RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Using nano-QSAR to determine the most responsible factor(s) in gold nanoparticles exocytosis

Arafeh Bigdeli[a], *Mohammad Reza Hormozi-Nezhad*[*a,b], *Hadi Parastar*[a]

[a]Department of Chemistry, Sharif University of Technology, Tehran, Iran

[b]Institute for Nanoscience and Nanotechnology (INST), Sharif University of Technology, Tehran, Iran

*Corresponding Author: Email: hormozi@sharif.edu , Tel.: +98 21 6616 5337

Abstract

There are, to date, few general answers to fundamental questions related to the interactions of nanoparticles (NPs) with living cells. Studies reported in the literature have delivered only limited principles about the nano-bio interface and thus the biological behavior of NPs is yet far from being completely understood. Combining computational tools to experimental approaches in this regard helps to precisely probe the nano-bio interface and allows the development of predictive and descriptive relationships between the structure and the activity of nanomaterials. In the present contribution, a nano-quantitative structure-activity relationship (nano-QSAR) model has been statistically established using Partial Least Squares Regression (PLSR) model. Also, variable importance on PLS projections (VIP) has been used to find the most responsible factors on NPs exocytosis. Physicochemical properties of a set of different sized gold NPs with different surface coatings were greatly correlated to their exocytosis in macrophages. The results suggest that among the pool of physicochemical properties defined as nano-descriptors, charge density and surface charge seem to be the paramount factors leading to higher exocytosis values. Furthermore, charge accumulation and circularity of NPs are in the next level of priority among other nano-descriptors. The regression based nano-QSAR model reported here is satisfactory in both statistical quality and interpretability. The results could serve as a quantitative framework for better understanding the mechanisms that govern the interactions at the nano-bio interface.

Keywords: Gold Nanoparticles, Exocytosis, Surface Charge, Nano-QSAR, Partial Least Squares

1. Introduction

The interaction of NPs with living cells is dictated by many factors, including size, shape, chemical composition, crystallinity, hydrophobicity, porosity, surface charge, aggregation state, surface coating, plus characteristics of the suspending media.[1-8]In addition to the type of surface chemical functions, their relative arrangement also plays a key role in nanoparticle-cell interactions.[9]Amongstthese numerous factors affecting NPs behavior at the "nano-bio" interface, the decision about their priority is far from being completely understood. Which factor is most responsible? How could one decide which factor to tune among the pool of parameters in order to adjust a specific cellular response? With the increasing number of medicinal and therapeutic applications of NPs, these questions gain even more attention. Answers to these questions could provide remarkable clues for preparing safe and efficacious NPs as diagnostic and drug delivery tools. Moreover, there is still a lot to find out about what exactly happens at the nano-bio interface. Knowing the exact contribution of each parameter in NP's behavior could help to deeply discover the interactions and to design NPs with enhanced efficiency. Meanwhile, it would be noteworthy if it was possible to quantitatively adjust and distinguish between the most important variables involved.

Various case studies have taken into account the role of different factors on NPs cellular uptake, cytotoxicity and exo/endocytosis. Chan et al.[10] investigated the impact of NP size on active and passive tumor targeting efficiency. Rotello et al.[11] quantified the exocytosis behavior of NPs with different surface functionalities. Crespy et al.[12] showed how shape can influence the uptake of anisotropic polymeric NPs. Kanaras et al.[13] reported that the penetration of gold NPs through skin is influenced by the charge, morphology and function of the NPs. In another study, Chan et

al.[14] investigated the role of size and surface chemistry in mediating serum protein adsorption to gold NPs and their subsequent uptake by macrophages.

All these studies together with other similar ones reported in the literature[15-19] have delivered only limited principles about how various parameters affect NPs behavior at the nano-bio interface. In other words, they have argued that the biological response to a NP is generally a complex function of multiple parameters and is still not fully understood. Thus, there is a need for a comprehensive tool that could gather this individual information and provide a broader framework for understanding the interactions beyond the nano-bio interface. This knowledge is also important from the perspective of developing quantitative relationships that are able to predict the biological response profiles of NPs from their physicochemical properties.

Recently, Park et al.[20]have investigated the effect of native surface chemistries of gold nanoparticles (GNPs) and their subsequent opsonozation by serum proteins on their exocytosis pattern in macrophages. They have reported the exocytosis rates of a set of different sized and different charged GNPs. It was demonstrated that between size and surface charge, the latter seems to play a more crucial role in determining the exocytosis pattern which is confirmed by other similar related works. Again, this was another case study that considered only two factors simultaneously, size and surface charge. As an alternative, what if it was possible to take a step further and make a more comprehensive look? Is surface charge still the most dominant factor if a wider pool of variables considered? To answer this question, a quantitative structure-activity relationship framework may be helpful.

In this study, we have proposed a nano-quantitative structure-activity relationship (nano-QSAR) to investigate the effects of various parameters on GNPs exocytosis in macrophages. The statistical model quantitatively detects the most premier factors affecting the exocytosis of GNPs

among a pool of physicochemical properties of NPs (called nano-descriptors). A set of morphological nano-descriptors have therefore been extracted from their corresponding TEM images by applying image processing methods (based on our previous work).[21] In addition, a set of experimental parameters, together with a combinatorial set of nano-descriptors (a combination of experimental and image extracted descriptors) have also been provided. Partial Least Squares Regression (PLSR) has been carried out to analyze the data both for predictive and descriptive purposes. Actually, from a predictive point of view, the regression model finds the best possible correlation between physicochemical properties of NPs and their cellular response profiles (here exocytosis).Consequently, the established regression model can be applied to predict the exocytosis for unknown GNPs. On the other hand, with a descriptive perspective, variable importance on PLS projections (VIP) have been used to indicate the most dominant factors that are responsible in determining the exocytosis of GNPs.

