ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2015, *60*(4), 367-370

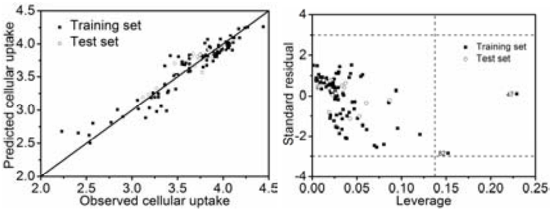
PREDICTION FOR CELLULAR UPTAKE OF MANUFACTURED NANOPARTICLES TO PANCREATIC CANCER CELLS

Wen DAI,[a] Xue-Ying SHAN,[a] Guang-Yu HE[b] and Hai-Qun CHEN[a,*]

[a] School of Environmental and Safety Engineering, Changzhou University, Changzhou, China
[b] Key Laboratory of Advanced Catalytic Materials and Technology, Changzhou University, Changzhou, China

Received February 27, 2015

A nano quantitative structure-activity relationship (nano-QSAR) was developed to predict the cellular uptake of metal oxide nanoparticles to pancreatic cancer cell. A set of 1666 molecular descriptors were calculated by E-Dragon software and used for each molecule in the data set. The major problem of QSAR is descriptor selection. In this paper, genetic algorithm-partial least square (GA-PLS) was used to select the best descriptors. Three descriptors, IC1, Hy and Mor12u, were selected and used as inputs for multiple linear regression (MLR). The established model confirmed by using leave-one-out validation and external validation test showed high statistical significance R^2=0.899, Q^2_{LOO}=0.887, Q^2_{ext}=0.834). The results showed the established model can be used as a powerful tool for prediction of cellular uptake of metal oxide nanoparticles to pancreatic cancer cell.



INTRODUCTION

Nowadays, nanomaterials with novel physicochemical properties have been widely used in many aspects of our daily life, such as cosmetics, self-cleaning textiles and supercapacitors.[1] However, the use of nanoscale particles may endanger human health due to the potential induction of cytotoxicity and ecotoxicity.[2] Therefore, these manufactured nanoparticles should be maintained with appropriate precautions. Çağın *et al.*[3] reported that metal oxide nanoparticles possessed higher toxicity than the bulk size ones did. Thus, it is necessary to evaluate the biological activity or toxicity of manufactured nanoparticles and prepare safe nanomaterials.

The evaluation of biological or toxic effects is usually performed using *in vitro* and *in vivo* studies. Compared with laborious, time-consuming and resource intensive experimental methods, the computational methods have advantages of lower cost and higher speed, which more importantly, will not be affected by the experimental conditions. And quantitative structure-activity relationship (QSAR) is an excellent method on toxicity prediction as it can take almost all relevant structural parameters into consideration. Thus, there is a strong need to extend the traditional QSAR paradigm.

The purpose of this study is to develop a model for cellular uptake prediction of a large set of nanoparticles to pancreatic cancer cell by using GA-PLS for descriptor selection and MLR for modeling. The leave-one-out cross validation and external validation were used to verify the statistical performance of obtained model.

---

[*] Corresponding author: dai.wei6@163.com

RESULTS AND DISCUSSION

To select the optimum number of descriptors, numbers of descriptors were investigated for the result of GA-PLS. As can be seen in Fig. 1, the models with four and five descriptors improved the statistics of the model very slightly, thus a maximum of three descriptors, IC1, Hy and Mor12u, is determined as the optimum subset. The chemical meaning and *MF* of these descriptors are shown in Table 1. The impact of various parameters on the targeted value is determined by *MF*, which is 0.822, 0.050 and 0.118, respectively. According to these values, the importance of the descriptors involved in the model decreases in the following order: IC1>Mor12u>Hy. All *VIF* values of descriptors are less than 5, which indicate that the obtained model has obvious statistical significance.

The selected descriptors by GA-PLS show different aspects of biotoxicity mechanism of metal oxide nanoparticles. It has been established that chemical toxicity is a combination of uptake into or through biological membranes and the interaction of toxicant with cells[4]. Uptake of most organic chemicals to the site of action is made by passive diffusion. Hydrophilicity is the most important factor in diffusion[5]. Selection of the Hy as the most repetitive descriptor supports the importance of this parameter. Valizadeh *et al.*[6] found that the toxicity of quantum dots decreased with their hydrophilicity increasing by surface modification, which is consistent with our results.

IC1 is one of the indices of neighborhood symmetry based on 1-order neighbor degrees and edge multiplicity. It has a considerable power of distinguishing different functional groups and dissimilar structures. As we all know, different electronic nature often drastically alters molecular properties. Therefore, substitution of a given atom of a molecule by another will change the biotoxicity of the compound. The negative coefficient implies the higher number of IC1 of a particular molecule, the lower cellular uptake of the compound.

The other selected descriptor is Mor12u from 3D-MoRSE descriptors which based on the idea of obtaining information from the 3D atomic coordinates by the transform used in electron diffraction. It reveals the effect of the molecular shape in the action of the toxicants and therefore the role of steric effects in interaction of these compounds with the site of action in pancreatic cancer cell.

