α and the $\gamma_{\alpha\beta}$ is the extent of that charge interaction between centres α and β. Finally, the $\gamma_{\alpha\beta}$ function is approximated using the Hubbard parameters U_α and U_β of the atoms (directly related to the chemical hardness of the element and used also in the DFT+U methodology) and the distance between those two centres. [4] The Hubbard parameters are obtained from DFT simulations, and therefore we do not introduce at this point any empirical data for the fitting of the computational methodology.

All the values parameterized for the V_{rep} of the different atoms required in the simulations, together with the symmetry transformations for the evaluations of the orbital energies, are saved in the Slater-Koster files (SK). [32]The SK files are tabulated for a large quantity of pairs of atoms (the full periodic table is still not available, but the database is increasing) and available to the general public in the DFTB project webpage, www.dftb.org. Alternatively, one can create his/her own files from ab-initio simulations, however at a large cost of computational and results analysis time. At present,

several codes and quantum chemistry packages include the DFTB methodology and are compatible with the SK files published for general use: DFTB+, [32,4] deMon, AMBER, ADF or Gaussian.

With DFTB, one can compute systems formed by thousands of atoms in relatively small computational clusters, being a reliable method. Nevertheless, the SK files available are still relatively limited to organic molecules, some inorganic solids and transition metals. Besides, the results that one obtains are also determined by the functionals used to implement the SK files. As a consequence, for instance, when computing liquid water systems, the hydrogen bonds are not properly described as in the case of the GGA functionals, and DFTB trends to overestimate the intermolecular forces in water, forming voids in the water bulk.[33]

As a final reminder of section 1.2, we present in table 1 a summary of those methods here explained, showing the advantages, possibilities and disadvantages of those methods.

Table 1. Summary of methods beyond classic DFT.

Approximation	Advantages	Disadvantages
DFT+U[1,2,24]	-Low cost methodology to accurately correct band gaps underestimation in GGA -It provides accurate results for strongly correlated systems: rare-earth compounds, transition metals, Mott insulators and impurity systems, superconductors	- DFT+U solutions can overestimate the degree of localization -The accuracy in results depends on the functional used in the DFT part
DFTB [4-6]	-Systems of tens of thousands of atoms at quasi-DFT level with small effort (x100 quicker on average) -Same accuracy as the DFT reference model -Easy to implement for both molecular and periodic systems, as well as including fields in the simulations	-The accuracy in results depends on the functional used for the parameterisation -Only databases for organic interactions, water and a few solids exist

2. Molecular Dynamics

Classical Molecular Dynamics (MD) methods are essentially based in Newton's mechanics to describe the nuclear motions. Treating atoms and molecules as classical particles gives valuable information about thermodynamics, structures and dynamical properties of the condensed matter from pure liquids to complex biomolecular systems. Molecular dynamics simulations compute motions, so it makes possible to describe position, velocities and changes versus time of individual molecules in solids, liquids o gases. In this thesis we have used MD methods to describe the physico-chemical behaviour of aqueous solutions of Nanoparticles.

The equilibrium in MD usually is applied to an isolated system containing N number of molecules and a fixed volume, V. The total energy, E, is also constant. E is the sum of the molecular kinetic and potential energies, all of these variables determining the thermodynamic state.

In a NVE-molecular dynamic position r^N is solved by Newton's equation:

$$F_i(t) = m\ddot{r}_i(t) = -\frac{dU(r^N)}{dr_i} \qquad (1.16)$$

where F_i is the force on i caused by the N-1 other molecules, m is the molecular mass, and U is the intermolecular potential energy.

Solving equation 1.16 yields the atomic momenta, atomic positions and individual atomic trajectories. Time averages ‹A› can be computed to obtain macroscopic properties:

$$\text{‹A›} = \lim_{t\to\infty} \frac{1}{t} \int_{t_0}^{t_0+1} A(\tau)d\tau \qquad (1.17)$$

When time averages are o are obtained from positions, time average (1.17) represents static properties as thermodynamic properties; from momenta we obtain dynamics properties, as transport coefficients. [35]

Some applications of molecular dynamics in different fields nowadays are listed herewith, in fundamental studies as kinetic theory, transport properties, size dependence. For collective behaviour, it can be used to calculate the decay of space and time correlation functions, vibration, spectroscopic measurements, and dielectric properties. In complex fluids, it can be used for modelling structure and dynamics of glasses, molecular liquids, pure water and aqueous solutions, fluid interfaces, films, etc. In solid cases is used to the study of defects, fractures, mechanical properties, etc. [35]

3. Radial distribution function.

The relative average position of particles in a liquid state is expressed by the Radial Distribution Function (RDF), in other words, RDF represents the spherical average local organisation around an atom, and it means the probability to find another N atom at a distance r from such an initial atom. The equation 1.16 represents the RDF between the molecules 1 and 2 in a fluid with N particles:

$$g(r_{12}) = \frac{\iint \dots \int e^{-\beta V_N} dr_3 dr_4 \dots dr_N}{N^2 \iint \dots \int e^{-\beta V_N} dr_1 dr_2 \dots dr_N} \qquad (1.18)$$

Where, β=1/kT and V_N is the potential energy of the N particles. In order to introduce the intermolecular potential, it is considered a box containing around 10^3 particles (this value depends on the system to study and the level of theory to evaluate V_N) and the rest of the liquid is simulated with infinite copies of this box surrounding the original. Therefore, the interactions between molecules are contemplated inside the box and if a molecule goes out the box where it was, its image arrives through the opposite face. [36]

In molecular dynamics, it is used the Newton law is used to estimate where will be each particle in a short period of time, around 10^{-15} s, a time period shorter than the collision time between particles. These calculations are repeated thousands of times giving snapshots of the liquid situatuion,[36,37] as we can see in the solubility chapter. The temperature of the system is inferred by computing the mean kinetic energy of the particles.

This process must be repeated for many complete shells over r to be significant. For example, in an ideal gas g(r) = $4\pi r^2$ gives a simple quadratic curve. [38]The fluid state is characterized by the absence of any permanent structure at medium and long range (RDF→1). Crystal solid state presents a very well defined long-range crystal structure.

The RDF is related to the experimental parameter S(k), the key quantity in interpreting x-ray scattering measurements, by Fourier Transformation. [35]

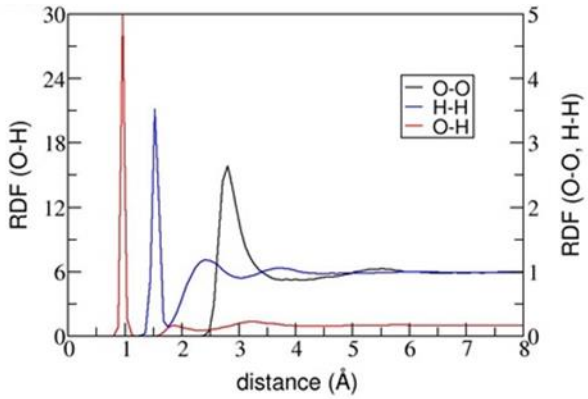


Figure 1. RDF of water box calculated with MD-DFTB

For example, the RDF corresponding to water bulk simulation in this thesis is displayed in Figure 1.

4. Solubility evaluation

The solubility of Metal Oxide (MeO) nanoparticles has been evaluated from MD energy simulations, but with two different approximations depending on the size of the nanoparticles. To evaluate the solubility of bulk and for the smallest NPs, it has been calculated the solubility product constant using the Helmholtz free energy; the extensive energy evaluation of the system and the solubility constant have been computed using what we called the "water box method" for small nanoparticles. For larger NPs, a methodology based on the Ostwald-Freundlich nonextensive thermodynamic approach was used.

4.1. Thermodynamic property: Helmholtz free energy.

The solubility product constant (K_{sp}) is a well-defined value for an ionic solid (in general a binary salt: A_nX_m) and is well defined from the ions in solution:

$$K_{sp} = [A^{+m}]^n[X^{+n}]^m \qquad (1.19)$$

However, in non-fully ionic solids, as in our case with metal oxides, this definition is more ambiguous since the solved product may not be an ion but a cluster or any other general species solved in water.[37]

Accordingly to our non-dissociating MeO systems in water as an ionic solution, another way is to define the K_{sp} is the reactivity constant, K, of the

dissolution chemical reaction. From statistical thermodynamics point of view, it is possible calculate this K using Helmholtz free energy (A) and the corresponding definition for a given chemical reaction:[35]

$$aA + bB \rightarrow cC + dD$$

$$K = \frac{\left(\frac{q^0_{C,m}}{N_A}\right)^c \left(\frac{q^0_{D,m}}{N_A}\right)^d}{\left(\frac{q^0_{A,m}}{N_A}\right)^a \left(\frac{q^0_{B,m}}{N_A}\right)^b} e^{-\Delta_r A/RT} \quad (1.20)$$

where q_m is the partition function of the components in the chemical reaction. In the systems simulated in this thesis, reactants were water, bulk materials and nanoparticles (in vacuum), and final products were defined as solved nanoparticles (in different aggregation models). This method has been used to calculate the K_{sp} of bulks and K_{sp} of small nanoparticle for both cases, ZnO and TiO_2. More details will be given in the results section for each simulated system.

4.2. Water box method

The periodic box concept is the way to represent the minimum expression of a solution, so a small nanoparticle of MeO is introduced in a cubic box of water. The cubic shape let to reproduce the boundary conditions of periodicity implicit in MD calculations with DFT-TB method.

The cubic box, shown in the Figure 2, allows to perform MD simulations of water to achieve the maximum real conditions as experimental ones. Given the motion perhaps some atoms go through the boundary of the box, but because of the periodic conditions in the system, this same atom is reproduced in the other side of the cell given by the periodic image.

Besides the “pure” water box, nanoparticles were included in the water box to simulate the interaction of the nanoparticles with water molecules. From these simulations, we computed the dissolution Helmholtz free energy of small nanoparticles.

Water box models were created using the VMD code. [40] On the other hand, spherical nanoparticles were created from bulk making use of the package Nanocut. [40]A Fortran code was developed to introduce pre-designed nanoparticles in the water box, which removed the overlapping water molecules with the particles at a certain critical distance (2.0 Å from any water molecules of the nanoparticle).

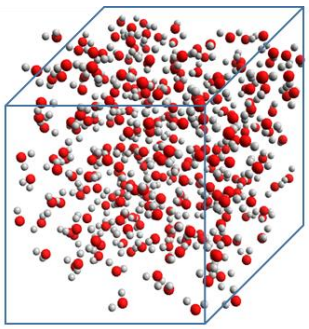


Figure 2. Cubic water box, 21 Å size.

The Water box method, has been chosen to obtain the Helmotz energy in the cases where the NP are bigger, from the energy of the total system (water and NP) is taking off either the energy of the water box without NP and the energy of the NP is the vacuum. More details and results are explained in the Solubility chapter of this thesis.

4.3. Ostwald-Freundlich thermodynamic approach.

As Ostwald and Freundlich postuled on the 20th century, solubility at nanometric scale becomes a magnitude which varies with size and shape of the solid. [41]In consequence, the solubility of a compound in a solvent at determined temperature, cannot be referenced as a thermodynamic property because it depends on its size. The theory postulated by Ostwald and Freundlich describes the variation of the concentration at saturation point, $C_{i(sat)}$, of a spherical particle of radius, r, of a molecular solid, i in a solvent as the equation:

$$\ln \frac{C_{i(sat)}}{C^*_{i(sat)}} = \frac{V_i^*}{RT}\frac{2\gamma^{SL}}{r} \qquad (1.21)$$

$C^*_{i(sat)}$ means concentration of i at saturation, when the solid i is in unlimited phase, no size effect. V_i^* molar volume of i, T is temperature, R perfect gas constant and γ^{SL} is the solid-liquid interfacial tension.

The application of this postulate to the solubility in spherical particles, which are represented by convention an Euler's function order, m, is equal to two-thirds, [43]if the area is compose by solid-liquid and the volum, V, is the volume of the spherical particle, the postulated for spherical nanoparticles is:

$$\ln \frac{C_{i(sat)}}{C^*_{i(sat)}} = \frac{V_i^*}{RT}\frac{2\gamma^{SL}}{3}\frac{4\pi r^2}{\frac{4}{3}\pi^3} = \frac{V_i^*}{RT}\frac{2\gamma^{SL}}{r} \qquad (1.22)$$

It is interesting to observe that at the increase of particle radius involves a decrease in the concentration of i at saturation equilibrium.

The same Ostwald Freundlich approach can be explained from the point of view of thermodynamics as:

$$2.3RTlog\frac{K_{SO,A}}{K_{SO}} = \frac{2\gamma M_{nano}}{\rho r} = \frac{2}{3}\gamma A \qquad (1.23)$$

Where $K_{SO,A}$ is the solubility constant for the bulk material (K_{SO}) and γ ($J \cdot m^{-2}$) is the surface tension of a specific area, A. ρ mass density (gm^{-3}) , M_{nano} is the molar mass ($g \cdot mol^{-1}$). [42]

Using the 1.23 equation, in this work we simulate a bulk of each MeO surrounded by two layers of water. Using MD simulation, we obtain the energy of the interaction between water and solid, this can be used to calculated the surface tension γ between the solids and water. From this point, it is possible the to achieve the K_{sp} and the solubility for each MeO. More details about these calculations are explained in the Solubility Chapter.

5. STATISTICAL METHODS

5.1. Nano-QSAR

Quantitative Structure-Activity Relationship (QSAR) is a common tool used to develop valuable *in silico* models to screen and predict the desired endpoint/s employing the correlation of some descriptors (experimental or theoretical). The endpoint gives a value or indicator of a parameter related to the activity of chemicals under specific conditions without kwon interaction beaten them. This computational technique is used for the assessment of industrial chemicals, by agencies as the European REACH and even Australian NICNAS.[44, 4]

The use of these models in industrial or other applications is strictly necessary to perform the OECD principles for the validation and regulatory purposes. These principles were dictated in the 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnologies in November 2004 and include:

- A defined endpoint: is defined as a measure of activity for chemicals made under specific conditions. This is refereeing to any physicochemical property, biological effect or environmental parameter that can be measured or modelled. [44]This means that

the endpoint for being predicted must be clear, since a given endpoint could be determined by different experimental or theoretic protocols.

- An unambiguous algorithm: transparency in the model algorithm that generates predictions is required. Accurate mathematical description of the algorithm used to define the relationship between the descriptors and the activity, even a clear description of the descriptors calculation.
- A defined domain of applicability: is the response and chemical structures space in which the model makes predictions with a given reliability. [44]QSAR is a reductionist model which is limited at the type of chemical structures, properties of mechanisms used for model generation.
- Appropriate measures of goodness-of-fit, robustness and predictivity: this principal is for setting the next principle, the goal is it simplify the set of principles without losing the distinction between the internal performance of a model and the predictively of it.
- A mechanistic interpretation, if possible: to find multiple mechanistic interpretation of a given model, then it is advisable to explain the interpretation. It not means that a model without an explanation is not valid in term of regulatory context.

The extension of QSAR methodology to nanoscale has developed the nano-QSAR modelling. In 2013, the European Union reported the importance of developing nano-QSAR models to predict relevant endpoints of toxicity and ecotoxicity for NP hazards assessment. [43] The nano-structure characteristics can be described experimentally measuring physicochemical properties, such as size, surface area, zeta potential and theorical characteristics can be molecular descriptors such as quantum –mechanics descriptors, SMILES.

At nano level, most of the descriptors used at micro or macro scales are not useful, because the reactivity of nanoparticles increases when the size is reduced and other properties also are being modified respect bigger scales as the solubility or hardnesses. Nanoparticle size also influences other important features of electronic and physico-chemical properties such as colour and electrical conductivity. Semiconducting materials become metallic and non-magnetic particles become magnetic due to the quantum-size effect and to the macro-quantum tunnelling effect related to nano-size. From the point of view of solid state physics, both the superposition of bulk states and the increase in the strength of the material may affect electronic properties of the material such as the band gap.[4]

5.2. Stepwise regression

To search the best nano-QSAR model, in this thesis, we used the method stepwise regression to fix the best combination of the descriptors as candidates of a regression variables. Stepwise regression is a way to automatically determine which variables should be included in the multiple linear regression model.

This classical statistical method calculates de F-value for the incremental inclusion of each variable in the regression. The F-value is equivalent to the square root of the student t-value, expressing how different two samples are from each other, where one sample includes the variable and the other samples does not. T-value is calculated:

$$F = \sqrt{t - value} \quad (1.24)$$

t = difference in the sample means/standard deviation of differences

F-value is sensitive to the number of variables used to calculate the numerator of this ratio and the denominator. Stepwise regression calculates the F-value both with and without using a particular variable and compares it with a critical F-value to either include the variable (forward stepwise selection) or to eliminate the variable from the regression; in a such way the algorithm can select the set of variables that meets the F-value criterion. It is assumed that the variables account for a sufficient amount of total variance of the variable at a given level of confidence specified in F-value (usually 95%).
[45]

6. Bibliography

1 B. Martorell Masip, Ph. D. Thesis, 2008, ISBN: 9788469197479.

2 P. Hohenberg and W. Kohn, *Phys. Rev. B*, 1964, **136**, B864–+.

3 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, 1133-.

4 E. Engel and R. M. Dreizler, *Theor. Math. Phys.*, 2011, **2011**, 499–515.

5 E. B. Tadmor and R. E. Miller, *Modeling materials: Continuum, atomistic and multiscale techniques*, 2011, ISBN: 9780521856.

6 J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka, *J. Chem. Phys.*, 2005, **123**.

7 L. Simón and J. M. Goodman, *Org. Biomol. Chem.*, 2011, **9**, 689-700.

8 S. F. Sousa, P. A. Fernandes and M. J. Ramos, *J. Phys. Chem. A*, 2007, **111**, 10439–10452.

9 J. P. Perdew and A. Ruzsinszky, *Int. J. Quantum Chem.*, 2010, **110**, 2801–2807.

10 V. Tschinke and T. Ziegler, *Can. J. Chem. Can. Chim.*, 1989, **67**, 460–472.

11 P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark and C. J. Pickard, *Philos. Trans. R. Soc. a-Mathematical Phys. Eng. Sci.*, 2014, **372**.

12 J. Perdew, J. Chevary, S. Vosko, K. Jackson, M. Pederson, D. Singh and C. Fiolhais, *Phys. Rev. B*, 1992, **46**, 6671–6687.

13 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

14 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

15 D. C. Langreth and M. J. Mehl, *Phys. Rev. B*, 1983, **28**, 1809–1834.

16 J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2005, **122,** 114102-7.

17 J. P. Perdew, S. Kurth, A. Zupan and P. Blaha, *Phys. Rev. Lett.*, 1999, **82**, 5179.

18 J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, M. L. Klein and J. P. Perdew, *Nat. Chem.*, 2016, **8**, 831–836.

19 J. P. Perdew and L. A. Constantin, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2007, **75**.

20 A. Roldán Martínez, Ph. D. Thesis, 2010, ISBN: 9788469376690.

21 P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953–17979.

22 D. Vanderbilt, *Phys. Rev. B*, 1990, **41**, 7892–7895.

23 A. Szabo and N. Ostlund, 1989, ISBN: 9780486691862.

24 E. Pavarini, E. Koch, F. Anders, *Correlated Electrons: From Models to Materials*, 2012, **2**, ISBN: 978-3-89336-796-2.

