

# NJC

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

## A closer look into the Ubiquitin Corona on Gold Nanoparticle by computational studies

*Francesco Tavanti<sup>a</sup>, Alfonso Pedone<sup>a</sup> and Maria Cristina Menziani<sup>a</sup>*

KEYWORDS: Ubiquitin, Protein Corona, Gold nanoparticle, Computational simulation, Coarse-Grained.

In this study, coarse-grained computational simulations of ubiquitins corona around gold nanoparticle have been carried out, and the effect of the nanoparticle size (10, 16, 20, and 24 nm diameter) and environment (bare nanoparticle surface, and citrate-coated surface, where citrate are treated with implicit and explicit models) has been analysed. The results showed that the corona is obtained after a slow reorientation step that occurs at the nanoparticle surface in order to optimize the nanoparticle-ubiquitins interaction. The ubiquitin binding modalities depend on the nanoparticle environment, while ubiquitins conformational changes upon binding and their aggregation propensity slightly depends on nanoparticle size.

## 1 Introduction

It has been long recognized that the contact of NPs with biological fluids rapidly results in the formation of a protein corona on the NP surface, whose composition changes over time<sup>1-8</sup>. A “soft corona” that reflects the relative abundance of individual medium proteins, is immediately formed. Then, protein-protein competition takes place, and displacement of low-affinity by high-affinity proteins occurs, at the origin of the “hard corona” (known as the Vroman effect). In their recent review, Walkey and Chan<sup>9</sup> showed that the plasma protein corona follows a general structure, with 2 to 6 proteins adsorbed at high abundance, and many more adsorbed at low abundance; moreover, they identify the family of 125 unique plasma proteins that have been associated with forming NP biocorona at many different nanomaterials. Therefore, the composition, structure, dynamics and stability of the protein corona (rather than the bare NP surface) control the interaction of the NP with the cell and the subsequent biological response of an organism to NP exposure<sup>1,9</sup>.

The characteristics of the protein corona are determined by the physiological environment, duration of exposure, and compatibility of the physico-chemical properties of the interacting proteins with those of the NP surface. Targeted NP usage in nanobiology, nanomedicine, and nanotoxicology could then be improved by manipulating the surface properties of NPs to bind proteins selectively in order to control signalling, kinetics, transport, accumulation, and toxicity<sup>10</sup>. However, the NP itself may alter the structure of the adsorbed proteins, leading to denaturation or significant conformational changes, with concomitant loss of their biological function and hazardous consequences<sup>11,12,13</sup>.

Despite the ever-increasing number of experimental studies, dedicated to uncovering the detailed relationships between the synthetic identity, biological identity, and physiological response of NP-protein complexes, a comprehensive picture is still missing due to the complexity inherent in the systems and the experimental dilemma of measuring without changing the nature of the original protein corona<sup>2</sup>.

Thus, the opportunity of assisting the experimental studies via computational simulations is of great importance in most nanotechnology applications, as demonstrated by the increasing number of studies, which are appearing in the literature<sup>3</sup>.

In this paper, the results of Coarse-Grained (CG) computer simulations carried out to gain insight into the gold NP-ubiquitin corona formation will be described.

Human ubiquitin (Ubq) is a convenient target for the analysis of its interaction with NP at the atomic level since it is a small protein composed of 76 amino acids, folded up into a compact globular structure comprising a long  $\alpha$ -helix, a short piece of 3(10)-helix, and a mixed  $\beta$ -sheet with five strands. Its three-dimensional structure is well characterized both in its crystallized form and in solution<sup>14,15</sup>. This highly conserved protein is found only in eukaryotic organisms, where it exerts a regulatory role: proteins that are to be degraded are first tagged by conjugating them with Ubq, then recognized and shuttled to the proteasome for degradation.

Experimental and computational data on ubiquitin corona on gold or silver NPs have become recently available in the literature. Unfortunately, the results obtained are often contradictory and can only be used for comparison (not for validation) of the CG approach used in the present paper.

Recently, Ding et al.<sup>16</sup>, studied the interaction between a number of Ubq molecules and the surface of a silver NP by means of a combined experimental and multiscale molecular dynamics simulations approach. Their results uncovered: a) a specific binding orientation between Ubq molecules and AgNP surfaces, driven by electrostatic interactions and regulated by a stretched

exponential kinetics, b) competition of Ubq molecules and citrate for binding to the NP surface, and c) a loss of  $\alpha$ -helical structure upon adsorption of Ubq onto the silver surfaces. The importance of taking into account citrate molecules for reproducing experimental results was pointed out also by Brancolini et al.,<sup>17</sup> who studied the adsorption of one Ubq molecule on a flat gold surface by means of a combination of simulation methods at different levels of theory (including Brownian dynamics, classical atomistic MD and quantum mechanical DFT calculations). They found that the leading terms for the encounter complex are short-range, non-electrostatic interactions (besides binding to citrate). All their reasoning on the binding orientations of Ubq on NP were based on the study by Calzolari et al.<sup>18</sup> who observed fast exchange ( $\sim 10$ – $100$   $\mu$ s) between the adsorbed and free protein states of Ubq. Perturbed chemical shifts associated to a few amino acids were considered to arise from the interaction of the protein with AuNP. Unfortunately, a subsequent study was not able to reproduce these observations, despite working at similar experimental conditions and with AuNPs of similar size (12 nm vs 15 nm in diameter)<sup>19</sup>. Moreover, experimental studies very recently carried out by Mangini et al.<sup>20</sup> on Ubq interacting with AgNP ( $\sim 10$ nm) added an additional layer of complexity. In fact, they invoked amyloid transition to explain small changes in the protein NMR spectra instead of protein-NP interaction. Moreover, they hypothesized that unlike citrate-coated NPs, naked NPs induce amyloid transition (i. e. after the protein corona formation on naked NPs, clustering into large aggregates held together by the proteins is observed).

In order to rationalize the results of these *in vitro* studies,<sup>16,18-20</sup> and to obtain a better understanding of Ubq-corona formation, the interaction between a large number of Ubqs and gold spherical NPs will be studied using molecular dynamics simulations. The effect of NP size (10, 16, 20, and 24 nm diameter) and environment (bare NP surface, and citrate-coated surface, where citrate are treated with implicit and explicit models) will be discussed. Due to the large number of atoms involved in this process, a CG model will be used to allow timescales of tenth of ns to be reached at a relatively small computational cost.

## 2 Methods

### 2.1 Coarse-Grained models of the system

Four gold NPs of diameter 10, 16, 20, and 24 nm were built by means of CG models. Each model consists of a neutral central bead associated to a steric repulsive sphere, representing the NP surface, on which a number of positively charged beads (250, 400, 500 and 600 beads with positive charge (+e), depending on the size of the NP) was randomly positioned.<sup>21</sup> A uniform density of beads on the AuNP surface was also ensured.<sup>22,23</sup>

The gold NPs were coated with citrate molecules modelled both implicitly and explicitly. In the case of implicit citrates 250, 400, 500 and 600 beads with negative charge (-e) were used for the NPs of 10nm, 16nm, 20nm and 24nm diameter, respectively. The procedure used in the case of explicit citrates, was as follow:

1. 42, 107, 167, and 240 charged beads (+3e) were assigned to the bare NP of 10nm, 16nm, 20nm, and 24nm diameter, in order to maintain the same density of charge for the AuNPs of different diameters, as in the implicit citrate model.<sup>21</sup>
2. each bare NP was surrounded by 300 citrates beads placed with a random distribution at 5nm from the surface. A containing sphere was used to avoid evaporation. The citrates were modeled with a single negative charged bead (-3e);

3. a 300K simulation run was performed for 2ns in order to obtain the maximum number of citrates adsorbed on the NP surface. In particular, the citrates beads adsorbed are 45, 119, 144 and 257 for the 10nm, 16nm, 20nm and 24nm diameter NP respectively;
4. all non-adsorbed citrates were then removed from the simulation box.

