Pablo del Pino[a,d], Beatriz Pelaz[a], Qian Zhang[a], Pauline Maffre[b], G. Ulrich Nienhaus[b,c], Wolfgang J. Parak[a,4d*]

[a]Fachbereich Physik, Philipps Universität Marburg, Marburg, Germany
[b] Institute of Applied Physics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[c] Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[d]CIC Biomagune, San Sebastian, Spain
[*]corresponding author: wolfgang.parak@physik.uni-marburg.de

Protein corona formation around nanoparticles - from the past to the future

Supporting Information

Detailed steps to obtain the equations as presented in the main manuscript

According to the law of mass action

$$NP + n\cdot P \leftrightarrow P_nNP \quad (1)$$

the dissociation equilibrium coefficient K_D (or also called the apparent dissociation coefficient) of Equation (1), in which one nanoparticle (NP) reacts with n proteins (P) to form a protein-NP complex (P_nNP), is expressed as

$$K_D = \frac{c(NP)\cdot c^n(P)}{c(P_nNP)} = \frac{k_{off}}{k_{on}} \quad (2)$$

Hereby c(NP), c(P), and $c(P_nNP)$ are the concentrations of free (naked) NPs without attached protein, of unbound protein, and of the protein-NP complex, respectively. k_{on} and k_{off} are the on- and off-rates. Thus, the total amount $c_0(NP)$ of NPs which are in the solution comprises the free NPs and the protein-NP complexes. In the same way the total amount of proteins in solution $c_0(P)$ is given by the free proteins and by the proteins bound to the NPs, whereby each NP binds n proteins to form one protein-NP complex P_nNP:

$$c_0(NP) = c(NP) + c(P_nNP) \quad (3)$$

$$c_0(P) = c(P) + n\cdot c(P_nNP) \quad (4)$$

$c_0(NP)$ and $c_0(P)$ can also be regarded as NP and protein concentrations before the reaction between them started, respectively. c(NP) and c(P) are the concentrations of naked, protein-free NPs and free proteins, respectively, after NPs and proteins have been brought into contact and an

equilibrium according to Equation (1) has been reached. One can now calculate the ratio of the number N of NPs with saturated protein shell (*i.e.* the amount of complexes P_nNP) to the total number N_{max} of NPs:

$$\frac{N}{N_{max}} = \frac{c(P_nNP)}{c_0(NP)} \xrightarrow{\text{Eq.(3)}} \frac{c(P_nNP)}{c(NP)+c(P_nNP)} = \frac{1}{\frac{c(NP)}{c(P_nNP)}+1} = \frac{1}{1+\frac{c(NP)}{c(P_nNP)}} \xrightarrow{\text{Eq.(2)}} \frac{1}{1+\frac{K_D}{c^n(P)}} = \frac{c^n(P)}{c^n(P)+K_D} \quad (5)$$

The Hill parameter n hereby describes the cooperativity. n > 1 indicates cooperative binding, *i.e.*, if NPs are already saturated with proteins, it is easier for the following NPs to become saturated. In other words, it is easier for several NPs to collectively become saturated with proteins than for single NPs to become independently saturated. In contrast, n < 1 refers to anti-cooperative binding, so that protein saturation of some NPs lowers the tendency for other NPs to become saturated with proteins. In the context of this simplified model of the reaction between proteins and NPs introduced with Eq. (5), the interpretation of the Hill coefficient can be subject to discussion, as the coverage state of a NP is not supposed to influence, other than through the law of mass action, the coverage state of another second NP. Despite this argument, Eq. (5) nevertheless provides a useful model for protein binding to NPs, if the involved parameters are suitably reinterpreted, as will be described in the following.

