The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators

Grant Agreement No 731032

Deliverable Report 5.6

Deliverable	D5.6 First corona simulation tools integrated into NanoCommons KnowledgeBase
Work Package	WP5: JRA3 - Analysis and Modelling tools
Delivery date	M18 - 22 February 2020 (Month 26)
Lead Beneficiary	NUID UCD
Nature of Deliverable	Demonstrator
Dissemination Level	Public (PU)

Submitted by	Konstantinos Kotsis (NUID UCD), Vladimir Lobaskin (NUID UCD), Dieter Maier (Biomax Ltd.)
Revised by	Iseult Lynch (UoB), Antreas Afantitis (NovaM)
Approved by	Iseult Lynch (UoB)

This project has received funding from European Union Horizon 2020 Programme (H2020) under grant agreement nº 731032

Table of contents

1. Summary

Interactions of a foreign nanomaterial (NM) with biological tissues control its fate and biological activity, including any eventual hazard. It is now well accepted that the interaction of NMs with biological fluids leads to the formation of a protein layer on the surface of the NM which is known as the protein corona. It has been established that the corona plays the central role in the bioactivity of NMs. The list of proteins present in the corona in a specific biofluid sample depends, on the one hand, on the NM chemistry and surface structure and reflects also the immediate interactions with the available proteins. One can hope to identify the biologically relevant physicochemical characteristics of the NM in the adsorbed proteins' statistics (i.e., the factions of the different Amino acids (AAs) in the protein corona).

This report describes integration of a corona modelling tool, based on the method developed within the SmartNanoTox project, into the NanoCommons research infrastructure knowledge base and computational platform. The main outcomes of the tool are adsorption energies for arbitrary proteins on selected NMs of different sizes, the adsorption energy heat map for various protein orientations (which provides insights into the most probable orientations for binding), and protein ranking by the binding affinity to the specified NM which may facilitate prediction of protein coronas in the future without the need for experimental inputs. Currently, the modelling is possible for selected set of NMs including gold, titanium oxide, and silica, and is now ready to be offered to users as a Transnational Access (TA) service.

2. Introduction

Interactions of a foreign nanomaterial (NM) with biological tissues control its fate and biological activity, including cellular attachment, uptake and any eventual hazard. Quantitative study of these interactions is extremely challenging due to the presence of multiple molecule types and material chemistries - as well as structure-specific effects, and while the immense system size prohibits the atomistic level modelling. It is now well accepted that the interaction of NMs with biological fluids leads to the formation of a protein layer on the surface of the NM which is known as the protein corona. It has been established that the corona plays the central role in the bioactivity of the NM. NMs of size of tens of nanometers can bind hundreds of different proteins [1-5].

The list of proteins present in the corona in a specific sample depends, on the one hand, on the NM chemistry and surface structure and reflects the immediate interactions. On the other hand, it also depends on the content of the biological fluid the NM is immersed in through the concentrations of the solutes and their relative affinities for the NMs' surface. Due to the presence of thousands of protein types in certain biological fluids, such as blood serum, the variability of the corona content may be immense. Still, one can hope to detect the influence of the physicochemical properties of the NM in the adsorbed proteins' statistics for different NMs, i.e., the count of a certain AA weighted by the protein abundance in the corona, thus the factions of different AAs in the corona. Certain features of proteins such as charged or hydrophobic patches, aromatic residues, etc. may tend to increase the propensity of molecules to adsorb on specific surfaces. Therefore, the characteristics of the preferentially adsorbing molecules, i.e. those with highest affinity for hte NM surface, form the NM fingerprint with respect to its bionano interactions [6]. These fingerprints appear to be useful for prediction of the biological activity of NMs, particularly NP-cell association,[6,7] which is a first step towards NM internalisation and can induce numerous responses in the cell including receptor activation.

Here, we present a simulation tool for evaluation of the adsorption energy of arbitrary proteins on a specific NM surface, which was developed in SmartNanoTox project [8,9], and has now been integrated into the NanoCommons computational platform to support its utilisation by the community, including via Transnational Access (TA) provision. The goal of the tool is to compare and rank biomolecules by their adsorption affinity to specific NMs and thus form a basis for producing NM biointeraction fingerprints. In this deliverable report, we describe the modelling techniques, as well as their coupling in Section 2. In Section 3, we describe the implementation of the tool into the NanoCommons knowledge base and computational platform and in Section 4 an example of using the tool for evaluation of adsorption energy of human serum albumin (HSA) protein onto gold NM is presented.

3. Methodology

3.1 Multiscale model of NM-protein interaction

Once an NM comes into contact with a biological medium, a protein corona forms on its surface [10]. It has been extremely challenging to develop a model that can predict the composition of the protein corona around an inorganic NM, as this depends on a multitude of physicochemical properties both of the protein, such as charge, isoelectric point and hydrophilicity/hydrophobicity and of the NM, such as size, shape, pH, hydrophilicity/hydrophobicity, and charge distribution.

Computer simulations of the interactions of NMs with proteins can offer a great support to experiments because of their great speed and flexibility [11]. Full-atomistic simulations have already proven to be a valuable tool in elucidating the binding mechanisms of proteins on metallic NMs [12-14]. However, their performance is severely hindered by the inefficiency in simulating systems with large NMs due to the high number of pair interactions that need to be evaluated. To speed-up the calculations, a cut-off of at the order of a nanometer is often introduced. However, this results in an underestimation of the adsorption energies of proteins on NMs due to the neglect of the influence from the core of the NM, which contributes much to their mutual attraction (especially at radii of over 10 nm) and so cannot be neglected. In this section, we describe a coarse grained (CG) model of protein-NM interactions that overcomes most of the challenges in the inclusion of NM core in the interaction.

While the number of atom-atom pairs in this problem is extremely large, the overall energy includes numerous instances of the same contributions, like the pair interaction of a given amino acid (AA) with a unit volume of the NM. One can dramatically reduce the amount of calculations by pre-computing those interactions for the specific materials in a specific medium. For a protein, one would need to pre-compute the interaction potentials between the AA and the NM through water, bearing in mind that different potentials will be needed for the interaction with the NM surface and the core. Assuming that these interactions do not depend on the position of the AA inside the protein and are additive (the two most significant approximations of our model), one can quickly scan multiple proteins once the potentials are known.

