

PRELIMINARY COMMUNICATION

Open Access



# Universal nanohydrophobicity predictions using virtual nanoparticle library

Wenyi Wang<sup>1</sup>, Xiliang Yan<sup>1,2</sup>, Linlin Zhao<sup>1</sup>, Daniel P. Russo<sup>1</sup>, Shenqing Wang<sup>2</sup>, Yin Liu<sup>3</sup>, Alexander Sedykh<sup>1,4</sup>, Xiaoli Zhao<sup>5</sup>, Bing Yan<sup>2,6</sup> and Hao Zhu<sup>1,7,8\*</sup>

## Abstract

To facilitate the development of new nanomaterials, especially nanomedicines, a novel computational approach was developed to precisely predict the hydrophobicity of gold nanoparticles (GNPs). The core of this study was to develop a large virtual gold nanoparticle (vGNP) library with computational nanostructure simulations. Based on the vGNP library, a nanohydrophobicity model was developed and then validated against externally synthesized and tested GNPs. This approach and resulted model is an efficient and effective universal tool to visualize and predict critical physicochemical properties of new nanomaterials before synthesis, guiding nanomaterial design.

**Keywords:** Nanohydrophobicity, Surface chemistry, Surface simulations, Nanomaterials design, Virtual nanoparticle library, Predictive model

Advances in nanotechnology and material sciences in the past decade have led to the rapid development of engineered nanomedicines in pharmaceutical sciences [1, 2]. The traditional development route of new nanomaterials solely depends on experimental testing, which is costly and time consuming. With rapidly rising experimental and labor costs, computational approaches have become promising low cost alternatives to study nanomaterials [3]. To date, computational modeling approaches are broadly applied to the research and development procedure of small molecules, but rarely for larger molecules like nanomaterials [4]. This is evidenced by the many available commercial software tools [5–7] capable of predicting physicochemical properties for new druggable small molecules but none are available for new nanomedicines. Compared to small molecules, the shape, size, composition and surface ligands of nanomaterials greatly increase nanostructure complexity. Due to this increased complexity, the biological activities and therapeutic effects of nanomaterials are more difficult to model than small molecules. As a key determinant of

drug pharmacokinetics, hydrophobicity influences drug solubility, absorption, distribution, and target binding characteristics, which are eventually associated with the drug efficacy, potency and toxicity [8, 9]. Therefore, it is critical to evaluate the hydrophobicity of nanomedicines in the early stages of development, even before chemical synthesis.

In previous studies, researchers have been devoted to building quantitative structure activity relationship (QSAR) models for various bioactivities of different nanomaterials but have had limited applicability for new nanomaterial development [10–13]. Namely, two major issues limited the applicability of the resulted models: (1) the lack of enough high quality nano-bioactivity data and (2) computational approaches to precisely quantify nanostructure diversity. Currently, the use of experimental values as descriptors [14, 15] prevents the predictions of new nanomaterials before chemical synthesis. On the other hand, computational calculation of descriptors allows for virtual nanoparticle generation and nano-bioactivity prediction with no chemical synthesis required. Some researchers found that descriptors calculated solely from the surface ligands of nanoparticles were useful in predicting properties. Although this is useful in predicting certain properties of nanoparticles, the effects of the

