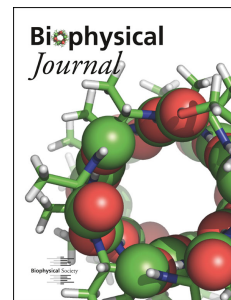


Journal Pre-proof

A hard-sphere model of protein corona formation on spherical and cylindrical nanoparticles

I. Rouse, V. Lobaskin



PII: S0006-3495(21)00736-0

DOI: <https://doi.org/10.1016/j.bpj.2021.09.002>

Reference: BPJ 11282

To appear in: *Biophysical Journal*

Received Date: 28 April 2021

Accepted Date: 2 September 2021

Please cite this article as: Rouse I, Lobaskin V, A hard-sphere model of protein corona formation on spherical and cylindrical nanoparticles, *Biophysical Journal* (2021), doi: <https://doi.org/10.1016/j.bpj.2021.09.002>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Biophysical Society.

A hard-sphere model of protein corona formation on spherical and cylindrical nanoparticles

I. Rouse^{1,*} and V. Lobaskin¹

¹School of Physics, University College Dublin, Belfield, Dublin 4, Ireland

*Correspondence: ian.rouse@ucd.ie

ABSTRACT

A nanoparticle (NP) immersed in biological media rapidly forms a corona of adsorbed proteins, which later controls the eventual fate of the particle and the route through which adverse outcomes may occur. The composition and timescale for the formation of this corona are both highly dependent on both the NP and its environment. The deposition of proteins on the surface of the NP can be imitated by a process of random sequential adsorption and, based on this model, we develop a rate-equation treatment for the formation of a corona represented by hard spheres on spherical and cylindrical NPs. We find that the geometry of the NP significantly alters the composition of the corona through a process independent of the rate constants assumed for adsorption and desorption of proteins, with the radius and shape of the NP both influencing the corona. We further investigate the roles of protein mobility on the surface of the NP and changes in the concentration of proteins.

SIGNIFICANCE The adsorption of proteins to nanoparticles is key to understanding their biological activity and fate. We present an improved version of the rate-equation based adsorption model that imitates random sequential adsorption and can be applied to curved nanoparticle surfaces. Our results demonstrate that the geometry of the nanoparticle alters both the evolution and the final state of the corona when proteins are modelled as rigid particles. In comparison to the previously employed mean-field model, in which proteins are taken to be arbitrarily deformable, we find a significantly longer timespan for the corona to form and reach the equilibrium state. We also demonstrate that the protein-nanoparticle interaction can change the corona formation kinetics and the resulting corona content, thus affecting the biological fate of the nanoparticle.

INTRODUCTION

Matter, even elements which are well-known on everyday size scales, takes on new properties on the nanoscale due to increased surface area and quantum mechanical effects, such that even a famously inert metal like gold may become highly reactive in the form of nanoparticles (NP) (1). These new properties, however, potentially lead also to hazards when the NPs are inhaled, absorbed through the skin, ingested, or otherwise find their way into the human body (2, 3). A key finding so far has been the observation that a NP in a biological medium rapidly forms a “corona” of adsorbed proteins and other molecules (4, 5). This corona is known to be linked to the transport of the NPs around the body, their uptake by cells, and potential adverse outcomes (6–8). Thus, the prediction of the content of this corona is key to predicting the safety of novel nanomaterials.

Qualitatively, the corona may be separated into two classes of bound protein: “hard” and “soft” (4). The hard corona refers to the proteins strongly bound directly to the surface of the NP, whereas the soft corona is a collection of proteins which are loosely bound on top of the adsorbed protein layer. The content of the hard corona is highly dependent on the material and morphology of the NP, and evolves through the adsorption of proteins upon their diffusion from the bulk, followed by their later desorption and replacement by other proteins over several hours (4, 5). The rate at which proteins desorb is linked to the strength with which they are bound to the NP. The weakly bound proteins are typically smaller and so diffuse faster through the biological medium. Thus, they have a higher rate of collisions with the NP and may feature more in the earlier stages of the corona before being replaced by more strongly-binding proteins which collide less frequently, but do not desorb as rapidly. The phenomenon of protein replacement in the adsorption layer is known as the Vroman effect (9). The chemical composition of the NP and any modifications to the surface directly influence the strength with which different proteins bind to it, for example, titania NPs are known to be strongly hydrophilic and thus bind more strongly to certain classes of proteins (10). Intuitively, the curvature of the surface of the NP also influences the strength of the binding: a flat NP will have more of its surface contacting

a given protein (assuming it is convex around the site of binding), and thus experience a stronger attractive force than a curved NP that makes a point contact. Likewise, at least for spherical and cylindrical NPs, a lower curvature implies a larger radius and thus a larger volume with a stronger van der Waals interaction between the protein and NP (11). We may readily expect therefore that the strength of the binding differs between NPs of the same material but different surface curvatures, i.e., as a function of the geometry of the NP, as is observed in calculations of the binding energy through coarse-grained methods (11, 12). Beyond this, we expect that the curvature of the surface plays a role in defining how tightly proteins may pack on the surface as a purely geometrical effect (13, 14). Indeed, as observed in recent work (15), spherical and cylindrical NPs of the same material may have differing corona compositions, which are attributed to a combination of these two effects. Likewise, spherical NPs of the same material but different radii have been observed to exhibit different corona compositions (16).

Given the relevance of the corona to biological outcomes, it is therefore essential to get insight into the mechanisms of its formation and predict its content. This can be done experimentally through a variety of methods, but these are typically highly sensitive to the exact concentration of proteins used, which may vary significantly between the samples of real biological fluids, and cannot be immediately applied to novel NPs. We therefore turn to theoretical and computational methods to predict both the evolution and the steady state. In principle, this could be achieved by performing a molecular dynamics simulation of the NP immersed in the biological liquid. A brute force solution of the problem is difficult due to the combination of extremely large numbers of atoms and long timescales involved. Atomistic simulations of individual proteins typically require several days to produce a few nanoseconds of simulated time, while the corona may consist of hundreds of proteins and reach equilibrium over several hours to days (16, 17). Thus, to model the actual adsorption-desorption kinetics it is necessary to employ a coarse-grained approach in which the biomolecules are represented by less complex structures that are more amenable to simulation. Even these coarse-grained simulations may be still limited in terms of the amount of computational time taken, and so to probe the dynamics of corona formation on biological timescales one may need to turn to alternate methods, e.g. rate-equation treatments or different simulation techniques (18–26).

Previous work on the formation of the corona based on rate equation treatments has primarily focused on a mean-field (MF) approach in which the rate of incoming proteins is reduced by a factor dependent on the total surface area occupied by proteins but further steric factors are not taken into account. That is, if 50% of the surface area of the NP is occupied by a protein, the binding rate for all incoming proteins is reduced by 50%. This assumption is justified if the proteins are sufficiently deformable that they pack perfectly tightly without any gaps. If this is not the case, then it is straightforward to envision a situation in which the total uncovered area is sufficiently large to allow the binding of further proteins, but is distributed across a number of gaps each of which are too small individually to admit a further protein. In general, proteins cannot arbitrarily adjust their bound configurations and so it is reasonable to assume that some parts of the NP will be uncovered but unavailable for binding. If we consider in particular globular proteins which may be approximated as spherical, then it is clear that there must be gaps between these proteins due to the impossibility to perfectly pack spheres. This leads us to consider a model of random sequential adsorption (RSA) with the addition of desorption and surface diffusion, see e.g. Ref(27) for an overview of models. In this paradigm, the proteins are modelled as hard spheres with well-defined positions on the NP. Intuitively, we may expect this hard-sphere (HS) model to produce different results to the MF approach. Indeed, numerical simulations have empirically shown that the HS model predicts effects arising purely from the geometry of the NP and deviations from the MF predictions for the adsorption of a single species of protein (13). Although these individual works each detail certain aspects of the protein-NP corona formation, none completely describe the realistic case of a mixture of rigid proteins binding to curved surfaces. There is therefore a need to develop a theory capable of accounting for the effects of curvature when multiple species of protein are present to enable a more physically realistic modelling of the corona formation on curved NPs. However, rate equation treatments have a limitation in that the medium in which the NP is immersed may contain a very large number of different protein species, each of which may bind at an essentially arbitrary orientation relative to the NP with an affinity unique to each orientation. Numerical integration of such a large number of

dependent variables rapidly becomes infeasible, while models more complex than the MF typically lack analytical solutions for even a single spherically symmetric protein. Thus, we must consider also computational approaches capable of simulating large periods of time, e.g., kinetic Monte Carlo approaches in which the time between the adsorption and desorption of proteins is not directly simulated.

In this work, we employ a combined analytical and computational approach to investigate the effects of the geometry of the NP on the corona composition in terms of how this alters the packing efficiency and the rate at which proteins adsorb to the surface. We demonstrate that the MF and hard sphere (HS) models predict significant differences in the timescales and numbers of adsorbed proteins for a simple system of three common blood serum proteins previously employed in studies of corona formation on spherical NPs (18, 19). Using an approach based on scaled particle theory as previously proposed for adsorption onto planar surfaces (28), we develop a rate-equation treatment for the HS model that may be numerically integrated to obtain the evolution of the corona on spherical and cylindrical NPs, finding that the geometry of the NP alters the composition of the corona even when the change in binding affinity due to the surface curvature is neglected. We confirm these results using Kinetic Monte Carlo simulations, finding excellent agreement. Our model highlights the importance of geometric effects, binding strength, and the surface diffusion of the bound proteins, and extends previous work for adsorption on planar surfaces by demonstrating how it may be adapted for adsorption to curved surfaces for sets of multiple adsorbing particles.