We calculate the coarse-grained adsorption energy for a protein molecule as a sum of energies of non-bonded (van der Waals + excluded volume) and electrostatic interactions between the AA and segments of the NM as:

$$U = U_{nb} + U_{el} \quad (1)$$

$$U_{nb}(d,\theta,\phi) = \sum_{i=1}^{N_{AA}} U_i\left(h_i(d,\theta,\phi)\right) \quad (2)$$

$$U_{el}(d,\theta,\phi) = \sum_{i=1}^{N_{AA}} \frac{\psi_S q_i R}{R+h_1(d,\theta,\phi)} e^{-\kappa h_i(d,\theta,\phi)} \quad (3)$$

where, $R$is the NM radius, $h_i(d,\theta,\phi)$ is the distance between the AA center and the NP surface, which depends on the distance d from the protein center of mass (COM) to the surface and orientation angles (θ,ϕ)), $\kappa = \sqrt{8\pi l_B I_0}$ is the inverse Debye length where $l_B = \frac{e^2}{4\pi\epsilon k_B T}$ is the Bjerrum length and I_0 is the solution ionic strength, ψ_s is the electrostatic surface potential of the NP, q_i is the charge of the AA, and $N_{AA}$is the number of AAs. The non-bonded interaction energy includes contributions from the surface and the core of the NM:

$$U_i\big(h_i(d,\theta,\phi)\big) = (U_s)_i + (U_c)_i \quad (4)$$

where $U_{s,}U_c$ are the potential energies of interaction between the i-th AA and surface or core parts of the NM, respectively. The surface part is calculated using atomistic molecular dynamics (MD) simulation with a fixed cut-off and thus includes occasional hydration effects, local charge effects or surface geometry, while the core part covers the long-range contributions to the interaction from the rest of the NM, beyond the MD cut-off distances to ensure that the proteins sees the entire nanoparticle.

In the remainder of this section, we describe the systematic procedure of evaluation of the required potentials of mean force. In the final section, we test the model by simulating the adsorption of some common blood serum proteins on gold NMs.

3.1.1 Coarse-grain protein model

In contrast to NMs, which populate a whole universe of chemistries and structures, the description of biomolecules can be significantly compressed due to their chemical uniformity, e.g. the same AAs are present in all proteins or the same nucleic acids in all DNA. The AA sequence alone, however, does not provide exhaustive information as the shape and functionality of the protein depends on its 3D structure. The 3D structure of the molecule, where possible, can be retrieved from the Protein Data Bank (PDB). For our calculations, we consider the proteins as rigid 3D globules with fixed conformations corresponding to their crystal structures as presented in the PDB. The model thus preserves two main structural features that guide the binding mechanisms, i.e., the overall shape of the protein and the charge distribution. Although proteins may change conformation upon interaction with the NM surface, this only happens when the attraction forces are sufficiently strong, as, for example, can be observed for dense metals like gold, silver, or large nanoparticles of size of tens of nanometres, where the adsorption energies may exceed 10^2 kJ/mol per molecule. As these strong attractions make adsorption irreversible, we expect that the conformational changes do not affect the corona content even if they lead to even stronger adherence of the proteins to the NM surface, so they can be safely neglected.

We here use a one-bead-per-AA (united atom – UA) model of globular proteins, which is suitable for estimating the adsorption energy [8]. Our UA protein model, in which every AA in the protein is substituted by a single bead whose center is placed at the position of the α-carbon atom, is illustrated by Figure 1.

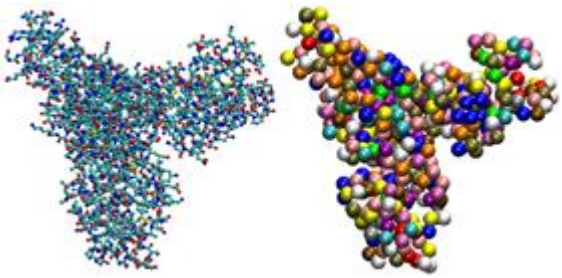

Figure 1. All-atomistic (left) and united atom – UA (right) models of lung surfactant protein D (SP-D).

3.1.2 Coarse-grain NM

The protein model described above allows us to reduce the number of components in treating the protein. NM size, however, also plays a fundamental role in the formation of the corona and in the interactions at the bio-nano interface. The number of atoms needed to represent an NM is again a severe limitation to all atomistic calculations. Simulation of NMs of size greater than 10 nm in biological media is an unfeasible task even for modern computers and the CG model for describing NMs is therefore highly needed.

Our model starts by considering the contributions that different atoms in the NM (e.g. Ti versus O in a TiO_2 NM) give to the binding interaction. Note that, as is the case experimentally, the parameterisation is different for different crystal phases and different surfaces / faces, from which a weighted average is calculated over the different crystalline phases to produce the energy for the specific NM (e.g. rutile or anatase). Based on the nature of these contributions, we partition the NM into a core segment and a surface segment. The outer layer on the NM surface is directly in contact with the solvent and the pair interaction with the protein residues must include solvent effects as well as the chemical composition, charge, and hydrophilicity/ hydrophobicity of the NM surface. Therefore, the interaction of each residue with the nearest part of the surface must be parameterized to reflect these details, using full-atomistic simulations. The size of the surface segment is thus determined by the cutoff, r_c, used in the full-atomistic simulation (typically, 1 to 2 nm). Geometrically, the surface segment is a lens formed by the intersection of a sphere of radius r_c, centered on the AA bead, and a sphere of radius of the NM, R, centered on the NM itself (Figure 2). The core comprises the majority of the atoms, but these only interact with the protein via long-range forces, for which we assume that a continuum-level description is sufficient. The core of the NM is then modeled as a single bead of the shape of a sphere of radius R with a cut-out surface lens. The potential between the core and the AA beads in our model is calculated using the Lifshitz theory [15] for interaction between two macroscopic bodies. In the next two subsections, we describe how the potentials are parameterized for the AA-NM interaction.

3.1.3 Generation of surface potentials

Adsorption free energy profiles (or potentials of mean force - PMFs) can be calculated using adaptive well-tempered metadynamics (AWT-MetaD) [9,16-19]. This method uses a time-dependent bias term which is evolved according to:

$$\frac{dV(z,t)}{dt} = \omega e^{-V(z(t),t)/\Delta T} \exp\left[-\left(z-\bar{z}(t)\right)^2/2\,\sigma^2(t)\right] \quad (5)$$

withωbeing the initial filling rate and ΔT being a temperature boost factor that determines how large a part of the free energy space is reached, z is the distance from the AA centre of mass to the surface. The positions and widths of the added Gaussians are changed on-the-fly as exponentially weighted averages, as follows:

$$\bar{z}(t) = \frac{1}{\tau_D}\int \int_0^t dt\,' z(t') e^{(-(t-t\prime)/\tau_D)} \quad (6)$$

and

$$\sigma^2(t) = \frac{1}{\tau_D}\int_0^t dt\,'(z(t') - \bar{z}(t'))^2 e^{-(t-t\prime)/\tau_D} \quad (7)$$

The single parameter τ_D determines the time window used to estimate fluctuations in z. The free energy is recovered in a single simulation from the expression [14]:

$$F_\Omega(z) = -\lim_{t\to\infty} [V(z,t) + k_B T \ln \Omega(z,t) \text{ const.}\times t] \quad (8)$$

where $\Omega(z,t)$is the accumulated histogram of the reaction coordinate z up to time t. The constant on the right-hand side is independent of z and t. The logarithmic divergence as $t \to \infty$ is irrelevant as long as we are interested in free energy differences only along z-axis. The AWT-MetaD simulations need to be run for ca. 200 ns to allow the system to visit all states many times for each AA or lipid segment.