Table 1

Molecular descriptors selected by GA-PLS

Descriptor	Chemical meaning	*MF*	*VIF*
IC1	Information content index	0.822	1.185
Hy	Hydrophilic factor	0.050	1.443
Mor12u	3D MoRSE-signal 12/unweighted	0.118	1.290

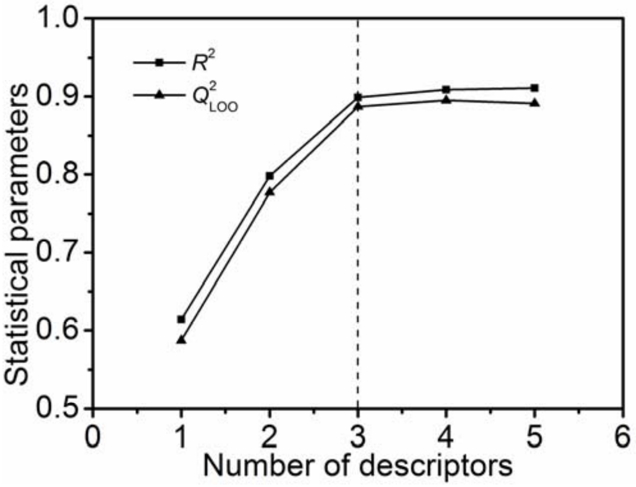


Fig. 1 – Effects of the number of descriptors on the statistical data.

In the presented model, using GA-PLS, all important parameters of hydrophobicity, steric in toxicant diffusion and interaction with the active site are selected.

The model established by MLR on training set is shown as follows.

$$\lg[NP]/\text{cell} = 4.603 - 0.321\text{IC1} - 0.077\text{Hy} + $$
$$+ 0.173\text{Mor12u}$$

$$R^2 = 0.899,\ F = 253.4,\ SEE = 0.146,$$
$$Q^2_{\text{ext}} = 0.834,\ SEP = 0.104,\ Q^2_{\text{LOO}} = 0.887$$

The correlation coefficient for both training and test sets were quite higher than the acceptable limit of 0.5 and close to each other, suggesting the model is highly stable and reliable. The established model was then used to predict the test set data. As can be seen from Fig. 2, the predicted values of lg*[NP]*/cell are in good agreement with the experimental ones, representing good correlations between the parameters.

In order to check the reliability and the stability of QSAR model established by MLR method, the internal and external validation were both conducted. The Q^2_{LOO} was 0.887, exhibiting good robustness. Moreover, predictions realized on the test set were in good agreement with experimental values (Q^2_{ext} 0.834, *SEP* 0.104).

The applicability domain (Fig. 3) analysis indicated there were only two chemicals (47 and 82 in the training set) having leverage higher than the warning h^* value of 0.135 which can be regarded as structural outliers. Fortunately, in this case the data predicted by the model are still good for them as standard residuals less than 3δ.

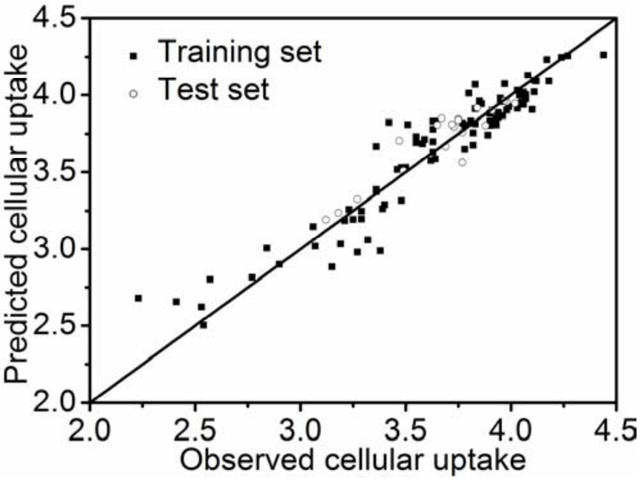


Fig. 2 – Comparison between the predicted and experimental values.

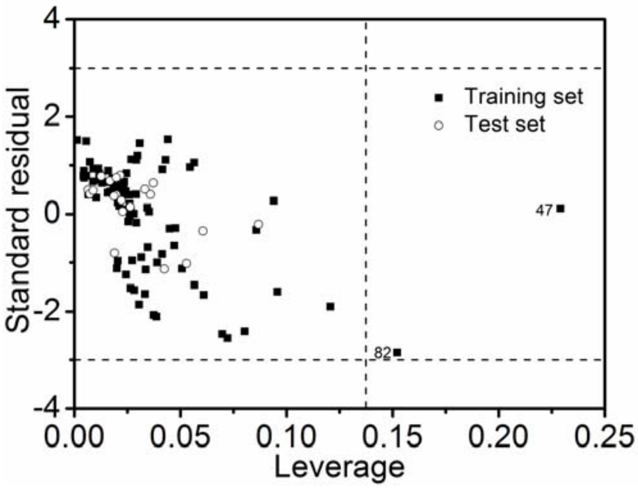


Fig. 3 – William's plot for the regression-based nano-QSAR model.