METHODS

Rate equations

We consider a model of reversible adsorption of proteins to the surface of an NP in which the protein is in physical contact with the surface and occupies an area A_i thus making it inaccessible for other proteins. The surface area of the NP is A_{NP} , and we denote the number of available binding sites for a given protein as $n_i = A_{NP} / A_i$. We define the average number of proteins of type i which are bound to the NP to be N_i and further define the surface coverage of this type to be $s_i = N_i / n_i$. Each protein is further characterised by a rate constant for adsorption, $k_{a,i}$ (in units $s^{-1} M^{-1}$) and desorption $k_{d,i}$ and a bulk molar concentration $[C_i]$. We assume that this bulk concentration remains fixed during the evolution of the corona due to either diffusion of proteins from a reservoir or a sufficiently low concentration of NPs that the effect of binding on the concentration of unbound proteins is negligible.

Once a NP is immersed in a protein solution, the evolution of the resulting corona using a rate-equation treatment is given by

$$\frac{dN_i}{dt} = k_{a,i} n_i [C_i] P_i(\mathbf{N}) - k_{d,i} N_i, \quad (1)$$

where proteins of type i collide with the NP at a rate given by $k_{a,i} n_i [C_i]$ and $P_i(\mathbf{N})$ is the probability that a colliding protein of this type successfully binds to the surface of the NP as a function of the set of bound proteins, denoted by $\mathbf{N} = (N_1, N_2 \dots N_i)$. We further assume that the protein binds to the surface with unit probability if there is sufficient space for it to do so, with any other factor which might prevent binding subsumed into $k_{a,i}$, e.g., the protein only binding in certain orientations. Thus, the probability of binding is determined by the probability that a protein randomly positioned on the surface of the NP does not overlap with other proteins. In the MF approach of Refs.(18, 19), this probability is simply equal to the fraction of the surface area of the NP that is not already covered by proteins,

$$P_i^{\text{MF}}(\mathbf{N}) = 1 - \sum_j \frac{N_j}{n_j}. \quad (2)$$

This function is independent of the identity of protein i and thus the rate of binding of all proteins is reduced by the same amount. Using this definition and changing variables from N_i to s_i we obtain

$$\frac{ds_i}{dt} = k_{a,i}[C_i] \left(1 - \sum_j s_j \right) - k_{d,i}s_i, \quad (3)$$

which is independent of the set of values n_i . Thus, in the MF model, the corona composition defined in terms of the surface coverages depends only on the concentration of unbound proteins and the rate constants for adsorption and desorption. The set of linear first-order differential equations given by (3) can be solved analytically using standard techniques. Here, we simply give the steady-state corona composition¹:

$$N_i(t = \infty) = n_i \frac{k_{a,i} / k_{d,i} [C_i]}{1 + \sum_j k_{a,j} / k_{d,j} [C_j]}. \quad (4)$$

It should be noted that this result is essentially equivalent to the Langmuir isotherm, and that although we have obtained it through a kinetic approach it may also be obtained as the thermodynamical equilibrium. Under this model, if the proteins are reasonably strongly binding or have a high concentration then the occupied fraction of the NP surface rapidly approaches unity, that is, the NP becomes completely covered in proteins with no gaps between them. This simple model is valid if the area occupied by a protein can be deformed to enable this optimal packing. In general, however, proteins possess some degree of rigidity preventing this deformation. Consequently, gaps will likely exist between proteins bound to the surface. In the mean field model, the sum of these gaps represents valid area for the binding of further proteins, even if each gap is individually too small to admit a protein. A further limitation of this model is that it assumes that the coverage of the NP reduces the binding rate of each protein identically. Intuitively, we might expect that small proteins are less effected by the increased coverage due to their ability to fit better into gaps between proteins already present on the surface. To quantify these effects, we investigate a model in which the projections of the proteins are taken to be rigid and the proteins are modelled as spherical. We refer to this as the hard-sphere (HS) model, although strictly speaking it is the projections of the proteins which are hard in the sense they cannot overlap or deform. As we are interested only in the monolayer hard corona, the binding process is equivalent to the two-dimensional random sequential adsorption and desorption model which has been thoroughly investigated for planar surfaces. In particular, it has been shown that the acceptance probability for an incoming particle in this model can be approximated by (28)

$$P_i^{\text{HS}}(\mathbf{N}) = \exp(-\mu_i^{\text{ex}}(\mathbf{N}) / k_B T), \quad (5)$$

where μ_i^{ex} is the excess chemical potential for the insertion of a new protein of type i as a function of the current state, under the assumption that the state is currently in equilibrium. As before, a protein successfully binds to the NP with probability P_i^{HS} and rejected with probability $1 - P_i^{\text{HS}}$. In the case of the adsorption of disks onto a planar surface, it was shown that this assumption is generally quite good even for non-equilibrium states and that the required acceptance probability can be obtained from scaled particle theory (SPT) (28). The probability for a freely-rotating convex particle to have sufficient space to bind is given by

¹Unlike in Ref.(19), this is the exact solution to the steady-state as we have made the approximation that the concentration of unbound proteins remains fixed.

$$P_i^{\text{HS}}(\mathbf{s}) = \left(1 - \sum_j s_j\right) \exp \left[-\frac{A_i \left(\sum_j s_j p_j / A_j\right)^2}{4\pi \left(1 - \sum_j s_j\right)^2} - \frac{\sum_j s_j \left(A_i / A_j + p_i p_j / (A_j 2\pi)\right)}{1 - \sum_j s_j} \right], \quad (6)$$

where A_i and p_i are the area and perimeter of the projection of protein i onto the surface, $\mathbf{s} = (s_1, s_2, \dots, s_i)$. This expression implicitly assumes that the bound particles diffuse on the surface, and we investigate the effects of this assumption later.

In general, the acceptance probability in the HS model is reduced in comparison to that in the MF model, reflecting the fact that an incoming protein is not necessarily able to find a suitably large gap to bind to the NP even if there is sufficient space in total. To illustrate this effect, we consider the special case of a single type of spherically symmetric protein binding to a spherical NP, for which we can assume the projection of the molecule is circular. In this case, Eq.(6) becomes independent of the actual sizes of the protein and NP, but does not reduce to the MF case. In Figure 1 we plot the average time taken to adsorb a new protein, $1/(k_a[C_i]n_iP_i)$ as a function of the surface coverage for both the MF and HS cases, setting $k_a[C_i]n_i = 1\text{s}^{-1}$ such that one potential adsorption occurs per second. It can immediately be seen that the time taken to adsorb a single new protein rapidly increases even at modest surface coverages in the HS model compared to the MF model. Since the formation of the corona likely involves many adsorption events, it follows that even if individual adsorption attempts proceed quicker than the rate of 1s^{-1} considered in this simple example, the overall formation of the corona will proceed slowly once a moderate surface coverage is reached. Consequently, the steady-state may not be reached on experimental timescales unless desorption of bound proteins is very fast and the equilibrium surface coverage is low. A rough estimate of the total collision rate $k_{\text{coll}} = k_a n_i [C_i]$ for a spherical protein to a spherical NP in terms of proteins per unit time can be obtained from diffusion theory (29),

$$k_{\text{coll}} \approx 4\pi (D_{\text{NP}} + D_i) (R_i + R_{\text{NP}}) [C_i] N_A \quad (7)$$

where N_A is Avogadro's constant and D_j is the diffusion coefficient. For a typical protein diffusion coefficient on the order $1\mu\text{m}^2\text{s}^{-1}$ (and neglecting the diffusion of the larger NP), protein concentration $1\mu\text{M}$, with $R_i + R_{\text{NP}} = 20\text{ nm}$, this implies a few hundred proteins collide with a given NP per second. With this collision rate, we may reasonably expect surface coverages on the order of 0.7 in the timeframe of a few hours in the HS model, with a comparable coverage in the MF model reached within seconds.

We hypothesise that we can employ Eq.(6) to estimate the probability for the successful insertion of globular proteins into the corona of curved NPs, provided that we correctly calculate A_i , p_i to reflect the projected area of the protein onto the surface of the NP when evaluating the insertion probability. Under the assumption that this procedure is valid, which we test in this work, the evolution of the corona in the HS model is given by

$$\frac{ds_i}{dt} = k_{a,i} [C_i] P_i^{\text{HS}}(\mathbf{s}) - k_{d,i} s_i, \quad (8)$$

where P_i^{HS} will in general be a function of both the geometry of the NP and of the proteins. We may therefore expect that, unlike the MF case, the corona composition in this model will depend on the geometry of the NP even if the parameters $k_a, k_d, [C_i]$ are kept fixed, since this will alter the values of A_i , p_i used to calculate the value of P_i^{HS} . For a single protein, we can numerically obtain the steady-state surface coverage as a function of the equilibrium constant $K_{eq} = k_a / k_d$ and

concentration, as shown in Fig. 1 and compared to the equivalent prediction for the MF model. It is immediately apparent that this steady-state coverage is much less in the HS model than in the MF, with two key implications. Firstly, we can generally expect fewer proteins to be bound in the HS model, even at very high concentrations or if the binding is particularly strong. Secondly, given an observed surface coverage (e.g. obtained from experiment) and initial concentration of proteins, different equilibrium constants would be obtained depending on which of the two models was employed, as noted in Ref (13).

