Preprint Title	Multiscale modelling for biomolecular corona formation on metallic surfaces
Authors	Parinaz Mosaddeghi Amini, Ian Rouse, Julia Subbotina and Vladimir Lobaskin
Publication Date	17 Oct 2023
Article Type	Full Research Paper
Supporting Information File 1	Supporting material.zip; 3.8 MB
ORCID® iDs	Parinaz Mosaddeghi Amini - https://orcid.org/0009-0003-6003-6830; Ian Rouse - https://orcid.org/0000-0002-3686-7701; Vladimir Lobaskin - https://orcid.org/0000-0002-5231-0639

Multiscale modelling for biomolecular corona formation on metallic surfaces

Parinaz Mosaddeghi Amini*, Ian Rouse, Julia Subbotina and Vladimir Lobaskin[1]

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

Email: Parinaz Mosaddeghi Amini - parinaz.mosaddeghiamini@ucdconnect.ie

* Corresponding author

Abstract

Background: In the realm of food industry, the choice of non-consumable materials used plays a crucial role in ensuring consumer safety and product quality. Aluminum is widely used in food packaging and food processing applications, including dairy products. However, the interaction between aluminum and milk content requires further investigation to understand its implications.

Results: In this work, we present the results of multiscale modeling of the interaction between various surfaces (100,110,111) of FCC aluminum with the most abundant milk proteins and lactose. Our approach combines atomistic molecular dynamics, a coarse grained United Atom (UA) model, and kinetic Monte Carlo (KMC) simulations to predict the protein corona composition in the deposited milk layer on aluminum surfaces. We consider a simplified model of the milk, which was composed of the six most abundant milk proteins found in natural cow milk and lactose, which is the most abundant sugar found in dairy. Through our study, we ranked selected proteins and lactose adsorption affinities based on their corresponding interaction strength with aluminum surfaces and predicted the content of the naturally forming biomolecular corona.

Conclusion: Our comprehensive investigation sheds light on the implications of aluminum in food processing and packaging, particularly concerning its interaction with the most abundant milk proteins and lactose. By employing a multiscale modeling approach, we simulated the interaction between metallic aluminum surfaces and the proteins and lactose, considering different crystallographic orientations. The results of our study provide valuable insights into the mechanisms of lactose and proteins deposition on aluminum surfaces, which can aid in the general understanding of protein corona formation.

Keywords

bionano interface, multiscale modeling, milk protein, lactose, aluminum, protein corona, all atomistic, coarse grained model, kmc model

Introduction

The interface between biological systems and engineered materials has gained significant attention in recent years due to its wide range of applications, spanning from food to medicine and environmental science [1,2]. This interface plays a crucial role in ensuring the safety and quality of processed and packaged products. The selection of packaging materials and their interaction with biological components have emerged as critical determinants impacting the preservation, shelf life, and overall acceptability of dairy products [3]. Consequently, the interface between biologically relevant molecules and nanoscale materials, such as aluminum, has become an increasingly important and intriguing area of research [4]. For long-term storage and preservation of prepared food, the choice of containers and utensils made from specific materials is essential [5]. For example it was shown that ripened cheese and cheese spreads acquire a higher aluminum content as compared to other milk products [6]. Aside from wrapping and container packaging, aluminum has found a wide popularity in other applications, such as manufacturing of kitchen utensils, cosmetics, components for medical and scientific equipment [7]. Figure 1 presents a schematic contamination cycle of dairy products, showcasing potential sources and pathways of aluminum pollution. It illustrates the journey of milk from a cow grazing on the grass, contaminated with heavy metals, highlighting the crucial role of metallic containers, metal-based equipment, and kitchen utensils in maintaining product integrity. The figure further demonstrates the potential to introduce heavy metal contamination, including iron and aluminum, during processing and emphasizes the formation of a milk layer in form of protein/lactose corona at the outer surface of macroscropic, micro- and nano-sized particulate during post-packaging period. It also highlights the dynamic interactions at the bio-nano interface associated with potential human health hazards. Through biomolecule adsorption, change of conformation, and surface chemistry, foreign materials engage in a complex interplay of dynamic physicochemical interactions, kinetics, and thermodynamic exchanges that can lead to undesirable outcomes [1,8-10].

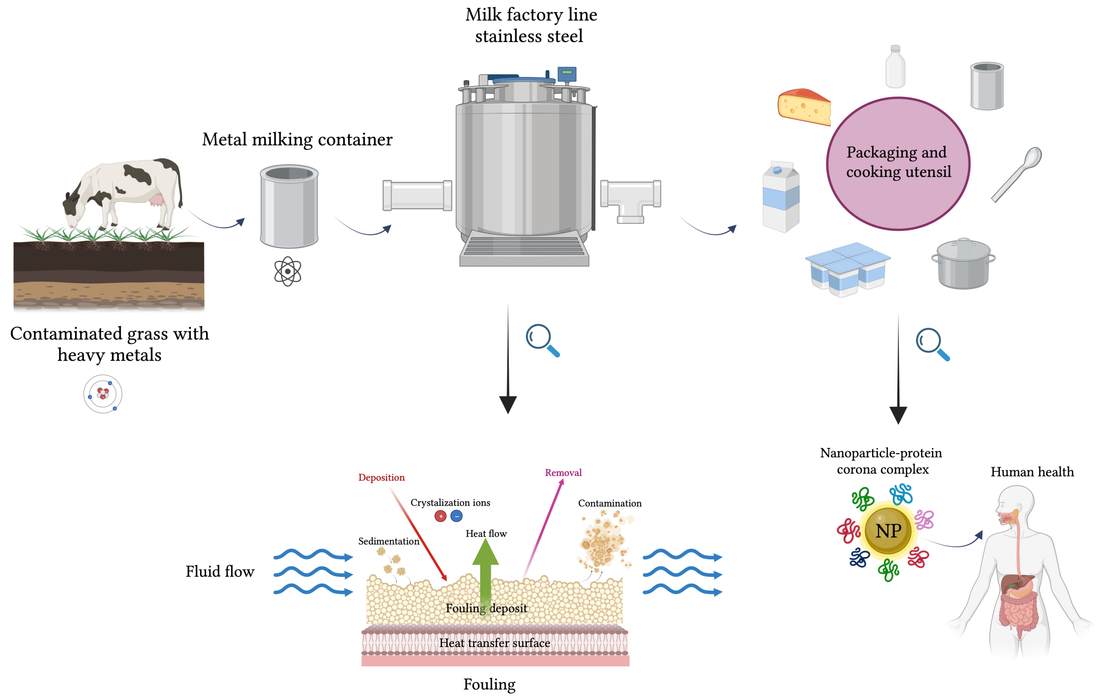

Figure 1: Schematic representation of the contamination cycle of dairy products, showcasing potential sources and pathways of contamination. The diagram illustrates the journey of milk from a cow grazing on grass contaminated with heavy metals, through collection, processing, and packaging stages. It highlights the crucial role of metallic containers, stainless steel equipment, and kitchen utensils in maintaining product integrity. The figure further illustrates the fouling process and the potential for milk contamination during processing. Additionally, it emphasizes the formation of a protein corona around nanoparticle (NP) post-packaging, while demonstrating the potential uptake of these NP by cells, thus highlighting the dynamic interactions at the bionano interface. Figure 1 was created with BioRender.com, https://biorender.com/. This content is not subject to CC BY 4.0.

In a more general context, the importance in understanding the mechanism of bionano interactions arises from the increasing awareness and concerns regarding the safety of nanoparticles (NPs) in relation to human and animal health. The toxicity of NPs is closely linked to its chemical aggressiveness and varies with its physicochemical properties, including surface area, charge, and reactivity. Understanding the intricate interplay between these properties and the biological systems is vital for assessing and mitigating any potential adverse effects associated with exposure to NPs [11]. To advance in this field, it is crucial to comprehend the underlying forces and molecular constituents that govern these interactions between biomolecules and metals. However, traditional safety assessment methods can be costly, time-consuming, and often involve animal studies. In this regard, *in silico* modeling offers a promising alternative that can predict the interactions of NPs with living organisms. By leveraging computational approaches, *in silico* modeling provides a humane and cost-effective means of obtaining the necessary information, thus aiding in the evaluation of NP safety and reducing reliance on animal experimentation [12-14]. Data-driven methods that rely on statistical analysis are employed for this purpose, particularly when sufficient data are available. These methods leverage the power of large datasets to identify patterns, trends, and correlations between metal properties and their interactions with biomolecules [15-18]. In recent years, researchers have focused on using physics-based models to understand the mechanisms underlying the formation of NP protein corona, a complex layer of biomolecules that surrounds NPs upon their exposure to biological fluids [19,20]. It is widely recognized that the composition and configuration of the protein corona play a crucial role in determining the biochemical reactivity, sensitivity of NPs, as well as their cellular uptake and systemic transfer [21]. However, in order to develop predictive models, a deeper understanding of the interactions at the bionano interface and their relationship to material and protein properties is necessary. Gathering more information on these intricate interactions will facilitate the development of accurate predictive models, thereby advancing our ability to assess the behavior and potential implications of NPs in biological systems. The bionano interface can be broken down into three interconnected components: (i) the surface of the NP, which is influenced by its physicochemical composition; (ii) the interface between the solid NP and the surrounding liquid environment, where notable changes occur upon interaction; and (iii) the contact zone between the solid-liquid interface and biological substrates (Figure 2) [22].

