Bionano interactions: A key to mechanistic understanding of nanoparticle toxicity

David Power[1], Stefano Poggio[1], Hender Lopez[2], Vladimir Lobaskin[1]

[1] *School of Physics, University College Dublin, Belfield, Dublin 4, Ireland*

[2] *School of Physics, Technological University Dublin, Dublin 8, Ireland*

Vladimir.Lobaskin@ucd.ie

The new paradigm in the assessment of toxicity of nanomaterials relies on a mechanistic understanding of the organism's response to an exposure to foreign materials from the initial, molecular level interactions to signaling and regulatory cascades. Here, we present a methodology to quantify the essential interactions at the bionano interface, which can be used in combination with the adverse outcome pathway analysis to build mechanism-based predictive schemes for toxicity assessments. We introduce a set of new, advanced descriptors of the nanomaterials, which refer to their ability to bind biomolecules and trigger the pathways via the molecular initiating events.

1.1 INTRODUCTION

Predictive toxicology is now experiencing a transition from descriptive histopathological analyses to a data-rich science with a much greater focus on the understanding of biological mechanisms down to the molecular level. The mechanistic approach to toxicity assessment involves analysis of pathways based on particle tracking and on transcriptomics or metabolomics data, reflecting the activated system level responses. In this paradigm, one assesses the possibility of initiating an Adverse Outcome Pathway (AOP), which covers the evolution of a toxic process, from its molecular initiating event (MIE), to downstream cascading key events (KEs), leading eventually to a pathology or adverse outcome. The pathways are

recognised via biomarkers in specific bioassays, which are developed to test each of these events.

In this mechanistic picture, AOPs are triggered and steered by molecular-level interactions including those at the bionano interface, a nanoscale layer where biological fluids meet foreign materials. Quantitative understanding of the interactions and structure of the bionano interface is, therefore, crucial to our ability to predict the probability of a MIE for the specific AOP and to relate it to the nanomaterial's properties. For most nanomaterials studied until now, the standard physicochemical descriptors are not sufficiently informative to predict the outcome of their interactions with biomolecules and the likelihood of the MIEs. The size, charge, and chemical composition cannot not immediately predict whether a particle can produce free radicals, bind to certain cell receptor or penetrate the cell membrane. Therefore, the mechanistic paradigm demands the development of novel characterisation techniques and new pathway-oriented descriptors.

1.2 ADVANCED DESCRIPTORS OF THE BIONANO INTERFACE

1.2.1 *Protein corona*

The exposure of nanoparticle (NPs) to biological fluids leads to the formation of a protein layer on the surface of the NP which is known as protein corona. This is a central concept in the description of the bionano interface as it has been established that the NP-protein complex is what determines the bioactivity of the nanomaterial. NPs of size of tens of nanometers can contain hundreds of different biomolecules and its structure can be quite complex and include two layers which are known as hard and soft corona [1-5]. Irreversible (or strong) binding of proteins on the NP is associated with the concept of a "hard corona" whereas quick reversible binding of proteins that have faster exchange rates are defined as "soft corona". The list of proteins present in the corona depends, on the one hand, on the NP chemistry and reflects the bionano interactions. On the other hand, it also depends on the content of the biological fluid the NP is immersed in. The corona formed in blood is extremely sensitive to the patient's health state, i.e. to the biological environment surrounding the NP, and cannot be predicted from the intrinsic properties of the NP alone [6, 7]. Nevertheless, while it is clear that the variability of the corona content is immense, one can hope to find regularities

if not in the list of molecules as such then at least in the adsorbed proteins' statistics. Certain properties of proteins such as the presence of charged or hydrophobic patches, aromatic residues, etc. may tend to increase the propensity of molecules to adsorption on specific materials. This observation led to an idea of construction of nanomaterial fingerprints that would contain the essential relevant information about the interactions and thus discriminate between different materials [8]. These fingerprints appeared to be useful for the prediction of the biological action of a nanomaterial, namely the NP-cell association. Although the initial statistical quantitative structure-activity relationships (QSAR) model developed in [8] were heavy (64 parameters), with a more thorough analysis it was possible to demonstrate that only few of them are actually important [9], e.g. cell association of gold NPs correlated well with the sequence descriptors responsible for protein charge (such as basic, acidic, and aspartic amino acids percentage) as well as with molecular weight, and propensity of the protein to aggregation. Despite that it is generally known that the abundances of any particular protein in the corona varies for different nanomaterials (see a review presented in Ref. [10]), the properties responsible for this variance have not been identified. We believe that a significant improvement in the predictive power of the bioinformatics-based models, as compared to the sequence-only descriptor sets, can be achieved with structure descriptors characterising size, shape, and charge distribution on the protein, i.e. the properties relevant for the interactions with NPs.

1.2.2 *Nanoparticle descriptors and QSARs*

Although it is clear that the physicochemical properties of NPs determine their interactions with proteins in biological matrices (e.g. blood plasma and alveolar fluid) and with the immune cells [11], there are only few known structure–activity relationships between the physicochemical properties of NPs and their effects on the immune system that lead to the most common types of immunotoxicity.

QSARs have been considered as a promising step in building the novel NP toxicity assessment strategies [12,13]. Their predictive power has been demonstrated, for example, for metal oxide NPs, where the conduction band energy levels have shown to be correlated with their toxicological potential at cellular and systemic levels [14]. However, further progress in the

construction of predictive QSARs is seen in connection to the mechanism-of-action of nanomaterials, which depends on a provision of relevant descriptors based on the AOPs. Advanced material characterization that includes both intrinsic and extrinsic properties, such as hydrophobicity, protein adsorption affinity, dissolution rates, ability to generate reactive oxygen species, etc. is necessary to elucidate the molecular level mechanisms of toxicity, to identify the properties of concern and thus provide a basis for material grouping and read-across techniques, which can be used by regulators.

Some descriptors can be derived directly from the physicochemical properties of the NPs, such as: band gap energy (related to probability of electron transfer and catalytic activity of the material), heat of formation of a NP, energies of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO), total energy, solvent accessible surface area, dipole moment, molecular weight, polarizability and dielectric constant [15]. But the activity of a NP can be modified by the medium, which can change its surface charge, surface reactivity, and in the case of biological fluids leads to the formation of a protein corona. Therefore, a set of extrinsic parameters should complement the description. So, to achieve the goal of formulating descriptors these should not only be based on properties of the material, but also extrinsic properties of the NPs in exposure conditions and properties of relevant biomolecules present in the medium, which together are responsible for the interactions and the structure of the bionano interface.