25 C. Loschen, J. Carrasco, K. M. Neyman and F. Illas, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2007, **75**.

26 B. Himmetoglu, A. Floris, S. De Gironcoli and M. Cococcioni, *Int. J. Quantum Chem.*, 2014, **114**, 14–49.

27 S. Bajaj, C. Sevik, T. Cagin, A. Garay, P. E. A. Turchi and R. Arryave, *Comput. Mater. Sci.*, 2012, **59**, 48–56.

28 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260–7268.

29 P. Koskinen and V. Mäkinen, *Comput. Mater. Sci.*, 2009, **47**, 237–253.

30 G. Seifert, in *Journal of Physical Chemistry A*, 2007, **111**, 5609–5613.

31 D. A. Papaconstantopoulos and M. J. Mehl, *J. Phys. Condens. Matter*, 2003, **15**, 413-440.

32 B. Aradi, B. Hourahine and T. Frauenheim, in *Journal of Physical Chemistry A*, 2007, **111**, 5678–5684.

33 T. H. Choi, R. Liang, C. M. Maupin and G. A. Voth, *J. Phys. Chem. B*, 2013, **117**, 5165–5179.

34 J. M. Haile, I. Johnston, A. J. Mallinckrodt and S. McKay, *Comput. Phys.*, 1993, **7**, 625.

35 D. C. Rapaport, *The art of molecular dynamics simulation*, 2004, ISBN: 9780511816581

36 P. Atkins and J. De Paula, *Atkins' physical chemistry*, 2009.

37 L. Escorihuela, A. Fernández, R. Rallo and B. Martorell, *Food Chem. Toxicol.*, 2018, **112**, 518–525.

38 F. L. Somer, *J. Chem. Educ.*, 2004, **81**, 1573.

39 W. Humphrey, A. Dalke and K. Schulten, *J Molec Graph.*, 1996, **14**, 33–38.

40 Copyright (c) 2009, Florian Uekermann, Sebastian Fiedler, Bálint Aradi.

41 P. Letellier, A. Mayaffre and M. Turmine, *J. Phys. Condens. Matter*, 2007, **19**, 436229.

42 T. Hiemstra, *Geochim. Cosmochim. Acta*, 2015, **158**, 179-198.

43 Y. Pan, T. Li, J. Cheng, D. Telesca, J. I. Zink and J. Jiang, *RSC Adv.*, 2016, **6**, 25766–25775.

44 T. Puzyn, N. Jeliazkova, H. Sarimveis, R. L. Marchese Robinson, V. Lobaskin, R. Rallo, A. N. Richarz, A. Gajewicz, M. G. Papadopulos, J. Hastings, M. T. D. Cronin, E. Benfenati and A. Fernández, *Food Chem. Toxicol.*, 2017, **112**, 478-494.

45 R. Nisbet, J. Elder and G. Miner, *Handbook of Statistical Analysis and Data Mining Applications*, 2009.

Chapter 3. ZnO, case of study

1. Introduction

Zinc oxide is a highly versatile material with a unique combination of electronic, catalytic, piezoelectric and pyroelectric properties, which lend it to a vast number of emerging technological applications. Zinc oxide may be fabricated in a wide variety of possible nanostructures, and nanostructured ZnO lends itself to a wide range of possible applications including nanogenerators, sensors, cosmetics, UV-absorbers and photovoltaics. ZnO exhibits a wide bandgap (3.37 eV), large exciton binding energy (60 meV) and low lasing threshold, which is applicable to optoelectronics, sensors, transducers and biomedical science. [1]

Mechanical and electromechanical properties of *1-D* ZnO nanostructures have been the subject of numerous investigations to form part of the fundamental components of nanopiezotronics in nanotechnology in which piezoelectric nanostructures are employed for harvesting energy in self-

powered wireless nanodevices. Size dependence of Young's Modulus in ZnO *nanotubes* (NTB) has been experimentally and theoretically revealed. In particular, it has been shown that the Young's modulus of ZnO Nanowires *(Nws)* with diameters smaller than about 120 nm increases dramatically with decreasing diameters. The reported ZnO nanotubes usually grow in experiments along the [0001] direction. [2,3]

The crystal structures shared by ZnO are wurtzite, zinc blende, and rocksalt. Most of the group II–VI binary compound semiconductors crystallize in either cubic zinc blende or hexagonal wurtzite (Wz) structure, where each anion is surrounded by four cations at the corners of a tetrahedron, and vice versa. This tetrahedral coordination is typical of sp^3 covalent bonding nature, but these materials also have a substantial ionic character that tends to increase the bandgap beyond the one expected from the covalent bonding. ZnO is a II–VI compound semiconductor whose ionicity resides at the borderline between the covalent and ionic semiconductors.

In this thesis we centered our studies in ZnO wurtzite crystalline phase. Under ambient and near-ambient conditions, ZnO forms with a wurtzite crystal structure and nano-structures. The wurtzite structure has three primary cleavage planes, the {1010}, {1120}, and {0001}. Cleavage along the {0001} axis results in two polar surfaces, the (0001)-Zn terminated and (0001)-O terminated surfaces, while cleavage along the (1010) and (1120) planes results in nonpolar surfaces. [4]

The wurtzite structure has a hexagonal unit cell formed by 4 atoms, with lattice parameters [5] *a* and *c* showed in Figure 1:

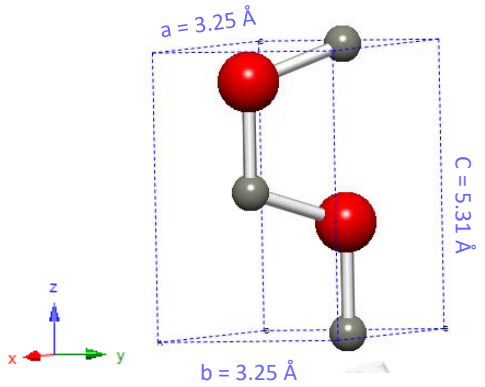


Figure 1. ZnO hexagonal 4-atoms unit cell (5th atom is the periodic image). Spatial group P_{6m3}.

2. DFT calculations

2.1. Computational details.

Density functional theory calculations were conducted using the Quantum espresso (QE) program. [6,7] The self-consistent calculations were performed under the Gradient Generalised Approximation (GGA) with the functional developed by Perdew-Burke-Ernzerhof (PBE)[8] .

The effect of the core electrons on the valence electron density was described by the projector augmented wave (PAW) method. The cut-off for the kinetic energy of the plane-waves was set to 85 Ry throughout, which ensures total energy convergence to better than 10^{-6} eV. A Gaussian smearing technique with a 0.005 eV width was applied to enhance convergence but all energies presented in the following were obtained by extrapolating to zero smearing (0K).

Some calculations were carried out on parallel computers in an in-house facility or in the ARCCA –Raven Supercomputing Cluster at the University of Cardiff, thanks to the collaboration of Professor Peter Kille. However, for the largest species here computed, a high performing computer (HPC) was required. The work of this thesis was been awarded twice with CSUC–RES grants QCM-2017-2-0025 and QCM-2017-3-0040, where calculations with up to 256 cores were performed.

The analysis of electronic structure properties and the comparison of ZnO species (band gap, formation energies…) started with an analysis of properties, going from 3 dimensions material to 0 dimensions, this means from bulk, to surfaces, nanotubes and finally spherical and non-spherical nanoparticles. In the case of the bulk, nanotubes, surfaces and non-spherical nanoparticles, structures modelled were cut from the 3D expansion of the wurtzite-type structure (zincite) with the experimental lattice parameters of a=b=3.25 and c=5.20 Å. [5] The expansion was built with Avogadro, [9] an open-source molecular builder and visualization tool. For the construction of spherical nanoparticles, it was used the software Nanocut, a program designed to cut out various objects from three dimensional crystal structures for atomistic simulations (https://aradi.bitbucket.io/nanocut/index.html).

2.2. Bulk optimisation

The energy of experimental crystallographic parameters of ZnO bulk crystal was minimized at various k-points grids and energy cut-offs to establish the optimum values for those parameters and for the bulk crystallographic parameters. Simulation cell was formed by 4 atoms as shown in Figure 1. Results converged for the bulk structure at plane-wave cut-off energy of 85 Ry and k-points 5x5x3, see Table 1. Results show good agreement with experimental and GGA values from literature.

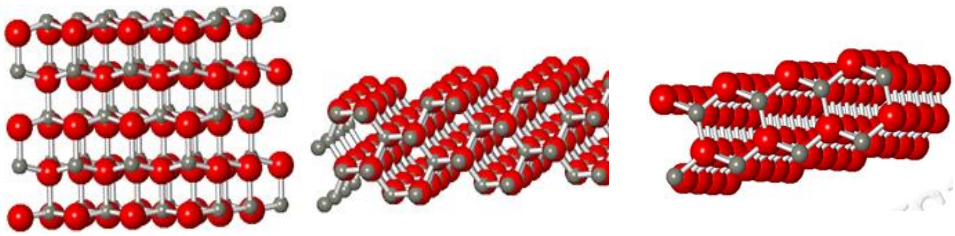
Table 1. ZnO bulk computational and experimental cell parameters.

CELL PARAMETERS (Å)		
	This thesis	GGA/ Experimental
a & b	3.30	3.25 [10]/ 3.23, 3.283[2]
c	5.31	5.21 [10]/ 5.27[2]
c/a	1.61	1.60 [10]/ 1.63[2]

The band gap obtained for the bulk is 0.7 eV, clearly underestimated at this level of theory but a result expected for GGA optimisation; other computational studies at the same level of theory are around 0.74 eV. [10]

2.3. Surfaces

Next step for the thesis was the building and optimization of three different surfaces geometries for ZnO wurtzite type structure, (100), (001) and ($1\bar{1}0$). These surfaces were found in the literature to be the most stable ones, [10–12] and therefore the most adequate candidates to be exposed in nanoparticles or other reactive media. The geometry of these three surfaces are showed in Figure 2a, 2b and 2c.



a. Surface (001) **b. Surface (100)** **c. Surface ($1\bar{1}0$)**

Figure 2. General representation of the three different surfaces of ZnO studied.

The work done for the ZnO surfaces was focused in the study of the total energy of the computed system, the calculation of the surface formation energy and band gap for each face. Since the slab can be formed by the repetition of unit cells normal to the surface, we explored the effect of increasing the number of those unit cells (it is, the thickness of the slab); for the surface (001), the surface unit cell; was (2x1), for the other two surfaces, the surface unit cell repeated was (1x1). The effect of varying the vacuum (d_v) gap between the layers in calculations was also explored to ensure there was no interaction between adjacent slabs (i.e., between top surface and the bottom surface of two consecutive slabs). Vacuum regions (d_v) of 6 to 15 Å in case of (001) face were explored, results showed in Table 2. In Figure 3 is represented a scheme of one surface of 3 slabs.

Table 2. DFT optimization results for (001) face.

	d_v (Å)	E_{total} (Ry)/slab	ΔE band gap (eV)	$E_{surface\ formation}$ (J/m^2)
single slab	6	-571.101	1.37	0.5139
	8	-571.101	1.37	0.5167
	10	-571.101	1.4	0.5157
	12	-571.101	1.39	0.5174
	15	-571.101	1.4	0.5170
double slab	6	-571.109	1.2	0.4286
	8	-571.109	1.21	0.4285
	10	-571.109	1.2	0.4301
	12	-571.109	1.19	0.4306
	15	-571.109	1.19	0.4309
triple slab	6	-571.110	metallic	0.4130
	8	-571.111	1.08	0.4010
	10	-571.111	1.14	0.3996
	12	-571.111	1.1	0.4031
	15	-571.111	1.11	0.4019

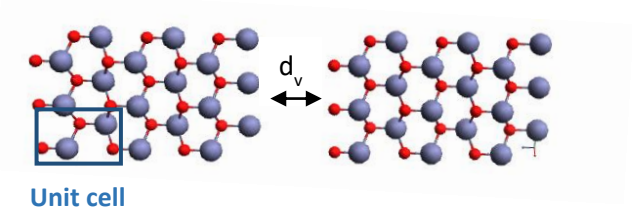


Figure 3. Representation of (001) surface geometry with three slabs. In blue the initial unit cell of ZnO bulk

A Similar work was done for slabs on the (100) structure. However, after the findings for (001) surface, and single and double slabs in the (001) surface, it was concluded that a vacuum gap of 12 Å was enough to avoid any slab interactions (even 10 Å have given almost the same results). As a consequence, for the triple repetition (100) slab, only the calculation of 12 Å vacuum was performed. The results for (100) face are showed in Table 3 and for the face (11 0) in Table 4.

Table 3. DFT optimization results for (100) face.

	d_v (Å)	E_{total} (Ry)/slab	ΔE band gap (eV)	$E_{surface\ formation}$ (J/m^2)
single slab	6	-1142.102	1.80	0.684
	8	-1142.101	1.85	0.686
	12	-1142.100	2.35	0.688
	15	-1142.100	3.35	0.689
double slab	6	-1142.174	1.90	0.846
	8	-1142.173	1.41	0.853
	12	-1142.173	1.42	0.853
	15	-1142.173	1.42	0.853
triple slab	12	-1142.213	2.25	0.853

Table 4. DFT optimization results for $(1\bar{1}0)$ face

	d_v (Å)	E_{total} (Ry)/slab	ΔE band gap (eV)	$E_{surface\ formation}$ (J/m^2)
single slab	6	-1,142.174	3.95	0.732
	8	-1,142.174	3.05	0.734
	12	-1,142.173	2.06	0.735
	15	-1,142.173	2.3	0.736
double slab	12	-1,142.226	1.3	0.816
triple slab	12	-1,142.248	2.2	0.811

In all cases, the thicker the slab, the more stable is the surface, because of the presence of more saturated atoms, less presence of dangling bonds in average, and also because the dangling bonds are further on the two exposed surfaces. We also want to mention that the small band gaps are also

an effect of the use of GGA functionals, and this effect will be corrected in the next section.

Comparing results of $E_{surface\ formation}$ between Table 2, 3 and 4 structures, the different surface (001) geometries are the most favourable to growth because of the lowest values in surface formation energy. The surface (100) is much less stable than the (001) and (1$\bar{1}$0). As a consequence, these two last ones will be predominant in nanomaterials. If we compare three structures of exposed surface with the same vacuum distance, we obtain the graphic of 4, where the stability of the surfaces can be observed easily.

Results of optimized surfaces by DFT (B3LYP functional) found in bibliography are for the face (100) for slabs from 12 to 18 layers with values around 1.3 to 1.4 J/m^2. [13] It is also noteworthy they report values for band gap in surfaces in the range of 2.8 to 3.1 eV using this hybrid functional. On the other hand, for the same surface, we have found values in literature from 0.84 to 0.91 J/m^2 from different works using GGA functional, PBE, 1.58 J/m^2 using LDA or 0.94 J/m^2 with GGA-PBE+U. [14] Then our results are in good agreement with bibliography.

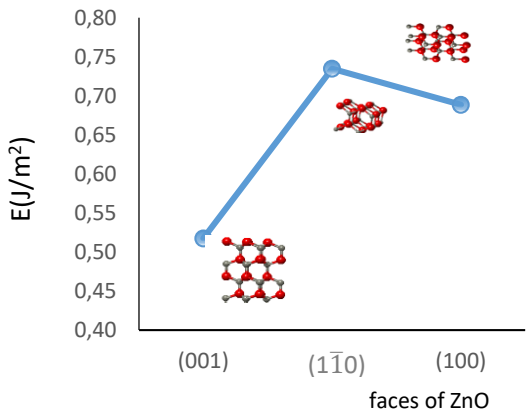


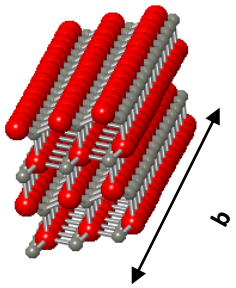
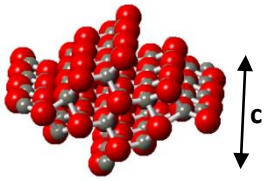
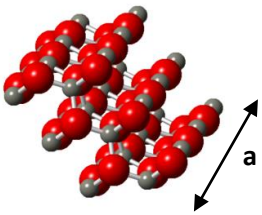
Figure 4. Surface formation energy for double slabs.

2.4. Nanotubes of ZnO, 1D structures.

The ZnO nanotubes were constructed from wurtzite ZnO with optimized lattice parameters and d_v of 10 Å. In the case of nanotubes, and after the analysis of the energy interaction in slabs, we decided to use 10 Å instead of 12 Å because results were converged and it also helps to save

computational time. The model started from cutting a 3x3x3 unit cell with plane directions of interest and then expanding the minimum repetition unit cell until double or triple size a, b or c as required. With this, we obtained 3 different combinations of faces, and in consequence, 3 different geometries of nanotubes. Nanotubes type 1 are combining the faces (100) & (001), nanotubes type 2 are combining the faces (100) & $(1\bar{1}0)$ and finally nanotubes type 3 are combining (001) & $(1\bar{1}0)$ face geometries. General view of each type of NTB are in Table 5.

Table 5. General representation of the three different types of nanotubes studied with the direction expanded.

Type 1	Type 2	Type 3
		

The edge formation energies and the band gap of the nanotubes were evaluated to study their variation depending of the geometry used. Results of these studies are summarised in Table 6

Table 6. Nanotubes results.

		Number of atoms	a (Å)	b (Å)	c (Å)	ΔE band gap (eV)	E edges formation (eV/ nm)
Type 1	**NTB1**	24	9.75	3.25	10.41	1.12	0.42
	NTB2	72	9.75	9.75	16.42	0.99	11.50
Type 2	**NTB1**	16	6.60	6.60	5.33	1.55	0.50
	NTB2	32	6.60	6.60	10.66	1.55	0.73
	NTB3	36	9.91	9.91	5.33	0.00	2.19
	NTB4	108	9.91	9.91	15.99	1.45	2.14
Type 3	**NTB1**	16	3.30	3.30	5.29	1.80	1.98
	NTB2	36	3.25	9.46	15.74	1.60	4.12
	NTB3	32	6.60	6.60	10.57	1.80	2.13
	NTB4	108	9.91	9.90	15.74	0.85	4.98

During this thesis we have tried to find some bibliography related to the study of the formation energies of ZnO nanotubes or similar results to compare, but it has been tricky and few results have been found. For nanotubes type 2, with different faces orientation, Wilson *et al.* [4] gave results between 0.71 and 0.93 eV/nm, which are in very good agreement with values computed in this work. The rest of results are in this order of magnitude or slightly higher values. Taking into account the reconstruction of surfaces for different size of the surfaces in the nanotubes, these E_{edge} variations can be explained according to this effect.