In this work, we study bionano interactions involving metallic aluminum and common dairy biomolecules: lactose and six most abundant milk proteins from these six major groups [23]. The

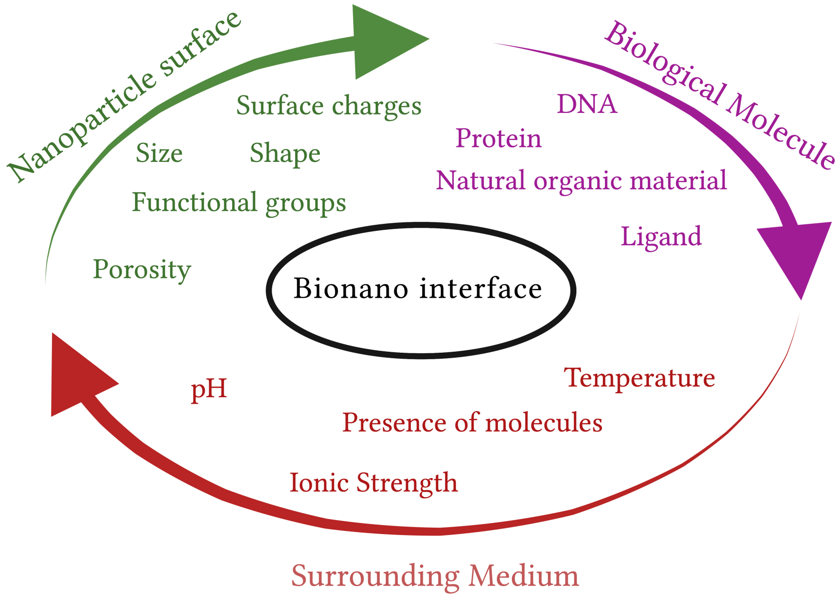

Figure 2: The bionano interface comprises three essential aspects, symbolized as components within a circular representation. These components consist of the surface properties of the nanomaterial, the characteristics of the surrounding medium, and the biological factors at play. These parameters collectively govern the intricate interactions and dynamics occurring at the interface.

main objective of our analysis is to computationally quantify the relative binding of these proteins on zero-valent aluminum surfaces based on their energy of adsorption and orientation. We employ a three-level multiscale method (as shown in Figure 3 to calculate the energies of adsorption and the content of the corona for these proteins on the selected surfaces. In the next section (Section "Results and Discussion"), we provide a detailed explanation of the theoretical model developed to study the interaction between protein and lactose with metals, as well as the rationale behind the parameterization scheme used. Subsequently, we discuss the simulation results and analyze the individual adsorption affinities predicted for molecules representing the biological aspect of the interface, including amino acids (AAs), milk proteins, and carbohydrates. Additionally, we examine the preferred orientations of these molecules upon adsorption and investigate the kinetics of competitive adsorption among the proteins and lactose, aiming to understand the process of protein deposition on metallic surfaces. Finally, the key insights gained from this study are summarized, highlighting the implications and potential applications of the findings.

Results and Discussion

In this section, we present the results and discuss the findings of our study of milk protein and lactose interaction with metallic aluminum surfaces using a multiscale molecular simulation. Our methodology employs a coarse-grained (CG) kinetic Monte Carlo (KMC) method [16] to simulate competitive adsorption of biomolecules onto the aluminum surface. To achieve this, we evaluate individual binding energies at various orientations (represented by heatmaps) for each selected protein immobilized on different FCC planes of the aluminum surface. These heatmaps for individual proteins are acquired through UA simulations [24,25]. While the UA method has been parameterized for a range of rigid surfaces, including metals (Ag, Au, Cu, and Fe), oxides (TiO_2, SiO_2, and Fe_2O_3), organic NPs (graphene, carbon nanotubes, and carbon black), semiconductors (CdSe) [26], and polymers [27], it lacks the set of short-range potentials required for calculating milk protein–aluminum adsorption energies. Here, we compute potentials of mean force (PMF) for Al surfaces derived from explicit all-atom molecular dynamics simulations utilizing a previously established scheme [2,24,28]. These PMFs provide the input required to determine the adsorption energies between milk proteins and aluminum surfaces by using multiscale UA CG model, spanning from the atomistic level of description to the complete mesoscale model of the corona. Figure 3 shows the parameterization and simulation workflow, outlining different stages and components involved in the study.

All-atoms short-range interaction modelling results

All-atom metadynamics simulations were conducted using GROMACS-2018.6 and PLUMED (PLUMED2-2.5.1.conda.5) software packages [29-31]. CHARMM-GUI/Nanomaterial Modeler was employed to construct the topology and force fields of three FCC surfaces of Al: (100), (110),

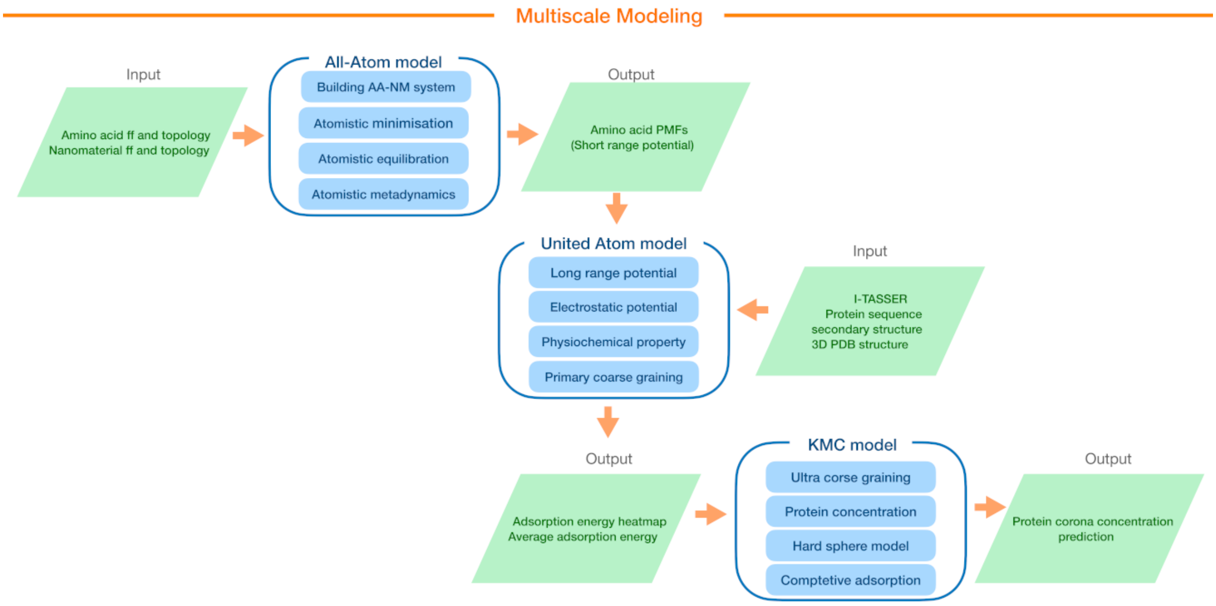

Figure 3: The multiscale modeling approach employed in this study. Our methodology begins by computing potentials for individual amino acids using an all-atom model. Next, it utilizes these to determine the adsorption energy for each protein, employing the coarse-grained United Atom model. Finally, it predicts the composition of the biomolecule corona using the coarse-grained kinetic Monte Carlo model. The figure provides an overview of input and output data at each scale.

and (111) [32]. The General Amber Force Field (GAFF) was utilized to model side-chains analogues (SCA) within the system [33,34]. The AMBER force field is a widely recognized and extensively validated force field that provides accurate descriptions of molecular systems [35]. We investigated the short-range PMF between 22 SCAs and a AlNP slab within a solvent environment comprising water and salt ions. The system's pH was maintained at a neutral level, and the salt concentration was set to match the biological salt concentration of 150 mM, equivalent to one salt molecule per 10 nm^3. The system underwent equilibration for 1.0 nanosecond under constant pressure conditions at 1.0 bar and a temperature of 300 K, following the NPT ensemble, employing Berendsen weak coupling method [36]. Subsequently, a pre-equilibration phase was conducted for 10 nanoseconds within the NVT ensemble. For the short-range van der Waals (vdW) interactions, the cut-off distance was defined as 1.0 nm. The adaptive well-tempered metadynamic (AWR-MetaD), the energy of adsorption was carried out at 300 K and pressure 1.0 bar and neutral pH in the NVT ensemble. Additionally, we measured the adsorption energy as a function of surface separation distance (SSD) as a collective variable, enabling a comprehensive analysis of the AA-NP interactions. For a detailed explanation of the method used in this study, please refer to previous reports [2,24,28] where the method has been described in depth. Figure 4 and dataset [37] shows the obtained free energy of adsorption in units of $k_B T$.