\*Correspondence: hao.zhu99@rutgers.edu

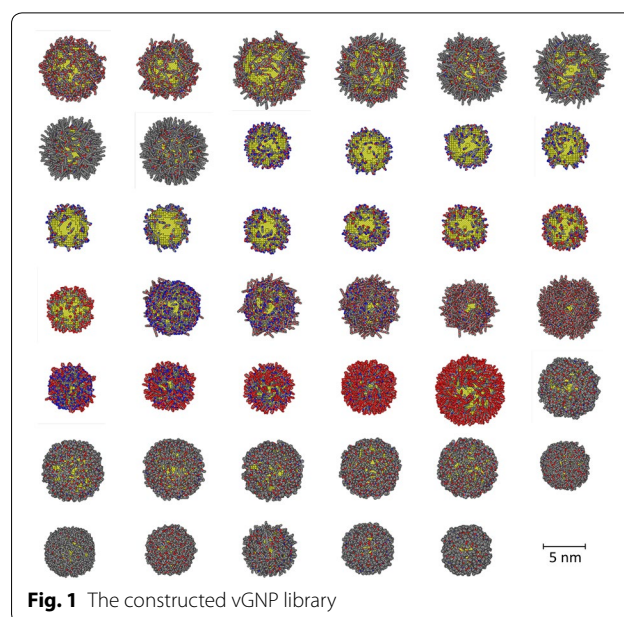
<sup>7</sup> Department of Chemistry, Rutgers University, 315 Penn St., Camden, NJ 08102, USA

Full list of author information is available at the end of the article



nanoparticle size and surface ligands density, position, distribution, were not considered in these studies and likely also contribute to the nano-bioactivity. More recently, however, some researchers have utilized some of these properties in addition to the general descriptor set from surface ligands, e.g., electronic properties, [16] ionic characteristics, [17] and others [18–20]. The major drawback of these available modeling studies is the lack of approaches to correctly quantify and represent nanostructure diversity during the modeling procedure. In our previous studies, we have shown that surface chemistry was the most critical factor in determining the bioactivities of gold nanoparticles (GNPs), including nanohydrophobicity [21]. Furthermore, correctly simulating surface chemistry can result in novel nanodescriptors which can be used to develop quantitative nanostructure–activity relationship (QNAR) models, showing superior advantages than traditional modeling studies [22]. Here, we report a novel approach to develop a virtual gold nanoparticle (vGNP) library with surface simulations precisely predicting nanohydrophobicity for new nanomaterials. Using this approach, a nanohydrophobicity model was developed based on surface chemistry simulation of a set of GNPs with various surface ligands. The model predictivity was further proved by experimentally synthesizing and testing nine new GNPs, and comparing their experimental/predicted logP values. The predicted nanohydrophobicity showed high correlations with experimental results, indicating the applicability of using this universal predictive modeling approach to design and select new GNPs with desired hydrophobicity.

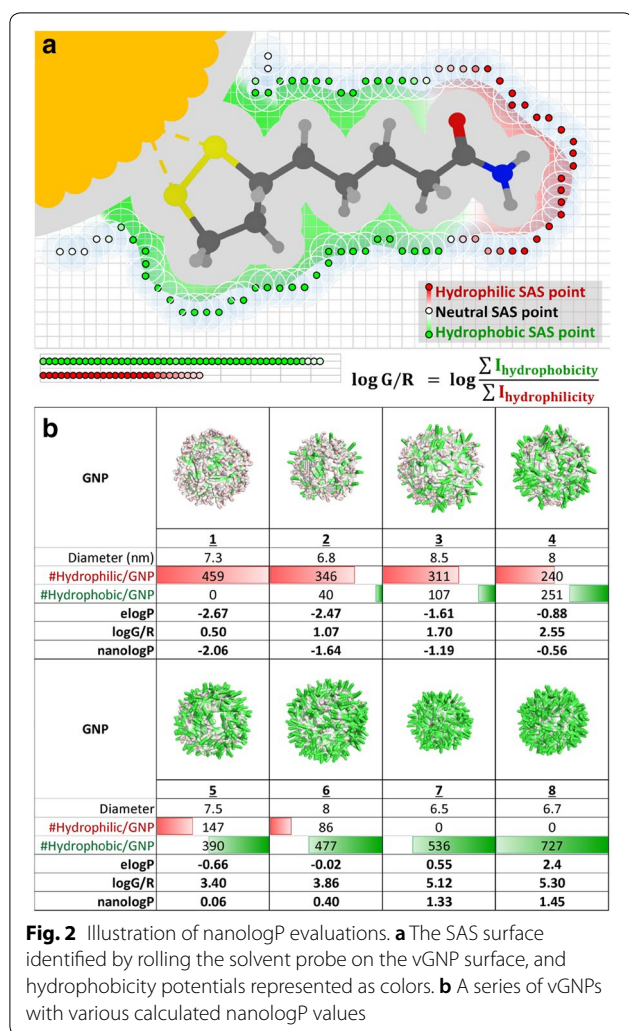
In a recent study, we developed a novel method to construct vGNP libraries [22]. Using this approach, we constructed the vGNP library with a dataset of 41 GNPs, as shown in Fig. 1. Specifically, using the structural information of surface ligands, ligand density of each GNP, and the GNP size, the virtual structure for each of the GNPs in the library was constructed as follows. First, the gold core was constructed based on the GNP size. Then, the surface ligands, with ligand density information, were randomly attached to the gold core to simulate the experimental conditions. These 41 GNPs were synthesized and tested for their hydrophobicity. The high nanostructure diversity of these 41 GNPs, including various surface ligands, different ligand densities per GNP and various GNP sizes, and high hydrophobicity diversity (experimental logP values range from  $-3$  to  $3$ ) make this dataset suitable for modeling purposes. This dataset was used as the modeling set to develop nanohydrophobicity models. The experimental approaches to synthesize this GNP library and test the logP values are described in our previous study [22]. All the experimental data used to