2. Methods

2.1. Nano-QSAR approach

There are a lot of parameters to control when investigating the biological behavior of NPs interacting with living cells. Many of these parameters are strongly inter-correlated. Besides, it may be difficult to vary individual properties while keeping others constant. To overcome these constraints, attempts have been made to develop predictive models which relate physicochemical properties of NPs to their biological response(s).[22-30]However, this is a great challenge and requires the use of both powerful computational tools and experimental methods. Applying QSAR models to comprehensive datasets (collected from reliable experimental parameters and

several physicochemical properties of NPs) could manifest the real parameters underlying a considered interaction and therefore provides substantial details about the nano-bio interface. The more the data provided the more the authenticity and accuracy of the constructed model.

2.2. Descriptor generation

To gather the required data set for the nano-QSAR approach, three subsets of variables were provided: a) TEM extracted nano-descriptors including size, surface area, aspect ratio, corner count, curvature, aggregation state, and shape; b) experimental parameters including zeta potential, hydrodynamic diameter, and maximum wavelength both before and after protein coating; and c) combinatorial nano-descriptors including charge density, adjusted aspect ratio, charge accumulation, spectral size, spectral surface area, spectral aspect ratio and spectral aggregation. The first subset was prepared by performing image processing on TEM images shown in **Figure 1**. The details on extracting these twelve image nano-descriptors can be found elsewhere.[21]The second subset was inserted from the characterization information of surface-functionalized GNPs before and after serum coating.[20] In order to extract more information on the morphology and surface of the NPs, an idea was to combine image nano-descriptors with experimental ones. Consequently, a set of ten new meaningful combinatorial nano-descriptors were provided as a combination of some previously defined image extracted and experimental descriptors. The aim of defining the third subset was to provide maximum informative data for the corresponding GNPs. The complete set of nano-descriptors is shown in **Table1**.The last column comprises the numerical values of the exocytosis of different sized and coated GNPs extracted from the referencearticle.[20] In addition, **Table 2** demonstrates the calculation of the new defined combinatorial descriptors: charge density accounts for the amount of charge per unit

surface area which can be calculated based on the zeta potential before or after protein coating (ChDensB and ChDensA); adjusted aspect ratio was defined to distinguish between NPs with similar aspect ratios but different lengths which can be calculated from multiplying aspect ratio with either size extracted from TEM or hydrodynamic size before protein coating (AdjAR1, AdjAR2); charge accumulation takes into account the amount of available charge surrounding aggregated NPs. It is obvious that the amount of free charge is different between the situations in which the particles are either dispersed in the media or have formed an aggregate. Again, this combinatorial variable can be calculated from multiplying the aggregation state value by the zeta potential of the NPs before or after protein binding (ChAccumB, ChAccumA).The last combined nano-descriptors have been defined based on the strong dependency of NPs' plasmon shift to their size, surface area, aspect ratio and amount of aggregation (SpecSize, SpecSA, SpecAR, SpecAgg).

2.3. Data Analysis

PLSR model was utilized to analyze the data. PLS as a powerful chemometric tool in data analysis has been widely applied to numerous datasets in order to predict a set of dependent variables (responses) from a set of independent variables (predictors or descriptors). PLS finds the best correlation between these two datasets by extracting a small number of latent variables (LVs). The LVs form a new set of basis vectors that span a new space of the original variables and are aligned in directions in which both the captured variance and the correlation between **x** and **y** are maximized. A detailed explanation of PLS can be found elsewhere.[31]The PLS model is also able to reveal the most important variables that participated in the construction of LVs and actually, measures the variable significance. This is done by assessing PLS model parameters

such as weights, regression coefficients, selectivity ratios (SR) and the scores of variables important for projection (VIP).[32] In the present contribution, a PLS model with four LVs was built according to the minimum value of root mean square error of cross-validation (RMSECV) to find the best correlation between the data presented in Table 1 and the exocytosis of GNPs in macrophages reported by Park *et al*. PLSR was performed using the PLS Toolbox *v.* 5.8 (Eigenvector Research, Inc., Wenatchee, WA) and the resulting statistical model was constructed using the calibration set (X-block (12×28) and Y-vector (12×1)). The model was then validated by different cross-validation methods. The aim was to mathematically correlate the X-block to the Y-block using PLSR. The descriptor pool was preprocessed previous to modeling by autoscaling, as a common preprocessing method. This approach is necessary when the data consists of variables with different scaling and is applied to equalize the scale of different descriptors. Therefore the model can be then constructed based on the relative changes in the variables rather than being concerned with their absolute values. Autoscaled data have a mean expression of zero and a standard deviation of one. This is achieved by subtracting the column mean from each descriptor and dividing with the standard deviation of each column (descriptor). For values of each nano-descriptor, the corresponding auto scaled column can be achieved by following:

$$\text{Xautosc}=\frac{\text{X}-\bar{\text{X}}}{\text{SD(X)}} \quad (1)$$

where $\bar{\text{X}}$ and SD respectively stand for the mean value of each column in the data set and its corresponding standard deviation.