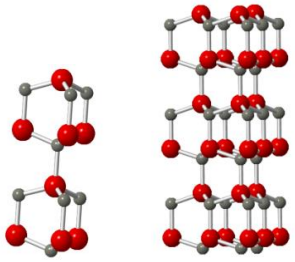
About the band gap for nanotubes here showed, Band gaps of NTBs here presented are in the same order of magnitude than in bulk calculations, because is the same degree of theory.

2.5. Nanoparticles of ZnO, 0D structures.

For the study of the 0 D structures, several types of nanoparticles (NP) were built. Non-spherical nanoparticles were built from the basis unit cell of wurzite type and cutting until the number of atoms desired. On the other hand, the spherical nanoparticles were built with the Nanocut software. In Table 7, the parameters of non-spherical nanoparticles and the band gap (eV) obtained after their optimisation. In Figure 5 are shown each non-spherical NP. In Table 8 we display the parameters of spherical NP and the band gap (eV), finally we can observe in Figure 6 the different views for each one.

Table 7. Parameters and optimised results for non-spherical nanoparticles

Parameters	NP5	NP6
a (Å)	3.676	8.728
c (Å)	3.343	6.587
Number of atoms	16	54
ΔE Band gap (eV)	metallic	1.0

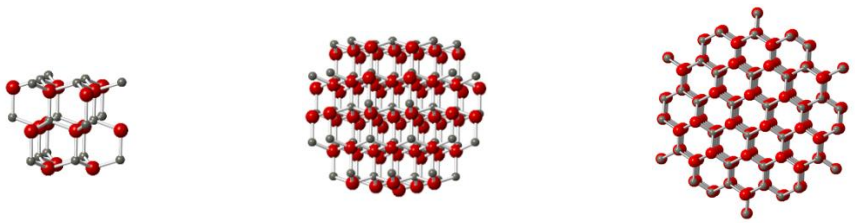


NP5 NP6

Figure 5. Representation of non-spherical nanoparticles, NP5 and NP6, optimized.

Table 8. Parameters for spherical nanoparticles

Parameters	NP10	NP15	NP20
d (nm)	1.0	1.5	2.0
Number of atoms	33	16	327
ΔE Band gap (eV)	1.2	0.25	0.52



NP10 NP15 NP20

Figure 6. Representation for each type of spherical nanoparticles optimized NP10, NP15 and NP20.

In the case, of NP we have obtained small results of band gaps given the quantity of dangling bonds and small peaks in the density of state representation. In the case of spherical nanoparticles, it is not possible to calculate the nanoparticle formation energy given that the quantity of Zn and O are not stechiometric.

3. ZnO at DFT+U level

As we mentioned in the Introduction chapter, one of the porpoise of this thesis is to get comparable computational results of band gap and solubility properties with the experimental results found.

While satisfactorily reproducing structural parameters with DFT at GGA level, in some cases electronic properties substantially underestimate the bandgap of semiconductor oxides, with ZnO bulk representing a dramatic case. Due to the limitation of the band gap calculation in this type of metal oxides with the pure DFT GGA functional, we can introduce the Hubbard corrections to the GGA, GGA+U, for improving the bandgap determination.

The on-site Coulomb correlation energy Hubbard-type interactions are included through the Hubbard parameters U and J in our calculations, the values used were 10.5 and 7.0 (eV) respectively for the metal centre. [15,16] The electronic structure of O atoms was not corrected.

In order to obtain better band gap results, we ran the optimized geometry structures in DFT with DFT+U correction. Results are shown in Table 9:

Table 9. Resume of DFT+U results obtained.

		Band gap Average (eV)	
		DFT	DFT+U
(001)	single slab	1.39	3.58
	double slab	1.20	3.46
	triple slab	1.11	3.42
(100)	single slab	1.42	5.60
	double slab	1.42	4.05
	triple slab	2.25	3.40
$(1\bar{1}0)$	single slab	2.68	4.10
	double slab	1.30	4.25
	triple slab	2.20	3.25

As a difference with the band gap determination by DFT, no metallic results appeared. These results are in good agreement with similar results published either theoretical of 2.8-3.1 eV for surfaces, [13] 3.36 eV [16] or 3.3-3.4 theoretical and experimental [17,18] for ZnO bulk. In such a manner, results more similar to the experimental ones were achieved. Since the goal is to simulate computationally this parameter as a property for toxicity evaluation, we need

as much accuracy as possible when comparing with experimental results for further QSAR models development.

The optimisation of NP with this functional was done using then so much time for running calculations, the results obtained are shown in Table 10. As in the case of surfaces, the band gap obtained is better for smaller NP, the NP5, but for the rest of them is far of the experimental value for the bulk.

4. DFTB, Optimization with Density functional Tight-Binding methodology.

The next step in this thesis it was to implement DFTB for our simulations. The reason to use this method was because with pure DFT (or even worst with DFT+U) there are limitations in time consuming resources and with the capacity of the same resources. This means that for structures with more than 200 atoms, it is necessary lots of CPU time and memory capacity for the optimization of geometries and band gap calculations. The only way to achieve such results for bigger NP is running calculation in High Performance Computing (HPC) systems, where each calculation needs 256 cores or more and many days.

An alternative to use HPC systems is DFTB methodology. The use of DFTB let us to increase the size of nanoparticles and run optimizations faster and with accurate results. [19,20]. The results obtained comparing DFTB with DFT+U are showed in Table 10 (DFT results are neglected because of the poor results showed).

Table 10. NP results comparison between DFTB and DFT+U. Band gaps in eV.

	Name of NP	DFTB Band gap	DFT+U Band gap	Number of atoms
NP no spherical	NP5	3.03	2.95	16
	NP6	1.35	1.8	54
	NP8	3.11	0.54	568
NP spherical	NP10	2.78	0.85	33
	NP15	2.12	0.29	144
	NP20	3.61	0.65	327

The results showed in Table 10 for the DFT+U techniques for bigger sizes in both types of nanoparticles are far of the band gap found in literature

for ZnO NPs, as it can be seen in Table 1 of chapter 6. However, this effect is not happening for the DFTB simulations, which are around the values expected from experiments. So, after this work it is shown that DFTB has some advantages for band gap calculations respect the pure DFT and DFT+U in terms of results, as well as for the time consuming for running the calculations and the resources needed for that porpoise.

7. Bibliography

1 A. Kołodziejczak-Radzimska and T. Jesionowski, *Materials (Basel).*, 2014, **7**, 2833–2881.

2 S. Haffad, G. Cicero and M. Samah, in *Energy Procedia*, 2011, **10**, 128–137.

3 H. Xu, F. Zhan, A. L. Rosa, T. Frauenheim and R. Q. Zhang, *Solid State Commun.*, 2008, **148**, 534–537.

4 H. F. Wilson, C. Tang and A. S. Barnard, *J. Phys. Chem. C*, 2016, **120**, 9498–9505.

5 H. Morkoç and Ü. Özgür, *Zinc Oxide: Fundamentals, Materials and Device Technology*, Wiley-Vch, 2009, ISBN: 978-3-527-40813-9.

6 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *J. Phys. Condens. Matter*, 2009, **21**, 395502.

7 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu and S. Baroni, *J. Phys. Condens. Matter*, 2017, **29**, 465901.

8 N. M. Harrison, Computational Materials Science, 2003, **187**, 45-70 .

9 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminform.*, 2012, **4**, 17.

10 B. Meyer and D. Marx, *Phys. Rev. B*, 2003, **67**, 35403.

11 A. Wander and N. Harrison, *An ab-initio study of ZnO(1120)*, 2000, **468**, 851-855.

12 J. E. Jaffe, N. M. Harrison and A. C. Hess, *Phys. Rev. B*, 1994, **49**, 11153–11158.

13 N. L. Marana, V. M. Longo, E. Longo, J. B. L. Martins and J. R. Sambrano, *J. Phys. Chem. A*, 2008, **112**, 8958–8963.

14 C. Tang, H. F. Wilson, M. J. S. Spencer and A. S. Barnard, *Phys. Chem. Chem. Phys.*, 2015, **17**, 27683–27689.

15 R. M. Sheetz, I. Ponomareva, E. Richter, A. N. Andriotis and M. Menon,

Phys. Rev. B, 2009, **80**, 195314-4

16 H. C. Wu, Y. C. Peng and T. P. Shen, *Materials (Basel).*, 2012, **5**, 2088–2100.

17 E. S. Goh, J. W. Mah and T. L. Yoon, *Comput. Mater. Sci.*, 2017, **138**, 111–116.

18 A. Consiglio and Z. Tian, *Sci. Rep.*, 2016, **6**, 1–7.

19 N. H. Moreira, G. Dolgonos, B. Aradi, A. L. Da Rosa and T. Frauenheim, *J. Chem. Theory Comput.*, 2009, **5**, 605–614.

20 L. T. H. L. Nguyen Viet Minh, Vu Ngoc Tuoc, *Commun. Phys.*, 2011, **21**, 235–243.

Chapter 4. TiO_2, validation case.

1. Introduction

TiO_2 has been commercially manufactured by millions of tons to be widely utilized such as pigment, paint additive, in sunscreen, due to its photostability and light dispersion, yet simultaneously strong UV light filtering, also serving as the electron transport scaffold in photovoltaic applications due to its electrical properties and for other advanced applications such as sensing, biomedicine, and electronics. [1,2]

Solid TiO_2 presents in nature three common atomic arrangements: the anatase, which has a tetragonal crystallographic structure with space group $I4_{1/amd}$, brookite, which is orthorhombic with space group P_{bca} and rutile with tetragonal phase and the space group is $P4_{2/mnm}$. All aforementioned applications of TiO_2 are possible thanks to the large band gap of 3.2, 3.02, and 2.96 eV for the anatase, rutile and brookite phases, respectively. [1] Rutile is the

most stable structure in nature and in most temperatures and pressures, and for nanoparticles with sizes bigger than 14nm,[1] though anatase is very close in energy and it is also easily observed.

This thesis is centred in anatase phase and the derived nanoparticles from this structure, because it is preferred over other polymorphs for its high application in technology fields given its large band gap (better photocatalyst than rutile)[1] and higher chemical reactivity. [3,4] Besides, anatase structure is preferred for nanostructures of TiO_2 in the range of size that we are capable of simulating. [5,6]

The anatase phase has a tetragonal structure with a unit cell of 12 atoms and lattice parameters a = b = 3.7842 Å and c = 9.5146 Å, α, β, γ = 90º. The basis cell of TiO_2 anatase phase is shown in Figure 1, where blue balls represent Ti and red O.

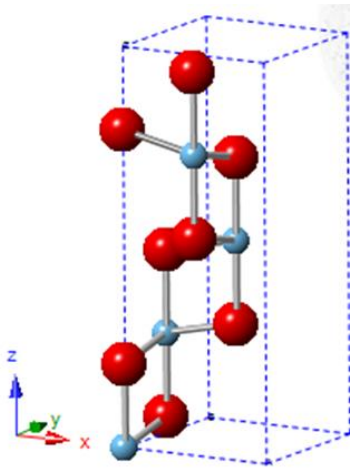


Figure 1. Tetragonal cell of anatase phase. In blue Ti atoms, in red O atoms.

As previously commented, the objective in studying TiO_2 is to confirm the good agreement of results in this thesis in band gap determination and solubility by *in silico* methods with existing experimental results, trying to avoid more expensive experimental work and data dispersion given the experimental conditions. TiO_2 anatase phase has been optimised by DFT at Gradient Generalised Approximation (GGA) level, as well as it was done with the ZnO; later with DFT+U to obtain improved results for the electronic band gap; finally, DFTB was used in order to compare results of this quasi pure DFT with the DFT and DFT+U electronic band gap values in case of TiO_2 as in ZnO.

This was done given that the time consuming and resources needed for DFTB are much scarce in comparison with DFT and DFT+U.

2. DFT and DFT+U Calculations

2.1. Computational details

Density functional theory calculations are conducted using the Quantum espresso (QE) program. [7,8] The self-consistent calculations were performed under the (GGA) with the functional PBE. The effect of the core electrons on the valence electron density was described by the projector augmented wave (PAW) method. The cut-off for the kinetic energy of the plane-waves has been set to 90 Ry throughout, which ensures a total energy convergence better than 10^{-6} eV. A Gaussian smearing technique with a 0.005 eV width has been applied to enhance convergence but all energies presented in the following have been obtained by extrapolating to zero smearing (0K).

For the DFT+U calculations the on-site Coulomb correlation energy Hubbard-type interactions are included through the Hubbard parameters U and J for Ti atoms, the values used were 4.0 and 4.2, respectively. [9,10]

For DFTB calculations, Self-Consistent-Charge calculations were performed with electronic convergence energy of 10^{-6} eV. The Sloter-Koster files tiorg-0-1 for Ti and mio-0-1 for O used in current simulations were obtained from DFTB+ webpage distribution (http://www.dftb.org/parameters/download/).

2.2. Bulk optimization

Table 1 summarizes optimized parameters for anatase bulk phase at DFT and DFT+U levels. For comparison, we also include literature values both at DFT(+U) and at experimental level values.

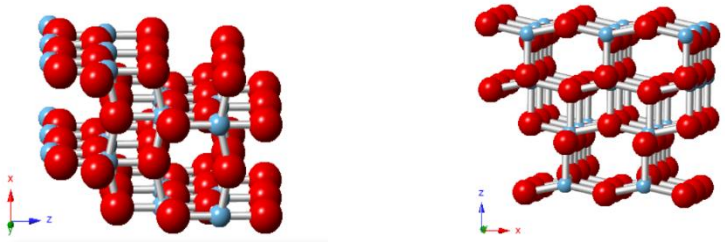
Table 1. Optimized cell parameters and band gap.

	DFT (this work)	DFT+U (this work)	DFT	DFT+U	Experimental[1]
a & b (Å)	3.801	3.827	3.78,[11] 3.737[10]	3.809, 3.795[10]	3.784
c (Å)	9.737	9.646	9.777,[11] 9.721[10]	9.641, 9.595[10]	9.515
Band gap (eV)	2.35	3.15	2.27,[12] 2.11[10]	3.03, 3.35[10]	3.20

Our results at DFT+U level of theory are very close to experimental parameters, both for lattice vectors and for band gap. They also compare extremely well with previously published DFT parameters, being differences in the range of 0.03 Å and 0.2 eV for lattice parameters and energy band gaps, respectively.

2.3. Results for Anatase surfaces at DFT level

The following step for the thesis was, as it was done for ZnO, the optimization of two surface geometries, the (100) because is considered the most stable for anatase phase given the high surface free energy and the (001) is the next surface more stable, for TiO_2 anatase, [2] as can be found in bibliography and the results obtained will show. The geometries of the surfaces are displayed in Figure 2, exposing both faces to vacuum. Surfaces were represented by periodically repeated slabs consisting of two atomic layers and separated by the vacuum. Simulation supercells in our calculations corresponded to the 3x1 surface unit cell and to the 3x3 surface unit cell, for the (100) and (001) surfaces, respectively.



a) Surface (100) **b) Surface (001)**

Figure 2. Representation of the surface structures studied.

The work done with TiO_2 surfaces was centered in the study of the surface formation energy (E_{surf}) of the computed surfaces and band gaps. After a set of simulations with different distances of vacuum (d_v), we present results for 10 Å gap between consecutive layers. This separation showed up to be enough to ensure there was no interaction between adjacent slabs, from the top of a slab and the bottom of the periodic slab image.

E_{surf} for the (100) surface was 0.27 J/m^2 (see Figure 2a). Although larger, this result is in good agreement with the 0.58 J/m^2 found in bibliography for the work of Vittadini *et al.* [13]. The discrepancy is possibly given by the use of larger unit cell larger than in the referenced work. This gives our systems more degrees of freedom for stabilitzation in the observed small surface reconstruction. For the case of the (001) surface (structure Figure 2b) the computed value is 0.52 J/m^2. This result is in good agreement with the range of values showed in Table 2 of the work by Eduard Araujo-Lopez *et al.* for the same type of pseupotentials and DFT functionals. [10] Therefore, taking into account the presented results, it is clear to state that the surface (100) is more stable and less reactive, as previously observed in literature. [13]

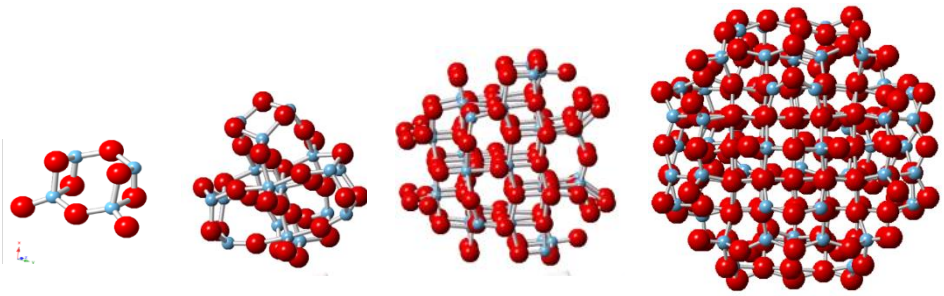
As you can state, the study of the TiO_2 anatase surfaces was not as deep as in the case of ZnO because this thesis is more centered in the NP optimization for the calculation of toxicity descriptors. The work for validating methodology was previously done with ZnO, and this part of the thesis is focused in the use of the validated methods for TiO_2.

3. TiO_2 Nanoparticles

For the study of the 0-D limit of TiO_2, a set of different kinds of nanoparticles were built as in the case of ZnO. The non-spherical nanoparticles were built from the basis unit cell of anatase type and cutting until the number of atoms desired. In Figure 3 there are represented the NPs calculated. It displays from the NP1, which is the smallest for TiO_2 cluster for non-spherical nanoparticles and the cluster was formed by 12 atoms in the stechiometric ratio $(TiO_2)_4$, until the spherical NP 4 of 261 atoms. For this last NP4 an alternative NP4a structure was optimized with stechiometric proportion of TiO_2 for the sake of comparison.

The spherical nanoparticles were built with the nanocut software. As in the case of ZnO, this type of structure did not keep the stechiometric ratio, and the formation energy of the NP respect to cannot be obtained. In Figure 3, there are the representative figures for the studied nanoparticles. All five NPs have been optimized at DFT, DFT+U and DFTB levels, but as it can be seen

in Table 2, for the case of NP 4 and NP4a it was not possible to run the calculations because of memory capacity necessity in the computational clusters at this sizes.



NP 1 **NP 2** **NP 3** **NP 4**

Figure 3. Representation of the TiO_2 nanoparticles studied.

Table 2. Results for the TiO_2 nanoparticles studied by the three methods of DFT. E_{form} is the formation energy in H/atom. Band gap in eV.

Nº of NP	Number of atoms	DFT		DFT+U		DFTB	
		E_{form}	Band gap	E_{form}	Band gap	E_{form}	Band gap
1	12	0.028	1.550	0.041	3.000	0.030	2.610
2	51	0.020	1.500	0.023	2.700	0.017	2.600
3	143	3.345	0.500	3.429	2.400	-0.075	1.140
4	261					0.031	metallic
4a	258					0.031	metallic

In Table 2 is observed that the band gap for DFT+U and DFTB methods are much closer than for DFT. Formation energy was computed from the bulk and per number of atoms. For non-stechiometric structures (NP3 and NP4), this value must be taken with great care, because the contribution of Ti and O atoms may be different in the different type of methods. NP3 presents an excess of 10% in O atoms, and NP4 an excess of 4% in Ti atoms. In the case of stechiometric structures (NP1, NP2 and NP4a), results for formation energy are consistent, and the trend that band gaps follows a decreasing line. In any case, we expect that for larger NP, the O excess or deficiency trends to 0,

because the largest part of the system will be the stechiometric core of the NP, and the surface contribution to formation energy per atom becomes stable.