The crystal structure of Ubq solved at 1.8 Å resolution was retrieved from the Protein Data Bank (PDB) (PDB id: 1UBQ), and was used for the all-atom simulation. The Gō-type CG model of Ubq molecules consisted of a single bead per amino acid residue centered at the α-carbon position. Thus, a linear chain of beads characterized by bonds with distance fixed at 3.8Å, and angles and dihedral angles that reproduce the Ubq secondary structure was obtained. The Gō-type model<sup>24</sup> used showed to furnish good performance in previous studies<sup>25,26,27,28</sup>.

For both all-atom and CG simulations all treatable protein residues were assigned their standard protonation state at pH 7.

For each simulation run carried out with NP of different diameters (bare NP and citrate-coated NP, with citrate ions treated implicitly and explicitly), 48 Ubq molecules were randomly distributed 5nm far from the NP surface. A test with 83 Ubq molecules was also carried out for the 10nm AuNP diameter with implicit citrates model. The system was then soaked in a containing sphere of 60nm diameter in order to ensure that proteins interact with the NP, avoiding evaporation during dynamics.

## 2.2 Force Field

The functional form of the Force Field applied to the Ubq molecules represents all possible interactions among beads, and comprises both short-term ( $U_{\text{bonded}}$ ) and long-term ( $U_{\text{non-bonded}}$ ) interactions:

$$U = \sum U_{\text{bonded}} + \sum U_{\text{non-bonded}} \quad (1)$$

where :

$$\begin{aligned} \sum U_{\text{bonded}} = & \sum_{i,j} U_{\text{bonds}}(i,j) + \\ & \sum_{i,j,k} U_{\text{angles}}(i,j,k) + \sum_{i,j,k,l} U_{\text{dihedrals}}(i,j,k,l) \end{aligned} \quad (2)$$

$$\begin{aligned} \sum U_{\text{non-bonded}} = & \sum_{i,j} U_{\text{non-bonded}}^{\text{local}}(i,j) \\ & + \sum_{i,j} U_{\text{non-bonded}}^{\text{non-local}}(i,j) \end{aligned} \quad (3)$$

Bonds, angles and dihedrals describe links between two, three and four consecutive amino acids respectively. These terms are responsible of the secondary structure of the protein. Non bonded terms describe both local, i.e. short-range, and non-local, i.e. long-range, interactions. Local non-bonded interactions are calculated between beads that are not linked by a bond, an angle or a

dihedral angle and that are closer than the cutoff of 8.5Å. These interactions are stronger and more anisotropic than non-local non bonded interactions and they are based on the native reference structure, giving a bias to the model. Non-local non-bonded interactions, given by VdW and electrostatic, describe unbiased interactions between amino acids. For these interactions the cutoff used is the one reported in paragraph 2.4.

Protein FF parameters are reported in ref.s<sup>25,26,27</sup>. The FF for citrate interactions is adapted from the work of Ding et al.<sup>16</sup>. All these parameters are summarized in Table 1.

**Table 1.** Analytical expressions of the force field used and associates parameters.

Potential	Type	Formula	Parameters
<b>Ubiquitin</b>			
$U_{\text{bonds}}$	constraints		$r_0 = 3.8\text{\AA}$
$U_{\text{angles}}$	harmonic	$\frac{1}{2}k_{\theta}(\theta - \theta_0)^2$	$k_{\theta} = 20 \text{ Kcal/mol}$
$U_{\text{dihedrals}}$	cosine	$k_{\phi}[1 - \cos(m(\phi - \phi_0))]$	$k_{\phi} = 100 \text{ Kcal/mol}$
$U_{\text{local}}$	12-6	$\left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^6}\right)$	$A = R_{ij}^{12} *$ $B = 2R_{ij}^6 *$
$U_{\text{non-local}}$	12-6	$\left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^6}\right)$	$A = 3.8^{12}$ $B = 0$
<b>Citrates</b>			
$U_{\text{non-local}}$	12-6	$\left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^6}\right)$	$A = 3.8^{12}$ $B = 0$

\* R is the distance between two beads in the reference structure

The Ubq-gold NP-citrate interactions are ruled by the CG Hamiltonian developed by Li et al.<sup>21</sup>, which consists of the exclude volume (VdW), and electrostatic terms. The VdW term is:

$$U_{\text{VdW}} = \sum_{i,j} \left( \frac{\sigma}{r_{ij}} \right)^{12} \quad (4)$$

where  $\sigma$  is the reference radius set to 3.8Å, and  $r_{ij}$  is the distance between two beads. The Electrostatic term is computed by the Debye-Hückle potential

$$U_{\text{elec}} = \sum_{i,j} \frac{z_i z_j e^2}{4\pi\epsilon_0 \epsilon_r r_{ij}} e^{-\frac{r_{ij}}{l_D}} \quad (5)$$

Where  $z_i$  and  $z_j$  are the bead charges involved in the interaction,  $r_{ij}$  is the distance between them, and  $l_D$  is the Debye length:  $l_D = (8\pi l_b I)^{-\frac{1}{2}}$  for monovalent cations. Here,  $l_b \approx 7 \text{ \AA}$  is the Bjerrum length at room temperature, the ionic strength,  $I$ , is 15 mM, and the dielectric constant,  $\epsilon$ , is 10.

### 2.3 Mixed All-Atom/Coarse-Grained model of the system

In order to evaluate the effect of a CG treatment of Ubq, an all-atom simulation was also carried out. The system comprises one Ubq (all-atom), a 20nm diameter gold NP (CG), and implicit citrates. The Amber FF<sup>29</sup> was used for Ubq; the ProMetCS atomistic force field, described by Kokh et al.,<sup>30</sup> was used for Ubq-AuNP VdW interactions. Interactions between each Ubq atom and the AuNP are codified by a 12-6 potential parametrized by Iori et al.<sup>31</sup>.

### 2.4 Molecular Dynamics Simulations

All simulations were performed with the DL\_POLY\_2.20 software<sup>32</sup>. Classical MD simulations were performed at constant volume and temperature ( $T = 300 \text{ K}$ ) by using the Berendsen thermostat ( $\tau=1\text{fs}$ ). The cutoff radius used for both electrostatic and VdW interactions is set to the NP radius plus  $10\text{\AA}$  in order to make proteins and citrates interact with the repulsive sphere. The timestep used is 1fs and trajectories are 10ns long. In all simulations, the Au-NP center was kept fixed during the simulations and gold charged beads were frozen in their starting position.

Three independent runs were performed for each MD simulation, each starting with different and random initial velocities.

