ACS APPLIED MATERIALS & INTERFACES

www.acsami.org

Research Article

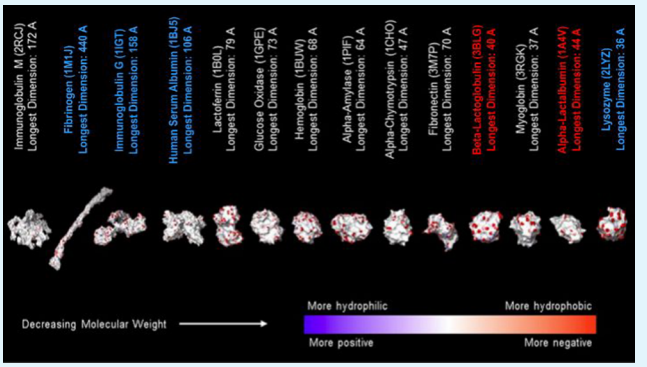
Protein Adsorption on Solid Surfaces: Data Mining, Database, Molecular Surface-Derived Properties, and Semiempirical Relationships

Matthew Cho,[‖] Zahra Mahmoodi,[‖] Prasad Shetty,[‖] Lauren R. Harrison, Maru Arias Montecillo, Ayyappasamy Sudalaiyadum Perumal, Gerardin Solana, Dan V. Nicolau, Jr., and Dan V. Nicolau*

Cite This: https://doi.org/10.1021/acsami.4c06759

Read Online

ABSTRACT: Protein adsorption on solid surfaces is a process relevant to biological, medical, industrial, and environmental applications. Despite this wide interest and advancement in measurement techniques, the complexity of protein adsorption has frustrated its accurate prediction. To address this challenge, here, data regarding protein adsorption reported in the last four decades was collected, checked for completeness and correctness, organized, and archived in an upgraded, freely accessible Biomolecular Adsorption Database, which is equivalent to a large-scale, ad hoc, crowd-sourced multifactorial experiment. The shape and physicochemical properties of the proteins present in the database were quantified on their molecular surfaces using an in-house program (ProMS) operating as an add-on to the PyMol software. Machine learning-based analysis indicated that protein adsorption on hydrophobic and hydrophilic surfaces is modulated by different sets of operational, structural, and molecular surface-based physicochemical parameters. Separately, the adsorption data regarding four "benchmark" proteins, i.e., lysozyme, albumin, IgG, and fibrinogen, was processed by piecewise linear regression with the protein monolayer acting as breakpoint, using the linearization of the Langmuir isotherm formalism, resulting in semiempirical relationships predicting protein adsorption. These relationships, derived separately for hydrophilic and hydrophobic surfaces, described well the protein concentration on the surface as a function of the protein concentration in solution, adsorbing surface contact angle, ionic strength, pH, and temperature of the carrying fluid, and the difference between pH and the isoelectric point of the protein. When applying the semiempirical relationships derived for benchmark proteins to two other "test" proteins with known PDB structure, i.e., β-lactoglobulin and α-lactalbumin, the errors of this extrapolation were found to be in a linear relationship with the dissimilarity between the benchmark and the test proteins. The work presented here can be used for the estimation of operational parameters modulating protein adsorption for various applications such as diagnostic devices, pharmaceuticals, biomaterials, or the food industry.



KEYWORDS: *protein adsorption, database, molecular surface, atomic hydrophobicity, Langmuir isotherm, multilinear regression with breakpoint*

INTRODUCTION

Protein adsorption is an ubiquitous biological process, such as in protein interaction with biological membranes,[1] and formation of microbial biofilms.[2] The adsorption of proteins on abiotic solid surfaces is also very important for many applications, such as medical and dental practice, e.g., implants,[3] catheters;[4] biomedical research, e.g., drug delivery and release;[5] and devices for diagnostics and drug discovery, e.g., assays,[6] microarrays,[7] and lab-on-a-chip.[8,9] Importantly, in most of these instances, and depending on the specific interest, protein adsorption can be either beneficial or deleterious. For instance, high-throughput proteomic microarrays[10] require both a robust immobilization of various proteins on the chip surface and the preservation of their native conformation—often contradictory desiderata. Nanoparticles are efficient carriers for proteins to be transported inside cells for therapeutic purposes,[11] but they can also be toxic[12,13] via specific inactivation of other essential proteins. Finally, protein adsorption on biomaterials can be undesirable,[14] e.g., if it elicits host immune response, but it is also important in

Received: April 24, 2024
Revised: May 8, 2024
Accepted: May 14, 2024

tissue engineering applications[15] since it modulates cell activation, adhesion, and wound healing.