Taking into account that DFT+U is the most reliable method, but with the heaviest calculation time, DFTB is a good alternative for large NPs band gap evaluation. As a general trend, it is also observed that the bigger the NP, the smaller the values of band gap, describing a more metallic behaviour of the TiO_2 NP.

These optimized values at DFTB level will be used in the following chapters to create a band gap model for prediction as it was done in the case of ZnO.

As a conclusion, and in accordance to the case of the ZnO, given the good results obtained with DFTB, this is the technique chosen to create more results for the descriptors used for QSAR modelling in Chapter 6 of this thesis and in the Chapter 5 for the solubility study. With this, we have validated this methodology of combining DFT methods to improve the quality of our results, and computing systems beyond the limits of classic DFT.

4. Bibliography

1 S. M. Gupta and M. Tripathi, *Chinese Sci. Bull.*, 2011, **56**, 1639–1657.
2 L. Sang, Y. Zhao and C. Burda, *Chem. Rev.*, 2014, **114**, 9283-9318.
3 M. Posternak, A. Baldereschi and B. Delley, *J. Phys. Chem. C*, 2009, **113**, 15862–15867.
4 C. Dette, M. A. Pérez-Osorio, C. S. Kley, P. Punke, C. E. Patrick, P. Jacobson, F. Giustino, S. J. Jung and K. Kern, *Nano Lett.*, 2014, **14**, 6533–6538.
5 D. Wen, Y. Yinping, L. Hongxuan, F. Zhao, L. Ji, Q. Weilong, J. Chen and Z. Huidi, *RARE Met. Mater. Eng.*, 2012, **41**, 263–267.
6 O. Van Overschelde, G. Guisbiers and M. Wautelet, *J. Phys. Chem. C*, 2009, **113**, 15343–15345.
7 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu and S. Baroni, *J. Phys. Condens. Matter*, 2017, **29**, 465901.
8 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *J. Phys. Condens. Matter*, 2009, **21**, 395502.
9 B. J. Morgan and G. W. Watson, *J. Phys. Chem. C*, 2010, **114**, 2321–2328.
10 E. Araujo-Lopez, L. A. Varilla, N. Seriani and J. A. Montoya, *Surf. Sci.*, 2016, **653**, 187–196.
11 D. Selli, G. Fazio and C. Di Valentin, *Catalysts*, 2017, **7**, 357.
12 Y. Yu, P. Zhang, L. Guo, Z. Chen, Q. Wu, Y. Ding, W. Zheng and Y. Cao, *J. Phys. Chem. C*, 2014, **118**, 12727–12733.
13 A. Vittadini, M. Casarin and A. Selloni, *Theor. Chem. Acc.*, 2007, **117**, 663–671.

Chapter 5. Solubility

1. Introduction

This chapter is dedicated to the determination of the solubility of metal oxides nanoparticles (in particular ZnO and TiO_2 were the centre of our work) in aqueous environment. The work started with ZnO NP in deep detail, determining the most reliable method as a function of composition and NP size. Later on, the experience acquired in the work with ZnO was extrapolated to TiO_2 NP in order to standardize the methodology used, and as a confirmation of the fruitful strategy in the evaluation of NP solubility.

One extra skill introduced in this chapter is the creation of QR codes to facility the observation of the molecular dynamics simulations. The visualization of molecular dynamics is a powerful tool in understanding chemical processes, and in particular, the chemistry happening on NP surfaces

and the interaction with water molecules. The QR code is directly linked to a youtube channel created for this purpose, the caption of the pictures related to the MD simulations are direct links to the simulation too.

As stated in the introduction of this thesis, one key parameter for the toxicology evaluation of nanoparticle is the solubility in the media. In this chapter, a new methodology for the solubility estimation of metal oxide nanoparticles based on MD simulations and *in silico* methods is presented. With the use of MD simulations coupled with DFTB, which can incorporate the direct description of liquid water, we can go a step further in the field of toxicology and calculate physical and chemical properties that can be subsequently used as descriptors for Quantitative Structure-Activity Relationship (QSAR) toxicology models.

Current work focuses on the computational prediction of ZnO solubility as a first step to establish a purely computational workflow for *in silico* nanotoxicity assessment of metal oxides based on molecular dynamic (MD) simulations. The development and improvement of electronic structure calculations methods based on quantum mechanics, as is the case of the Density Functional Tight Binding (DFTB) theory, and the ever-growing computational capabilities in supercomputing in the last years, makes it possible nowadays to simulate dynamics models that were prohibitive a few years ago. As a step further, we have extrapolated the solubility estimation method used for ZnO to TiO_2, in order to validate the methodology based on pure *in silico* data obtained from DFTB simulations and using non-extensive thermodynamic models.

This chapter is divided as follows. First, we present how to evaluate K_{sp} for small nanoparticles directly form MD simulations for ZnO NP. After this, results obtained using this method are shown and compared with experimental values. Finally, in order to evaluate the solubility of large nanoparticles, we couple this methodology with the use on non-extensive methods based on thermodynamic properties for ZnO and TiO_2 NP.

2. Solubility evaluation for nanoparticles up to 2 nm

2.1. Computational details for MD Simulations and K_{sp} evaluation

The solubility product constant (K_{sp}) is a well-defined value for an ionic solid (in general a binary salt: A_nB_m) and is well defined from the ions in solution: $K_{sp} = [A^{m+}]^n[B^{n+}]^m$. However, in non-ionic solids, this definition is more ambiguous since the solved product may not be an ion but a cluster or any

other general species interacting with water. In the case of ZnO (it can be the bulk or any large nanoparticles), the general dissolution equation is:

$$ZnO_{(s)} + H_2O_{(l)} \rightarrow ZnO_{(aq)} \quad (5.1)$$

Large ZnO nanoparticles, in solid state or already in solution, can react in water and become more dissolved in smaller nanoparticles or clusters as:

$$ZnO_{(large\ NP,\ s)} + H_2O_{(l)} \rightarrow ZnO_{(small\ NP,\ aq)} \quad (5.2)$$

Another way to define the K_{sp} is as the reactivity constant, K_r, of the previous reactions. From statistical thermodynamics, one can define the reactivity constant of a reaction as:

$$K_r = \prod_i Q_i e^{-\Delta E/RT} \quad (5.3)$$

where Q_i are the partition functions of products and reactants, and ΔE is the free energy exchange in the dissolution reaction.[1] In this work, the Q was approximated as 1, because in the dissolution reaction, the most important factor contributing to the partition function is the entropy of the liquid phases, both in the reactants and the products. Even if this value is not close to one, it will not be the predominant factor in the reaction constant and will barely change the order of magnitude or the K_r, since the exponential energy related factor is the most predominant in determining the order of magnitude of K_r.

MD simulations were performed to model the solution of ZnO in water. Since liquid water is a reactive dynamic system, the use of quantum-based MD simulations is mandatory at finite temperatures. In order to compute interatomic interactions, we made use of the Density Functional Tight Binding (DFTB) approximation as implemented in the DFTB+ code. This code uses Verlet's algorithm to integrate the classical Newton's equations of motion. A time step of 1.0 femtosecond (fs) was used for the integration. Simulations were performed in the canonical ensemble (at constant temperature, volume, and number of atoms, also named as NVT ensemble) using the Nosé-Hover thermostat, with a frequency of oscillation of the thermostating particles of 4000 cm^{-1}. A convergence criterion of 10^{-4} eV in the energies of the k-point was used on the Self Consistent Charge (SCC) cycles. For charged systems, the Ewald summation technique was used to account for the periodic dipole images of the super unit cells, where DFTB+ automatically detects the optimum summation value for the computed systems. [2]

Simulations were performed at nominal room temperature (300 K) for all systems. Simulations were run for 10 picoseconds (ps) for most systems; 20

ps runs were performed on two reference systems to ensure the convergence of the energy and structural properties of the models. The average properties were obtained discarding the 2 first picoseconds of simulation to allow the system to equilibrate after the kinetic energy injection at the start of the simulation. Finally, Radial Distribution Functions (RDF) were used to follow the water and ZnO crystal structures during the simulations period.

As a starting point in this work, we simulated the reactants in equation (1) using the aforementioned methodology. The solid ZnO bulk model consisted of a 5x5x3 supercell (figure 1a) of the wurtzite-type structure (zincite) with the experimental lattice parameters of 3.25 and 5.20 Å. RDF of solid ZnO is reproduced in Figure 1b, clearly depicting the interatomic distances and the coordination of Zn and O atoms. The cubic water simulation boxes of nominal 20 (figure 2a) 25, and 30 Å were obtained using the water creation plugging in the VMD code, [3] which produces water boxes of randomly distributed water molecules with a formal density around 0.9-0.95 g·cm^{-3}.

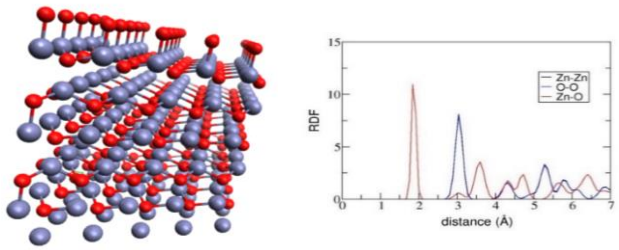


Figure 1. (a) 5x5x3 supercell ZnO wurtzite type. (b) Radial Distribution Function of the 5x5x3 unit cell of the wurzite-ZnO bulk at 300 K for 10 ps simulation. Zn-Zn and O-O curves overlap due to the crystal symmetry.

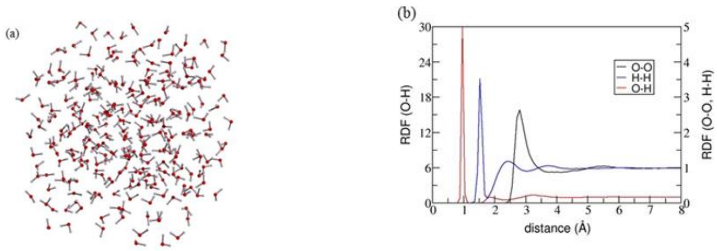


Figure 2. (a) Model of a cubic box of 268 water molecules (20 Å) for simulations at 300 K and DFTB level and (b) its corresponding Radial Distribution Function of water molecules for a 10 ps simulation. In (a) oxygen is red and hydrogen is light grey.

Water is a system that has been previously studied by means of SCC-DFTB simulations. The system is very special from the atomistic simulations point of view due to the H-bond interactions between molecules. These interactions represent a big challenge, not only at SCC-DFTB level, but also in normal DFT. DFT functionals (at LDA, GGA and meta-GGA levels) do not describe properly the long range coulombic and exchange interactions[4] limiting the description of structural and thermodynamic properties of liquid water.

The RDF corresponding to water bulk simulation is displayed in Figure 2(b). At 0.9 Å appears the peak assigned to the O–H bonds in water molecules, at 1.8 Å there is the peak of to the O···H contact of the H-bonds of the first solvation shell and at 3.2 Å is a smooth peak of the other H atoms of this first solvation shell. The RDF of O-O interactions shows a pronounced peak at 2.8 Å, corresponding to the first solvation shell, and then a broad peak at 5.5 Å, produced by an almost non-existing second coordination sphere. The integration of the first peak gives values of 2.3-3.6 water molecules, depending on how the first peak is deconvoluted, in accordance with results of 2.4-3.7 water molecules in the SCC-DFTB simulations of 128 water molecules.[5]

The average properties resulting from a water box simulation of 20 Å up to 20 ps were used to check the convergence of the energy and RDF structural parameters in 10 ps simulations. Average energies (Table 1) per water molecule differ only 1 thousandth of eV in the longer simulation respect to the 10 ps one. The RDF analysis of the extended simulation is very similar to the one of 10 ps; however, the maximum of the first peak of the O-O interaction moves down ~0.15 Å, indicating a slightly more compressed water system. This effect was also observed in longer simulations times (50 ps) of water with SCC-DFTB, where water created some voids in the water boxes.[6] Because we are mostly interested in the ZnO-water interactions to describe the solubility, and not the water itself, the simulation period of 10 ps is enough for non-reactive systems, avoiding also the formation of those voids in the water box.

Table 1. Computational data for water models. Size of boxes in Å; simulation time (τ) in ps; density (ρ) in g·cm^{-3}; number of water molecules (N); average energy (E) in hartrees per water molecule; computational time (t) per MD step in s (calculations were performed on a DELL PowerEdge R610 platform, with 32 threads Intel Xeon E5530 2.40GHZ and 90 GB of RAM).

Nominal size	Real size	τ	N	ρ	E	t
20	20.5	10	268	0.93	-4.0792	14.88
20	20.5	20	268	0.93	-4.0790	14.97
25	25.9	10	526	0.91	-4.0791	64.5
30	31.9	10	995	0.92	-4.0793	354.3

Studies of bulk water (in relatively small unit cells in comparison to our work, between 128-512 atoms) at SCC-DFTB level have shown that there is an overcoordination of water molecules in the first solvation shell, and loss of order at the second and third shells; this has as a consequence that, in variable volume cell simulations, one overestimates the water density and underestimates the heat of vaporisation of water dimers. [6,5,7] Some corrections have been introduced in SCC-DFTB to account for the H-bond interactions in what is called DFTB3 approximation, and although improving matters, this still lacks the right description of bulk water.[5,6] Despite these limitations in bulk water, the method has been applied quite successfully in studying the proton exchange reaction in water [6–8] or in anionic systems in water, [9] which still shows the possibilities of this method in complex reactions in solution.

Table 1 also presents the average energies computed for larger cubes of water (nominal edges of 25 and 30 Å), and differences only differ in the fourth decimal of the energy in Hartrees (10^{-4} H are ~2.7 meV), which is below the limit of precision of the DFTB+ energy evaluation for water boxes and proton transfers in water boxes.[6] Another important factor here is the scalability of the system. The computational time increase in our simulations is in the order of $N^{2.2\text{-}2.4}$ (N is the number of atoms); this is the classic behaviour of DFTB methodology.[8]

2.2. Zn^{2+}, O^{2-} and NPs in water solution

As a first model to describe the solubility of ZnO and its large nanoparticles, the fully dissolved system, i.e., the Zn^{2+} cation and O^{2-} anion in water boxes were simulated and taken as the final products of the solving process. Furthermore, as we know that in those systems, the oxygen anion is

especially unstable in water solutions, the simulation of the smallest ZnO cluster consisting of only two atoms was performed.

Figure 3 shows the RDF analysis of those species in solution. For the RDF of O^{2-} in the water box (Figure 3a) there is no substantial difference respect to Figure 2(b) of pure water, because the averaged effect of the O^{2-} in the total number of water molecules is tiny. However, the O^{2-} is tremendously reactive in water, and the immediate effect after starting the simulation was to start proton exchanges with other water molecules. This had as a result that the simulation ended being a box containing two solvated OH^- groups.

The case of Zn^{2+} is slightly different, since this cation is stabilized in water solutions due to the interaction with the lone-pair electrons of O-atoms of water. Beyond the features assigned to water (O-O, O-H and H-H contacts), the RDF of this system shows a clear Zn-O contact at 2.0 Å produced by a first solvation shell (Figure 3b), as well as a more diffused second solvation shell around 3.75 Å. The first peak integrates 6 water molecules, corresponding to an octahedral solvation shell around the Zn^{2+} cation. This same behaviour was observed both at DFTB and DFT level [10] in their tests calculations when studying the water/ZnO interfaces, or in QM/MM models when studying divalent cations in water. [11]

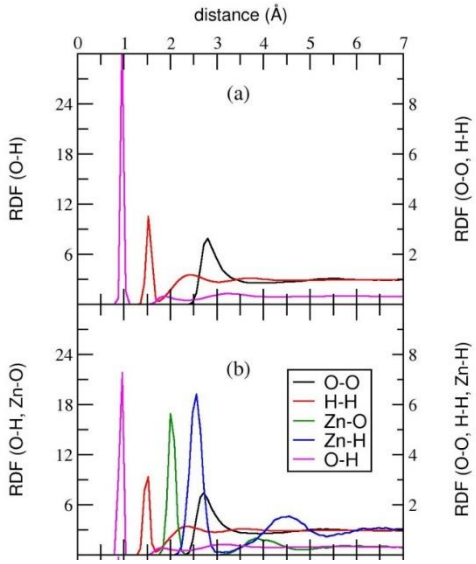


Figure 3. Radial Distribution Function of the (a) O^{2-}, (b) Zn^{2+} species in a water box of nominal 20 Å, at 300 K, for 10 ps simulation.

The first NP introduced in water was the NP1 model, i.e. a linear cluster formed only by a Zn and an O atom. However, as in the case of the O^{2-}, this system showed a high reactivity vs. water, and once the simulation started, it hydrolysed immediately and formed the neutral $Zn(OH)_2$ species. The RDF of this system is displayed in Figure 4**,** presents a first solvation shell of the Zn atom between 1.6-2.2 Å (Zn-O contact) with two peaks. The peak centred at 1.7 Å corresponds to the two Zn-OH contacts, and the one at 1.9 Å to three contacts to water molecules. This $Zn(OH)_2 \cdot (H_2O)_3$ species adopts a trigonal bipyramidal structure, with the two OH groups in the equatorial plane during all the simulation period. This structure is different to the $[Zn(H_2O)_6)]^{2+}$ species because there exists charge transfer between the OH^- groups to the Zn^{2+} centre and it can stabilize another symmetry structure. Moreover, Zn^{2+} is a metallocenter known to easily stabilize different coordination numbers and structures. [12]

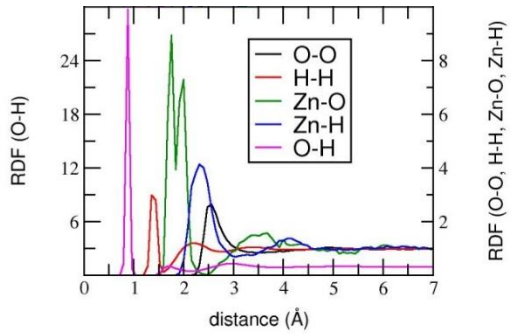


Figure 4. *Radial Distribution Function of NP1 specie in a water box of nominal 20 Å, at 300 K, for 10 ps simulation.*

Because ZnO is an ionic solid with high covalent contributions, the final dissolved structure is probably a cluster of ZnO groups. More clusters of $(ZnO)_n$ were modelled as possible final species in the dissolution process, where n=3, 6 and 9 (NP3, NP6, and NP9 particles). The initial simulation clusters and water boxes are shown in Figure 5, where it can be observed that the NP6 and NP9 are the stacking of several NP3 clusters, which are layers of the (0001) wurzite-ZnO planes.