Binding areas

To apply this model, we must define the area A_i occupied by a protein on the surface of an NP. We consider two morphologies of NP: spherical NPs of radius R_{NP} and rod-like cylindrical NPs of radius R_{NP} and effectively infinite length. Planar surfaces can be obtained as a limit of either of these with $R_{NP} \rightarrow \infty$. We assume that all proteins are approximately spherical and characterised by a radius R_i , and that a bound protein is in physical contact with the NP. We project the protein onto a curved surface as shown schematically in Fig. 2. For a spherical NP, this occupied area is given in terms of

the angle $\theta_m = \sin^{-1}\left(\frac{R_i}{R_i + R_{NP}}\right)$ by

$$A_i = R_{NP}^2 \int_0^{2\pi} \int_0^{\theta_m} \sin\theta d\theta d\phi, \quad (9)$$

such that

$$A_i = 2\pi R_{NP}^2 \left[1 - \sqrt{1 - \left(\frac{R_i}{R_i + R_{NP}}\right)^2} \right]. \quad (10)$$

For a very large NP, the area tends towards πR_i^2 , i.e., the maximum cross-section of the protein.

Conversely, for an NP much smaller than the protein, the number of available binding sites approaches the limiting value of 2, in agreement with the geometrical constraint that only two spheres may touch at a single point. In all cases, the number of binding sites is given by the surface area of the sphere divided by the area occupied by a protein, $n_i = 4\pi R_{NP}^2 / A_i$. For the HS model we also require the projected perimeter of the protein. By reasons of symmetry, it follows that the projection should be circular and thus the perimeter can be obtained from the area by $p_i = 2\pi\sqrt{A_i / \pi}$.

Next we turn to a cylindrical NP. The irreversible adsorption of monodisperse spheres onto a cylinder has been recently investigated in Ref.(14) and demonstrated to be equivalent to the deposition of approximately ellipsoidal shapes onto a planar surface, as depicted in Fig. 3. The exact area of the projection of the protein onto the cylinder is given by integrating the expression in Ref.(14) and re-scaling to restore the dependence on R_{NP} ,

$$A_i = 4R_i R_{NP} \sqrt{\frac{R_{NP}}{R_i} \left(\frac{R_{NP}}{R_i} + 2\right)} \left(E\left(\sqrt{-\frac{1}{\frac{R_{NP}^2}{R_i^2} + \frac{2R_{NP}}{R_i}}}\right) - K\left(\sqrt{-\frac{1}{\frac{R_{NP}^2}{R_i^2} + \frac{2R_{NP}}{R_i}}}\right) \right) \quad (11)$$

where $E(m)$ and $K(m)$ are the complete elliptic functions of the second and first kind respectively and are given in terms of the parameter $m = k^2$.

The area A_i can be used to calculate the number of available binding sites for a cylinder of finite (but large) length L and convert the surface fraction into the number of bound proteins according to $n_i = 2\pi R_{NP} L / A_i$. To apply the HS model, we must take into account the fact that the projections are not freely-rotating and thus Eq. (6) does not apply. Following Refs. (30, 31), the acceptance probability for shapes with a fixed orientation is given by

$$P_i^{HS}(\mathbf{N}) = \left(1 - \sum_j s_j\right) \exp \left[- \left[\frac{A_i \sum_k \sum_j \frac{s_j s_k A_{k,j}^{ex}}{A_j A_k}}{(1 - \sum_j s_j)^2} + \frac{\sum_j \frac{s_j A_{i,j}^{ex}}{A_j} + \sum_j \frac{s_j A_{j,i}^{ex}}{A_j}}{1 - \sum_j s_j} + \frac{A_i \sum_j \frac{s_j}{A_j}}{1 - \sum_j s_j} \right] \right], \quad (12)$$

where $A^{ex} = \frac{1}{2}(A_{ij} - A_j - A_i)$ is the excess excluded area for protein j due to protein i , defined in terms of the co-area A_{ij} which is the area of space in which the centre of protein j cannot be placed without causing an overlap with protein i . For the particular case of spherical proteins binding to a cylinder, for which all the projections are aligned along the same axis, we approximate this area by the area of an ellipse with axes $R_i + R_j, \tilde{R}_i + \tilde{R}_j$, where

$$\tilde{R}_i = R_{NP} \sin^{-1} \left(\frac{R_i}{R_i + R_{NP}} \right), \quad (13)$$

and $A_{ij} = \pi(R_i + R_j)(\tilde{R}_i + \tilde{R}_j)$.

We may demonstrate how the curvature of the NP affects the probabilities for an incoming protein to be accepted into the corona. We consider a simple system consisting of two types of proteins with radii $R_1 = 5$ nm and $R_2 = 2.5$ nm, and in Fig. 4 we plot P_i as a function of s_2 for two values of s_1 on spherical NPs of radius 5 nm. We additionally plot the results obtained for planar NPs in the HS model and in the MF model. In all HS cases, the insertion probability for $s_2 > 0$ is significantly lower than the MF model, and so we may immediately expect the rate of corona evolution to be slower in the HS model. We also observe that the curvature of the NP plays a role, with the spherical NP of radius 5 nm exhibiting a greater acceptance probability for the larger protein than the planar HS model predicts. Conversely, the smaller protein exhibits a lower acceptance probability to the curved NP than to a planar NP if the larger protein is already adsorbed. Thus, larger proteins are not only inserted into the corona more efficiently for curved NPs, but also more effectively screen out smaller proteins. We may therefore expect that the corona content of small, spherical NPs is more strongly biased towards larger proteins than would be predicted for planar NPs. We find very similar effects for cylindrical NPs and do not show these for reasons of space.

It is important to note that, although we have presented the binding areas assuming a spherical protein, the general model proposed here depends only on the projection of the protein onto the surface and not its exact shape. Since Eq.(6) is applicable to freely-rotating convex projections and Eq.(12) to projections of fixed orientation, a wide range of proteins can be modelled using this approach provided that their projected area and perimeter can be calculated and it is reasonable to model their exclusion of other proteins based on their projected area onto the NP. Thus, a complex protein can be treated by estimating binding constants and an associated binding area for each possible orientation relative to the surface of the NP. In this work, we limit ourselves solely to considering spherical proteins to simplify the computational approach described later. We also note that it is the projections of the proteins which cannot overlap, which is a more strict requirement than simply that the proteins themselves do not overlap. This assumption is necessary to allow small proteins to block larger ones and vice-versa (32). Physically, this model of blocking based on occupied area represents that the protein makes contact with the NP over a surface greater than a

single point. In this way, small proteins cannot adhere in the region between a larger protein and the NP which would be permitted in a purely HS model. Likewise, large proteins cannot bind to the NP in regions occupied by smaller proteins by simply making a point contact with the NP, which in reality would not be likely to produce stable adsorption and lead to a very rapid desorption of the protein.

Computational methods

The set of equations in the MF approximation can in principle be integrated analytically to produce the evolution of the corona as a function of concentrations and shapes of the proteins. The more complex HS, however, must be integrated numerically. In practice, we also numerically integrate the rate equations in the MF model for consistency and to simplify the code. We employ the numerical integration routines (“NDSolve”) in Mathematica 12.1 for both HS and MF models (33). For the parameter sets considered here, we find that the default options for the automatic selection of integration methods and step sizes perform sufficiently well that we do not need to manually optimise these. We have found that the numerical integration for the MF rate equations produces results in very good agreement with the analytical solution with absolute errors in the surface coverage of less than 5×10^{-8} for a typical set of proteins, further validating this approach.

To validate the rate-equation approach, we also perform simulations of the adsorption process. As discussed previously, even highly coarse-grained models of corona formation using molecular dynamics take an extreme amount of computational time. Thus, we instead employ a kinetic Monte Carlo (KMC) method (34) that can readily be extended to large numbers of proteins. This algorithm samples a set of events – adsorption or desorption of proteins – advancing from one to the next without requiring the evaluation of the time inbetween them, allowing for the efficient simulation of a large number of events and thus an extended amount of time. At a given time t , the algorithm generates a list of possible events, randomly selects one with a probability weighted by the rate at which that event occurs and the system is updated accordingly. Afterwards, the time is advanced by a random amount given by

$$\Delta t = -\frac{\ln(u)}{k_{tot}} \quad (14)$$

where k_{tot} is the sum of all event rates and u is a random number uniformly distributed in the interval $[0,1]$. This process iterates until the time reaches a pre-determined stopping point. The possible events correspond to proteins colliding with (and potentially adsorbing to) or desorbing from the surface of the NP. Collisions occur with a rate given by $k_{a,i}n_i[C_i]$ and desorption with a rate $N_i k_{d,i}$.

