

UC Riverside

UC Riverside Previously Published Works

Title

Prediction of protein corona on nanomaterials by machine learning using novel descriptors

Permalink

<https://escholarship.org/uc/item/8cm2z8nx>

Authors

Duan, Yaokai
Coreas, Roxana
Liu, Yang
[et al.](#)

Publication Date

2020

DOI

10.1016/j.impact.2020.100207

Peer reviewed



Published in final edited form as:

NanoImpact. 2020 January ; 17: . doi:10.1016/j.impact.2020.100207.

Prediction of protein corona on nanomaterials by machine learning using novel descriptors

Yaokai Duan^a, Roxana Coreas^b, Yang Liu^b, Dimitrios Bitounis^c, Zhenyuan Zhang^c, Dorsa Parviz^d, Michael Strano^d, Philip Demokritou^c, Wenwan Zhong^{a,b,*}

^aDepartment of Chemistry, University of California, Riverside, CA 92507, United States

^bDepartment of Environmental Toxicology Graduate Program, University of California, Riverside, CA 92507, United States

^cCenter for Nanotechnology and Nanotoxicology, HSPH-NIEHS Nanosafety Center, Department of Environmental Health, Harvard T. H. Chan School of Public Health, Harvard University, 665 Huntington, Boston, MA 02115, USA

^dDepartment of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue 66-570b, Cambridge, MA 02139, USA

Abstract

Effective in silico methods to predict protein corona compositions on engineered nanomaterials (ENMs) could help elucidate the biological outcomes of ENMs in biosystems without the need for conducting lengthy experiments for corona characterization. However, the physicochemical properties of ENMs, used as the descriptors in current modeling methods, are insufficient to represent the complex interactions between ENMs and proteins. Herein, we utilized the fluorescence change (FC) from fluorescamine labeling on a protein, with or without the presence of the ENM, as a novel descriptor of the ENM to build machine learning models for corona formation. FCs were significantly correlated with the abundance of the corresponding proteins in the corona on diverse classes of ENMs, including metal and metal oxides, nanocellulose, and 2D ENMs. Prediction models established by the random forest algorithm using FCs as the ENM descriptors showed better performance than the conventional descriptors, such as ENM size and surface charge, in the prediction of corona formation. Moreover, they were able to predict protein corona formation on ENMs with very heterogeneous properties. We believe this novel descriptor can improve in silico studies of corona formation, leading to a better understanding on the protein adsorption behaviors of diverse ENMs in different biological matrices. Such information is essential for gaining a comprehensive view of how ENMs interact with biological systems in ENM safety and sustainability assessments.

Keywords

ENMs; Protein corona; Descriptor; Machine learning; Nanotoxicology; Modeling

*Corresponding author at: Department of Chemistry, University of California, Riverside, CA 92507, USA. wenwanz@ucr.edu (W. Zhong).

1. Introduction

Increasing applications of engineered nanomaterials (ENMs) in industry, medical care, agri-food and consumer goods raise concerns over the effects of ENMs on living systems (Wagner et al., 2006; Shi et al., 2017; Riehemann et al., 2009; Pirela et al., 2017; Vaze et al., 2019a; Vaze et al., 2019b; Watson-Wright et al., 2017), triggering tremendous research efforts to study how ENMs behave in cells and tissues for the benefits of sustainable developments and safe implementation of ENMs (McClements et al., 2016; Stueckle et al., 2017; Sotiriou et al., 2014; Gass et al., 2013; Gajewicz et al., 2017). While the detailed correlations between the properties of ENMs and their biological outcomes are not yet fully understood (Lynch et al., 2014; Walczyk et al., 2010), one important aspect that has been widely recognized as able to impact the ENM's functions in biosystems is the protein layer adsorbed by the ENMs upon entering a biological environment, termed as the "protein corona" (Van Hong Nguyen, 2017; Corbo et al., 2017; Chen et al., 2017; Konduru et al., 2017; Tsuda and Venkata, 2016). Full surface coverage by the protein corona establishes a new biological identity for the ENM (Monopoli et al., 2012; Caracciolo et al., 2017), which can be seen by other matrix molecules, cells, and tissues; the corona contributes to the biological behaviors of the ENM, including stability, cellular uptake, distribution, immune response, toxicity induction, etc. (Walczyk et al., 2010; Van Hong Nguyen, 2017; Monopoli et al., 2011; Lesniak et al., 2012; Yan et al., 2013; Saha et al., 2014; Lee et al., 2015; Walkey et al., 2014; Liu et al., 2015). Thus, protein corona has become an important property for ENM characterization in order to better evaluate ENMs' biological outcomes. Although comprehensive "imaging" of the protein corona established on a particular type of ENM can be carried out by mass spectrometry (MS), sophisticated sample processing and complex data analysis are required, the speed and throughput of which are not aligned well with the ultrafast pace of ENM evolution (Li et al., 2010a; Smits and Vermeulen, 2016). Therefore, *in silico* methods have been explored to build prediction models for rapid characterization of the protein adsorption behaviors and corona formation (Findlay et al., 2018).

The majority of the previously developed prediction models have focused on the structure-activity relationship (SAR) between corona formation and ENM properties (Ashby et al., 2014; Durán et al., 2015; Clemments et al., 2015; Lundqvist et al., 2011; Natte et al., 2013; Hadjidemetriou and Kostarelos, 2017). In these studies, the physicochemical properties of ENMs, including core materials, sizes, surface charges, and functional groups, were precisely controlled; protein corona composition comparison was carried out by changing one property at a time while keeping the others the same. But some physicochemical properties of ENMs are entangled with each other, making it impossible to change only one without affecting others (Xu et al., 2018). Thus, despite successes in obtaining certain correlations between selected ENMs' properties and the compositions of protein corona, these correlations have limited applicability on more diverse ENMs.

In order to overcome such challenges, inclusion of multiple physicochemical properties of ENMs in one prediction model have also been attempted, in which modern machine learning methods, e.g., random forest and hierarchical clustering, were applied to cope with large-scale and high dimensional data (Findlay et al., 2018; Xu et al., 2018). For example, two properties of ENMs, the size and surface charge, were used as the descriptors of ENMs to

build a quantitative SAR (QSAR) model using random forest classification, while the properties of proteins were used as protein descriptors simultaneously (Findlay et al., 2018). This model could reasonably predict which protein in a complex biofluid (i.e., serum) could be enriched in the corona of the ENMs, however, the descriptors of the ENMs chosen (i.e., size, surface charge) were still a bottleneck, since relatively low contributions of these descriptors to the prediction accuracy of the model were observed (Findlay et al., 2018).

On the other hand, numerous kinds of modern descriptors for ENMs, including detailed structural and chemical properties computed by quantum mechanical calculation, have been proposed (Vaze et al., 2019b; Smits and Vermeulen, 2016; Findlay et al., 2018). However, they have shown to be successful at predicting the interactions of ENMs with small molecules, but not with proteins (Lynch et al., 2014; Apul et al., 2012; Mahmoudi, 2018). Proteins are much larger and more complex than small molecules, suggesting that the interactions between proteins and ENMs are not based on a single factor, but rather are established on the cooperative effect of multiple factors, including diverse types of interaction forces between the surface groups on the ENMs and the amino acid residues located at the ENM-protein binding interface. Moreover, proteins could undergo conformational changes upon interactions with ENMs, further increasing the difficulty to predict by models (Shang et al., 2007; Deng et al., 2011; Dominguez-Medina et al., 2016). Correspondingly, to improve prediction performance, these models require ENM descriptors that can represent the coordination and spatial distribution of different interaction forces with proteins, rather than focusing on one or two individual factors. This issue has been noted for the study of protein-ligand interactions with 3-dimensional (3-D) QSAR models, which have been developed to incorporate the spatial distribution of different factors (Kubinyi, 1997). However, it is very challenging, if not impossible, to routinely obtain the detailed surface structure of ENMs, which are prerequisites for the application of the 3-D QSAR models to study the interactions between ENMs and proteins.

One potential solution to characterize the overall effect of all spatially distributed factors on the surface of ENMs, is to employ suitable probes that can map the surface properties of ENMs with measurable responses and then use such responses as the descriptors of the ENMs. For instance, the biological surface adsorption index (BSAI) system was developed to use small molecules with diverse physicochemical properties as probes, and measure their adsorption coefficients on the ENM as the descriptors of the ENM's surface chemistry (Xia et al., 2010; Chen et al., 2014). However, small molecules are very different from proteins, and therefore exploiting the interactions between small molecules and ENMs to describe protein adsorption on ENMs is unsuitable. In contrast, the application of proteins, instead of small molecules, as probes could be advantageous: the complex interaction between an ENM and each protein probe could be evaluated as an integrated entity, and subsequently could be used to represent the interactions of all other proteins with similar structural properties. If a set of proteins with reasonable diversity are selected as probes, a prediction model with decent generality could be obtained for the protein corona, which can be established via adsorption equilibria of heterogeneous proteins with different binding affinities and kinetics.

Conversely, no endeavor has successfully utilized proteins as probes, partially due to the lack of methods that can screen such interactions rapidly, robustly and economically. Common methods that measure protein-ENM interactions, including analytical ultracentrifugation, capillary electrophoresis, size exclusive chromatography, surface plasma resonance, quartz crystal microbalance, isothermal titration calorimetry, enzyme-linked immunosorbent assay, etc., involve lengthy procedures, technically demanding operations, and/or expensive reagents (i.e., antibodies) (Casals et al., 2011; Cedervall et al., 2007; Li et al., 2010b; Hoshino et al., 2012; Yonamine et al., 2012). Recently, our group developed a method for high-throughput screening of protein-ENM interactions with the fluorogenic dye, fluorescamine (Ashby et al., 2015; Duan et al., 2017). The non-fluorescent fluorescamine rapidly reacts with primary amines on protein surfaces and becomes fluorescent. Because the interaction with the ENM could either block the primary amines on the protein surface from reacting with fluorescamine, or expose more amines to the surface by inducing a protein conformational change, the resultant fluorescence would differ before and after the protein binds to the ENM. We have proven that the fluorescence change (FC) of a set of selected proteins upon binding to various types of ENMs could differentiate the ENMs by their size or surface property and reflect binding strength (Ashby et al., 2015; Duan et al., 2017). The close relationship between FCs and protein-ENM interactions yields FCs as a potentially new category of ENM descriptors to help build an accurate model for the prediction of the protein corona formed in biological matrices.

Herein, a series of proteins with an assortment of properties, were selected to interact with various types of ENMs. After the FCs were measured, a prediction model with machine learning algorithm was established (i.e., random forest classification or regression) using the FCs as the descriptors for the ENMs in addition to their physicochemical properties. We demonstrated that the inclusion of the FCs as the ENM descriptors could improve the prediction accuracy for protein corona composition. We also validated the power of the model by predicting the composition of corona formed on different ENMs in a biological matrix.