1.2.3 *Biomolecule descriptors*

In contrast to NPs, the development of universal descriptors for biomolecules is relatively straightforward due to their chemical uniformity, e.g. the same aminoacids present in all proteins or nucleic acids in all DNA. For proteins, the simplest descriptors can be constructed using their aminoacid (AA) sequence. These can include counts of AAs of different types, net charge or total mass. Already this characterisation is very rich and capable of predicting complex events at the bionano interface [8,9]. Moreover, obtaining descriptors from AA sequences can be done by using a wide range of software tools such as the EMBOSS PepStats tool [16]. More advanced descriptors for proteins can be built by analyzing their structure. In some cases, starting with the AA sequence of the protein the 3D structure of the

molecule can be retrieved from the Protein Data Bank (PDB) and then used to construct the descriptors. When the structure is not available, one can then use a structure prediction software. There are multiple automated tools available for this task, such as i-Tasser [17]. Using the measured or predicted 3D structure of the protein, several advanced descriptors can be calculated. We previously developed a one-bead-per-aminoacid (united atom – UA) model of globular proteins, which is suitable for this purpose. Some examples of advanced descriptors that can be calculated include protein globule dimensions (radius of gyration and hydrodynamic radius), aspect ratio, dipole moment, rotational inertia, dielectric constant, hydrophobicity, surface charge at different pH and salt concentrations. In addition, protein charge at different pH can be calculated using the Poisson-Boltzmann cell model with charge regulation as reported by da Silva et al. [18].

1.2.4 *Interaction descriptors*

With the known 3D structure of the protein and the nanomaterial, bionano interaction descriptors can be systematically calculated based on how the proteins adsorb onto the surface of the NPs. While a calculation of the precise conformation of adsorbed molecules and a careful evaluation of ensemble averages is definitely a challenging task, several relevant quantities can be calculated using a simplified approach. We here make two major approximations: assume additivity of the interactions between the building blocks of the biomolecule and the NP and neglect the change of conformation for adsorbed molecules. While these assumptions prevent us from obtaining accurate adsorption energies, they create a possibility of a uniform screening of thousands of molecules and ranking them based on how strongly they will attach to the surface of the NP. This ranking is a statistical measure of the content of the biomolecular corona and constitutes a unique fingerprint of a NP. Using the UA model [19], one can compute preferred adsorbed orientation and evaluate mean adsorption energy at different conditions. Moreover, using the same bottom-up construction approach, one can engineer an ultra-coarse-grained model (united AA - UAA) that closely reproduces the total protein-protein pairwise interaction energy profiles obtained in the UA model. In the UAA mode, one typically needs between 5 and 30 united-aminoacid beads to capture the geometry and reproduce the adsorption characteristics of the original protein. This second coarse-graining can be based on the mass distribution in the complete protein and

can be optimized by tuning the protein diffusion coefficients to those obtained using UA model. The UAA model is then suitable for modelling competitive protein adsorption and formation of protein corona. An example of the all-atom, UA and UAA CG models for the same protein is shown in Fig. 1.

1.3 MULTISCALE MODELLING OF BIONANO INTERFACE

1.3.1 *General Methodology*

In order to model the bionano interface and generate descriptors to predict the toxicity of NPs with molecular simulations one must address length scales of up to hundred nanometers and timescales of at least microseconds. Despite the growth of the computational power in the last decade, these length-time scales are not achievable with atomistic detailed simulations, and for this reason coarse-grained (CG) models must be used. Coarse-graining of a molecule consists of reducing the number of degrees of freedom by representing the molecule with CG particles (often called "beads") instead of one particle per atom in the molecule. This reduced representation of the structures of the molecules in the simulation allows one to perform simulations with longer length and time scales than with detailed atomistic models. The main challenge in the development of the CG representation of the molecules and the interactions between the segments of the system is how to reduce the degrees of freedom while ensuring that the reduced representation keeps enough detail to capture the relevant physicochemical processes that occur at the bionano interface. For the task of building CG models there are two main approaches: the top-down approach in which parameters from macroscopic experimental data are used to calibrate the models [20-22] and the bottom-up in which atomic simulations are used as input to generate the CG structures and force-fields [9,23-25]. We refer the reader to the paper of Noid [26] for an overview of these CG techniques.

In the case of bionano interactions, top-down approaches have been used to study the formation of the NP-biomolecule corona [19,27,28] but regardless of their success the parameterisation process relies on the availability of experimental data such as adsorption energies or affinity constants for any protein-surface interaction to be simulated. For practical applications, this constrain is certainly the main drawback as for example a real NP-protein

corona can be composed of hundreds of different proteins. This disadvantage can be overcome by a systematic bottom-up CG strategy in which the basic building blocks of the model are chosen general enough that they can then be used to construct a wide range of molecules. In the case of modelling proteins, the obvious choice is to select the CG beads to be AA. Then, using atomistic simulations of the interaction of amino acids with a surface of interested and techniques such as iterative Boltzmann inversion [29-30] or force matching [21,32] the AA-surface interactions can be parameterised. In this way, the obtained AA-surface CG force fields can be used to model the interaction of any protein present if the structure or sequence of the molecule are known.

1.3.2 *Coarse-Grained Protein Model*

As mentioned previously, once a NP is in contact with a biological medium, a protein corona forms on its surface [33], and that the nature of the corona is what regulates the interaction between the NP and the other biomolecules. It has been extremely challenging to develop a model that could predict the composition of the protein corona around an inorganic NP, as this depends on a multitude of physiochemical properties of both the protein and the NP, such as NP size, shape, pH, hydrophilicity/hydrophobicity, and electrostatic effects.

Computer simulations of the interactions of NPs with proteins can offer a great support to experiments because of their great speed and flexibility [34]. Full-atomistic simulations have already proven to be a valuable tool in elucidating the binding mechanisms of proteins on metallic NPs [35-37]. However, their performance is severely hindered by the inefficiency in simulating systems with large NPs due to the high number of pair interactions that need to be evaluated. To speed-up the calculations, a cut-off of the order of a nanometre is often introduced but this results in an underestimation of the adsorption energies of proteins on NPs mainly because the core of the NP (especially at sizes of over 10 nm) contributes much to their mutual attraction. In this section, we describe a CG model of protein-NP interactions that overcomes most of the challenges in inclusion of the bulk part of NPs in the interaction. In the final section, we test the model by simulating the adsorption of some common proteins such as Human Serum Albumin and Lysozyme on TiO_2 NPs.

Our UA protein model [19], in which every amino acid in the protein is substituted by a single bead whose center is placed at the position of the α-carbon atom, is illustrated in Fig. 1. The aminoacid beads in this model are connected by rigid harmonic springs, if the distance between them is less than 0.7 nm. The springs take care of the molecule shape, so the conformation does not change during the simulation. The model preserves two main structural features that guide the binding mechanisms, i.e., the overall shape of the protein, which is treated as a rigid body, and the overall charge. The super-coarse-grained UAA model that typically contains 5-20 beads presents a further simplification and allows one to model a competitive protein adsorption is also shown in Fig. 1 along with the full-atom and UA presentations.

