I&EC research
Industrial & Engineering Chemistry Research

Article

pubs.acs.org/IECR

Modeling the Cellular Uptake of Magnetofluorescent Nanoparticles in Pancreatic Cancer Cells: A Quantitative Structure Activity Relationship Study

Mehdi Ghorbanzadeh,*,† Mohammad H. Fatemi, and Masoumeh Karimpour

Chemometrics Laboratory, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran

Supporting Information

ABSTRACT: An artificial neural network was employed to predict the cellular uptake of 109 magnetofluorescent nanoparticles (NPs) in pancreatic cancer cells on the basis of quantitative structure activity relationship method. Six descriptors chosen by combining self-organizing map and stepwise multiple linear regression (MLR) techniques were used to correlate the nanostructure of the studied particles with their bioactivity using MLR and multilayered perceptron neural network (MLP-NN) modeling techniques. For the MLR and MLP-NN models, the correlation coefficient was 0.769 and 0.934, and the root-mean-square error was 0.364 and 0.150, respectively. The results obtained after a leave-many-out cross-validation test revealed the credibility of MLP-NN for the prediction of cellular uptake of NPs. In addition, sensitivity analysis of MLP-NN model indicated that the number of hydrogen-bond donor sites in the organic coating of a NP is the predominant factor responsible for cellular uptake.

INTRODUCTION

Nanoparticles are microscopic particles having one or more dimensions of 100 nm or less. They are used in a variety of fields due to their unique physical and chemical properties.[1,2] In cosmetics, for instance, nanosized titanium oxide and zinc oxide particles are employed in place of bulk materials to generate products with improved texture and greater skin penetration.[3] In environmental monitoring, NP-based sensors can be used as potential detectors of toxins, heavy metals, and organic pollutants in air, water, and soil.[4] In the field of medicine or biology, NPs have been used for the detection of proteins[5] and probing of DNA structure.[6] They are also of great interest for use as imaging tools,[7] phototherapy agents,[8] gene delivery carriers,[9,10] drug and energy delivery vehicles for therapeutics and theranostics,[11,12] cell separation systems for in vitro diagnostics,[13] and fluorescent biological labels.[14,15] A summary of nanoparticle technologies in medicine and biology can be found elsewhere.[16,17] The extraordinary properties such as shape, large surface area, surface charge, size, porosity, and surface chemistry give NPs new biological activities, which can be either beneficial (e.g., penetration of cellular barrier for drug delivery) or harmful (e.g., toxicity). There are some reports on adverse effects of NPs on humans and the environment. For example, Klaine et al. reviewed toxic effects of NPs on freshwater and soil organisms, and bacteria.[18] Moller and colleagues reported in vitro high toxicity and the ability of copper oxide NPs to cause DNA damage and oxidative lesions in human lung epithelial cell line A549.[19] Drezek et al. also reviewed cytotoxicity of carbon-based, metal-based, and semiconductor-based NPs developed for biomedical application.[20]

To find the effect of the NP's structure on its activity/property and also to accelerate the process of designing safe and efficient NPs, developing chemoinformatics methods such as quantitative structure activity/property relationship (QSAR/QSPR) seems to be useful. This method is based on the assumption that the variation in the properties or biological activities of a NP can be correlated with changes in its molecular structure. This method can be used to predict the activity/property of newly synthesized NPs without resorting to experimentation. With the help of QSAR/QSPR method as a time- and money-saving technique, the number of animal experiments would be reduced as well. There are some reports on the application of the quantitative structure property relationship technique to predict the solubility of C60 in different solvents and Young's modulus values for inorganic nanoparticles.[21−29] However, there are a very limited number of reports on the establishment of a quantitative structure activity relationship model to correlate molecular nanostructures with activities of NPs. Tropsha and colleagues have recently developed a QSAR model to predict the cellular uptake of 109 NPs in pancreatic cancer cells (PaCa2).[30] Each NP possessed the same metal core but different organic coatings.[31] The authors calculated 150 molecular descriptors for all 109 organic molecules and employed the *k*-nearest neighbors (kNN) method to correlate biological activity of NPs with the calculated descriptors. The values of squared correlation coefficient obtained for external prediction sets ranged from 0.65 to 0.80. They concluded that QSAR is a useful tool to predict the cellular behavior of a NP library based on a common core and that such a model can be useful to design a specific coating material to control the cellular uptake. In the same study, a support vector machine (SVM) was used to build a binary classification model using a set of 51 NPs with different metal core and surface modifications.[32] They used an arbitrary