We further used the free energies $F_\Omega(z)$ for each AA as estimates of the surface terms $U_s(z), U_s(z) = F_\Omega(z)$ with the integration constant set to zero. Upon evaluating the PMFs, we also calculated the mean adsorption free energies for each AA as:

$$E_{AA} = -k_B T \ln \left[\frac{1}{\delta}\int_0^\delta \exp\left(-\frac{U_s(z)-U_s(\delta)}{k_B T}\right) dz\right] \quad (9)$$

The integration cut-off distance δ is set at 0.8 nm, as all the PMFs usually vanish at that distance. Generally, the interaction with a convex surface of a NM of finite radius is less than that for the flat slab due to the lesser number of atoms of the NM within the interaction cut-off distance. To account for this reduction, we correct the PMFs for the flat surface by a distance-dependent multiplicative function $f(h)$ that reflects also the cut-off radius (r_c) used in the calculations as well as the radius R of the NM:

$$U_s(h,R) = U_s(h,\infty) f(h) \quad (10)$$

Here, $U_s(h,R), U_s(h,\infty)$ are the PMFs for the curved and flat surfaces, respectively, and$h$is the minimum distance between the AA-bead center and the NM surface. By taking the appropriate limits for R, we can calculate a correction factor for any geometry. A diagram showing how an AA bead interacts with a CG NM is shown in Figure 2 [9].

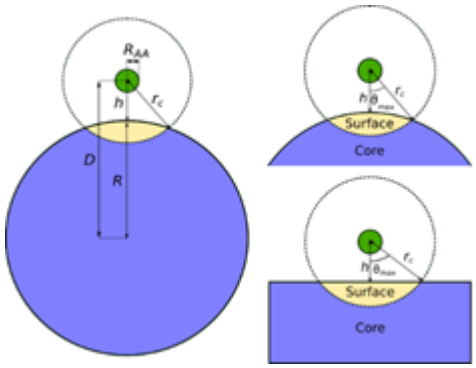

Figure 2. Left: A schematic of an amino acid (AA) bead of radius R_{AA} interacting with a NM of radius R, where the distance between the centres of the two particles is given by D and the distance between the surface of the NM and the centre of the bead is given by h. The NM is divided into a surface region (yellow) containing the volume within a distance r_c of the centre of the bead, and a core region (blue) containing the volume outside this cutoff range. Right: Schematics of the surface region for a spherical NM (top) and a planar slab (bottom), indicating the angle θ_{max} used for the calculation of the surface correction factor.

We assume that the relevant point-point interaction is dominated by dispersion forces that scale as r^{-6}, where r is the distance between the interaction centers (e.g. atoms). We then assume that AAs are much smaller than the NM. In this case, the attraction energy for a small particle of volume V to a large sphere of radius R, such that $R \gg V^{1/3}$, given the finite cut-off r_c, can be calculated as:

$$U_s(h,R) = \varepsilon V \int_h^{r_c} \int_0^{\vartheta_{max}} \int_0^{2\pi} \frac{r^2 \sin\theta}{r^6} d\phi d\theta dr \tag{11}$$

$$\vartheta_{max} = \cos^{-1}\left(\frac{r^2 - R^2 + (R+h)^2}{2r(R+h)}\right)_{max} \tag{12}$$

$$U_s(h,R) = -\frac{\pi\varepsilon V}{h+R}\left(\frac{h-2R}{12h^3} + \frac{-6r_c^2 + 8r_c(h+R) - 3h(h+2R)}{12r_c^4}\right) \tag{13}$$

Here, ε is the interaction energy per unit volume. For $R \to \infty$ Eq. (13) reduces to that of a flat surface:

$$U_s(h,\infty) = \pi\varepsilon V\left(\frac{1}{6h^3} - \frac{2}{3r_c^3} + \frac{h}{2r_c^4}\right) \tag{14}$$

and the correction factor is calculated as:

$$f = \frac{U_s(h+R)}{U_s(h,\infty)} = -\frac{r_c^2(h-2R) + 2r_c h(h-2R) - 3h^2(h+2R)}{2\left(r_c^2 + 2r_c h + 3h^2\right)(h+R)} \tag{15}$$

Figure 3 [9] shows how the volume correction factor changes with the distance from the AA to the surface for a set of NM radii. One can see that, as the radius increases approaching the flat surface limit, $f \to 1$ for all values of h.

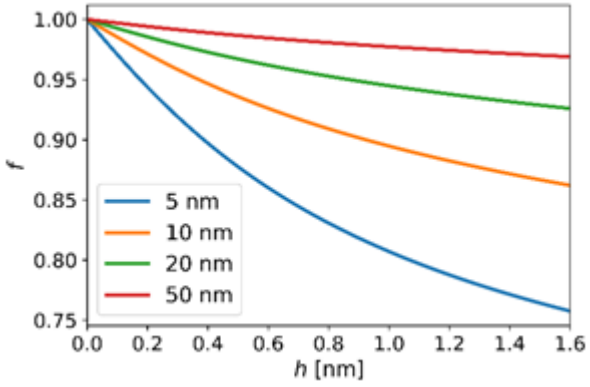

Figure 3. Correction factor $f(h)$ vs. distance from the surface h, Eq. (15), for a range of NM radii.

3.1.4 Generation of the core potential

The NN core plays a crucial role in the protein adsorption as it contains most of the NM. A serious limitation of all-atom models is the difficulty of obtaining a correct accounting of the attraction by the core atoms. This problem is mainly due to the short-range cut-off employed in simulations leading to a considerable underestimation of the adsorption energies. The latter, however, can be easily calculated in the continuum approximation, which is commonly used in colloid science. The correction

we propose in this report is to evaluate the contribution of the core of the NM at distances $r > r_c$ by treating the remote part of the NM as a single sphere less the part within the cut-off distance from the specific AA. The interaction energy between an AA and the NM core can be computed using the Hamaker method for dispersion forces. We take into account only the part of the NM that is beyond the reach of the PMF cut-off distance r_c. Then, for two spheres or radii R, R_{AA} at a distance D between their centers, such that the interaction energy is the Hamaker potential:

$$U_{c,long}(D) = \frac{-A_{123}}{12}\left(\frac{4RR_{AA}}{D^2-(R+R_{AA})^2} + \frac{4RR_{AA}}{D^2-(R-R_{AA})^2} + 2\ln\left(\frac{D^2-(R-R_{AA})^2}{D^2-(R+R_{AA})^2}\right)\right) \quad (16)$$