In Equation (5) we so far assumed a scenario in which, in equilibrium, N NPs are saturated with proteins (*i.e.*, these NPs form complexes P_nNP), and (N_{max}-N) NPs are naked (*i.e.*, with no protein attached). Of course, the scenario where NPs can only be naked or saturated does not reflect the reality, where NPs can also be partially covered by proteins. To describe the reaction between NPs and proteins exactly, we should consider the Adair scheme, where each partially covered state of the NPs is considered. However, for a large number of binding sites per NP, this scheme becomes very fast very complicated and we instead choose here to consider a second scenario, in which all NPs have the same (partially) coverage state. In other word, all NPs have an average number of N proteins per NP (*i.e.*, there are no naked, protein-free NPs). In this scenario, N_{max} would be the maximum number of binding sites for proteins *per* NP. In the following, we show that the fraction of saturated NPs N/N_{max} from the first scenario is equivalent to the fraction N/N_{max} in this second scenario. The first scenario considers the ratio x of the number of saturated NPs (*i.e.* N_{max} proteins per NP) to the total number of NPs, $x = N/N_{max}$. The ratio of the number of NPs without protein shell (*i.e.*, 0 proteins per NP) is then given by $(1-x) = (N_{max}-N)/N_{max}$. Thus, the total amount of proteins attached to the NPs is equal to the proteins on the saturated NPs and on the bare NPs: $x \cdot N_{max} + (1-x) \cdot 0 = \frac{N}{N_{max}} \cdot N_{max} + \frac{N_{max}-N}{N_{max}} 0 = N$. In the second scenario all NPs have N proteins attached and there are no "naked" NPs, which leads to the same total number of proteins per total number of NPs as in the first scenario. In the following we will continue our discussion according to the interpretation of the second scenario. Thus, N/N_{max} can be interpreted as the fraction of occupied protein sites on the NP surface.

Equation (5) suggests also reinterpretation of n. In a modern view, n is not the stoichiometric factor, and it can also have fractional numbers (instead of only integer values). According to the Hill model, n is the so-called Hill coefficient. This coefficient n describes the cooperativity of binding ligands (in our case proteins) to their substrate (in our case the surface of NPs). $n > 1$ describes cooperative binding. Adhesion of proteins already present on the NP surface is enhanced if more proteins bind next to them. $n < 1$ refers to anti-cooperative binding. Adhesion of proteins already present on the NP surface is suppressed if more proteins bind next to them. $n = 1$ describes non-cooperative binding. Each protein binds individually to the NP surface and does not recognize the proteins that are already bound there.

Instead of the dissociation constant K_D, often, the concentration at which half of protein coverage is achieved (*i.e.* the protein concentration producing half occupation of the NP surface) K'_D is considered. Let us assume half of the NPs are saturated with proteins at a concentration of free proteins $c_{1/2}(P)$:

$$\frac{1}{2} = \frac{N}{N_{max}} \xrightarrow{Eq.(5)} \frac{c^n_{1/2}(P)}{c^n_{1/2}(P)+K_D} \Rightarrow 2\cdot c^n_{1/2}(P) = c^n_{1/2}(P) + K_D \Rightarrow c^n_{1/2}(P) = K_D \Rightarrow c_{1/2} = K_D^{1/n} := K'_D \quad (6)$$

Thus, Equation (5) can be rewritten as $\frac{N}{N_{max}} \xrightarrow{Eq.(5,6)} \frac{c^n(P)}{c^n(P)+K'^n_D} = \frac{1}{1+(\frac{K'_D}{c(P)})^n}$. (7)

Now, we consider a situation (according to the interpretation with the second scenario) in which, in equilibrium, N proteins are bound to each NP on average. Thus, all NPs have on average N proteins bound *per* NP:

$$c_0(NP) = c(P_N NP). \quad (8)$$

However, there might also be free proteins (= ligands; c(P)) in solution. Thus the total amount of proteins is

$$c_0(P) = c(P) + N\cdot c(P_N NP) \xrightarrow{Eq.(8)} c(P) + N\cdot c_0(NP) \xrightarrow{Eq.(7)} c(P) + \frac{N_{max}}{1+(\frac{K'_D}{c(P)})^n} c_0(NP) \quad (9)$$