Received: March 15, 2012
Revised: June 18, 2012
Accepted: July 24, 2012
Published: July 24, 2012

dx.doi.org/10.1021/ie3006947 | *Ind. Eng. Chem. Res.* 2012, 51, 10712−10718

threshold to divide the data set into two groups, each containing the same number of NPs. Four experimental descriptors including size, R1 and R2 relaxivities (representing NP's magnetic properties), and zeta potential (representing the intensity of charge on their surface) were used to classify the NPs into two clusters. The authors reported external prediction power of 73% for the classification modeling. They finally indicated that QSAR analysis can be used to predict the bioactivity of NPs based on descriptors computed for either organic coatings (in the case of NPs with the same metal core) or whole NP (in the case of NPs with diverse core and surface modifier). In a more recently published study, Puzyn et al. developed and validated a model to describe the relationship between the structures of 17 metal oxide NPs and their cytotoxicity to bacteria *Escherichia coli*.[33] They calculated 12 descriptors quantitatively describing the variability of the NPs' structure. Multiple linear regression was used to predict EC_{50} values of the studied nanomaterials. The best MLR model was built using one descriptor selected by genetic algorithm. The MLR model was given by the following equation:

$$\log(1/EC_{50}) = 2.59 - 0.50\Delta H_{Me^+} \quad (1)$$

where ΔH_{Me^+} is the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure:

$$Me(s) \rightarrow Me^{n+}(g) + n{\cdot}e^- \Delta H_{Me^+} \quad (2)$$

The authors reported the statistics of $R^2 = 0.85$, with the cross-validated regression coefficient = 0.77, and the externally validated regression coefficient = 0.83 for the MLR model (eq 1). The results indicated satisfactory goodness of fit, robustness, and predictive ability of the model.

A certain number of modeling techniques have been found useful for the establishment of the relationship between the molecular structures and biological activities of NPs such as MLR, SVM, and kNN.[30,33] However, the artificial neural network as a powerful nonlinear mapping technique has not been used for predicting the biological activity of NPs so far. Neural networks are machine learning techniques applicable to both classification and regression. Because of flexibility, they are capable of discovering complex nonlinear relationships in experimental data. The aim of this investigation was to provide an application of multilayered perceptron neural network (MLP-NN) to predict the cellular uptake of NPs in PaCa2 using calculated molecular descriptors as inputs. To select the optimal combination of the molecular descriptors, a two-step variable selection procedure was performed. First, the self-organizing mapping (SOM) method was used as an unsupervised feature reduction technique to eliminate redundant descriptors. In the second step, a stepwise MLR method was used as a feature selection technique to choose the most relevant descriptors correlating nanostructures and their bioactivity. Results obtained by MLP-NN were compared to those given by MLR. The satisfactory results in training and test sets proved MLP-NN to be a useful and powerful technique in the field of QSAR analysis of nanomaterials.

MATERIALS AND METHODS

Experimental Data. The experimental values of cellular uptake of 109 magnetofluorescent NPs in PaCa2 were taken from ref 30 and are listed in Table S1 (Supporting Information). All NPs in the data set have exactly the same metal core decorated with different synthetic small molecules. Each NP is represented by the structure of organic surface modifier, which in turn is characterized by conventional molecular descriptors. Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell, which varies from 2.23 to 4.44. The compounds in the data set were randomly divided into the training, internal, and external test sets, consisting of 90, 19, and 10 members, respectively. The training set and internal test set participated in the generation of the MLP-NN model and adjusted its parameters. The external test set was used to validate the predictive performance of the model. In the case of MLR modeling, both internal and external test sets were considered as the prediction set.

Molecular Descriptors. Molecular descriptors are the simple mathematical representation of a molecule and are used to encode significant structural features of molecules. To calculate descriptors, the Hyperchem program (ver. 7)[34] was applied to construct all molecular structures. Molecular geometry was optimized with the Austin Model 1 (AM1) semiempirical method. Dragon[35] was then used for calculation of a wide variety of molecular descriptors. Calculated descriptors were analyzed for the existence of a constant or near constant values, and the descriptors with low variation were removed from more investigation. The remaining descriptors were gathered in a 109 × 256 matrix, where 109 and 256 are the number of NPs and molecular descriptors, respectively. For further reduction of descriptor space, SOM network was then trained with transpose of the data matrix, and the number of descriptors was reduced from 256 to 66. Finally, the stepwise MLR technique was used for selection of the most significant descriptors from the remaining molecular descriptors. In the stepwise MLR technique, a set of models is examined for stability and validity through a variety of statistical methods. The best MLR model is the one with high multiple correlation coefficient (R), high F-statistic, low standard error (SE), least number of descriptors, and high ability for prediction. On the basis of these criteria, six descriptors were selected and used for linear and nonlinear model generation.

Multiple Linear Regression. MLR is a technique used to model the linear relationship between a dependent variable y (here cellelar uptake) and one or more independent variables x_i, that is, molecular descriptors, as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (3)$$

The coefficients vector b is calculated using descriptor matrix X, containing an additional column with ones to calculate coefficient b_0, according to the following equation:

$$b = (X^TX)^{-1}X^Ty \quad (4)$$

It is worth noting that MLR is based on least-squares; that is, the model is fitted such that the sum of squares of differences of experimental and predicted values is minimized.