If a desorption event rate is selected, a randomly chosen protein of the specified type is removed from the surface of the NP. Conversely, if a collision occurs, the protein is either accepted or rejected with criteria depending on the model employed. In the MF model, the current occupied surface fraction

$\sum_j s_j$ is calculated, and a random number in the interval $[0,1]$ is generated. If this random number is greater than $\sum_j s_j$, the protein is accepted and added to the state of the system. If not, the simulation

proceeds to the next event. We note that this may allow a temporary “over-saturation” of the surface of the NP by a single protein since it does not test if the space remaining on the surface is sufficiently large to allow for the admission of this protein, but this effect is likely to be significant only for very small NPs in which an individual protein occupies a large amount of the surface of the NP.

In the HS model, the location of each protein is explicitly tracked by generating a location for each colliding protein, where these positions are uniformly distributed on the surface of the NP. In the spherical case, a pair of angles θ_i, ϕ_i are randomly generated to produce a uniform distribution of points on the surface of the sphere. This is achieved by drawing two random numbers u_1, u_2 from the uniform distribution $[0,1]$ and taking $\phi_i = 2\pi u_1, \theta_i = \cos^{-1}(2u_2 - 1)$ (35). A protein of the

corresponding type is then inserted at the Cartesian coordinates given by $(R_{\text{NP}} + R_i)[\cos\phi_i\sin\theta_i, \sin\phi_i\sin\theta_i, \cos\theta_i]$ provided that it does not overlap with any existing proteins. An overlap is deemed to have occurred if the condition,

$$\cos^{-1}(\sin\theta_i\sin\theta_j + \cos\theta_i\cos\theta_j\cos(\phi_j - \phi_i)) < \sin^{-1}\frac{R_i}{R_{\text{NP}} + R_i} + \sin^{-1}\frac{R_j}{R_{\text{NP}} + R_j} \quad (15)$$

is found to hold for any protein j already adsorbed to the NP and where protein i is the new protein. This ensures that the projections of the proteins onto the surface do not overlap, as assumed in the rate-equation model, and is derived by projection of the centre of the two proteins onto the sphere and calculating the minimum angular distance between them allowable without overlap occurring. In both models, regardless of whether the protein successfully adsorbs or not the simulation is then moved to the next timestep. In this way we ensure that the rate of successfully adsorbing proteins is consistent with the rate-equation model.

For the cylindrical NP, the position of the protein is defined by the coordinates ρ_i, ϕ_i, z_i , where $\rho_i = R_{\text{NP}} + R_i$ such that the protein is in contact with the NP and binding occurs only to the curved surface. The angular coordinate ϕ_i is chosen from a uniform distribution in $[0, 2\pi]$. The final coordinate z is chosen uniformly from the interval $[-L/2, L/2]$, where L is the length of the cylinder. To minimise edge effects, we ensure that L is large compared to the typical radius of the proteins and employ periodic boundary conditions at $\pm L/2$. Unlike in the spherical case, we cannot write down a simple expression to test if a pair of proteins overlap due to the more complex shape of the projection of proteins onto the cylinder. Instead, for a pair of proteins of radii R_i, R_j , we determine if overlap occurs by setting the radial co-ordinate for both to $\rho_{i,j} = R_{\text{NP}} + \max(R_i, R_j)$ and determining if this causes a physical overlap between the proteins, i.e. if the distance between their centres is less than $R_i + R_j$. This procedure ensures that a smaller protein cannot bind in the region between a larger protein and the NP, which would be possible if physical overlap was only checked for at their actual location, is consistent with the definition used for the spherical NP, and avoids a computationally expensive test for overlapping based on the projections of the two proteins onto the surface. By projecting the locations of proteins obtained from simulations onto a planar surface as depicted in Fig. 3 we have determined that this test correctly prevents the overlap of the projected areas of proteins for those considered here, although in general this method may fail if the proteins are much larger than the NP. To implement the periodic boundary conditions with respect to the ends of the cylinders, overlap is tested for the incoming protein against the original set of bound proteins and copies of this set translated by $\pm L$ along the axis of the cylinder. For both spherical and cylindrical NPs, these overlap criteria are expected to produce equivalent results to the theoretical expressions Eq. (6) and Eq. (12) respectively, but the KMC program does not evaluate these probabilities when assessing if a protein is accepted or not. Thus, this method provides a means to test the validity of these expressions for the systems considered here.

For the majority of the simulations, we assume that once a protein has adsorbed to the NP it remains fixed in place until it desorbs. It is known, however, that the diffusion of proteins on the surface of the NP increases the packing efficiency (36), and the rate-equation model assumes that the adsorbed proteins are effectively in a fluid phase (28). The desorption and readsorption of proteins produces a restructuring of the surface equivalent to the diffusion, justifying the exclusion of this process. If, however, proteins are sufficiently tightly bound that they do not escape the NP but are still free to diffuse across the surface, then it is of interest to investigate how the surface restructuring impacts the results. To implement this, after each event the proteins are selected in a random order and the angular position of each molecule perturbed by a small random amount. These perturbations are randomly drawn from zero-mean normal distributions with standard deviations given by (37)

$$\sigma_\theta = \frac{2D_s\Delta t}{R_{NP}^2} \quad (16)$$

$$\sigma_\phi = \frac{2D_s\Delta t}{R_{NP}^2} \left(\frac{1}{\theta^2} + \frac{1}{(\theta - \pi)^2} \right) \quad (17)$$

where Δt is the time between the most recently applied event and the event before and D_s is the surface diffusion coefficient, which we set to $10^5 \text{ nm}^2 \cdot \text{s}^{-1}$ for all proteins. A test move is generated and accepted if it does not result in an overlap between proteins and otherwise rejected, with up to four test moves attempted before moving to the next protein, with proteins selected in a random order. Proteins which are not successfully moved are not re-tested following the moving of other proteins, but are tested again on the next timestep. The inclusion of this restructuring requires re-checking for collisions following each trial perturbation and so dramatically slows the simulation. Thus, we do not employ this procedure for simulations except where noted and have implemented it in the code only for spherical NPs. We note that this procedure is not a completely physical representation of diffusion but is employed to produce a restructuring of the adsorbed layer consistent with the effects produced by diffusion while maintaining a reasonable computational cost.

The code used for the KMC simulations is implemented in Python 2.7 and available for download as part of the UnitedAtom software package (38).

RESULTS

In this work, we are primarily interested in determining how the geometry of the NP influences the corona, independently of the strength with which a given protein binds to the surface of the NP. Thus, we first vary the size and shape of the NP while keeping the concentrations and rate constants for adsorption and desorption for each protein fixed and present results comparing the KMC simulations to the rate-equation model. For consistency with the literature, we primarily employ the same system of proteins as in Ref. (18) using the same rate constants and concentrations quoted there as summarised in Table 1. This system consists of three proteins: HDL, HSA, and fibrinogen (Fib). Of these, HDL and HSA are globular and so can be reasonably modelled as spherical. Fibrinogen is a rod-like protein, which can be treated using the analytical model but not in the current implementation of the CoronaKMC program for side-on orientations. For the purposes of this work, we simply assume the binding is achieved through an end-on configuration, such that the binding profile is approximately equivalent to that of a globular protein. The exact identity of these proteins is immaterial as we are focused on the mathematical treatment rather than attempting to match experimental results, and we simply require a range of concentrations and binding constants. Thus, Fib can also be considered to be a large but weakly binding globular protein, with the name used purely to ensure consistency with Ref. (18). Of the proteins considered here, HDL binds significantly more strongly than the other two, with a characteristic lifetime $1/k_d$ on the order of about 9 hours compared to ca. 8 minutes for HSA and Fib. In certain simulations, we also consider two fictional proteins corresponding to one comparable to Fib but with a higher concentration, and one which is a slightly larger form of HDL. To enable a comparison between NPs of different sizes and morphologies, we present all results in terms of the surface coverages s_i , using the relation

$s_i = N_i / n_i$ to convert the numbers of bound proteins observed in KMC simulations to the surface coverage.

First, we investigate the MF model, with results shown for both the rate-equation model and the KMC simulations for spherical NPs of radius both 5 and 50 nm in Fig. 5. The agreement between the results obtained from the rate-equation and KMC simulations is excellent, and it can be seen that the mean surface fraction is independent of the radius of the NP, with HSA rapidly absorbed and then replaced by HDL, and with Fib essentially excluded entirely at all times. The exclusion of Fib is due to the preferential adsorption of the other proteins, since if these are excluded we find $s_{Fib} = 0.90$ in the

steady-state. Although the mean coverages are independent of the radius of the NP, it can be seen that the fluctuations around the mean value are larger for the smaller NP. The coverage for single-particle adsorption on infinite surfaces is known to follow a Poisson distribution (39), for which

$$\sqrt{\langle N_i^2 \rangle - \langle N_i \rangle^2} \propto \sqrt{N_i}, \text{ such that fluctuations in the surface coverage scale as } \sqrt{s_i / n_i}.$$

The true distribution for multi-component adsorption on finite surfaces is of course likely to be much more complex, but heuristically we can expect similar scaling of fluctuations, especially when the steady-state corona consists of essentially only a single protein as in the present case.