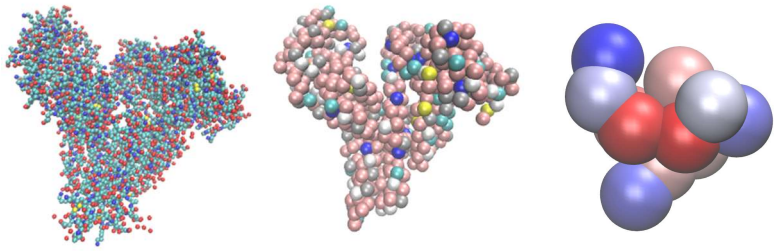

Figure 1 All-atomistic (left), united atom – UA (centre), and united aminoacid (UAA) models (right) of human serum albumin (HSA).

1.3.3 *Coarse-Grained Nanoparticles*

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

Our model starts by considering the contributions that different atoms in the NP give to the binding interaction and propose to partition the NP into two segments – a core and a surface segment. The outer layer on the NP 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 NP 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 $r_c \ll R$ used in the full-atomistic simulation (typically, 1 to 2 nm).Geometrically, the surface segment is a lens formed by an intersection of a sphere of radius r_c, centered on the AA bead, and a sphere of radius of the NP, R, centered on the NP itself (Figure 2).The core comprises 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 NP is then modeled as a single bead of the shape of a sphere of radius with a cut-out surface lens. The potential between the core and the AA beads in our model is calculated using the Lifshitz theory [38] for interaction between two macroscopic bodies. The overall interaction energy between the NP and protein is then estimated by a sum over all AAs in the protein

$$U(D,\theta,\phi) = \sum_{i=1}^{N_{AA}} U_i\big(h_i(D,\theta,\phi)\big) \tag{1.1}$$

where the energy contributions from individual AAs are evaluated as a sum of the interaction potentials with the core and the surface of the NP:

$$U_i\big(h_i(D,\theta,\phi)\big) = (U_s)_i + (U_c)_i \tag{1.2}$$

Here, h_i, D, θ, ϕ are the variables desribing the distance between the AA centre of mass (COM) and the surface of the NP, distance between the protein‘s COM and the NP‘s COM, and orientation of the protein globule, respectively.

In the next two subsections, we describe how the potentials are parameterized for the AA-NP interaction.

1.3.4 *Generation of surface pair potentials*

Here, we assume a pairwise additivity of AA-NP interactions. To calculate the adsorption energy of a whole protein, we need to determine the pair interactions for each AA type with the NP. The potentials of mean force (PMF) for an AA-NP interaction must be first calculated at the atomistic level in

order to capture the surface and protein specific details. These potentials can be obtained using Metadynamics method [39]. Our starting point is the PMF of an AA with a flat slab of the material of interest. Details of how the atomistic PMFs used in our model were calculated, as well as the parameters used to model the slabs, are described in [40]. Note that the PMFs obtained in this way include solvation effects.

Generally, the interaction with a convex surface of a NP of finite radius is less than that for the flat slab due to the lesser number of atoms of the NP 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 cutoff radius r_c used in the calculations as well as the radius R of the NP.

$$U_s(h, R) = U_s(h, R \to \infty) f(h) \tag{1.3}$$

Here, h is the minimum distance between the AA-bead center and the NP surface By taking the appropriate limits for R, it is possible to calculate a correction factor for any geometry. A diagram showing how an AA bead interacts with a CG NP is shown in Figure 2.

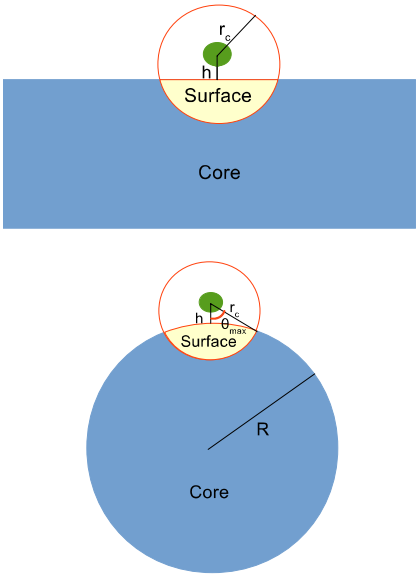

Figure 2 Schematic representation of an AA bead (green sphere) interacting with a slab (left) and with a NP of radius *R* (right) with a cut-off radius of r_c. *h* is the minimum distance between the AA and the NP surface.

We assume that the relevant point-point interaction is dominated by dispersion forces that scales as r^{-6}, where r is the distance between the interaction centers (e.g. atoms).In this case, the attraction energy for a particle to a sphere of radius R, given the finite cut-off r_c can be calculated via the following equations:

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

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

$$U_s(h,R) = -\frac{\pi\varepsilon V}{h+R}\left(\frac{h-2R}{12h^3} + \frac{-6r_c^2+8r_c(h+R)-3h(h+2R)}{12\ r_c^4}\right) \qquad (1.4)$$

Here, ε is the interaction energy per unit volume. For $R \to \infty$, Eq. 1.4 reduces to that of a flat surface:

$$U_s(h, R \to \infty) = \pi\varepsilon V\left(\frac{1}{6h^3} - \frac{2}{3r_c^3} + \frac{h}{2r_c^4}\right) \quad (1.5)$$

and the correction factor is calculated as:

$$f = \frac{U_s(h,R)}{U_s(h,\infty)} = -\frac{r_c^2(h-2R)+2r_c h(h-2R)-3h^2(h+2R)}{2\left(r_c^2+2r_c h+3h^2\right)(h+R)} \quad (1.6)$$

In the equations above, V , is the volume of the AA, which can be represented by a volume of equivalent sphere of radius R_2: $V = \frac{4\pi R_2^3}{3}$, reflecting the size of the respective AA.

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

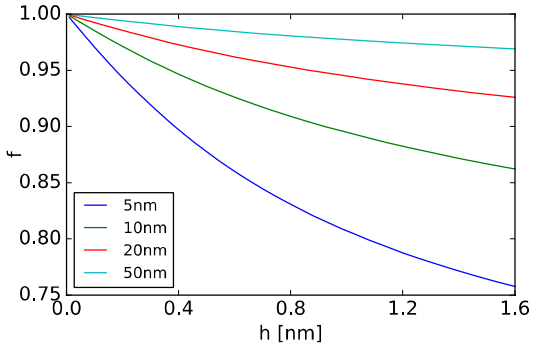

Figure 3 Correction factor f vs. distance from the surface h for a range of NP radii.