2. Materials and methods

2.1. Synthesis and characterization of ENMs

In this study, a panel of 22 ENMs were used (Xia et al., 2010; Chen et al., 2014; Casals et al., 2011; Cedervall et al., 2007; Li et al., 2010b; Hoshino et al., 2012; Yonamine et al., 2012; Ashby et al., 2015; Duan et al., 2017; Zimmerman et al., 2019; Ahn et al., 2018; Beltran-Huarac et al., 2018; Righettoni et al., 2010). Specifically, 15 metallic, 3 cellulose-based, and 4 two-dimensional ENMs were used for this work. The synthesis and characterization for each material are described below. These ENMs are part of the reference ENM repository established at Harvard as part of the Nanotechnology Health Implications Research Consortium established by the National Institute of Environmental Health Sciences (NIEHS).

2.1.1. Metallic ENMs—Fifteen metallic ENMs of near-spherical shape were used in this study. CuO, TiO₂ (P25), ZnO were acquired from Sigma Aldrich, Acros Organics, and

Meliorum Technologies, Inc., respectively. V_2O_5 and ZnS NPs were both obtained from NanoShel LLC. The synthesis and characterization of the not-commercially available ENMs have been described by the authors elsewhere. In brief, citrate-capped Au ENMs, with a nominal diameter of 15 nm, were prepared using the Turkevich method as described by Zimmerman et al. (2019). Citrate-capped Ag ENMs, with a nominal diameter of 15 nm, were synthesized following a hydrothermal method as described by Ahn et al. (2018). CeO_2 (10 and 30 nm), SiO_2 , 1%, and 10% Ag-doped SiO_2 (15, 11, and 8 nm, respectively), Al_2O_3 (30 nm), Fe_2O_3 (10 nm), and Ag ENMs (20 nm) were synthesized using high-precision flame spray pyrolysis (FSP), as previously described in details by Beltran-Huarac et al. (2018). Tungsten oxide (WO_3) was also synthesized by FSP, following a protocol introduced by Righettoni et al. (2010) with diethylene glycol monobutyl ether and ethanol as the solvents, and ammonium tungstate hydrate as the metal-containing precursor. All chemicals used in ENM preparation were purchased from Sigma-Aldrich and used without further processing. FSP took place using an open flame with the spraying conditions optimized to yield particles with desired diameters, while the content of oxygen and the molecular ratio of the precursor were adjusted in order to minimize the residual organic species on the surface of the particles.

2.1.2. Cellulose-based ENMs—Details on the exact synthesis parameters of all CNF-50, CNF-80, and CNC-250 have been previously published by Pyrgiotakis et al. (2018). In brief, cellulose nano-fibrils with nominal, single-fibril diameters of 50 and 80 nm (CNF-50 and CNF-80, respectively) were prepared by the controlled mechanical disintegration of bleached softwood fibers using an ultra-fine friction grinder. Cellulose nano-crystals with nominal length and diameter of 250 nm and 25 nm, respectively, were synthesized by the hydrolysis of the same starting material using 72% w/w H_2SO_4 .

2.1.3. 2D ENMs—Graphene, hexagonal boron nitride (hBN), and molybdenum disulfide (MoS_2) sheets were prepared by the authors using a liquid-phase exfoliation synthesis of the respective bulk materials in the presence of Na-cholate, similarly to what has been previously described for the graphene-based materials by Yi and Shen (2015). The reduced graphene oxide (rGO) used in this study was prepared following a two-step approach: first, endotoxin-free graphene oxide (GO) was synthesized as described by Parviz and Strano (2018) next, GO flakes were reduced in the presence of ascorbic acid, similar to what has been described by Fernández-Merino et al. (2010).

2.2. Characterization of ENMs

Following their synthesis or procurement, all ENMs employed in this study were characterized using various state-of-the-art analytical techniques and instruments. In brief, transmission and scanning electron microscopy (SEM and TEM, respectively) were used for size and shape measurements. X-Ray diffraction (XRD) was used to study the crystal structure of powders; N_2 adsorption to measure particle density and surface area according to the Brunauer–Emmett–Teller (BET) theory; and X-Ray photoelectron spectroscopy (XPS) and attenuated total reflection Fourier-transform infrared spectroscopy (FTIR) to study their surface chemistry. The metal trace purity of the synthesized particles was measured by inductively coupled plasma mass spectrometry (ICPMS); and the residual carbon species

were determined by quantifying the elemental and organic carbon on the particles. It is also worth noting that the biological sterility of ENMs was assessed following the U.S. Pharmacopeia protocol for sterility (WHO document QAS/11.413) and their endotoxin load was assessed by using the Recombinant Factor C (rFC) assay with the Lonza PyroGene® kit following the manufacturer's instructions. For the characterization of 2D ENMs, Raman spectroscopy was employed to record the unique vibrational modes of each material; ultra-violet visible spectroscopy (UV-Vis) was measured to acquire the extinction spectra of 2D ENMs in suspension; and atomic force microscopy (AFM) was carried out to evaluate the thickness and number of sheets. Finally, single particle tracking (SPT) was used to obtain the hydrodynamic size distribution of the 2D ENMs in water and their mean lateral size was then calculated by applying an empirical model developed by Lotya et al. (2013).

2.3. ENM dispersion preparation and colloidal characterization

The dispersion of metallic ENMs in deionized water was performed following a detailed dispersion protocol by the authors (Cohen et al., 2014; DeLoid et al., 2014; Cohen et al., 2018; DeLoid et al., 2017). Briefly, each ENM powder was added to deionized water at a final concentration of 0.5 mg/ml and the particles were then sonicated in a cup-horn sonicator until their respective critical sonication energy (DSE_{cr}) had been delivered (i.e.: until there was no considerable (< 5%) change in their z-average, as measured by dynamic light scattering (Zetasizer Nano ZS)). At this point, the particles' ζ -potential was also measured using folded capillary cells in the same instrument. All suspensions of metallic ENMs in deionized water were then diluted to a final concentration of 0.1 mg/ml and used as described below.

The dispersion of cellulose nano-fibrils and nano-crystals in deionized water were performed according to a protocol provided by Bitounis et al., 2019. In brief, high-speed vortexing was used to disperse the fibrils and crystals at 0.5 mg/ml. The z-potential values of the dispersed cellulose nanomaterials (CNMs) in deionized water were obtained as described above. The as-prepared suspensions were then diluted to a final concentration of 0.1 mg/ml with the addition of deionized water and were used as described below. Finally, 2D ENMs were synthesized in suspension and only had to be manually agitated before use. In particular, the as-synthesized suspensions were shaken until any deposited material on the bottom of the container was fully resuspended in the liquid volume. Their colloidal characterization was performed using SPT (NanoSight, Malvern) and their z-potential was measured using the method described above. Their final concentration before being used in this work was adjusted at 0.1 mg/ml with the addition of deionized water.

2.4. Chemicals and reagents

Proteins, fluorescamine, SYPRO Orange, urea, ammonium bicarbonate, 1,4-dithiothreitol (DTT), iodoacetamide (IAM), formic acid (FA), trypan blue, and pooled human serum were purchased from Sigma-Aldrich (St. Louis, MO). Ultrapure water with electric resistance > 18.2 M Ω was produced in-house by the Millipore Milli-Q water purification system (Billerica, MA).

2.5. Fluorescamine labeling at different temperatures

The ENMs at a concentration of 0.1 mg/ml were mixed with 0.2 mg/ml of protein in 1 × PBS buffer (pH 7.4) and incubated at 37 °C for 1 h. The mixture was then incubated at 37, 60, or 80 °C for 5 min. After the mixture was quickly cooled to room temperature, an aliquot of fluorescamine in acetone was added to the solution at a final concentration of 1 mM. The solution was then diluted 10-fold with 1 × PBS and its fluorescence was measured in a Victor II plate reader without removing the unbound proteins.

2.6. Thermal stability screening

The same incubation step (stated above) was conducted by mixing 0.1 mg/ml of ENMs and 0.2 mg/ml of the protein at 37 °C for 1 h. Next, SYPRO Orange dye was added to the solution at a final concentration of 4×. Then, the mixture was transferred into the CFX Real-Time PCR instrument (Bio-Rad) and subjected to a temperature gradient increasing from 37 to 98 °C, with an incubation period of 20s at each temperature before recording the fluorescence intensity. The excitation wavelength was 488 nm, and the range for the emission filter was 515-545 nm.

2.7. Serum protein corona identification

The ENMs at 0.1 mg/ml in 1× PBS were mixed with the same volume (20 μL) of human serum, and the mixtures were incubated for 1 h at 37 °C. Centrifugation at 15,000 ×g for 15 min was used to pellet the ENMs carrying the adsorbed proteins. After two wash cycles with 1 × PBS, 8 M urea in 50 mM ammonium bicarbonate was used to resuspend the ENMs. DTT was added into the solution at a final concentration of 5 mM and incubated for 40 min. at 56 °C. The solution was cooled to room temperature before 10 mM IAM was added. Then 50 mM of ammonium bicarbonate was used to dilute the solution 8 folds, and trypsin was added to digest the proteins at a trypsin: protein mass ratio of 1:50. The ENMs were removed by centrifugation at 15,000 ×g for 15 min. After being lyophilized and desalted, the resultant peptides were injected into the Waters CapLC system, which was connected to a Finnigan LTQ MS with a nano-ESI ion source. Collision induced dissociation (CID) was used for fragmentation, and the mass range was set to 300–2000 Da. MSGF+ was used to search against the human or rabbit proteome downloaded from UniProt. Reversely ordered protein sequences were used as decoys, and false discovery rate (FDR) was set to 0.01. Spectra counting (SC), as a label-free semi-quantitative method, was used to calculate the relative abundance (RA) of the corresponding protein *i* with Eq. (1):

$$\%RA_i = \frac{SC_i}{\sum SC} \times 100\% \quad (1)$$

The similarity or overlap of the protein corona between two ENMs (a, b) was calculated with Eq. (2):

$$Similarity = \sum \min(RA_a, RA_b) \quad (2)$$

2.8. Correlation and clustering

The FC values were calculated using Eq. (3), in which $F_{\text{protein-ENM}}$ and F_{protein} represent the fluorescamine labeling intensity of each protein-ENM mixture and the protein itself, respectively. The influence of ENMs on fluorescence, e.g., quenching during of the inner filter effect (IFE), was evaluated by the ratio of $F'_{\text{HAS-ENM}}$ and F_{HSA} , the fluorescence of the prelabeled HSA incubated with ENMs and that of HSA labeled by fluorescamine, respectively. This ratio was then used as the correction factor in our method to eliminate the IFE caused by ENMs.