Next, we perform simulations for this same set of proteins in the HS model. The results for spherical NPs of size 5 and 50 nm are shown in Fig. 6. It is immediately apparent that this significantly alters both the surface coverage and the rate at which the system approaches equilibrium compared to the MF model. The analytical model using the surface coverage given by Eq. 6 is in good agreement with the numerical results. It can be seen that, unlike in the MF model, the surface coverage in the steady-state depends on the radius of the NP. Since we have left the rate constants unchanged, this implies that this is a purely geometric effect arising solely from the different packing efficiencies of spheres on curved surfaces. Likewise, in Fig. 7 we present a comparison between the numerical simulations and the numerically-integrated rate equation treatment for cylinders of radius $R = 1$ and $R = 5$ nm, again finding that the radius of the NP plays a role in defining the evolution of the corona even when rate constants are not altered. In both cases, the time taken to reach the steady-state is on the order of several hours despite the relatively quick desorption of individual HSA molecules, which occurs on the order of minutes. This is most likely caused by the preferential re-adsorption of HSA into the gaps left by its desorption compared to the adsorption of the larger HDL molecules. As with the MF case, Fib is essentially excluded from the binding.

In Figure 8, we plot the steady-state surface coverage obtained from the rate-equation treatment for HDL as a function of the radius of the NP for both spherical and cylindrical cases. It can clearly be seen that for both the coverage increases as the radius approaches zero due to the increased packing efficiency of spheres around curved surfaces relative to planar surfaces. For both geometries, the selectivity of HDL over HSA decreases as the size of the NP increases. In contrast, the MF model is radius independent and predicts almost complete saturation of the NP in HDL, $s_{\text{HDL}} \approx 0.99$. We note that the total surface coverage also exhibits a slight dependency on the curvature of the NP, but for the protein set considered here this is a very minor effect and the total coverage falls within the range $0.62 - 0.65$. As a measure of the time taken for the corona to evolve, Figure 8 also shows the time at which the surface coverage switches from predominantly HSA to HDL as a function of the radius. In general, it appears that NPs with smaller radii approach this point more rapidly than larger ones, indicating that the corona evolves faster for these high-curvature NPs. Likewise, the effects are more strongly pronounced for the spherical geometry in comparison to cylinder of the same radius, reflecting the higher local curvature of the sphere. Convergence towards the planar limit is very slow in both cases.

With the geometric effects established, we next consider how varying the binding parameters alters the outcome. Certain nanomaterials, for example gold, are known to adsorb proteins so strongly that the binding is essentially irreversible, $k_{d,i} \approx 0$ (11). In this case, we may intuitively expect that the corona primarily represents the proteins which collide with the NP more frequently than others. The rate-equation model proposed previously is in principle still valid if we set $k_d = 0$ for all proteins, however, the probability for successful adsorption of a protein in the HS model assumes a fluid-like phase of bound proteins (28). If proteins cannot desorb, then the bound layer of proteins does not meet this requirement and so we may expect deviations between the simulations and rate-equation model. The fluid-like behaviour can also be obtained by allowing bound proteins to diffuse on the surface of the NP and so we perform simulations including a process based on surface diffusion which enables the layer of bound proteins to restructure. The results in Figure 9 show a comparison between the simulations with (points) and without (open circles) surface diffusion for the protein set in the absence of desorption. Clearly, the rate-equation model remains accurate when surface diffusion is enabled for irreversible adsorption, but fails to accurately describe the results if the adsorbed proteins are fixed on

the surface. Due to the limitations of computational resources, only a small number of simulations were performed with restructuring, but these confirm the fact that the corona has effectively reached its steady state within 360 s. Formally, we expect the rate-equation model in the limit of non-desorbing proteins to produce inaccurate results after an extreme period of time, as it can straightforwardly be seen that a steady-state solution with $k_{d,i} = 0$ can only exist when the surface of the NP is completely covered in proteins. Geometrically, this optimum packing cannot be achieved for spherical proteins. In practice, however, we find that for the set of proteins considered here the numerical integration of the rate-equation treatment provides physically realistic results even for timespans of up to one year, far exceeding experimental timescales.

For small but non-zero desorption rates, we find again that the evolution of the corona is significantly delayed. Even a relatively minor decrease in the desorption rates by a factor of 10 (with diffusion disabled) as shown in Fig. 9 results in the steady-state corona taking several days to form in the HS model. A slightly worse agreement between the KMC and rate-equation models can be observed in the initial stages of the corona formation. Taking into account the results found for irreversible adsorption, it is likely that this is due to the inability of the layer of bound proteins to restructure sufficiently quickly during the initial adsorption to provide the fluid-like properties necessary for the model to hold. We have performed further simulations of the initial stage of formation including diffusion and find a better agreement when this is included.

The adsorption rate constants listed in Table 1 are quite low, corresponding to (on average) less than one potential adsorption event per second for Fib interacting with an NP of radius 10 nm, and even the highly abundant HSA undergoes less than one hundred events per second. It is of interest to investigate how the predicted corona content varies if these adsorption rates are varied, in particular to see if increased adsorption rates alter the steady-state corona composition. Although in principle we may simply scale the values of $k_{a,i}$, it is more physically reasonable (and mathematically equivalent) to instead scale the concentrations of each protein by a global factor while keeping their relative abundances fixed, representing adding more or less solvent during the preparation of the protein solution. The results of numerically integrating² the rate equations for both the HS and MF models with the protein concentrations increased by a factor of 1000 are shown in Fig. 10. In the MF model for high scale factors, there is essentially no effect on the evolution of the corona compared to the results obtained using the original rate constants beyond it reaching the peak coverage of HSA more rapidly. In contrast, however, we observe a significant difference in the HS model in this high concentration regime: the steady state corona is now primarily composed of the more abundant HSA rather than the more strongly binding HDL. Furthermore, the corona still takes a significantly long time to reach this steady state, with equilibrium not reached after 12 hours, although only minor changes take place after this point. Although the underlying model is less physically realistic at such high concentrations, it nonetheless demonstrates that there is a significant change in the corona content caused by this scaling. To further investigate this, we calculate the corona composition at 24 hours for a range of concentration scaling factors, as shown in Fig. 10. The surface coverage of HDL can be seen to vary with this scaling factor in a non-trivial way. At concentrations somewhat lower than the initial set, s_{HDL} is increased, while lowering the concentration even further then causes s_{HDL} to decrease. Even over the wide range of concentrations considered, Fib. does not appear to meaningfully bind to the NP. We have performed equivalent calculations in the MF model and find almost no changes in the steady-state coverages, which we discuss further later.

As a further test of the model, we add two additional proteins to the set originally considered and recalculate the expected corona composition. The first of these two proteins, FP1, is simply the original set of parameters for Fib with the concentration increased to be equal to that of HSA, while the second additional protein FP2 is equivalent to HDL but with a slightly greater radius $R_i = 5.5$ nm in comparison to the original radius of 5 nm. Results are plotted in Fig. 11 for both HS and MF models. In the HS model, it can be seen that even significantly increasing the concentration of Fib is not

²Due to the high rate of adsorption, simulating the corona using the KMC method for the timespans required to reach steady-state for systems with a high concentration of proteins is impractical, even with diffusion disabled.

sufficient to cause it to be expressed in the long-term corona, as although it is initially adsorbed it is rapidly replaced by more strongly adsorbing proteins (HDL, FP2) or those with similar adsorption characteristics but a smaller size (HSA). Indeed, although both HSA and FP1 have nearly identical values of $k_a[C]$ and k_d , it is notable that of the two only HSA is exhibited in the steady-state corona. Meanwhile, FP2 has adsorption kinetics quite similar to that of the original protein HDL, but due to its increased size is less strongly expressed in the corona. Conversely, in the MF model, HDL and FP2 have identical surface coverages as a result of the fact in this model these coverages are independent of the radius of the protein. Thus, the difference in surface coverage observed in the HS model is due to the radius-dependent insertion probabilities. Likewise, in the MF treatment HSA and FP1 exhibit very similar profiles, in contrast to the size effects observed in the HS model.

DISCUSSION

From the results shown in the previous section, it is clear that the timescale for the formation of the corona – both the initial adhesion of proteins to the surface of the NP and the formation of the steady state – varies significantly with the model used to describe how the rate of binding is reduced by pre-existing proteins. At one extreme, if proteins are allowed to move freely around the surface and deform to an effectively arbitrary degree as in the mean field model, the corona approaches equilibrium rapidly. At the other extreme in which the proteins are rigid and fixed in place, the time taken for the corona to reach equilibrium is dramatically increased for the set of proteins considered here. In both cases, the results of the KMC simulations can be adequately explained using analytical models of the adsorption process. Strictly speaking, the model of Ref. (28) used for the HS approach is valid for an equilibrium state, but appears sufficiently accurate for the dynamical states observed here, even taking into account the lack of diffusion in the majority of the simulations. Indeed, in Ref. (28) the adsorption probability is shown to be relatively accurate even in the absence of desorption. Thus, although strictly speaking the processes with and without diffusion are physically different and are expected to show different physics, we find that including either diffusion or desorption in the KMC simulations produces a sufficiently good agreement between these and the rate-equation model to allow their cross-validation.