At the shorter distances, the Hamaker potential must be corrected for the interaction with the lens cut out of the NM by a sphere of radius r_c centered at the AA center of mass, as it is covered by the PMF:

$$U_c(D) = U_{c,long}(D) + U_{c,lens} \approx U_{c,long}(D) + \frac{A_{123}}{12}\left(\frac{4\pi^2 R_{AA}^3}{3D}\left(\frac{D-3R}{(D-R)^3} + \frac{-6r_c^2+8r_cD-3h(D+R)}{r_c^4}\right)\right) \quad (17)$$

In the above equations, A_{123} is the Hamaker constant for interaction between material 1 (e.g. protein) with material 2 (e.g. the NM) through material 3 (e.g. water), which is the only material dependent term in the equation. The Hamaker constant for the AA (phase 1)-material (phase 2) interactions through a medium (phase 3) can be obtained as follows:

$$A_{123} = \frac{3}{4}k_BT\frac{(\varepsilon_1-\varepsilon_3)(\varepsilon_2-\varepsilon_3)}{(\varepsilon_1+\varepsilon_3)(\varepsilon_2+\varepsilon_3)} + \frac{3h\nu_e}{8\sqrt{2}}\frac{(n_1^2-n_3^2)(n_2^2-n_3^2)}{(n_1^2+n_3^2)^{1/2}(n_2^2+n_3^2)^{1/2}\left((n_1^2+n_3^2)^{1/2}+(n_2^2+n_3^2)^{1/2}\right)} \quad (18)$$

Where n_i are the refractive indices of the materials in the visible region, ν_e is the main electronic absorption frequency in the UV range (typically 3×10^{15} s^{-1}) for material 2 and ε_i are the dielectric permittivities of the materials, which are equal to n^2 in the visible part of the spectrum. In the case where one of the materials is a conductor, Eq. (18) must be modified to take into account the high values of the polarisability and therefore of the dielectric constant. The equation for a dielectric-conductor interaction in a medium is given by [15]

$$A_{123} = \frac{3}{8\sqrt{2}}\left(\frac{n_1^2-n_3^2}{n_1^2+n_3^2}\right)\frac{h\sqrt{\nu_1\nu_3}\cdot\nu_2}{\sqrt{\nu_1\nu_3}+\frac{\nu_2}{\sqrt{n_1^2-n_3^2}}} \quad (19)$$

where ν_i are the frequencies of maximum absorption for the material in the UV region; for metals this corresponds to the plasma frequency.

3.1.5 Evaluation of the adsorption free energy

To evaluate the average adsorption energy for a protein globule on a NM we scan the configurational space (i.e. all possible orientation in which a protein can adsorb on the surface of the NM) by a systematic rotation of the protein and calculate the Boltzmann-averaged energy. There are three degrees of freedom (DOF) that must be scanned [8,21]: The orientation of the protein (Figure 4) can be described by a vector from the center of mass (COM) to an arbitrary point of the molecule. It is characterized by two angles: ϕ and θ and by rotating the molecule an angle $-\phi$ about the z-direction and then by an angle $-\theta + 180°$ about the y-axis will make the position vector point towards the surface (along the negative z-axis). The third DOF is the distance from the COM to the closest point of the surface, d_{COM}. Instead of obtaining the actual adsorption free energy by calculating the PMF for

all orientations and conformations of the protein molecule, we calculate a composite energy U , which is the sum of all the pairwise interactions between the surface and the AAs. In the energy U, some of the interactions (van der Waals) are represented by the potential energy while the other ones (electrostatic and surface PMFs) include ensemble averages over the positions of water molecules, ions, and segment orientations, and in this sense represent the free energy of adsorption. For each configuration (ϕ_i, θ_j), the total energy is calculated as a function of distance of the COM, $U(D, \phi_i, \theta_j)$, to the surface for the case of a slab or to the center of the NM for the case of a NM. Following a similar approach as in Kokh et al. [21], and denoting the reaction coordinate $D = z$, the mean interaction energy between a protein with orientation (ϕ_i, θ_j) and a spherical NM is given by:

$$E(\phi_i, \theta_j) = -k_B T \ln\left[\frac{3}{(R+a)^3 - R^3}\int_R^{R+a} D^2 \exp\left(-\frac{U(D,\phi_i,\theta_j)}{k_B T}\right) dD\right] \quad (20)$$

where $a = a(\phi_i, \theta_j)$ is the maximum interaction distance from the center of mass of the protein to the NM surface for the given orientation. The mean adsorption free energy is obtained by averaging this interaction energy over the distribution of orientations,

$$E_{ad} = \sum\sum P_{ij}\, E(\phi_i, \theta_j) \quad (21)$$

The weighting factor in Eq. (21) is given by:

$$P_{ij} = \sin\theta_j \exp\left(-\frac{E(\phi_i,\theta_j)}{k_B T}\right) \quad (22)$$

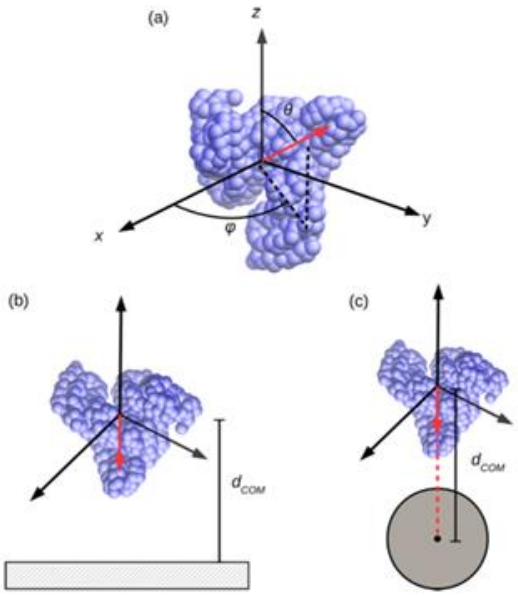

Figure 4. Definition of protein orientation. (a) Any "atom" of the protein can be described by a position vector from the COM, whose orientation is characterized by two angles ϕ and θ. The angles correspond to azimuthal and polar rotations that would turn the original vector towards the surface (along the negative z-axis). The remaining degree of freedom is the distance of the COM, d_{COM}, to (b) the surface for a slab or (c) to the center of the NM [8].

3.1.6 Simulation parameters

We use AWT-MetaD to compute PMFs of AA residues at NM surfaces. There are 20 naturally occurring AAs. We also consider two protonated versions of histidine (HID and HIE) so there are 22 biomolecules

in total, as per Figure 5 below. The backbone fragments of amino acids are terminated by neutral NH_2 and COOH groups in order to mimic their behavior in a long peptide chain. AAs are described by the AMBER03 force field [19]. Here we used parameters from the force field by Heinz *et al*. [22]. We used the Lorentz-Berthelot combination rules to describe the interaction of NM atoms with water and AA. For water, the TIP3P model is used.