$$FC = \frac{F_{\text{protein-ENM}}/F_{\text{protein}}}{F'_{\text{HAS-ENM}}/F_{\text{HSA}}} \quad (3)$$

The FC of the standard proteins used in the fluorescamine screening of all ENMs tested were put into one individual array. Similarly, the relative abundance of each protein in the protein corona of one type of ENM (i) was considered y_i , and the values for all ENMs were combined into another array (Y). Pearson's correlation coefficient (r) between FC and Y was calculated by Eq. (4):

$$r(FC, Y) = \frac{\sum (FC_i - \overline{FC})(y_i - \bar{y})}{\sqrt{\sum (FC_i - \overline{FC})^2 \sum (y_i - \bar{y})^2}} \quad (4)$$

2.9. Machine learning model for serum protein corona prediction

The isoelectric point (pI), molecular weight (Mw), grand average of hydropathy (GRAVY), and percentage of negative/positive/aromatic amino acids, were used as the descriptors for each protein. The fluorescence changes of each ENM with all standard proteins measured at 37, 60, 80 °C were used as the descriptors for ENMs. The RA of each protein (i) in the corona of ENMs (RA_i^{NM}) was compared to that in the serum control (RA_i^{serum}), and the abundance change (AC) of protein (i) was calculated in Eq. (5):

$$AC_i = \log_2 \left(\frac{RA_i^{NM}}{RA_i^{serum}} \right) \quad (5)$$

Student's *t*-test was performed to test if AC was significantly different from 0 (i.e.: no change). For the classification model, the proteins with AC larger than 0 and *p*-value smaller than 0.05, were considered as enriched in the protein corona (i.e., "positive" or "1"); while the proteins with AC smaller than 0 and *p*-value smaller than 0.05, were labeled as decreased in the protein corona (i.e., "−1"); and all other proteins were classified as no change in the protein corona (i.e., "0"). For the regression model, the AC values were used as targets or dependent variables. Random forest was used for both classification and regression models. The running environment included python 2.7, scikit-learn v0.19.1, NumPy v1.15.0, and Pandas v0.23.4. The minimum number of samples in each leaf node was set to 3. One thousand trees were grown for the bootstrap. The data was randomly split into two sets:

training set containing 80% of the data, and testing set encompassing the other 20%. A 5-fold cross validation was performed during the training of the model.

3. Results and discussion

3.1. Properties of synthesized and procured ENMs

This study employed diverse ENMs, including metal and metal oxide nanoparticles (NPs), cellulose-based ENMs, and 2D semi-conducting materials. Table S-1 summarizes the hydrodynamic diameter and zeta-potential properties for the ENMs used in this study, which were measured after the materials were dispersed in deionized water at 0.5 mg/ml following the protocol reported previously by the authors (DeLoid et al., 2017). Details on the characterization of the Fe₂O₃, CeO₂, SiO₂, Ag doped SiO₂, Al₂O₃, and Ag NPs were reported elsewhere (Beltran-Huarac et al., 2018); characterization of the citrate-capped Ag and TiO₂ NPs can be found in the work by Ahn et al. (2018), while that of the citrate-capped Au ENMs in Zimmerman et al. (2019). Complete morphological and physicochemical summary reports of the CuO, V₂O₅, WO₃, ZnS, and ZnO NPs are shown in Tables S-3 to S-7, respectively. Most of these NPs were smaller than 50 nm in diameter, except for SiO₂, ZnS, and V₂O₅ which were around 100–300 nm. Several of the metal oxides (i.e.: ZnO, CeO₂, Fe₂O₃, and Al₂O₃) exhibited positive ζ -potential when dispersed in water, ranging from 19 mV for ZnO to 65 mV for the 30 nm CeO₂.

Table S-1 includes the dimensions and surface potential of the cellulose-based ENMs: CNF-50, CNF-80, and CNC-250. Characterization of these ENMs has previously been reported by Pyrgiotakis et al. (2018) the mean single fibril diameter for CNF-50 and CNF-80 were 64 ± 29 nm and 78 ± 25 nm, respectively whereas the mean crystal length and diameter for CNC-250 were measured at 267 ± 91 nm and 25 ± 9 nm, respectively.

As for the 2D materials, their lateral dimensions and surface charges are also reported in Table S-1. Complete summary reports of the morphological and physicochemical characterization of graphene, RGO, hBN, and MoS₂ are presented in Tables S-8 to S-11, respectively. In brief, AFM proved that these ENMs are mostly organized in single or few-layer sheets. The extrapolation of the average lateral dimensions from the SPT analyses suggested that graphene, rGO, hBN, and MoS₂ had a lateral dimension of 109 nm, 411 nm, 149 nm, and 428 nm, respectively.

Although such characterizations of ENM dispersed in the protein solutions were not performed because of technical limitations, it has been established that ENMs sonicated at DSEcr and immediately dispersed in protein-rich media are expected to form stable suspensions for at least 24 h due to steric stabilization of the particle agglomerates (Cohen et al., 2013).

3.2. Fluorescence changes being the descriptors of ENMs

Our previous works have confirmed that the fluorescence signal from fluorescamine labeling of a protein can change after the protein interacts with ENMs, like polystyrene, silica, and iron oxide nanoparticles with diameters ranging from 10 to 100 nm (Ashby et al., 2015; Duan et al., 2017). We proved that the FCs were indicative of the NP's aptitude in protein

interactions and closely correlated with the NP's physiochemical properties (Ashby et al., 2015; Duan et al., 2017). Still, for the FCs to be effective ENM descriptors to build corona prediction models, they should be universally measurable on diverse types of ENMs and should preserve close correlations with the ENMs' properties. Thus, in the present work, we acquired a good collection of ENMs from the HSPH-NIEHS Nanosafety Center, including metallic NPs, cellulose nanofibrils (CNFs) and nanocrystals (CNCs), two-dimensional (2D) ENMs, and a series of SiO₂ NPs doped with different amounts of Ag (Table S-1). This collection of ENMs spanned a wide range of core compositions, sizes, shapes, and aspect ratios. These ENMs cannot be used to build conventional prediction models because they require well-controlled changes in only one or two physicochemical properties. In addition, a total of 11 proteins were judiciously chosen to represent the abundant proteins found in human serum (i.e., serum albumin, transferrin, and γ globulin (γ G)) as well as those with a wide range of molecular weights (Mw), isoelectric points (pI), and hydrophobicity (GRAVY) (Table S-2).

Each protein was incubated with one type of the ENMs at a time, in a 2:1 mass ratio, for 1 h, which was found to be long enough to establish the binding equilibrium in our previous works (Tenzer et al., 2013). Following a short 5-min incubation at three different temperatures (37, 60, and 80 °C), the protein mixture was labeled by fluorescamine (Ashby et al., 2015; Duan et al., 2017). One advantage of this labeling method was the dispensability of the washing step, since the bound proteins were labeled by a different amount of fluorescamine compared to the unbound ones, due to surface blockage and unfolding. Moreover, the influence of the ENMs on the fluorescence signal was accessed and subtracted (Fig. S-1). Hence, the FC should represent different interactions between proteins and ENMs. As demonstrated in our previous work, the external pressure from heating could alter the FC values measured in our screening method, because protein-ENM interactions reduce or enhance protein thermal stability, thereby increasing the impact of interaction on protein conformation (Duan et al., 2019). Such phenomena were also observed with the ENMs employed in this work (Fig. S-2). As a result, larger FC values could be more valuable descriptors to discriminate different ENMs in modeling. This not only increases the predictability of the resulting model, but also minimizes the activity cliffs (Danishuddin; Khan, A. U, 2016). The resulting FCs measured for two proteins, with similar Mw (14.3 and 14.1 kDa), but different pI values (11.3 and 4.5), lysozyme and lactalbumin, after incubation with individual ENMs are shown in Fig. 1. FCs for the remaining proteins are displayed in Fig. S-3. Each protein displayed distinct labeling profiles with the ENMs, agreeing with our previous work (Duan et al., 2017).

In order for the FCs to be good descriptors of ENMs in modeling, they should be reflective of the ENMs' properties. Thus, we checked whether the FCs could differentiate different ENMs using principal component analysis (PCA). The scatter plots, using the first two principal components (PC), are shown in Fig. S-4. Interestingly, using FCs at either temperature, ENMs can be separated based upon their shapes (e.g. spherical ENMs, 2D ENMs, and cellulose based nanofibrils). These results are corroborated by our previous finding that FCs are correlated with the physical properties of ENMs (Ashby et al., 2015; Duan et al., 2017). The separation also suggested that the shape of the ENMs could constrain the applicability domain of FC usage for QSAR modeling. However, protein

corona compositions are evaluated based on individual ENMs, so superior discrimination effects of FCs are desired for them to serve as descriptors of ENMs. To represent the discriminability of the FCs, the averaged Euclidean distance between any two pairs of ENMs on the PCA plot were calculated. We found that, the FCs at 37 °C gave out an average Euclidean distance of 0.32, but the value for those acquired at 60 and 80 °C increased to 0.42 and 0.67, respectively, indicating higher discrimination power. The increase of averaged Euclidean distance was not caused by outliers, but by the overall shifted distributions (Fig. S-5). The contributions of the FCs of different proteins to ENM discrimination changed at elevated temperatures, as their component coefficients for PC1 and PC2 varied substantially among the three temperatures, confirming the benefit of obtaining FCs across a wide temperature range. Indeed, the best discrimination of all ENMs tested was obtained using the FCs at all three temperatures, resulting in an averaged Euclidean distance of 0.83 (Fig. S-4D). For some ENMs that were reported as dissolvable, including ZnO and CuO NPs, the ratio of dissolved ions was quantified by ICP-MS (Fig. S6). There was negligible dissolution for either ENM after 1 h incubation in 1× PBS. Thus, the NPs remained intact during FC measurement and corona study.

We also attempted to correlate ENM properties with FCs. Due to the diverse properties of the consortium ENMs, systematic investigation of size-, surface charge-, or other well defined properties was not possible. Therefore, we evaluated Pearson's correlations between any of the two ENMs using the FCs obtained from all proteins tested (Fig. S-7A). Some 2D ENMs, including the reduced graphene oxide (RGO), vanadium pentoxide flakes (V_2O_5) and molybdenum disulfide (MoS_2), showed relatively low correlation coefficients (< 0.5) with any other ENMs; ENMs with at least one property similar between each other, like CNF-50 and CNF-80, which share the same core composition and comparable diameter but different lengths, exhibited a relatively high correlation coefficient (> 0.75). Moreover, the correlations between the size/ ζ -potential of ENMs and FCs were evaluated (Table S-12), and the size of the ENMs showed a correlation coefficient larger than 0.5, suggesting the close relationship between the FCs and the property of ENMs, particularly the size.