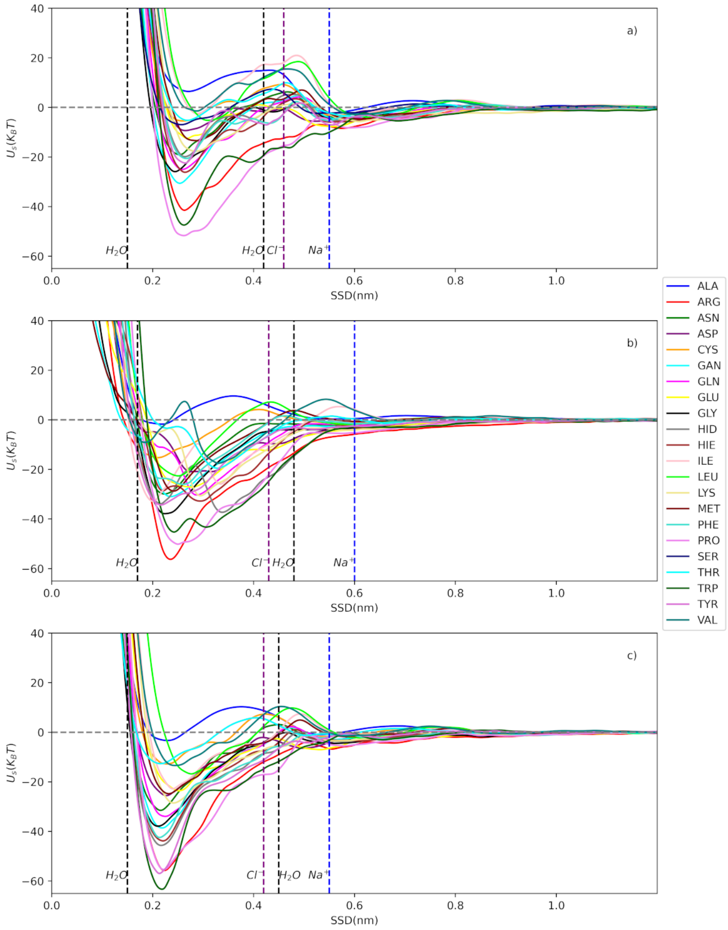

Figure 4: Adsorption free energy profiles of side chain amino acids on three aluminum FCC slabs as a function of Surface Separation Distance (SSD). These profiles were calculated using all-atomistic AWT-MetaD. The vertical lines indicate the positions of water and ion layers. **(a)** Al-100 **(b)** Al-110 and **(c)** Al-111.

The water density profiles obtained from MD simulations for the slab-water system in the context of Al surfaces revealed characteristics which were previously observed for other simulated metallic surfaces [2,28]. The profiles exhibited two distinct regions with elevated water density located approximately 0.15-0.18 nm and 0.42-0.48 nm away from the aluminum surface. These regions corresponded to the first and second water layers adjacent to the metal surface, respectively (as depicted in Figure S1). Further examination of the ion density profiles indicated the presence of sodium ions within a range of 0.55-0.60 nm and chloride ions within a range of 0.42-0.46 nm from the Al surface. Notably, the positions of the sodium and chloride ions align closely with the second water layer and the first water layer, respectively, as marked by the blue and purple vertical dashed lines in Figure 4. This alignment suggests that the sodium ions position themselves in proximity to the oxygen atoms of the first water adlayer, while the chloride ions integrate into the network of water molecules comprising the second adlayer. Additionally, the analysis of the PMF energy minima revealed a significant minimum at a distance of 0.21-0.25 nm, indicating a significant structuring of the adjacent water layer. Figure 5 shows the minimum energy values obtained for each AA on different facets of the aluminum surface (100, 110, and 111) in a bar chart.

A comparison of the adsorption energies on aluminum and iron surfaces reveals distinct preferences for different AA types. On aluminum surfaces, ARG, PRO, TRP, TYR AAs show the strongest attraction ($-63.32 k_B T$ to $-41.46 k_B T$), followed by HIE, GLN, PHE, GAN ($-43.86 k_B T$ to $-20.85 k_B T$). VAL, THR, SER, CYS, ALA exhibit the weakest attraction ($-19.51 k_B T$ to $-1.76 k_B T$). On iron surfaces, charged and aromatic PRO, TYR, ARG, HIS AAs are strongly attached ($-91.29 k_B T$ to $-43.34 k_B T$), while hydrophobic VAL, LEU, ALA AAs show a weaker adhesion ($-21.70 k_B T$ to $2.86 k_B T$). The PMF for glucose with aluminum surfaces computed using the PMFPredictor software is shown in Figure 6 [38].

Protein-NP interactions

To further understand the adsorption energy and orientation of each individual protein, a primary coarse-graining step was performed. In this part, we use the UA model to predict the protein–NP binding energies. This model takes into account various factors, such as the material's chemical composition, size, shape, surface roughness, charge, functionalization, and hydrophobicity, when constructing CG models for bionano interface. The UA model simplifies the protein-NP interactions by representing proteins as rigid structures composed of 20 AA types, each represented by

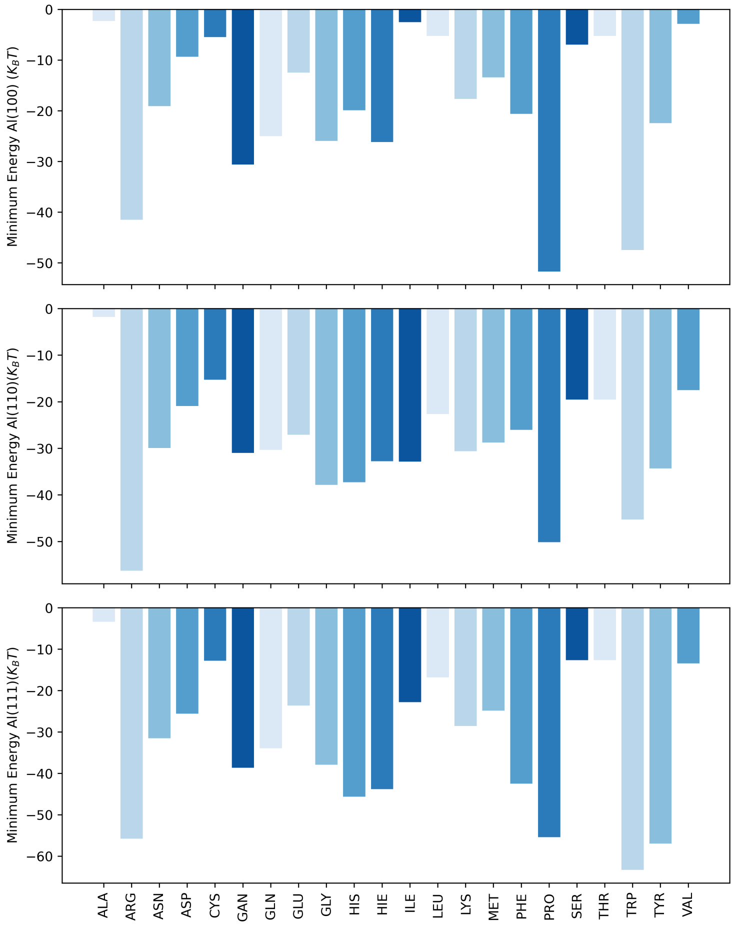

Figure 5: Minimum energy of adsorption values (k_BT) for each side chain amino acid on three Al FCC slabs (100, 110, 111) obtained through all-atomistic simulation. Notably, Al-111 exhibits stronger binding affinity in comparison to Al-100 and Al-110.

