CrossMark

RESEARCH PAPER

Advantages and limitations of classic and 3D QSAR approaches in nano-QSAR studies based on biological activity of fullerene derivatives

Karolina Jagiello · Monika Grzonkowska · Marta Swirog · Lucky Ahmed · Bakhtiyor Rasulev · Aggelos Avramopoulos · Manthos G. Papadopoulos · Jerzy Leszczynski · Tomasz Puzyn

Abstract In this contribution, the advantages and limitations of two computational techniques that can be used for the investigation of nanoparticles activity and toxicity: classic nano-QSAR (Quantitative Structure–Activity Relationships employed for nanomaterials) and 3D nano-QSAR (three-dimensional Quantitative Structure–Activity Relationships, such us Comparative Molecular Field Analysis, CoMFA/Comparative Molecular Similarity Indices Analysis, CoMSIA analysis employed for nanomaterials) have been briefly summarized. Both approaches were compared according to the selected criteria, including: efficiency, type of experimental data, class of nanomaterials, time required for calculations and computational cost, difficulties in the interpretation. Taking into account the advantages and limitations of each method, we provide the recommendations for nano-QSAR modellers and QSAR model users to be able to determine a proper and efficient methodology to investigate biological activity of nanoparticles in order to describe the underlying interactions in the most reliable and useful manner.

Keywords Nano-QSAR · 3D QSAR · CoMFA · Nanomaterials · Toxicity · Environmental, health and safety effects

Introduction

There has been a significant increase in computational studies related to nanoparticles' activity and toxicity in the last few years (Ahmed et al. 2013; Durdagi et al. 2008b; Epa et al. 2012; Gajewicz et al. 2015;

K. Jagiello · M. Grzonkowska · M. Swirog ·
T. Puzyn (✉)
Laboratory of Environmental Chemometrics, Faculty of Chemistry, Institute for Environmental and Human Health Protection, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland
e-mail: t.puzyn@qsar.eu.org

L. Ahmed · B. Rasulev · J. Leszczynski
Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 JR Lynch Street, Jackson, MS 39217-0510, USA

B. Rasulev
Center for Computationally Assisted Science and Technology, North Dakota State University, 1805 NDSU Research Park Drive, Post Office Box 6050, Fargo, ND 58108, USA

A. Avramopoulos · M. G. Papadopoulos
Institute of Biology, Pharmaceutical Chemistry and Biotechnology, National Hellenic Research Foundation, 48 Vas. Constantinou Ave., 11635 Athens, Greece

Mikolajczyk et al. 2015; Puzyn et al. 2011b; Salahine-jad 2015; Sizochenko et al. 2014, 2015; Toropov et al. 2012, 2013; Tzoupis et al. 2011; Winkler et al. 2013). The majority of these contributions are based on the main chemistry principle that similar compounds will have similar biological properties (Hansch et al. 1963). The most important group of these techniques is represented by Quantitative Structure–Activity Relationships (QSAR) modelling (Gajewicz et al. 2015; Mikolajczyk et al. 2015; Puzyn et al. 2011b; Salahinejad 2015; Sizochenko et al. 2015; Toropov and Toropova 2015; Toropov et al. 2010; Winkler et al. 2013).

Classical QSAR approach, known also as Hansch Analysis (Hansch et al. 1963), is based on the assumption that biological activity of chemicals is correlated with their physicochemical properties and/or so-called structural descriptors (Puzyn et al. 2010). These descriptors encode certain structural features, such as polarizability, electronic properties and steric parameters. In this case, the developed model includes a set of selected variables (descriptors) that are statistically important and allow providing useful insights and understanding of the mode of studied interaction. However, this approach does not consider the 3D geometric features of the molecules, which leads to some difficulties in adequately describing ligand–receptor interactions. For this type of interactions, better results one can obtained by applying 3D QSAR methodology (Cramer et al. 1988; Klebe et al. 1994; Sippl 2010).

The first application of 3D QSAR technique was proposed in 1988 by Cramer et al. (1988). Their program, the Comparative Molecular Field Analysis (CoMFA) (Cramer et al. 1988), assumes that differences in biological activity correspond to changes in shapes and strengths of non-covalent interaction fields surrounding the molecules (Sippl 2010). Other techniques that also allow to describe 3D interactions in a quantitatively manner include: Comparative Molecular Similarity Indices Analysis (CoMSIA) proposed by Klebe et al. (1994) and the GRID/GOLPE program developed by Reynolds et al. (1989). Both could be considered as the extensions of CoMFA methods that propose to expand its applicability, and in many cases are applied as an alternative to the original CoMFA approach. Taking into account that 3D QSAR techniques consider the ligand properties calculated in its bioactive conformation, it is more suitable than classic

approach to study the ligand–receptor interactions (Sippl 2010).

Recently, both classical QSAR and 3D QSAR methodologies are widely applied to study biological activity of nanoparticles (Ahmed et al. 2013; Puzyn et al. 2011b; Tzoupis et al. 2011). Thus, the question: "How to select the best approach in order to properly describe the biological activity of nanomaterials in the most reliable and efficient manner?" may be raised. In this contribution, we compare the efficiency and applicability of both techniques: nano-QSAR (the classic Hansch approach applied for nanomaterials) with 3D nano-QSAR (CoMFA/CoMSIA approach applied for nanomaterials), in order to provide recommendations for QSAR modellers and the models users, to determine the right methodology for investigating nanoparticles' biological activity.

Methods

Nano-QSAR model

Objects

Fullerene derivatives were previously studied in order to understand their binding mode to HIV-1 protease based on the 3D-QSAR approaches (Durdagi et al. 2008b; Tzoupis et al. 2011). The CoMFA/CoMSIA models proposed by Tzoupis et al. (2011) were developed for binding energy (BE, kJ/mol) of 74 fullerene derivatives to HIV-1 protease. Among the studied inhibitors, there are 54 compounds, for which BE was calculated with docking simulations and 20 compounds for which the binding energy was obtained experimentally. This is a source of additional variance in the dataset. The average value of the binding energy in the first set (for which BE was calculated with docking simulations) is of an order of magnitude higher than the average binding energy in the second set (for which the binding energy was obtained experimentally).