1.3.5 *Generation of the core potential*

The NP core plays a crucial role in the protein adsorption as it contains most of the nanomaterial. A serious limitation of the all-atom models is the difficulty of a correct account for the attraction by the core atoms. This problem is mainly due to the short-range cutoff 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 NP at distances $r <$

r_c by treating the remote part of the NP as a single sphere less the part within the cut-off distance from the specific AA. The interaction energy between an AA and the core can be computed using the Hamaker method for dispersion forces [38]. For this, we assume that the spheres representing AAs (sphere 2) are small compared to those representing the NP core (sphere 1) and perform an integration over the volume of AAs first. Then, depending on the distance D between the sphere centers, the interaction energy becomes:

$$U_C(D) = \begin{cases} -\frac{16A_{123}}{9}\frac{R_2^3R^3}{(D^2-R^2)^3}, & R + r_c < D \\ -\frac{4A_{123}R_2^3}{9}\left(\frac{4R^3}{(D^2-R^2)^3} - \frac{8Dr_c - 6r_c^2 - 3D^2 + 3R}{4Dr_c^4} - \frac{3R-D}{4D(R-D)^3}\right), & R < D < R + r_c \end{cases} \quad (1.7)$$

i.e. where the distance to the AA centre exceeds the reach of the PMFs, $R + r_c$, the interaction is just the Hamaker potential for two spheres in the approximation $R_2 \ll R$ while where a part of the NP is covered by the PMF, the van der Waals force on the AA is calculated only for the part of the sphere beyond the cut-off distance. In these expressions A_{123} is the Hamaker constant for interaction between material 1 (e.g. the NP) with material 2 (e.g. the protein) through material 3 (e.g. water), 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:

$$A_{123} = \frac{3}{4}k_BT\frac{(\varepsilon_1-\varepsilon_3)}{(\varepsilon_1+\varepsilon_3)}\frac{(\varepsilon_2-\varepsilon_3)}{(\varepsilon_2+\varepsilon_3)} + \frac{3h\nu_e}{8\sqrt{2}}\frac{(n_1^2-n_3^2)(n_2^2-n_3^2)}{(n_1^2+n_3^2)^{\frac{1}{2}}(n_2^2+n_3^2)^{\frac{1}{2}}\left((n_1^2+n_3^2)^{\frac{1}{2}}+(n_2^2+n_3^2)^{\frac{1}{2}}\right)} \quad (1.8)$$

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 (3×10^{15} s^{-1}) for the and ε_i are the dielectric constants of the materials, which are equal to n^2 in the visible part of the spectrum. In the case one of the materials is a conductor, Eq. 1.8 must be modified to take into account the high values of the polarizability and therefore of the dielectric constant. The equation for a dielectric-conductor interaction in a medium is given by:

$$A_{123} = \frac{3}{8\sqrt{2}}\left(\frac{n_1^2-n_3^2}{n_1^2+n_3^2}\right)\frac{h\sqrt{\nu_1\nu_3}\cdot\nu_2}{\sqrt{\nu_1\nu_3}+\frac{\nu_2}{\sqrt{n_1^2-n_3^2}}} \quad (1.9)$$

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

1.3.6 *Calculation of the adsorption energy*

To evaluate the average adsorption energy for a protein globule on a NP we scan the configurational space (i.e. all possible orientation in which a protein can adsorb on the surface of the NP) by a systematic rotation of the protein and calculate the Boltzmann-averaged energy. There are three degrees of freedom (DOF) that must be scanned [19]. The orientation of the protein 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 $-\theta = z$ the mean interaction energy for any particular orientation for the case of a protein interacting with a spherical NP is given by:

$$E(\phi_i,\theta_j) = -k_B T \ln\left[\frac{3}{\left(R+a(\phi_i,\theta_j)\right)^3-R^3}\int_R^{R+a(\phi_i,\theta_j)} D^2 \exp\left(-\frac{U(D,\phi_i,\theta_j)}{k_B T}\right) dD\right] \quad (1.10)$$

$$E_{ad} = \frac{\sum_i\sum_j P_{ij}E(\phi_i,\theta_j)}{\sum_i\sum_j P_{ij}} \quad (1.11)$$

$E(\phi_i,\theta_j)$ is the total adsorption energy for a fixed set of angles (ϕ_i,θ_j). P_{ij} in eq. 1.11 is given by:

$$P_{ij} = \sin\theta_j \exp\left(-\frac{E(\phi_i,\theta_j)}{k_B T}\right) \quad (1.12)$$

$a(\phi_i,\theta_j)$ in Eq. 1.10 is the maximum interaction distance from the center of mass of the protein to the surface for the given orientation.

1.3.7 *From united-atom to united aminoacid description*

As mentioned above, to study the formation of a realistic protein corona a more aggressive coarse-graining approach is needed which we defined as UAA model. To parameterize the UAA beads, the data output from the UA model described in the previous section is used as a reference. The required output from this UA model consists of a two-dimensional adsorption energy map for the average adsorption energies $E(\phi_i, \theta_j)$.

The generation of the UAA model is done in two steps: (i) a reduced representation of a protein is created based on its all-atom 3D structure, and (ii) the effective bead-bead potentials are derived for the UAA beads and NP beads. In the first part, we used VMD Coarse-Grain Builder (CGB) to create the reduced representation. For this we chose CGB shape-based method, where a neural network learning algorithm is used to determine the placement of neurons (or CG beads). The CG beads have masses correlated to the clusters of atoms which the beads are representing [41].

At the second stage, a pseudo-random piece-wise potential, describing the interaction with the NP, is created for each of these beads. A population of these test potentials is generated and a genetic algorithm is used to minimize the overall difference between the original CG adsorption map from the UA model and the new CG adsorption map. The genetic algorithm minimizes the function:

$$S = \Sigma_{i,j} \left(E_{trial}(\phi_i, \theta_j) - E(\phi_i, \theta_j) \right)^2 \qquad (1.13)$$

Once our CG UAA solution had reached the optimum, we are left with a simplified model of the original protein that has the same (or, at, least similar) characteristic adsorption map as the original UA model.

1.4 APPLICATION OF THE METHOD

1.4.1 *Protein descriptors*

To perform a simulation of adsorption, we use a 3D folded globular structure of the molecule. The 3D structures of the proteins can be obtained from the

Protein Data Bank or calculated using an automated predictor system. Here, we demonstrate how advanced descriptors can be calculated. For illustration, we take the most common plasma protein: Human serum albumin (HSA). The structure of the HSA protein used in this work corresponds to the Protein Data Bank ID 1N5U. We assume physiological conditions, which correspond to a monovalent electrolyte concentration of 100 mM and pH 7. Residue charges at these conditions are $+e$ for LYS and ARG, $-e$ for ASP and GLU, and $+0.5e$ for HIS. The rest of the residues are neutral. Thus, the total charge of the HSA molecule is $-6e$.

Table 1 Descriptors of HSA protein.