The prediction ability of the PLS regression model was evaluated using the leave-one-out cross validation (LOO-CV) method. Iteratively, eleven out of twelve NPs were used to generate a PLS

model (as the training set) and the left-out NP was tested as an unknown sample (as the test set). This process was repeated until each sample was left out once, and the results were compiled to determine the mean cross-validation regression coefficient (R^2_{CV}) and root mean square error ($RMSE_{CV}$) values. In order to assure preventing over fitting in the model, the adjusted R-squared (R^2_{adj}) value was also calculated. These statistical parameters were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$R^2_{adj} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / (n-p)}{\sum_{i=1}^{n}(y_i - \bar{y})^2 / (n-1)} \quad (4)$$

Where y_i and $\hat{y}_i$ stand for the observed and predicted exocytosis value for each *i*th sample, respectively. *n* and *p* are the total number of samples and the number of parameters in the model. $\bar{y}_i$ is the mean exocytosis value of the GNPs. It must be noticed that the proper number of LVs was chosen based on the maximum explained variance and minimum RMSEcv value corresponding to each LV (See **Figure S1**).

3. Results and discussion

Regression results are illustrated in **Table 3**. The statistical significance of the developed model is reflected from the acceptable values of R^2_{Cal} (0.971) and RMSEC (3.45) together with other

parameters reported in this table. Least possible deviations of the predicted exocytosis endpoints from the corresponding observed/measured ones is further implied from the satisfactory values of R^2_{CV} (0.707) and R^2_{adj} (0.780). **Figure 2** shows the percent of GNPs exiting the macrophages measured by ICP/MS *vs* their corresponding predicted values estimated by the PLS regression model. As can be seen, an acceptable correlation[33-35] is obtained and the resulting graph shows that the points are close to the line of fit. This again implicated the predictive ability of the developed PLSR model. The PLSR model uses latent variables (linear combinations of initial descriptors) to predict the exocytosis values. The exact amount of the contribution of the descriptors in each predictor (latent variable) can be extracted from the Loadings plot (**Figure S2b**). The more loading value of the descriptor means the more descriptor contributes in predicting the response (exocytosis). In the present study, 4 latent variables were chosen to build the PLS model. The amounts of loadings for each descriptor on these latent variables can be seen from their corresponding loading plot. Furthermore, the order of importance of the descriptors towards the exocytosis of the GNPs has been demonstrated using the Variable Importance Projection. VIP scores calculated for individual variables are demonstrated in **Figure 3**. Descriptors with VIP scores over the cutoff contribute higher in the regression model. Actually, the descriptors are ranked according to the descending order of VIP scores. Similarly, the calculated values of SR and Regression Vectors for all the variables can be seen in **Figure S3-b and c**. Another informative plot in the PLSR model is called the "Biplot" that graphically demonstrates the association between samples (GNPs) and model variables (nano-descriptors) (**Figure 4**). Showing the scores and loadings in one plot helps to interpret significant variables while looking at the samples location.

The interpretation and importance of the descriptors appearing in the regression are discussed below. The results revealed that among the nano-descriptors inserted to the model, the charge density and zeta potential, together with charge accumulation and circularity have the highest influence on gold NPs exocytosis. This conclusion confirms previous studies[20,36,37] and suggests that the amount of charge density is a better predictor for the exocytosis of a particular GNP than the amount of its surface charge. Moreover, the positive sign of the regression vector for the most important surface charge related descriptors (ZP_B and ChAccum_B) in Figure S3b, indicates that surface charge has a direct effect on exocytosis of the corresponding NPs, i.e. positive zeta potential leads to higher number of GNPs leaving the macrophages (the exocytosis value), which again is consistent with the results published by Park et al.[20] In addition, the results in Figure 2 clearly represent the importance priority of the selected descriptors. It must be noticed that, as it can be seen, all the important descriptors derived from the PLS model attribute to parameters before protein coating. This finding suggests that the properties of the NPs previous to entering the biological media intensely control their behavior, compared to the parameters after protein coating. This assumption can be supplemented by comparing the model based variable importance derived from VIP scores, SR and regression vectors. As expected, important variables from all these viewpoints are in common and belong to parameters "before protein coating". Therefore, as a complement to previous findings[38-40] which discuss that the protein corona (formed after entering NPs into biological media), determines NPs' fate and transport, one might conclude from the results displayed here that the initial conditions of the NP previous to entering the cell can actually have the same level of importance. In other words, the properties of NPs prior to moving into the cell can influence the nano-bio interface by dictating the protein corona formed around a special NP.

Moreover, the low VIP scores for the size and surface area nano-descriptors quite below the cut off value indicates that they do not seem to be significant factors in the exocytosis of GNPs within this size range (about 10-70nm) in comparison to other descriptors. This result was further investigated and confirmed based on the low SR and regression vector values appeared for these variables (Figure S-3). It should be mentioned here that in contrast to the endocytosis which has been reported to be impressed mainly by the size and shape of GNPs,[36,41-43]the results herein reveal that GNPs exocytosis is rarely dependent to size.

On the other hand, the relatively high importance value for aggregation state and its derivatives (charge accumulation and spectral aggregation) manifest the effect of these features that had been poorly mentioned and discussed in previous reports.[37,44,45]It can be concluded that the amount of variables' contribution for a specific biological endpoint might vary when looking into a wider framework. From another viewpoint, a negative but large regression vector value of the circle index favors lower exocytosis of GNPs in macrophages. Compared to variables with positive regression vectors, this variable contributes to the exocytosis in a reverse manner. The more the circularity, the less the exocytosis of GNPs. It is noteworthy to mention that this shape-type descriptor that has appeared among the set of important contributing factors, along with the rest of the newly defined descriptors in this study(columns 3, 5-12 and 19-28 of Table 1) have been introduced and investigated for the first time in such a nano-QSAR approach. It can be noticed from Figures 3and S-3 that circularity and charge accumulation before protein binding approximately contribute equally (but with opposite directions) in the exocytosis of GNPs, signifying the important effect of the newly defined shape descriptor. Among the other shape descriptors, square-like feature also displays a pretty high VIP score. Appearing the circle and square descriptors above the VIP cut off value can be explained from the TEM images of the

considered GNPs, displaying rather spherical particles. As a result, these two shape descriptors are expected to disclose higher impacts compared to others.