According to the OECD QSAR recommendations (OECD 2004), data for the modelled property (endpoint) should be obtained with the same methodology/protocol. Thus, theoretically, we would develop classic QSAR model either for the first set of 54 fullerene derivatives or by using the second set of 20 compounds. Tzoupis's model was calibrated based on

the compounds with the computed BE and then validated based on the only 3 compounds with the computed BE and the 20 compounds with the experimentally measured BE. Therefore, we have decided to take the derivatives with the computed BE for calibrating our model.

Thus, the dataset used in this study contains 54 fullerene analogues that were tested for interaction with HIV-1 protease. All data have been taken from the literature (Tzoupis et al. 2011). The structures of the molecules are given in Table 1. Binding energies (BE) for these chemicals to HIV-1 protease have been calculated from docking simulations (Tzoupis et al. 2011).

Quantum–mechanical and descriptors calculations

To obtain optimal geometries of the investigated fullerene derivatives, we applied Density Functional Theory (DFT) approach employing the hybrid meta exchange–correlation functional M06-2X (Zhao and Truhlar 2008). All calculations were performed with the Gaussian 09 code. The 6-31G(d,p) basis set was used throughout the computations (Rassolov et al. 2001).

The following quantum-chemical descriptors were computed for all optimized structures: the total dipole moment, the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}) and the total energy. In addition, we have calculated structure-based additive descriptors for the fullerene substituents by applying a Dragon (version 6.0) software (Talete 2014). We have assumed, as it was previously described and adopted (Ahmed et al. 2013) that functional groups have a major contribution to the change of the properties of fullerenes since the fullerene core remains constant in each fullerene molecule. Thus, at first, we have calculated a full series of descriptors available in the Dragon software (about 4500 descriptors), then we excluded: (1) descriptors with standard deviation less than 0.0001, (2) descriptors with at least one missing value and (3) descriptors with cross-correlation larger than or equal to $r = 0.95$. The final set of Dragon-generated and quantum–mechanical descriptors contained 1419 descriptors. Then, the whole set of descriptors was used to develop a nano-QSAR (Hansch) model.

Nano-QSAR modelling procedure

First, the binding energies of fullerenes to HIV-1 protease were sorted according to the increasing values of the energy (response variable). Then, the dataset was split into two subsets: training and validation ones. The splitting was performed using a 3:1 algorithm according to which, after sorting, every third compound was assigned to a validation set (Puzyn et al. 2011a). Thus, we obtained a training set containing 36 (68 %) compounds and a validation set containing 17 (32 %) of them.

To select the most optimal combination of descriptors, we have used a genetic algorithm (Gramatica et al. 2014) implemented in the QSARINS software (Gramatica et al. 2013) (population size: 300, mutation rate: 65). In addition, descriptors with cross-correlation coefficient values higher than 0.85 have been excluded during final model construction. The chosen descriptors were then used to develop the model using a multiple linear regression analysis (MLR), as it is implemented in QSARINS software (Gramatica et al. 2013). The goodness-of-fit of the developed model was assessed through the squared correlation coefficient (R^2) and Root-Mean-Square Error of Calibration ($RMSE_C$). The Leave-One-Out (LOO) cross-validation technique was utilized for internal validation. The predictivity of the model was assessed by using the squared external validation coefficient (Q^2_{Ext}) and Root-Mean-Square Error of Prediction ($RMSE_P$) (Gramatica 2007). The applicability domain was verified by applying the Williams plot analysis (Gramatica 2007).

In addition to the above-described internal and external validation tests, an advanced statistical procedure was performed according to recommendation provided by Roy et al. (2015).

3D nano-QSAR model

As it was mentioned in the Introduction, the 3D QSAR methodology was recently applied to study the interaction of nanomaterials with their biological targets (Durdagi et al. 2008a, b; Tzoupis et al. 2011). For this contribution, we have adopted CoMFA/CoMSIA approach developed for interaction of fullerene derivatives with HIV-1 protease described in Tzoupis et al. (2011).

Table 1 Chemical structures of fullerene derivatives and the values of binding energy (BE) for these carbon-based nanoparticles to HIV-1 protease
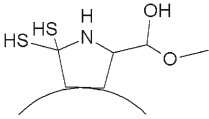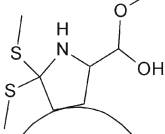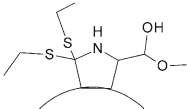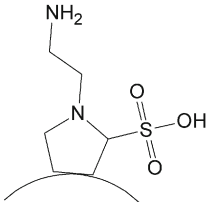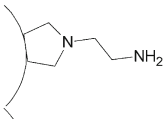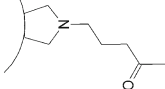
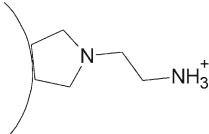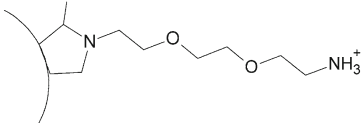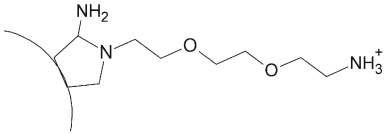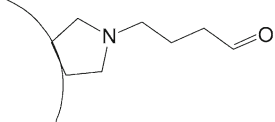| No.	Chemical structure of the substituent	BE (kcal/mol)
S1		−5.03
S2		−5.54
S3		−4.80
S4		−4.36
S5		−4.60
S6		−4.70
S7		−5.02
S8		−5.54
S9		−5.80
S10		−6.05