### 2.5 Contact probability calculation

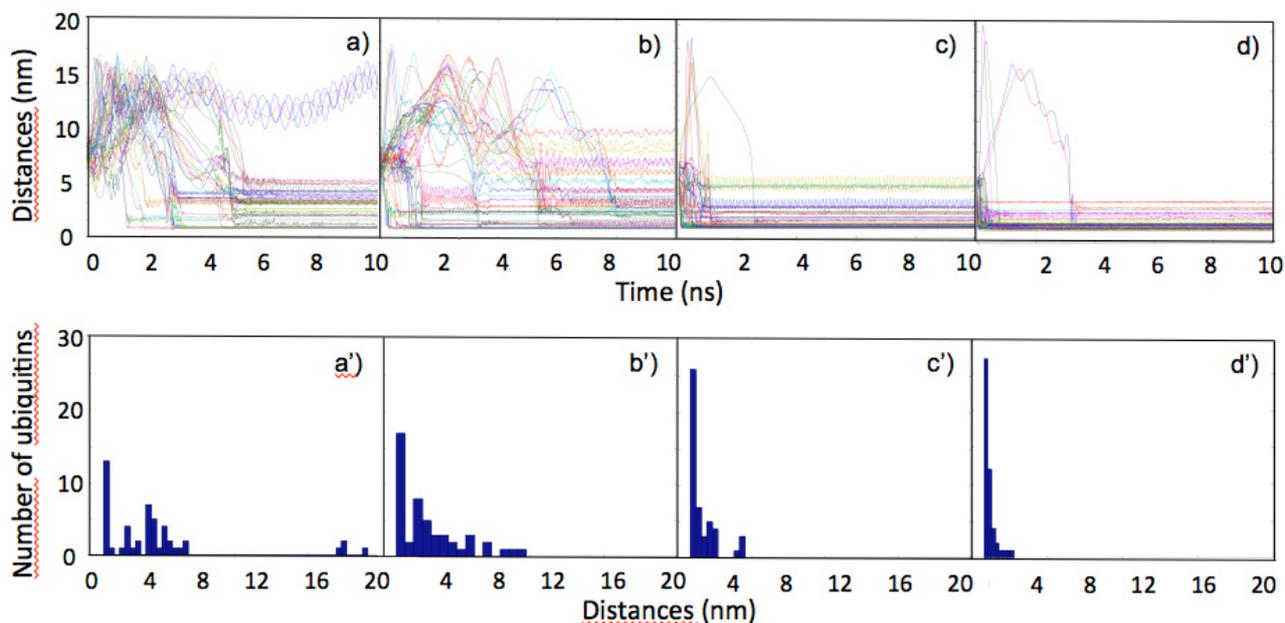
The contact probability for each amino acid residue is defined as the ratio of the amount of time in which the distance of a given amino acid residue of the protein X from the NP surface is less than a given cutoff ( $6.5\text{\AA}$ ), and the amount of time in which the protein X remains adsorbed on the gold NP surface. These values are normalized in order to obtain the sum of all probabilities equals to 1. A protein is considered adsorbed on the gold NP surface if its center of mass is closer than  $17.5\text{\AA}$  from the NP surface. This value is given by the distance of the center of mass from the Ubq surface plus the bonding distance, assumed to be  $7.5\text{\AA}$ . The choice of using the center of mass instead of calculating each residue-surface distance is justified by the spherical shape of Ubq.

## 3 Results and Discussion

The results obtained by a CG computational simulation study of 48 Ubq molecules in interaction with gold spherical NPs (10, 16, 20, and 24 nm diameter), in different environment (bare nanoparticle surface, and citrate-coated surface, treated as implicit and explicit citrate models) are here discussed in terms of: Ubq re-orientation upon corona formation, NP adsorption capacity, Ubq binding site, Ubq structural changes upon NP binding, and Ubq clustering.

### 3.1 Corona formation and Ubiquitin re-orientation

The variation of the distances of each Ubq protein from the AuNPs as a function of time can be used to monitor the formation of the corona. The results obtained for the 10, 16, 20, and 24 nm NPs coated with explicit citrates are reported in Fig. 1. These can be considered representative of the results of computational simulations on the bare and implicit citrate-coated NP. In all the computational simulations carried out, a consistent number of proteins (from 23% to 96%, depending on the size of the NP) is able to make persistent interactions with the NP already after few nanoseconds. The free space available in the simulation box for protein migration (see paragraph 2.1) and the aggregation phenomena which will be discussed in paragraph 3.5 are mainly responsible for the different behavior observed for small (10, and 16 nm) and big (20, 24 nm) NPs. Overall, a first layer of proteins of  $\sim 2$  nm (corona) is formed on the AuNP surface, as also visualized by the first peaks in Fig. 1, panels a') - d'). A smaller number of proteins (from 10 to 30%) interacts with the protein corona forming a second layer, at approximately 4 nm, whereas only a few proteins remain free to move around the AuNP during the whole simulation, and are occasionally organized into clusters (Fig. 1 panel a')).



**Fig. 1** Distances between the centers of mass of each Ubq and the surface of the AuNPs versus the computational simulation time; panels a) - d) refer to NPs of 10, 16, 20, and 24, nm diameter, respectively. Simulations with explicit citrate-coated NPs are considered. Histograms of the distances showing the distribution of Ubq proteins around the NP of different diameters (panels a') - d')). The first peaks show the corona layer. Other peaks refer to Ubqs not directly in contact with the AuNP.

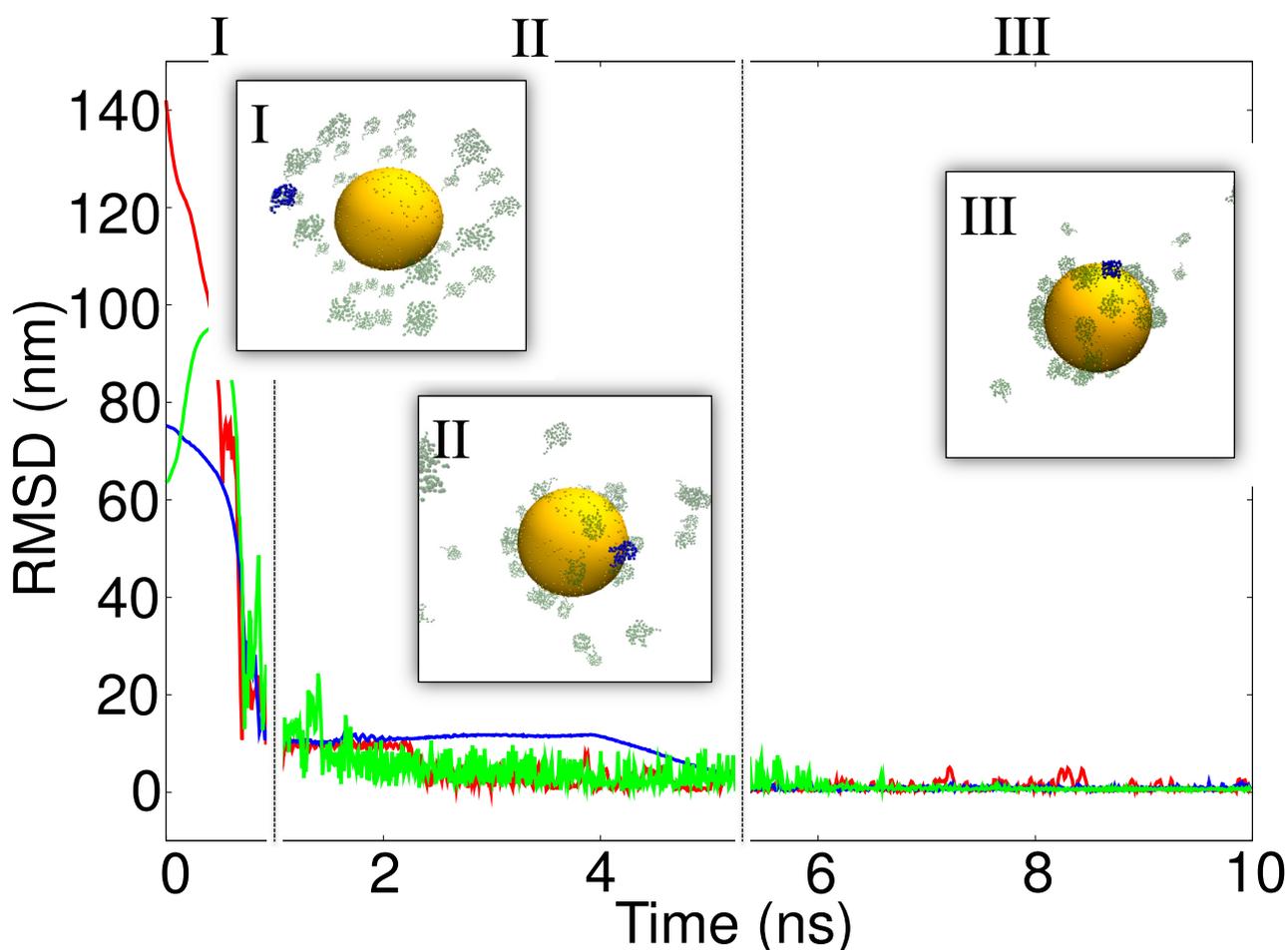
The size of the NP did not affect the corona thickness in a relevant manner, in fact the diameter of free Ubq is around 2.5 nm (See Fig. S1 in ESI). Although, this result could be due to the relative small number of Ubqs used in the computational simulations (see paragraph 3.2), it is in line with the experimental findings on Au and AgNPs of comparable sizes<sup>18,19</sup>.