a single bead. This interaction is described through a short-range surface non-bonded potential (U_s^{nb}) (including vdW repulsion and solvent effects), a long-range core vdW potential (U_l^{vdW}), and an electrostatic potential (U^{el}). Through interaction potentials for specific AAs with the NP, the overall interaction potential between the NP and the complete protein (U_{p-NP}) is expressed in a pairwise additive manner:

$$U_{p-NP} = \sum_{i=1}^{N_{AA}} U_i(d_i(\theta,\phi)) = \sum_{i=1}^{N_{AA}} U_i^{el}(d_i(\theta,\phi)) + \sum_{i=1}^{N_{AA}} U_i^{nb_s}(d_i(\theta,\phi)) + \sum_{i=1}^{N_{AA}} U_i^{vdW_l}(d_i(\theta,\phi)) \tag{1}$$

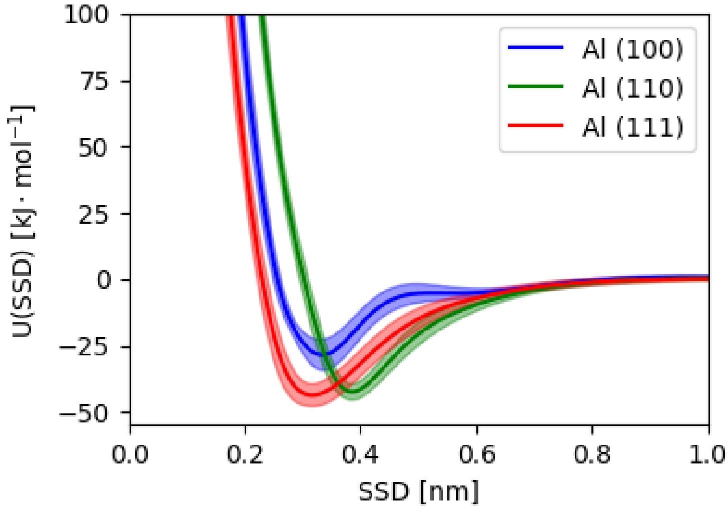

Figure 6: The interaction energy of glucose with the three Al surfaces considered in this work, predicted using the machine-learning method implemented in the PMFPredictor Toolkit. The solid lines give the ensemble average of ten versions of the model while the shaded regions indicate the 95% confidence intervals.

The potential U_{p-NP} depends on the distance d_i between the centers of mass (COMs) of the NP and each AA in the protein. This distance is determined by the protein's orientation with respect to the NP's surface, which is defined by two rotational angles (θ, ϕ) relative to the protein's initial orientation. This initial orientation is set by performing a principle axis transformation such that the axis associated with the smallest moment of inertia is aligned to the z-axis and the second smallest to the y-axis, i.e., the z-axis is now typically associated with the greatest extent of the protein. Since this does not uniquely specify the orientation, further rotations of 180° are then applied if necessary such that the electric dipole moment is positive along these two axes. This produces a convenient reference state by which other orientations are defined. The specific orientation ϕ, θ is generated by applying a rotations of $-\phi$ around the z-axis followed by $180° - \theta$ around the y-axis. The short-range surface non-bonded potentials are extracted from all-atom adaptive well-tempered metadynamics (AWR-MetaD) simulations that were described in Section "All-atoms short-range interaction modelling results" . The Hamaker technique is used to approximate the long-range term that results from the vdW forces working through the aqueous medium between the NP core and the ith AA. The electrostatic interaction between the NP and AA is represented by the screened Coulomb potential. More comprehensive information about the theoretical aspects of the UA model can be found in our previous publications [2,25,28,39,40]. The output of the UA simulations contains a collection of rotational configurations and their corresponding $E(\theta_k, \phi_l)$ values. By employing Boltzmann averaging and weighting factors based on the potential energy as a function of distance for each angle, we calculate the average adsorption energy of these configurations. Using this approach, we evaluate the adsorption energies of the entire proteins on aluminum surfaces. To predict the three-dimensional (3D) structures of proteins, we utilize the I-TASSER (Iterative Threading ASSEmbly Refinement) 5.1 software [41], which uses the protein's AA sequences as an input.

For this study, we have chosen 6 representative cow milk proteins and lactose that constitute most of the non-fat milk solids. Table 1 displays properties of the chosen milk molecules. It includes their UniProt IDs, molecular weights, charges, and the number of AAs in each protein. The charge data was determined through the PROPKA method [42,43] at a pH of 7.0. As we will discuss in Section "Protein-NP interactions", we model the lactose molecule as a pair of glucose beads, and it does not possess a UniProt ID or a count of AA residues. We estimated the concentration of each protein and lactose based on their weight fraction in milk and considering the fact that cow milk has 30–39 g/L of protein and 45-55 g/L of lactose in total. The molar mass of each protein was taken from AlphaFold database [44]. Following this, all proteins underwent a 50 ns equilibration in water using NVT and NPT ensembles.

Table 1: Characteristics of the selected milk proteins and lactose.

Abbreviation	UniProt ID	Molecule Name	MW^a, Da	Charge, e	Res^b	$C^c\,[10^{-4}]$, mol/L
AS1C	P02662	αs1-casein	24528.00	-8.5	214	4
AS2C	P02663	αs2-casein	26018.69	4.5	222	1
BC	P02666	β-casein	25107.33	-4.5	224	4
ALAC	P00711	α-lactalbumin	16246.61	-5	142	0.9
BLAC	P02754	β-lactoglobulin	19883.25	-6	178	2
BSA	P02769	bovine serum albumin	69293.41	-4.5	607	0.1
LAC	-	lactose	342.3	0	-	1300

(a) Molecular weight, (b) Number of residues, (c) Concentrations [mol/L] of the molecules in milk that were used in KMC calculations.

The UA computations were conducted using nine different Al NPs with varying radii, namely 2 nm, 5 nm, 10 nm, 20 nm, 30 nm, 40 nm, 50 nm, 80 nm, and 100 nm, to investigate the influence of size and curvature on the adsorption energies. The results and detailed information on the calculation can be found in Figures S2 and S3, which illustrate the variations in adsorption energies as a function of NP size. Within the range of 2-20 nm the binding energy of ALAC, BLAC, BC, and BSA shows an initial increase on all surfaces, followed by a stabilization at larger NP sizes. In contrast, AS1C and AS2C exhibit a continuous rise in binding energy across the entire size spectrum, ranging from $-48.0 k_B T$ at 2 nm to $-281.09 k_B T$ at 100 nm for AS1C and $-15.26 k_B T$ at 2 nm to $-275.60 k_B T$ at 100 nm for AS2C, with AS2C exhibiting the most dramatic changes in binding energy as a function of size. This strong size dependence in binding energy for AS2C can be attributed to its rod-like 3D structure and the rigidity assumption in our model. As the size of the NP increases, AS2C can make more extensive contact with the surface. This increased contact area leads to enhanced binding affinity, resulting in the observed stonger in binding across the size range. This is not the case for other proteins on the list as they are small and compact and therefore reach the maximum number of contacts at relatively small NP sizes. Regarding the binding affinity rankings, for the smallest NPs (2 nm), the order from weakest to strongest is observed as AS2C, BSA, ALAC, BLAC, AS1C, and BC on Al-100, with similar rankings observed on Al-110 and Al-111 surfaces. However, for the largest (flattest) NPs (100 nm), the binding affinity ranking changes to ALAC, BLAC, BSA, BC, AS2C, and AS1C on Al-100, BC, ALAC, BLAC, BSA, AS2C, and AS1C on Al-110, and BLAC, ALAC, BC, BSA, AS2C, and AS1C on Al-111 (see Figure S2). In reality, proteins' structure is not rigid, allowing them to adapt to the surfaces upon immobilisation. Thus this can potentially affect their binding behavior. This can be especially significant for caseins, as they belong to the group of flexible milk proteins with no tertiary structure. Globular milk proteins (lactoglobulin and lactalbumin) are expected to be less prone to this shortcoming of UA model.

Figure 7 shows the output of the UA model for the selected milk proteins on Aluminum NP surface size 80 nm with zeta potential -5 mV at pH 7.0. The heatmaps display the adsorption energies for all values of θ and ϕ. Blue areas with lower energies indicate more favorable orientations of the proteins. Each heatmap is accompanied by a 3D representation of the protein on the NP surface, with the AAs closest to the NP's surface marked. The AAs that are most likely to make contact with the metal surfaces, according to analysis, are LYS, TYR, PHE, GLU, ARG, ASP.