Table 1 continued
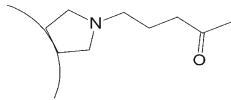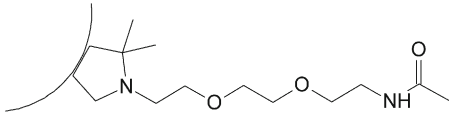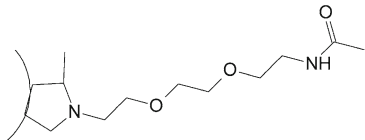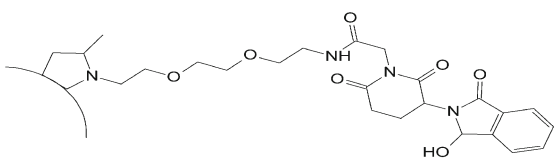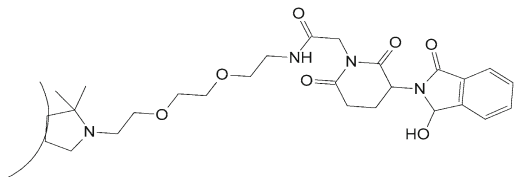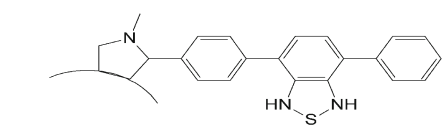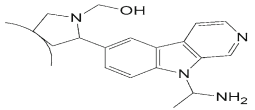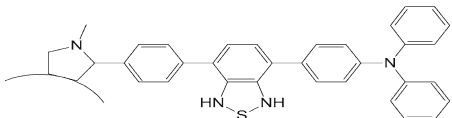
No.	Chemical structure of the substituent	BE (kcal/mol)
S11		−5.90
S12		−8.03
S13		−8.50
S14		−8.89
S15		−7.13
S16		−8.43
S17		−8.36
S18		−10.80
S19		−7.65

Table 1 continued
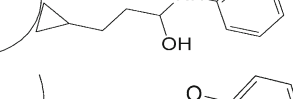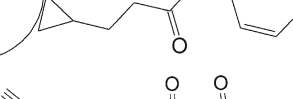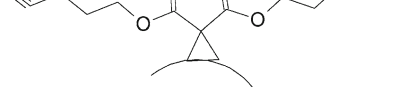
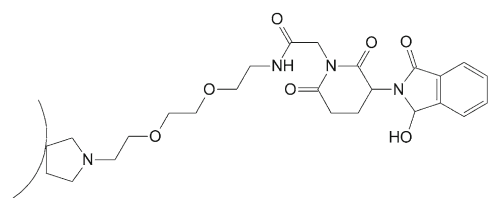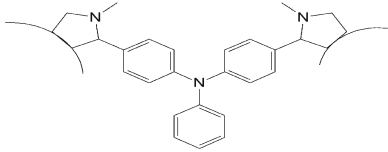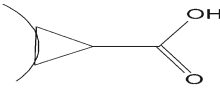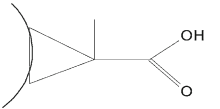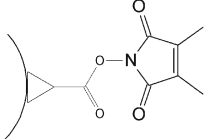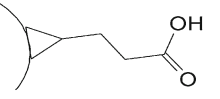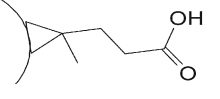| No.	Chemical structure of the substituent	BE (kcal/mol)
S20		−8.22
S21		−8.41
S22		−7.40
S23		−8.12
S24		−7.17
S25		−7.71
S26		−8.20
S27		−7.91
S28		−8.29
S29		−7.91
S30		−8.17

Table 1 continued
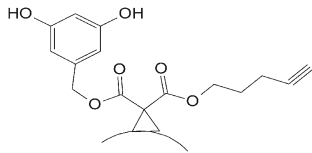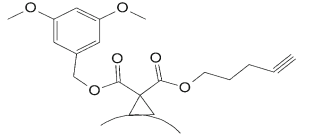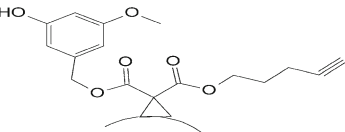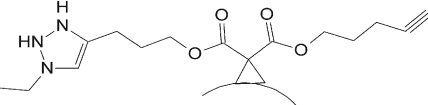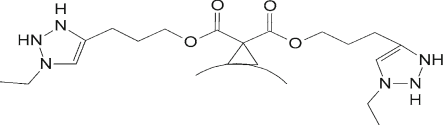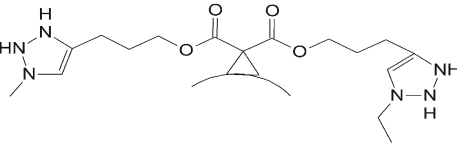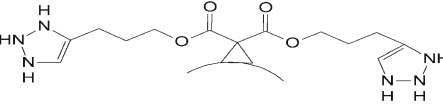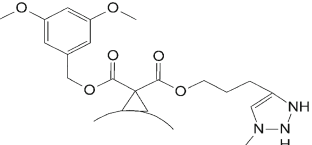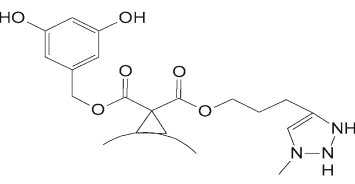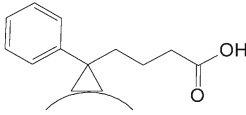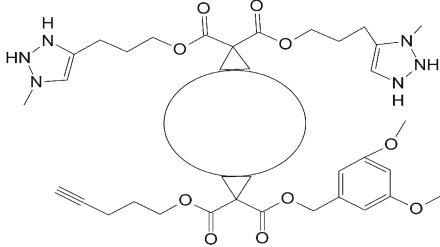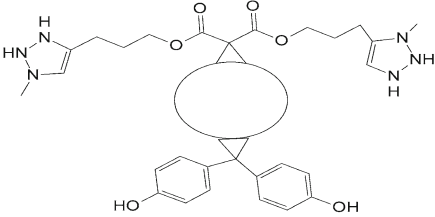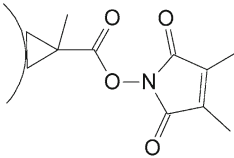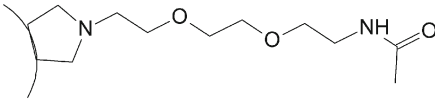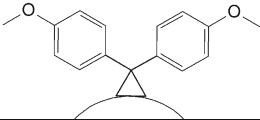
No.	Chemical structure of the substituent	BE (kcal/mol)
S31		−7.92
S32		−6.83
S33		−6.58
S34		−6.88
S35		−8.63
S36		−7.68
S37		−8.36
S38		−8.22
S39		−8.12