Lastly, we tested the correlation between the FCs obtained by pairing two proteins randomly (Fig. S-7B), since highly correlated variables are not useful for modeling. Low correlation coefficients (< 0.5) were observed for most of the protein pairs' FCs, suggesting that the proteins chosen in this study can provide unique FCs to be used as the descriptors of ENMs for modeling. Although relative high correlations (> 0.5) were observed for the FCs of few protein pairs, FCs obtained from all proteins tested were kept for modeling to maximize the multiplicities of the proteins.

3.3. FCs correlated with protein corona compositions

After confirming the capability of FCs as suitable descriptors for the ENMs, correlations of FCs with the abundance of individual proteins found in the ENMs' protein coronas were examined. We incubated the ENMs with human serum for 1 h at 37 °C to form a stable protein corona. The corona proteins were identified by LC-MS/MS, and the relative abundance (RA) of each identified protein was calculated using Eq. (1) as the individual component of the corona profile. As shown in the clustergram (Fig. S-8), the protein profile

in the corona of each type of ENMs was obviously different from that of the serum control: abundant serum proteins, such as transferrin, were diminished in the corona, while hydrophobic proteins, including apolipoprotein A, were enriched. To better express the overall differences, the global similarity of the protein corona profiles was calculated with Eq. (2) and is shown in Fig. 2. For a majority of the ENMs, the protein corona shared a similarity < 40% with the serum control, suggesting that adsorption of the corona proteins by the ENMs was not simply due to the relatively higher abundance of specific proteins in serum. Instead, the adsorption was controlled by the properties of the ENMs and proteins. Moreover, ENMs with similar physicochemical properties shared relatively higher similarity in their protein corona compositions and were clustered closer together in terms of correlation distance (Fig. 2), such as the two cerium oxide NPs with different sizes (CeCO₂-10 and CeCO₂-30), the cellulose nanocrystals and nanofibrils (CNC-250, CNF-50, and CNF-80), the SiO₂ NP and its Ag-doped variants (1% and 10% Ag-SiO₂), ZnS and ZnO, as well as the citrate-capped Ag and Au NPs. The 2D materials, like hBN, graphene, and reduced graphene oxide (rGO), were grouped separately from the others (i.e.: spherical, metal-based ENMs and cellulose-based, and anisotropic ENMs). These similarity and clustering results reiterated the importance of surface chemistry, size and shape of the ENMs to the formation of their unique protein corona, agreeing with literature reports (Walkey et al., 2014; Xu et al., 2018).

Because both the FC and corona profile are determined by the properties of the ENMs, it was anticipated that there would be certain correlations between the FC of one protein, measured with a specific type of ENM, and its abundance in the corona if the protein is a component in the matrix. Several of the standard proteins used to acquire the FC profiles were also found in serum, like HSA and transferrin. Thus, we examined the correlations between FC and RA values of these proteins (Fig. 3). Interestingly, the absolute value of the Pearson's correlation coefficient ($|r|$, from Eq. (4)) for the FC and RA values at one temperature was larger than 0.5, indicating certain levels of correlation. For example, the RA values of HSA and immunoglobulin (Ig), including IgG and IgM, in the protein corona were correlated with the FCs of HSA and γ -globulin at 60 °C with a $|r|$ value of 0.55 and 0.54, respectively. The $|r|$ between the RA of transferrin and the FC of transferrin at 37 °C was 0.52. However, poor correlations were also observed, for example, between the RA of HSA and the FC of HSA at 37 or 80 °C, suggesting that the FCs by one single protein measured at one temperature was not enough to help reveal the compositions of the complex corona formed in biological matrices, probably because the competitive adsorption of different proteins on the ENMs could heavily impact the abundance of others. While it is impossible to obtain the pure form of each of the proteins found in biological matrices and measure their FCs for better prediction of the corona composition, we remained optimistic that a sophisticated prediction model, using machine learning approaches, and the FCs of a group of judiciously selected and representative proteins could be obtained.

3.4. Building machine learning model to predict corona compositions

The goal of this work was to develop a prediction model for protein corona compositions using FCs as ENM descriptors. We adopted random forest (RF), a supervised machine learning algorithm, in our model, because RF is an ensemble of decision trees, and has

shown less overfitting and better performance compared to other advanced machine learning algorithms, such as artificial neural networks (ANN) and support vector machine (SVM) (Findlay et al., 2018; Panapitiya et al., 2018).

Fig. 4 depicts the workflow for running the prediction model, in which the RA values from the proteins were used as the targets, while the descriptors for proteins and ENMs were used as the features. The physicochemical properties of the proteins identified in the corona, including Mw, pI, GRAVY, and the percentage of negative/positive/aromatic amino acids, were calculated by ProtParam and used as the descriptors of proteins, while the FCs were used as the descriptors of ENMs. All data points (i.e.: ENM-protein pairs) were randomly divided into the training set (80%) and the testing set (20%), so the performance of the model could be tested on an independent dataset to avoid overfitting. Both classification and regression tasks were implemented. As for classification, the model was to predict whether a protein could be enriched in the protein corona of the ENMs. The RA values of the corona proteins were compared with those measured in the serum control, and the abundance changes (ACs) were calculated using Eq. (5). Proteins were classified into three categories (e.g. $AC > 1$ (class “1”), $AC < 1$ (class “-1”), and all other proteins (class “0”)), which were used as targets for the classification model. In the regression model, a more challenging task, the AC value for each protein in the protein corona was predicted. Features selection was used to optimize the model in both phases, and five of the important FCs were selected (Figs. S-9, S-10).

3.4.1. FCs outperform physicochemical properties of ENMs in corona

prediction—Since the physicochemical properties of ENMs (e.g. primary size, surface charge, ζ potential), are commonly used as descriptors for prediction models (Findlay et al., 2018; Pan et al., 2016; Singh and Gupta, 2014), they were used as ENM descriptors to build the benchmark model to evaluate the performance of using FCs as the descriptors of ENMs for the modeling. Typically, multiple specific properties are needed for modeling to ensure that the selected properties include the factors that truly influence the behaviors of ENMs. In addition, one of the physicochemical properties is kept the same so that the structure-activity relationship can be revealed. For example, in keeping the charge or hydrodynamic diameter (HD) constant, and changing the shape of gold NPs (i.e.: gold nanospheres vs. gold nanorods), it was revealed that the shape was a more crucial determinant of the protein corona than the other two properties (Xu et al., 2018). The ENMs included in this study encompass a wide range of physicochemical properties, with differences in size, shape, core material, and coating. To simplify the modeling, and to evaluate the effectiveness of our model compared to results obtained with conventional approaches, one particular group of ENMs (i.e.: nanospheres such as metallic and metal oxide NPs) was chosen to build the benchmark model using physicochemical properties of ENMs as the descriptors. Data obtained from the cellulose materials and 2D ENMs, like graphene, rGO, hBN, MoS₂, and V₂O₅, were not included in this step of the modeling. Another benchmark model was established using only the descriptors of proteins (DPs) and not the descriptors of ENMs.

The performance of the classification model was evaluated by the conventional confusion metrics including precision, recall, and f1 score (Table 1, Table S-14). Precision is the ratio of the correctly predicted proteins to the total predicted proteins, whereas recall measures

the ratio of the correctly predicted proteins to the proteins that should be in that class. The f1 score is the weighted average of precision and recall; higher values represent more accurate prediction. The values for precision, recall, and f1 score were all equal to 0.80 using only protein descriptors. This result suggests that protein properties are very important for the formation of the protein corona and using them as the descriptors can already provide a certain level of accuracy. By including the size/charge of the ENMs, the values for precision, recall, and f1 score increased by 18%, 7%, and 11%, respectively, for the prediction of whether a protein could be enriched in the corona (class “1”). Substituting the size/charge descriptors with FCs further increased these values by 6%, 38%, and 22%; such benefits were verified by the slightly larger areas under the Receiver Operating Characteristic (ROC) curves (AUC) (Fig. S-11), which measure how well a parameter can distinguish between groups. But, using both size/charge and FCs as the ENM descriptors did not provide additional improvement in the prediction performance. A similar phenomenon was also observed for the regression model, which was evaluated by R^2 (coefficient of determination), EVS (explained variance score), MAE (median absolute error), and MSE (mean squared error) (Table 1, Table S-15). The R^2 increased from 0.77 with only protein descriptors to 0.78 by adding the size/charge of ENMs. Using FC as the ENM descriptor instead of size/charge resulted in the highest R^2 (0.82), but no further improvement in prediction accuracy was observed by using both size/charge and FC. These results indicate that FC outperforms size/charge as an ENM descriptor in these two prediction models, hinting the possibility of modeling the corona formation on ENMs with diverse physicochemical properties once FC values are acquired.

3.4.2. Prediction model using FCs as descriptors works for a wide range of ENMs—The model was extended to include not only spherical, but also anisotropic ENMs, like cellulose-based crystals and fibrils, as well as 2D ENMs. The prediction models were built on the same procedure illustrated in Fig. 4. Model 1 was built using both FCs and the descriptors of proteins, and compared with Model 2, which was trained with only the protein descriptors. To optimize the performance of the model, feature selection was performed based on the relative importance of all features (Figs. S-9, S-10): the top 5 most important FCs – Tf-80 (Tf at 80 °C), γ G-60, α C-37, La-80, and α Cd-80 – were kept, while all other FCs were discarded.

For the classification task, the confusion metrics for both models at the threshold of 0.5 are shown in Table 1 and Table S-16. The precision, recall, and f1 score of Model 1 was 0.85, 0.85, and 0.84, respectively, and higher than those resulting from Model 2 (0.80, 0.79, 0.79). The improved performance by using FCs was also verified by the ROC curves (Fig. 5). The area under ROC (AUROC) of Model 2 was 0.9, 0.85, and 0.93 for class “1”, “0”, and “-1”, respectively, and these values increased to 0.94, 0.92, and 0.97 by adding FCs as the ENM descriptors.

As for the regression task (Fig. 6, Table S-17), incorporating the FCs as descriptors (Model 3) also greatly improved modeling accuracy, with the EVS, R^2 , MAE, and MSE being 0.81, 0.81, 0.40, and 0.82, respectively. In contrast, the prediction accuracy without the FCs (Model 4) dropped substantially, with values changing to 0.71 (EVS), 0.71 (R^2), 0.55 (MAE), and 1.24 (MSE). The decreased EVS and R^2 suggest a worse predictability, while

larger MAE and MSE suggest a more obvious deviation between the predicted and true values. Additionally, fitting the distributions of residual errors in prediction with the Gaussian function (Fig. 6C) resulted in a mean residual error for Model 3 of -0.07 , much closer to 0 compared to that of Model 4 (-0.17), suggesting that the predicted AC values by Model 3 were less deviated from the actual values. Moreover, the distribution of the residual errors of Model 3 was narrower than that of Model 4, shown as a smaller σ value (0.38 vs 0.54) and a higher maximum count. All these comparisons verified that the enhanced prediction performance of Model 3, compared to Model 4, were a result of the inclusion of FCs as the ENM descriptors.