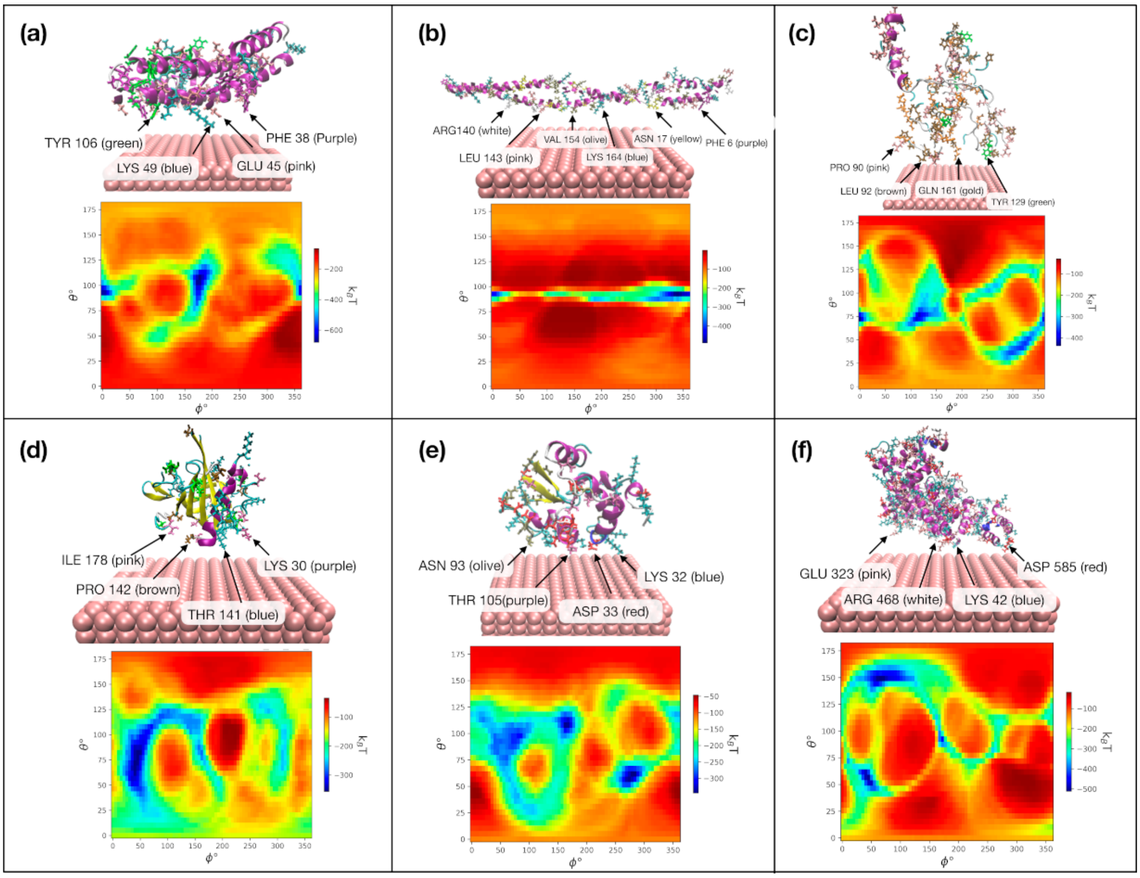

Figure 7: Heatmap results obtained from United Atom model and corresponding 3D representations of the interactions of (**a**) αs$_1$-casein, (**b**) αs$_2$-casein, (**c**) β-casein, (**d**) β-lactoglobulins, (**e**) α-lactalbumin, and (**f**) bovine serum albumin with Al-110 on the preferred orientation. The figure highlights the closest amino acids to the surface of the material.

The rankings of protein adsorption on each aluminum surface are shown in Table 2 , highlighting the variations in adsorption energies (E_{ad}/k_BT) and the particular protein-surface interactions (θ and ϕ in degrees). Moreover, the minimum distance (r_{min} in nm) indicates the closest approach of the protein to the aluminum surface during the adsorption process.

The ranking of adsorption energies highlights the distinct adsorption behaviors of various proteins on different aluminum FCC surfaces. Particularly noteworthy is the consistently high adsorption

Table 2: Comparison of milk proteins' binding affinities and orientations on Al-100, Al-110, and Al-111 with NP size of 80 nm, derived from the United Atom model and ordered by the binding strength on each surface.

Individual protein adsorption description on Al-100

Protein,	E_{ad}/k_BT	ϕ,°	θ,°	r_{min}, nm
AS1C	−145.65	175	100	0.19
BC	−108.13	305	40	0.13
AS2C	−96.12	315	95	0.05
BSA	−91.11	45	60	0.11
BLAC	−67.35	65	90	0.19
ALAC	−49.12	125	35	0.20

Individual protein adsorption description on Al-110

Protein,	E_{ad}/k_BT	ϕ,°	θ,°	r_{min}, nm
AS1C	−278.37	175	100	0.32
AS2C	−224.01	345	90	0.10
BSA	−173.77	40	60	0.23
BLAC	−157.70	50	95	0.28
ALAC	−155.17	70	90	0.29
BC	−132.52	0	70	0.20

Individual protein adsorption description on Al-111

Protein,	E_{ad}/k_BT	ϕ,°	θ,°	r_{min}, nm
AS1C	−242.93	175	100	0.15
AS2C	−181.65	330	90	0.11
BSA	−137.46	45	60	0.13
BC	−131.93	140	110	0.15
ALAC	−125.76	75	90	0.17
BLAC	−113.39	45	75	0.20

energy of AS1C across all surfaces, indicating a strong binding affinity with aluminum. On the other hand, ALAC and BLAC exhibited the lowest adsorption energies on most surfaces, suggesting weaker interactions. Meanwhile, BC, AS2C, and BSA displayed moderate adsorption energies, indicating intermediate binding strengths with aluminum. The Table S2 of the Supplementary Material reports the preferred orientations of all 820 milk proteins based on the lowest energy from the UnitedAtom output. In our investigation of these proteins, we focused on identifying the most strongly adsorbing proteins when exposed to Fe and Al. These proteins, including P19660, A6QP30, G3X745, F1MMI6, E1BBY7, A6QLY7, and Q9N2I2, demonstrated remarkable similarity in their binding behavior towards Fe-100 and Al-100 surfaces, E1BGJ4, A5D7M6, F1MMI6, A6QP30, G3X745, and F1N1C7 on Fe-110 and Al-110 surfaces, and F1MMI6 and E1B748 and A6QP30 on Fe-111 and Al-111 surfaces.

In the subsequent step, we predict the composition of the milk protein layer at the aluminum surfaces. For this analysis, we consider the Al surface as a spherical NP with the protein layer uniformly adsorbed on its entire surface, forming the protein corona.

Competitive adsorption and biomolecular corona

Kinetic Monte Carlo (KMC) simulations as implemented in the CoronaKMC tool [26] were employed to investigate competitive adsorption and determine the composition of the protein corona. This method models adsorbates as hard-spheres which adsorb and desorb to the surface of the NP, with different orientations of each protein treated as different potential adsorbates to allow for a more physically realistic model of corona formation for anisotropic proteins. In brief, a standard kinetic Monte Carlo routine is used to advance the simulation from one event – collision of an incoming adsorbate with the NP or desorption of an adsorbed species – to the next, with events occurring with a probability proportional to their rate. In the initial form of the model, adsorption is assumed to occur with unit probability if the incoming species does not overlap with any currently adsorbed species and fails to take place otherwise. We parameterize this model using adsorption and desorption rate constants extracted from UnitedAtom results as described previously [16,45]. In brief, each potential adsorbate (e.g. a small molecule or a particular orientation of a protein) is projected onto the surface of the NP and a convex hull procedure used to estimate the area of the NP occupied by that adsorbate, A_i. The adsorbate is then assigned an effective radius R_i such that a sphere projected onto the NP would produce the same radius [16]. The per-site adsorption rates are calculated using kinetic theory for the rate of collisions between two spheres in solution, normalized by the number of binding sites for that protein,

$$k_a = \frac{A_i}{4\pi R_{NP}^2} \left[4\pi D N_A \left(R_{NP} + R_i\right)\right] \tag{2}$$

where R_{NP} is the radius of the NP, N_A is Avogadro's number, R_A is the effective adsorbate radius, D is the pair diffusion coefficient given by,

$$D = \frac{k_B T}{6\eta} \left(R_{NP}^{-1} + R_A^{-1} \right) \tag{3}$$

taking the viscosity $\eta = 8.9 \times 10^{-4}$ Pa $\cdot$s. We employ SI units in the above calculation, noting that k_a must then be multiplied by 1000 to convert from units $\mathrm{m}^3 \cdot \mathrm{mol}^{-1}$ to $\mathrm{L} \cdot \mathrm{mol}^{-1}$. Desorption rates are found by requiring that $k_a / k_d = K_{eq} = 1 \frac{\mathrm{L}}{\mathrm{mol}} e^{-E_{ads}/k_B T}$, where E_{ads} is the value obtained for that orientation using UnitedAtom [45]. A concentration is then assigned to the adsorbate based on the bulk concentration of that adsorbate, weighted by the relative abundance of that orientation of the adsorbate if necessary, i.e., for protein i with a bulk concentration of C_i and set of orientations θ_k an orientation θ_j is assigned a concentration,

$$C_{i,j} = C_i \frac{\sin \theta_j}{\sum_k \sin \theta_k}, \tag{4}$$

to ensure that orientations are correctly weighted and the total concentration summed over orientations is correctly reproduced. Scripts to automate this parameterization based on UA output and adsorbate structure files are available as part of the UnitedAtom repository [26].