Table 1 continued

No.	Chemical structure of the substituent	BE (kcal/mol)
S40		−8.76
S41		−10.30
S42		−10.74
S43		−10.80
S44		−11.02
S45		−11.07
S46		−10.99
S47		−8.54
S48		−10.52

Table 1 continued

No.	Chemical structure of the substituent	BE (kcal/mol)
S49		−12.17
S50		−8.24
S51		−8.38
S52		−8.27
S53		−6.38
S54		−5.51

Calculated binding energies from docking simulation, values taken from Tzoupis et al. (2011)

Classic nano-QSAR versus 3D nano-QSAR

Nano-QSAR and 3D nano-QSAR were then compared according to the following criteria: (1) efficiency, which is measured by the obtain statistics, (2) type of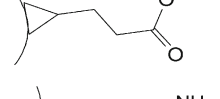 experimental data: receptor-base response, cell-based response and tissue-based response, (3) type of nanomaterials: their chemical nature (organic, non-organic, metals) and the homogeneity of their structures, (4) computational cost and time required

to perform whole procedure and (5) software availability.

Results and discussion

Nano-QSAR model

Based on the values of the binding energy (BE, kcal/mol) and the calculated descriptors, we have

developed a classic nano-QSAR model by employing the MLR methodology. The obtained model Eq. (1) is as follows:

$$
\begin{aligned}
\mathbf{BE}[\mathrm{kcal/mol}] = &-7.84 + 0.73 * \mathbf{MAXDN} \\
&+ 1.77 * \mathbf{GATS2e} - 0.51 * \mathbf{HNar} - 0.51 * \mathbf{C} \\
&- \mathbf{0.07} - 0.95 * \mathbf{B08}[\mathbf{C - O}]
\end{aligned} \quad (1)
$$

$N_{\text{train}} = 36$, $N_{\text{test}} = 17$, $R^2 = 0.80$, $\text{RMSE}_c = 0.85$, $Q_{\text{CV}}^2 = 0.74$, $\text{RMSE}_{cv} = 0.98$, $Q_{\text{ext}}^2 = 0.73$, $\text{RMSE}_p = 0.92$, $\text{CCC} = 0.86$, $r_m^2(\text{training})\text{scaled} = 0.80$, $r_{\text{maver}}^2(\text{training})\text{scaled} = 0.72$, $\Delta r_m^2(\text{training})\text{scaled} = 0.16$, $r_m^2(\text{test})\text{scaled} = 0.66$, $r_{\text{maver}}^2(\text{test})\text{scaled} = 0.65$, $\Delta r_m^2(\text{test})\text{scaled} = 0.02$, $r_m^2(\text{overall})\text{scaled} = 0.76$, $r_{\text{maver}}^2(\text{overall})\text{scaled} = 0.70$, $\Delta r_m^2(\text{overall})\text{scaled} = 0.12$.

As it can be noticed, the developed model represents a linear combination of five descriptors. Two of them belong to topological indices, namely: maximal electrotopological negative variation (**MAXDN**) (Todeschini and Consonni 2009) and Narumi harmonic topological index (**HNar**) (Todeschini and Consonni 2009). Next three descriptors are: Geary autocorrelation of lag 2 weighted by Sanderson electronegativity (**GATS2e**) (Todeschini and Consonni 2009), the descriptor of atom-centred CH2X2 fragment (**C-007**) (Todeschini and Consonni 2009), the presence or absence of C–O at topological distance 8, and **B08**—descriptor that belongs to 2D atom pairs group (Todeschini and Consonni 2009).

The Narumi harmonic topological index (HNar) is related to molecular branching and is proposed as the number of non-hydrogen atoms divided by the reciprocal vertex degree sum (Todeschini and Consonni 2009). According to Eq. 1, higher values of this descriptor decrease the binding energy. The presence of aromatic rings in the structure increases the value of this descriptor. Most probably, the presence of HNar in the model means that the higher number of rings (for example, aromatic cycles) and connected to them functional groups increase the interaction with protein in comparison to structures with the same number of atoms composed of long, linear alkyl chains only. Second topological descriptor, the maximal electrotopological negative variation (MAXDN), describing the maximum negative intrinsic state difference in the molecule is related to the nucleophilicity of the molecule. The positive value (0.726) of the coefficient (Eq. 1) implies that this descriptor is correlated

positively with the value of BE. Fullerenes that have more electronegative elements in theirs structures, are characterized by higher values of MAXDN. Next descriptor: Geary autocorrelation of lag 2 weighted by Sanderson electronegativity (GATS2e) represents 2D-autocorrelation classes of descriptors. This descriptor is related to the topology of the structure or its parts that are in association with a given physicochemical property (in this case—Sanderson electronegativity) (Todeschini and Consonni 2009). According to Eq. 1, an increase in GATS2e descriptor's value increases the binding energy. Another descriptor that is indirectly connected to electronegativity is C-007, which encodes the presence of CH2X2 atom-centred fragment. The X stands for a highly electronegative atom, like oxygen, nitrogen, sulphur, phosphorus and various halogens (Todeschini and Consonni 2009). Only four of the considered structures (S18, S41-S44) display negative value for this descriptor.