If the concentration of free proteins is much smaller than the concentration of half saturation, according to Equation (9) one gets:

$$c_0(P) \xrightarrow{Eq.(9)} c(P) + \frac{N_{max}}{1+(\frac{K'_D}{c(P)})^n} c_0(NP) \xrightarrow{c(P) << K'_D} c(P) + \frac{N_{max}}{(\frac{K'_D}{c(P)})^n} c_0(NP)$$
$$= c(P) + (\frac{c(P)}{K'_D})^n \cdot N_{max}\, c_0(NP) \quad (10)$$

On the other hand, if the concentration of free proteins is much bigger than the concentration of half saturation, then all NPs are saturated with proteins and additional free proteins remain in solution, see also Figure S1.

$$c_0(P) \xrightarrow{\text{Eq.(9)}} c(P) + \frac{N_{max}}{1+(\frac{K'_D}{c(P)})^n} c_0(NP) \xrightarrow{c(P) >> K'_D} c(P) + \frac{N_{max}}{1+0} c_0(NP) = c(P) + N_{max}\, c_0(NP)$$
$$\xrightarrow{c(P) >> N_{max} \cdot c_0(NP)} c(P) \qquad (11)$$

According to Equation (10), at very low protein concentrations c(P), a significant part of the proteins is bound to the surface of NPs: $c_0(P)\text{-}c(P) \rightarrow (\frac{c(P)}{K'_D})^n \cdot N_{max} \cdot c_0(NP)$. On the other hand, at very high protein concentrations c(P), basically all proteins are free proteins, *cf.* Equation (11), as the absolute amount of proteins which is bound to NPs ($c_0(P)\text{-}c(P) \rightarrow 0$) is small, due to the fact that NPs cannot exist at very high concentrations $c_0(NP)$ well above the mM regime. In other words, since all NPs are already saturated with protein, all of the excess of added protein will remain free, *cf.* Figure S1.

Equation (9) is very useful, as experimentally, $c_0(P)$ is the easily accessible protein concentration, *i.e.*, the concentration of protein which has been added to solution, and not c(P), the concentration of protein which remains free in solution, after an equilibrium according to Equation (1) has been reached. It can be used to calculate the fraction of proteins which is bound to NPs (*i.e.*, the ratio of the amount of bound proteins to the total amount of proteins):

$$\frac{c_0(P) - c(P)}{c_0(P)} \xrightarrow{\text{Eq.(9)}} \frac{N_{max}}{1+(\frac{K'_D}{c(P)})^n} \frac{c_0(NP)}{c_0(P)} \qquad (12)$$

In Equation (9) the total protein concentration $c_0(P)$ is given as a function f depending on the concentration of free protein: $c_0(P) = f(c(P))$. Equation (9), however, also provides an implicit equation which allows for calculating c(P) in dependence on $c_0(P)$ as inverse function f^{-1}: $c(P) = f^{-1}(c_0(P))$. In general, the inverse function can be calculated only numerically (Figure S1 shows that, in order to do this, the c(P) and $c_0(P)$ axes simply need to be switched), but not analytically. However, in some special cases, an analytical solution is also possible (such as for n = 0.5, 1, 2). In the case of n= 1, Equation (9) is simplified to

$$c_0(P) \xrightarrow{\text{Eq.(9)}} c(P) + \frac{N_{max}}{1+(\frac{K'_D}{c(P)})^n} c_0(NP) \xrightarrow{n=1} c(P) + \frac{N_{max}}{1+\frac{K'_D}{c(P)}} c_0(NP) \qquad (13)$$