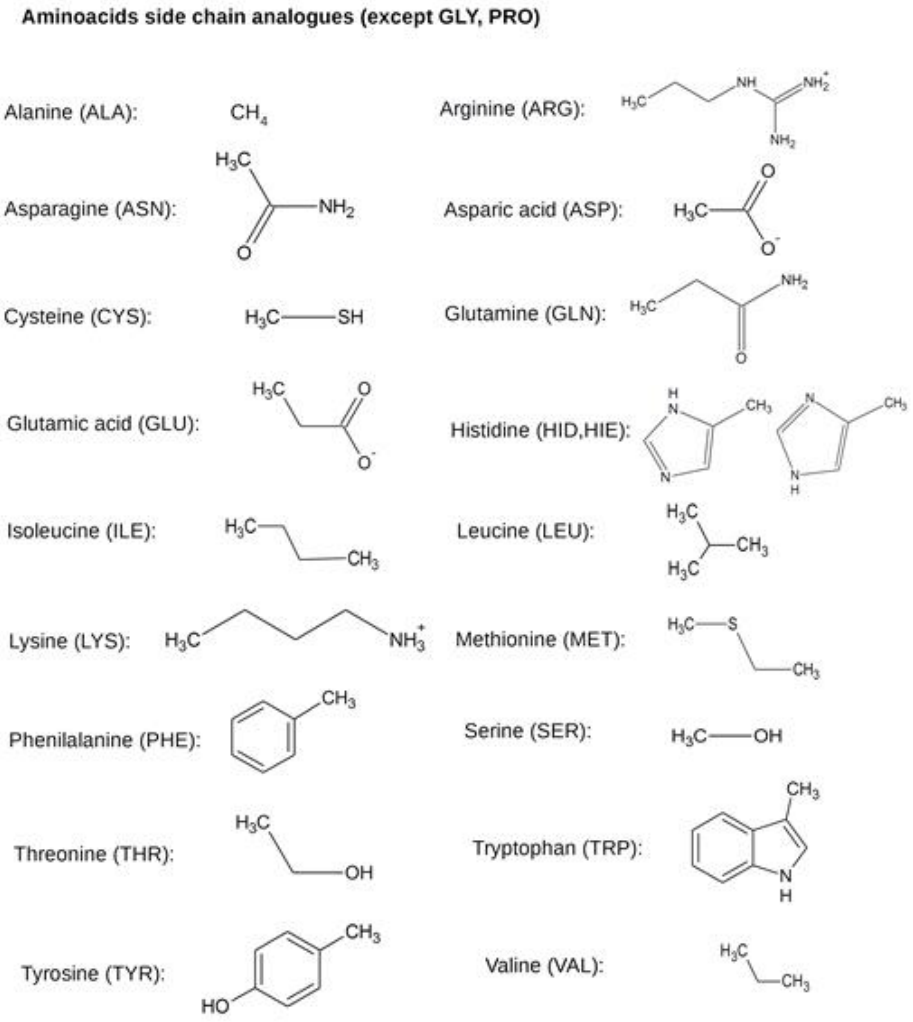

Figure 5: AA side chain analogues, showing also the protonated histidines (HID, HIE).

The adsorption energy calculations described in Section 3.1.5 are implemented in a C++ program, UnitedAtom, which handles the summation of the interaction potential over all the AA beads and the numerical integration of Eq. (20). Briefly, the potential as a function of distance is tabulated for each bead, and this data is used together with the known structure of the protein to tabulate the total potential as a function of the distance from the centre of mass of the protein to the NM. The protein orientations are sampled from 0 to 355° in steps of 5° in the azimuthal direction and from 0 to 175° in steps of 5° in the polar direction. Energies are thus evaluated for 2592 different orientations. A more detailed account of the rotation procedure is given in references [8,9]. To improve the sampling of orientations, the energy at each nominal orientation is calculated as the average of 16 randomly-sampled orientations.

4. Software tool

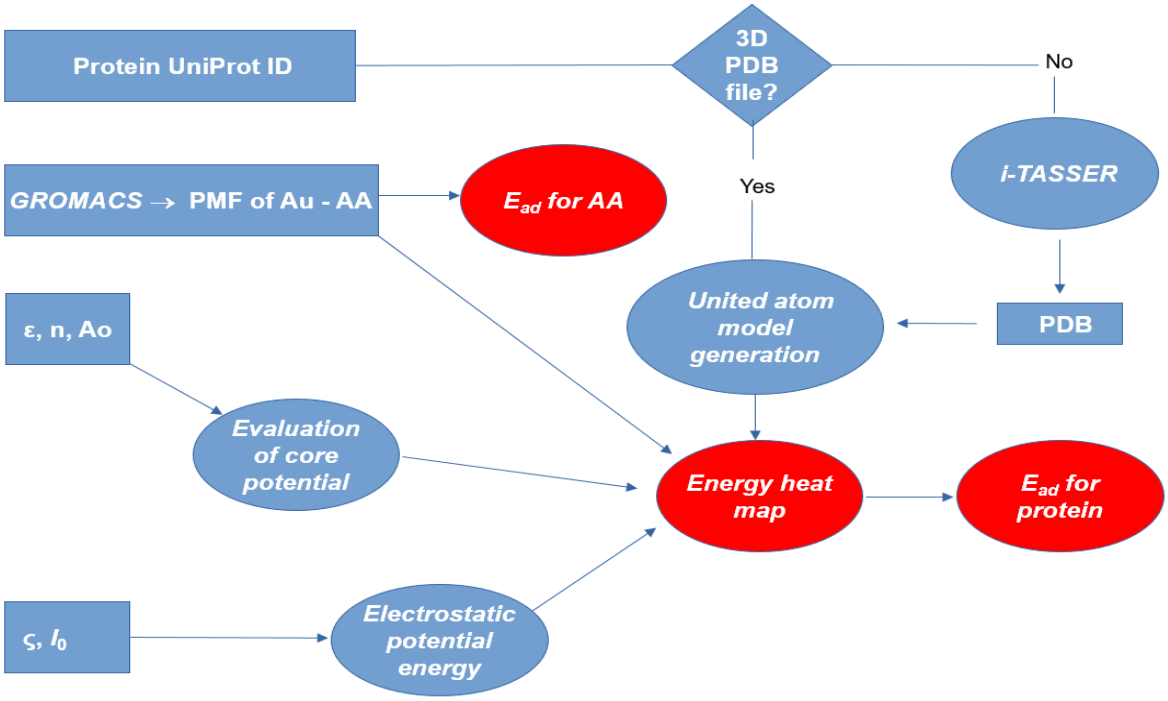

Figure 6. The workflow of the corona modelling tool developed in SmartNanoTox and integrated into NanoCommons.