Though extending the findings of this study to wider sets of NPs and diverse biological endpoints requires taking into account larger data sets, but the method developed in this study has revealed the potential benefits of using chemometric approaches, such as nano-QSAR modeling to obtain both predictive and interpretative knowledge for a sample set of GNPs entering macrophages. This knowledge can be utilized to improve the experimental design of safe and effective GNPs for specific purposes and can be further applied to other sets of NPs. Considering all these points, the proposed model provides useful information for screening NPs libraries seeking different biological end points.

4. Conclusion

The biological behavior of NPs is a complicated function of multiple parameters and requires powerful tools to derive accurate correlations between the nanostructures and their biological responses. To gain an in depth understanding of the relationship between physicochemical properties of gold NPs and their exocytosis in macrophages, PLS regression model as a powerful data analysis tool was proposed. The nano-QSAR PLS model was derived from a pool of nano-descriptors consisting of image extracted features, experimental parameters and combinatorial descriptors. In addition to the predictive ability of the developed PLS regression model, major contributing features for the exocytosis of GNPs were also identified. Inspection of the results suggest that surface charge, charge density, circularity and charge accumulation seem to exhibit the highest impact on the exocytosis of GNPs among other variables. In addition, the results

revealed that parameters attributing to the NPs "before" protein binding have more influence on their exocytosis, compared to the ones "after" protein binding. Thus, controlling the initial conditions of the NPs previous to entering into the cell media is of great importance. The constructed model that quantitatively correlates the exocytosis of gold NPs to their basic physicochemical properties could allow researchers to predict biological response profiles (cellular uptake, cytotoxicity…) of NPs and design potential NPs for particular means.

References

1 A. Verma,F. Stellacci, *Small,*2010, **6**, 12.

2 A. E. Nel, L. Madler, D. Velegol, T. Xia, E. M. V. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, M. Thompson, *Nat. Mater.,*2009, **8**, 543.

3 X. Duan, Y. Li, *Small,*2013, **9**, 1521.

4 M. Zhu, G. Nie, H. Meng, T. Xia, A. Nel, Y. Zhao, *Acc. Chem. Res.,*2013, **46**, 622.

5 S. T. Kim, K. Saha, C. Kim, V. M. Rotello, *Acc. Chem. Res.,*2013, **46**, 681.

6 L. A. Dykman, N. G. Khlebtsov, *Chem. Rev.,*2014, **114**, 1258.

7 D. Bartczak, S. Nitti, T. M. Millar, A. G. Kanaras, *Nanoscale,*2012, **4**, 4470.

8 A. Albanese, P. S. Tang, W. C. W. Chan, *Annu. Rev. Biomed. Eng.,*2012, **14**, 1.

9 A. Verma, O. Uzun, Y. Hu, Y. Hu, H. S. Han, N. Watson, S. Chen, D. J. Irvine, F. Stellacci, *Nat. Mat.,*2008, **7**, 588.

10 E. A. Sykes, J. Chen, G. Zheng, W. C. W. Chan, *ACS Nano,*2014, **8**, 5696.

11 C. S. Kim, N. D. B. Le, Y. Xing, B. Yan, G. Y. Tonga, C. Kim, R. W. Vachet, V. M. Rotello, *Adv. Healthcare. Mater.,*2014, **3**, 1200.

12 L. Florez, C.Herrmann, J. M.Cramer, C. P. Hauser, K. Koynov, K. Landfester, D.Crespy, V. Mailänder, *Small,*2012, **8**, 2222.

13 R. Fernandes, N. R.Smyth, O. L.Muskens, S.Nitti, A. Heuer-Jungemann, M. R. Ardern-Jones, A. G.Kanaras, *Small,2014,* **11**, 713.

14 C. D. Walkey, J. B. Olsen, H. Guo, A. Emili, W. C. W. Chan, *J. Am. Chem. Soc.,*2012, **134**, 2139.

15 E. Oh, J. B. Delehanty, K. E. Sapsford, K. Susumu, R. Goswami, J. B. Blanco-Canosa, P. E. Dawson, J. Granek, M. Shoff, Q. Zhang, P. L. Goering, A. Huston, I. L. Medintz, *ACS Nano,*2011, **5**, 6434.

16 A. K. Suresh, D. A. Pelletier, M. J. Doktycz, *Nanoscale,*2013, **5**, 463.

17 E. C. Cho, L. Au, Q. Zhang, Y. Xia, *Small,*2010, **6**, 517.

18 K. Huang, H. Ma, J. Liu, S. Huo, A. Kumar, T. Wei, X. Zhang, S. Jin, Y. Gan, P. C. Wang, S. He, X. Zhang, X. J. Liang, *ACS Nano,*2012, **6**, 4483.

19 P. Rivera-Gil, D. Jimenez de Aberasturi, V. Wulf, B. Pelaz, P. del Pino, Y. Zhao, J. M. de la Fuente, I. Ruiz de Larramendi, T. Rojo, X. J. Liang, W. J. Parak, *Acc. Chem. Res.,*2013, **46**, 743.