More information on the modalities of the corona constitution can be derived from the analysis of the trajectories of the proteins during dynamic simulations. The Root Mean Square Deviation (RMSD) of the center of mass of one Ubq, representative of the protein ensemble, is reported in Fig. 2 for the 20nm AuNP embedded in the different environments. During the first 0.8ns of the

simulation (Fig. 2) the Ubq moved around the NP without interacting with it and large changes in RMSD are observed. In the range 0.8-2.5 ns the protein is trapped by the electrostatic potential generated by the NP, the RMSD fluctuations decrease and Ubq moves on the NP surface. The presence of citrates treated in an explicit manner extends this period to ~5 ns. Finally, the Ubq optimal interaction site for binding is found, and the corona is formed.

It can be noticed that the absence of citrates (Fig. 2, green line) slightly accelerates the reorientation process due to the only presence positive charges on the NP surfaces. On the contrary, the explicit citrates retard the reorientation process. In fact, from 9% to 44% citrates move from the surface of the NP and bind to the Ubq positive charged areas (formed by R42 and R72, and, in a lesser extent, K6 and K11). Moreover, Ubqs bind both to the AuNP and to citrates anchored on the surface, as already pointed out by Brancolini et al.<sup>17</sup> in their study of Ubqs binding to an Au flat surface.

These observations are in agreement with the three-step model for protein adsorption hypothesized by Wang et al.,<sup>19</sup> which consists of an initial reversible association step, a rearrangement/reorientation step on the AuNP surface, and a final “hardening” step, which in the case of cysteine-containing proteins becomes irreversible.



**Fig. 2** Root Mean Square Deviations (RMSD) of a representative Ubq interacting with AuNP of 20nm diameter, computed during dynamics simulation. Data for the bare NP, implicit citrate, and explicit citrate simulations are reported in green, red, and blue, respectively. In the first period of the simulation (I) the Ubq moves around the NP without interacting with it; the black dotted line represents the time in which the first contact between Ubq and the AuNP surface is established. In the second period (II) Ubq moves on the NP surface; finally, in the third period (III) the Ubq optimal interaction site is found and very small RMSD fluctuations are observed. Three snapshots

of the simulation run representing each step of protein-NP interaction are reported in insets I, II, and III. The Ubq, whose motion has been traced during the dynamics run, is coloured in blue; transparent beads represent the other Ubqs; the NP surrounded by charged beads is coloured in gold.

### 3.2 Nanoparticle adsorption capacity

The maximum number of Ubq molecules that the AuNP is able to bind at his surface can not be obtained directly from the computational simulations, since the number of Ubqs in the simulation system is limited by the computational costs.

However, previous studies<sup>16,33</sup> showed that the nanoparticle adsorption capacity can be estimated by fitting the data regarding the variation of the number of bound proteins ( $N_{\text{bounds}}$ ) with time, by means of a stretched exponential function (Eq. 6) or a cumulative log-normal function (Eq. 7) (See ESI for details).

The stretched exponential function is given by:

$$N_{\text{bound}} = N_{\text{max}} \left( 1 - e^{-\left(\frac{t}{\tau}\right)^\alpha} \right) \quad (6)$$

where  $N_{\text{max}}$  is the maximum number of proteins that can be adsorbed on the NP,  $t$  is the simulation time,  $\alpha$  controls the width of the corresponding rate distribution and  $\tau$  is the characteristic time constant.

The log-normal function is given by

$$N_{\text{bound}} = \frac{N_{\text{max}}}{2} \left[ 1 + \text{erf} \left( \frac{\ln(t) - \mu}{\sqrt{2}\sigma} \right) \right] \quad (7)$$

Where,  $\mu$  and  $\sigma$  are respectively the mean and the standard deviation of the variable's natural logarithm and  $\text{erf}$  is the error function.

Very similar  $N_{\text{max}}$  values were obtained by using the two formulas for AuNPs of 10 and 16 nm diameter. Moreover, a computational simulation test carried out with 84 Ubqs and the 10 nm NP yielded the  $N_{\text{max}}$  values of 78, in full agreement with the values derived by Eq.s 6 and 7 (Table 2). As expected, the maximum number of Ubqs increases with the AuNP diameter, due to the increase in the available binding sites on the NP surface.

It is worth to note that this approach can not be applied to bigger NPs, because of the consistent difference between the limited number of Ubqs that we are forced to used in the simulation and the ones that should be really employed to reach an overall coverage of the NP surface.

The maximum number of Ubqs ( $N_{\text{max}}$ ) extracted by fitting our data using the cumulative log-normal function is compared with those obtained in the literature by using other methods in Table 2.

Wang et al.<sup>17</sup> quantified the stoichiometry of molecular binding to AuNPs ( $N_{\text{max}}$ ) by an NMR-based approach and found good agreement with the data values predicted using the radius of gyration ( $R_G$ ) calculated from the protein structure.  $R_G$  is considered to represent roughly the occluded surface on the AuNP itself. Using this metric,  $N_{\text{max}}$  is given by:

$$N_{\text{max}} = \frac{4R_{\text{AuNP}}^2}{R_G^2} \quad (8)$$

where  $R_{\text{AuNP}}$  is the radius of the nanoparticle, and  $R_G = 12.04 \text{ \AA}$  for Ubq. For AuNPs of 15 nm diameter, they found a maximum number of Ubqs absorbed of  $156 \pm 12$ , which compares very well with that predicted (155) by using eq. 8.

Calzolari et al.<sup>18</sup> used dynamic light scattering methodologies coupled with simple steric

considerations to estimate  $N_{\max}$  on the surface of each spherical gold nanoparticle:

$$N_{\max} = 0.65 (R_{\text{complex}}^3 - R_{\text{AuNP}}^3) / R_{\text{Ubq}}^3 \quad (9)$$

where  $R_{\text{complex}}$  is the radius of the AuNP-Ubq complex,  $R_{\text{AuNP}}$  is the radius of AuNP, and  $R_{\text{Ubq}}$  is the radius of Ubq.