All the results discussed above show that the presented MLR model can be effectively used to predict the biological activity of nanoparticles to PaCd2 cell.

With applying three descriptors which were selected by GA-PLS, the established MLR model was better than those reported by Epa[7] for these compounds. In the Epa's report, with 19 descriptors, only an R^2 value of 0.76 was obtained. This improvement in toxicity predictability of our model over previously published models was due to the use of GA-PLS as the descriptor selection method in combination with MLR as a powerful modeling tool.

MATERIALS AND METHODS

The data set of cellular uptake expressed as decadic logarithm of the concentration of nanoparticles per cell (lg*[NP]*/cell) has been taken from literature.[8] All samples in the dataset have exactly the same metal core of CLIO-NH_2 decorated with different synthetic small molecules.

Geometries of all molecules attached on the surface of CLIO-NH_2 were optimized by MM+ molecular mechanics force field and the semi-empirical AM1 method.[9] Afterwards, up to 1666 molecular descriptors were calculated by using E-Dragon 1.0 software for each molecule. The molecular descriptors stayed constant for all molecules were eliminated. Then examine pairwise correlations between descriptors so that only the one with the highest correlation was retained (correlation coefficient > 0.95). With hundreds of descriptors remained GA-PLS was used to find the molecular descriptors closely related to cellular uptake. The GA-PLS programs were implemented using the software package PLS-Algorithm Toolbox written by Leardi and Lupiáñez.[10]

The multiple linear regression was performed using the statistical software SPSS. The statistical qualities of MLR equation were judged by parameters as the coefficient of determination (R^2), Fischer statistic (F) and standard deviation (SEE). Also, the coefficient of leave-one-out cross-validation (Q^2_{LOO}) for the training set, the external validation metrics Q^2_{ext} for test set and the standard deviation of prediction (SEP) were used to assess the predictive ability of the proposed model. Multi-collinearity between the selected descriptors was detected by calculating their VIF values,[11] and the contribution of each descriptor was evaluated by calculating the value of membership functions (MF) parameters.[12]

The William's plot was used to assess applicability domain, *i.e.*, the plot of the standardized residuals versus the leverage.[13] A compound with leverage value higher than warning value seriously affects the regression performance, while a value of 3 for standardized residual is commonly used as a cut-off value for accepting predictions.

CONCLUSIONS

GA-PLS was used to search for the molecular descriptors closely related to cellular uptake of CLIO-NH_2 nanoparticles to PaCd2 cell. A set of three optimum descriptors, IC1, Hy and Mor12u, were finally selected and used to establish the model by MLR. The established model confirmed by using leave-one-out validation and external validation test showed high statistical significance (R^2=0.899, Q^2_{LOO} 0.887, Q^2_{ext} 0.834). The results showed the established model can be used as a powerful tool for prediction of cellular uptake of metal oxide nanoparticles to pancreatic cancer cell.

Acknowledgements: The financial support from the National Natural Science Foundation of China (No. 51202020, No. 51472035) is gratefully acknowledged.

REFERENCES

1. J. W. Zhu, S. Chen, H. Zhou and X. Wang, *Nano Res.*, **2012**, *5*, 11-19.
2. A. Gajewicz, B. Rasulev, T. C. Dinadayalane, P. Urbaszek, T. Puzyn, D. Leszczynska and J. Leszczynski, *Adv. Drug Deliver. Rev.*, **2012**, *64*, 1663-1693.
3. T. Çağın, J. W. Che, Y. Qi, Y. H. Zhou, E. Demiralp, G. H. Gao and W. A. Goddard III, *J. Nanopart. Res.*, **1999**, *1*, 51-69.
4. J. W. McFarland, *J. Med. Chem.*, **1970**, *13*, 1192-1196.
5. K. Zarei, M. Atabati and K. Kor, *Bull. Environ. Contam. Toxicol.*, **2014**, *92*, 642-649.
6. A. Valizadeh, H. Mikaeili, M. Samiei, S. M. Farkhani, N. Zarghami, M. Kouhi, A. Akbarzadeh and S. Davaran, *Nanoscale Res. Lett.*, **2012**, *7*, 480-494.
7. V. C. Epa, F. R. Burden, C. Tassa, R. Weissleder, S. Shaw and D. A. Winkler, *Nano Lett.*, **2012**, *12*, 5808-5812.
8. R. Weissleder, K. Kelly, Y. E. Sun, T. Shtatland and L. Josephson, *Nat. Biotechnol.*, **2005**, *23*, 1418-1423.
9. Hypercube HyperChem, Inc., http://www.hyper.com.
10. R. Leardi and A. L. González, *Chemometr. Intell. Lab.*, **1998**, *41*, 195-207.
11. C. Rücker, G. Rücker and M. Meringer, *J. Chem. Inf. Model*, **2007**, *47*, 2345-2357.
12. M. Dolatabadi, M. Nekoei and A. Banaei, *Monatsh. Chem.*, **2010**, *141*, 577-588.
13. P. Gramatica, *QSAR Comb. Sci.*, **2007**, *26*, 694-701.