In Figure 6, the workflow in the corona modelling tool is presented. Starting from a Protein UniProt ID and where a 3D PDB structure file is available the United atom tool will, with the use of certain parameters such as the Hamaker constant, the zeta potential, the refractive index and the PMF of the specific NM of interest with AAs, produce an energy heat map as given by equation (20). From this, the mean interaction energy between an NM and a protein, i.e., the mean adsorption free energy, is obtained by averaging the interaction energy over the distribution of orientations as explained in more detail above. If the 3D PDB structure is not known or note available from the database, the code will call

wget https://www.uniprot.org/uniprot/P02768.fasta

to download the protein structure and subsequently the i-Tasser program [26] is run to create a PDB structure file which is fed into the United atom tool to follow the procedure as described above.

A user with access to the NanoCommons Knowledge Base User Interface (UI) selects a specific NM and a (single) protein of interest visualise their interaction (see Figure 2) with the selection steps and UI shown in Figure 7 below.

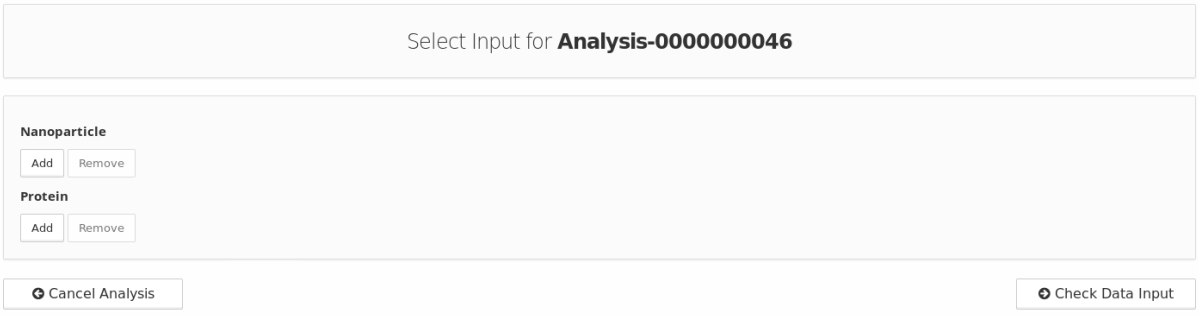

Figure 7. NanoCommons KnowledgeBase UI for data input to perform corona analysis.

The selection of the NM is restricted to types with existing force fields from atomistic simulations; force fields for TiO_2 rutile and anatase, silica, gold NMs and CNTs are available to date. A dropdown list as shown in Figure 8 of pre-parameterised NMs is provided so that users can choose one of the parameterized NMs. In the future, this library will be extended by new materials. The user Input is checked for correctness (see Figure 9) before running the calculations.

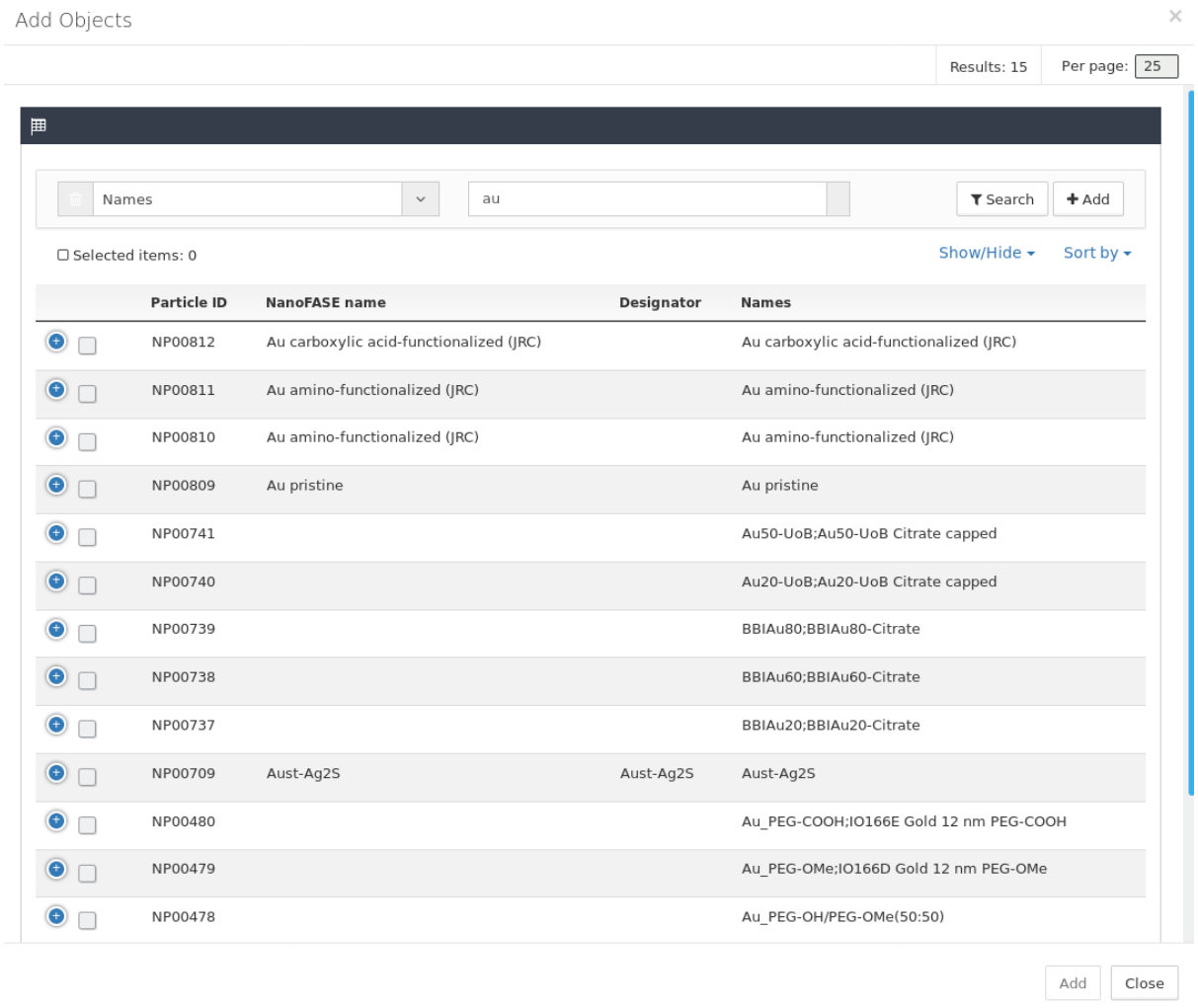

Figure 8. Snapshot showing the list of pre-parametrized NMs available in the NanoCommons KnowledgeBase.