Table 2 shows that there is consensus among the various methods used for the assessment of the maximum number of Ubqs bounded to NPs of small radii, but a consistent divergence for greater diameter. However, it must be taken into account Calzolari et al.<sup>18</sup> assume a close packing of proteins on the AuNP surface, therefore the actual number of Ubq molecules per nanoparticle is most likely lower.

**Table 2** Maximum number of Ubqs (with relative errors) that can bind to the AuNP with different diameters as derived from the MD simulations of the bare (B), and citrate-coated NP models, with citrates treated implicitly (I) and explicitly (E). Data computed according to Calzolari et al.<sup>18</sup> (Semi-empirical) and Wang et al.<sup>19</sup> (Radius of gyration) are also listed.

Model	NP diameter (nm)			
	10	16	20	24
<b>B</b>	78±4	177±2	-	-
<b>I</b>	70±20	160±15	-	-
<b>E</b>	77±18	165±21	-	-
<b>Semi-empirical</b> <sup>18</sup>	75	241	433	705
<b>Radius of gyration</b> <sup>19</sup>	69	177	276	397

### 3.3 Ubiquitin binding site

The frequency of interaction (contact probability) between each Ubq amino acid residues and the AuNP during the dynamics run is reported in Fig. 3 as averages over the 10, 16, 20, and 24 nm NPs for the bare, citrate implicit, and explicit models.

The analysis of the results shows different binding modalities for the three models, suggesting that the presence of citrates and the way in which they are taken into account in the simulations influence the Ubq-NP binding in a relevant manner. In fact, the Debye-Hückle potential felt by the proteins in the computational simulations is homogeneously positive for the bare NP, and homogeneously negative for the implicit citrates NP coated. Instead, the presence of citrates treated in an explicit manner implies the contemporaneous formation on the NP surfaces of areas of positive and negative potential. The importance of electrostatic interactions for binding orientations has been recently demonstrated experimentally by Treuel et al.,<sup>2</sup> who investigated the consequences for protein corona formation upon different NP surface charge distributions by using fluorescence correlation spectroscopy.

Residues S20, D21, N25 and G53 are the main responsible for the interaction of Ubqs with the bare AuNP. These residues are located in a region of highly negative potential generated by the spatial proximity of amino acid residues E16, E18, D21, E24, E51, and D52 (Fig. 3a).

The implicit citrate AuNP model shows a different behaviour. Ubqs feel the negative potential given by citrates at the NP surface (Fig. 3b) and interact with their hydrophobic moiety comprising residues L8, G10, A46, G47, K63, and G76 that are embedded into the positive charged area

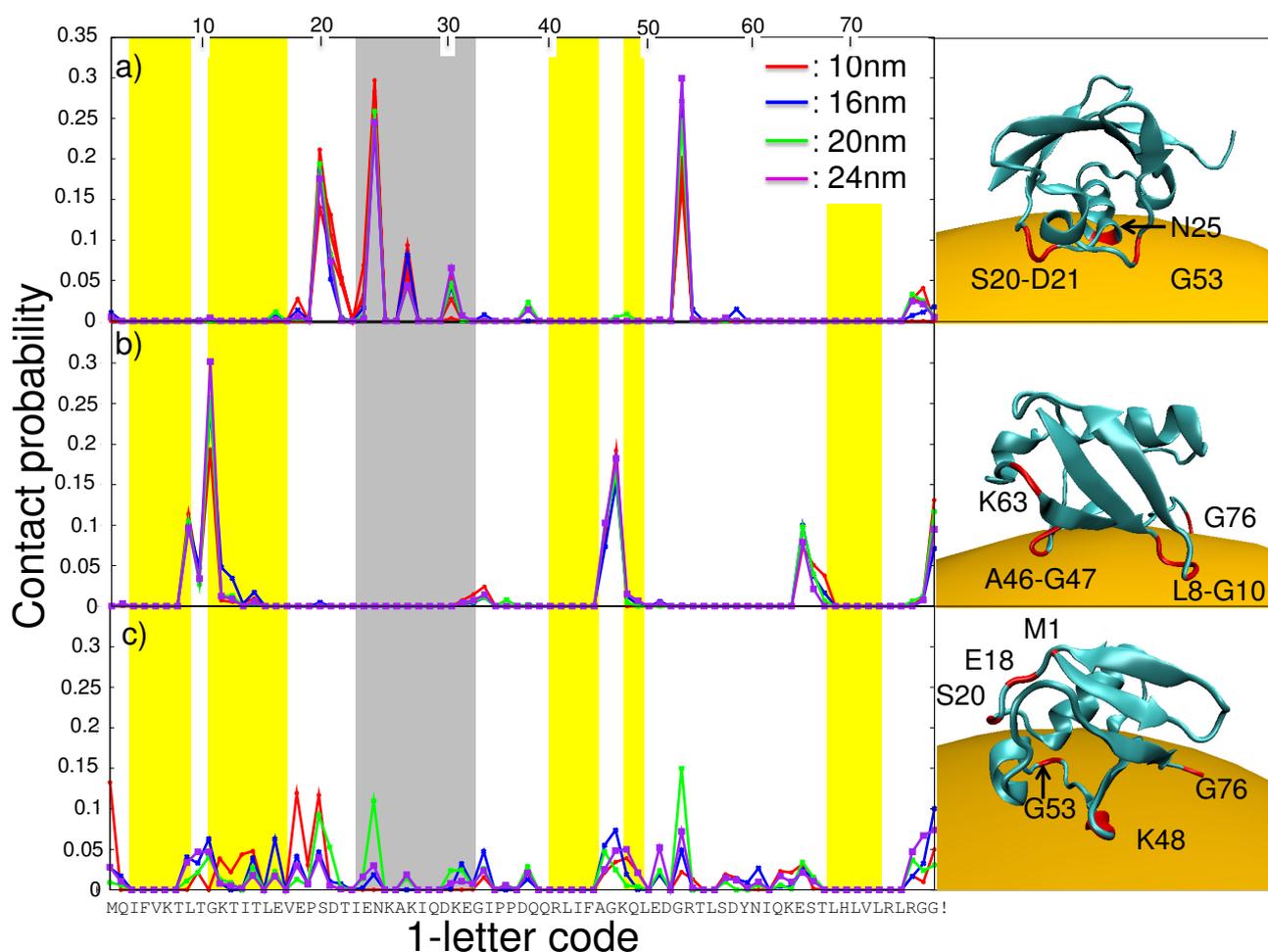
composed by the charged N-terminal, and K10, K48, R72, and R74 residues. The same pattern of interaction is found for the simulation carried out as a test (see Methods) on the all-atom Ubq model interacting with a 16 nm NP, within the implicit citrate model.

Finally, the explicit citrates model shows an Ubq interaction site mainly constitutes by M1, E18 S20, K48, G53, and G76 (Fig. 3c). This site is large and much less specific than the ones described in the previous cases, since Ubqs feel simultaneously both the positive and negative charges of gold and citrates, respectively. In addition, as underlined in the previous paragraph, Ubqs bind with both citrates and Au beads.