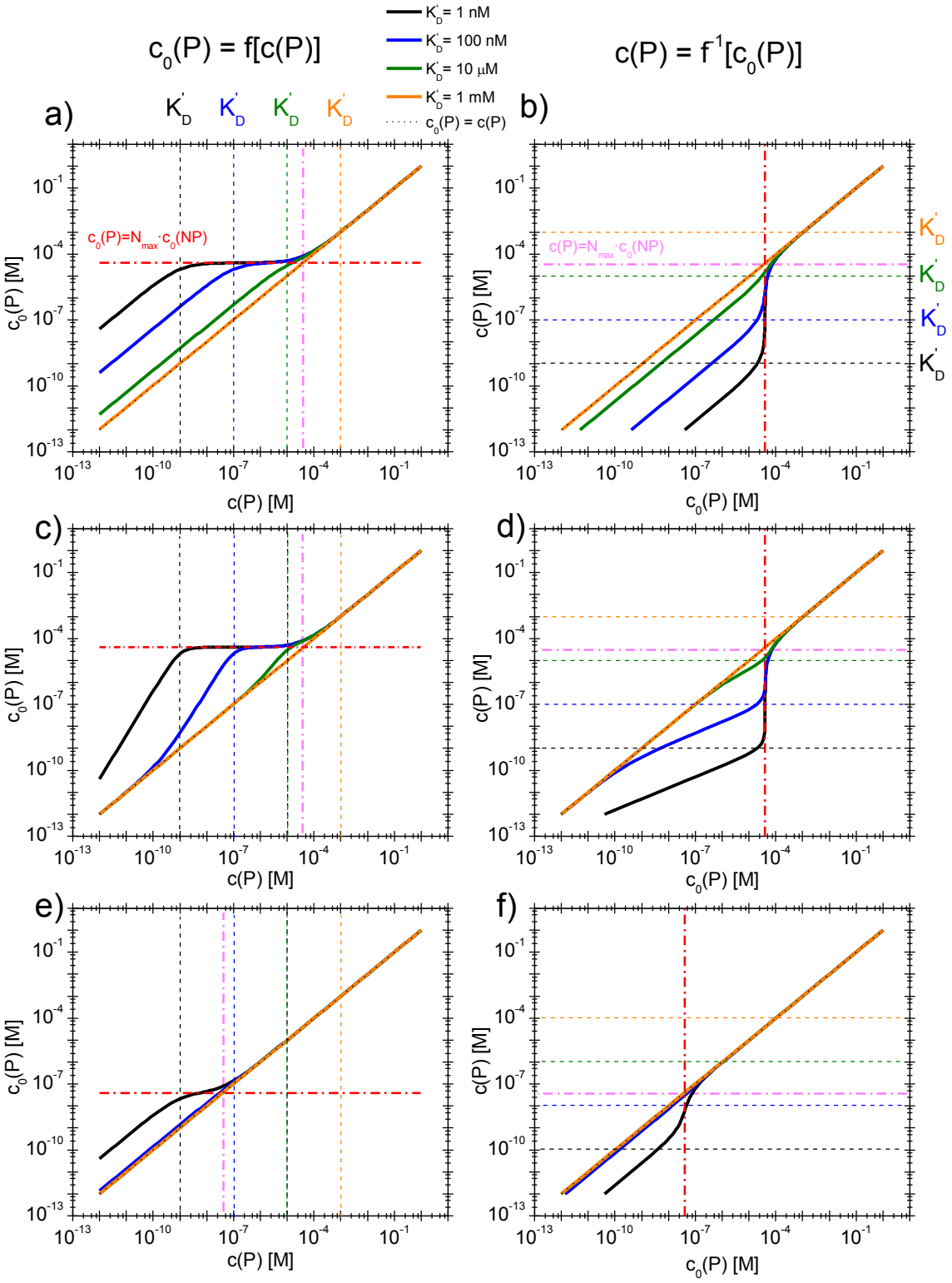

Figure S1. a, c, f) Direct graphical representation of Equation (9) in which $c_0(P)$ is displayed in dependence of c(P). for $K_D' = 1$ nM, 100 nM, 10 μM, and 1 mM with the following additional parameters: a) N_{max}= 40, $c_0(NP)$= 1 μM, n=1, c) N_{max}= 40, $c_0(NP)$= 1 μM, n=2, and e) N_{max}= 40, $c_0(NP)$= 1 nM, n=1. b, d, e) display the corresponding inverse functions, c(P) versus $c_0(P)$, with the same parameters. The dotted/dashed red lines (parallel to the c(P) axes) are guides to the eye and represent $c_0(P) = N_{max} \cdot c_0(NP)$. The dotted/dashed pink lines (parallel to the $c_0(P)$

axes) represents guides to the eyes for c(P)= N_{max}·c_0(NP). The dotted grey lines represent the diagonals c_0(P)= c(P), which can be hardly seen because they are beneath the graphs for K'_D = 1 mM.