$\frac{I_{xx}}{I_{yy}}$	$\frac{I_{yy}}{I_{zz}}$	$\frac{I_{zz}}{I_{xx}}$	Rotational Inertia (g Å^2/mol)	Total Charge (e)	Dipole moment, d_x (e Å)	Dipole moment, d_y (e Å)	Dipole moment, d_z (e Å)	Solvent accessible surface area (nm^2)
4.955	0.3233	0.6240	2.883×10^7	-6.0	386.0	128.0	98.7	315.9

In the above example, the three descriptors are the principle moments of inertia, obtained by calculating the three eigenvalues of the following inertial matrix:

$$\hat{\mathbf{I}} = \begin{bmatrix} \sum_k m_k(y_k^2 + z_k^2) & \sum_k m_k x_k y_k & \sum_k m_k x_k z_k \\ \sum_k m_k x_k y_k & \sum_k m_k(x_k^2 + z_k^2) & \sum_k m_k y_k z_k \\ \sum_k m_k x_k y_z & \sum_k m_k y_k z_k & \sum_k m_k(x_k^2 + y_k^2) \end{bmatrix} \quad (1.14)$$

Here, m_k are the masses of the aminoacids, (x_k, y_k, z_k) are the coordinates of their centers of mass. The eigenvectors now form a new reference system, which is protein specific and independent from the arbitrary axes used within each PDB file. It is along these axes which we calculate the dipole moments, which is simply the sum of all the residue charges by the position vector from the origin along each axis

$$\mathbf{d} = \sum_k q_k \mathbf{r}_k \quad (1.15)$$

The solvent accessible surface area is calculated using the Shrake-Rupley algorithm.

1.4.2 *Bionano-interface descriptors*

The model has been validated for a set of common proteins in blood plasma on TiO_2 NPs of radii ranging from 2.5 nm to 640 nm, see Ref. [19] for details on the validation procedure. We sample protein orientations ϕ from 0 to 355° in steps of 5° and θ from 0 to 175° in steps of 5°. Table 2 summarizes the Hamaker constants for AA-TiO_2 interaction in water calculated from equation 1.8, the UV absorption frequencies for AA are taken from [43] and the refractive indices from the procedure described in [44], the refractive index of TiO_2 was taken to be 2.5.

Table 2 Hamaker constants for AA-TiO_2 interaction through water. All values are reported in units of 10^{-20}J

ARG	HIS	LYS	ASP	GLU	SER	THR	ASN	GLN
5.36	5.87	4.65	5.87	5.22	5.53	4.69	5.74	5.44
CYS	GLY	PRO	ALA	VAL	ILE	LEU	MET	PHE
5.66	5.66	4.37	4.52	4.00	3.96	3.91	5.10	5.61
TYR	TRP							
5.06	6.62							

The potentials generated were then used to calculate the binding energies of a test protein, human serum albumin (HSA, PDB code: 1N5U) on TiO_2 NPs of radii ranging from 2.5 nm to 640 nm, without surface charges. The binding energies were calculated using the ESPResSo package [45] as described in section 1.3.6. In order to evaluate the energy for each binding orientation, the protein was systematically rotated along two angles (ϕ_i, θ_j) around the position vector that joined the protein's center of mass to the center of the NP. Energies were thus evaluated for 2592 different orientations. A more detailed account of the rotation procedure is given in Refs. [19] and [46].

In Table 3, we present a set of descriptors based on the AA-TiO_2 PMFs from Ref. [40].

Table 3 Mean adsorption energies for AA-TiO_2 interaction (uncharged flat surface) through water, calculated using Eq. 1.10. All values are reported in units of k_BT.

ALA	ARG	ASN	ASP	CYS	GLN	GLU
0.327	-0.864	-0.493	-0.197	0.145	-0.696	-0.193
GLY	HIS	ILE	LEU	LYS	MET	PHE
-0.024	-0.156	-0.127	-0.138	0.054	0.000	-0.169
PRO	SER	THR	TRP	TYR	VAL	
-0.040	-0.710	-0.234	-0.871	-1.220	0.107	

Figure 4 below shows the binding energy for one particular orientation of HSA on a 5 nm and 50 nm TiO_2 NP. The contributions from the core and the surface have been separated.

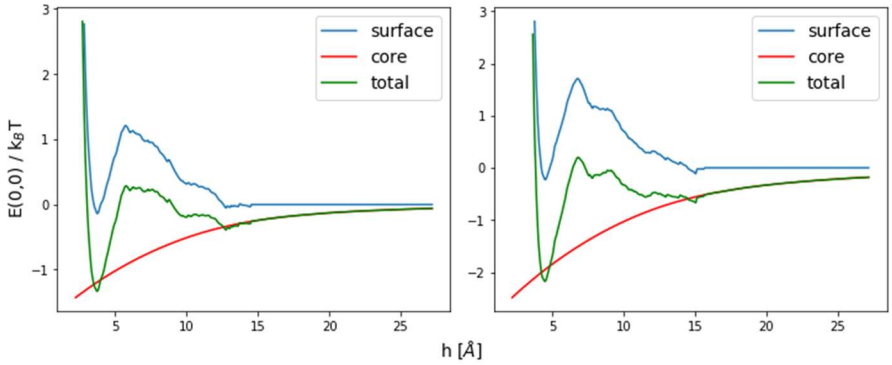

Figure 4 HSA $E(\phi, \theta)$ profile on 5 nm (left) and 50 nm (right) TiO_2 NP. The protein orientation corresponds to the PDB configuration.

The contribution from the NP core is evident in both NP sizes, however it increases as NP radius is increased and for some orientations becomes the dominant contribution. We also see a much-structured short-range part of the potential, which includes a contact minimum at about zero energy

compared to the bulk, a maximum of $1k_BT - 2k_BT$ that corresponds to a hydration layer between the protein and the NP surface, and a much shallower secondary minimum.

Figure 5 shows a binding energy map ($E(\phi_i, \theta_j)$) for HSA on 5 nm and 50 nm TiO_2 particles. We see that the binding energy changes in magnitude with the NP size as the large NP produces deeper minima. Moreover, the map for the 50 nm particle shows more features: in addition to the minima at (230°, 40°) and (340°, 80°) for the 5 nm NP, we find a preferred orientation at (200°, 140°) and another minimum at (70°, 130°).

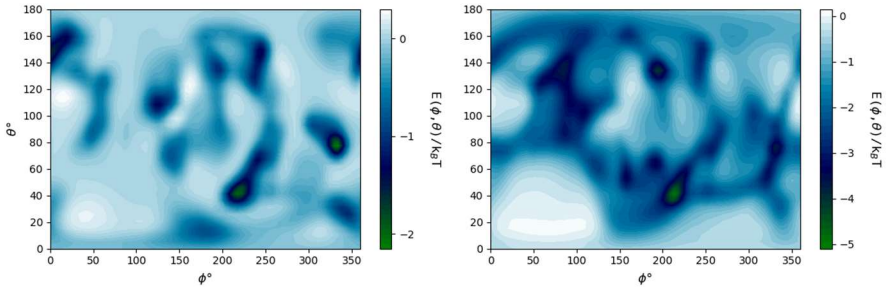

Figure 5 $E(\phi_i, \theta_j)$ map of HSA on the surface of a 5 nm TiO_2 NP (left) and 50 nm TiO_2 NP (right).