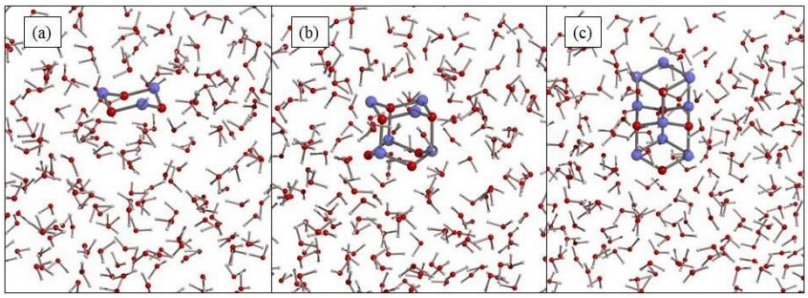


Figure 5. Models of (a) NP3, (b) NP6 and (c) NP9 in water boxes. Zinc atoms in electric blue, oxygen atoms in red and hydrogen atoms in light grey.

During simulation of these nanoparticles, especially in the case of NP6 and NP9, some of the surrounding water molecules hydrolysed. The OH groups linked to the Zn of the cluster and the protons of split water bonded to the O atoms of the nanoparticle. This reaction helps to stabilize the NP in the water environment, and therefore, increases the solubility of the ZnO.

In the RDF of these three clusters (Figure 6), the first peaks of ZnO cluster, corresponding to the Zn-Zn and Zn-O contacts, show a large similarity with those of ZnO bulk, being the first Zn-O contact ~2 Å, and the fist Zn-Zn ~3 Å. However, the second Zn-O peak is much more diffuse in the clusters, being the centre between 3.5-4 Å, indicating the liquid behaviour of the solution.

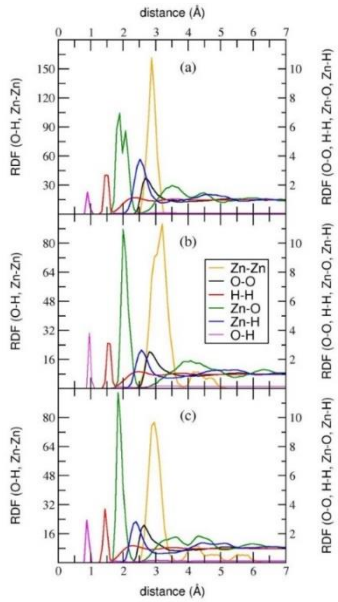


Figure 6. Radial Distribution Function of the (a) NP3, (b) NP6 and (c) NP9 in a water box of nominal 20 Å, at 300 K, for 10 ps simulation.

Finally, it is important to mention that the NP6 simulation was extended up to 20 ps to compare the structural and energetic evolution. Results show, as in the case of water, no significant differences exist and the energies are well-equilibrated after 10 ps of simulation for our purposes in a system where we do not expect any important chemical reactivity beyond the dissolution.

2.3. Spherical nanoparticles in water solution

Although the solubility of the bulk material is an important issue, the solubility of nanoparticles is also of crucial relevance. For this reason bigger NP were created, in order to evaluate the solubility using equation (2). First a NP of 15 Å diameter, NP15 and secondly a NP of 20 Å diameter, NP20 were simulated using the Nanocut software were built, both within the water box and without the water box, as models to determine if the proposed methodology is also applicable for nanoparticles. Figure 7 shows the initial structures of these nanoparticle simulations in water boxes of nominal 30 Å edges.

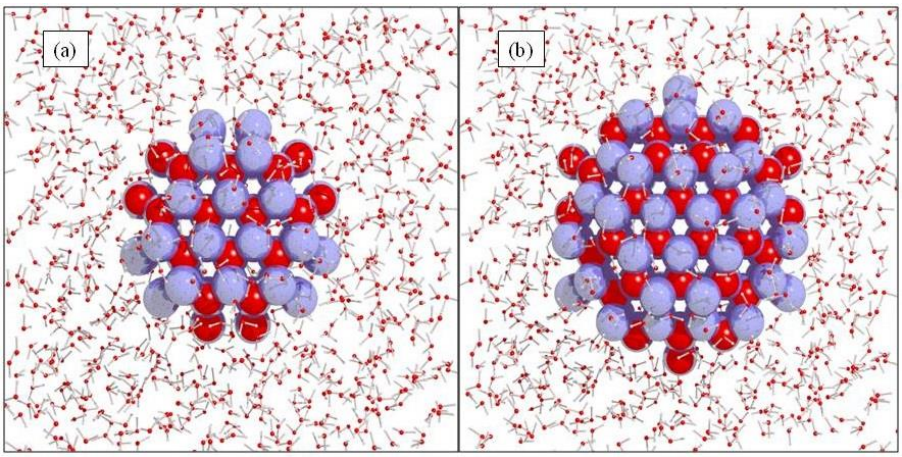


Figure 7. Models of (a) NP15, and (b) NP20. Zinc atoms in electric blue, oxygen atoms in red and hydrogen atoms in light grey.

These two nanoparticles expose a larger surface to water than previous small ZnO clusters, as well as edges and vertices that could be the

source of possible chemical reactions with water or other species present in biological solutions (not modelled here). Both NP15 and NP20 have shown to be stable during the 10 ps of simulation, with no disaggregation into smaller parts.

The RDF for NP15 and NP20 in a water box of 30 Å (Figure 8) are very similar to the respective RDF of NP6 and NP9, but with the peaks of Zn-Zn and Zn-O contacts of the solid nanoparticle present at longer distances, indicating that the internal structure of the nanoparticle conserves the longer range order of the solid. The Zn-O and Zn-H bands are very similar to those of NP6 and NP9 showing a similar mechanism of the water interaction with the ZnO surfaces, for small and larger nanoparticles.

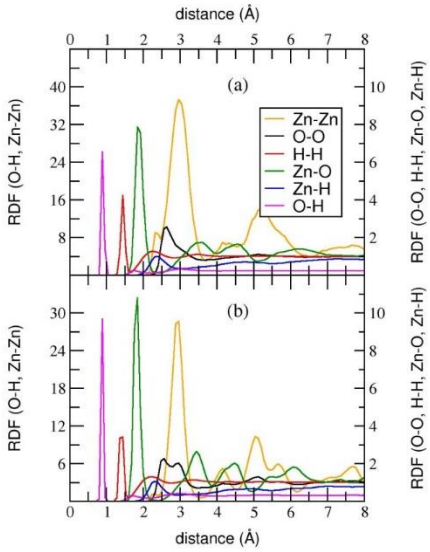


Figure 8. Radial Distribution Function of the (a) NP15, and (b) NP20 in a water box of nominal 30 Å, at 300 K, for 10 ps simulation.

2.4. Solubility evaluation of bulk and nanoparticles

Table 2 summarizes both the solubility and solubility constant product (as computed from equation (3)) for the different reaction models that we have proposed, both from solid bulk to dissolved nanoparticles, as well as from nanoparticles in free space to the corresponding nanoparticulated solution. The dissolution of ZnO bulk reaches maximum values in our NP6 and NP9 models as final products of the dissolution process of 0.2-0.5 $mg \cdot L^{-1}$. This value

differs by an order of magnitude from the experimental bulk solubility of 8-15 $mg \cdot L^{-1}$, [13,14] which is in excellent agreement in terms of a theoretical evaluation for this value. As commented in the methods section, due to the exponential factor in the K_r formula, any small change in the free energy evaluation of the system will have a tremendous effect in the results.

One has also to take into account that the solubility in experimental works is given as the equivalent $[Zn^{2+}]$ in the system, which can be in this form in solution or aggregated with other ZnO clusters and forming a large $(Zn_{x+1}O_x)^{2+}$ cation. This can explain why we find in literature values of solubility for ZnO bulk in the range 8-15 $mg \cdot L^{-1}$ of equivalent $[Zn^{2+}]$ in the system, [13,14] and at the same time a K_{sp} of $1.58 \cdot 10^{-17}$.[15] In order to obtain the solubility from the K_{sp} value, one has to assume a completely ionic model (as we have done in our values to obtain solubility from K_{sp}), and in the case of David *et al.* [15] that would correspond to a low solubility of only 0.3 $mg \cdot L^{-1}$, a large variation respect to other experimental evaluations.

The dissolution of ZnO bulk in other models (as the free ions in solution, to NP1 or to NP3) gives a much lower solubility, because of the instability of the final structures relative to the NP6 and NP9 models. In particular, the fully ionized system is highly unstable, therefore forming the incorrect conclusion that no Zn^{2+} or O^{2-} exist in solution. This is just a consequence of this model, because the high instability of O^{2-} in solution does not permit Zn^{2+} to appear in our models. In order to avoid this issue, we have studied another possibility. Since NP6 is the most stable small cluster we found, we decided to remove a Zn^{2+} atom from this structure, and take as the final results in ZnO bulk solution the system composed by solvated Zn^{2+} and solvated $(Zn_5O_6)^{2-}$. Results show that this ionic system is much more stable for ZnO bulk dissolution than the Zn^{2+} or O^{2-} mixture in solution (K_{sp} of $2.9 \cdot 10^{-19}$ vs $1.52 \cdot 10^{-31}$ for Zn^{2+}+ $(Zn_5O_6)^{2-}$ and Zn^{2+} or O^{2-}, respectively) to release ionic Zn in the systems; with this new model values of 45 $\mu g \cdot L^{-1}$ of free Zn^{2+} are found in solution.

The evaluation of the solubility of ZnO bulk has only taken as final steps, for the dissolution reaction in equation 1, the small clusters and we have not evaluated it for NP15 or NP20 as final dissolution products. The reason why we have not done that is simple. NP15 and NP20 were obtained using the nanocut software, imposing a restriction only in the diameter of the particle, but not imposing the stoichiometry of ZnO, and NP15 and NP20 do not possess the same number of Zn and O atoms (as shown in Table 2). This has as a consequence that free energy change of the reaction cannot be defined unambiguously, and the corresponding solubility of the bulk is not evaluable.

On the other hand, a key parameter would be the evaluation of the solubility of the nanoparticulated material. We have chosen as the final

products for the dissolution reaction the solvated NP15 and NP20, since smaller nanoparticles do not have the same Zn/O ratio. For initial products of the dissolution reaction we decided to choose the nanoparticles in the same conditions as in the water box, but with no water, i.e., the nanoparticles against the vacuum in free space and, of course, a water box separately. Results for this dissolution reaction (or maybe just a suspension reaction), are given in Table 2. K_{sp} is extremely high, in the order of 10^8. This means that NP15 and NP20 would dissolve completely in water.

Table 2. Computational data for ZnO nanoparticles and ionic Zn^{2+} and O^{2-} in water boxes. Nominal size of boxes in Å; N is the number of ZnO units in the cluster (for NP15 and NP20 in parenthesis the excess of Zn atoms); E_{sol} (in hartrees) are formation energies of the clusters per ZnO unity as defined in reaction (1) for ionic, NP1, NP3, NP6 and NP9, and E_{sol} (in hartrees) are the solution energies per ZnO units of reaction (1) from solid NP to aqueous NP for NP15 and NP20; [ZnO] is the nominal concentration of ZnO in the simulation boxes (g ZnO / g solution); K_{sp} as defined in equation (3) from the bulk to the corresponding NP for ionic, NP1, NP3, NP6 and NP9, and K_{sp} as defined in equation (3) from the solid NP to the aqueous NP for NP15 and NP20; Solubility (s) in $mg \cdot L^{-1}$.

Nanoparticle	Box size	N	E_{sol}	[ZnO]	K_{sp}	s
$Zn^{2+}+O^{2-}$	20	/	0.088	0.017	$4.97 \cdot 10^{-41}$	$5.73 \cdot 10^{-16}$
NP1	20	1	0.0674	0.017	$1.52 \cdot 10^{-31}$	$3.17 \cdot 10^{-11}$
NP3	20	3	0.0367	0.049	$1.74 \cdot 10^{-17}$	$3.39 \cdot 10^{-4}$
NP6	20	6	0.0247	0.095	$5.27 \cdot 10^{-12}$	0.187
NP6 (20 ps)	20	6	0.023	0.095	$3.21 \cdot 10^{-11}$	0.461
NP9	20	9	0.025	0.136	$3.57 \cdot 10^{-12}$	0.154
NP15	30	66(12)	-0.0177	0.278	$1.28 \cdot 10^{8}$	/
NP20	30	159(9)	-0.0177	0.484	$1.23 \cdot 10^{8}$	/
Expt (bulk)[a]					$1.58 \cdot 10^{-17}$	
Expt (bulk)[b]						~15
Expt (bulk)[c]						~8.45
Expt (NP 6nm)[15]					$8.13 \cdot 10^{-13}$	
Expt (NP)[13]						~15
Expt (NP 25-70 nm)[14]						~34.2

Experiments show that, although NPs are more soluble than ZnO bulk, they do not dissolve in such an extension completely. Although suspensions of 1 $g \cdot L^{-1}$ of ZnO NP can be obtained, it is not a complete dissolution,[16] and properties of the systems are different. The reason why we obtain such an

enormous value is because the choice of the initial reference system. Nanoparticles are not in free space, and this nanoparticulated material interacts with the adjacent nanoparticles in real systems, creating aggregated systems that are much more stable than our free space simulated NPs (NPs of 15-20 nm in diameter of ZnO have been observed to aggregate forming large systems of 0.15-1.5 µm, as observed by Li *et al.*[16] Therefore, the solubility of these NPs is more similar to the values of the bulk, than our values obtained from free space NPs. Although this last fact seems to be discouraging, the trends we have obtained are in the right direction, since we observe the qualitative solubility increase from the bulk to the nanoparticulated solid.

2.5. TiO_2 small NP in water solution.

TiO_2 bulk shows a small solubility in form of equilibriums quantified at level of trace in aqueous media at room temperature and pH>1. Perhaps this is the reason that only experimental publications determinate aqueous solubility of Ti or hydrous species of TiO_2. [17] In the work of Schmidt *et al.,* we can find the study of the solubility of anatase nanoparticles in 0.1 $mol \cdot L^{-1}$ NaCl respect time and different pH values; at pH = 3 the solubility has values around 10^{-3} $mol \cdot L^{-1}$ beyond the 500 h of experiment, and at pH=7 it is between 10^{-8} to 10^{-9} $mol \cdot L^{-1}$ for the same equilibration time. These results are modelled under the assumption that the total concentration of dissolved Titanium(IV) is distributed among mononuclear titanium hydroxo complexes of general formula $[Ti(OH)_n]^{(4-n)+}$. [17]

For experiments of 2h of assay, Avramescu *et al*. [18] give values of solubility for anatase bulk around $2 \cdot 10^{-9}$ $mol \cdot L^{-1}$ and for the anatase nanoparticles around $2.5 \cdot 10^{-9}$ at neutral pH, which are in the same order of magnitude as the aforementioned work

In this thesis, the K_{sp} of TiO_2 bulk has been evaluated as the same way of the ZnO from equation 1 (with the appropriate substitution of ZnO by TiO_2), using for MD simulations a small NP of 10 Å diameter of TiO_2. This NP was introduced in a cubic water box of 30 Å and the simulation was equilibrated and then ran for 10 ps. Using the equation 3 of this chapter, the K_{sp} for the anatase bulk obtained was $3.72 \cdot 10^{-18}$, which using an ionic solubility model gives a values of solubility of $9.76 \cdot 10^{-7}$; this result compares very well with the experimental values mentioned above in the work of Schmidt and Vogelsberger and the one of Avramescu *et al*., which are in the same order of magnitude and we had to use some assumptions in the equilibrium model, as the ionic components and the final particle for dissolution (a $(TiO_2)_4$ cluster). Figure 9 shows a 10 Å NP of TiO_2 mentioned above.

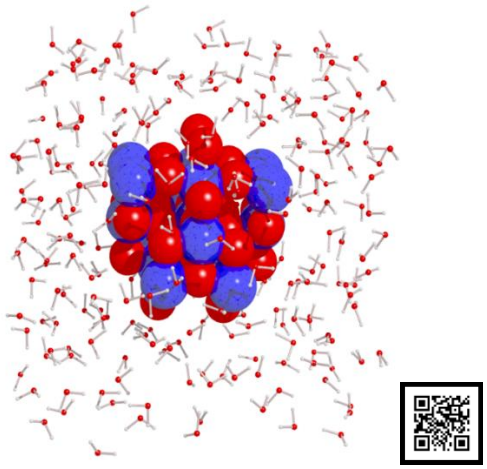


Figure 9. *10 Å diameter NP of TiO_2 inside a water cubic box*

3. Solubility evaluation for nanoparticles beyond 2 nm

After the limitations found in the methodology presented in the section 2 of this chapter, i.e. computational limitations in the size of particles, limitations in the solubility chemical models due to the break in the stochiometry and to the same ionic definition of solubility, an alternative methodology had to be searched. In the literature search, we found a possible solution in non-extensive thermodynamic models.

3.1. Non-extensive thermodynamic model: Oswalt-Freundlich equation

Given the limitation found in our first model for bigger nanoparticles due to the computational time consumption and computational limits of the simulations, we decided to make use of a new model based on the Oswalt-Freundlich approximation.[19] In this approximation, they proposed that the variation of concentration at saturation, $C_{i(sat)}$, of spherical particles of radius r, of a molecular solid i, in a solvent is described by the equation 5.4:

$$\ln \frac{C_{i(sat)}}{C^*_{i(sat)}} = \frac{V_i^*}{RT} \frac{2\gamma^{SL}}{r} \qquad (5.4)$$

$C^*_{i(sat)}$ is the concentration of i at saturation when the solid i is an unlimited phase, no size effect. V_i^* is the molar volume of i, T is the

temperatures, R the perfect gas constant and γ^{SL} is the solid-liquid interfacial tension or interaction energy.

This equation means that the decrease in particle radius evoques an increase in the concentration of i at saturation equilibrium. With this relation we can calculate, as is experimentally done, the solubility of the nanoparticles from the simulated γ^{SL} parameter from small nanoparticles of size known or for water-MeO interfaces in 2D.

3.2. Computational details to evaluate γ^{SL}

The solid-liquid interface for a fluid and planar surface has been analysed using Molecular dynamics (MD). [20] The approach used involves calculating the thermodynamic free energy difference between two systems to estimate the interfacial tension.

The method employed to evaluate the solubility has started with the definition of the equilibrium reaction for the ZnO in water, in our case we use the equilibrium equation described in this chapter (1), and the solubility is calculated by the equation:

$$s = \left(\frac{Ksp}{4}\right)^{1/3} \tag{5.5}$$

This relationship comes from the following equilibrium:

$$ZnO_{(s)} + H_2O_{(l)} \rightarrow Zn^{2+} + 2\ OH^-_{(aq)}$$

We decided to use this dissolution equilibrium instead of the pure ionic system (Zn^{2+} + O^{2-}), because we found in our simulations that O^{2-} was tremendously unstable in water and formed 2 OH^- anions.