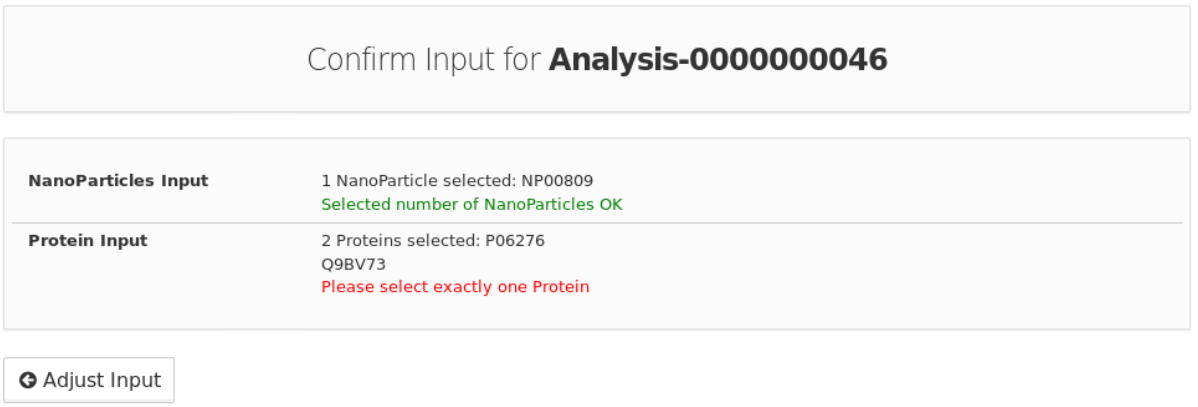

Figure 9. Confirmation step of the chosen by the user in order to perform corona analysis via the NanoCommons KnowledgeBase.

If the input is correct the user can start the analysis (see Figure 10).

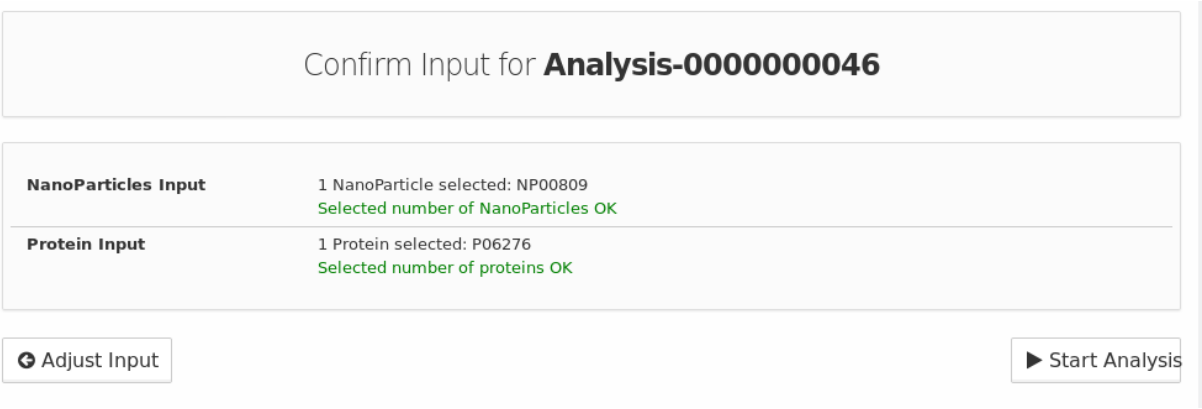

Figure 10. Start button to run the corona analysis via the KnowledgeBase.

The NanoCommons Knowledge Base then calls the simulation docker with the following input information for NM and protein:

NMe: material (ID), size, zeta potential value

Protein - PDB file including 3D structure, if no PDB entry is available UniProt FASTA sequence file is called and fed into I-Tasser routine [26] to generate the 3D structure. Then, the following output is generated:

A heat map data file with a corresponding image in png format and the average Boltzmann adsorption energy. The output will be stored in the KnowledgeBase linked to NM and protein and available in tabulated form (csv, txt). Ranking or comparison to another specific protein can be done. To predict corona compositions (i.e., abundances of the given proteins), initial solution concentrations must be provided.

The columns of the heat map output are:

Phi: The orientation of the protein around the *Z* axis (between 0° and 360° degrees) with respect to the starting orientation in the PDB file.

Theta: The orientation of the protein around the rotated Y axis (between 0° and 180°) with respect to the starting orientation in the PDB file.

Adsorption Energy: The average adsorption energy between the protein and the NM when the protein is rotated by ϕ and θ above (in units of k_BT).

Error in Adsorption Energy: The code calculates the adsorption energy several times, with slight variations to θ and ϕ (a random value between -2.5° and 2.5° to each). This value is the standard deviation of those energies (in units of k_BT). Figure 10 shows an example of a heat map of HSA adsorbed onto an Au NM [9].

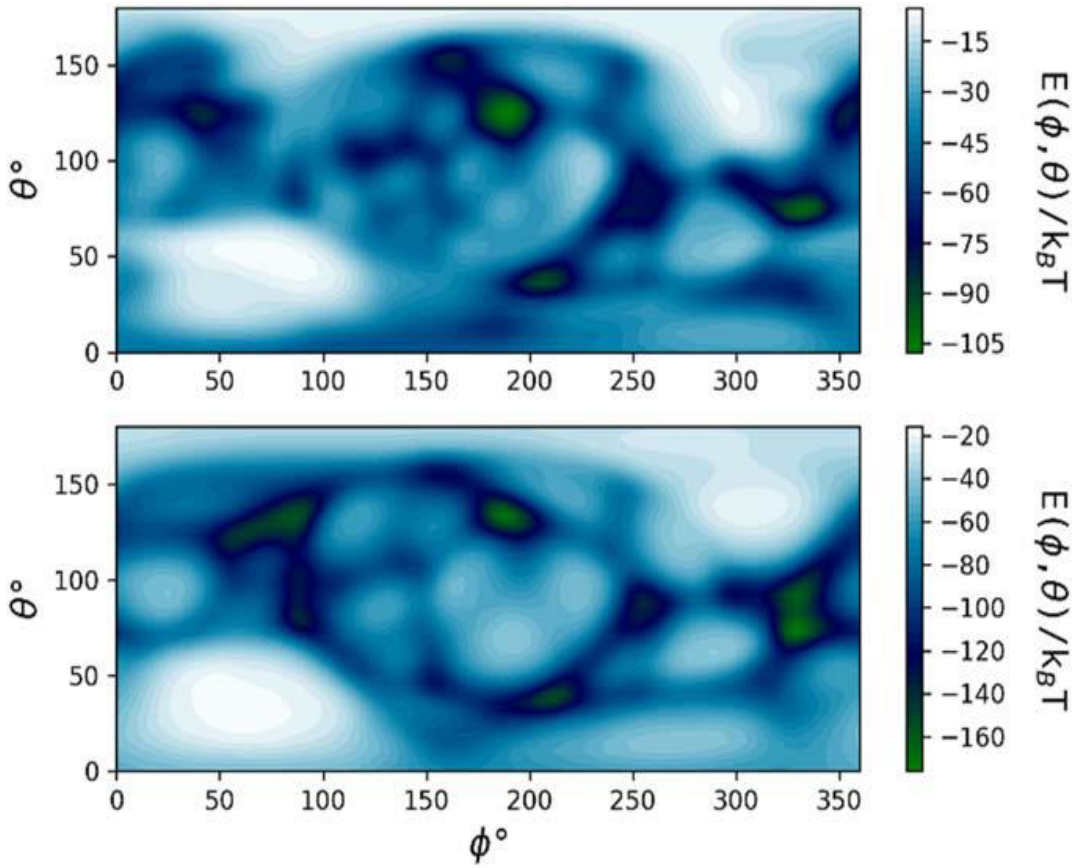

Figure 10. Orientation-specific binding energy E of HSA on an R = 5 nm (top) and R = 50 nm (bottom) Au NM [9].