To further validate the robustness of modeling with FCs as the ENM descriptors, Jackknife resampling was carried out; certain ENMs were chosen as the testing sets and checked against the models established using the rest of the ENMs as the training set for corona prediction. The 2D ENMs (i.e.: graphene, rGO, hBN, V_2O_5 , or MoS_2) or the cellulose-based ENMs (i.e.: CNF-50, CNF-80, or CNC-250) were chosen as the testing sets, because they are significantly different from other ENMs (i.e.: metal/metal oxide nanospheres) allowing the exploration of the applicability domain (AD) of the model. The remaining ENMs were used as the training sets to build the model. Both classification and regression models were established.

The prediction performance of the classification and regression models are shown in Table 2, Tables S-18/19, and Figs. S-12/13. Improvements in most of the performance parameters (e.g. larger f1 score and R^2) were observed after using the FCs as ENM descriptors, and there were more obvious improvements on the regression model compared with the classification model. The improvements were not the same for different ENMs, due to the high heterogeneity of the ENMs included. Prediction performance of our model is better when using the ENMs with similar shapes as the ones in question to form the training set. If using cellulose and spherical ENMs as the train set, the accuracy of the model was significantly lower (Table 2). Such an outcome was consistent with that of the PCA analysis which revealed FCs were different for ENMs with various shapes: Fig. S-4 clearly shows that FCs of 2-Ds ENMs are quite different from those of cellulose or spherical ENMs, which thus were displayed as distinct points in the PCA scatter plots.

Nevertheless, the advantage of using FCs for modeling was still obvious. For example, graphene, hBN, and MoS_2 had very different corona compositions compared to other ENMs (Fig. S-7A), but our modeling approach still resulted in good prediction accuracy (Table 2) and demonstrated the capability of the model to use FCs to make predictions on heterogeneous ENMs.

4. Conclusions

Protein corona composition is very important for the biological outcomes of ENMs. Herein, we developed a well performed prediction model using changes in fluorescence (FCs) as the novel descriptors of ENMs, which are capable of quickly and accurately predicting the protein corona compositions on diverse ENMs as demonstrated in our work. The FC values, obtained from fluorescamine labeling on a series of judiciously selected proteins at different

temperatures, were highly correlated with the protein interaction behaviors of the ENMs. The FC values were acquired in a high-throughput screening manner, and the labeling method was applicable to diverse ENMs. These features facilitate the evaluation of FCs compared to the ENM properties, like size and surface charge, which are measured by sophisticated instruments. In addition, FCs are superior descriptors of ENMs in corona modeling in contrast to traditional descriptors, and can successfully predict the protein corona formed on heterogeneous ENMs, which is very challenging for conventional modeling approaches that use well defined physicochemical properties of ENMs as descriptors. We expect this modeling approach can be employed to gain insights into the corona formation properties of newly synthesized ENMs before proceeding with detailed corona characterization. In silico comparison of the corona formed on a wide range of ENMs can potentially help focus follow-up studies to a handful of ENMs with the desired corona, instead of wasting time and efforts on a large number of ENMs. Future exploration will also be devoted to reveal potential correlation between protein corona and biological responses to ENMs, with the collegial efforts from research groups involved in the Nanotechnology Health Implications Research Consortium.

Still, the present model has its limitations, partially due to the very rough, semi-quantitative protein quantification method (i.e.: spectral counting employed in the present work for corona identification). We expect that a more accurate measurement of the corona composition could further improve our modeling approach. Collection of FC values for more representative proteins and inclusion of additional ENMs to build the model could also help enhance prediction accuracy. Moreover, with further understanding on how ENMs impact FCs or protein corona formation using first-principle calculation or molecular modeling techniques, this model can be used to guide the virtual design or optimization of ENMs for desired biological outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Research reported in this publication was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number U01ES027293 for W. Zhong as part of the Nanotechnology Health Implications Research (NHIR) Consortium. All the ENMs presented in this publication were procured/ developed, characterized, and provided by the Engineered Nanomaterials Resource and Coordination Core established at Harvard T. H. Chan School of Public Health (NIH grant # U24ES026946) as part of the Nanotechnology Health Implications Research Consortium. R. Coreas was supported by the Research Training Grant in Environmental Toxicology from the National Institute of Environmental Health Sciences (T32ES018827). Furthermore, the authors acknowledge the valuable contribution of Dr. Pyrgiotakis and Dr. Beltran-Huarac. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declaration of competing interest

Drs. Zhong and Demokritou report grants from the National Institutes of Health during the conduct of the study.

Appendix A.: Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.impact.2020.100207>.

References

- Ahn S, Ardonna HAM, Lind JU, Eweje F, Kim SL, Gonzalez GM, Liu Q, Zimmerman JF, Pyrgiotakis G, Zhang Z, Beltran-Huarac J, Carpinone P, Moudgil BM, Demokritou P, Parker KK, 2018 Mussel-inspired 3D fiber scaffolds for heart-on-a-chip toxicity studies of engineered nanomaterials. *Anal. Bioanal. Chem* 410, 6141–6154. [PubMed: 29744562]
- Apul OG, Wang Q, Shao T, Rieck JR, Karanfil T, 2012 Predictive model development for adsorption of aromatic contaminants by multi-walled carbon nanotubes. *Environmental science & technology* 47, 2295–2303. [PubMed: 22747100]
- Ashby J, Pan S, Zhong W, 2014 Size and surface functionalization of iron oxide nanoparticles influence the composition and dynamic nature of their protein Corona. *ACS Appl. Mater. Interfaces* 6, 15412–15419. [PubMed: 25144382]
- Ashby J, Duan Y, Ligans E, Tamsi M, Zhong W, 2015 High-throughput profiling of nanoparticle-protein interactions by fluorescamine labeling. *Anal. Chem* 87, 2213–2219. [PubMed: 25587850]
- Beltran-Huarac J, Zhang Z, Pyrgiotakis G, DeLoid G, Vaze N, Demokritou P, 2018 Development of reference metal and metal oxide engineered nanomaterials for nanotoxicology research using high throughput and precision flame spray synthesis approaches. *NanoImpact* 10, 26–37. [PubMed: 30035243]
- Bitounis D, Pyrgiotakis G, Bousfield D, Demokritou P, 2019 Dispersion preparation, characterization, and dosimetric analysis of cellulose nano-fibrils and nano-crystals: implications for cellular toxicological studies. *NanoImpact* 15, 100171.
- Caracciolo G, Farokhzad OC, Mahmoudi M, 2017 Biological identity of nanoparticles in vivo: clinical implications of the protein corona. *Trends Biotechnol.* 35, 257–264. [PubMed: 27663778]
- Casals E, Pfaller T, Duschl A, Oostingh GJ, Puntès VF, 2011 Hardening of the nanoparticle-protein corona in metal (Au, Ag) and oxide (Fe₃O₄, CoO, and CeO₂) nanoparticles. *Small* 7, 3479–3486. [PubMed: 22058075]
- Cedervall T, Lynch I, Lindman S, Berggård T, Thulin E, Nilsson H, Dawson KA, Linse S, 2007 Understanding the nanoparticle–protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. *Proc. Natl. Acad. Sci. U. S. A* 104, 2050–2055. [PubMed: 17267609]
- Chen R, Zhang Y, Darabi Sahneh F, Scoglio CM, Wohlleben W, Haase A, Monteiro-Riviere NA, Riviere JE, 2014 Nanoparticle surface characterization and clustering through concentration-dependent surface adsorption modeling. *ACS Nano* 8, 9446–9456. [PubMed: 25133703]
- Chen F, Wang G, Griffin JI, Brenneeman B., Banda NK, Holers VM, Backos DS, 2017 Complement proteins bind to nanoparticle protein corona and undergo dynamic exchange in vivo. *Nanotechnol* 12, 387–393. [PubMed: 27992410]
- Clemments AM, Botella P, Landry CC, 2015 Protein adsorption from biofluids on silica nanoparticles: corona analysis as a function of particle diameter and porosity. *ACS Appl. Mater. Interfaces* 7, 21682–21689. [PubMed: 26371804]
- Cohen J, Deloid G, Pyrgiotakis G, Demokritou P, 2013 Interactions of engineered nanomaterials in physiological media and implications for in vitro dosimetry. *Nanotoxicology* 7, 417–431. [PubMed: 22393878]
- Cohen JM, Teeguarden JG, Demokritou P, 2014 An integrated approach for the in vitro dosimetry of engineered nanomaterials. *Part Fibre Toxicol* 11, 20. [PubMed: 24885440]
- Cohen JM, Beltran-Huarac J, Pyrgiotakis G, Demokritou P, 2018 Effective delivery of sonication energy to fast settling and agglomerating nanomaterial suspensions for cellular studies: implications for stability, particle kinetics, dosimetry and toxicity. *NanoImpact* 10, 81–86. [PubMed: 29479575]