construct the vGNP library, including the structure information of surface ligands, are provided in Additional file 4: Table SI.

Besides providing a large nanohydrophobicity dataset in this study, a new surface chemistry simulation approach was developed based on the constructed vGNP library to evaluate hydrophobicity of GNPs. The core of this technique was to evaluate the solvent accessible surface (SAS) of GNPs and to calculate the nanohydrophobicity accordingly. The SAS, also named the Connolly Surface, [23] was identified for each GNP using a grid based method [24]. The cross section (grey area) of a vGNP surface ligand was constructed in a 2D grid as shown in Fig. 2a. The SAS was determined by rolling a solvent probe, simulated by the size of a water molecule of radius  $1.4 \text{ \AA}$ , over the surface of the vGNP. Probes were placed on grid points surrounding the vGNP surface ligand. A grid point was identified as a SAS point of this vGNP when the probe was within one grid unit distance to at least one vGNP atom, and does not overlap with any other vGNP atoms [24].

Once the SAS, with all identified grid points, was constructed for a vGNP, its hydrophobicity potential was evaluated by calculating the octanol–water partition coefficient from a distance-dependent weighting function of atomic contributions [25, 26]. The hydrophobic/hydrophilic potential of an identified SAS point was determined by nearby atoms and weighted by their distances to the SAS point. As shown in Fig. 2a, hydrophilic SAS points were colored with red while hydrophobic SAS points were colored with green. The hydrophilic/hydrophobic potential for each SAS was represented as the intensity of the corresponding



color—red as hydrophilic and green as hydrophobic. As an example, the hydrophobic potentials of eight vGNPs can be visualized in Fig. 2b. This series of GNPs were constructed with two types of surface ligands with different hydrophobicities: one ligand was hydrophilic and the other was hydrophobic. The ratio of these two types of surface ligands among the eight GNPs was gradually changed to modulate the nanohydrophobicity from low to high. From Fig. 2b, this series of GNPs showed a clear trend of hydrophobicity change with an increased ratio of hydrophilicity/hydrophobicity surface ligands. Thus, the surface colored vGNPs could be a representation of nanohydrophobicity of GNPs.

The nanohydrophobicity was then quantified using the colored vGNP. The nanohydrophobicity of a vGNP can be calculated as:

$$\log G/R = \log \frac{\sum I_{\text{hydrophobicity}}}{\sum I_{\text{hydrophilicity}}} \quad (1)$$

where G and R represent the hydrophobic potential (green) and hydrophilic potential (red) for each SAS point, and I is the intensity of hydrophobic/hydrophilic potential.