HS and MF models predict qualitatively similar behaviour for the evolution of the corona – a sharp initial rise in HSA followed by a slow replacement by HDL until the steady-state is reached, although on significantly different timescales. This is in agreement with typical experimental results demonstrating the Vroman effect, in which small, weakly-binding proteins adsorb first and are later replaced by larger proteins which bind more strongly but collide less frequently with the NP (9). We observe an enhancement of this effect for highly curved NPs, i.e., those with small radii, and a stronger effect for spheres than for cylinders. With the set of rate constants used here, the NP corona in the HS model takes several hours to reach equilibrium as a consequence of the decreased rate of adsorption at higher coverages. Experimentally, it has been observed that the corona continues to develop for a period of several hours and so this is not an unreasonable estimate of the equilibration time (16). The difference in timescale for the HS approach compared to the MF approach arises from two main factors. Firstly, the adsorption probabilities in the HS model are in general lower than that of the MF model, and so the rate of adsorption is slowed down. Secondly, the two models behave differently when the corona is close to saturation with proteins. In the rate-equation MF model, adsorption of the larger HDL can essentially always take place as the total coverage $S = \sum_i s_i$ never

reaches unity and the adsorption or desorption of a fraction of a molecule is permitted. Thus, the amount of HDL in the corona can continuously increase and gradually replace HSA. In the KMC implementation of the MF model, adsorption of proteins is only blocked completely when $S > 1$. The desorption of a single HSA molecule is usually sufficient to restore $S < 1$, enabling the subsequent adsorption of the larger HDL even if strictly speaking there is not sufficient room on the surface for it, enabling a stepwise replacement of HSA by HDL. In the HS model, however, adsorption of HDL in the KMC simulations frequently requires the sequential desorption of multiple HSA molecules in order to free up sufficient room on the surface for it to bind. The time scale for the evolution of the corona is therefore set by the rate at which a sufficient amount of HSA can desorb without

replacement by more HSA to permit the adsorption of HDL, which is accurately captured by the HS rate equations. We have tested a form of the MF KMC algorithm in which adsorption of proteins is allowed only if this does not lead to over-saturation and found this leads to a decrease in the rate at which the corona evolves, supporting this interpretation.

The above effect explains the observation that increasing the protein concentrations globally by a factor of up to one thousand (see Fig. 10) both slows the rate at which the corona approaches the steady-state and alters the steady-state corona composition. Increasing the concentration of proteins leads to a significantly faster adsorption of HSA such that desorption of HSA is likely followed by re-adsorption of this rather than adsorption of HDL. Conversely, decreasing the overall concentration of proteins initially favours the adsorption of HDL, as it results in a decrease in the amount of adsorbed HSA while maintaining a sufficiently high rate of incoming HDL to take advantage of this increase in available binding area. As the concentration decreases further, there is no longer a high enough flux of HDL to maintain a high surface coverage, and so s_{HDL} reaches a maximum and then decreases. We next consider the corresponding effects in the MF model, for which we may obtain analytical results. From Eq. 4, if we apply a scaling factor α to each concentration such that $[C_i] = \alpha[C_{i,0}]$ then the equilibrium surface coverages are given by,

$$s_i = \frac{\alpha K_{eq,i}[C_{i,0}]}{1 + \alpha \sum_i K_{eq,i}[C_{i,0}]} \quad (18)$$

where we have defined $K_{eq,i} = k_{a,i} / k_{d,i}$. The ratio of two surface coverages, s_i / s_j , in the MF model is given by:

$$s_i / s_j = \frac{K_{eq,i}[C_{i,0}]}{K_{eq,j}[C_{j,0}]} \quad (19)$$

and is independent of α . By differentiation of Eq. 18 with respect to α , it can also be shown that s_i is a monotonic function of α . Taken together, these two points imply that much of the dependency of s_i on α observed for the HS case cannot be replicated in the MF case: it is not possible for the relative abundances of two proteins to change by making the medium more concentrated or dilute, nor can the coverage of a given protein exhibit a maximum as a function of α . Indeed, for $\alpha \sum_i K_{eq,i}[C_{i,0}] \gg 1$ the

term of unity in the denominator of Eq.18 can be neglected, such that s_i in the MF model is

practically independent of α . For the present system of proteins, $\sum_i K_{eq,i}[C_{i,0}] \approx 16000$, and so α

must be extremely small for any effect to be noticeable. These observations provide a means by which the MF and HS models can be distinguished experimentally: diluting the protein solution will lead to changes of the relative abundance of adsorbed proteins in the HS model but not in the MF model, provided that the concentration of proteins remains much larger than the concentration of NPs to ensure that the model considered here remains valid.

In contrast to the MF model, the HS model predicts that the size of the NP relative to the size of the proteins significantly alters the evolution of the corona. This is especially apparent in the results shown in Fig. 11, where simply increasing the radius of a protein by 0.5 nm significantly reduces the extent to which it features in the corona. Likewise, keeping protein sizes fixed and varying the size of the NP produces a variation in the coverages of specific proteins and the time taken to reach equilibrium as shown in Fig. 8, while the MF model is radius-independent. Over the range of radii considered, there does not appear to be a point at which the two models completely converge for the set of proteins used here. Qualitatively, for spherical NPs of very small radii the HS model predicts results which are somewhat closer to the MF model than obtained for larger radii. For these cases, the crossover point at which $s_{\text{HDL}} = s_{\text{HSA}}$ is reached at around 0.3 hours, comparable to the result

obtained for the MF model of 0.25 hours, and likewise for small NPs in the HS model the steady-state corona consists almost entirely of HDL, as it does in the MF model. However, the actual value of the surface coverage still remains significantly lower at around 0.65 rather than the MF values of close to 1, and we find analytically that the insertion probability in the limit $R_{NP} \rightarrow 0$ does not reduce to the simple form used in the MF model. From Fig. 1, we can predict that the MF and HS models are likely to agree completely only when all proteins present have either an extremely small concentration or bind exceedingly weakly to the NP to ensure the total coverage is on the order 0.1–0.2 at most. A further difference between the two models is that in the MF model the more strongly-binding protein essentially entirely excludes a weaker-binding one, whereas in the HS model a weakly-binding but small protein may co-exist in the steady-state corona with a larger, more strongly binding protein. The question therefore remains as to which model – MF or HS – is more appropriate to describe the formation of protein coronas under biological conditions. In reality, proteins are neither completely deformable as assumed for the MF approach, nor completely rigid as in the HS model, and so we may expect that the actual timescales are not as slow as predicted in the HS model.

A key limitation of the analytical approach – whether MF or HS – is the fact that obtaining the solution to the set of differential equations rapidly becomes difficult as the number of proteins considered increases. For the results shown here, we have included a set of only three to five proteins, for which the equations may be numerically integrated using e.g. Mathematica. Realistic systems, however, may consist of a much greater number of proteins, each of which may bind through multiple orientations. For the MF model at fixed concentrations of unbound proteins as considered here, an analytical time-dependent solution can be straightforwardly obtained using matrix methods, producing solutions which are sums of exponential functions. For a total of N_T types of protein, this requires the diagonalisation of an $N_T \times N_T$ matrix. The computational cost of the numerical integration of the rate equations, meanwhile, scales as $O(N_T^2)$ for the spherical case, since for each protein we must evaluate a sum over all types of protein to obtain $P_i(\mathbf{N})$. In the cylindrical case, the scaling is of order $O(N_T^3)$ due to the presence of the double sum in Eq. 12. The KMC simulations, in contrast, are limited instead by the total number of proteins bound N_b due to the requirement to check an incoming protein for collisions with each pre-existing protein. The number of proteins bound is proportional to R_{NP}^2 , and so overall the computational time can be expected to scale as $O(R_{NP}^2)$. The task becomes significantly more challenging with surface diffusion enabled, since the collision tests there require that every existing protein is tested against every other protein, thus scaling as $O(N_b^2)$, i.e. $O(R_{NP}^4)$. This is manageable for small NPs, but for larger NPs it may be necessary to instead rely on numerical integration with a limited selection of proteins present.

We must also mention the limitations of the current model. We have assumed for the HS model that the proteins can be represented by hard disks with diameters calculated from that of the corresponding hard sphere. It has previously been shown in the RSA model that a true HS model exhibits different dynamics as a result of the decreased ability of smaller proteins to block larger ones, and, conversely, the increased ability of smaller proteins to bind in gaps between the NP and larger proteins (32). However, that work suggests that for proteins a hard disk model is more reasonable, and it is unlikely that proteins make only a single point contact with the NP as the true hard sphere model would require. For the system of proteins considered here, the sizes are sufficiently close that there is not likely to be a significant difference caused by the differences between these two models. A further complication is that protein shapes may deviate significantly from spherical and deform to improve their binding ability, with alternate binding profiles shown to result in increased coverage (40). While not accounted for in the present model, we note that the surface function (Eq. 6) derived from scaled particle theory is in principle valid for an arbitrary convex binding area. It is therefore likely that more complex protein shapes can be modelled by finding the area and perimeter of the projection of the protein onto the surface of the NP, treating different orientations as being effectively different

proteins. This would further allow for a differentiation between different binding profiles, e.g. head-on vs. side-on for ellipsoidal proteins. Here, we do not investigate this further, but the CoronaKMC tool provided in the UnitedAtom package supports proteins with multiple binding profiles by representing these as hard spheres of varying size (38). For rod-like proteins or those with more complex shapes, it will likely be necessary to update this software to be able to detect overlaps between arbitrary projections on the surface of the NP. This, together with testing the analytical approach for these non-globular proteins, is an ideal topic for further research.