- Corbo C, Molinaro R, Tabatabaei M, Farokhzad OC, Mahmoudi M, 2017 Personalized protein corona on nanoparticles and its clinical implications. *Biomater. Sci* 5, 378–387. [PubMed: 28133653]
- Danishuddin; Khan AU, 2016 Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discov. Today* 21, 1291–1302. [PubMed: 27326911]
- DeLoid G, Cohen JM, Darrah T, Derk R, Rojanasakul L, Pyrgiotakis G, Wohlleben W, Demokritou P, 2014 Estimating the effective density of engineered nanomaterials for in vitro dosimetry. *Nat. Commun* 5, 3514. [PubMed: 24675174]
- DeLoid GM, Cohen JM, Pyrgiotakis G, Demokritou P, 2017 Preparation, characterization, and in vitro dosimetry of dispersed, engineered nano materials. *Nat. Protoc* 12, 355. [PubMed: 28102836]
- Deng ZJ, Liang M, Monteiro M, Toth I, Minchin RF, 2011 Nanoparticle-induced unfolding of fibrinogen promotes Mac-1 receptor activation and inflammation. *Nat. Nanotechnol* 6, 39. [PubMed: 21170037]
- Dominguez-Medina S, Kisley L, Tausin LJ, Hoggard A, Shuang B, Indrasekara AS, Chen S, Wang LY, Derry PJ, Liopo A, Zubarev ER, Landes CF, Link S, 2016 Adsorption and unfolding of a single protein triggers nanoparticle aggregation. *ACS Nano* 10, 2103–2112. [PubMed: 26751094]
- Duan Y, Liu Y, Shen W, Zhong W, 2017 Fluorescamine labeling for assessment of protein conformational change and binding affinity in protein–nanoparticle interaction. *Anal. Chem* 89, 12160–12167. [PubMed: 29083159]
- Duan Y, Liu Y, Coreas R, Zhong W, 2019 Mapping molecular structure of protein locating on nanoparticles with limited proteolysis. *Anal. Chem* 91, 4204–4212. [PubMed: 30798594]
- Durán N, Silveira CP, Durán M, Martínez DST, 2015 Silver nanoparticle protein corona and toxicity: a mini-review. *Journal of nanobiotechnology* 13, 55. [PubMed: 26337542]
- Fernández-Merino MJ, Guardia L, Paredes JI, Villar-Rodil S, Solís-Fernández P, Martínez-Alonso A, Tascón JMD, 2010 Vitamin C is an ideal substitute for hydrazine in the reduction of graphene oxide suspensions. *J. Phys. Chem. C* 114, 6426–6432.
- Findlay MR, Freitas DN, Mobed-Miremadi M, Wheeler KE, 2018 Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. *Environmental Science: Nano* 5, 64–71. [PubMed: 29881624]
- Gajewicz A, Jagiello K, Cronin MTD, Leszczynski J, Puzyn T, 2017 Addressing a bottle neck for regulation of nanomaterials: quantitative read-across (Nano-QRA) algorithm for cases when only limited data is available. *Environmental Science: Nano* 4, 346–358.
- Gass S, Cohen JM, Pyrgiotakis G, Sotiriou GA, Pratsinis SE, Demokritou PA, 2013 Safer formulation concept for flame-generated engineered nanomaterials. *ACS Sustain. Chem. Eng* 1, 843–857. [PubMed: 23961338]
- Hadjidemetriou M, Kostarelos K, 2017 Nanomedicine: evolution of the nanoparticle corona. *Nat. Nanotechnol* 12, 288. [PubMed: 28383044]
- Hoshino Y, Koide H, Furuya K, Lee SH, Kodama T, Kanazawa H, Oku N, Shea KJ, 2012 The rational design of a synthetic polymer nanoparticle that neutralizes a toxic peptide in vivo. *Proc. Natl. Acad. Sci. U. S. A* 109, 33–38. [PubMed: 22198772]
- Konduru NV, Molina RM, Swami A, Damiani F, Pyrgiotakis G, Lin P, Andreozzi P, Donaghey TC, Demokritou P, Krol S, Kreyling W, Brain JD, 2017 Protein corona: implications for nanoparticle interactions with pulmonary cells. *Part Fibre Toxicol* 14, 42. [PubMed: 29084556]
- Kubinyi H, 1997 QSAR and 3D QSAR in drug design part 2: applications and problems. *Drug Discov. Today* 2, 538–546.
- Lee YK, Choi E-J, Webster TJ, Kim S-H, Khang D, 2015 Effect of the protein corona on nanoparticles for modulating cytotoxicity and immunotoxicity. *Int. J. Nanomedicine* 10, 97. [PubMed: 25565807]
- Lesniak A, Fenaroli F, Monopoli MP, Åberg C, Dawson KA, Salvati A, 2012 Effects of the presence or absence of a protein corona on silica nanoparticle uptake and impact on cells. *ACS Nano* 6, 5845–5857. [PubMed: 22721453]
- Li L, Mu Q, Zhang B, Yan B, 2010a Analytical strategies for detecting nanoparticle-protein interactions. *Analyst (Cambridge, United Kingdom)* 135, 1519–1530.
- Li N., Zeng S, He L, Zhong W, 2010b Probing nanoparticle – protein interaction by capillary electrophoresis. *Anal. Chem* 82, 7460–7466. [PubMed: 20672831]

- Liu R., Jiang W, Walkey CD, Chan WC, Cohen Y, 2015 Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale* 7, 9664–9675. [PubMed: 25959034]
- Lotya M, Rakovich A, Donegan JF, Coleman JN, 2013 Measuring the lateral size of liquid-exfoliated nanosheets with dynamic light scattering. *Nanotechnology* 24, 265703. [PubMed: 23732310]
- Lundqvist M, Stigler J, Cedervall T, Berggård T, Flanagan MB, Lynch I, Elia G, Dawson K, 2011 The evolution of the protein corona around nanoparticles: a test study. *ACS Nano* 5, 7503–7509. [PubMed: 21861491]
- Lynch I, Weiss C, Valsami-Jones E, 2014 A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. *Nano Today* 9, 266–270.
- Mahmoudi M, 2018 Debugging nano–bio interfaces: systematic strategies to accelerate clinical translation of nanotechnologies. *Trends Biotechnol.* 36 (8), 755–769. [PubMed: 29559165]
- McClements DJ, DeLoid G, Pyrgiotakis G, Shatkin JA, Xiao H, Demokritou P, 2016 The role of the food matrix and gastrointestinal tract in the assessment of biological properties of ingested engineered nanomaterials (iENMs): state of the science and knowledge gaps. *NanoImpact* 3–4, 47–57.
- Monopoli MP, Walczyk D, Campbell A, Elia G, Lynch I, Baldelli Bombelli F, Dawson KA, 2011 Physical – chemical aspects of protein corona: relevance to in vitro and in vivo biological impacts of nanoparticles. *J. Am. Chem. Soc* 133, 2525–2534. [PubMed: 21288025]
- Monopoli MP, Åberg C, Salvati A, Dawson KA, 2012 Biomolecular coronas provide the biological identity of nanosized materials. *Nat. Nanotechnol* 7, 779. [PubMed: 23212421]
- Natte K, Friedrich JF, Wohlrab S, Lutzki J, von Klitzing R, Osterle W, Orts-Gil G, 2013 Impact of polymer shell on the formation and time evolution of nanoparticle-protein corona. *Colloids and Surfaces B-Biointerfaces* 104, 213–220.
- Pan Y, Li T, Cheng J, Telesca D, Zink JI, Jiang J, 2016 Nano-QSAR modeling for predicting the cytotoxicity of metal oxide nanoparticles using novel descriptors. *RSC Adv.* 6, 25766–25775.
- Panapitiya G, Avendaño-Franco G, Ren P, Wen X, Li Y, Lewis JP, 2018 Machine-learning prediction of CO adsorption in thiolated, Ag-alloyed Au nanoclusters. *J. Am. Chem. Soc* 140, 17508–17514. [PubMed: 30406644]
- Parviz D, Strano M, 2018 Endotoxin-free preparation of graphene oxide and graphene-based materials for biological applications. *Current Protocols in Chemical Biology* 10, e51. [PubMed: 30285316]
- Pirela SV, Martin J, Bello D, Demokritou P, 2017 Nanoparticle exposures from nano-enabled toner-based printing equipment and human health: state of science and future research needs. *Crit. Rev. Toxicol* 47, 678–704. [PubMed: 28524743]
- Pyrgiotakis G, Luu W, Zhang Z, Vaze N, DeLoid G, Rubio L, Graham WAC, Bell DC, Bousfield D, Demokritou P, 2018 Development of high throughput, high precision synthesis platforms and characterization methodologies for toxicological studies of nanocellulose. *Cellulose* 25, 2303–2319. [PubMed: 31839698]
- Riehemann K, Schneider SW, Luger TA, Godin B, Ferrari M, Fuchs H, 2009 Nanomedicine—challenge and perspectives. *Angew. Chem. Int. Ed* 48, 872–897.
- Righettoni M, Tricoli A, Pratsinis SE, 2010 Thermally stable, silica-doped ϵ -WO₃ for sensing of acetone in the human breath. *Chem. Mater* 22, 3152–3157.
- Saha K, Moyano DF, Rotello VM, 2014 Protein coronas suppress the hemolytic activity of hydrophilic and hydrophobic nanoparticles. *Materials horizons* 1, 102–105.
- Shang W, Nuffer JH, Dordick JS, Siegel RW, 2007 Unfolding of ribonuclease A on silica nanoparticle surfaces. *Nano Lett.* 7, 1991–1995. [PubMed: 17559285]
- Shi J, Kantoff PW, Wooster R, Farokhzad OC, 2017 Cancer nanomedicine: progress, challenges and opportunities. *Nat. Rev. Cancer* 17, 20. [PubMed: 27834398]
- Singh KP, Gupta S, 2014 Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. *RSC Adv.* 4, 13215–13230.
- Smits AH, Vermeulen M, 2016 Characterizing protein–protein interactions using mass spectrometry: challenges and opportunities. *Trends Biotechnol.* 34, 825–834. [PubMed: 26996615]