The amino acids residues listed above are also found to be responsible for Ubq-AgNP interactions in the recent study by Ding et al.<sup>16</sup> carried out by means of discrete molecular dynamics on both atomistic and CG models. In addition, their results revealed a competition between Ubqs and citrate ions with Ubq replacing citrate molecules bound at the AgNP surface. However, they found a specific binding orientation between Ubq molecules and AgNP surfaces, driven by electrostatic interactions. This discrepancy in the specificity of the interactions is due to the constraints imposed by Ding et al.<sup>16</sup> in order to enhance the interaction between amino acid residues 18, 19, 20, 21, 22, 24 25, 28, 52, 54, 55, 57 and 58 with AgNP, whereas no additional residue-specific parameters are used here.

A validation of the results obtained in the present study on the basis of detailed comparisons with the experimental findings available in the literature on similar systems<sup>18,19,20</sup> is hard to achieve.

In fact, the results obtained by means of chemical shift perturbation analysis, and dynamic light scattering in the pioneering study by Calzolari et al.<sup>18</sup> (i.e. two polypeptide fragments comprising G2, I3 and L15, V17, E18 show significant chemical shift perturbation upon NP addition) were not confirmed by subsequent investigations<sup>19,20</sup>. In particular, despite the similarity in the experimental condition and AuNPs size (12 nm vs 15 nm NP diameter), no significant chemical shift changes or line broadening for Ubqs in the presence of AuNPs were observed by Wang et al.<sup>19</sup>. Moreover, the chemical shift changes observed for the Ubq-bare AgNP system (i.e. residues 3, 6, 7, 13, 14, 17, 39, 42–44, 55, 61, 64, 66, 67, 69, 70), were attributed by Mangini et al.<sup>20</sup> to protein–protein interaction starting on the AgNPs surface and leading to amyloid transition, rather than to adhesion of Ubqs to the AgNPs surface.



**Fig 3** Contact probability between AuNP and each Ubq amino acid residues for the three bare (panel a), implicit citrates (panel b), and explicit citrates (panel c) NP models. The area coloured in yellow represents Ubqs zones of  $\beta$ -sheets secondary structure, and the area coloured in grey represents the Ubq  $\alpha$ -helical regions. Data for the 10, 16, 20, and 24 nm diameter AuNPs are reported in red, blue, green, and purple, respectively. Contact probability greater than 0.1, 0.1, and 0.05 can be considered significant for bare, implicit and explicit citrates NP models, respectively. A cartoon representation of the Ubq-AuNP contact probability for each model is also showed: protein red areas represent residues with higher contact probability, which are labelled with the respective one-letter code and sequence number.

### 3.4 Ubiquitin structural changes upon NP binding

The relationship between the relative size of the protein and NP (and, consequently, the perception of the NP curvature by the protein) and conformational changes upon binding has been investigated by comparing the percentage of secondary structures of Ubqs after adsorption with the one of the X-ray (PDB ID: 1UBQ<sup>14</sup>) and NMR (PDB: 2MI8<sup>15</sup>) resolved structures, see Table 3).

The overall fold of the protein is maintained during the computational simulations, the percentage of unfolded Ubqs being 4.17%, a figure comparable with the results of 4.54% reported in the computational simulation work by Ding et al.<sup>16</sup>

This behaviour is consistent with a combination of widely recognized favourable properties of Ubq, i.e. small size, tightly-packed globular structure, and remarkable intrinsic stability,<sup>34</sup> and it is confirmed by the very good resolution of the two-dimensional [<sup>15</sup>N-<sup>1</sup>H]-HSQC spectrum of human

Ubq in the presence of citrate coated AuNP obtained both by Calzolari et al.<sup>18</sup> and by Wang et al.,<sup>19</sup> which is indicative of a properly folded protein.

However, an overall small increase in the percentage of coil structure at the expenses of the  $\alpha$ -helix content is observed as a function of the model used. Moreover, while the variation in the secondary structure elements of Ubq are independent from the NP size for the explicit citrate models, a decrease of 7 to 5 % in the  $\alpha$ -helix content is observed as a function of NP size for the implicit citrate and bare NP models. The degree of destabilization of the  $\alpha$ -helix secondary structure found for Ubq interacting with AuNP is consistent with the 5% reduction in  $\alpha$ -helix content detected by circular dichroism measurements of a 10nm AgNP - Ubqs system reported by Ding et al.<sup>16</sup>. On the contrary, none of the computational models used in the present study is able to reproduce the increase in  $\beta$ -sheet content experimentally detected, although a dim tendency is observed for the 10nm bare NP. This seems not to be due to the CG model, since the all-atom simulation of Ubq (20 nm NP, implicit citrates) yields a % in  $\beta$ -sheets (15.2%) very similar to the one obtained for the corresponding CG simulation, the lower content % in  $\alpha$ -helix (11.2%) being compensated by an increase in the coil (43.5%), and turn (30.1%) structures.

**Table 3** Secondary structures characterization of isolated Ubq as detected from X-ray crystallography and NMR structures, and of Ubq interacting with AuNP as derived from the MD simulations of the bare (B), and citrate-coated NP models, with citrates treated implicitly (I) and explicitly (E).

Secondary structure (%)	X-ray (1UBQ)	NMR (2MI8)	MD simulations							
			10nm		16nm		20nm		24nm	
			B/	I / E	B/	I / E	B/	I / E	B/	I / E
$\alpha$ -helix	19	23	11.2 / 11.7 / 14.0	11.8 / 11.7 / 14.6	13.2 / 13.7 / 14.4	13.8 / 14.2 / 14.5				
$\beta$ -sheet	29	31	32.2 / 30.8 / 29.7	29.0 / 30.9 / 29.5	29.6 / 29.8 / 29.7	30.5 / 30.0 / 29.4				
Turn	19	22	19.7 / 16.6 / 19.4	19.1 / 17.4 / 19.5	18.4 / 22.0 / 19.3	19.2 / 20.8 / 19.2				
Coil	33	24	36.8 / 40.9 / 36.9	40.1 / 40.0 / 36.4	38.8 / 34.4 / 36.6	36.5 / 35.1 / 36.9				

### 3.5 Ubiquitin clustering

A marked propensity of Ubqs to aggregate into clusters, which progress from dimers to large Ubq conglomerates up to 20 proteins, is observed during the dynamics runs. The computational models with explicit citrates yield the highest number of proteins in the clusters favoured by the electrostatic field generated by the charge of citrates bounded to the proteins.