The concentration of unbound proteins in this work and the rate at which they diffuse to the NP surface is here assumed to be constant. Previous work in the MF model has indicated that finite rates of diffusion to the surface of the NP influence the dynamics of the corona formation (41), which could be represented here by a decrease in the adsorption rate k_a . Likewise, it is reasonable to assume that the dynamics will be altered if the concentration of NPs $[NP]$ is sufficiently high that substantial protein depletion in the environment occurs, i.e., if $[NP]n_i$ is of the same order of magnitude as the concentration of the protein. The adsorption in this case is governed by the diffusion-limited aggregation. The protein depletion effect can be accounted for in the rate equation model by replacing $[C_i] \rightarrow [C_i](0) - [NP]n_i(t)$, and in the KMC simulations by dynamically updating the concentrations of proteins based on the number currently bound to the NP, i.e. essentially by treating the system via an NVT ensemble rather than grand-canonical one. In this work, we have simply assumed that $[NP] \ll [C_i]$ such that these effects can be neglected. We have performed a preliminary exploration of protein depletion, including the necessary modifications to the KMC simulations, and found that for a spherical NP of radius 20 nm interacting with the present system of proteins these effects can be neglected for NP concentrations of less than $0.1 \mu\text{M}$. A further possible extension would be to include effects such as three-body interactions in the KMC simulations to produce better agreement with experiment as in Ref (22). Our model also does not capture more complex interactions between the proteins and their environment as considered in the dynamic density functional theory of Angioletti et al. (25, 26), which indicates that for an accurate model of corona formation it may be necessary to compute more realistic models of the protein density distribution near a NP, especially for NPs surrounded by a permeable gel. In the present case, however, we operate under the assumption that the NP and proteins in solution are sufficiently mobile that no significant concentration gradients develop.

It is important to note that the HS model, in which the proteins are represented as hard spheres, is conceptually quite similar to mesoscopic molecular dynamics simulations of the formation of protein coronas as in (22, 23, 42, 43). In future work, it would be interesting to compare the predictions of this model to such simulations. However, the timescales predicted for the system to reach equilibrium are significantly longer compared to the timescales for individual events. Here, we assumed that proteins collide with the NP at a typical rate on the order of 1/second, and equilibrium takes about 20000 seconds to reach. To get more realistic estimates of the corona equilibration times, the collision rates from real protein and NP concentrations should be used. On the other hand, it may therefore be possible to simulate only the initial stages of formation of the corona using such simulations and employ the HS model to extrapolate to greater timespans. A similar approach has been employed in Ref (22), where the MF model has been fitted to the results of the simulation to extrapolate to greater time periods than can be achieved through MD simulation. Based on the present work, in which the MF and HS models predict both very different timescales for the corona equilibration based on the same rate constants and concentrations and differing steady-state corona compositions, we conclude that extraction of rate constants by fitting to numerical data (obtained through experiment or simulation) must be done with care to ensure that the correct underlying model is chosen. That is, attempting to fit the MF model to results obtained in the HS model would result in significantly different rate constants to the underlying values, and vice versa. Thus, it is vital that experimental results are interpreted by applying the correct model to obtain physically meaningful rate constants.

CONCLUSION

Based on the random sequential adsorption and desorption paradigm, we have developed a rate equation model for the evolution of the NP protein corona for proteins modelled as hard spheres binding to spherical and cylindrical NPs. We have demonstrated that the geometry of the NP directly influences the composition of the corona through altering the efficiency with which proteins can pack and new proteins can adsorb to the NP, and that this results in geometry-dependent corona compositions even when the rates at which proteins adsorb or desorb from surface sites are kept fixed. We have found that these effects can be explained using scaled-particle theory for both spherical and cylindrical NPs as long as there is either desorption or diffusion of adsorbed proteins to ensure a fluid-like surface layer. Our results pave the way for an improved understanding of experimental results and the ab initio prediction of protein corona compositions.

AUTHOR CONTRIBUTIONS

IR and VL designed the research. IR developed the KMC software, carried out all simulations, analyzed the data. IR and VL wrote the article.

ACKNOWLEDGMENTS

We are grateful for funding from EU H2020 grants SmartNanoTox (grant agreement No. 686098) and NanoSolveIT (grant agreement No. 814572).

REFERENCES

1. Phala, N. S., and E. Van Steen, 2007. Intrinsic reactivity of gold nanoparticles: Classical, semi-empirical and DFT studies. *Gold Bulletin* 40:150–153.
2. Valsami-Jones, E., and I. Lynch, 2015. How safe are nanomaterials? *Science* 350:388–389.
3. Love, S. A., M. A. Maurer-Jones, J. W. Thompson, Y.-S. Lin, and C. L. Haynes, 2012. Assessing nanoparticle toxicity. *Annual review of analytical chemistry* 5:181–205.
4. Kharazian, B., N. Hadipour, and M. Ejtehadi, 2016. Understanding the nanoparticle–protein corona complexes using computational and experimental methods. *The international journal of biochemistry & cell biology* 75:162–174.
5. Monopoli, M. P., C. Åberg, A. Salvati, and K. A. Dawson, 2012. Biomolecular coronas provide the biological identity of nanosized materials. *Nature nanotechnology* 7:779–786.
6. Monopoli, M. P., D. Walczyk, A. Campbell, G. Elia, I. Lynch, F. Baldelli Bombelli, and K. A. Dawson, 2011. Physical- chemical aspects of protein corona: relevance to in vitro and in vivo biological impacts of nanoparticles. *Journal of the American Chemical Society* 133:2525–2534.
7. Walkey, C. D., J. B. Olsen, F. Song, R. Liu, H. Guo, D. W. H. Olsen, Y. Cohen, A. Emili, and W. C. Chan, 2014. Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. *ACS nano* 8:2439–2455.
8. Nienhaus, K., H. Wang, and G. Nienhaus, 2020. Nanoparticles for biomedical applications: exploring and exploiting molecular interactions at the nano-bio interface. *Materials Today Advances*

5:100036.

<https://www.sciencedirect.com/science/article/pii/S2590049819301109>.

9. Slack, S. M., and T. A. Horbett, 1995. The Vroman Effect. *In* Proteins at Interfaces II, ACS Publications, chapter 8, 112–128. <https://pubs.acs.org/doi/abs/10.1021/bk-1995-0602.ch008>.

10. Lundqvist, M., J. Stigler, G. Elia, I. Lynch, T. Cedervall, and K. A. Dawson, 2008. Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. *Proceedings of the National Academy of Sciences* 105:14265–14270.

11. Power, D., I. Rouse, S. Poggio, E. Brandt, H. Lopez, A. Lyubartsev, and V. Lobaskin, 2019. A multiscale model of protein adsorption on a nanoparticle surface. *Modelling and Simulation in Materials Science and Engineering* 27:084003.

12. Lopez, H., and V. Lobaskin, 2015. Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles. *The Journal of chemical physics* 143:243138.

13. Manzi, B. M., M. Werner, E. P. Ivanova, R. J. Crawford, and V. A. Baulin, 2019. Simulations of protein adsorption on nanostructured surfaces. *Scientific reports* 9:1–13.

14. Memet, E., N. Tanjeem, C. Greboval, V. N. Manoharan, and L. Mahadevan, 2019. Random sequential adsorption of spheres on a cylinder. *EPL (Europhysics Letters)* 127:38004.

15. Madathiparambil Visalakshan, R., L. E. González García, M. R. Benzigar, A. Ghazaryan, J. Simon, A. Mierczynska-Vasilev, T. D. Michl, A. Vinu, V. Mailänder, S. Morsbach, et al., 2020. The influence of nanoparticle shape on protein corona formation. *Small* 2000285.

16. Lima, T., K. Bernfur, M. Vilanova, and T. Cedervall, 2020. Understanding the lipid and protein corona formation on different sized polymeric nanoparticles. *Scientific reports* 10:1–9.

17. Casals, E., T. Pfaller, A. Duschl, G. J. Oostingh, and V. Puentes, 2010. Time evolution of the nanoparticle protein corona. *ACS nano* 4:3623–3632.

18. Dell’Orco, D., M. Lundqvist, C. Oslakovic, T. Cedervall, and S. Linse, 2010. Modeling the time evolution of the nanoparticle-protein corona in a body fluid. *PloS one* 5:e10949.

19. Sahneh, F. D., C. Scoglio, and J. Riviere, 2013. Dynamics of nanoparticle-protein corona complex formation: analytical results from population balance equations. *PloS one* 8:e64690.

20. Shao, Q., and C. K. Hall, 2016. Protein adsorption on nanoparticles: model development using computer simulation. *Journal of Physics: Condensed Matter* 28:414019.

21. Zhdanov, V. P., 2019. Formation of a protein corona around nanoparticles. *Current Opinion in Colloid & Interface Science* 41:95–103.