20 N. Oh, J. H. Park, *ACS Nano,*2014, **8**, 6232.

21 A. Bigdeli, M. R. Hormozi-Nezhad, M. Jalali-Heravi, M. R. Abedini, F. Sharif-Bakhtiar, *RSC Adv.,*2014, **4**, 60135.

22 D. Fourches, D. Pu, C. Tassa, R. Weissleder, S. Y. Shaw, R. J. Mumper, A. Tropsha, *ACS Nano,*2010, **4**, 5702.

23 V. C. Epa, F. R. Burden, C. Tassa, R. Weissleder, S. Shaw, D. A. Winkler, *Nano Lett.,*2012, **12**, 5808.

24 T. Puzyn, D. Leszczynska, J. Leszczynski, *Small,*2009, **5**, 2494.

25 B. Rasulev, A. Gajewicz, T. Puzyn, D. Leszczynska and J. Leszczynski, Towards Efficient Designing of Safe Nanomaterials: Innovative Merge of Computational Approaches and Experimental Techniques,RSC, 2012, Ch. 10, pp 220-256.

26 T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, *Chem. Rev.*, 2012, **112**, 2889.

27 X. R. Xia, N. A. Monteiro-Riviere, J. E. Riviere, *Nat. Nanotech.,*2010, **5**, 671.

28 X. R. Xia, N. A. Monteiro-Riviere, S. Mathur, X. Song, L. Xiao, S. J. Oldenberg, B. Fadeel, J. E. Riviere, *ACS Nano,*2011, **5**, 9074.

29 X. Hu, S. Cook, P. Wang, H. M. Hwang, *Sci. Total Environ.,*2009, **407**, 3070.

30 K. P. Singh, S. Gupta, *RSC Adv.,* 2014, **4**, 13215.

31 S. Wold, M. Sjostrom, L. Eriksson, *Chemometr. Intel. Lab.,*2001, **58**, 109.

32 C. M. Andersen, R. Bro, *J. Chemometr.,*2010, **24**, 728.

33 D. M. Hawkins, C. B. Basak, D. Mills, *J. Chem. Inf. Comput. Sci.,*2003, **43**, 579.

34 P. P. Roy, K. Roy, *QSAR Comb. Sci.*2008, **27**, 302.

35 J. H. Kalivas and P. Gemperline, Practical Guide to Chemometrics, Taylor and Francis, 2006, Ch. 5, pp 125-131.

36 N. Oh, J. H. Park, *Int. J. Nanomed.,*2014, **9**, 51.

37 R. Sakhtianchi, R. F. Minchin, K. B. Lee, A. M.Alkilany, V. Serpooshan, M. Mahmoudi, *Adv. Colloid Interface Sci.,*2013, **201-202**, 18.

38 M. Rahman, S. Laurent, N. Tawil, L. Yahia, M. Mahmoudi, Protein-Nanoparticle Interactions, Springer-Verlag, 2013, Ch. 2, pp 21-44.

39 I. Lynch, A. Salvati, K. A. Dawson, *Nature Immunol.,*2009, **4**, 546.

40 M. Mahmoudi, S. N. Saeedi-Eslami,M. A. Shokrgozar,K. Azadmanesh, M. Hassanlou, H. R. Kalhor, C. Burtea, B. Rothen-Rutishauser, S. Laurent, S. Sheibani,H. Vali, *Nanoscale,*2012, **4**, 5461.

41 B. D. Chithrani, W. C. W. Chan, *Nano Lett.,*2007, **7**, 1542.

42 W. Jiang, B. Y. S. Kim, J. T. Rutka, W. C. W. Chan, *Nat. Nanotechnol.,*2008, **3**, 145.

43 W. G. Kreyling, S. Hirn, W. Moller, C. Schleh, A. Wenk, G. Celik, J. Lipka, M. Schaffler, N. Haberl, B. D. Johnston, R. Sperling, G. Schmid, U. Simon, W. J. Parak, M. Semmler-Behnke, *ACS Nano*, 2014, **8**, 222.

44 M. Tarantola, A, Pietuch, D.Schneider, J. Rother, E. Sunnick, C.Rosman, S.Pierrat, C. Sönnichsen, J.Wegener, A. Janshoff, *Nanotoxicology*. 2011, **5**, 254.