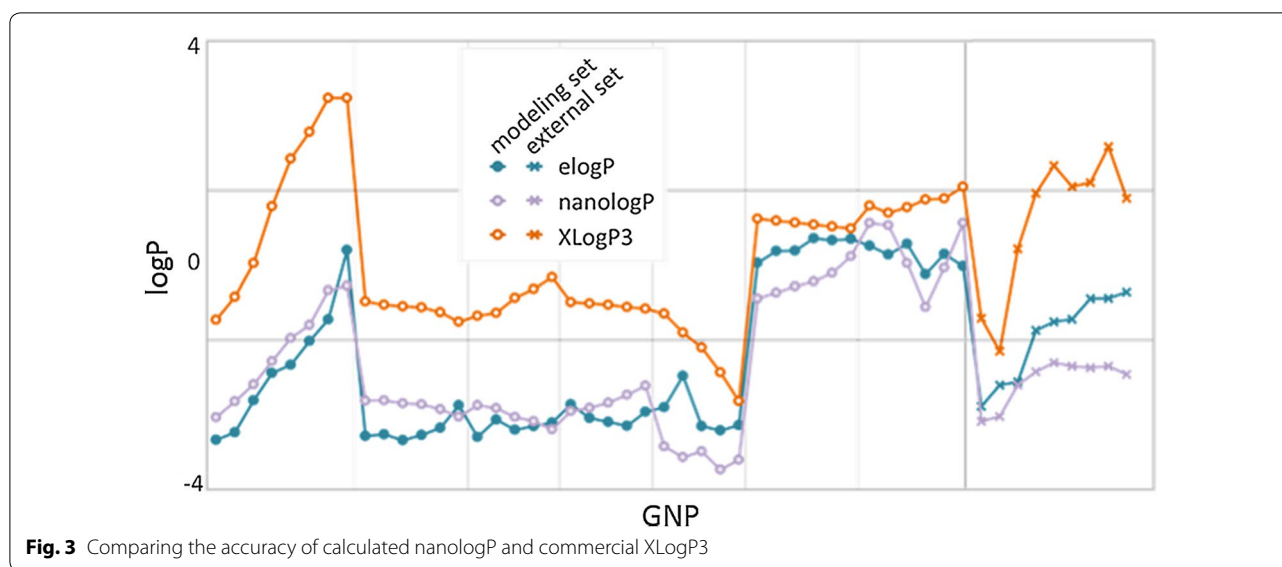
Then, with a linear regression analysis between logG/R and logP values of these 41 GNPs, the following equation was generated and can be used to calculate nanologP (i.e. logP values of GNPs) values for new GNPs from their logG/R results, which were obtained from vGNP simulations:

$$\text{nanologP} = 0.7334 * \log G/R - 2.4306 \quad (2)$$

The calculated logP values of all the 41 nanoparticles (nanologP), obtained from the above equation, were compared to their experimental logP results (eLogP), which were obtained by experimentally testing the partition coefficients in n-octanol and water solutions.

The step by step instruction of vGNP generations and log G/R calculations were described in the Additional file 3 (vGNP logP Supplementary demo file) and all source code files were also shared as Python files (see details in the Additional files 1, 2, and 3).