Interestingly, we noticed that all descriptors selected by the genetic algorithm into the model (1), although not internally correlated, were connected to the electronegativity to some extent. Certainly, the presence of the electronegative elements results in polarization of a molecule, changes the binding energy, and thereby increases or decreases the interaction between investigated ligands and protein.

Goodness-of-fit, robustness and predictive ability of the developed QSAR model have been confirmed by high values of R^2, Q_{CV}^2, Q_{Ext}^2, CCC and relatively low values of the errors represented by: RMSE_C, RMSE_{CV}, RMSE_P. Moreover, the visual correlation between the observed and predicted values of BE for the training (T) and validation (V) sets confirmed the good quality of the model (Fig. 1a). Additional metrics (summarized under Eq. 1) confirmed the robustness of the developed model as well.

Since the error values (RMSE_C, RMSE_{CV}, RMSE_P) are at the same level and there are no significantly large residual values for the validation set, one can conclude that the model has not been over fitted (Jagiello et al. 2014). This means that the model predictions are correct not only for the training compounds, but also for external set of compounds.

In order to verify the applicability domain of the nano-QSAR model, we have applied the leverage approach (Gramatica 2007). So-called the Williams plot (Fig. 1b) presents the relationship between leverage values (expressing similarity of a given compound

Fig. 1 a Docking-based versus predicted binding energy plot for the MLR model; **b** Williams plot: standardized residuals versus leverages
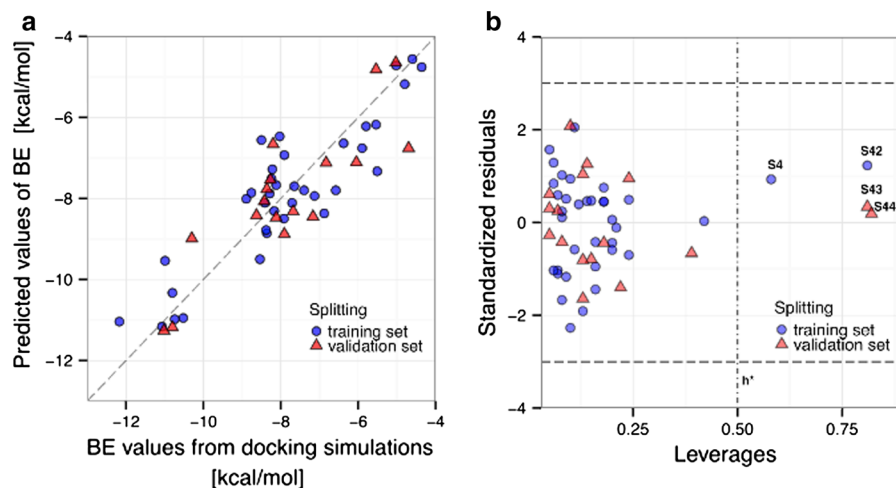
to the training set) and the standardized cross-validated residuals (prediction errors observed for particular compounds). Analysis of the plot confirmed that because the prediction errors for all compounds from the training and validation sets do not exceed the square area between ± 3 standard deviation units, there are not outlying predictions observed. The formal leverage (similarity) threshold value h^* is equal to 0.51. Interestingly, two compounds from the training set (S4 and S42) and two from validation set (S43 and S44) are characterized by the leverages greater than the threshold value, but—simultaneously—they have small residuals. Such compounds are called "good high leverage points", and—as it has been previously demonstrated by Jaworska et al. (Jagiello et al. 2014; Jaworska et al. 2005)—compounds having h_i greater than h^*, stabilize the model and make it predictive for new compounds differing structurally from the training set. Obviously, this is the true only when the residuals observed for the training compounds are small.

3D nano-QSAR model

Recently, the contributions aimed at the application of 3D QSAR approach for evaluation of the binding energy of fullerene inhibitors to HIV-1 protease have been published (Durdagi et al. 2008a, b; Tzoupis et al. 2011).

In order to compare the classic nano-QSAR (Hansch Analysis) approach with the 3D nano-QSAR, we have adopted contribution proposed by Tzoupis et al. (2011). The main objective of these studies was to

design a series of fullerene-based inhibitors for HIV-1 protease by employing the CoMFA/CoMSIA approach. Models proposed by Tzoupis et al. (2011) were developed for the same set of fullerenes as we have used in developing nano-QSAR (Hansch) model (Table 1). Moreover, Tzoupis et al. (2011) obtained models with better statistics than those previously presented by Durdagi et al. (2008a).

Comparison between the classic nano-QSAR (Hansch Analysis) and 3D nano-QSAR

In order to assess the efficiency of the classic nano-QSAR versus the 3D nano-QSAR, we have made a comparison of the statistics characterizing quality of the predictions for each approach (Table 2). The obtained statistics are very close to each other. Similar statistics were also obtained in previously published contributions related to fullerenes activity against HIV-1 protease (Ahmed et al. 2013; Toropov et al. 2010). Nano-QSAR model developed by Toropov et al. (2010) displays statistics, as follow: $R^2 = 0.9769$, $Q_{cv}^2 = 0.9646$, similar to model developed by Ahmed et al. (2013): $R^2 = 0.882$, $Q_{cv}^2 = 0.738$. This could suggest that both approaches have similar efficiency and can be applied to study this phenomenon. Which one is more suitable depends on particular task, i.e. a type of information one would like to gather.