Multilayered Perceptron Neural Network. Generally, each neural network is built from several layers: one input layer, one or more hidden layers, and one output layer. The node in each layer is connected to the nodes of the next layer by weights. During training, these weights are iteratively adjusted to minimize the network errors. The flexibility of ANNs enables them to discover nonlinearity in experimental data. A detailed description of the theory behind neural networks is described elsewhere.[36,37] In MLP-NN, the inputs are fully connected to the hidden layer, and hidden layer neurons are fully connected to the outputs. In the learning phase, the sequenced input

patterns, presented to the network, are propagated in the forward direction layer by layer until the final layer output is computed. The error is computed as the difference between the calculated output and the target value. These calculated errors are taken as input pattern for feedback connections from which the synaptic weights are adjusted in backward direction layer by layer.[38] The trained network was used as an analytical tool to predict the cellular uptake values of NPs. A detailed description of this type of ANN modeling has been published elsewhere.[39−41]

RESULTS AND DISCUSSION

Self-organizing mapping network was used as a feature reduction technique to eliminate redundant descriptors. The stepwise MLR method was then applied to the remaining descriptors to select the relevant ones for further investigation. Descriptors selected by this procedure (SOM-MLR) are number of donor atoms for H-bonds (N and O) (nHDon), Geary autocorrelation of lag 1 weighted by van der Waals volume (GATS1v), 3D-MoRSE-signal 29/unweighted (Mor29u), D total accessibility index/weighted by Sanderson electronegativity (De), 3D-MoRSE-signal 14/unweighted (Mor14u), and mean electrotopological state (Ms). nHDon belongs to count descriptors, which are simple molecular descriptors based on counting the defined element of a compound. It is a measure of the hydrogen-bonding ability of a molecule expressed in terms of number of possible hydrogen-bond donors. In particular, it is calculated as the count of hydrogen atoms bonded to oxygen and nitrogen atoms in the molecule.[42] Inspection of the descriptor values shows that the NPs with low cellular uptake have a high number of heteroatoms (i.e., oxygen and nitrogen) possessing a proton that can be donated. The second descriptor is Geary autocorrelation descriptor at a topological distance equal to 1 and weighted by van der Waals volume.[42] It is among the 2D-autocorrelation descriptors, which encode structural features such as branching patterns and the presence of the substituents and rings. Mor29u and Mor14u are 3D-MoRSE (molecule representation of structures based on electron diffraction) descriptors.[42] They are representations of the three-dimensional structures of a molecule and encode structural features such as mass and amount of branching. The next descriptor (De) belongs to three-dimensional WHIM (Weighted Holistic Invariant Molecular) descriptors, which contain information about the whole three-dimensional molecular structure in terms of size, shape, symmetry, and atom distribution.[42] The last descriptor (mean electrotopological state) is calculated by dividing the sum of Kier−Hall electro-topological states by the number of non-hydrogen atoms. It is among the constitutional type of molecular descriptors that are used to describe the effect of different fragments of the molecules.[42] This kind of descriptors reflects the molecular composition of a compound without any information about its molecular geometry. The correlation matrix among the selected descriptors is presented in Table S2 (Supporting Information). Inspection of this matrix shows that there is not any high correlation between the selected results. These descriptors were used to build linear and nonlinear models.

To make sure the structures of the training and test sets represent those of the whole data set, diversity analysis[43,44] was performed on the data set. In this way, the mean distance of one chemical to the remaining ones ($\bar{d}$) was computed from the descriptor space matrix as follows:

$$\bar{d} = \frac{\sum_{j=1}^{n} d_{ij}}{n-1} i = 1, 2, ..., n \quad (5)$$

where d_{ij} is a distance score for two different compounds, which can be measured by the Euclidean distance norm based on the compound descriptors (x_{ik} and x_{jk}):

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})} \quad (6)$$

Next, the mean distances were normalized within the interval [0, 1]. The closer to 1 is the distance, the more diverse each compound is from the others. For the used data set, the mean distances of samples versus cellular uptake values are plotted in Figure 1, which illustrates the diversity of the molecules in training, internal test, and external test sets. As can be seen from this figure, the structures of compounds are diverse in all subsets. It warrants model stability and that the external test set is suitable to assess the predictive performance of the developed model.

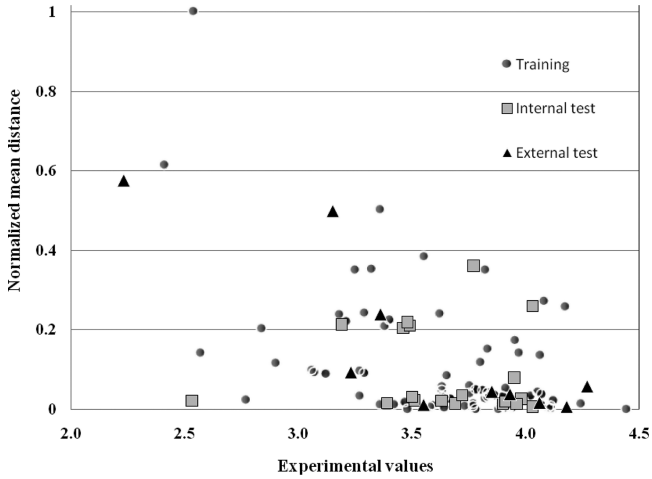

Figure 1. Scatter plot of normalized mean distance of samples versus experimental cellular uptake.