By restructuring some terms it is possible to express Equation (12) as a quadratic equation:

$$c_0(P) = c(P) + \frac{N_{max}}{1+\frac{K'_D}{c(P)}}\, c_0(NP) \;\Leftrightarrow\; c_0(P) = c(P) + \frac{c(P)\cdot N_{max}}{c(P)+K'_D}\, c_0(NP)$$

$$\Leftrightarrow (c(P) + K'_D)\cdot c_0(P) = (c(P) + K'_D)\cdot c(P) + c(P)\cdot N_{max}\, c_0(NP)$$

$$\Leftrightarrow c(P)^2 + c(P)\cdot[N_{max}\cdot c_0(NP) + K'_D - c_0(P)] - [c_0(P)\cdot K'_D] = 0 \qquad (14)$$

Equation (13) is a quadratic equation of the form x^2+bx+c=0 for which the solution is $x = \frac{-b\pm\sqrt{b^2-4c}}{2}$. In the present case, only the "+" is a physically correct solution. Thus, Equation (14) can be solved and therefore, in the case of n = 1, we get the dependence of c(P) on c_0(P) as analytical expression:

$$c(P) = \frac{1}{2}\left[(c_0(P) - N_{max}\cdot c_0(NP) - K'_D) + \sqrt{(N_{max}\cdot c_0(NP) + K'_D - c_0(P))^2 + (4\cdot c_0(P)\cdot K'_D)}\right] \qquad (15)$$

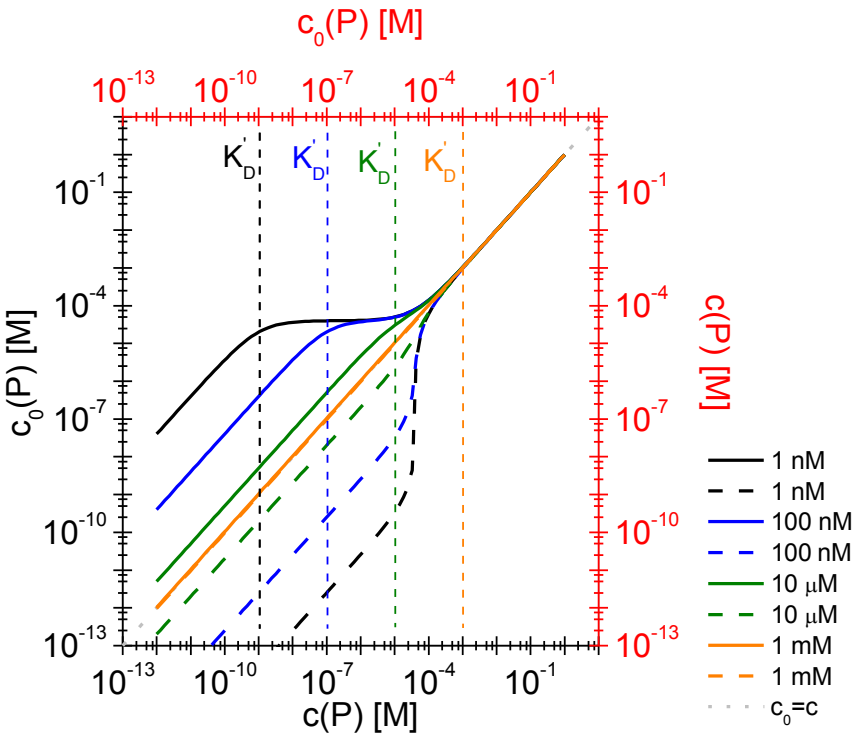

Figure S2. Graphical representation of Equations (13) and (15) for the parameters N_{max}= 40, c_0(NP)= 1 µM, n=1, and K'_D = 1 nM, 100 nM, 10 µM, and 1 mM. The solid lines (with the corresponding black axes) represent c_0(P) versus c(P), and the dashed lines (with the corresponding red axes) represent c(P) versus c_0(P).