22. Vilanova, O., J. J. Mittag, P. M. Kelly, S. Milani, K. A. Dawson, J. O. Raedler, and G. Franzese, 2016. Understanding the kinetics of protein–nanoparticle corona formation. *ACS nano* 10:10842–10850.
23. Vilaseca, P., K. A. Dawson, and G. Franzese, 2013. Understanding and modulating the competitive surface-adsorption of proteins through coarse-grained molecular dynamics simulations. *Soft Matter* 9:6978–6985.
24. Rabe, M., D. Verdes, and S. Seeger, 2011. Understanding protein adsorption phenomena at solid surfaces. *Advances in colloid and interface science* 162:87–106.
25. Angioletti-Uberti, S., M. Ballauff, and J. Dzubiella, 2014. Dynamic density functional theory of protein adsorption on polymer-coated nanoparticles. *Soft Matter* 10:7932–7945.
<http://dx.doi.org/10.1039/C4SM01170H>.
26. Angioletti-Uberti, S., M. Ballauff, and J. Dzubiella, 2018. Competitive adsorption of multiple proteins to nanoparticles: the Vroman effect revisited. *Molecular Physics* 116:3154–3163.
<https://doi.org/10.1080/00268976.2018.1467056>.
27. Talbot, J., G. Tarjus, P. Van Tassel, and P. Viot, 2000. From car parking to protein adsorption: an overview of sequential adsorption processes. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 165:287–324.
28. Talbot, J., X. Jin, and N.-H. Wang, 1994. New equations for multicomponent adsorption kinetics. *Langmuir* 10:1663–1666.
29. Chandrasekhar, S., 1943. Stochastic problems in physics and astronomy. *Reviews of modern physics* 15:1.
30. Martinez-Raton, Y., 2011. Phase behaviour of mixtures of hard ellipses: a scaled particle density functional study. *Liquid Crystals* 38:697–707.
31. Heyes, D. M., and A. Santos, 2018. Chemical potential of a test hard sphere of variable size in hard-sphere fluid mixtures. *The Journal of chemical physics* 148:214503.
32. Talbot, J., and P. Schaaf, 1989. Random sequential adsorption of mixtures. *Physical Review A* 40:422.
33. Inc., W. R. Mathematica, Version 12.1. Champaign, IL, 2020.
34. Voter, A. F., 2007. Introduction to the kinetic Monte Carlo method. *In* Radiation effects in solids, Springer, 1–23.
35. Weisstein, E. W., 2002. Sphere point picking. <https://mathworld.wolfram.com/>.

36. Ravichandran, S., and J. Talbot, 2000. Mobility of adsorbed proteins: a Brownian dynamics study. *Biophysical journal* 78:110–120.
37. Apaza, L., and M. Sandoval, 2017. Brownian self-driven particles on the surface of a sphere. *Physical Review E* 96:022606.
38. UnitedAtom code repository.
<https://softmattergroup@bitbucket.org/softmattergroup/unitedatom.git>.
39. Adler, S. L., and I. Mitra, 2000. Equilibrium distribution of gas molecules adsorbed on an active surface. *Physical Review E* 62:4386.
40. Maffre, P., S. Brandholt, K. Nienhaus, L. Shang, W. J. Parak, and G. U. Nienhaus, 2014. Effects of surface functionalization on the adsorption of human serum albumin onto nanoparticles—a fluorescence correlation spectroscopy study. *Beilstein journal of nanotechnology* 5:2036–2047.
41. Zhdanov, V. P., and N.-J. Cho, 2016. Kinetics of the formation of a protein corona around nanoparticles. *Mathematical biosciences* 282:82–90.
42. Tavanti, F., A. Pedone, and M. C. Menziani, 2015. Competitive binding of proteins to gold nanoparticles disclosed by molecular dynamics simulations. *The Journal of Physical Chemistry C* 119:22172–22180.
43. Lee, H., 2020. Effects of Nanoparticle Electrostatics and Protein–Protein Interactions on Corona Formation: Conformation and Hydrodynamics. *Small* 1906598.

Figure 1: Left: The average time taken for a new protein to successfully adsorb to a spherical NP for the HS (blue) and MF (red) models as a function of the occupied area of the NP, for an average adsorption time of 1 second for a bare NP and assuming all proteins are the same size and have a circular binding cross-section. Right: The steady-state surface coverage s_i of a protein with a circular binding cross-section on a spherical NP, defined as a function of the binding equilibrium constant K_{eq} and bulk protein concentration $[C]$. Surface coverages are given for the HS (blue solid line) and MF (red dashed line) models.

Figure 2: The projection of a spherical protein (blue) of radius R_i onto an NP (gray) of radius R . The angle θ_m is defined as the angle between the line connecting the centres of the NP and protein and the line tangent to the protein and passing through the centre of the NP. The area shaded in pale blue is assumed to be occupied by this protein and unavailable for binding by other proteins, blocking off a region of the surface of the NP. Under this projection, spherical proteins adsorbed to the NP as shown in the centre image can be mapped to two-dimensional regions on the surface of the NP as shown on the right.

Figure 3: A cylinder with multiple adsorbed proteins shown in 3D (top) and the projection of these proteins onto the 2D ϕz plane (bottom). Positions of the proteins on the cylinder were obtained by kinetic Monte Carlo simulations and their projections in the plane determined following Ref.(14), with the periodic boundary conditions shown by included translated protein projections where necessary.

Figure 4: Probability for a protein of type i to be successfully inserted into the corona of a NP for a two protein system consisting of a large protein $R_1 = 5$ nm and small protein $R_2 = 2.5$ nm as a function of the surface coverage of the smaller protein s_2 , taking $s_1 = 0$ (left) and $s_1 = 0.3$ (right). Results for the HS model for a spherical NP of radius 5 nm, HS model on a planar NP, and the MF model are shown. The MF model predictions are not visible in the scale used in the plot on the right, likewise, the $i = 2$ planar case overlaps with the results for $i = 2, R_{NP} = 5$ nm in the plot on the left.

Figure 5: Evolution of the corona in the MF model, showing the surface fraction occupied by HDL (blue), HSA (red) and Fib (green). The points show the results obtained from the kinetic Monte Carlo simulations for NPs of radius 5 nm (left, 100 trajectories) and 50 nm (right, 50 trajectories), with error bars indicating the standard deviation of these results. The solid lines indicate the analytical prediction obtained by integration of the rate-equation model. The protein Fib. is plotted for consistency but is essentially non-binding.

Figure 6: Evolution of the corona in the hard sphere model on a spherical NP of radius 5 nm (left) and 50 nm (right). The points show the mean surface fraction of the adsorbed proteins from the KMC simulations averaged over 100 runs with the error bars indicating the standard deviation. The solid lines indicate the predictions of the rate-equation model.

Figure 7: Time-dependent surface coverage of proteins on a cylindrical NP of radius 1 (left) and 5 (right) nm in the hard sphere model. The lengths of the cylinders are set to 100 nm with periodic boundary conditions applied. Solid lines indicate the numerical integration of the rate-equation model and points indicate the results of numerical simulations. Error bars indicate $\pm$ one standard deviation as obtained from the simulations.

Figure 8: Left: Steady-state ($t = 24$ hours) surface coverages of HDL as a function of the radius of the NP in the HS model for the spherical and cylindrical NPs obtained via numerical integration of the rate equation model. The asymptotic limit for $R_{NP} \rightarrow \infty$, i.e. a planar NP, is shown as a dashed line. Right: time-point at which the surface coverage crosses over from consisting mainly of HSA to HDL as a measure of the time taken for the corona to evolve.

Figure 9: Left: Time-dependent surface coverage of proteins on a spherical NP of radius 20 nm in the hard sphere model in which the adsorption is taken to be irreversible. Solid lines indicate the rate-equation HS model, solid points indicate the results of simulations with surface diffusion, and open symbols indicate the simulations excluding surface diffusion. 10 runs of the simulation are performed for each case, with error bars indicating the standard deviation and points the mean. Right: As left, except the desorption is only slowed by a factor of 10 relative to the standard set of proteins, diffusion is disabled, and $R_{NP} = 10$ nm, with 100 KMC simulations performed.

Figure 10: Left: Time-dependent surface coverage of proteins on a spherical NP of radius 10 nm in the hard sphere model (solid lines) and mean field model (dashed lines and darker shades) obtained from the numerical integration of the rate equations. Concentrations for all proteins are increased by a factor of 1000 compared to the values presented in Table 1. Right: Surface coverages after 24 hours as a function of a global scaling factor by which the concentration of each protein is multiplied, calculated for a spherical NP of radius 10 nm.

Figure 11: Time-dependent surface coverage of proteins on a spherical NP of radius 10 nm in the hard sphere model (left) and mean field model (right), including additional proteins: FP1, a variant of Fib with a much greater concentration, and FP2, a variant of HDL with the radius increased by 0.5 nm. Lines show the results of numerical integration of the rate equation treatment and points show mean results from 100 trajectories obtained from the KMC simulations, with error bars indicating the standard deviation.

Table 1: Summary of the parameters for the three proteins HDL, HSA and Fib used for the numerical investigation taken from Ref (18). The proteins FP1 and FP2 are fictional proteins based on modifications of the original three proteins to highlight the effects of varying parameters in the model.

	[C] [μ M]	R [nm]	k_a [$\times 10^3 \text{ M}^{-1} \text{ s}^{-1}$]	k_d [$\times 10^{-3} \text{ s}^{-1}$]
HDL	15	5	30	0.03
HSA	600	4	2.4	2
Fib	8.8	8.3	2	2
FP1	600	8.3	2	2
FP2	15	5.5	30	0.03