We further analyze the results for adsorption of milk components obtained from KMC simulations, specifically focusing on the mean absolute and relative abundance of proteins (10^{-3} nm^2) adsorbed on Al surfaces per unit area (nm^2). Table 3 shows the abundances of the proteins and lactose on Al surfaces.

The simulations were performed using NPs with a radius of 80 nm, and the data is presented in Table 3. It presents the number concentration and mass abundance of proteins adsorbed on three different Al surfaces: Al-100, Al-110, and Al-111. Each protein's adsorption behavior is quantified in terms of its number concentration (expressed in units of 10^{-3} nm^{-2}) and mass abundance (represented as a percentage of the total adsorbed mass). These calculations were performed utilizing the most recent KMC method modifications, including an alternative mode in which the acceptance-

Table 3: Mean amounts of proteins adsorbed on Al surfaces per unit area: number concentration (per nm^2), mass abundance obtained from KMC simulations with NPs of radius 80 nm. These calculations have been done using the latest modifications to the KMC approach.

	Al-100	**Al-100**	**Al-110**	**Al-110**	**Al-111**	**Al-111**
Protein	N_{ads}[10^{-3}, nm^{-2}]	M_{ab},%	N_{ads}[10^{-3}, nm^{-2}]	M_{ab}, %	N_{ads}[10^{-3}, nm^{-2}]	M_{ab}, %
AS1C	12.26	57.16	16.70	67.82	27.21	83.19
BC	4.45	21.24	3.38	14.07	1.91	5.84
BLAC	2.91	10.99	2.97	9.79	1.00	2.43
LAC	96.59	6.28	89.13	5.05	84.50	3.62
ALAC	1.14	3.51	1.13	3.05	1.84	3.60
AS2C	0.11	0.55	0.04	0.16	3.00	1.09
BSA	0.02	0.25	0.00	0.05	0.02	0.21

rejection criteria for incoming adsorbates are tweaked to allow for some overlap with pre-existing adsorbates. We should note that the Al-111 has the lowest energy of all three, according to the Materials Project data, so we expect the adsorption profile in real systems be similar to that predicted for Al-111.

We also compared the protein composition in the corona on aluminum and iron [2], obtained in our previous work using the original KMC approach without molecular displacements. This comparison is shown in Figure 8. AS1C exhibited the highest abundance on both iron and aluminum among the studied proteins, indicating a strong affinity for both metals with both KMC methods. Following AS1C, BC and BLAC and ALAC also showed fairly equal abundances on the surfaces of iron and aluminum. On the other hand, BSA displayed the lowest abundance on both metals due to its larger size and the relatively low molar fraction in milk as compared with other proteins. Figure 8 shows the mass abundance of each protein on both aluminum (Al-100, Al-110, Al-111) and iron (Fe-100, Fe-110, Fe-111) surfaces. We can also observe that AS1C, BLAC, and ALAC display significantly enhanced presence on Fe surfaces in contrast to their Al counterparts. Conversely, AS2C shows greater adsorption on Al surfaces as compared to Fe. Overall, we expect a somewhat different corona formed on these metallic surfaces.

Realistic organic media do not consist only of proteins, but it also includes many other molecules, e.g. sugars and other organic compounds, that may bind to an NP along with proteins. It can reasonably be assumed that these molecules may alter both the kinetics and equilibrium state of the

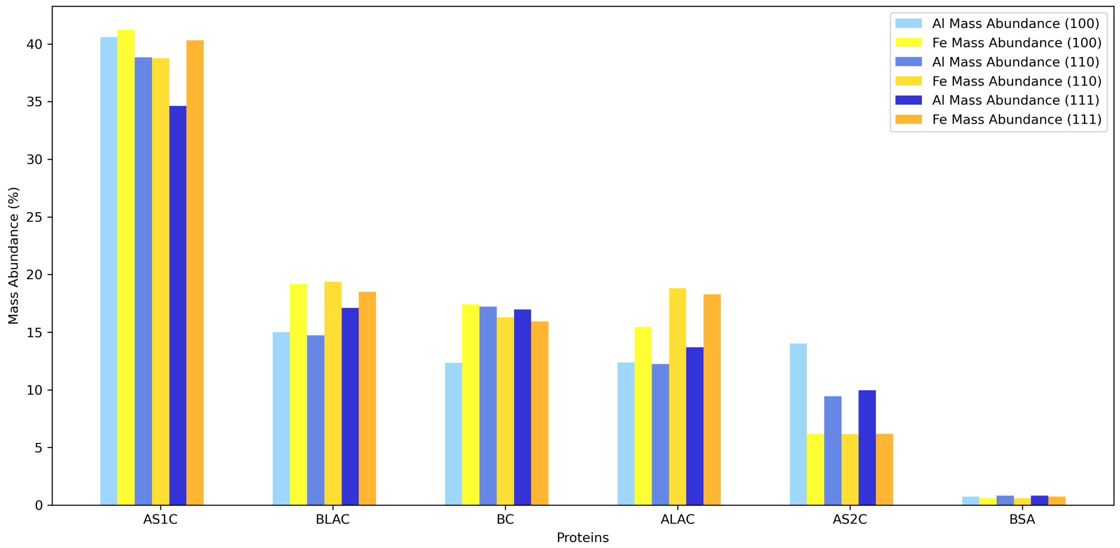

Figure 8: Mass Abundance of Proteins on Al (blues) and Fe (yellows) Surfaces (100, 110, and 111) using the original KMC approach without molecular displacements, with a NP radius size of 80 nm.

corona, and moreover may play a role in biological outcomes. Thus, it is of interest to include these small molecules in the corona simulation to not only gain further insight into this particular case of aluminum in milk, but also to establish a methodology by which more general molecules can be included in these simulations. We choose lactose as a prototypical example of a small molecule capable of binding to an NP, since this is present at a high concentration in milk. We model the lactose molecule as a pair of glucose beads separated by a distance determined by the equilibrium structure of lactose. Although this is not completely rigorous, it demonstrates how the UnitedAtom software can be adapted to model larger molecules other than proteins using the same fragment-based approach. To avoid the need to run a time-consuming parameterization protocol based on metadynamics simulations, we produce PMFs for the glucose bead using a machine-learning technique (PMFPredictor) trained on previous metadynamics results [38]. For the lactose molecule, each constituent glucose bead is assigned a charge of 0 and the Hamaker term is neglected due to the small size of these beads. Following this parameterization, the coarse-grained lactose molecule is processed identically to proteins using the same automated pipeline, i.e., UnitedAtom is run to produce a table of orientation-specific binding energies and these mapped to rate constants for adsorption and desorption. We stress that this procedure is sufficiently generic that essentially arbitrary organic molecules can be included in the simulation by performing a fragment-based decomposition, generating PMFs via traditional or machine-learning approaches, and constructing a coarse-grained representation for input to UA. To simplify this procedure for more complex molecules, we have developed a Python script (MolToFragments.py) employing RDKit [46] to automate splitting larger molecules into suitable fragments and producing coarse-grained input files suitable for UnitedAtom and included with this repository [26].