Linear Modeling. The MLR model was obtained using the descriptors selected by the SOM-MLR procedure. Table S3 (Supporting Information) shows the specification of the MLR model. It can be seen from the table that P values are smaller than 0.05 and the tolerance values are greater than 0.1, which means that both of them fulfill the criteria necessary for a MLR model applicable for validation.[45] The regression coefficients obtained from training of MLR model were used to calculate the cellular uptake of NPs for prediction set. The calculated values of cellular uptake for training and prediction sets are shown in Table S1 (Supporting Information). This model gave a RMSE of 0.364, and the corresponding correlation coefficient was 0.769 for the whole data set. Also, the leave-many-out cross-validation test was performed to validate the MLR model. The statistical parameters of this calculation are shown in Table 1.

Nonlinear Modeling. Nonlinear behavior of descriptors can be assessed by visual inspection of three-dimensional surface plots. The network output was plotted against two input descriptors to generate a functional dependence surface. This gives the idea of how the network output alters in response to the two selected input variables.[46] Figure 2 shows functional

Table 1. Statistical Parameters of MLR and MLP-NN Models[a]

	model				
	MLR		MLP-NN		
	training	prediction	training	int. test	ext. test
R	0.782	0.755	0.934	0.945	0.943
RMSE	0.369	0.357	0.146	0.121	0.214
F	122.430	35.890	531.138	141.796	64.717
SE	0.258	0.3035	0.148	0.123	0.217
Q^2	0.577		0.655		
SPRESS	0.278		0.251		

[a]Int. test and ext. test refer to internal and external test sets, respectively.

dependence surfaces of descriptors used for building nonlinear model. On the basis of the resulting plots, the biological activity of NPs has nonlinear dependence on selected descriptors, suggesting the complex relationship between the input descriptors and cellular uptake values. Thus, the neural network mapping technique, which is capable of mapping complex patterns of nonlinearity, is useful to model the nonlinear relationship between selected descriptors and cellular uptake of NPs in PaCa2. Therefore, a three-layer network with a hyperbolic tangent transfer function in both hidden and output layers was designed and trained by Quasi-Newton algorithm. The number of neurons in the hidden layer was optimized according to the procedure explained in our previous work.[47] The developed 6-10-1 network was trained for adjusting the weights and bias values. In this step, an internal test set was used to prevent overtraining. In the next step, the trained network was used to predict the cellular uptake values for the compounds of the internal and external test sets, respectively. The MLP-NN calculated values of cellular uptake for the training, internal, and external test sets are presented in Table S1 (Supporting Information). The calculated RMSE and R values for the whole data set were 0.150 and 0.934, respectively. The statistical parameters obtained for the MLP-NN model are also shown in Table 1. Comparison among the values in this table implies that the MLP-NN model produces better statistical results in terms of R, RMSE, and Fisher statistic value over the MLR model. This strongly suggested a nonlinear relationship between the selected descriptors and cellular uptake values of NPs. To further evaluate the MLP-NN model, internal validation (leave-many-out cross-validation test) was performed, and the values of the cross-validation correlation coefficient (Q^2) and standard deviation based on predicted residual sum of square (SPRESS) were calculated for the constructed MLP-NN model. The obtained values of Q^2 and SPRESS are shown in Table 1. Inspecting these values reveals the successful predictive power of the neural network model. To further assess the predictive ability of the model, the following criteria suggested by Golbraikh and Trospsha[48,49] and Roy and Roy[50] were calculated:

$$q^2 > 0.5 \tag{7}$$

$$R^2 > 0.6 \tag{8}$$

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \text{or} \frac{(R^2 - R_0'^2)}{R^2} < 0.1 \tag{9}$$