By employing classic nano-QSAR, the certain parts of molecules, that are responsible for the biological activity, have been identified statistically. For

Table 2 Comparison of statistics obtained in nano-QSAR and 3D nano-QSAR (CoMFA and CoMSIA) approaches

	R^2	Q^2_{cv}	References
nano-QSAR	0.80	0.74	This work
CoMFA	0.84	0.613	Tzoupis et al. (2011)
CoMSIA	0.92	0.763	Tzoupis et al. (2011)

example, changes in the electronegativity have been determined to be the driving force of the interactions between fullerenes and HIV-1 protease (this work). CoMFA/CoMSIA analysis of fullerene-based inhibitors of HIV-1 protease pointed out that the highest contribution to the binding energy is associated with the electrostatic interactions, where highly electronegative groups increase (decrease) the affinity of fullerene derivatives to the protein. This is in agreement with the classic nano-QSAR results (Tzoupis et al. 2011). However, the applied 3D QSAR approach indicates also other effects that favour these interactions, such as: hydrophobic interactions and H-bonding (Tzoupis et al. 2011). Thus, one can conclude that classic nano-QSAR allows gathering general knowledge about the mechanism of the studied interaction, while 3D nano-QSAR describes the ligand-based response in more details, relying on their three-dimensional structures.

Additionally, 3D nano-QSAR approach provides a clear visualization, i.e. allows obtaining 3D graphics image superimposed on a core molecule of the dataset. This permits to determine more precisely functional groups of the molecules involved in interactions with residues within the binding pocket of the protease (see details in (Tzoupis et al. 2011)). Such information facilitates appropriate modifications of fullerene derivatives that appropriately improve their binding affinity to the HIV-1 protease.

The literature studies (Kim et al. 1998; Klebe et al. 1994; Podlogar and Ferguson 2000; Puzyn et al. 2010) allow us to provide comprehensive comparison between both techniques. Thus, the advantage of 3D QSAR over Hansch analysis includes also applicability of this approach to evaluate a set of structurally diverse compounds, as long as they act within the same mechanism (Kim et al. 1998; Kubinyi 1998). This advantage, in the case of nanomaterials, is of a high value. Since the classic nano-QSAR is

developed with the application of various statistical techniques, they require experimental data measured for sufficient number of considered species. Thus, in order to apply QSAR one needs to have at least 15–20 experimentally measured values of biological activity for chemicals that are located within the applicability domain of the model, which means that they are structurally similar. This principle is often difficult to be fulfilled in case of nanomaterials. Moreover, the application of nano-QSAR is limited also by insufficient set of tools to describe the uniqueness of nanoparticles. More appropriate types of descriptors should reflect not only the molecular structure of these species but also their supra-molecule pattern (e.g. size, shape, porosity, morphology, etc.), and very often their system dependent properties (e.g. agglomeration, formation of protein coronas etc.).

On the other hand, there is no limitation for the application of classic nano-QSAR considering the type of the experimental endpoint values (in vivo and in vitro) and the type of chemicals for which this model could be applied (organic, inorganic, etc.). The various quantum-chemical descriptors can provide useful insight into mode of cytotoxic/toxic action of nanoparticles involving metal oxides (Gajewicz et al. 2015; Toropov et al. 2013), as well as fullerene derivatives (Ahmed et al. 2013; Toropova et al. 2010). Nevertheless, 3D nano-QSAR is much more applicable for organic nanomaterials.

Both techniques require similar calculation time and computational costs. In case of classic nano-QSAR, computer resources are limited mainly by calculations of appropriate nano-descriptors. In 3D QSAR approach, the proper orientation of the ligand to its biological target becomes a crucial factor of success (Kim et al. 1998). The bioactive conformation of ligand can be obtained experimentally, by NMR spectroscopy or X-ray crystallography, or theoretically by means of molecular docking. For both techniques, commercial tools with user-friendly interface are available (Gramatica et al. 2013; TRIPOS Inc. 2001).

In Table 3, we have summarized a comparison of the two considered approaches: nano-QSAR and 3D nano-QSAR according to the described above criteria: type of experimental data; type and characterization of nanomaterials; required software, time and computational costs.

Recommendations

Taking into account the advantages and limitations of the nano-QSAR and the 3D nano-QSAR technique, we provide some recommendations for nanomodellers as well as for the users of these methodologies, Fig. 2, to better understand and more efficient predict the biological activity of nanomaterials.

According to the decision tree, shown in Fig. 2, the recommendation which approach should be applied, either classic nano-QSAR or 3D nano-QSAR, in order to better understand the biological activity of nanomaterials, require to answer to the following questions:

1. What types of experimental (response) data are available considered?
2. What types of nanomaterials are considered?
3. What is the major goal of the study?

It is obvious that the adequate experimental data are essential to obtain proper models, both in case of classic nano-QSAR and 3D approach. Appropriate data should fulfil two main principles: (1) should be measured according to the same protocol (ideally if they could be from the same source) and (2) should be symmetrically distributed around their mean value and their precision should be distributed over its range of variation (Kubinyi 1998). The more extensive discussion on biological data for nanomaterials could be found in the literature (Hristozov et al. 2012; Puzyn et al. 2010). Besides the quality of the data, the type of measured response is important to answer the first above-listed question. In this point, it worth to emphasize that classic nano-QSAR represents more universal approach. There are models that have been developed for particular molecular targets response (Ahmed et al. 2013), cell response (Toropov et al. 2013), or the response measured on higher level of organism organization (Toropova et al. 2015). On the other hand, in the 3D nano-QSAR approach the receptor-based response is required. This knowledge one can obtain directly by performing the proper experimental studies (e.g. X-ray crystallography, NMR) or indirectly by applying classic QSAR studies. Defining the type of descriptors that are correlated with the modelled activity, in many cases allows finding the molecular target of the process. Development of 3D nano-QSAR model is not recommended, unless one expects that the analysis will reveal insights into 3D interaction between ligand and receptor in its binding pocket (Puzyn et al. 2010).