In some previous studies, logP of nanomaterials were calculated based only on surface ligand structures [15, 18, 19, 21, 27]. For comparison purposes, logP values of these 41 GNPs were calculated using four calculators, XlogP3, [28] AlogPS 2.1, [29] ClogP calculated in ChemDraw 17.0 [30] and the logP model in MOE 2016 [31]. These four logP calculators were built by either chemical atom/fragment contribution methods (XlogP3, ClogP and logP in MOE) or QSAR modeling (AlogPS). These calculators are commonly used to calculate the surface ligand logP and are based on various linear and non-linear modeling approaches. For example, XlogP3, AlogPS 2.1, ClogP and logP in MOE were based on a nearest neighbor approach combined with linear additive model [28], associated neural networks [29], fragmental additive approach [30] and atom additive approach [31], respectively. When modeling mixtures, the weighted average according to the component fractions was used for calculating the chemical descriptors [32]. Similarly, in this study, for a GNP with two different surface ligands, its logP value was calculated by averaging two ligand logP values weighted by the number of the two types of ligands. As shown in Fig. 3 and Additional file 4: Table SI, the best obtained logP results from commercial software, XlogP3, which yielded a low correlation with eLogP with a coefficient of determination ( $R^2$ ) = 0.577, and large prediction errors as Mean Absolute Error (MAE) = 2.633 and root mean square error (RMSE) = 3.00 [33, 34]. These results were much worse than that of nanologP developed in



this study ( $R^2=0.884$ ,  $MAE=0.719$  and  $RMSE=0.81$ ). A five-fold cross-validation was performed for nanologP and the results are similar ( $R^2=0.832$ ,  $MAE=0.75$  and  $RMSE=1.28$ ). The summary table of training and validation sets and the original GNP library file for calculation can be viewed in Additional files 4 and 5.

To further validate the performance of the proposed nanologP method, we synthesized nine new GNPs with different surface ligands compared to the modeling set and experimentally obtained their elogP values. The calculated nanologP values show high predictivity for this external set with  $R^2_{ext}=0.762$ ,  $MAE_{ext}=1.182$  and  $RMSE_{ext}=1.24$ , similar to the modeling set result. In comparison, the best calculated logP values from commercial software (XlogP3) show much worse prediction accuracy with  $R^2_{ext}=0.534$ ,  $MAE_{ext}=3.097$  and  $RMSE_{ext}=3.49$ .

In this study, an applicable nanohydrophobicity computational method was developed. The results showed that precisely simulated nanostructures using the vGNP library technique was the key to the accurate calculation of physicochemical properties of GNPs, such as hydrophobicity. There is potential to adapt the approach for other nanoparticles (e.g., carbon nanotubes and silver nanoparticles). The logG/R can be calculated by simulating the new type of nanoparticles with the designated core and shape, and the same hydrophobicity/hydrophilicity evaluation strategy. This is an ongoing work when more experimental data becomes available in the future. Furthermore, this approach can also be applied to the modeling and evaluation of other critical properties or bioactivities (e.g., interaction potentials with the environment, permeability through cell membranes, etc.).

## Additional files

**Additional file 1.** Python codes for the calculation of nanologP.

**Additional file 2.** Python codes for the calculation of nanologP.

**Additional file 3.** A demo for calculation of nanologP.

**Additional file 4.** Summary table of the GNP library in the training and validation sets.

**Additional file 5.** The original GNP library input file for the demo.

## Abbreviations

GNP: gold nanoparticle; vGNP: virtual gold nanoparticle; QSAR: quantitative structure activity relationship; QNAR: quantitative nanostructure-activity relationship; SAS: solvent accessible surface.

## Authors' contributions

HZ and WW conceived and designed the study and wrote the manuscript. WW, XY, LZ and DPR performed the three repeats of the nanologP calculation. XY calculated the logP from commercial software as control. SW and YL performed the experimental testing. AS assisted the data analysis. XZ and BY provided experimental support. All authors read and approved the final manuscript.