The addition of lactose (or other small molecules) to the corona simulation poses a challenge for the form of the CoronaKMC algorithm previously employed due to the high concentration and very small binding area of this small molecule relative to proteins [16,45]. As a consequence of these factors, the original form of the algorithm results in rapid coverage of the NP with a very large quantity of lactose which greatly increases the required computational time, which scales as $O(N^2)$ for N adsorbed particles. Moreover, in this original form of the model a single adsorbed lactose molecule inhibits the adsorption of a large protein, no matter how strongly the protein may adsorb. To counteract these issues, the following features were added to the new version of the CoronaKMC software. Firstly, we implemented a method to accelerate the simulation by adjusting rate constants for quasi-equilibriated processes (e.g. the adsorption of lactose) according to the methodology of Dybeck et al. [47]. Secondly, we added an optional mode in which the acceptance-rejection criteria for an incoming adsorbate is modified such that an incoming adsorbate is no longer immediately rejected if it overlaps with a pre-existing adsorbate. Instead, the incoming adsorbate is accepted with a probability p given by,

$$p(\Delta E) = \frac{\exp[-\Delta E/k_B T]}{1 + \exp[-\Delta E/k_B T]} \tag{5}$$

where ΔE is the difference in energy between the two states,

$$\Delta E = E_{ads} - \sum_j E_j \tag{6}$$

where j is the set of all adsorbed particles which would overlap with this particle, taking ΔE =

E_{ads} if no overlaps are found. If the adsorbate is accepted then all the overlapping particles are removed from the NP. We note that this breaks the principle of detailed balance in that it allows for the replacement of a set of adsorbates by a single molecule, but does not allow for the converse in which a set of incoming molecules can displace an adsorbate. We justify this neglect on the basis that the required event of multiple simultaneous collisions on a single target would occur so rarely that it would essentially not be sampled in the course of a simulation. The probabilistic acceptance to regions of the NP without explicit adsorbates present effectively multiplies the adsorption rate by a factor of $p(E_{ads})$ and so to maintain the same equilibrium constant, we must multiply the desorption rate by this same factor, noting that this correction is only significant for very weakly adsorbing particles with $E_{ads} \gtrsim -3k_BT$. This methodology does not treat adsorption of water to the NP explicitly but it is assumed that all binding energies are defined relative to the adsorption of water which is assigned an affinity $E_{ads} = 0k_BT$, and that the concentration of water is sufficiently high such that any region of the NP without an explicit adsorbate can be assumed to be covered in water. The results of simulations obtained with the updated CoronaKMC (i.e. including the molecule displacement) are shown in Table 3 and they suggest a notable variation in the abundances of proteins and lactose among different Al crystallographic orientations. Notably, on all surfaces studied, AS1C and BC consistently exhibited the highest protein abundances, while BLAC, LAC, and ALAC demonstrated moderate adsorption levels. In contrast, AS2C and BSA consistently displayed the lowest adsorption among the proteins considered in our simulations. Furthermore, when considering different Al facets, it is evident that surface 110 consistently exhibited the weakest average adsorption across all proteins. When the displacement is allowed, AS1C gains much more space in the corona by replacing other proteins, mostly BLAC, ALAC, and AS2C.

Figure 9 presents a comparison between the protein abundances in the corona on Al and Fe obtained using the enhanced version of the KMC algorithm with molecular displacements. As discussed earlier, this improved algorithm addresses computational efficiency concerns and more accurately represents long-term scenarios during protein corona formation. As shown in the Figure, these algorithmic improvements have a profound impact on the mass concentration of milk proteins on metallic surfaces, particularly on iron. In the original algorithm (Figure 8), proteins showed comparable mass abundances on both metals. However, the enhanced algorithm reveals a distinct change in the adsorption behavior of the AS1C protein on Fe and Al surfaces, characterized by a substantial increase in mass concentration compared to other proteins. The data in Table 3 show that in terms of mass abundance lactose ranks fourth among the corona components (see Figure S3). As compared to the algorithm without displacement [2], the protein abundance ranking on iron (NP radius 80 nm) surfaces changes to: AS1C $\gg$ BC $\geq$ BLAC $\geq$ ALAC $>$ AS2C $\approx$ BSA. Comparable affinity ranking is also now observed for aluminum surfaces (80 nm) studied in current work: AS1C $\gg$ BC $\geq$ BLAC $\geq$ ALAC $>$ AS2C $\approx$ BSA.

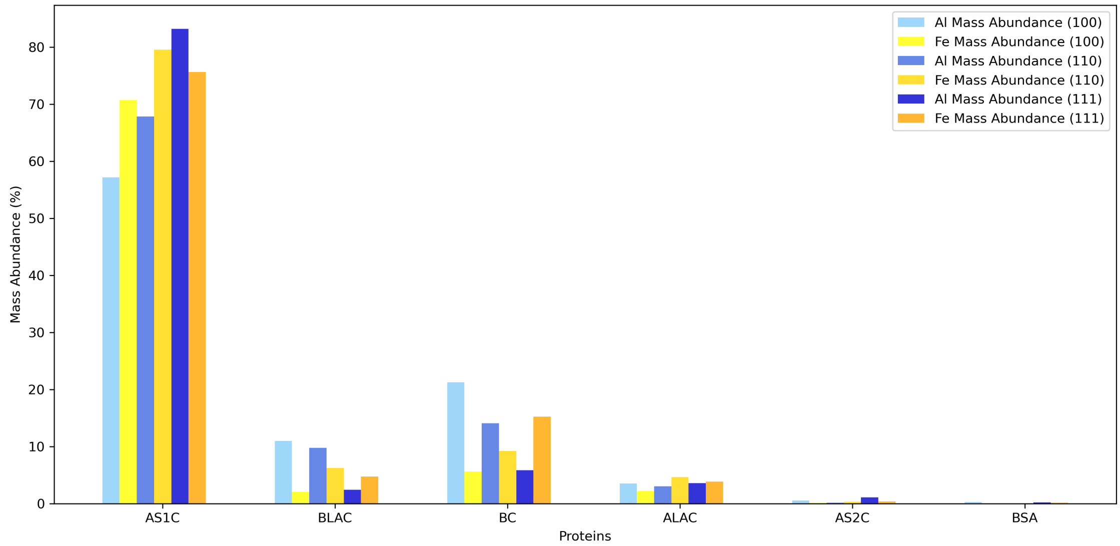

Figure 9: Mass abundance of proteins on Al (blues) and Fe (yellows) surfaces (100, 110, and 111) using the KMC model with molecule displacement and NP radius size of 80 nm.

Conclusions

In this work, we applied a multiscale computational model to study adsorption of milk solids on the metallic surfaces of aluminum, widely used in food processing/packaging. The milk model was composed of six most common milk proteins and lactose. To account for the size differences of selected milk constituents we used an improved competitive adsorption algorithm that can potentially achieve a realistic description of biocorona formation processes with diverse adsorbates (e.g. for predicting eco-corona).

Our computational model predicts strong binding of milk proteins to pure aluminum surface, which is in agreement with our previous observations for metallic iron surfaces [2]. For aluminum, we also found that αs1-casein and αs2-caseins exhibited the strongest binding to the metal, followed by bovine serum albumin, β-casein, β-lactoglobulins, and α-lactalbumin, which displayed weaker adsorption. We also found similar protein abundances in the corona for the two metals demonstrated by KMC simulation results. α s_1-casein dominates the adsorption as the most abundant protein on aluminum surfaces, with bovine serum albumin being the least abundant. We found a small difference in the predicted corona content between the two metals: β-casein and β-lactoglobulin prefer Al-100 and Al-110 to iron, while αs1-casein prefers Fe-100 and Fe-110 over aluminum. Although the adsorption energy regulates the interaction strength between proteins and surfaces, the mass concentration of proteins in the solution has a major effect on the amount of protein adsorbed onto the surface. Expanding the milk model by adding lactose into the mix did not alter the ranking of protein abundance in the corona. Despite the high concentration in the milk, lactose does not exceed the mass abundance of specific proteins such as AS1C due to its small size. In our model, it essentially makes a thin monolayer on the surface.

Overall, our freely accessible multiscale computational model [26] allows us to make predictions of the binding strength, preferred orientations, and relative abundance of the specified molecules on the specified material surfaces or nanoparticles and thus gives an insight into the mechanisms of fouling. We can compare different materials in terms of the protein binding affinity and corona content and optimize the processes in food and chemical industry. The presented methodology can be easily extended to other molecules, materials, and contexts involving the bionano interface such as environmental safety, health, medical devices, or toxicology.