For the present example (n = 1), Figure S2 displays Equation (13) and (15) for different K'_D values. The results are the same as in Figure S1, but now obtained analytically and not numerically. However, as already mentioned, in general no analytical solution is possible and thus, c(P) in dependence on c_0(P) has to be numerically derived from the implicit Equation (9).

In the special case n = 1, also Equation (12) can be solved analytically in dependence on c_0(P), by putting Equation (15) into Equation (12). Some results are displayed in Figure S3.

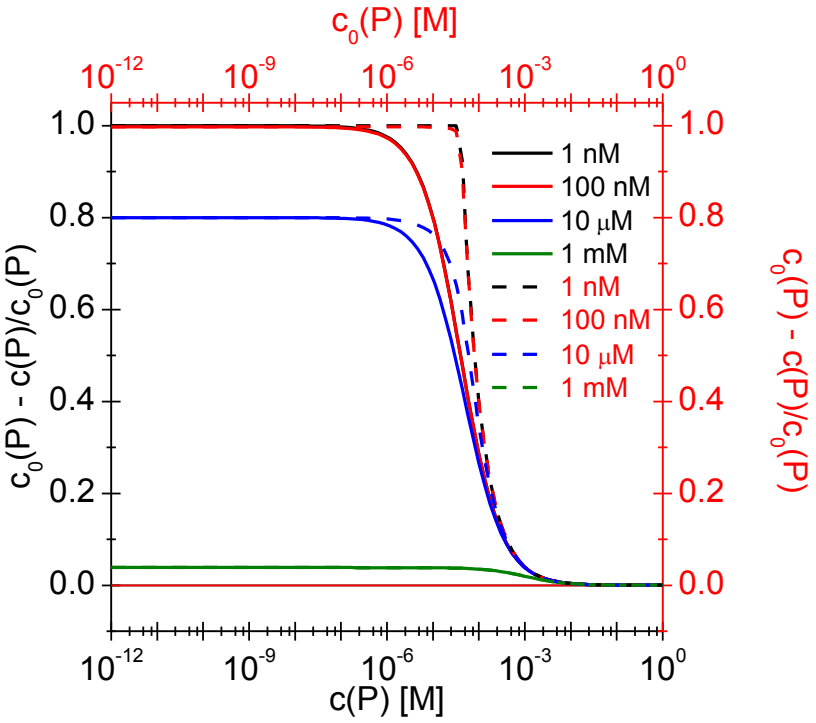

Figure S3. Fraction of the proteins which are bound to the NPs (c_0(P)-c(P))/c_0(P)) in dependence on the free protein concentration c(P) (solid lines, corresponding to the black axes), and in dependence on the total protein concentration c_0(P) (dashed lines, corresponding to the red axes), according to Equations (12) and (15), using the parameters N_{max} = 40, c_0(NP) = 1 µM, n = 1. The protein concentrations at half saturation K'_D were varied: K'_D = 1 nM, 100 nM, 10 µM, and 1 mM.

The same concept as described above can now be applied also for two different types of proteins $P_{(1)}$ and $P_{(2)}$:

$$NP + n_{(1)}\cdot P_{(1)} \leftrightarrow P_{n(1)}NP \qquad K_{D(1)} = \frac{c(NP)\cdot c^{n_{(1)}}(P_{(1)})}{c(P_{n(1)}NP)};\ K'_{D(1)} = (K_{D(1)})^{1/n_{(1)}} \qquad (16)$$

$$NP + n_{(2)}\cdot P_{(2)} \leftrightarrow P_{n(2)}NP \qquad K_{D(2)} = \frac{c(NP)\cdot c^{n_{(2)}}(P_{(2)})}{c(P_{n(2)}NP)};\ K'_{D(2)} = (K_{D(2)})^{1/n_{(2)}}$$