45 J. A. Yang, S. E. Lohse, C. J. Murphy, *Small,* 2014, **10**, 1642.

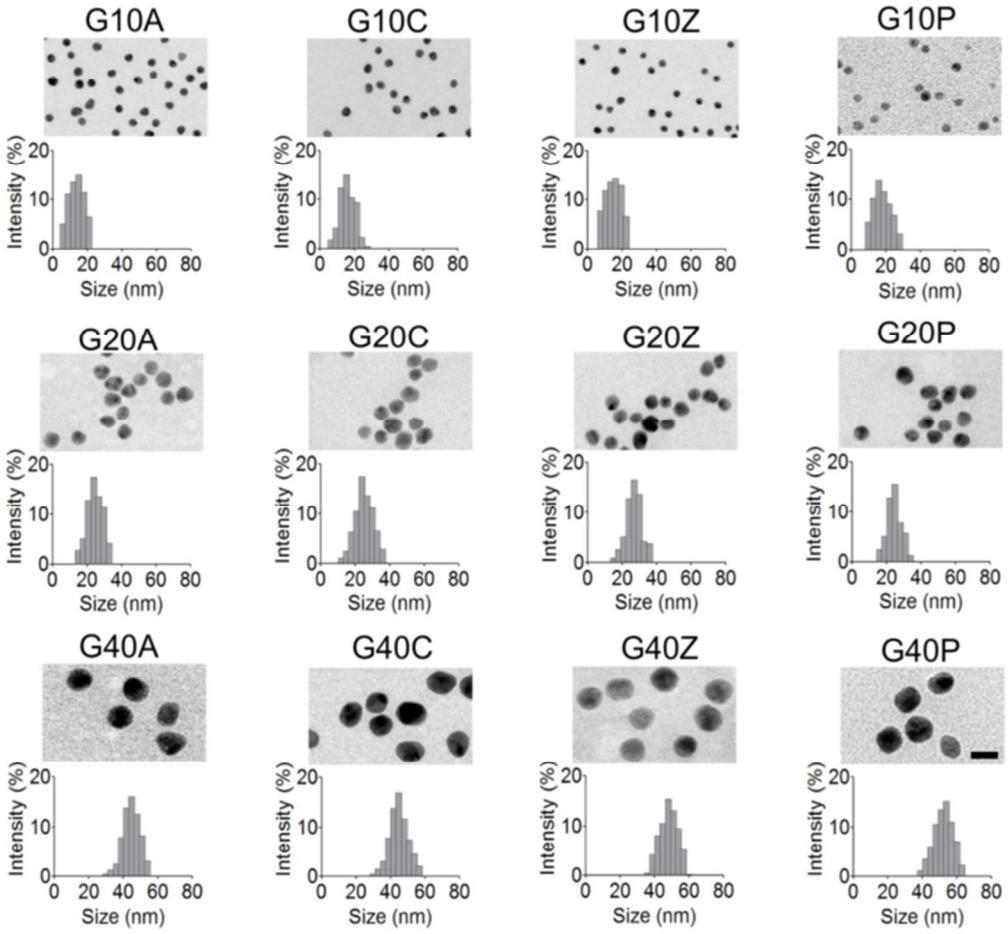
Figure Captions

Figure 1. TEM visualization (top) and size distribution histograms (bottom) of all GNPs in the dataset with different sizes and different coatings.[20] The number and letter after the GNP identifier designate the size and type of coating (anionic, cationic, zwitterionic and PEGylated surfaces), respectively. Scale bar is 50nm. *Reprinted with permission from ref. [20]. Copyright 2014 ACS.*

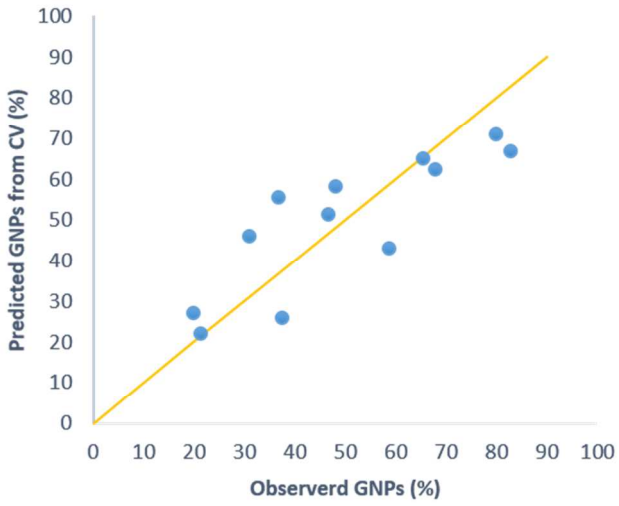
Figure 2. The predicted versus observed exocytosis values of gold nanoparticles displayed as the percent of GNPs leaving the macrophages.

Figure 3. Variables Important for Projection (VIP) scores calculated for all the nano-descriptors in the PLS model. The descriptors with VIP scores higher than the cutoff (VIP=1) are important.

Figure 4. Biplot (combined Score and Loading plots) showing the samples (red rectangles) and the nano-descriptors (blue squares) together in one plot.



--Figure 1.TEM visualization (top) and size distribution histograms (bottom) of all GNPs in the dataset with different sizes and different coatings.[20] The number and letter after the GNP identifier designate the size and type of coating (anionic, cationic, zwitterionic and PEGylated surfaces), respectively. Scale bar is 50nm. *Reprinted with permission from ref. [20]. Copyright 2014 ACS.*



405

406 **Figure 2.** The predicted versus observed exocytosis values of gold nanoparticles displayed as the
407 percent of GNPs leaving the macrophages.