The decision on which approach should be applied depends also on the chemical nature of nanomaterials. There is no limitation in application of classic nano-QSAR considering type of chemicals for which this model could be applied (organic, inorganic, metals, etc.). However, 3D nano-QSAR is rather applicable for organic nanomaterials.

The third question refers to the major task of the study. If the biological target is not known, and the objective is to find this target or gather general information about the biological activity of nanomaterials, the classic nano-QSAR would be the right choice. But, if one knows the binding pocket of the studied materials, the 3D nano-QSAR might provide more useful information regarding investigated activity.

Table 3 Applications and requirements of classic nano-QSAR and 3D nano-QSAR

Methods criteria	Nano-QSAR	3D nano-QSAR
Experimental data	Cell-based response, tissue-based response, etc	Ligand-based response
Nanomaterials	Inorganic, organic, metals	Organic
(a) Homogeneity of the chemical structure	homogenous set	Heterogeneous data with the same mode of action
(b) Data preparation	Calculation of nanodescriptors	Knowledge on the bioactive conformation of each molecule (docking)
Statistics obtained	Determination coefficients for calibration and validation, root-mean-square errors	Determination coefficients for calibration and validation, root-mean-square errors
Time	Limited by descriptors' calculation	Limited by docking procedure
Computational costs	Limited by descriptors' calculation	Limited by docking procedure
Software	Commercially available in user-friendly software	Commercially available in user-friendly software

Fig. 2 Decision tree for determining application classic or 3D nano-QSAR
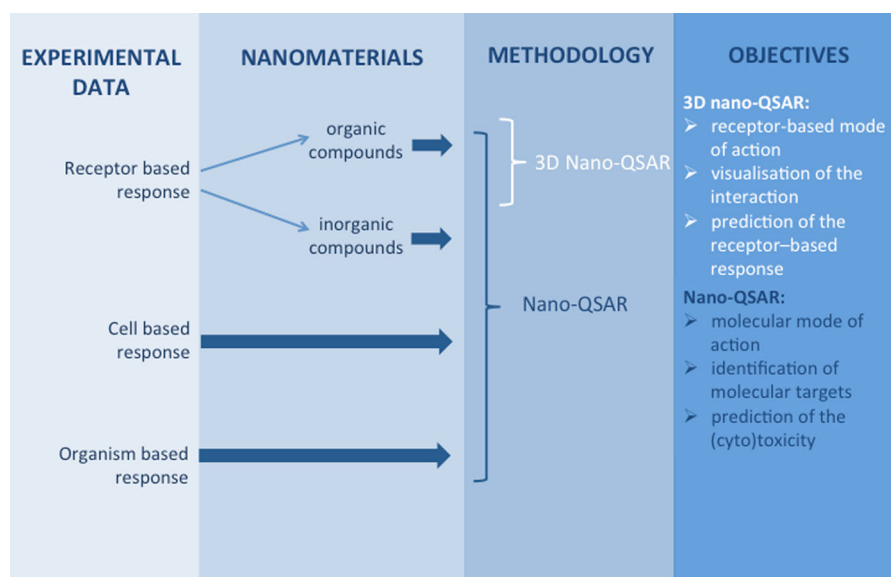
Conclusions

Summing up, we have developed nano-QSAR model allowing to predict the activity of fullerenes derivatives against HIV-1 protease. Developed model was compared with previously published contribution describing the same interaction by means of 3D QSAR approach. Taking into account this case study and literature available studies, the limitations and advantages of each methodology have been discussed. We have developed the recommendation tree for determining a proper methodology to investigate biological activity of nanoparticles. We do believe that both approaches, nano-QSAR and 3D nano-QSAR, could be used simultaneously, if it is possible. Application of classic nano-QSAR model, which is more universal approach, would allow gathering general information about the mode of biological activity of nanomaterials. Then, the 3D QSAR application would help in understanding this activity in detail.

References

Ahmed L, Rasulev B, Turabekova M, Leszczynska D, Leszczynski J (2013) Receptor- and ligand-based study of fullerene analogues: comprehensive computational approach including quantum-chemical, QSAR and molecular docking simulations. Org Biomol Chem 11:5798–5808. doi:10.1039/c3ob40878g

Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967. doi:10.1021/ja00226a005

Durdagi S, Mavromoustakos T, Chronakis N, Papadopoulos MG (2008a) Computational design of novel fullerene analogues as potential HIV-1 PR inhibitors: analysis of the binding interactions between fullerene inhibitors and HIV-1 PR residues using 3D QSAR, molecular docking and molecular dynamics simulations. Bioorgan Med Chem 16:9957–9974. doi:10.1016/j.bmc.2008.10.039

Durdagi S, Mavromoustakos T, Papadopoulos MG (2008b) 3D QSAR CoMFA/CoMSIA, molecular docking and

molecular dynamics studies of fullerene-based HIV-1 PR inhibitors. Bioorg Med Chem Lett 18:6283–6289. doi:10.1016/j.bmcl.2008.09.107

Epa VC, Burden FR, Tassa C, Weissleder R, Shaw S, Winkler DA (2012) Modeling biological activities of nanoparticles. Nano Lett 12:5808–5812. doi:10.1021/nl303144k

Gajewicz A, Cronin MTD, Rasulev B, Leszczynski J, Puzyn T (2015) Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: nano-read-across. Nanotechnology. doi:10.1088/0957-4484/26/1/015701

Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26:694–701. doi:10.1002/qsar.200610151

Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S (2013) QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. J Comput Chem 34:2121–2132. doi:10.1002/jcc.23361

Gramatica P, Cassani S, Chirico N (2014) QSARINS-chem: insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. J Comput Chem 35:1036–1044. doi:10.1002/jcc.23576

Hansch C, Streich M, Geiger F, Muir RM, Maloney PP, Fujita T (1963) Correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. J Am Chem Soc 85:2817–2824. doi:10.1021/Ja00901a033