In case NPs which are saturated with protein species $P_{(1)}$ are brought into contact with free protein ligands $P_{(2)}$, an exchange (ligand exchange) can take place:

$$P_{n(1)}NP + n_{(2)}\cdot P_{(2)} \leftrightarrow P_{n(2)}NP + n_{(1)}\cdot P_{(1)} \text{ with } K_D = \frac{c^{n_{(2)}}(P_{(2)})\cdot c(P_{n(1)}NP)}{c(P_{n(2)}NP)\cdot c^{n_{(1)}}(P_{(1)})} = \frac{K_{D(2)}}{K_{D(1)}} \quad (17)$$

If the NPs are incubated with both protein species at the same time, then both of them can bind to the NP surface. For reasons of simplicity, according to the first scenario, we are considering here only NPs with either $P_{(1)}$ or $P_{(2)}$, but not mixtures of both proteins on the same NP, though they certainly exist (as we will later use them in the second scenario):

$$NP + n_{(1)}\cdot P_{(1)} + n_{(2)}\cdot P_{(2)} \leftrightarrow P_{n(1)}NP + P_{n(2)}NP \quad (18)$$

Following the strategy of Equation (5) we can then determine how many NPs will be covered with $P_{(1)}$ and with $P_{(2)}$. Following the first scenario from above, the total amount of NPs is given by free NPs, NPs saturated with protein species $P_{(1)}$, and NPs saturated with protein species $P_{(2)}$:

$$c_0(NP) = c(NP) + c(P_{n(1)}NP) + c(P_{n(2)}NP) \quad (19)$$

Thus, the ratio of number $N_{(1)}$ of NPs with saturated shell of protein $P_{(1)}$ to the total number of NPs N_{max} can be expressed as:

$$\frac{N_{(1)}}{N_{max}} = \frac{c(P_{n(1)}NP)}{c_0(NP)} \xrightarrow{Eq.(19)} \frac{c(P_{n(1)}NP)}{c(P_{n(1)}NP)+c(P_{n(2)}NP)+c(NP)} = \frac{1}{1+\frac{c(P_{n(2)}NP)}{c(P_{n(1)}NP)}+\frac{c(NP)}{c(P_{n(1)}NP)}}$$
$$= \frac{1}{1+\frac{c(P_{n(2)}NP)}{c(P_{n(1)}NP)}\cdot\frac{c(NP)}{c(NP)}+\frac{c(NP)}{c(P_{n(1)}NP)}} = \frac{1}{1+\frac{K_{D(1)}}{c^{n_{(1)}}(P_{(1)})}\cdot\frac{c^{n_{(2)}}(P_{(2)})}{K_{D(2)}}+\frac{K_{D(1)}}{c^{n_{(1)}}(P_{(1)})}}$$
$$= \frac{1}{1+\frac{K_{D(1)}}{c^{n_{(1)}}(P_{(1)})}\cdot\left(1+\frac{c^{n_{(2)}}(P_{(2)})}{K_{D(2)}}\right)} \xrightarrow{Eq.(16)} \frac{1}{1+\left(\frac{K_{D\prime(1)}}{c(P_{(1)})}\right)^{n_{(1)}}\cdot\left(1+\left(\frac{c(P_{(2)})}{K_{D\prime(2)}}\right)^{n_{(2)}}\right)} \quad (20)$$