## Author details

<sup>1</sup>The Rutgers Center for Computational and Integrative Biology, Camden, NJ 08102, USA. <sup>2</sup>School of Chemistry and Chemical Engineering, Shandong University, Jinan 250100, China. <sup>3</sup>Research Center for Eco-Environmental Science, Chinese Academy of Sciences, Beijing 100085, China. <sup>4</sup>Sciome, Research Triangle Park, NC 27709, USA. <sup>5</sup>Department of Physiological Sciences, Eastern Virginia Medical School, Norfolk, VA 23507, USA. <sup>6</sup>School of Environment, Jinan University, Guangzhou 510632, China. <sup>7</sup>Department of Chemistry, Rutgers University, 315 Penn St., Camden, NJ 08102, USA. <sup>8</sup>College of Life Science and Bio-Engineering, Beijing University of Technology, Beijing 100124, China.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The structural information of the GNPs and the predicted nanologP values are available in Additional file 4: Table S1. The codes and the data used for calculating nanologP are available as additional files 1, 2, 5. Codes include gnpprep.py, for generating vGNPs as PDB files from an SDF file; and colorQuantification.

py, for calculating nanologP. Data include univ\_multi.sdf, the original surface ligand and GNP structure of the library of 41 modeling and 9 external vGNPs.

### Funding

This work was supported by the National Key R&D Program of China (2016YFA0203103), the National Natural Science Foundation of China (91543204 and 91643204), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB14030401).

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 May 2018 Accepted: 10 January 2019