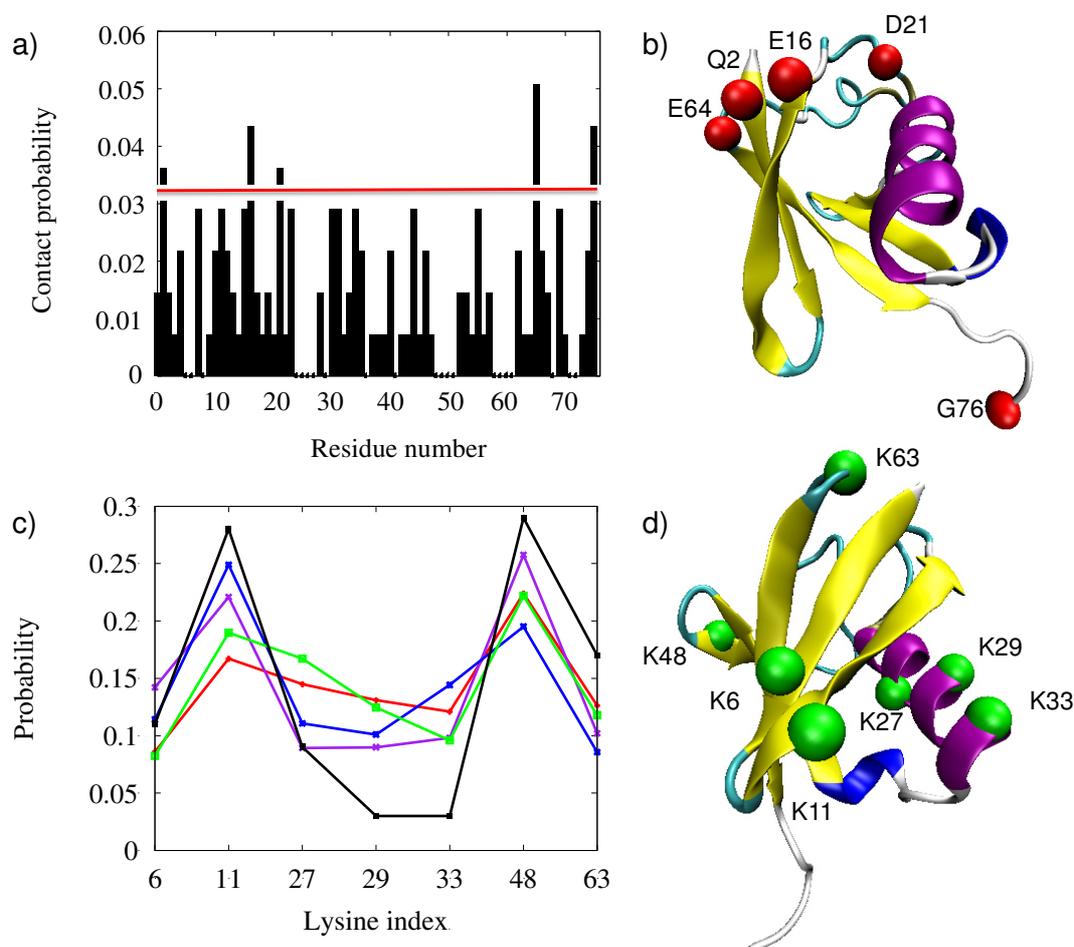
Aggregation phenomena occur since the first steps of the dynamics runs, and take place during all the phases of protein adsorption on AuNP surface described in paragraph 3.1.

Susceptibility to aggregation of Ubq represents a potential risk factor for disease onset or progression while cells attempt to tag and process toxic substrates<sup>35</sup>. Aggregation-prone regions of Ubq have been identified by Arnesano et al.<sup>36</sup> using the program PASTA (Prediction of Amyloid SStructure Aggregation),<sup>37</sup> as located in the polypeptide regions 1-15, and 65-70, which roughly

correspond to the regions involved in the protein-protein interactions more frequently observed during the molecular simulations runs carried out in the present study.

In fact, the results show that electrostatic interactions are established among amino acid residues Q2, E16, D21, E64, G76 and K11, K48 (highlighted in Figure 4a and in Figure 4b) belonging to two different Ubqs. These contacts orient the proteins allowing the  $\beta$ -sheets moieties of the two proteins to be close in the space.

In the computational models with implicit citrates the interaction of Ubqs with AuNP occurs mostly by means of the  $\beta$ -sheet regions, (See Section 3.3) thus subtracting potential binding sites for other Ubqs, therefore a smaller number of clusters is systematically observed for the implicit with respect to the explicit citrates models.



**Fig 4** a) Contacts of the lysine residues with all the amino acids of other Ubqs. A contact probability  $>0.03$  can be considered of relevance. b) Cartoon representation of Ubq; red spheres represent the residues with higher contact probability with lysine residues, labeled with their sequence numbers. The  $\alpha$ -helix is coloured in purple,  $\beta$ -sheets in yellow,  $3_{10}$  helix in blue, and coils in cyan. c) Contact probability of each lysine (labeled with their respective amino acid numbering in the protein sequence) computed for the simulations with explicit citrates and different AuNP diameter: 10nm (red), 16nm (purple), 20nm (blue), 24nm (green). Data for *S. cerevisiae* Ubq adapted from work of Komander<sup>39</sup> are reported in black. d) Cartoon representation of Ubq showing the location of the lysine residues, represented by green spheres and labeled with their sequence numbers.

Ubiquitination is the process by which Ubqs link to proteins or to others Ubqs via a lysine, forming clusters or chains. Ubqs contain seven lysine amino acids (K6, K11, K27, K29, K33, K48 and K63) and linkages with different lysines bring to different clusters and chains conformations. We found that K11 and K48 show the higher probability to make contact with other Ubqs in all the computational simulations carried out for the bare and coated AuNPs of different size, see Figure 4c and Figure 4d. Small differences are observed for the simulations with explicit citrate coating due to the direct Ubq-citrate interactions. Interesting, K11 and K48 are the residues most frequently implicated in the covalent polymerization with other Ubqs (28 and 29%, respectively) playing a fundamental role in protein ubiquitination<sup>38,39</sup>(see Fig. 4 c, black line): Lys11-linked polyUb chains is involved in the cell cycle regulation, and Lys48-linked polyUbq chains target their substrates to the proteasome for degradation.

In addition, the modality of Ubq-Ubq interaction observed during the dynamics runs is also supported by the results of the investigation on E16V, and E18V Ubq mutants carried out by Mangini et al.<sup>20</sup> on Ubq interacting with AgNP (~10nm). They recognized E16 as dispensable for adsorption of Ubq on the AgNP surface, but required for amyloid conversion, because either it is directly involved in the protein-protein aggregation interface or just needed to orient properly the adsorbed protein molecules so to prime their amyloidogenic interactions. In contrast, they recognized E18 as neither required for protein-AgNP interaction, nor for subsequent amyloid conversion.

#### 4. Conclusions

Coarse-grained computational simulations have allowed the treatment of a large number (48) of ubiquitin molecules in interaction with gold spherical nanoparticles, and the study of the influence of nanoparticle size (10, 16, 20, and 24 nm diameter) and environment (bare nanoparticle surface, and citrate-coated surface, treated as implicit and explicit citrate models) on the bio-corona formation.

The results show that, a consistent number of proteins is able to make persistent interactions with the nanoparticles already after few nanoseconds, depending from the nanoparticle size. Nevertheless, a proper protein corona is obtained only after a slower reorientation step that occurs at the nanoparticle surface in order to optimize the nanoparticle-ubiquitins interaction. The presence of citrates treated in an explicit manner protracts this step, competing with nanoparticle for binding ubiquitins.

A consensus among the various methods (based on fitting of the binding kinetic at saturation, or geometrical considerations) used for the assessment of the maximum number of ubiquitins bounded to nanoparticles of small radii is found, but a consistent divergence for greater diameter hampered the estimation of adsorption capacity for nanoparticle greater than 16 nm diameter.

Different binding modalities are found depending on the absence/presence of citrates and the way in which they are taken into account in the simulations. In particular, in the case of explicit citrates ubiquitins bind both to the gold nanoparticle and to citrates (free and/or anchored on the nanoparticle surface), therefore the ubiquitin-nanoparticle binding is less specific than in the other cases. This result is in agreement with experimental study of Wang et al.<sup>19</sup>. On the contrary, Calzolari et al.<sup>18</sup> showed a very specific binding site that is very similar to the one found in our simulations using a bare AuNP.

The overall fold of the protein is maintained during the computational simulations. However, an overall small decrease of 7 to 5 % in the  $\alpha$ -helix content is observed as a function of nanoparticle size depending on the model used to treat the environment, consistently with experimental findings by Ding et al.<sup>16</sup> for ubiquitins interacting with silver nanoparticles. Moreover, the all-atom simulation yields comparable results, suggesting that the coarse grained model adopted is able to reproduce correctly the main features of these systems.