- Sotiriou GA, Watson C, Murdaugh KM, Darrah TH, Pyrgiotakis G, Elder A, Brain JD, Demokritou P, 2014 Engineering safer-by-design silica-coated ZnO nanorods with reduced DNA damage potential. *Environmental Science: Nano* 1, 144–153. [PubMed: 24955241]
- Stueckle TA, Davidson DC, Derk R, Kornberg TG, Schwegler-Berry D, Pirela SV, Deloid G, Demokritou P, Luanpitpong S, Rojanasakul Y, Wang L, 2017 Evaluation of tumorigenic potential of CeO₂ and Fe₂O₃ engineered nanoparticles by a human cell in vitro screening model. *NanoImpact* 6, 39–54. [PubMed: 28367517]
- Tenzer S, Docter D, Kuharev J, Musyanovych A, Fetz V, Hecht R, Schlenk F, Fischer D, Kiouptsi K, Reinhardt C, 2013 Rapid formation of plasma protein corona critically affects nanoparticle pathophysiology. *Nat. Nanotechnol* 8, 772. [PubMed: 24056901]
- Tsuda A, Venkata NK, 2016 The role of natural processes and surface energy of inhaled engineered nanoparticles on aggregation and corona formation. *NanoImpact* 2, 38–44. [PubMed: 29202111]
- Van Hong Nguyen B-JL, 2017 Protein corona: a new approach for nanomedicine design. *Int. J. Nanomedicine* 12, 3137. [PubMed: 28458536]
- Vaze N, Pyrgiotakis G, Mena L, Baumann R, Demokritou A, Ericsson M, Zhang Y, Bello D, Eleftheriadou M, Demokritou P, 2019a A nano-carrier platform for the targeted delivery of nature-inspired antimicrobials using engineered water nanostructures for food safety applications. *Food Control* 96, 365–374.
- Vaze N, Pyrgiotakis G, McDevitt J, Mena L, Melo A, Bedugnis A, Kobzik L, Eleftheriadou M, Demokritou P, 2019b Inactivation of common hospital acquired pathogens on surfaces and in air utilizing engineered water nanostructures (EWSN) based nano-sanitizers. *Nanomedicine* 18, 234–242. [PubMed: 30904585]
- Wagner V, Dullaart A, Bock A-K, Zweck A, 2006 The emerging nanomedicine landscape. *Nat. Biotechnol* 24, 1211. [PubMed: 17033654]
- Walczyk D, Bombelli FB, Monopoli MP, Lynch I, Dawson KA, 2010 What the cell “sees” in bionanoscience. *J. Am. Chem. Soc* 132, 5761–5768. [PubMed: 20356039]
- Walkey CD, Olsen JB, Song F, Liu R, Guo H, Olsen DWH, Cohen Y, Emili A, Chan WC, 2014 Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. *ACS Nano* 8, 2439–2455. [PubMed: 24517450]
- Watson-Wright C, Singh D, Demokritou P, 2017 Toxicological implications of released particulate matter during thermal decomposition of nano-enabled thermoplastics. *NanoImpact* 5, 29–40. [PubMed: 29333505]
- Xia X-R, Monteiro-Riviere NA, Riviere JE, 2010 An index for characterization of nanomaterials in biological systems. *Nat. Nanotechnol* 5, 671. [PubMed: 20711178]
- Xu M, Soliman MG, Sun X, Pelaz B, Feliu N, Parak WJ, Liu S, 2018 How entanglement of different physicochemical properties complicates the prediction of in vitro and in vivo interactions of gold nanoparticles. *ACS Nano* 12, 10104–10113. [PubMed: 30212621]
- Yan Y, Gause KT, Kamphuis MM, Ang C-S, O’Brien-Simpson NM, Lenzo JC, Reynolds EC, Nice EC, Caruso F, 2013 Differential roles of the protein corona in the cellular uptake of nanoporous polymer particles by monocyte and macrophage cell lines. *ACS Nano* 7, 10960–10970. [PubMed: 24256422]
- Yi M, Shen Z, 2015 A review on mechanical exfoliation for the scalable production of graphene. *J. Mater. Chem. A* 3, 11700–11715.
- Yonamine Y, Hoshino Y, Shea KJ, 2012 ELISA-mimic screen for synthetic polymer nanoparticles with high affinity to target proteins. *Biomacromolecules* 13, 2952–2957. [PubMed: 22813352]
- Zimmerman JF, Ardoña HAM, Pyrgiotakis G, Dong J, Moudgil B, Demokritou P, Parker KK, 2019 Scatter enhanced phase contrast microscopy for discriminating mechanisms of active nanoparticle transport in living cells. *Nano Lett.* 19, 793–804. [PubMed: 30616354]

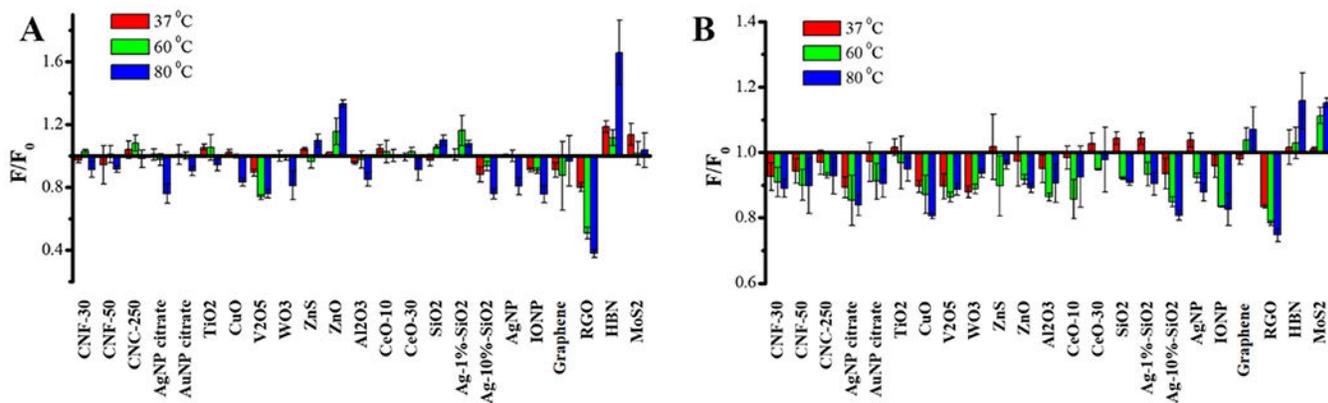


Fig. 1. Fluorescence changes (F/F_0) of (A) lysozyme and (B) lactalbumin after incubation with ENMs at different temperatures. The fluorescence signals of free proteins at different temperatures, were used as controls (F_0), and the signals of protein-ENM (2:1 mass ratio) mixtures were measured as F .

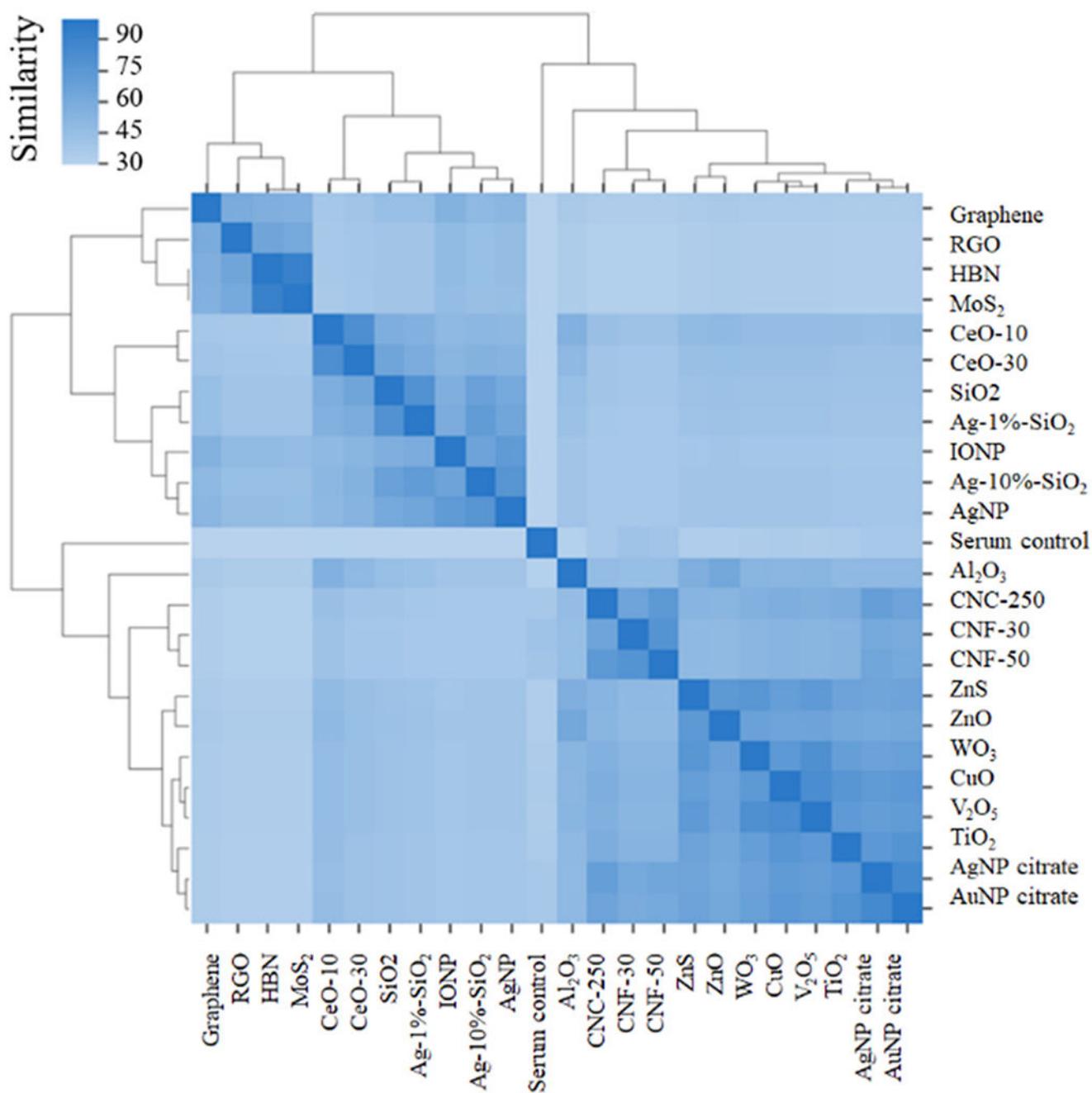


Fig. 2. The heatmap and hierarchical cluster of similarities of serum protein corona compositions for different ENMs. Human serum without ENMs was used as the control.

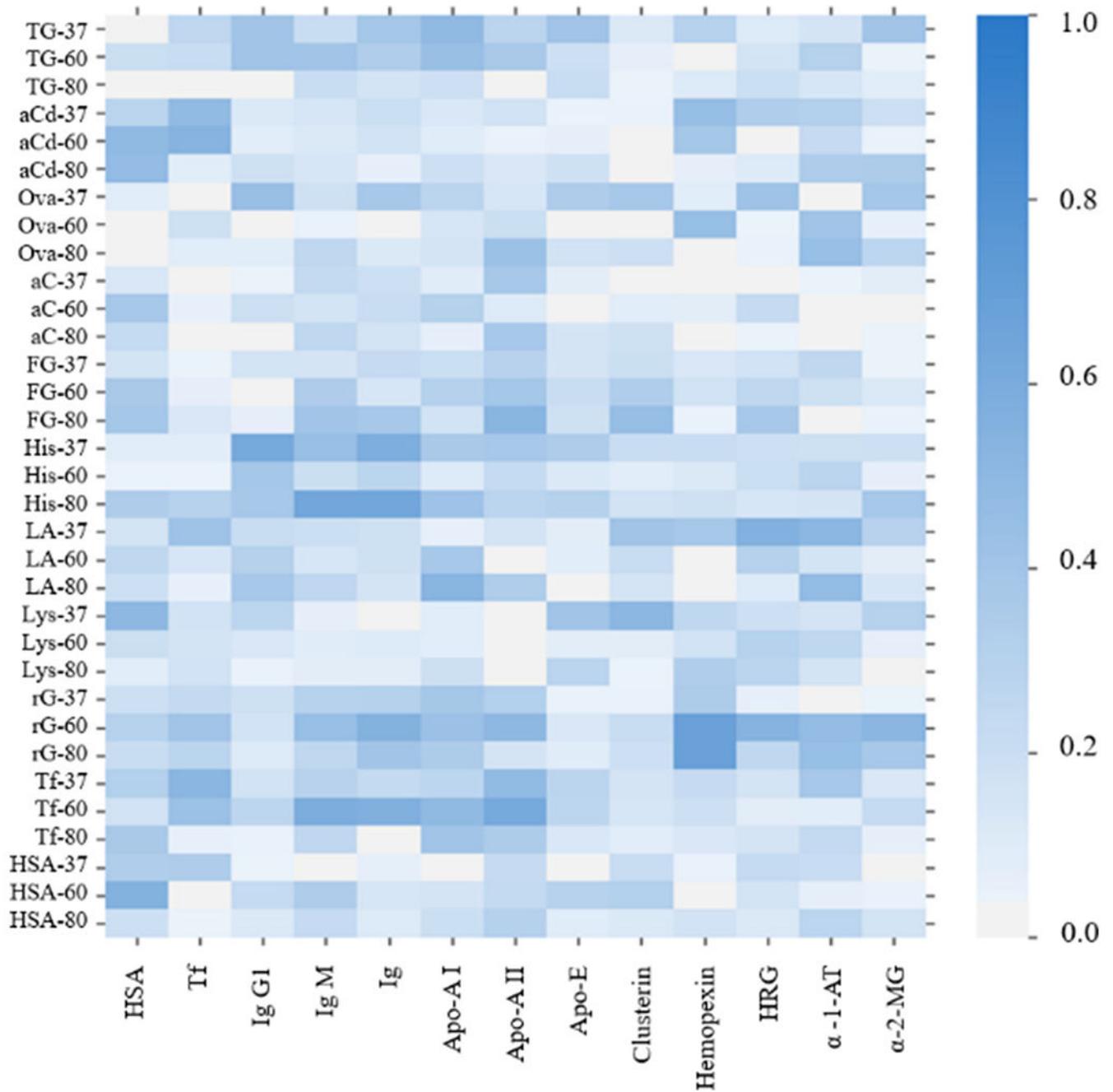


Fig. 3.