408

409

410

411

412

413

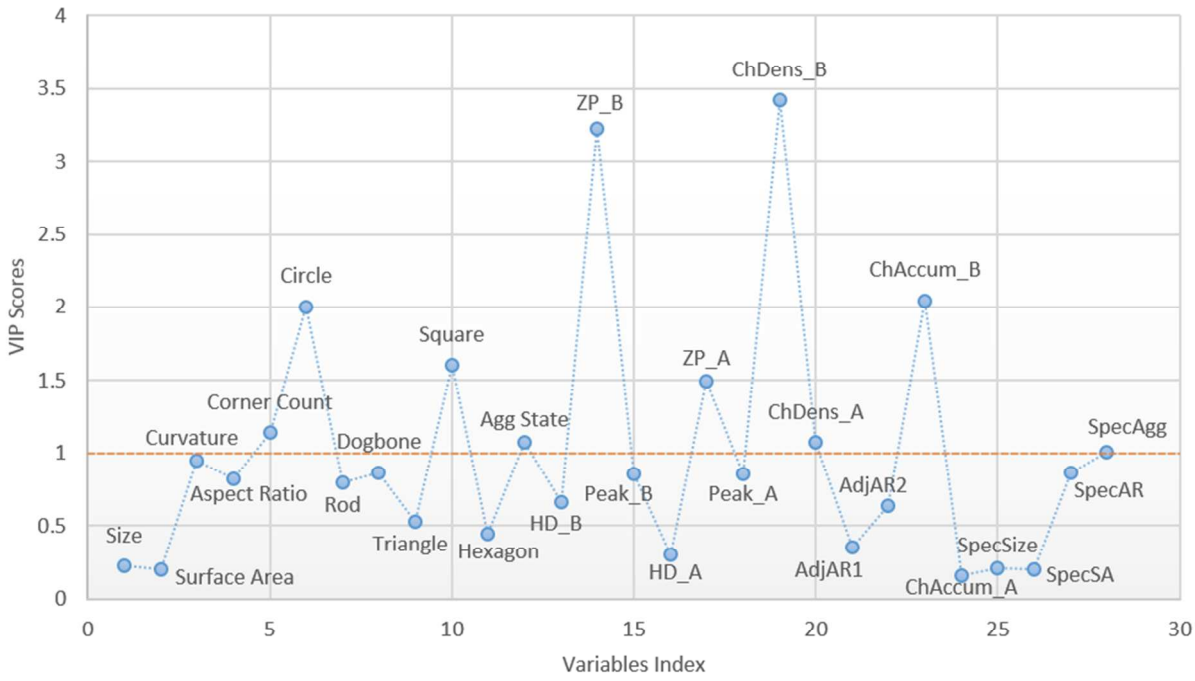


Figure 3. Variables Important for Projection (VIP) scores calculated for all the nano-descriptors in the PLS model. The descriptors with VIP scores higher than the cutoff (VIP=1) are important.

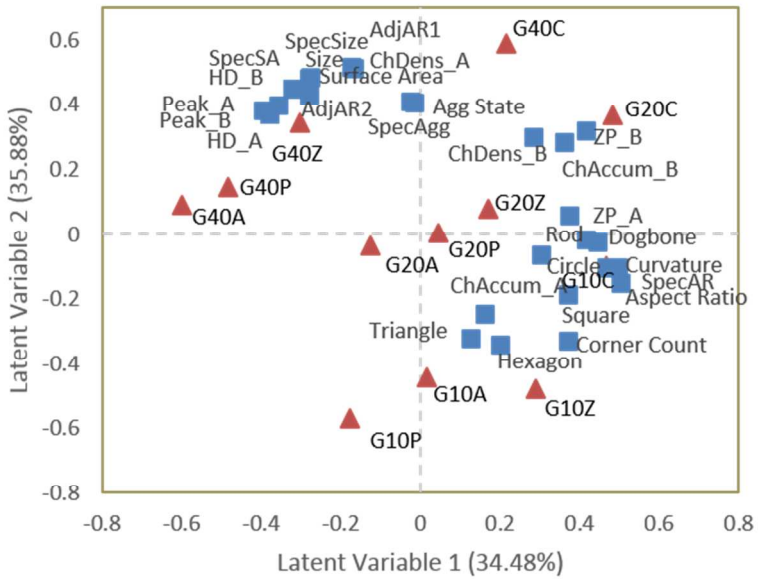


Figure 4. Biplot (combined Score and Loading plots) showing the samples (red rectangles) and the nano-descriptors (blue squares) together in one plot.

Table 1. Whole set of nanodescriptors and the last column consists of the observed exocytosis values.

NP ID	Nano-descriptors																												Observed Exocytosis	Predicted Exocytosis
	Image Extracted[a]												Experimental[b]						Combinatorial[c]											
	Size	Surface Area	Curvature	Aspect Ratio	Corner Count	Circle	Rod	Dogbone	Triangle	Square	Hexagon	Agg State	HD_B	ZP_B	Peak_B	HD_A	ZP_A	Peak_A	ChargeDensity_B	ChargeDensity_A	AdjAR1	AdjAR2	ChargeAccum_B	ChargeAccum_A	SpecSize	SpecSA	SpecAR	SpecAgg		
G10A	18.800	267.200	0.793	1.056	7.133	0.655	0.168	0.151	0.754	0.578	0.777	0.121	14.410	-29.280	520.000	36.030	-28.390	525.000	-0.110	-0.106	19.864	15.219	-3.552	-3.444	9780.160	138944.0	549.182	63.087	46.7	51.6
G10C	16.700	227.600	1.280	1.061	7.176	0.712	0.138	0.143	0.533	0.663	0.798	0.065	13.480	26.330	520.000	30.640	-20.090	525.000	0.116	-0.088	17.698	14.300	1.721	-1.313	8674.952	118345.7	551.642	33.992	80	71.1
G10Z	14.000	165.800	1.519	1.073	7.455	0.794	0.132	0.139	0.545	0.670	0.832	0.041	16.260	-0.920	520.000	36.120	-21.270	525.000	-0.006	-0.128	14.981	17.440	-0.037	-0.863	7263.100	86225.36	557.736	21.110	36.7	55.5
G10P	19.600	274.100	0.687	1.044	7.571	0.669	0.118	0.120	0.450	0.579	0.714	0.052	25.420	-27.540	525.000	37.220	-29.610	530.000	-0.100	-0.108	20.497	26.546	-1.436	-1.544	10304.38	143925.0	548.263	27.368	20	27.2
G20A	31.800	816.800	0.831	1.008	6.833	0.653	0.102	0.110	0.545	0.649	0.798	0.230	26.860	-38.400	525.000	48.220	-40.010	530.000	-0.047	-0.049	32.058	27.071	-8.849	-9.220	16698.99	428793.7	529.132	120.983	58.6	43.2
G20C	36.800	999.700	0.981	1.113	7.444	0.752	0.173	0.166	0.442	0.566	0.671	0.238	29.020	47.120	525.000	40.790	-23.580	530.000	0.047	-0.024	40.923	32.296	11.207	-5.608	19305.35	524825.1	584.262	124.861	82.7	66.9
G20Z	28.800	679.600	0.924	1.007	6.909	0.700	0.110	0.119	0.529	0.629	0.792	0.161	27.230	0.120	525.000	47.510	-15.560	530.000	0.000	-0.023	29.024	27.420	0.019	-2.512	15132.07	356808.9	528.665	84.762	65.5	65.2
G20P	37.300	1027.100	0.840	1.072	7.333	0.751	0.193	0.182	0.464	0.572	0.731	0.221	35.930	-13.550	530.000	44.090	-34.740	535.000	-0.013	-0.034	39.953	38.521	-2.993	-7.673	19751.24	544368.3	568.213	117.068	31	46.2
G40A	46.100	1925.200	0.560	0.843	6.200	0.655	0.049	0.050	0.437	0.522	0.674	0.137	45.490	-35.500	535.000	66.280	-36.450	540.000	-0.018	-0.019	38.845	38.362	-4.861	-4.991	24643.86	1029982.	451.168	73.255	37.5	26.1
G40C	51.100	1956.600	1.105	1.035	6.625	0.679	0.185	0.180	0.407	0.582	0.638	0.211	47.790	40.030	535.000	62.310	-22.070	540.000	0.020	-0.011	52.901	49.483	8.448	-4.657	27334.11	1046791.	553.950	112.902	67.8	62.5
G40Z	50.800	2131.800	0.549	0.923	6.375	0.641	0.051	0.052	0.482	0.582	0.741	0.141	48.840	0.130	535.000	65.380	-25.890	540.000	0.000	-0.012	46.856	45.069	0.018	-3.647	27165.10	1140486.	493.695	75.369	48.2	58.4