Interestingly, a marked propensity of ubiquitins to aggregate into clusters is observed during the dynamics runs. Aggregation occurs since the first steps of dynamics, and takes place during all the phases of protein adsorption on the gold nanoparticle surface. Aggregation-prone zones (located in the  $\beta$ -sheets region formed by the first and last  $\beta$ -strands of each polypeptide chain) are brought to close spatial proximity by electrostatic interactions among Lysine11, Lysine48 and Glutamic Acid16, Glutamic Acid18, Aspartic Acid21, and Serine65 belonging to two different ubiquitins. In the implicit citrates computational models, where ubiquitins interaction with gold nanoparticle occurs mostly with the same sheet regions, a smaller number of clusters is observed.

The results show that computational approaches provide molecular insights into the possible mechanism of interactions and allow the identification of important patterns in the nanoparticle-protein interactions, thus emerging as a powerful tool for guiding the design of meaningful experiments.

## Acknowledgements

This work was supported by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) through the "Programma di ricerca di rilevante interesse nazionale" (PRIN) Grant 2010C4R8M8\_002 entitled "Nanoscale functional Organization of (bio)Molecules and Hybrids for targeted Application in Sensing, Medicine and Biotechnology" and the "Futuro in Ricerca" (FIRB) Grant RBFR1248UI 002 entitled "Novel Multiscale Theoretical/Computational Strategies for the Design of Photo and Thermo responsive Hybrid Organic-Inorganic Components for Nanoelectronic Circuits".

## Notes

<sup>a</sup>Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Via G. Campi 183, 41125, Modena, Italy.

## References

1. M. P. Monopoli, C. Aberg, A. Salvati and K. Dawson, *Nat. Nanotechnol.* 2012, **7**, 779–786.
2. P. Del Pino, B. Pelaz, Q. Zhang, P. Maffre, G. U. Nienhaus and W. J. Parak, *Mater. Horiz.*, 2014, **1**, 301-313.
3. L. Treuel, K. A. Eslahian, D. Docter, T. Lang, R. Zellner, K. Nienhaus, G. U. Nienhaus, R. H. Stauber and M. Maskos, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15053–15067.
4. M. Mahmoudi, I. Lynch, M. Reza Ejtehadi, M. P. Monopoli, F. Baldelli Bombelli and S. Laurent, *Chem. Rev.*, 2011, **111**, 5610–5637.
5. C. C. Fleischer and C. K. Payne, *Acc. Chem. Res.*, 2014, **47**, 2651-2659.
6. E. Casals, T. Pfaller, A. Duschl, G. J. Oostingh, and V. Puentes, *ACS Nano* 2010, **4**, 3623–3632.
7. D. Dell'Orco, M. Lundqvist, C. Oslakovic, T. Cedervall and S. Linse, *PLoS One* 2010, **5**, e10949.
8. M. Lundqvist, J. Stigler, T. Cedervall, T. Berggård, M. B. Flanagan, I. Lynch, G. Elia and K. Dawson. *ACS Nano* 2011, **5**, 7503–7509.
9. C. D. Walkey and W. C. W. Chan, *Chem. Soc. Rev.*, 2012, **41**, 2780-2799.
10. P. R. Leroueil, S. Hong, A. Mecke, J. R. Baker Jr., B. G. Orr and M. M. B. Holl, *Acc. Chem. Res.*, 2007, **40**, 335–342.
11. A. A. Shemetov, I. Nabiev and A. Sukhanova, *ACS Nano*, 2012, **6**, 4585–4602.
12. R. Cukalevski, M. Lundqvist, C. Oslakovic, B. Dahlbäck, S. Linse and T. Cedervall, *Langmuir* 2011, **27**, 14360–14369.

13. V. Marino, A. Astegno, M. Pedroni, F. Piccinelli and D. Dell'Orco, *Nanoscale* 2014, **6**, 412–423.
14. S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531-544.
15. T. Huang, J. Li and R. A. Byrd, *Protein Sci.*, 2014, **5**, 662-667.
16. F. Ding, S. Radic, P. Choudhary, R. Chen, J. M. Brown and P. C. Ke, *Nanoscale* 2013, **5**, 9162–9169.
17. G. Brancolini, D. B. Kokh, L. Calzolari, R. C. Wade and S. Corni, *ACS Nano*, 2012, **6**, 9863–9878.
18. L. Calzolari, F. Franchini, D. Gilliland and F. Rossi, *Nano Lett.*, 2010, **10**, 3101–3105.
19. A. Wang, K. Vangala, T. Vo, D. Zhang and N. C. Fitzkee, *J. Phys. Chem. C*, 2014, **118**, 8134–8142.
20. V. Mangini, M. Dell'Aglio, A. De Stradis, A. De Giacomo, O. De Pascale, G. Natile and F. Arnesano, *Chem. Eur. J.*, 2014, **20**, 10745–10751.
21. R. Li, R. Chen, P. Chen, Y. Wen, P. Chun Ke and S. S. Cho, *J. Phys. Chem.*, 2013, **43**, 13451-13456.
22. J. M. Cook, *Math. Tables Aids Comput.* 1957, **11**, 81–82.
23. G. Marsaglia, *Ann. Math. Stat.*, 1972, **43**, 645–646.
24. V. Tozzini, *Q. Rev. Biophys.*, 2010, **43**, 333–371.
25. C. Clementi, H. Nymeyer and A. N. Onuchic, *J. Mol. Biol.*, 2000, **298**, 937–953.
26. R. D. Hills and C. L. Brooks, *Int. J. Mol. Sci.*, 2009, **10**, 889–905.
27. D. L. Pincus, S. S. Cho, C. Hyeon and D. Thirumalai, *Prog. Mol. Biol. Transl. Sci.*, 2008, **84**, 203–250.
28. F. Tavanti and V. Tozzini, *Molecules*.2014, **19**, 14961-14978.
29. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J Wang and P. Kollman, *J. Comp. Chem.*, 2003, **24**, 1999-2012.
30. D. B. Kokh, S. Corni, P. J. Winn, M. Hoefling, K. E., Gottschalk and R. C. Wade, *J. Chem. Theory Comput.*, 2010, **6**, 1753–1768.
31. F. Iori, R. D. I. Felice, E. Molinari and S. Corni, *J. Comp. Chem.*, 2008, **30**, 1465–1476.
32. W. Smith and T. R. Forester. *J. Mol. Graphics*, 1996, **3**, 136–141.
33. G. C. Faas, B. Schwaller, J. L. Vergara and I. Mody, *PLoS Biology* 2007, **5**, e311.
34. S. E. Jackson, *Org. Biomol. Chem.*, 2006, **4**, 1845–1853.
35. A. H. P. Jansen, E. A. J. Reits and E. M. Hol, *Front. Mol. Neurosci.*, 2014, **7**, 1-14.
36. F. Arnesano, S. Scintilla, V. Calò, E. Bonfrate, C. Ingrosso, M. Losacco, T. Pellegrino, E. Rizzarelli and G. Natile, *PLoS one*, 2009, **9**, e7052.
37. A. Trovato, F. Chiti, A. Maritan and F. Seno, *PLoS Computat. Biol.*, 2006, **12**, e170.
38. D. Komander and M. Rape, *Ann. Rev. Biochem.*, 2012, **81**, 203-229.
39. D. Komander, *Biochem. Soc. Trans.*, 2009, **37**, 937–953.