Equation (20) can also be interpreted according to the second scenario in which, in average, each NP has its N_{max} available binding sites covered with $N_{(1)}$ proteins of species $P_{(1)}$ and $N_{(2)}$ proteins of species $P_{(2)}$. Knowing how many proteins are bound per NP on average, one can calculate the hydrodynamic radius r_h of one protein-NP complex. In case N proteins are adsorbed per NP, the volume V(N) of the NP with the protein corona is

$$V(N) = V_{NP} + N\cdot V_P. \quad (21)$$

Hereby, V_P is the volume of one protein (which can be estimated from protein databases) and

$$V_{NP} = V(0) = \frac{4\pi}{3} r_h{}^3(0) \quad (22)$$

is the volume of one NP without attached proteins (*i.e.*, without proteins in solution). Thus, the hydrodynamic radius $r_h(N)$ of one NP with N adsorbed proteins is

$$r_h(N) = \sqrt[3]{\frac{3}{4\pi}V(N)} = \sqrt[3]{\frac{3}{4\pi}(V_{NP} + N \cdot V_P)} = \sqrt[3]{r^3(0) + \frac{3}{4\pi}N \cdot V_P} \qquad (23)$$

with $r_{NP} = r_h(0)$ being the hydrodynamic radius of one plain NP without adsorbed proteins. Using Equation (7), this becomes

$$r_h(c(P)) = \sqrt[3]{r_{NP}^3 + \frac{3}{4\pi} \cdot N_{max} \cdot \frac{1}{1+(\frac{K_{D'}}{c(P)})^n} \cdot V_P}\ , \qquad (24)$$

where c(P) is the concentration of free protein, which however can be expressed in terms of the added concentration of protein $c_0(P)$, which is the concentration experimentally accessible. Thus, in order to calculate $r_h(c_0(P))$ in Equation (24), c(P) has to be expressed in terms of $c_0(P)$ *via* the implicit Equation (9). Only in some special cases (such as n = 1), an analytical solution can be directly given by inserting Equation (15) in Equation (24), *cf.* Figure S4.

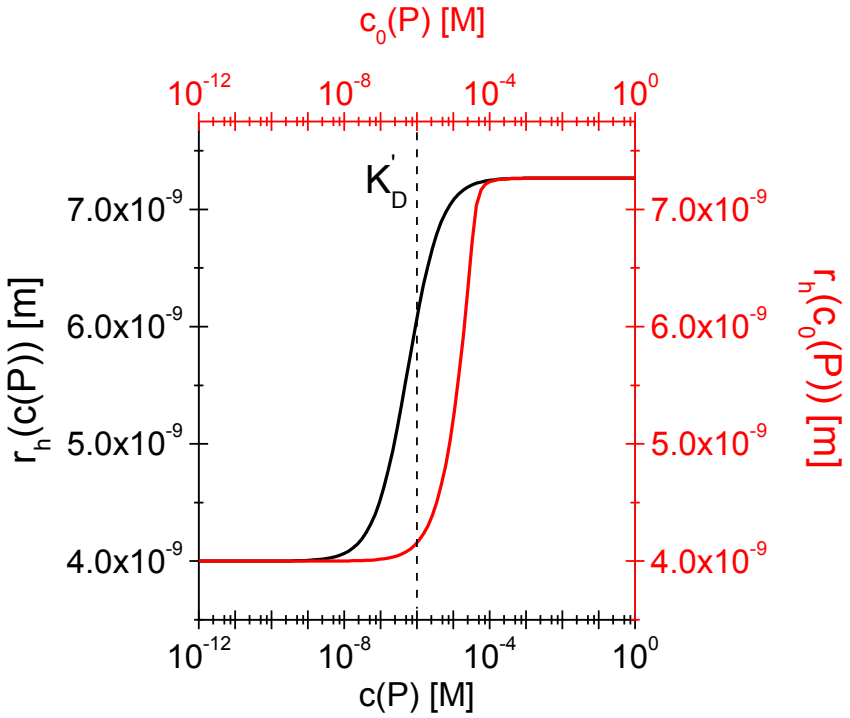

Figure S4. Hydrodynamic radius r_h of NPs in terms of c(P) (black line), and c_0(P) (red line), according to Equation (24) with the parameters N_{max} = 40, n = 1, c_0(NP) = 1 μM, r_{NP} = 4 nm, $V_P = \frac{4}{3}\pi r_p^3$ with r_P = 2 nm (idealized globular proteins were considered in this example) and K'_D = 1 μM.