The relation between interfacial tension and solubility is introduced by the Ostwald Freundlich equation[21]:

$$\frac{S}{S(bulk)} = \exp\left(\frac{4\gamma V}{RTd}\right) \tag{5.6}$$

Where S is the solubility (mol/kg) of spherical particles, d (m) diameter, gamma (mJ/m^2) is the interaction energy, V is the molar volume (m^3/mol), R gas constant (mJ/molK) and T(K) temperature. This equation is related to the K_{sp} of the bulk by the expression:

$$\ln K_{sp} = \ln K_{sp}^{bulk} + \frac{12\gamma V}{RTd} \tag{5.7}$$

The key part of this equation is the evaluation of interaction energy between the solid and the liquid γ^{SL}. As previously done in literature,[22] we

evaluated the γ^{SL} as the Helmholtz free energy of interaction between water and ZnO or TiO_2 surfaces. Since the γ^{SL} values depend on the exposed faces on the NP, and we want to study large NP, where the most exposed surfaces will be those that are more stable, we decided to compute γ^{SL} for stable surfaces using the slab approximation.

Computational method consists in the NVT (canonical ensemble) MD simulations using the Velocity Verlet algorithm in the DFTB+ programme. Simulations were run at 300K for 10 and 30 ps in the case of ZnO and 10ps for TiO_2. The method consists to run the MD simulation for the whole system (slab and water layer), and later on, we run the same MD calculation for bulk without water and separately the water without the bulk using the trajectories of the whole system. To do this, a new python code was developed. The final γ^{SL} is the difference of the total free energy of the system and the isolated parts.

The simulation model used for ZnO is displayed in Figure 10; it consists of a 5x5x2 ZnO slab, exposing the (001) surface enclosed by two layers of water containing 73 water molecules that represent a water density of ~1 $g \cdot cm^{-3}$ in the interslab region of periodic images. We chose this surface because it is the most stable one found in the study shown in the chapter 3. Scanning the QR code, it is possible to observe how the simulation evolves and how the water system is restructured due to the presence of the slab. One can clearly observe the formation of a first water layer on the surfaces followed by a more fluid layer in the inner part of water slab.

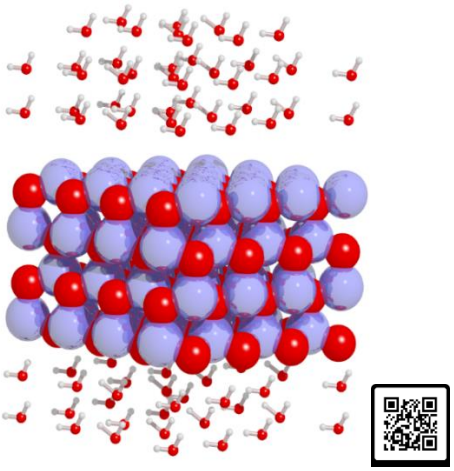


Figure 10. Initial structure of the slab of ZnO surrounded by water layer for periodic MD simulations.

For the evaluation of γ^{SL} in the TiO_2 case, we made use of a 3x3x2 TiO_2 slab, exposing the (100) surface, with 108 molecules of water that represent a water density of ~1 $g \cdot cm^{-3}$ in the interslab region of periodic images. This surface was chosen because in our study in chapter 4 was the most stable one Scanning the QR code, it is possible to observe how the simulation it evolves and how the water system is restructured due to the presence of the slab.

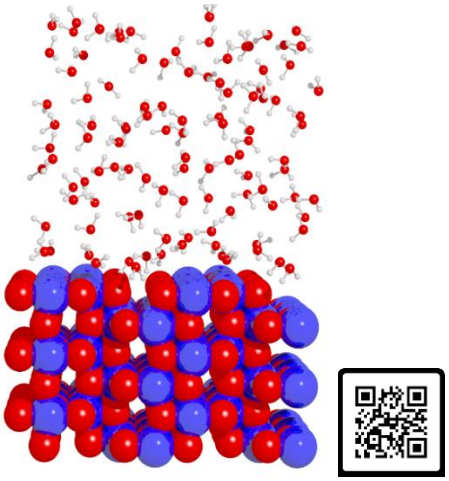


Figure 11. Structure of the slab of TiO_2 NP surrounded by a water layer for periodic MD simulations.

After simulations, γ^{SL} was computed as follows:

$$\gamma^{SL} = \frac{E_{MeO+H2O} - E_{MeO} - E_{H2O}}{A} \quad (5.8)$$

where $E_{MeO+H2O}$ is the average total energy of the system after 10 ps calculation, E_{MeO} is the average total energy of the MeO without water, and E_{H2O} is the average total Energy of the water layer without MeO. For the isolated systems, energy was evaluated in the same structures obtained in the MD trajectories of the total system. A is the area of the interface (2 interfaces existed in simulations per superunit cell).

In addition to the simulations previously mentioned, we also evaluated the interfacial force for ZnO with an extra layer of water, which is the double from Figure 10. This was done with the aim to check if the distance between slabs makes any variation to our γ^{SL} evaluation. The representation of this system can be seen in Figure 12.

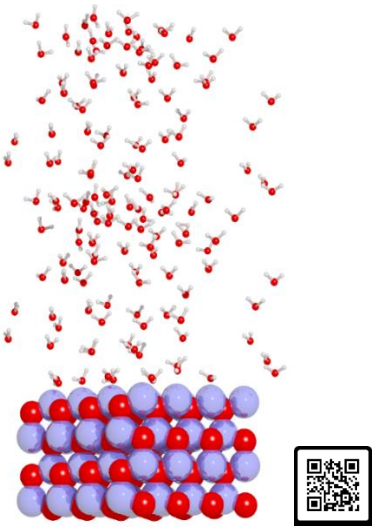


Figure 12. Structure of the slab of ZnO surrounded by a double water layer for periodic MD simulations.

4. Results

4.1. Evaluating the solid-liquid interaction

Table 3. Results of the interfacial tension solid-liquid for TiO_2 and ZnO. Number of Water Molecules in the simulation (N H_2O). Area of the interface in $Å^2$. γ in $mJ \cdot m^{-2}$. See section 2.3 for details.

	N H_2O	**Area**	**ϒ**
ZnO	73	292.60	1516
ZnO (double water layer)	146	292.60	1579
TiO_2	108	432.06	1011

Table 3 presents the results obtained from the MD simulations for evaluating the interaction solid-liquid between ZnO/TiO_2 slabs and water. It is noteworthy to mention the computational effort included in the evaluation of γ^{SL} to get proper solubility results. The solid-liquid interaction here computed is not only taking into account the interaction of a single water molecule with the surface, but we have evaluated with this methodology: i) the interaction of water molecules with the surface; ii) the effect of inter-water molecules interaction; ii) the possible reactivity on the surface (proton exchange); iii) the

entropy effect of the finite temperature at room conditions (300 K); and finally, iv) the own phonon structure of the slab. This means that, finding other theoretical works to compare our γ^{SL} results at this level has not been easy, because, or they focused at the static level interaction of water molecules (no MD simulation), or they were simulations performed at Molecular Dynamics level, focusing more in the structural parameters of the interaction that in the value of the interaction. [23,24]

The results obtained for ZnO surfaces are 1516 and 1578 $mJ \cdot m^{-2}$. The change in doubling the water bulk is only of 4%, only 62 $mJ \cdot m^{-2}$, which lays in the limit of the precision of the computational method. Both cases show, that even modifying the system, the Υ^{SL} is always in the same order of magnitude, meaning that the solid-liquid interaction is almost converged for the small system. Frequently, for this work it was very difficult to find bibliography to compare the results; for ZnO, we found that the solid-liquid interactions are in the range of 770±300 $mJ \cdot m^{-2}$ from thermodynamic measurements of the solubility in solid bulk, [25] and 1310 $mJ \cdot m^{-2}$ from calorimetric measurements in nanocrystals with different shapes and structures. [21] These values found show that the results obtained with our methods are in the same order of magnitude and have a very good agreement with experimental results, if we consider that we have only explored one surface in the case of the slab, whereas experimental results are for structures with random surfaces exposed to water. An extremely low value for ZnO nanoparticles (>20 nm and <130nm) of 60 $mJ \cdot m^{-2}$ was found in literature; however, the authors point out that probably the determination of their particle surface area is not accurate enough due to the formation of aggregates, and consequently the interfacial interaction found is too low.

Given that for ZnO the smallest separation between slabs showed a high accuracy, in the work for TiO_2 just the water bulk with a thickness of 18.9 Å have been run. Results show a good agreement with the bibliography found, 100 to 1000 $mJ \cdot m^{-2}$ depending on the solvent molecules and the ions absorbed [26] and 1030 $mJ \cdot m^{-2}$ for the hydrated surface (101) molecular water absorbed calculated by DFT.[27]

4.2. Solubility evaluation with the Ostwald-Freundlich model.

Using the non-extensive thermodynamic model developed by Ostwald-Freundlich, we estimated the solubility of large nanoparticles (>2 nm from our definition of large) from the solubility of the bulk and from the Υ^{SL}, both of them calculated from *ab initio* methodology in this thesis.

At this point, we estimated the concentration of ZnO NP in solution versus radius using the equation 5.4 and $\Upsilon = 1515\ mJ \cdot m^{-2}$ from previous section (using the value of 1578 $mJ \cdot m^{-2}$ makes no significant difference). We also evaluated the K_{sp} for each NP radius size. The representation of the ZnO NP concentration results versus radius from the solving of the Ostwald–Freundlich equation is shown in Figure 12. One observes the exponential behaviour for small nanoparticles, as well as the trend to the bulk solubility for large nanoparticles.

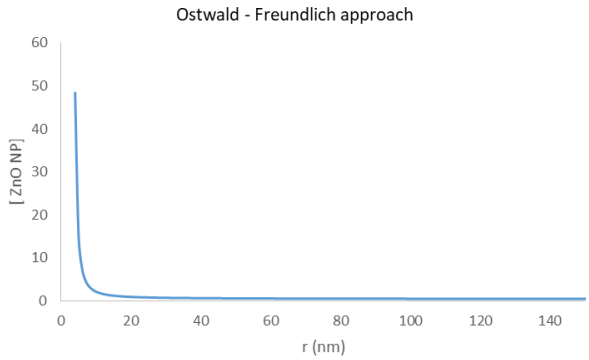


Figure 12. Graphical representation of the solubility showing the concentration of ZnO NP in (mol/l) using the Ostwald-Freundlich model using the slab model Υ^{SL}.

Table 4. Results ZnO NP solubility using the Ostwald-Freundlich model using the slab model Υ^{SL}. Solubility s in $mol \cdot L^{-1}$.

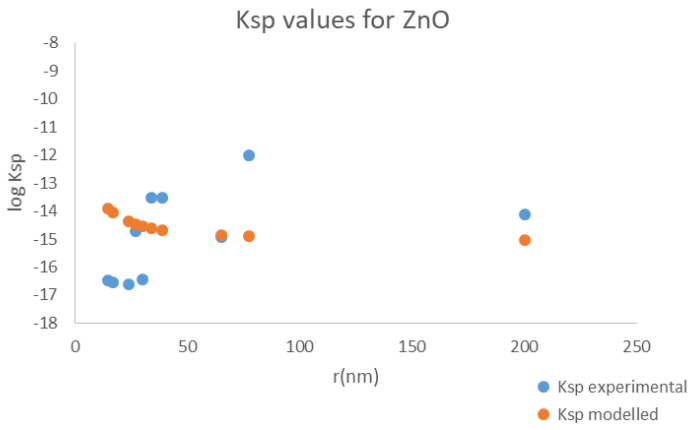
r(nm)	s	K_{sp}
2	$3{,}85 \cdot 10^{-02}$	$2{,}28 \cdot 10^{-04}$
3	$2{,}03 \cdot 10^{-03}$	$3{,}36 \cdot 10^{-08}$
5	$1{,}93 \cdot 10^{-04}$	$2{,}89 \cdot 10^{-11}$
15	$1{,}84 \cdot 10^{-05}$	$2{,}48 \cdot 10^{-14}$
20	$1{,}37 \cdot 10^{-05}$	$1{,}03 \cdot 10^{-14}$
35	$9{,}38 \cdot 10^{-06}$	$3{,}30 \cdot 10^{-15}$
50	$8{,}06 \cdot 10^{-06}$	$2{,}10 \cdot 10^{-15}$
80	$7{,}06 \cdot 10^{-06}$	$1{,}41 \cdot 10^{-15}$
100	$6{,}76 \cdot 10^{-06}$	$1{,}23 \cdot 10^{-15}$

The results of solubility found in literature, apart from the commented in the Table 2 of this chapter, have good agreement with results obtained by

Ostwald-Freundlich approximation. For example, Mudunkotuwa *et al.* [21] give, for pH around 7.8, values of solubility of ZnO between $1.8 \cdot 10^{-4}$ mol·L^{-1} to $4.1 \cdot 10^{-5}$ mol·L^{-1} for a range of sizes between 4 and 130 nm. The work of Avramescu *et al.* [28] presents solubility values for ZnO bulk $5.6 \cdot 10^{-5}$ mol·L^{-1}, and a range between $1.2 \cdot 10^{-4}$ and $9.6 \cdot 10^{-5}$ mol·L^{-1} for ZnO NP with sizes less than 50 nm and 100 nm, respectively. Therefore, the results we obtained with the *in silico* approximation combining the Quasi- *Ab Initio* Molecular Dynamics simulations with the Ostwald-Freundlich model are very close to the experimental solubilities found in these works.

We have represented in a graphic (Figure 13) experimental results found in literature for ZnO NP solubility and compared with results obtained with our combination of MD simulations with the Ostwald-Freundlich approach. Collected experimental K_{sp} results show a trend in values, though it is not so clear as in the theoretical evaluation given the variability that the experimental proceedings have implicated. With several experimental conditions; it has been a difficult task to add new values for comparing due to the variability in experiments. Also, it has to be taken into account the experimental difficulty given the low solubility of ZnO NPs form aggregates, or the definition of sizes of these NPs from microscopic techniques, adding a new variability in the K_{sp} calculation. For this reason, one conclusion in our work is that the modeling results are more homogeneous than the experimental ones helping to predict the solubility easily.

Figure 13. Graphical representation of the K_{sp} modelled in this thesis vs. K_{sp} experimental found in literature of ZnO NPs.



.

Then, using the previous mentioned parameters we have evaluated the solubility for TiO_2 nanoparticles. Results as a function of spherical diameter are shown in Table 5. We again appreciate the exponential behavior of the model for small nanoparticles. However, in this case, because TiO_2 is more insoluble than ZnO, the TiO_2 nanoparticles also present a degree of solubility lower than ZnO, and the slope of the curve is different due to the lower γ^{SL} of TiO_2-water system.

Table 5. Results of Ostwald-Freundlich model for TiO_2 NP using values from MD simulations in this thesis Radius r is in nm. Solubility s in $mol \cdot L^{-1}$.

r	s
2	$6.01 \cdot 10^{-7}$
3	$4.04 \cdot 10^{-8}$
5	$4.66 \cdot 10^{-9}$
15	$5.37 \cdot 10^{-10}$
20	$4.10 \cdot 10^{-10}$
35	$2.90 \cdot 10^{-10}$
50	$2.52 \cdot 10^{-10}$
80	$2.23 \cdot 10^{-10}$
100	$2.14 \cdot 10^{-10}$

These results are in good concordance with values found in bibliography. For example, for TiO_2 NP of diameter less than 25 nm a solubility of $2.5 \cdot 10^{-9}$ $mol \cdot L^{-1}$ have been found, which is in good agreement with our size of 5 nm, the disagreement with one unit in the exponential is not considered high given the error of the experimental measurements and the approximations made in the theoretical framework. Other results found are $2.8 \cdot 10^{-9}$ $mol \cdot L^{-1}$ for NP with diameter of less than 100 nm. [17] Also in the work of Schimdt *et al.* [17] are given results of solubility of anatase nanoparticles around 10^{-9} $mol \cdot L^{-1}$ for neutral pH.

In the case of TiO_2, it has been considered not necessary to evaluate K_{sp}. Because this MeO is clearly non soluble in neutral pH, the reaction then is not an ionic reaction but it is can be an equilibrium where multiple species can exist at the same time, for this reason and for the purpose of this thesis is not justified to evaluate the K_{sp} for this case.

5. Final Remarks

As an extra in this chapter, I include this paragraph to highlight the results achieved in this voluminous chapter. The study of the interaction liquid-solid has been performed using a well established method based on the combination of DFTB and MD simulations; however, it is not so known the method for the solid-liquid interfacial tension characterization, from the Ostwald Freundlich equation. Results obtained are encouraging to implement this methodology for other MeO and compute the interaction energy and interfacial tension for the solubility evaluation. Moreover, instead of concentrating all efforts in the interfacial tension evaluation for the solubility calculation, we also have estimated the k_{sp} for bulks, and in addition to this, the great results in the study of small NP inside a water cubic box, taking also in account the limitations found of this method for big NPs or aggregates.

6. Bibliography

1 P. Atkins and J. De Paula, *Atkins' physical chemistry*, 2009.

2 L. Escorihuela, A. Fernández, R. Rallo and B. Martorell, *Food Chem. Toxicol.*, 2018, **112**, 518-525.

3 W. Humphrey, A. Dalke and K. Schulten, *J Molec Graph.*, 1996, **14**, 33–38.

4 M. J. Gillan, D. Alfè and A. Michaelides, *J. Chem. Phys.*, 2016, **144,** 130901-13.

5 C. M. Maupin, B. Aradi and G. A. Voth, *J. Phys. Chem. B*, 2010, **114**, 6922–6931.

6 T. H. Choi, R. Liang, C. M. Maupin and G. A. Voth, *J. Phys. Chem. B*, 2013, **117**, 5165–5179.

7 P. Goyal, M. Elstner and Q. Cui, *J. Phys. Chem. B*, 2011, **115**, 6790–6805.

8 H. Nakai, A. W. Sakti and Y. Nishimura, *J. Phys. Chem. B*, 2016, **120**, 217–221.

9 S. Jahangiri, L. Cai and G. H. Peslherbe, *J. Comput. Chem.*, 2014, **35**, 1707–1715.

10 S. Große Holthaus, S. Köppen, T. Frauenheim and L. Colombi Ciacchi, *J. Chem. Theory Comput.*, 2012, **8**, 4517–4526.

11 S. Riahi, B. Roux and C. N. Rowley, *Can. J. Chem. Can. Chim.*, 2013, **91**, 552–558.

12 T. Dudev and C. Lim, *J. Am. Chem. Soc.*, 2000, **122**, 11146–11153.

13 N. M. Franklin, N. J. Rogers, S. C. Apte, G. E. Batley, G. E. Gadd and P. S. Casey, *Environ. Sci. Technol.*, 2007, **41**, 8484–8490.

14 R. B. Reed, D. A. Ladner, C. P. Higgins, P. Westerhoff and J. F. Ranville, *Environ. Toxicol. Chem.*, 2012, **31**, 93–99.

15 C. A. David, J. Galceran, C. Rey-Castro, J. Puy, E. Companys, J. Salvador, J. Monné, R. Wallace and A. Vakourov, *J. Phys. Chem. C*, 2012, **116**, 11758–11767.

16 M. Li, S. Pokhrel, X. Jin, L. Mädler, R. Damoiseaux and E. M. V Hoek, *Environ. Sci. Technol.*, 2011, **45**, 755–761.