$$0.85 \ll k \ll 0.1 \text{or} 0.85 \ll k' \ll 1.15 \tag{10}$$

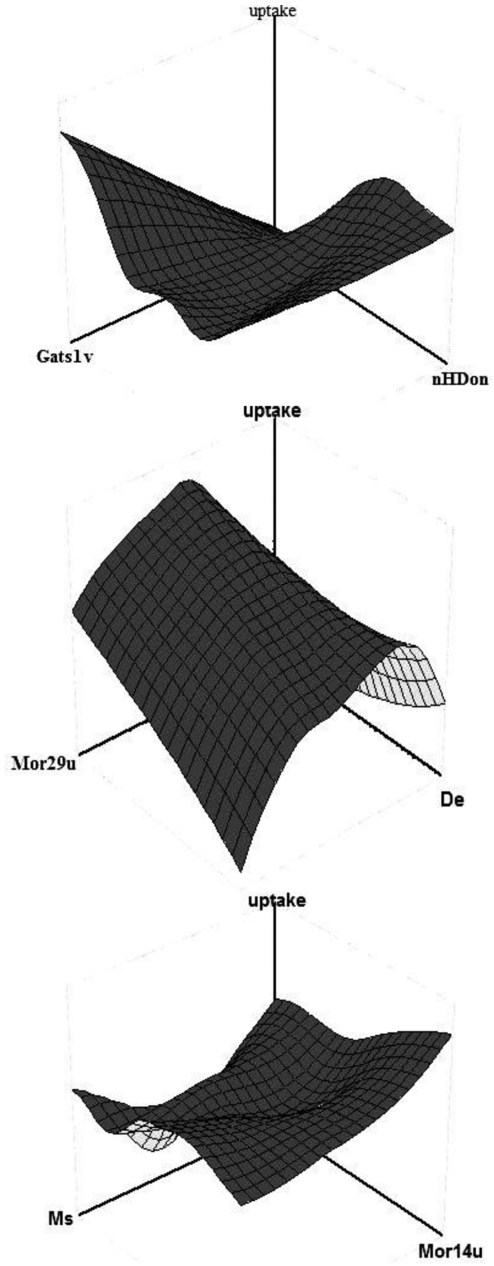

Figure 2. Functional dependence surfaces for selected descriptors. On each surface, the values of input descriptors were varied through their range, while the other four inputs were held constant. On the basis of these plots, it seems that cellular uptake of NPs in PaCa2 is a nonlinear function of selected descriptors.

$$R_m^2 = R^2(1 - |\sqrt{R^2 - R_0^2}|) > 0.5 \tag{11}$$

$$|R_0^2 - R_0'^2| < 0.3 \tag{12}$$

Definitions of calculated parameters are presented obviously in ref 38. The statistical values for the external validation set used in MLP-NN modeling were $q^2 = 0.977$, $R^2 = 0.889$, $k = 1.016$, $k' = 0.981$, $(R^2 - R_0^2)/R^2 = R_m^2 = 0.607$, and $|R_0^2 - R_0'^2 = 0.004|$. The obtained values of the model are in good agreement with the limits described above, demonstrating once again the high predictive ability of MLP-NN model. As this model fulfills the above criteria and also positively passes internal and external validation, it can be used to predict the cellular uptake of newly synthesized and untested NPs. Figure 3 shows the plot of the

dx.doi.org/10.1021/ie3006947 | Ind. Eng. Chem. Res. 2012, 51, 10712−10718

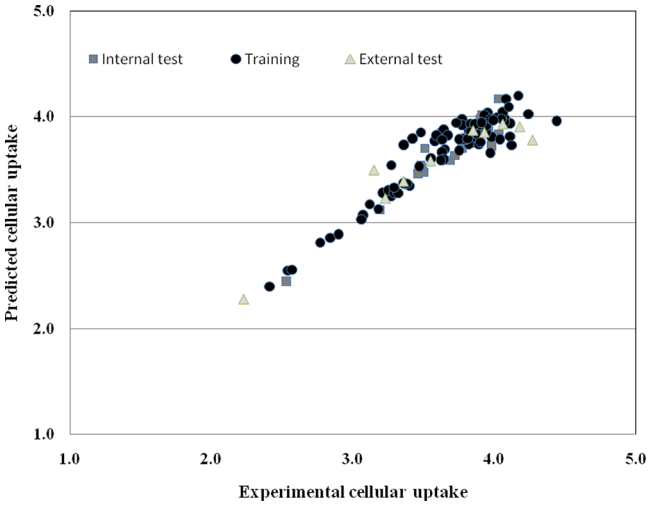

Figure 3. Plot of MLP-NN calculated versus experimental values of cellular uptake.

MLP-NN predicted values of cellular uptake against the experimental values. As can be seen, the agreement between measured and predicted results across the entire range of values is very good. The residuals of the predicted values are plotted against the experimental values in Figure 4. Propagation of the

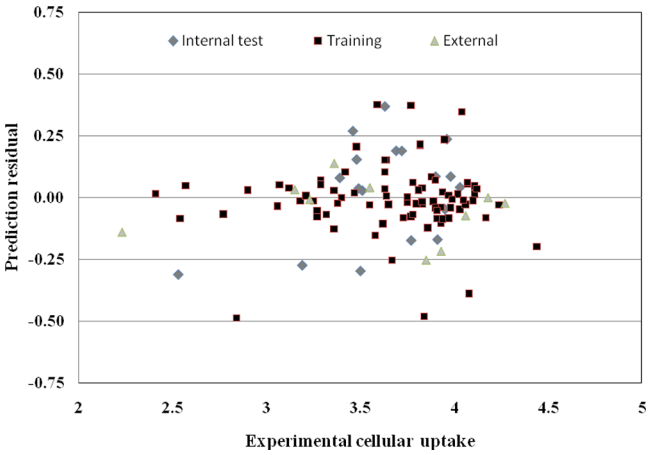

Figure 4. Plot of prediction residuals versus experimental values of cellular uptake.

residuals on both sides of the zero line indicates that no systematic error exists in the MLP-NN model. The better prediction results of MLP-NN method support our assumption that a nonlinear relationship may exist between selected descriptors and cellular uptake values of magnetofluorescent NPs. To determine the relative importance of each descriptor in the nonlinear MLP-NN model, sensitivity analysis was performed. In this approach, the sum of squares of residuals for the model is computed when the respective descriptor is eliminated from the MLP-NN model. Next, differences between the full model and the reduced models are calculated, and the descriptors are sorted by their importance for the neural network. According to the results of sensitivity analysis, the order of descriptor importance was nHDon ≫ GATS1v > Mor29u > Mor14u > De > Ms. On the basis of this result, the number of hydrogen-bond donors was found to be the most important factor, which should be considered to synthesize a new organic modifier to control PaCa2 cellular uptake of NPs. Also, molecular shape and size, and amount of branching in organic caotings, can be effective factors in cellular uptake of studied NPs in pancreatic cancer cells.