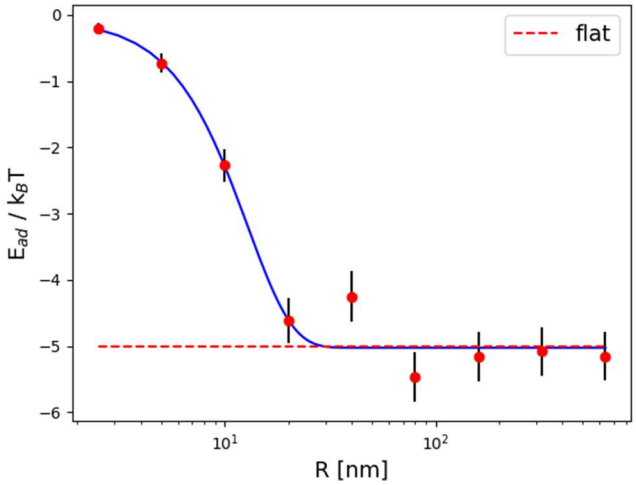

Figure 6 Average E_{ad} for HSA (left) on (100) rutile TiO_2 with increasing NP radius. The dots represent a multiscale calculation according to Eq (1.11), red dashed line is for the flat surface ($R = \infty$). The blue solid line is a fit with $E_{ad} = -5.0\ (1.0 - \mathrm{e}^{aR^2})$

The dependence of the adsorption energy of HSA on the NP size is shown in Fig. 6. The maximum binding energy is about $-5k_BT$, which corresponds to relatively weak adsorption. This means that the HSA molecules, if adsorbed, will be easily displaced by larger ones, according to the Vroman effect. As we can see from the data in Table 3, the contributions to the adsorption energy from most AAs are positive or slightly negative. The largest negative figures come from the charged ARG, and large SER, TYR and TRP. These observations are in agreement with results of direct atomistic and *ab initio* simulations of AA and peptide adsorption on titania surfaces, which confirm that the adsorption is driven by charged residues [47-49]. We also note two other important features of the size dependence of the adsorption energy: the saturation at large NP radii, such that the limiting value corresponds to a flat surface, and the stronger binding at large radii due to the greater van der Waals attraction.

1.4.3 *United-aminoacid model*

Table 4 Ultra-coarse-grained (UAA) model for HSA protein as obtained from VMD CG builder.

Bead	X Position (nm)	Y Position (nm)	Z Position (nm)	Radius (nm)
Bead 1	5.8252	0.9764	2.1574	1.2905
Bead 2	4.0639	0.9955	1.2708	1.1097
Bead 3	0.2843	0.7511	4.7677	1.1667
Bead 4	1.1033	-0.4196	2.1635	1.1767
Bead 5	1.4744	0.4070	-1.2432	1.1042
Bead 6	4.8305	-0.6400	2.6579	1.1598
Bead 7	0.9128	1.3064	1.5903	1.1918
Bead 8	3.2678	1.2892	3.5817	1.1686
Bead 9	2.5456	0.0073	0.6177	1.1631
Bead 10	-0.2788	-0.2642	3.5709	1.1036
Bead 11	2.8617	1.8590	0.0009	1.1923

The UAA model of the HSA protein as obtained from our UA model using the procedure from section 1.3.7 is described in Table 4. The presentation we obtained contains 11 beads of radius of about 1 nm. This model is visualized in Fig. 1. The corresponding interaction potentials for the UAA blocks with a TiO_2 NP are shown in Fig. 7. All the obtained potentials are all smooth and most of them have a single minimum.

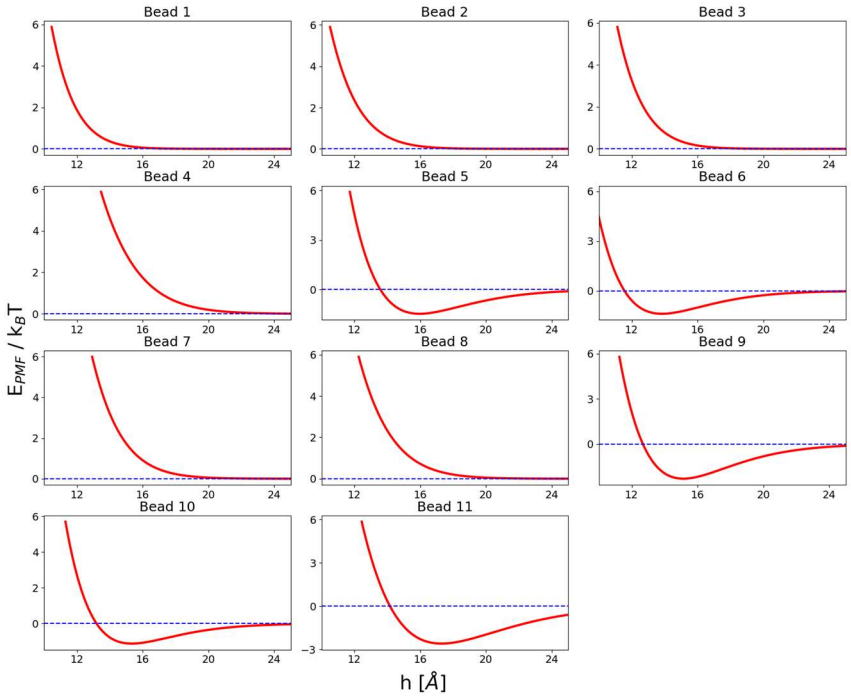

Figure 7 PMFs for beads 1-11 of UAA presentation of HSA protein (see Table 3) with a TiO_2 nanoparticle of size 50 nm.

The UAA presentation preserves the shape of the protein and is sufficient to reproduce the detailed structure of the adsorption energy map. The energy maps obtained within UA and UAA presentations are shown in Figs. 8-9. A comparison of these shows that the UAA models captures all the main the energy minima (corresponding to the preferred orientations of the protein on the NP surface): (230°, 40°) and (340°, 80°) for 5 nm NP, although they become more diffuse (Fig. 8). For 50 nm NP (Fig. 9), the structure of the energy surface in the UAA model is also more diffuse: the two minima at (230°, 40°) and (150°, 50°) are now represented by a single extensive minimum spanning an area of 100°×20°. Although the minima seen in the UAA presentation are more diffuse, they match the magnitude of those in the UA model and ranked in the same order.