17 J. Schmidt and W. Vogelsberger, *J. Solution Chem.*, 2009, **38**, 1267–

1282.

18 M.-L. Avramescu, P. E. Rasmussen, M. Chénier and H. D. Gardner, *Environ. Sci. Pollut. Res.*, 2017, **24**, 1553–1564.

19 D. R. Ely, R. Edwin García and M. Thommes, *Powder Technol.*, 2014, **257**, 120–123.

20 A. R. Nair and S. P. Sathian, *J. Chem. Phys.*, 2012, **137**, 084702-9.

21 I. a Mudunkotuwa, T. Rupasinghe, C.-M. Wu and V. H. Grassian, *Langmuir*, 2012, **28**, 396–403.

22 T. Hiemstra, *Geochim. Cosmochim. Acta*, 2015, **158**, 179-198.

23 S. Hosseinpour, F. Tang, F. Wang, R. A. Livingstone, S. J. Schlegel, T. Ohto, M. Bonn, Y. Nagata and E. H. G. Backus, *J. Phys. Chem. Lett.*, 2017, **8**, 2195–2199.

24 M. Sumita, C. Hu and Y. Tateyama, *J. Phys. Chem. C*, 2010, **114**, 18529-18537.

25 W. J. Wu and G. H. Nancollas, *J. Solution Chem.*, 1998, **27**, 521–531.

26 G. Oskam, A. Nellore, R. L. Penn and P. C. Searson, *J. Phys. Chem. B*, 2003, **107**, 1734–1738.

27 A. S. Barnard, P. Zapol and L. A. Curtiss, *J. Chem. Theory Comput.*, 2005, **1**, 107–116.

28 M.-L. Avramescu, P. E. Rasmussen, M. Chénier and H. D. Gardner, *Environ. Sci. Pollut. Res.*, 2017, **24**, 1553–1564.

Chapter 6. Statistical Analysis

1. Introduction

One of the goals of this thesis is to highlight the experimental variability that exists when measuring experimental properties such as band gap and/or solubility, and the lack of detailed information on experimental conditions that sometimes is found in literature. It is difficult to compare different experiments when variables such as time of assays, intervals of time during which the diverse biological systems have been exposed to the NP, temperatures, etc., are missing. On the other hand, modelling toxicological properties is a challenging task and computational methods can help in the prediction of these properties, avoiding the variability that the experimental process suffers.

Another objective of this thesis is to improve existing QSAR models adding computational descriptors, such as band gap and solubility, calculated *in silico* following the methodology we have explained in previous chapters. This is very helpful to understand better the toxicological behaviour of NP, what is essential for their safe use under regulations such as the REACH legislation. However, the lack of information and experimental data for this type of materials makes the development of QSAR and QNAR methods a more challenging task.

TiO_2 and ZnO NP are probably the most studied metal oxide NP given their widely applicability and their association of carrying toxicological effects. As a matter of fact, the Scientific Commission on Consumer Safety (SCCS) of the European Commission does not recommend these two types of NP in spray products that could be exposed to consumers by inhalation.[1]

With the aim to achieve the goals of this thesis, predictive models have been created for the prediction of ZnO and TiO_2 band gap and solubility and, later on, improving of QSAR existing models introducing our *in silico* descriptors.

2. Band gap prediction models

2.1. ZnO NP case

For band gap determination different experimental methods exist. In this thesis, around 60 values of experimental band gap have been gathered together from literature and compared with ZnO band gap determination by *in silico* prediction. The same procedure has been done in case of solubility, where the results are shown in Chapter 4.

In the case of ZnO, we present in Table 1 the different experimental values obtained from literature for spherical ZnO NP and also in Table 2 some results from computational works. What we can see is that the determination of the NP sizes can be done by two techniques, with microscopy (TEM, SEM, etc) or using X-Ray diffraction (XRD) by Debye-Scherrer's formula[2]. Another important difference between the experimental data is the technique used to calculate or measure the band gap. For this purpose, some experimental groups use the UV-absorption spectroscopy, where the direct band gap of ZnO is estimated from the absorption graph of $h\nu$ versus $(\alpha h\nu)^2$ through the absorption coefficient α, which is related to the band gap energy (E_g) as $(\alpha h\nu)^2=k(h\nu-E_g)$, where $h\nu$ is the incident light energy and k is a constant. The extrapolation of the straight line to $(\alpha h\nu)^2=0$ gives the value of band gap energy E_g.

Another experimental technique used to calculate the optical band gap from UV-visible spectra is to equate E_g with the wavelength at which the absorption is 50% of that at the excitonic peak (or shoulder), called $\lambda_{1/2}$. [3]

From the E_g calculated is also possible to estimate the size of the particles (R), using the effective mass model[4,5] in Equation 6.1:

$$R = \sqrt{\frac{2\pi^2 h^2 E_{gb}}{m^*(E_{gn}^2 - E_{gb}^2)}} \qquad (6.1)$$

where R is the radius, m^* is the effective mass of the specimen (m^*=29.15×10^{-31} kg for ZnO), E_{gb} is bulk band gap, h is Planck's constant (6.626×10^{-34} J s) and E_{gn} is the band gap at strong absorption edge. E_{gn} can be calculated by the formula $E_{gn}=hc/\lambda_{gn}$, where c is the velocity of light (3×10^8 m s^{-1}) and λ_{gn} is the strong absorption edge (~298.8 nm).

With the *in silico* results obtained during this thesis for different sizes of spherical ZnO NP, we have developed a model to predict band gap for bigger NP. The same exercise has been repeated with the values obtained from bibliography, dividing the experimental data by the type of the technique used for size characterisation. Therefore, there is a model for microscopy techniques and another for XRD method. The results are shown in Figure 1.

What can be observed in Figure 1 is that, for NP sizes smaller than 10 nm, the XRD technique is giving better results in band gap determination, since the sizes calculated by this method are more precisse than the ones obtained with microscopy techniques, maybe caused by the tendency to form agglomarates in case of small NP in solution. For bigger sizes, three techniques are in good agreement in terms of size determination and band gap. However, in case of NP bigger than 60nm, the microscopy model is showing a negative tendency. This means that for bigger sizes it is difficult to identify the real size of NP because of the agglomaration, and then the band gap determination is giving low values.

Table 1. Experimental results of ZnO NP Band gap from literature

	size(nm)	Band gap (eV)	particle shape
Experimental	17.1	3.87	Spherical[6]
	13-16	3.38	Powder[7]
	8.5	3.63	Spherical[8]
	1.8	4.5	Spherical[9]
	2	4	Spherical[9]
	2.5	3.74	Spherical[9]
	2.6	3.35	Spherical[9]
	3.6	3.3	Spherical[9]
	22.6±5.1	3.31	Spherical[10]
	23.12	3.25	Spherical[11]
	16	3.5	Spherical[12]
	5	3.78	Nanorod[13]
	5	3.29	thin film[14]
	4.7	3.44	Spherical[14]
	2.5	3.85	Spherical[3]
	3	3.64	Spherical[3]
	3.6	3.58	Spherical[3]
	4	3.54	Spherical[3]
	4.5	3.46	Spherical[3]
	5	3.44	Spherical[3]
	6	3.39	Spherical[3]
	6.5	3.39	Spherical[3]
	2	>4	Spherical[5]
	3	3.35	Spherical[5]
	4	3.29	Spherical[5]
	6	3.27	Spherical[5]
	2	>4	Spherical[4]
	3	3.8	Spherical[4]
	4	3.55	Spherical[4]
	6	3.48	Spherical[4]
	8	3.45	Spherical[4]
	10	3.44	Spherical[4]
	12	3.43	Spherical[4]
	14	3.41	Spherical[4]
	100	3.76	Spherical[15]

Table 2. Computational results of ZnO NP Band gap from literature

	shape	Band gap (eV)	particle shape and size(nm)
Computational	unpuckered Zn18O18	1.7	Needles[16]
	puckered Zn18O18	1.67	Needles[16]
	unpuckered [Zn6O6]	1.6	Needles[16]
	puckered [Zn6O6]	1.84	Needles[16]
	Hexagonal 1	2.28	c(A)=5.379; D(A) =3.711[17]
	Hexagonal 2	1.55	c(A)=5.371; D(A) =9.86[17]
	Hexagonal 3	1.14	c(A)=5.34 ; D(A) =16.26[17]
	Triangular 1	1.81	c(A)=5.38; D(A) =7.344[17]
	Triangular 2	1.48	c(A)=5.356; D(A) =11.09[17]
	Triangular 3	1.25	c(A)=5.349; D(A) =14.786[17]
	$(ZnO)_{20}$	4.11	Cluster[18]

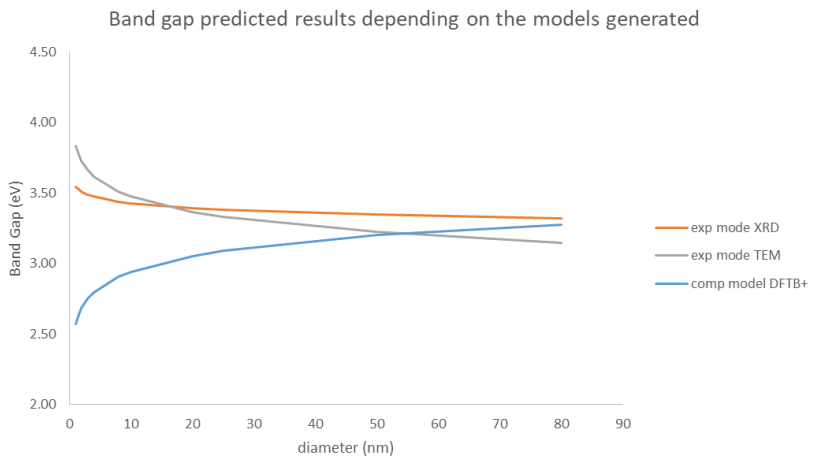


Figure 1. *Evolution of band gap prediction models versus NP sizes.*

The computational method gives good results in the range of 10 to 90 nm, and it is always much cheaper than experimental methods. So, we can say

that DFTB is a good methodology to create a model for band gap prediction, since this method always needs less resources than calculations based on pure DFT. There is no need to say that, the more resources available from the computational point of view, the bigger sizes can be calculated and the more refined the prediction model can be.

After the analysis and the demonstration how computational methods can help in this prediction, another statistical analysis has been applied in order to implement computational methods in the prediction of toxicological properties, i.e. the QSAR modelling.

2.2. TiO_2 NPs case of study

After the work done in the ZnO case and the good results obtained from the *in silico* prediction model, a similar work has been done for TiO_2 NP.

In this case, not so many experimental data have been found but enough to create a good experimental prediction model and compare it with the computational model created with this thesi's results. In this case, for the computational model, the results used have been a mixture between DFTB and DFT+U, since both techniques have given similar results.

In Table 3 are shown the experimental data found in literature. And in Figure 2, both prediction models are represented.

Table 3. Experimental data found in bibliography for TiO_2 NPs.

size (nm)	Band gap (eV)	particle shape
22.5	3.15	spherical[19]
39	3.2	spherical[20]
7.00	3.36	spherical[20]
7.9	3.32	spherical[20]
6.3	3.47	spherical[21]
4.8	3.79	spherical[22]
35	3.27	spherical[23]
100	3.2	spherical[24]

As it can be observed in the Figure 2, the computational prediction model gives very similar results to the experimental prediction model as bigger as the particles are. The maximum discrepancy in prediction results for small NP is about 5%, taking into account that it is a comparison between prediction

models, not true values. For this reason the computational model is considered good enough to generate data for the QSAR band gap evaluation.

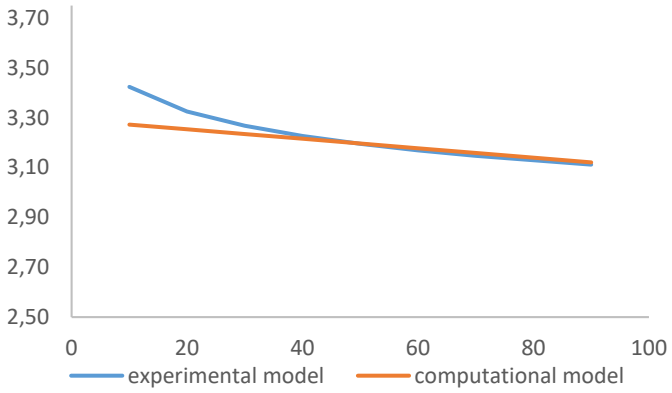


Figure 2. Data representation of prediction models for TiO_2 case. Y axis the band gap (eV) and in the X axis the radius of TiO_2 NP.

3. QSAR modelling

For this part of the work, we based our study in the data obtained in the study by Sayes and Ivanov[6] and the multiple linear regression models made by E. Papa *et al*.[1] for ZnO NP and TiO_2 NP. Membrane disruption was used as the toxicity indicator and the experimental descriptors used were the ones published by Sayes and Ivanov. In addition to these descriptors, in this work we added two more: the solubility and the band gap calculated by computational tools as it is explained in this thesis. In Table 4 are shown the data available for the model of ZnO NP.

Table 4. Data used to build the QSAR models for ZnO NP.

x0, eng size (nm)	x1, size in water (nm)	x2, size in PBS (nm)	x3, size in CCM (nm)	x4,concentration [mg/l]	x5,Zpotential (mV)	y,membrane damage (units/l)	x6, band gap (eV)	x7,solubility (mol/l)
50	55	158	107	25	-55.00	1.1	3.20	$8.06 \cdot 10^{-06}$
60	68	208	145	25	-45.00	1.03	3.23	$7.60 \cdot 10^{-06}$
70	71	198	152	25	-50.00	1.08	3.26	$7.29 \cdot 10^{-06}$
50	56	258	203	50	-50.00	1	3.20	$8.06 \cdot 10^{-06}$
60	78	386	243	50	-50.00	0.92	3.23	$7.60 \cdot 10^{-06}$
70	95	279	261	50	-50.00	0.99	3.26	$7.29 \cdot 10^{-06}$
50	168	314	283	100	-25.00	1.12	3.20	$8.06 \cdot 10^{-06}$
60	151	385	250	100	-30.00	1.25	3.23	$7.60 \cdot 10^{-06}$
70	172	354	273	100	-29.00	1.19	3.26	$7.29 \cdot 10^{-06}$
1.000	1.245	1.319	1.102	25	-44.00	1.58	3.68	$5.77 \cdot 10^{-06}$
1.200	1.268	1.325	1.057	25	-33.00	1.69	3.71	$5.75 \cdot 10^{-06}$
1.000	1.243	1.925	1.372	100	-20.00	1.25	3.68	$5.77 \cdot 10^{-06}$
1.200	1.124	1.805	1.458	100	-21.00	1.39	3.71	$5.75 \cdot 10^{-06}$
1.500	1.269	2.109	1.578	100	-21.00	1.45	3.75	$5.72 \cdot 10^{-06}$
1.500	1.198	1.381	1.073	25	-25.00	1.59	3.75	$5.72 \cdot 10^{-06}$

The model postulated by E. Papa *et al.* for the ZnO case is given in Equation 2:

$$LDH_{(ZnO)} = 1.041 + 0.001X_1 - 0\text{-}001X_2 + 0.001X_4 \quad (6.2)$$

N=15, r^2= 0.91, Q^2_{loo}=0.80, $Q^2_{lmo\ 30\%}$=0.76, , F=35

Where N is the number of samples, r^2 is the coefficient of determination, F is the ratio test, Q^2_{loo} is the leave-one-out correlation coefficient and Q^2_{lmo} is the coefficient for leave-many-out. [7]

In order to look for the best combination between the existing descriptors and the new ones, solubility and band gap, a stepwise regression method (SWR) has been run using Matlab software[8]. SWR is a systematic method for adding and removing terms from a multilinear model based on their statistical significance in a regression. At each step, the p value of an F-statistic is computed to test models with and without a potential term. If a

term is not currently in the model, the null hypothesis is that the term would have a zero coefficient if it is added to the model. *P*-values are: the *p*-enter value, in the range from 0.05 to 0.5, representing the maximum *p*-value for a term to be added; and *p*-remove value, from 0.01 to 0.8, representing the minimum *p*-value for a term to be removed. Depending on the terms included in the initial model and the order in which terms are moved in and out, the method may build different models from the same set of potential terms. The method terminates when no single step improves the model.[9]

In order to measure the predictive ability and check the overfitting of the QSAR model, a cross validation (CV) method has been applied to describe the relationship between predictors and response. One can talk about overfitting when the r^2 value is larger than 25% of the cross validated correlation coefficient (Q^2) value, this difference between them should not be more than 0.3.[10,11] The Q^2 value is used to indicate the robustness and predictive ability of the model, where $Q^2 > 0.5$ corresponds to a model with a good predictive power. Q^2 values increase while useful variables are added to the variables of a model in a previous step.

One of the most used methods of CV is the leave-one-out method (LOO), this is the most used method to evaluate the internal validation.[12] LOO consists in excluding each sample once and constructing a new model without this sample.

Q^2 for LOO method has been calculated as:

$$Q^2_{LOO} = 1 - \frac{PRESS}{TSS} = 1 - \frac{\sum_{i=1}^{N}(y_{pred,i} - y_i)^2}{\sum_{i=1}^{N}(y_i - y_m)^2} \qquad (6.3)$$

where *PRESS* is the sum of the squared differences between the experimental response *y* and the response *predicted* by the regression model, and *TSS* is the Total Sum of Squares and represents the total variance that a regression model can explain. $y_{pred,i}$ indicates that the response is predicted by a model estimated when the *i* sample was left out from the training set. y_i is each experimental measure. Finally, y_m is the mean of the experimental values.

The root mean square error (RMSE) is also calculated to evaluate the predictive quality of the models generated, because it shows the error between the mean of the experimental values and the predictive subjects calculated, a good value of RMSE is < 0.3 [11]

3.1. ZnO NP models

After applying the SWR algorithm to our data set, we present our best QSAR models in Table 5 for the ZnO NP case:

Table 5. Best QSAR models created by SWR method for ZnO NP.

Model	Descriptors used	R^2	F	Q^2_{LOO}	RMSE Cross Val.
E. Papa *et al*.	x1, x2, x4	0.91	35.46	0.80	0.10
Model 1-ZnO	x2, x4, x6, x7	0.92	27.14	0.81	0.11
Model 2-ZnO	x1, x2, x4, x6, x7	0.94	25.99	0.79	0.11
Model 3-ZnO	x1, x2, x4, x6	0.93	31.19	0.82	0.10

After applying the SWR algorithm to our data set, we can see that the best models, where both solubility (x7) and band gap (x6) are included, are the models number 1 and number 2, because the *F*-value and Q^2_{LOO} obtained are higher than in other SWR models tested, and r^2 and Q^2_{LOO} are higher than in the model of reference by E. Papa *et al*. The model obtained where only the band gap is included, the model number 3, it could be considered even better than models 1 and 2 because Q^2_{LOO} and *F*-value are superior. It must be noted that none of these parameters is related in any way to the model prediction power, and they are only related to the goodness of fit.