Applicability Domain of the Model. In addition to validating the predictive ability of the MLP-NN model, analysis of the applicability domain was performed. This analysis is useful to determine if the QSAR model is capable of predicting the activity of new compounds with unavailable experimental data. The applicability domain of a QSAR model is defined as the response and chemical structure spaces in which the model makes predictions with a given reliability.[51] In this work, two approaches were used for the evaluation of the structural applicability domain. The first approach was based on the ranges of individual descriptors used for the building model. It is the simplest method for describing the applicability domain. According to this method, an NP with descriptor values within the range of those of the training set compounds is considered as being inside the applicability domain of the model.[52] The ranges of descriptors calculated for the compounds of training, internal and external test sets are shown in Table 2. It depicts that all NPs in internal and external test sets are inside the applicability domain of the proposed model. The second method was based on the leverage approach and Williams plot.[49,52,53] The leverage or hat value is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i (i = 1, \ldots, n) \quad (13)$$

where h_i is the leverage or hat value of the compound (i) in the descriptor space, x_i is the descriptor raw-vector of the query compound, and X is the descriptor matrix. The superscript T refers to the transpose of the matrix and vector. The observation that a chemical has a leverage value greater than the warning leverage (h^*) indicates that the chemical falls outside the applicability domain. The leverage value greater than h^* also means that the predicted response is the result of extrapolation of the model and, therefore, may not be reliable. The warning leverage is calculated as follows:

$$h^* = 3(p)/n \quad (14)$$

where p is the number of model parameters, and n is the number of training compounds. As can be seen from Figure 5, none of the compounds exceed the warning value of h^* = 0.263. On the other hand, the hat value less than h^* does not

Table 2. Ranges of Descriptors Calculated for Compounds[a]

	nHDon		GATS1v		Mor29u		De		Mor14u		Ms	
	min	max	min	max	min	max	min	max	min	max	min	max
training	0	8	0.52	1.73	−0.64	0.35	0.36	0.62	−1.04	4.24	1.05	5.57
int. test	0	5	0.80	1.63	−0.23	0.28	0.35	0.54	−0.34	3.07	1.87	3.62
ext. test	0	6	0.79	1.59	−0.31	0.32	0.39	0.53	−0.19	2.62	1.92	3.93

[a]Int. test and ext. test refer to internal and external test sets, respectively.

dx.doi.org/10.1021/ie3006947 | Ind. Eng. Chem. Res. 2012, 51, 10712−10718

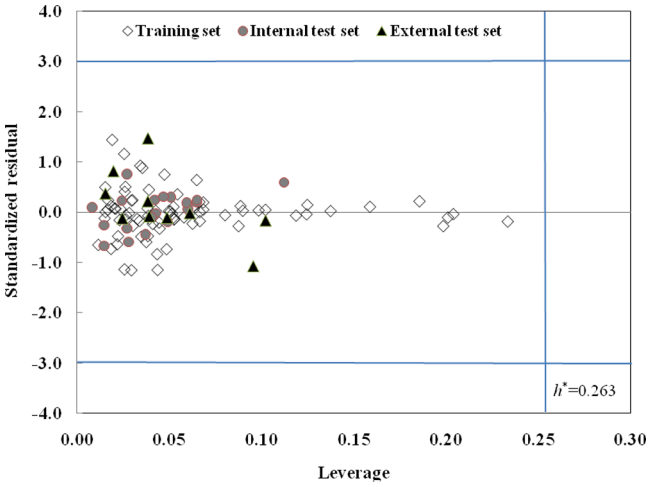

Figure 5. William plot; plot of standardized residuals versus hat values, with a warning leverage of 0.263.

necessarily guarantee that the chemicals fall within the applicability domain because compounds may be located outside the applicability domain due to large values of standardized residuals (values more than 3 times the standard deviation unit). To visualize the applicability domain based on both leverage and standardized residual, the Williams plot is used. It is a plot of standardized residuals versus leverage values with two boundary lines, the first for outliers and the second for high leverage compounds (Figure 5). As can be seen, no significant outlier and high leverage compound is detected by analyzing the William plot. Therefore, the MLP-NN model can be applied to predict the cellular uptake of NPs if the following conditions are satisfied:

(1) The values of the new compound's descriptors fall within the ranges of descriptors used for training the model.
(2) The calculated leverage value for the new compound is not higher than the warning value of $h^* = 0.263$.

■ ASSOCIATED CONTENT

Supporting Information

Additional tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +98 112 5342350. Fax: +98 112 5342350. E-mail: m.ghorbanzade@umz.ac.ir, mehdi.ghorbanzadeh@chem.umu.se.