The absolute value of the correlation coefficient between the fluorescence change of one protein to the ENMs and the %RA of the 12 most abundant proteins identified in the corona, including HSA, Tf, immunoglobulin (Ig), apolipoprotein A (I and II), clusterin, histidine-rich glycoprotein (HRG), α -1-antitrypsin (α -1-AT), and α -2-macroglobulin (α -2-MG). For the first column, characters before the dash line represent the protein name, and the adjacent number represents the temperature at which fluorescamine labeling was performed. Each cell was colored from dark blue to white based on decreasing values.

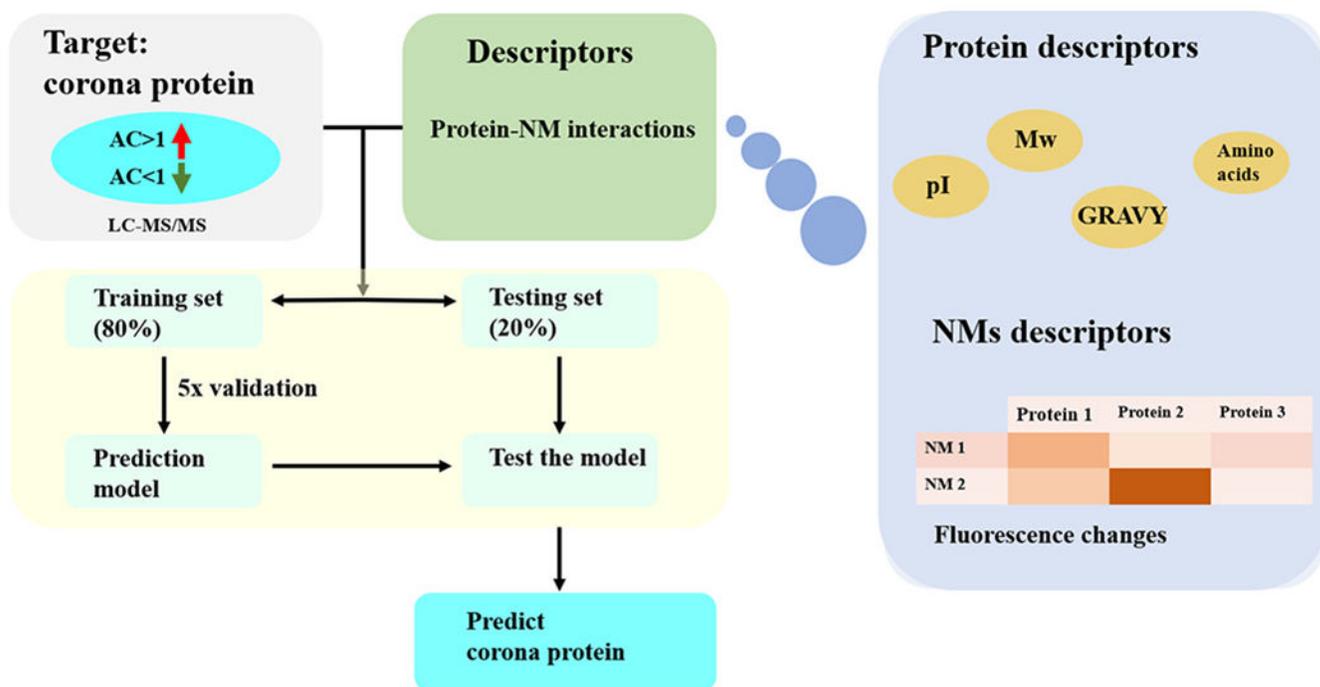


Fig. 4. The workflow followed to build the prediction model. The fluorescence changes of different proteins with ENMs at 37, 60, and 80 °C were used as descriptors for ENMs. The abundance change (AC) of each protein identified in serum protein corona was used as the target.

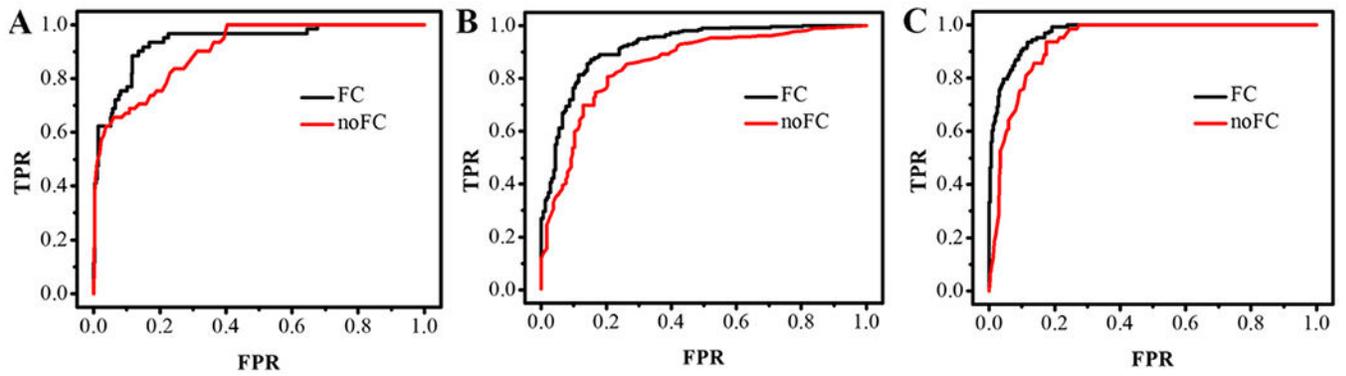


Fig. 5. Receiver operation characteristic (ROC) curve for the prediction of (A) proteins with abundance change (AC) > 1 , (B) proteins with AC between -1 and 1 , and (C) proteins with AC < -1 . The models with or without FCs were included for comparison. FPR: false positive rate. TPR: true positive rate.

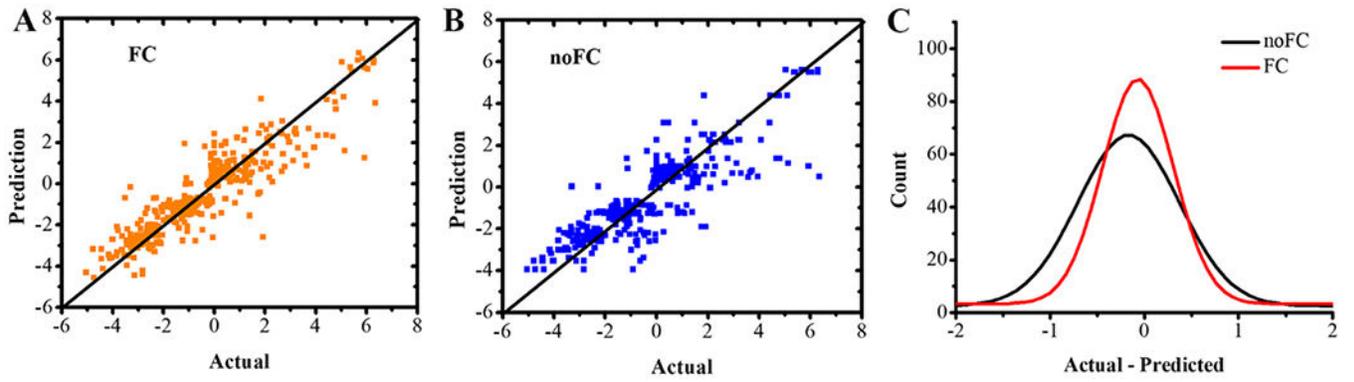


Fig. 6.

The values predicted by the regression model (A) with and (B) without FCs, were plotted against the actual values of the testing dataset. (C) The normal distribution of residual errors of the prediction values by both models.

Table 1

Performance of classification and regression models using different descriptors, for either only nanospheres or all ENMs.

ENMs	Descriptors	Classification		Regression
		f1 (average)	f1 (class 1)	R ²
Nanospheres	DP ^a	0.8	0.61	0.77
	DP + charge/size ^b	0.83	0.68	0.78
	DP + FC ^c	0.83	0.75	0.82
	DP + charge/size + FC	0.82	0.72	0.82
All ENMs	DP	0.79	0.67	0.71
	DP + FC	0.84	0.72	0.81

^aDP - descriptors of protein.

^bThe primary size and surface charge of NPs.

^cFluorescence changes.

Table 2:

Performance of both classification (evaluated by f1 score) and regression models (evaluated by R^2) using different descriptors for Jackknife resampling.

	DP^a		DP + FC^b	
	f1	R²	f1	R²
CNF ^c 50	0.76	0.65	0.81	0.85
CNF80	0.78	0.75	0.85	0.88
CNC ^d 250	0.78	0.62	0.84	0.75
Cellulose-ENM all	0.76	0.60	0.78	0.68
Graphene ^e	0.73	0.02	0.94	0.68
RGO ^f	0.64	0.57	0.67	0.56
hBN ^g	0.73	0.40	0.81	0.77
MoS2	0.71	0.39	0.84	0.68
V2O5	0.89	0.83	0.93	0.87
2D-ENM all	0.72	0.47	0.75	0.50

^aDescriptors of protein.

^bFluorescence changes.

^cCellulose nanofibril.

^dCellulose nanocrystal.

^eGraphene.

^fReduced graphene oxide.

^gHexagonal boron nitride.