As it can be seen in the Table 5, the models generated using band gap and/or solubility have similar values of RMSE to the model of reference, so the predictive quality is similar for all of them.

The coefficients for the model number 1 are:

$$Y = -12.20 - 7.53E\text{-}04\ X_2 + 3.06E\text{-}03\ X_4 + 3.86\ X_6 + 1.86E+05 X_7 \quad (6.4)$$

And the coefficients for the model number 3 are:

$$Y = -3.29 + 6.09E\text{-}04\ X_1 - 6.82E\text{-}04\ X_2 + 2.37E\text{-}03\ X_4 + 1.35\ X_6 \quad (6.5)$$

3.2. TiO_2 models

The model postulated by E. Papa *et al.* for the TiO_2 case is given in Equation 6:

$$LDH_{(TiO2)} = 0.599 + 0.003X_4 + 0.004X_0 \quad (6.6)$$

N=22, r^2= 0.84, Q^2_{loo}=0.79, $Q^2_{lmo\ 30\%}$=0.78, F=48

As in the case of the ZnO, for the generation of new QSAR models it has been introduced the solubility and the band gap of TiO_2 NP of the size indicated in Table 6 to the data used in the work of E. Papa *et al.*:

Table 6. Data set used for TiO_2 QSAR modelling.

x0	x1	x2	x4	x5	x6	x7	y
30	125	1250	25	-10	3.23	$3.13 \cdot 10^{-10}$	0.90
30	102	987	25	-12	3.23	$3.13 \cdot 10^{-10}$	1.00
30	281	1543	50	-15	3.23	$3.13 \cdot 10^{-10}$	0.75
30	101	1045	50	-9	3.23	$3.13 \cdot 10^{-10}$	0.70
30	299	1754	100	-11	3.23	$3.13 \cdot 10^{-10}$	1.04
30	134	961	100	-11	3.23	$3.13 \cdot 10^{-10}$	1.09
30	600	1876	200	-12	3.23	$3.13 \cdot 10^{-10}$	1.15
30	298	1165	200	-12	3.23	$3.13 \cdot 10^{-10}$	1.20
45	129	2567	25	-9	3.21	$2.61 \cdot 10^{-10}$	0.90
45	129	2309	25	-10	3.21	$2.61 \cdot 10^{-10}$	0.85
45	201	2431	50	-9	3.21	$2.61 \cdot 10^{-10}$	0.75
45	201	2987	50	-11	3.21	$2.61 \cdot 10^{-10}$	0.78
45	876	1965	200	-11	3.21	$2.61 \cdot 10^{-10}$	1.35
45	876	2109	200	-10	3.21	$2.61 \cdot 10^{-10}$	1.40
125	136	3215	25	-11	3.05	$2.08 \cdot 10^{-10}$	1.25
125	136	2667	25	-10	3.05	$2.08 \cdot 10^{-10}$	1.17
125	149	3782	50	-10	3.05	$2.08 \cdot 10^{-10}$	1.00
125	149	2144	50	-15	3.05	$2.08 \cdot 10^{-10}$	1.10
125	343	3871	100	-12	3.05	$2.08 \cdot 10^{-10}$	1.50
125	343	2890	100	-9	3.05	$2.08 \cdot 10^{-10}$	1.42
125	967	3813	200	-9	3.05	$2.08 \cdot 10^{-10}$	1.60
125	967	2671	200	-8	3.05	$2.08 \cdot 10^{-10}$	1.65

Applying the same methodology as before, i.e. using SWR, the best combinations of variables to create prediction models are found. For these models the statistics parameters to evaluate the fitting of the data and the predictive quality have been calculated too, and all of these data are shown in Table 7.

Table 7. Parameters values for the best TiO_2 QSAR models.

Model	Descriptors used	R^2	F	Q^2_{Loo}	RMSE Cross Val.
E. Papa et al.	x0, x4	0.84	48.44	0.79	0.13
Model 1-TiO2	x4, x6	0.84	48.89	0.79	0.10
Model 2-TiO2	x0, x1, x4, x7	0.84	23.09	0.76	0.11
Model 3-TiO2	x1,x2,x4,x6	0.85	24.14	0.72	0.10

At a glance, for the three best models calculated, the best one is the Model number 1 because it has almost all the statistical parameters equal to the reference model by E. Papa *et al.*, even the number of variables employed. The other two models could be good enough but having the model 1, there is no need to discuss which other could be taken into account.

The coefficients for the model number 1 are:

$$Y= 7.19 + 2.71E\text{-}03\ X_4 - 2.002\ X_6 \quad (6.7)$$

The coefficients for the model number 2 are:

$$Y= 8.08 + 1.96E\text{-}04\ X_0 + 1.98E\text{-}03\ X_1 - 2.36\ X_4 + 9.33E08\ X_7 \quad (6.8)$$

The coefficients for the model number 3 are:

$$Y= 8.12 + 2.03E\text{-}04\ X_1 - 4.56E\text{-}05\ X_2 + 1.97E\text{-}03\ X_4 - 2.26X_6 \quad (6.9)$$

3.3. ZnO and TiO_2 models

At the same work of E. Papa *et al.* we can find also a model generated for both metal oxides NP, the model is:

$$LDH_{(ZnO+TiO2)} = 0.66 + 0.003X_4 + 0.005X_0 - 4.46E\text{-}5X_2 \quad (6.10)$$

N=31, r^2= 0.82, Q^2_{loo}=0.76, $Q^2_{lmo\ 30\%}$=0.74, F=40

In this thesis, in order to generate a model for both MeO NP we have used the data shown in Table 8, the variables numeration corresponds at the same variables as in the cases before mentioned:

Table 8. Data set for ZnO + TiO_2 QSAR modelling

x0	x1	x2	x4	x5	x6	x7	y
50	55	158	25	-55	3.20	$8.06 \cdot 10^{-06}$	1.1
60	68	208	25	-45	3.23	$7.60 \cdot 10^{-06}$	1.03
70	71	198	25	-50	3.26	$7.29 \cdot 10^{-06}$	1.08
50	56	258	50	-50	3.20	$8.06 \cdot 10^{-06}$	1
60	78	386	50	-50	3.23	$7.60 \cdot 10^{-06}$	0.92
70	95	279	50	-50	3.26	$7.29 \cdot 10^{-06}$	0.99
50	168	314	100	-25	3.20	$8.06 \cdot 10^{-06}$	1.12
60	151	385	100	-30	3.23	$7.60 \cdot 10^{-06}$	1.25
70	172	354	100	-29	3.26	$7.29 \cdot 10^{-06}$	1.19
30	125	1.250	25	-10.00	3.23	$5.15 \cdot 10^{-05}$	0.9
30	102	987	25	-12.00	3.23	$5.15 \cdot 10^{-05}$	1
30	281	1.543	50	-15.00	3.23	$5.15 \cdot 10^{-05}$	0.75
30	101	1.045	50	-9.00	3.23	$5.15 \cdot 10^{-05}$	0.7
30	299	1.754	100	-11.00	3.23	$5.15 \cdot 10^{-05}$	1.04
30	134	961	100	-11.00	3.23	$5.15 \cdot 10^{-05}$	1.09
30	600	1.876	200	-12.00	3.23	$5.15 \cdot 10^{-05}$	1.15
30	298	1.165	200	-12.00	3.23	$5.15 \cdot 10^{-05}$	1.2
45	129	2.567	25	-9.00	3.21	$4.30 \cdot 10^{-05}$	0.9
45	129	2.309	25	-10.00	3.21	$4.30 \cdot 10^{-05}$	0.85
45	201	2.431	50	-9.00	3.21	$4.30 \cdot 10^{-05}$	0.75
45	201	2.987	50	-11.00	3.21	$4.30 \cdot 10^{-05}$	0.78
45	876	1.965	200	-11.00	3.21	$4.30 \cdot 10^{-05}$	1.35
45	876	2.109	200	-10.00	3.21	$4.30 \cdot 10^{-05}$	1.4
125	136	3.215	25	-11.00	3.05	$3.42 \cdot 10^{-05}$	1.25
125	136	2.667	25	-10.00	3.05	$3.42 \cdot 10^{-05}$	1.17
125	149	3.782	50	-10.00	3.05	3.42·10-05	1
125	149	2.144	50	-15.00	3.05	3.42·10-05	1.1
125	343	3.871	100	-12.00	3.05	3.42·10-05	1.5
125	343	2.890	100	-9,00	3,05	$3,42 \cdot 10^{-05}$	1,42
125	967	3.813	200	-9,00	3,05	$3,42 \cdot 10^{-05}$	1,6
125	967	2.671	200	-8,00	3,05	$3,42 \cdot 10^{-05}$	1,65

The best models obtained for the data in Table 8 are shown in Table 9:

Table 9. Best QSAR models for ZnO + TiO_2

Model	Descriptors used	R^2	F	Q^2_{LOO}	RMSE Cross Val.
E. Papa et al.	x0, x2, x4	0.82	39.87	0.76	0.08
Model 1-ZnO+TiO2	x4, x6, x7	0.80	35.61	0.74	0.09
Model 2-ZnO+TiO2	x0, x1, x2, x4, x6	0.83	24.23	0.75	0.08
Model 3-ZnO+TiO2	x0, x4, x7	0.81	38.50	0.76	0.08

The coefficients for the model number 1 are:

$$Y= 6.94 + 2.65E\text{-}03\ X_4 - 1.92\ X_6 + 2.39E\ 04\ X_7 \quad (6.11)$$

The coefficients for the model number 2 are:

$$Y= 2.74 + 3.54E\text{-}03\ X_0 + 2.11E\text{-}04\ X_1 - 7.09E\text{-}05\ X_2 + 1.91E\text{-}03\ X_4 - 6.20E\text{-}01\ X_6 \quad (6.12)$$

The coefficients for the model number 3 are:

$$Y= 0.61 + 3.81E\text{-}03\ X_0 + 2.66E\text{-}03\ X_4 + 1.13E\ 04\ X_7 \quad (6.13)$$

In this case there are two models, the 1 and 3, that could be considered as the best models, highlighting the model 1 where both computational variables are included, and the statistical parameters are in the same range of the proposed by E. Papa *et al*., then this is a clue result for the work developed in this thesis, showing that including more data and variability in the initial data to generate a QSAR model, computational descriptors give one of the best models. Model 2 is also good but has more variables in the model and the statistical parameter F is not so high, then a good quality in the predictions is not achieved.

In this chapter it has been shown how the inclusion of computational variables in the existing models has given either a good fitting in the QSAR models and predictive quality. It is important to highlight that in the cases of the TiO_2 model or the model for ZnO+TiO_2, an experimental variable can be substituted by a computational parameter, obtaining a model with the same number of variables without decreasing the predictive power. Then, at this point, it has been achieved the main goal of this work: to achieve the same quality of prediction in QSAR modelling using computational descriptors, saving experimental work, time, resources and animal testing for toxicology risk assessment.

4. Bibliography

1 E. Papa, J. P. Doucet and A. Doucet-Panaye, *SAR QSAR Environ. Res.*, 2015, **26**, 1–19.

2 A. R. Lang, *Acta Metall.*, 1956, **4**, 102.

3 E. A. Meulenkamp, *J. Phys. Chem. B*, 1998, **102**, 5566–5572.

4 K.-F. Lin, H.-M. Cheng, H.-C. Hsu, L.-J. Lin and W.-F. Hsieh, *Chem. Phys. Lett.*, 2005, **409**, 208–211.

5 E. M. Wong and P. C. Searson, *Appl. Phys. Lett.*, 1999, **74**, 2939–2941.

6 T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska and J. Leszczynski, *Nat. Nanotechnol.*, 2011, **6**, 175–8.

7 G. Varughese, P. W. Jithin and K. T. Usha, 2015, **5**, 146–154.

8 M. K. Debanath and S. Karmakar, *Mater. Lett.*, 2013, **111**, 116–119.

9 L. . Kukreja, S. Barik and P. Misra, *J. Cryst. Growth*, 2004, **268**, 531–535.

10 H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y.-P. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Maedler, Y. Cohen, J. I. Zink and A. E. Nel, *ACS Nano*, 2012, **6**, 4349–4368.

11 S. M. Hosseini, I. A. Sarsari, P. Kameli and H. Salamati, *J. Alloys Compd.*, 2015, **640**, 408–415.

12 S. A. Ansari, A. Nisar, B. Fatma, W. Khan and A. H. Naqvi, *Mater. Sci. Eng. B*, 2012, **177**, 428–435.

13 C. W. Cheng, E. J. Sie, B. Liu, C. H. A. Huan, T. C. Sum, H. D. Sun and H. J. Fan, *Appl. Phys. Lett.*, 2010, **96**, 2008–2011.

14 R. Viswanatha, S. Sapra, S. Sen Gupta, B. Satpati, P. V. Satyam, B. N. Dev and D. D. Sarma, *J. Phys. Chem. B*, 2004, **108**, 6303–6310.

15 P. K. Samanta, A. Saha and T. Kamilya, *J. nano- Electron. Phys.*, 2014, **6**, 040415 1-2.

16 C. Szakacs, E. Merschrod S. and K. Poduska, *Computation*, 2013, **1**, 16–26.

17 S. Haffad, G. Cicero and M. Samah, in *Energy Procedia*, 2011, vol. 10, pp. 128–137.

18 F. De Angelis and L. Armelao, *Phys. Chem. Chem. Phys.*, 2011, **13**, 467–475.

19 M. M. Khan, S. a. Ansari, D. Pradhan, M. O. Ansari, D. H. Han, J. Lee and M. H. Cho, *J. Mater. Chem. A*, 2014, **2**, 637–644.

20 K. Madhusudan Reddy, M. Sunkara and A. Ramachandra Reddy, *Mater. Chem. Phys.*, 2003, **78**, 239–245.

21 M. Saber Lassoued, A. Lassoued, S. Ammar, A. Gadri, B. S. Abdelhamid and S. García-Granda, *J. Mater. Sci. Mater. Electron.*, 2018.

22 L. Gnanasekaran, R. Hemamalini and K. Ravichandran, *J. Saudi Chem. Soc.*, 2015, **40**.

23 D. Reyes-Coronado, G. Rodriguez Gattorno, M. Espinosa Pesqueira, C. Cab, R. de Coss and G. Oskam, *Nanotechnology*, 2008, **19**, 145605.

24 S. M. Gupta and M. Tripathi, *Chinese Sci. Bull.*, 2011, **56**, 1639–1657.

25 C. Sayes and I. Ivanov, *Risk Anal.*, 2010, **30**, 1723–1734.

26 R. Todeschini, V. Consonni, A. Mauri and M. Pavan, *Anal. Chim. Acta*, 2004, **515**, 199–208.

27 MATLAB, *version 7.10.0 (R2010a)*, The MathWorks Inc., Natick, Massachusetts, 2010.

28 N. R. Draper and H. Smith, *Wiley Ser. Probab. Stat.*, 1998, **47**, 706.

29 R. Kiralj and M. M. C. Ferreira, *J. Braz. Chem. Soc.*, 2009, **20**, 770–787.

30 R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese and R. K. Agrawal, *Int. J. Drug Des. Disocovery*, 2011, **2**, 511–519.

31 N. Chirico and P. Gramatica, *J. Chem. Inf. Model.*, 2011, **51**, 2320–2335.

Chapter 7. Conclusions

In this thesis we performed a strict and deep study of the best methods to evaluate the band gap and the solubility of MeO NP from a computational point of view. The use of periodical-DFT methods has allowed us to optimise structures computing the electronic ground state energy for nanostructures as the nanotubes or spherical nanoparticles. To get more reliability for band gap determination, the exchange-correlation functional has been improved using the DFT+U methodology. This type of functional gives us a reasonable computational cost and accuracy in band gap calculations or geometry optimisation, without affecting the predictive capability on the influence of experimental environment. After that, to reach large systems up to 10^3 atoms in order to simulate more realistic biological systems, it has been used the DFTB development for band gaps determination in large nanoparticles. Furthermore, the coupling of DFTB and molecular

dynamics simulations has allowed the description of water-NP interaction, giving and extra value to this work.

As seen in bibliography, there is an important gap in experimental data, and this situation creates a use of time-consuming resources for which the combination of experimental with computational work can make this task more affordable. In my opinion, the implementation of a computational methodology for nanomaterials is still in primary stage, but as it is shown in this thesis, there is a promising way for these methods.

The computational results obtained with the methodology developed in this thesis for the ZnO case have been promising and, in order to make more robust the method employed, it has been tested for TiO_2 too, showing an excellent efficiency in the results.

This work is a clear example of how computational chemistry can provide data for toxicity risk assessment without the variability that the experimental work gives and without spending so much time and resources in animal testing.

The efforts were focused first in testing our theory that descriptors, physico-chemical properties that are size-dependent and considered the clue for nanomaterial behaviour description, can be calculated computationally and later added to experimental data for QNAR modelling as is shown.

Great part of the work has been spent in the solubility evaluation of the MeO NP, because there is a few quantity of bibliography about that (both experimentally and computationally), what indeed is reasonable because the MeO NP has a very low solubility and it is quite difficult its evaluation either experimentally or computationally. In my opinion, the results obtained with the implementation of the use of thermodynamical theory as the Ostwald–Freundlich approach in the Molecular dynamics framework are fixing a new starting point for the solubility evaluation of NP in the near future, and it is a proof for computational engagement of legislation authorities for toxicity risk assessment.

Finally, QSAR methods is the statistical tool chosen to give an accurate assessment in the toxicity behaviour of these NP. Indeed, QSAR modelling is an effective tool to assess biological activity of nanomaterials with a reduced cost and can help to understand the toxicity mechanism that exists in nanomaterials that are in the market; additionally, it can also provide a tool for safer nanomaterials design.

It is very difficult to find comparable experimental data in bibliography, for this reason it has been decided to use the QSAR methods created by Ester Papa *et al.,* as it is explained in the statistical chapter, with

the aim to show how the combination of both philosophies can provide more affordable and reliable results.

It is true that the data that may be used in these models is not enough to avoid statistical biases, but making use of Matlab software to implement statistical methods such as cross validation method, leave-on-out technique or the calculation of the root mean square error, we have made a deep study of the contributing factors for toxicity models. Even with the use of these statistical methods to avoid overfitting, I am not saying that is not necessary to use more data; totally on the contrary, it would be better if we could make the same QSAR models with thousands of values, but nowadays this is quite impossible.

Moreover, in the statistical chapter it has been shown how the inclusion of computational variables in the existing models has given either a good fitting in the QSAR models and predictive quality. It is important to highlight that in the cases of the TiO_2 model or the model for $ZnO+TiO_2$, an experimental variable can be substituted by a computational parameter, obtaining a model with the same number of variables without decreasing the predictive power. Then, at this point, it has been achieved the main goal of this work: to achieve the same quality of prediction in QSAR modelling using computational descriptors, saving experimental work, time, resources and animal testing for toxicology risk assessment.

Finally, I would like to highlight that the work of this thesis it has been carried out with the managing of different techniques, software and disciplines. So, this is an additional proof that the perfect combination of different disciplines can produce a great work.

UNIVERSITAT
ROVIRA i VIRGILI