Present Address

[†]Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Nel, A.; Xia, T.; Madler, L.; Li, N. Toxic potential of materials at the nanolevel. *Science* **2006**, *311*, 622−627.
(2) Auffan, M.; Rose, J.; Bottero, J.; Lowry, G.; Jolivet, J.; Wiesne, M. Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective. *Nat. Nanotechnol.* **2009**, *4*, 634−641.
(3) Newman, M. D.; Stotland, M.; Ellis, J. I. The safety of nanosized particles in titanium dioxide and zinc oxideebased sunscreens. *J. Am. Acad. Dermatol.* **2009**, *61*, 686−692.
(4) Wang, L.; Ma, W.; Xu, L.; Chen, W.; Zhu, Y.; Xu, C.; Kotov, N. A. Nanoparticle-based environmental sensors. *Mater. Sci. Eng., R* **2010**, *70*, 265−274.
(5) Nam, J. M.; Thaxton, C. C.; Mirkin, C. A. Nanoparticles-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **2003**, *301*, 1884−1886.
(6) Mahtab, R.; Rogers, J. P.; Murphy, C. J. Protein-sized quantum dot luminescence can distinguish between "straight", "bent", and "kinked" oligonucleotides. *J. Am. Chem. Soc.* **1995**, *117*, 9099−9100.
(7) Lewin, M.; Carlesso, N.; Tung, C.; Tang, X.; Cory, D.; Scadden, D. T.; Weissleder, R. Tat peptide-derivatized magnetic nanoparticles allow in vivo tracking and recovery of progenitor cells. *Nat. Biotechnol.* **2000**, *18*, 410−414.
(8) El-Sayed, I. H.; Huang, X. H.; El-Sayed, M. A. Selective laser photo-thermal therapy of epithelial carcinoma using anti-egfr antibody conjugated gold nanoparticles. *Cancer Lett.* **2006**, *239*, 129−135.
(9) Rosi, N. L.; Giljohann, D. A.; Thaxton, C. S.; Lytton-Jean, A. K. R.; Han, M. S.; Mirkin, C. A. Oligonucleotide-modified gold nanoparticles for intracellular gene regulation. *Science* **2006**, *312*, 1027−1030.
(10) Han, G.; You, C. C.; Kim, B.; Turingan, R. S.; Forbes, N. S.; Martin, C. T.; Rotello, V. M. Light-regulated release of DNA and its delivery to nucleiby means of photolabile gold nanoparticles. *Angew. Chem., Int. Ed.* **2006**, *45*, 3165−3169.
(11) Zhang, L.; Gu, F. X.; Chan, J. M.; Wang, A .Z.; Langer, R. S.; Farokhzad, O. C. Nanoparticles in medicine: therapeutic applications and developments. *Clin. Pharmacol. Ther.* **2007**, *83*, 761−769.
(12) Shubayev, V. I.; Pisanic, T. R.; Jin, S. Magnetic nanoparticles for theragnostics. *Adv. Drug Delivery Rev.* **2009**, *61*, 467−477.
(13) McCloskey, K. E.; Chalmers, J. J.; Zborowski, M. Magnetic cell separation: Characterization of magnetophoretic mobility. *Anal. Chem.* **2003**, *75*, 6868−6874.
(14) Chan, W. C. W.; Nie, S. M. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **1998**, *281*, 2016−2018.
(15) Wang, S.; Mamedova, N.; Kotov, N. A.; Chen, W.; Studer, J. Antigen/antibody immunocomplex from CdTe nanoparticle bioconjugates. *Nano Lett.* **2002**, *2*, 817−822.
(16) Murthy, S. K. Nanoparticles in modern medicine: State of the art and future challenges. *Int. J. Nanomed.* **2007**, *2*, 129−141.
(17) Salata, O. V. Applications of nanoparticles in biology and medicine. *J. Nanobiotechnol.* **2004**, *2*, 3−8.
(18) Klaine, S. J.; Alvarez, P. J. J.; Batley, G. E.; Fernandes, T. F.; Handy, R. D.; Lyon, D. Y.; Mahendra, S.; McLaughlin, M. J.; Lead, J. A. Nanomaterials in the environment: behavior, fate, bioavaibility, and effects. *Environ. Toxicol. Chem.* **2008**, *27*, 1825−1851.
(19) Karlsson, H. L.; Cronholm, P.; Gustafsson, J.; Moller, L. Copper oxide nanoparticles are highly toxic: A comparison between metal oxide nanoparticles and carbon nanotubes. *Chem. Res. Toxicol.* **2008**, *21*, 1726−1732.
(20) Lewinski, N.; Colvin, V.; Drezek, R. Cytotoxicity of nanoparticles. *Small* **2008**, *4*, 26−49.
(21) Sivaraman, N.; Srinivasan, T. G.; Vasudeva Rao, P. R. QSPR modeling for solubility of fullerene (C60) in organic solvents. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1067−1074.
(22) Liu, H.; Yao, X.; Zhang, R.; Liu, M.; Hu, Z.; Fan, B. Accurate quantitative structure-property relationship model to predict the solubility of C_{60} in various solvents based on a novel approach using a least-squares support vector machine. *J. Phys. Chem. B* **2005**, *109*, 20565−20571.
(23) Kiss, I.; Z.; Mandi, G.; Beck, M. T. Artificial neural network approach to predict the solubility of C_{60} in various solvents. *J. Phys. Chem. A* **2000**, *104*, 8081−8088.
(24) Danauskas, S. M.; Jurs, P. C. Prediction of C_{60} solubilities from solvent molecular structures. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 419−424.