Supporting Information

Table S1: Adsorption free energies for each SCA on Al surfaces; **Figure S1**: Water density profiles for aluminum slabs: (a) Al-100, (b) Al-110, (c) Al-111, **Figure S2**: Influence of the NP size on the adsorption energies; **Figure S3**: Milk molecules ranking based on mass abundance in the corona, **Figure S4**: Example of AlNP size-dependent interaction of α-lactalbumin: (a) 2 nm, (b) 5 nm, (c) 10 nm, (d) 20 nm, (e) 40 nm, (f) 50 nm, (g) 80 nm, (d) 100 nm; **Figures S5-S10**: Comparison of interaction of αs1-casein, αs2-casein,β-casein, α-lactalbumin, β-lactoglobulin, Bovine Serum Albumin, with different Al fcc surfaces: (a) Al-100, (b) Al-110, (c) Al-111; **Table S2**: Description of 820 milk proteins interaction with Al (100, 110, 111) based on the lowest energy values of the adsorption heatmaps.

Supporting Information File 1:
File Name: S1.pdf
File Format: PDF
Title: supporting material

Supporting Information File 2:
File Name: S2.pdf
File Format: PDF
Title: 820-milk-protein-table

Acknowledgements

The research presented in this study was supported by funding from Science Foundation Ireland under grant number 16/IA/4506, the European Union Horizon 2020 Programme grant 814572 (NanoSolveIT), and European Union Europe Programme grant 101092741 (nanoPASS). The authors extend their gratitude to the Irish Centre for High-End Computing (ICHEC) and the UCD Sonic High-Performance Computing Centre for providing the necessary computational resources for this research. The graphical abstract in this manuscript was created with https://biorender.com/ and it's content is not subject to CC BY 4.0.

References

1. Pulido-Reyes, G.; Leganes, F.; Fernández-Piñas, F.; Rosal, R. *Environ. Toxicol. Chem.* **2017**, *36* (12), 3181–3193.

2. Mosaddeghi Amini, P.; Subbotina, J.; Lobaskin, V. *Nanomaterials* **2023**, *13* (12), 1857.

3. Karaman, A.; Özer, B.; Pascall, M. A.; Alvarez, V. *Food Rev. Int.* **2015**, *31* (4), 295–318.

4. Lamberti, M.; Escher, F. *Food Rev. Int.* **2007**, *23* (4), 407–433.

5. Das, S.; Panda, S. H.; Bolem, M.; Pal, N.; Samantaray, B. R.; Thatoi, H. Facets of nanotechnology in food processing, packaging, and safety: an emerald insight. In *Bio-Nano Interface: Applications in Food, Healthcare and Sustainability*; Springer, 2021; pp 75–92.

6. Nabrzyski, M.; Gajewska, R.; Czuprynska-Rzepko, A.; Sandak-Bosak, K. *Rocz. Panstw. Zakl. Hig.* **1994**, *3* (45), 1181–1190.

7. Barabasz, W.; Albinska, D.; Jaskowska, M.; Lipiec, J. *Pol. J. Environ. Stud.* **2002**, *11* (3), 199–204.

8. Tang, L.; Thevenot, P.; Hu, W. *Curr. Top. Med. Chem.* **2008**, *8* (4), 270–280.

9. Williams, D. F. *Biomaterials* **2008**, *29*, 2941–2953.

10. Dobrovolskaia, M. et al. *Nat. Nanotechnol.* **2009**, *4*, 411–414.

11. Landsiedel, R.; Ma-Hock, L.; Kroll, A.; Hahn, D.; Schnekenburger, J.; Wiench, K.; Wohlleben, W. *Adv. Mater.* **2010**, *22* (24), 2601–2627.

12. Winkler, D. A.; Burden, F. R.; Yan, B.; Weissleder, R.; Tassa, C.; Shaw, S.; Epa, V. C. *SAR QSAR Environ. Res.* **2014**, *25* (2), 161–172.

13. Darabi Sahneh, F.; Scoglio, C.; Riviere, J. *PLOS ONE* **2013**, *8* (5), e64690.

14. Shao, Q.; Hall, C. K. *J. Phys. Condens. Matter* **2016**, *28* (41), 414019.

15. Angioletti-Uberti, S.; Ballauff, M.; Dzubiella, J. *Mol. Phys.* **2018**, *116* (21-22), 3154–3163.

16. Rouse, I.; Lobaskin, V. *Biophys. J.* **2021**, *120* (20), 4457–4471.

17. Rouse, I.; Power, D.; Brandt, E. G.; Schneemilch, M.; Kotsis, K.; Quirke, N.; Lyubartsev, A. P.; Lobaskin, V. *Phys. Chem. Chem. Phys.* **2021**, *23* (24), 13473–13482.

18. Wyrzykowska, E.; Mikolajczyk, A.; Lynch, I.; Jeliazkova, N.; Kochev, N.; Sarimveis, H.; Doganis, P.; Karatzas, P.; Afantitis, A.; Melagraki, G. et al. *Nat. Nanotechnol.* **2022**, *17* (9), 924–932.

19. Monopoli, M. P.; Walczyk, D.; Campbell, A.; Elia, G.; Lynch, I.; Baldelli Bombelli, F.; Dawson, K. A. *J. Am. Chem. Soc.* **2011**, *133* (8), 2525–2534.

20. Vilanova, O.; Mittag, J. J.; Kelly, P. M.; Milani, S.; Dawson, K. A.; Rädler, J. O.; Franzese, G. *ACS Nano* **2016**, *10* (12), 10842–10850.

21. Walkey, C. D.; Olsen, J. B.; Song, F.; Liu, R.; Guo, H.; Olsen, D. W. H.; Cohen, Y.; Emili, A.; Chan, W. C. W. *ACS Nano* **2014**, *8* (3), 2439–2455.

22. Nel, A. E.; Mädler, L.; Velegol, D.; Xia, T.; Hoek, E. M.; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M. *Nat. Mater.* **2009**, *8* (7), 543–557.

23. Eskin, N. A. M.; Shahidi, F. *Biochemistry of Foods*, 3rd ed.; Academic Press, 2012.

24. Brandt, E. G.; Lyubartsev, A. P. *J. Phys. Chem. C* **2015**, *119* (32), 18126–18139.

25. Power, D.; Rouse, I.; Poggio, S.; Brandt, E.; Lopez, H.; Lyubartsev, A.; Lobaskin, V. *Modelling Simul. Mater. Sci. Eng.* **2019**, *27* (8), 084003.

26. Physical- chemical aspects of protein corona: relevance to in vitro and in vivo biological impacts of nanoparticles. https://bitbucket.org/softmattergroup/ (accessed (accessed Dec 11, 2020)).

27. Subbotina, J.; Rouse, I.; Lobaskin, V. *Nanoscale* **2023**, *15* (32), 13371–13383.

28. Subbotina, J.; Lobaskin, V. *J. Phys. Chem. B* **2022**, *126* (6), 1301–1314.

29. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.

30. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D. et al. *Bioinformatics* **2013**, *29* (7), 845–854.

31. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.

32. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.

33. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modelling* **2006**, *25* (2), 247–260.

34. Sousa da Silva, A. W.; Vranken, W. F. *BMC Res. Notes* **2012**, *5*, 1–8.

35. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

36. Berendsen, H. J.; Postma, J. v.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

37. Potential of Mean Force (PMFs) for Zerovalent Aluminium (100-110-111) NanoParticles. https://doi.org/10.5281/zenodo.8334053 (accessed Sep 11, 2023).

38. Rouse, I.; Lobaskin, V. *Faraday Discussions* **2023**, *244*, 306–335.

39. Lopez, H.; Lobaskin, V. *J. Chem. Phys.* **2015**, *143* (24), 243138.

40. Rouse, I.; Power, D.; Brandt, E. G.; Schneemilch, M.; Kotsis, K.; Quirke, N.; Lyubartsev, A. P.; Lobaskin, V. *Phys. Chem. Chem. Phys.* **2021**, *23*, 13473–13482.

41. Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Protoc.* **2010**, *5* (4), 725–738.

42. Olsson, M.; Søndergaard, C.; Rostkowski, M.; Jensen, J. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

43. Søndergaard, C.; Olsson, M.; Rostkowski, M.; Jensen, J. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.

44. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. et al. *Nature* **2021**, *596* (7873), 583–589.

45. Hasenkopf, I.; Mills-Goodlet, R.; Johnson, L.; Rouse, I.; Geppert, M.; Duschl, A.; Maier, D.; Lobaskin, V.; Lynch, I.; Himly, M. *Nano Today* **2022**, *46*, 101561.

46. RDKit. https://www.rdkit.org/ (accessed (accessed Sep 01 2023)).

47. Dybeck, E. C.; Plaisance, C. P.; Neurock, M. *J. Chem. Theory Comput.* **2017**, *13* (4), 1525–1538.