Published online: 18 January 2019

### References

- Shi J, Votruba AR, Farokhzad OC, Langer R (2010) Nanotechnology in drug delivery and tissue engineering: from discovery to applications. *Nano Lett* 10:3223–3230
- Zhang L, Gu FX, Chan JM et al (2008) Nanoparticles in medicine: therapeutic applications and developments. *Clin Pharmacol Ther* 83:761–769
- Winkler DA, Mombelli E, Pietroiusti A et al (2013) Applying quantitative structure–activity relationship approaches to nanotoxicology: current status and future potential. *Toxicology* 313:15–23. <https://doi.org/10.1016/j.tox.2012.11.005>
- Oberdörster G (2010) Safety assessment for nanotechnology and nanomedicine: concepts of nanotoxicology. *J Intern Med* 267:89–105
- Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem* 58:4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>
- Hansen K, Biegler F, Ramakrishnan R et al (2015) Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. *J Phys Chem Lett* 6:2326–2331. <https://doi.org/10.1021/acs.jpclett.5b00831>
- Lagorce D, Sperandio O, Baell JB et al (2015) FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res* 43:W200–W207. <https://doi.org/10.1093/nar/gkv353>
- Krämer SD, Wunderli-Allenspach H (2001) Physicochemical properties in pharmacokinetic lead optimization. *Farmacol* 56:145–148
- Gleeson MP (2008) Generation of a set of simple, interpretable ADMET rules of thumb. *J Med Chem* 51:817–834. <https://doi.org/10.1021/jm701122q>
- Walkey CD, Olsen JB, Song F et al (2014) Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. *ACS Nano* 8:2439–2455. <https://doi.org/10.1021/nn406018q>
- Chen R, Zhang Y, Monteiro-Riviere NA, Riviere JE (2016) Quantification of nanoparticle pesticide adsorption: computational approaches based on experimental data. *Nanotoxicology* 10:1118–1128. <https://doi.org/10.1080/17435390.2016.1177745>
- Pathakoti K, Huang MJ, Watts JD et al (2014) Using experimental data of *Escherichia coli* to develop a QSAR model for predicting the photo-induced cytotoxicity of metal oxide nanoparticles. *J Photochem Photobiol B Biol* 130:234–240. <https://doi.org/10.1016/j.jphotobiol.2013.11.023>
- Fourches D, Pu D, Tassa C et al (2010) Quantitative nanostructure – activity relationship modeling. *ACS Nano* 4:5703–5712. <https://doi.org/10.1021/nn1013484>
- Mikolajczyk A, Malankowska A, Nowaczyk G et al (2016) Combined experimental and computational approach to developing efficient photocatalysts based on Au/Pd–TiO<sub>2</sub> nanoparticles. *Environ Sci Nano* 3:1425–1435. <https://doi.org/10.1039/c6en00232c>
- Fourches D, Pu D, Li L et al (2016) Computer-aided design of carbon nanotubes with the desired bioactivity and safety profiles. *Nanotoxicology* 10:374–383. <https://doi.org/10.3109/17435390.2015.1073397>
- Jagiello K, Chomicz B, Avramopoulos A et al (2017) Size-dependent electronic properties of nanomaterials: How this novel class of nanodescriptors supposed to be calculated? *Struct Chem* 28:635–643. <https://doi.org/10.1007/s11224-016-0838-2>
- Sizochenko N, Mikolajczyk A, Jagiello K et al (2018) How the toxicity of nanomaterials towards different species could be simultaneously evaluated: a novel multi-nano-read-across approach. *Nanoscale* 10:582–591. <https://doi.org/10.1039/c7nr05618d>
- Toropov AA, Toropova AP, Puzyn T et al (2013) QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells. *Chemosphere* 92:31–37. <https://doi.org/10.1016/j.chemosphere.2013.03.012>
- Luan F, Tang L, Zhang L et al (2016) A further development of the QNAR model to predict the cellular uptake of nanoparticles by pancreatic cancer cells. *Food Chem Toxicol* 112:571–580
- Mikolajczyk A, Pinto HP, Gajewicz A et al (2015) Ab initio studies of anatase TiO<sub>2</sub> (101) surface-supported Au<sub>8</sub> clusters. *Curr Top Med Chem* 15:1859–1867. <https://doi.org/10.2174/1568026615666150506151826>
- Li S, Zhai S, Liu Y et al (2015) Experimental modulation and computational model of nano-hydrophobicity. *Biomaterials* 52:312–317. <https://doi.org/10.1016/j.biomaterials.2015.02.043>
- Wang W, Sedykh A, Sun H et al (2017) Predicting nano-bio interactions by integrating nanoparticle libraries and quantitative nanostructure activity relationship modeling. *ACS Nano*. <https://doi.org/10.1021/acs.nano.7b07093>
- Connolly M (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713. <https://doi.org/10.1126/science.6879170>
- Sethian JA (1998) Fast marching methods and level set methods for propagating interfaces. *Proc Natl Acad Sci USA* 93(4):1591–1595
- Wildman SA, Crippen GM (1999) Prediction of physicochemical parameters by atomic contributions. *J Chem Inf Comput Sci* 39:868–873. <https://doi.org/10.1021/ci9903071>
- Heiden W, Moeckel G, Brickmann J (1993) A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J Comput Aided Mol Des* 7:503–514. <https://doi.org/10.1007/BF00124359>
- Moyano DF, Goldsmith M, Solfiell DJ et al (2012) Nanoparticle hydrophobicity dictates immune response. *J Am Chem Soc* 134:3965–3967. <https://doi.org/10.1021/ja2108905>
- Cheng T, Zhao Y, Li X et al (2007) Computation of octanol–water partition coefficients by guiding an additive model with knowledge. *J Chem Inf Model* 47:2140–2148. <https://doi.org/10.1021/ci700257y>
- Tetko IV, Tanchuk VY (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J Chem Inf Comput Sci* 42:1136–1145. <https://doi.org/10.1021/ci025515j>
- BioByte. <http://www.biobyte.com/>. Accessed 22 Feb 2018
- Chemical Computing Group ULC (2013) Molecular operating environment (MOE)
- Muratov EN, Varlamova EV, Artemenko AG et al (2012) Existing and developing approaches for QSAR analysis of mixtures. *Mol Inform* 31:202–221. <https://doi.org/10.1002/minf.201100129>
- 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. <https://doi.org/10.1021/acs.jpcc.5b05759>
- Mikolajczyk A, Sizochenko N, Mulkiewicz E et al (2017) Evaluating the toxicity of TiO<sub>2</sub>-based nanoparticles to Chinese hamster ovary cells and *Escherichia coli*: a complementary experimental and computational approach. *Bellstein J Nanotechnol* 8:2171–2180. <https://doi.org/10.3762/bjnano.8.216>