5. Summary

The corona simulation tool presented here performs calculations of the adsorption free energy for arbitrary proteins based on pre-calculated atomistic potentials of mean force for the specified NM and amino acids. The integration of this corona simulation tool into the NanoCommons KnowledgeBase allows any user interested in data on the biological activity of NMs to access libraries of the KnowledgeBase and calculate the corona on a specific (or different types of) NM and to store the outcome (data) of the simulation, in a form of a adsorption energy heat map (see Figure 10), into the KnowledgeBase. As more NMs are pre-parameterised the range of NMs included in the drop-down menu available to users will increase.

Abbreviations

3D: Three-Dimensional

AMBER03: Assisted Model Building with Energy Refinement 2003

AA: Amino Acid

AWT-MetaD: adaptive well-tempered metadynamics

CG: Coarse Grain

CNTs: Carbon NanoTubes

COM: Center of Mass

CSV: comma-separated values

DOF: Degrees of Freedom

DNA: Deoxyribo-Nucleic Acid

FASTA: FAST-All

GROMACS: GROningen MAChine for Chemical Simulation

HID: Histidine

HSA: Human Serum Albumin

i-Tasser: iterative-Threading ASSEmbly Refinement

MD: Molecular Dynamics

NM: Nanomaterial

NP: Nanoparticle

PDB: Protein Data Bank

PMF: Potential of Mean Force

SP-D: Lung surfactant protein D

UA: United Atom

UniProt: Universal Protein Resource

UV: Ultra Violet

TA: Transnational Access

TIP3P: Three-site Transferable Intermolecular Potential

TXT: Text

References

1. Cedervall T, Lynch I, Foy M, Berggad T, Donnelly S, Cagney G, Linse S and Dawson K 2007 Detailed identification of plasma proteins adsorbed on copolymer nanoparticles Angew. Chem. Int. Ed. 46 5754–6
2. Landsiedel R R, Ma-Hock L, Kroll A, Hahn D, Schnekenburger J, Wiench K and Wohlleben W 2010 Testing metal-oxide nanomaterials for human safety Adv. Mater. 22 2601–27
3. Lundqvist M, Sethson I and Jonsson B H 2004 Protein adsorption onto silica nanoparticles: conformational changes depend on the particles curvature and the protein stability Langmuir 20 10639–47
4. Lundqvist M, Sethson J and Jonsson B H 2005 Transient interaction with nanoparticles 'freezes' a protein in an ensemble of metastable near-native conformations Biochemistry 44 10093–9
5. Karajanagi S S, Vertegel A A, Kane R S and Dordick J S 2004 Structure and function of enzymes adsorbed onto single-walled carbon nanotubes Langmuir 20 11594–9
6. Walkey C D, Olsen J B, Song F, Liu R, Guo H, Olsen D W H, Cohen Y, Emili A and Chan W C W 2014 Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles ACS Nano 8 2439–55
7. Kamath P, Fernandez A, Giralt F and Rallo R 2015 Predicting cell association of surface-modified nanoparticles using protein corona structure—activity relationships (PCSAR) Curr. Top. Med. Chem. 15 1930
8. Lopez H, Lobaskin V 2015 Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles J. Chem. Phys. 143 243138
9. Power D, Rouse I, Poggio S, Brandt E, Lopez H, Lyubartsev A Lobaskin V 2019 A multiscale model of protein adsorption on a nanoparticle surface Modelling Simul. Mater. Sci. Eng. 27 084003
10. Rahman M, Laurent S, Tawil N, Yahia L and Mahmoudi M 2013 Protein-Nanoparticle Interactions (Springer Series in Biophysics) vol 15 (Berlin: Springer)
11. An D, Su J, Li C and Li J 2015 Computational studies on the interactions of nanomaterials with proteins and their impact Chin. Phys. B 24 120504
12. Brancolini G, Kokh D B, Calzolai L, Wade R and Corni S 2012 Docking of ubiquitin to gold nanoparticles ACS Nano 6 9863–78
13. Ding F, Radic S, Chen R, Chen P, Geitner N, Brown J and Ke P 2013 Direct observation of a single nanoparticle–ubiquitin corona formation Nanoscale 5 9162–9
14. Khan S, Gupta A and Nandi C 2013 Controlling the fate of protein corona by tuning surface properties of nanoparticles J. Phys. Chem. Lett. 4 3747–52
15. Israelachvili J N 2011 Intermolecular and Surface Forces 3rd edn (Israela: Academic)
16. Reith D, Pütz M and Müller-Plathe F 2003 Deriving effective mesoscale potentials from atomistic simulations J. Comput. Chem. 24 1624–36

17. Izvekov S and Voth G A 2005 A multiscale coarse-graining method for biomolecular systems J. Phys. Chem. B 109 2469–73
18. Noid W G, Chu J W, Ayton G S, Krishna V, Izvekov S, Voth G A, Das A and Andersen H C 2008 The multiscale coarse-graining method: I. A rigorous bridge between atomistic and coarse grained models J. Chem. Phys. 128 244114
19. Duan Y et al 2003 A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations J. Comput. Chem. 24 1999–2012
20. Brandt E and Lyubartsev A P 2015 Molecular dynamics simulations of adsorption of amino acid side chain analogues and a titanium binding peptide on the TiO2 (100) surface J. Phys. Chem. C 119 18126–39
21. Kokh D, Corni S, Winn P, Hoefling M, Gottschalk K and Wade R 2010 ProMetCS: an atomistic force field for modelling protein-metal surface interactions in a continuum aqueous solvent J. Chem. Theory Comput. 6 1753–68
22. Heinz H, Vaia R A, Farmer B L, and Naik R R 2008 Accurate Simulation of Surfaces and Interfaces of Face-Centered Cubic Metals Using 12−6 and 9−6 Lennard-Jones Potentials J. Phys. Chem. C 112 17281-17290.
23. Abraham M J, Murtola T, Schulz R, Páll S, Smith J C , Hess B and Lindahl E 2015 GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers SoftwareX 1–2 19–25
24. Tribello G A, Bonomi M, Branduari D, Camilloni C, Bussi G 2014 PLUMED2: New feathers for an old bird Comp. Phys. Comm. 185 604-13.
25. Sun Y, Welsh W and Latour R 2005 Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation Langmuir 21 5616–26
26. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y 2015 The I-TASSER Suite: Protein structure and function prediction Nature Methods 12 7-8