(25) Martin, D.; Maran, U.; Sild, S.; Karelson, M. QSPR modeling of solubility of polyaromatic hydrocarbons and fullerene in 1-octanol and *n*-heptane. *J. Phys. Chem. B* **2007**, *111*, 9853−9857.
(26) Toropov, A. A.; Toropova, A. P.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. Additive InChI-based optimal descriptors: QSPR modeling of fullerene C_{60} solubility in organic solvents. *J. Math. Chem.* **2009**, *46*, 1232−1251.
(27) Murray, J. S.; Gagarin, S. G.; Politzer, P. Representation of C_{60} solubilities in terms of computed molecular surface electrostatic potentials and areas. *J. Phys. Chem.* **1995**, *99*, 12081−12083.
(28) Toropov, A. A.; Leszczynska, D.; Leszczynski, J. QSPR study on solubility of fullerene C_{60} in organic solvents using optimal descriptors calculated with SMILES. *Chem. Phys. Lett.* **2007**, *441*, 119−122.
(29) Toropov, A. A.; Leszczynski, J. A new approach to the characterization of nanomaterials: Predicting Young's modulus by correlation weighting of nanomaterials codes. *Chem. Phys. Lett.* **2006**, *433*, 125−129.
(30) Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S. Y.; Mumper, R. J.; Tropsha, A. Quantitative nanostructure_activity relationship modeling. *ACS Nano* **2010**, *4*, 5703−5712.
(31) Weissleder, R.; Kelly, K.; Sun, E. Y.; Shtatland, T.; Josephson, L. Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nat. Biotechnol.* **2005**, *23*, 1418−1423.
(32) Shaw, S. Y.; Westly, E. C.; Pittet, M. J.; Subramanian, A.; Schreiber, S. L.; Weissleder, R. Perturbational profiling of nanomaterial biologic activity. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7387−7392.
(33) Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat. Nanotechnol.* **2011**, *6*, 175−178.
(34) *HyperChem Release 7.0 for Windows*; Hypercube, Inc.: 2002.
(35) http://www.disat.unimib.it/chem.
(36) Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*; Wiley-VCH: Weinheim, 1999.
(37) Bose, N. K.; Liang, P. *Neural Network Fundamentals*; McGraw-Hill: New York, 1996.
(38) Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice-Hall: NJ, 1999.
(39) Egmont-Petersen, M.; Talmo, J. L.; Hasman, A. Assessing the importance of features for multi-layer perceptrons. *Neural Networks* **1998**, *11*, 623−635.
(40) Bologna, G.; Pellegrini, C. Three medical examples in neural network rule extraction. *Phys. Med.* **1997**, *13*, 183−187.
(41) Rathbun, T. F.; Rogers, S. K.; DeSimio, M. P. MLP iterative construction algorithm. *Neurocomputing* **1997**, *17*, 195−216.
(42) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Germany, 2000.
(43) Rouvray, D. H. Definition and role of similarity concepts in the chemical and physical sciences. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 580−586.
(44) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B. T. Molecular similarity and diversity in chemoinformatics: from theory to applications. *Mol. Diversity* **2006**, *10*, 39−79.
(45) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69−77.
(46) Agatonovic-Kustrin, S.; Tucker, I. G.; Zecevic, M.; Zivanovic, L. J. Prediction of drug transfer into human milk from theoretically derived descriptors. *Anal. Chim. Acta* **2000**, *418*, 181.
(47) Ghorbanzad'e, M.; Fatemi, M. H.; Karimpour, M.; Andersson, P. L. Quantitative and qualitative prediction of corneal permeability for drug-like compounds. *Talanta* **2012**, *85*, 2686−2694.
(48) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graphics Modell.* **2002**, *20*, 269−276.
(49) Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inf.* **2010**, *29*, 476−488.
(50) Roy, P. P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **2008**, *27*, 302−313.
(51) Netzeva, T. N.; Worth, A. P.; Aldenberg, T.; Benigni, A.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D.; Schultz, T. W.; Stanton, D.; van de Sandt, J. J.; Tong, W.; Veith, G.; Yang, C. Current status of methods for defining the applicability domain of (quantitative) structure−activity relationships. *ATLA, Altern. Lab. Anim.* **2005**, *52*, 1−19.
(52) Kovarich, S.; Papa, E.; Gramatica, P.; Kovarich, S.; Papa, E.; Gramatica, P. QSAR classification models for the prediction of endocrine disrupting activity of brominated flame retardants. *J. Hazard. Mater.* **2011**, *190*, 106−112.
(53) Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set in descriptor space: A review. *ATLA, Altern. Lab. Anim.* **2005**, *33*, 445−459.