Hristozov DR, Gottardo S, Critto A, Marcomini A (2012) Risk assessment of engineered nanomaterials: a review of available data and approaches from a regulatory perspective. Nanotoxicology 6:880–898. doi:10.3109/17435390.2011.626534

Jagiello K et al (2014) Direct QSPR: the most efficient way of predicting organic carbon/water partition coefficient (log K (OC)) for polyhalogenated POPs. Struct Chem 25:997–1004. doi:10.1007/s11224-014-0419-1

Jaworska J, Nikolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set in descriptor space: a review. Atla Altern Lab Anim 33:445–459

Kim KH, Greco G, Novellino E (1998) A critical review of recent CoMFA applications. Perspect Drug Discov 12:257–315. doi:10.1023/A:1017010811581

Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indexes in a comparative-analysis (Comsia) of drug molecules to correlate and predict their biological-activity. J Med Chem 37:4130–4146. doi:10.1021/Jm00050a010

Kubinyi H (1998) Comparative molecular field analysis (CoMFA). Wiley, New York

Mikolajczyk A et al (2015) Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. Chem Mater 27:2400–2407. doi:10.1021/cm504406a

OECD (2004) OECD principles for the validation, for regulatory purposes, of (Quantitative) structure activity relationship models, 37th joint meeting of the chemicals committee and working party on chemicals, pesticides and biotechnology. Organisation for Economic Co-Operation and Development, Paris

Podlogar BL, Ferguson DM (2000) QSAR and CoMFA: a perspective on the practical application to drug discovery. Drug Des Discov 17:4–12

Puzyn T, Leszczynski J, Cronin MTD (2010) Recent advances in QSAR studies. methods and applications. Challenges and advances in computational chemistry and physics. Springer, New York

Puzyn T, Mostrag-Szlichtyng A, Gajewicz A, Skrzynski M, Worth AP (2011a) Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. Struct Chem 22:795–804. doi:10.1007/s11224-011-9757-4

Puzyn T et al (2011b) Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nat Nanotechnol 6:175–178. doi:10.1038/NNANO.2011.10

Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA (2001) 6-31G* basis set for third-row atoms. J Comput Chem 22:976–984. doi:10.1002/Jcc.1058.Abs

Reynolds CA, Wade RC, Goodford PJ (1989) Identifying targets for bioreductive agents: using grid to predict selective binding regions of proteins. J Mol Graphics 7:103–108. doi:10.1016/S0263-7855(89)80013-X

Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. Chemometr Intell Lab 145:22–29. doi:10.1016/j.chemolab.2015.04.013

Salahinejad M (2015) Nano-QSPR modelling of carbon-based nanomaterials properties. Curr Top Med Chem 15:1868–1886

Sippl W (2010) 3D-QSAR-application, recent advances and limitations. In: Puzyn T, Leszczynski J, Cronin MTD (eds) Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics. Springer, New York, pp 103–126

Sizochenko N, Rasulev B, Gajewicz A, Kuz'min V, Puzyn T, Leszczynski J (2014) From basic physics to mechanisms of toxicity: the "liquid drop" approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. Nanoscale 6:13986–13993. doi:10.1039/c4nr03487b

Sizochenko N, Jagiello K, Leszczynski J, Puzyn T (2015) How the "Liquid Drop" approach could be efficiently applied for quantitative structure-property relationship modeling of nanofluids. J Phys Chem C 119:25542–25547. doi:10.1021/acs.jpcc.5b05759

Talete (2014) Dragon (software for molecular descriptors calculation). http://www.talete.mo.it/. Milano

Todeschini R, Consonni V (2009) Molecular descriptors for chemoinformatics. Wiley, Weinheim

Toropov AA, Toropova AP (2015) Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes. Chemosphere 124:40–46. doi:10.1016/j.chemosphere.2014.10.067

Toropov AA, Toropova AP, Benfenati E, Leszczynska D, Leszczynski J (2010) InChI-based optimal descriptors: QSAR analysis of fullerene[C60]-based HIV-1 PR inhibitors by correlation balance. Eur J Med Chem 45:1387–1394. doi:10.1016/j.ejmech.2009.12.037

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

Chemosphere 89:1098–1102. doi:10.1016/j.chemosphere.2012.05.077

Toropov AA, Toropova AP, Puzyn T, Benfenati E, Gini G, Leszczynska D, Leszczynski J (2013) QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells. Chemosphere 92:31–37. doi:10.1016/j.chemosphere.2013.03.012

Toropova AP, Toropov AA, Benfenati E, Leszczynska D, Leszczynski J (2010) QSAR modeling of measured binding affinity for fullerene-based HIV-1 PR inhibitors by CORAL. J Math Chem 48:959–987. doi:10.1007/s10910-010-9719-x

Toropova AP, Toropov AA, Benfenati E, Leszczynska D, Leszczynski J (2015) QSAR model as a random event: a case of rat toxicity. Bioorgan Med Chem 23:1223–1230. doi:10.1016/j.bmc.2015.01.055

TRIPOS Inc. (2001) Sybyl Molexular Modelling Software Packages, ver. 6.8. St Luis

Tzoupis H, Leonis G, Durdagi S, Mouchlis V, Mavromoustakos T, Papadopoulos MG (2011) Binding of novel fullerene inhibitors to HIV-1 protease: insight through molecular dynamics and molecular mechanics Poisson-Boltzmann surface area calculations. J Comput Aided Mol Des 25:959–976. doi:10.1007/s10822-011-9475-4

Winkler DA, Mombelli E, Pietroiusti A, Tran L, Worth A, Fadeel B, McCall MJ (2013) Applying quantitative structure-activity relationship approaches to nanotoxicology: current status and future potential. Toxicology 313:15–23. doi:10.1016/j.tox.2012.11.005

Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor Chem Acc 120:215–241. doi:10.1007/s00214-007-0310-x