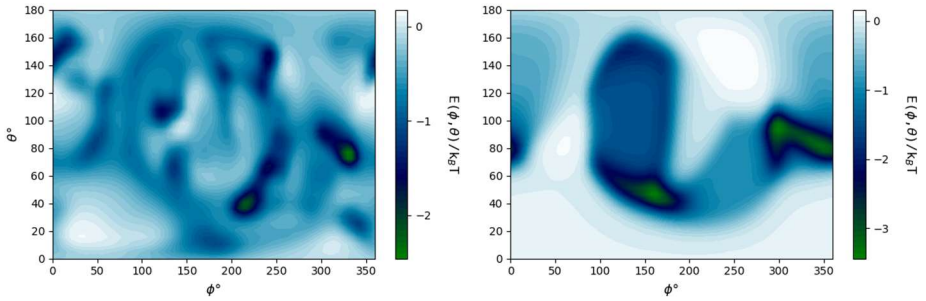

Figure 8 $E(\phi_i, \theta_j)$ map of HSA on the surface of a 5 nm TiO_2 NP from UA model (left) and UAA model (right).

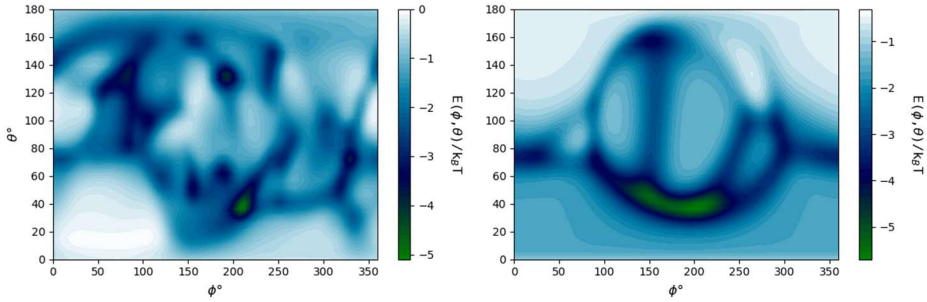

Figure 9 $E(\phi_i, \theta_j)$ map of HSA on the surface of a 50 nm TiO_2 NP from UA model (left) and UAA model (right).

1.5 CONCLUSIONS

We presented a universal methodology of evaluation of advanced descriptors of bionano interface, which can be used to characterise the interaction of a specific NP with essential biomolecules constituting the protein corona of the NP. Although we made a number of strong approximations on the way, the main benefit we achieved is that the calculation can be applied in a uniform fashion to a large number of biomolecules in a short time and to add more descriptors to elucidate further the mechanism of corona formation and its dependence on the protein's and NP's physicochemical properties. The new descriptors include those for proteins (principal moments of inertia, charge, dipole moment and the solvent accessible area) and those for interaction with NP (Hamaker constants for residues, their mean adsorption energies, and the overall adsorption energy for the protein globule). In addition to this, our method

delivers preferred orientations for proteins on a specific NP, thus systematically taking into account the NP shape and size, and allows one to rank arbitrary proteins by the adsorption affinity. Thus, dozens of essential descriptors can be routinely calculated. Another achievement of our methodology is that the core of the NP is included in the calculation, which makes the size-dependence of the interaction more credible. These descriptors can be used to produce interaction fingerprints for arbitrary nanomaterials with respect to specified interactions and, therefore, provide key information for gauging their biological activity, e.g. its ability to produce the molecular initiating events or disturb the key events of the adverse outcome pathways.

ACKNOWLEDGEMENTS

This research was funded by EU FP7 grant No. 310465 (MembraneNanoPart project) and EU H2020 grant No. 686098 (SmartNanoTox project).

REFERENCES

1. T. Cedervall, I. Lynch, M. Foy, T. Berggad, S. Donnelly, G. Cagney, S. Linse, K. Dawson, “Detailed identification of plasma proteins adsorbed on copolymer nanoparticles”, Angew. Chem. Int. Ed., 46, 5754–5756 (2007)
2. R. Landsiedel R, L. Ma-Hock, A. Kroll, D. Hahn, J. Schnekenburger, K. Wiench, W. Wohlleben, “Testing Metal-Oxide Nanomaterials for Human Safety”, Adv. Mater. 22, 2601–2627 (2010)
3. M, Lundqvist, I. Sethson, B. H. Jonsson, “Protein Adsorption onto Silica Nanoparticles: Conformational changes depend on the particles curvature and the protein stability”, Langmuir 20, 10639–10647 (2004).
4. M. Lundqvist, J. Sethson, B. H. Jonsson, “Transient interaction with nanoparticles “freezes” a protein in an ensemble of metastable near-native conformations”. Biochemistry (Mosc) 44, 10093–10099 (2005)
5. S. S. Karajanagi, A. A. Vertegel, R. S. Kane, J. S. Dordick, “Structure and function of enzymes adsorbed onto single-walled carbon nanotubes”, Langmuir 20, 11594–11599 (2004).
6. M. A. Dobrovolskaia, A. K. Patri, J. Zheng, J. D. Clogston, N. Ayub, P. Aggarwal, B. W. Neun, J. B. Hall, S. E. McNeil, “Interaction of colloidal gold nanoparticles with human blood: effects on particle size and analysis of plasma protein binding profiles”, Nanomedicine 5, 106–117 (2009).
7. M. A. Dobrovolskaia, B. W. Neun, S. Man, X. Ye, M. Hansen, A. K. Patri, R. M. Crist, “Protein corona composition does not accurately predict hematocompatibility of colloidal gold nanoparticles”, Nanomedicine: Nanotechnology, Biology and Medicine 10, 1453-1463 (2014)
8. C. D. Walkey, J. B. Olsen, F. Song, R. Liu, H. Guo, D. W. H. Olsen, Y. Cohen, A. Emili, W. C. W.