G40P	46.700	1917.400	0.613	0.882	6.600	0.724	0.086	0.089	0.402	0.485	0.647	0.180	55.150	-15.180	540.000	66.020	-30.980	545.000	-0.008	-0.016	41.170	48.620	-2.735	-5.582	25217.83	1035396.	476.057	97.300	21.4	22.2

a TEM extracted image nano-descriptors calculated by image analysis on TEM images in Figure 1. Please see [21] for more information on the image analysis process.

b Gold NPs characteristics extracted from Figure 1c in [20]

c Combinatorial descriptors calculated based on the formula presented in the last column of Table 2

Table 2. New combinatorial nano-descriptors

ID	Definition	Abbreviation	Calculation
1	Charge Density B	ChDenB	ZP_B/SurfaceArea
2	Charge Density A	ChDenA	ZP_A/SurfaceArea
3	Adjusted Aspect Ratio 1	AdjAR1	AspectRatio×Size
4	Adjusted Aspect Ratio 2	AdjAR2	AspectRatio×HD_B
5	Charge Accumulation B	ChAccumB	AggState×ZP_B
6	Charge Accumulation A	ChAccumA	AggState×ZP_A
7	Spectra Size	SpecSize	PeakB×Size
8	Spectra Surface Area	SpecSA	PeakB×SurfaceArea
9	Spectra Aspect Ratio	SpecAR	PeakB×AspectRatio
10	Spectra Aggregation State	SpecAgg	PeakB×AggState

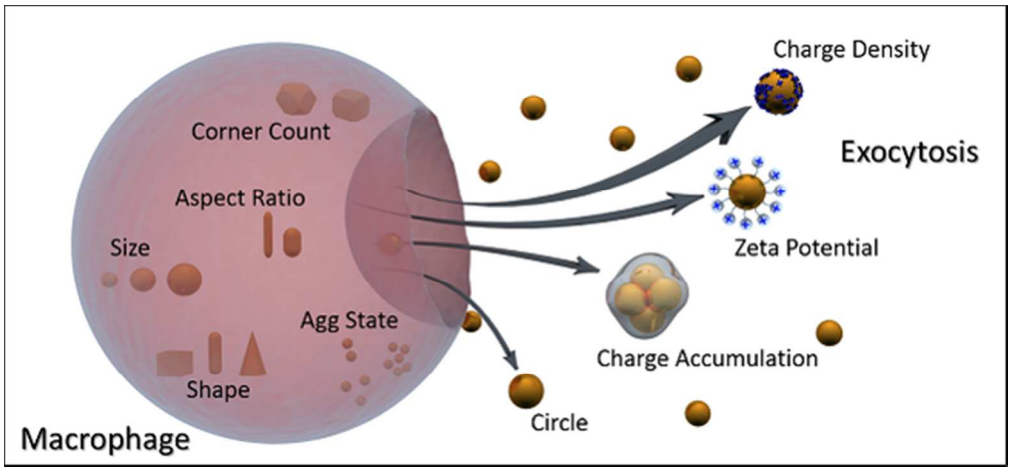
Table 3. Statistical results of the PLS model

Statistical parameter	LVs	R^2_{Cal}	R^2_{CV}	RMSEC	RMSECV	R^2_{Adj}	Rel. Error*
PLS model	**4**	**0.971**	**0.707**	**3.456**	**11.129**	**0.78**	**20.7%**

*Relative Error in percent is calculated from: $\text{Rel.Error}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i)^2}}\times 100$

514 **Table of Contents**

515



516

517

518 A nano-quantitative structure-activity relationship (nano-QSAR) model is proposed to indicate
519 the determining factors responsible in the exocytosis of gold nanoparticles in macrophages.