Chan, “Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles”, ACS Nano 8, 2439–2455 (2014)
9. P. Kamath, A. Fernandez, F. Giralt, and R. Rallo, “Predicting Cell Association of Surface-Modified Nanoparticles Using Protein Corona Structure - Activity Relationships (PCSAR)” Curr. Top. Med. Chem. 15, 1930 (2015).
10. S. R. Saptarshi, A. Duschl and A. L. Lopata, “Interaction of nanoparticles with proteins: relation to bio-reactivity of the nanoparticle”, J Nanobiotechnology 11, 26 (2013).
11. M.A. Dobrovolskaia, M. Shurin, A.A. Shvedova, “Current understanding of interactions between nanoparticles and the immune system”, Toxicol. App. Pharmacol., 299, 78-89 (2016)
12. T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H. M. Hwang, A. Toropov, D. Leszczynski, J. Leszczynski, “Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles”, Nature Nanotechnology 6, 175–178 (2011)
13. A. Gajewicz, N. Schaeublin, B. Rasulev, S. Hussain, D. Leszczynska, T. Puzyn, J. Leszczynski, “Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: hints from nano-QSAR studies”, Nanotoxicology 9, 313-25 (2015)
14. H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang ,Y. P. Liao , M. Wang , L. Li , R. Rallo, R. Damoiseaux, D. Telesca, L. Mädler, Y. Cohen, J. I. Zink, A. E. Nel, “Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation”, ACS Nano, 6, 4349-68 (2103)
15. EU US Roadmap Nanoinformatics 2030, doi: 10.5281/zenodo.1486011
16. P. Rice, I. Longden, A. Bleasby, “EMBOSS: The European Molecular Biology Open Software Suite”, Trends Genet., 16, 276-277 (2000)
17. A. Roy, A. Kucukural, Y. Zhang, “I-TASSER: a unified platform for automated protein structure and function prediction”, Nat. Protoc., 725-738 (2010)
18. F. L. B. da Silva, M. Boström, C. Persson “The effect of charge regulation and ion-dipole interactions on the selectivity of protein-nanoparticle binding”, Langmuir, 30, 4078-4083 (2014).
19. Power, D., Rouse, I., Poggio, S., Brandt, E., Lopez, H., Lyubartsev, A. and Lobaskin, V. “A multiscale model of protein adsorption on a nanoparticle surface”, Modelling Simul. Mater. Sci. Eng., 27, 084003 (2019).
20. M. Monopoli, C. Aberg, A. Salvati, and K. A. Dawson, “Biomolecular coronas provide the biological identity of nanosized materials” Nat. Nanotechnol., 7, 779 (2012).
21. I. Lynch, K. A. Dawson, and S. Linse, “Detecting cryptic epitopes created by nanoparticles”, Sci. Signaling, 2006, pe14
22. T. Cedervall, I. Lynch, S. Lindman, T. Berggard, E. Thulin, H. Nilsson, K. A. Dawson, and S. Linse, “Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles”, Proc. Natl. Acad. Sci. U. S. A. 104, 2050 (2007).
23. S. Lindman, I. Lynch, E. Thulin, H. Nilsson, K. A. Dawson, and S. Linse, “Systematic investigation of the thermodynamics of HSA adsorption to N-iso-Propylacrylamide/N-tert-Butylacrylamide copolymer nanoparticles. effects of particle size and hydrophobicity” Nano Lett. 7, 914 (2007).
24. L. Allen, M. Tosetto, I. Miller, D. O’Connor, S. Penney, I. Lynch, A. Keenan, S. Pennington, K. Dawson, W. Gallagher, “Surface-induced changes in protein adsorption and implications for cellular phenotypic responses to surface interaction”, Biomaterials 27, 3096 (2006).
25. O. V. Salata, “Applications of nanoparticles in biology and medicine”, J. Nanobiotechnol., 2, 3 (2004)
26. W. G. Noid, “Coarse-grained models for biomolecular systems”, J. Chem. Phys. 139, 090901 (2013)
27. P. Vilaseca, K. A. Dawson, G. Franzese, “Understanding and modulating the competitive surface-adsorption of proteins through coarse-grained molecular dynamics

simulations", Soft Matter, 9, 6978-6985 (2013)
28. O. Vilanova, J. J. Mittag, P. M. Kelly, S. Milani, K. A. Dawson, J. O. Rädler, G. Franzese, "Understanding the Kinetics of Protein–Nanoparticle Corona Formation", ACS Nano, 10, 10842-10850 (2016)
29. A. P. Lyubartsev, A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach", Phys. Rev. E 52, 3730 (1995)
30. D. Reith, M. Pütz, F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations", J. Comput. Chem., 24, 1624 –1636 (2003)
31. S. Izvekov, G. A. Voth, "A Multiscale Coarse-Graining Method for Biomolecular Systems", J. Phys. Chem. B, 109, 2469–2473 (2005)
32. W. G. Noid, J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models", J. Chem. Phys., 128, 244114 (2008)
33. M. Rahman, S. Laurent, N. Tawil, L. Yahia, and M. Mahmoudi, "Protein-Nanoparticle Interactions", volume 15 of Springer Series in Biophysics. Springer-Verlag (2013).
34. D. An, J. Su, C. Li, and J. Li, "Computational studies on the interactions of nanomaterials with proteins and their impact", Chin. Phys. B, 24, 120504 (2015).
35. G. Brancolini, D. B. Kokh, L. Calzolai, R. Wade, and S. Corni, "Docking of Ubiquitin to Gold Nanoparticles", ACS Nano 6, 9863-9878 (2012).
36. F. Ding, S. Radic, R. Chen, P. Chen, N. Geitner, J. Brown, and P. Ke, "Direct observation of a single nanoparticle–ubiquitin corona formation", Nanoscale 5, 9162-9169 (2013)
37. S. Khan, A. Gupta, and C. Nandi, "Controlling the fate of protein corona by tuning surface properties of nanoparticles", J Phys Chem Lett. 4, 3747-3752 (2013).
38. J. N. Israelachvili, "Intermolecular and Surface Forces", AP, 3rd edition, (2011)
39. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J., J. Wang, and Peter Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations", J. Comput. Chem, 24, 1999-2012 (2003).
40. E. Brandt and A. P. Lyubartsev, "Molecular dynamics simulations of adsorption of amino acid side chain analogues and a titanium binding peptide on the TiO_2 (100) surface", J. Phys. Chem. C, 119, 18126-18139 (2015).
41. D. Kokh, S. Corni, P. Winn, M, Hoefling, K. Gottschalk, R. Wade, "ProMetCS: an atomistic force field for modelling protein-metal surface interactions in a continuum aqueous solvent", J. Chem. Theory Comput., 6, 1753-1768 (2010)
42. A. Y. Shih, P. L. Freddolino, A. Arkhipov, K. Schulten, "Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations", J. Struct. Biol, 157, 579-592 (2007)
43. J. A. Glasel, "Validity of nucleic acid purities monitored by a260/a280 absorbance ratios" Biotechniques, 18, 62-63 (1995).
44. H. Zhao, P.H. Brown, and P. Schuck, "On the distribution of protein ractive index increments", Biophys. J., 100, 2309-2317 (2011).
45. H. Limbach, A. Arnold, B. Mann, C. Holm, "ESPResSo - An Extensible Simulation Package for Research on Soft Matter Systems", Comput. Phys. Commun., 174, 704 (2006).
46. Y. Sun, W. Welsh, and R. Latour, "Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation", Langmuir, 21, 5616-5626 (2005)
47. G. Raffaini, F. Ganazzoli, "Molecular modelling of protein adsorption on the surface of titanium dioxide polymorphs", Phil. Trans. R. Soc. A 370, 1444–1462 (2012)
48. M. Polimeni, L. Petridis, J. C. Smith, and C. Arcangeli, !Dynamics at a Peptide–TiO 2 Anatase (101) Interface", J. Phys. Chem. B 121, 8869–8877 (2017)
49. S. Liu, X.-Y. Meng, J. M. Perez-Aguilar, and R. Zhoua, "An In Silico study of TiO2 nanoparticles interaction with twenty standard amino acids in aqueous solution", Sci Rep. 6, 37761 (2016).