The ubiquity and importance of protein adsorption have translated into a very large number of studies. For instance, currently, there are more than 3000 articles with "protein adsorption" in the title and over 12,000 as a keyword. However, despite this large body of data, presently it is not possible to predict, with reasonable accuracy, the amount of a specific protein, adsorbed on a specific surface, from a specific fluid medium, and far less is known about the preservation of the bioactivity of the adsorbed protein. Three classes of problems appear to frustrate the prediction of protein adsorption. The first, and the most obvious, emerges from the high complexity of protein adsorption.[16] Indeed, the protein adsorption process couples the complexity of protein structure, which cannot be predicted from first principles even for medium-sized proteins,[17] with the complexity of the molecular structure of the adsorbing surface, ranging from very ordered, e.g., crystals, to very disordered and dynamic, e.g., polymer brushes. Also, the interactions between the proteins and the adsorbing surface nano/microstructure, and between proteins themselves, add further layers of complexity to protein adsorption. The second class of problems emerges from the evolutionary history of research methodologies and interests. For instance, while the initial studies of protein adsorption used radiolabeling and gravimetry and more recent studies increasingly use Quartz Micro Balance, the use of ellipsometry appears to be ageless. However, while it is generally accepted that these techniques provide similar results, they are rarely employed and compared in the same work. Moreover, initial studies on protein adsorption were focused on biomaterials applications, involving relatively large amounts of adsorbed proteins, whereas the present interests are more focused on the development of diagnostic devices, involving relatively low concentrations of immobilized proteins. Furthermore, the emergence of nanoparticles for various uses, such as in diagnostics and medical imaging, has shifted interest away from the "simpler" interaction between proteins and flat surfaces to the more complex interactions between individual protein molecules and artificial objects with similar sizes.[18] Finally, a third class of problems is "cultural" in nature. A previous effort[8] in mining the data in the literature on protein adsorption revealed that only a minority of scientific articles, i.e., approximately 10%, report information about *all* three classes of inputs modulating protein adsorption: (i) the protein (rarely identifying the PDB ID, or even its source); (ii) adsorbing surface (contact angle or surface tension, charging or zeta potential, nano/microtopography, surface chemistry); and (iii) fluid medium (pH, ionic strength, temperature), although it is expected that all of these parameters critically influence protein adsorption.[19,20] Moreover, while most studies correlated the protein concentration on the surface with that in solution, very few have investigated the impact of more than one other parameter, e.g., adsorbing surface hydrophobicity, ionic strength of the solution, etc.

Addressing the lack of predictive models for protein adsorption and capitalizing on the large size and variability of data in scientific literature, we have considerably upgraded the previously reported[8] Biomolecular Adsorption Database (BAD), with the data reported in the last 15 years. We then used this enlarged database, adding the quantification of shape and physicochemical properties on the molecular surfaces of proteins using an in-house built program (ProMS), to derive explicit relationships describing protein adsorption for four "benchmark" protein types, i.e., quasi-spherical lysozyme, quasi-ellipsoidal albumin, quasi-cylindrical fibrinogen, and Y-shaped IgG. Finally, we show how these explicit relationships (implemented in an app available online) can be used to estimate protein adsorption of other proteins, for which data is sparse or not existent, provided that the protein of interest is structurally similar to one of the four benchmark proteins.

METHODS

Collection of Data and Database Organization. The present Biomolecular Adsorption Database, (BAD 2.0), upgraded from BAD 1.0,[8] is an archive of the data regarding protein adsorption on *flat* solid surfaces, as reported in peer-reviewed literature. Designed as a web-orientated database, BAD 2.0 comprises only data that quantitatively report, completely, the descriptors of all three classes of descriptors relevant to protein adsorption: (i) the adsorbed protein, preferably identifiable in the Protein Data Bank (PDB);[21] to allow further derivation of protein descriptors; (ii) the *flat* surface (type, water contact angle or surface tension); and (iii) fluid descriptors, i.e., protein concentration in solution, pH, ionic strength, and temperature. BAD 2.0 also comprises auxiliary data, i.e., method of measurement, and the DOI of the relevant references for easy retrieval.

BAD 2.0 is an open and freely accessible database (https://www.bionanoinfo.com/bad/), which can be continuously upgraded, either via the inclusion of new data during scheduled maintenance or via new entries proposed by individual researchers and vetted by database maintenance team. An image of the entry web portal to the BAD is presented in Figure SI 5, and the entity relationship diagram for BAD 2.0. is presented in Figure SI 6.

The source data of BAD 2.0 comprise experimental results from adsorption isotherms, plateaus of adsorption kinetics experiments, and single adsorption experiments (when surfaces are incubated in a protein solution with a known initial protein concentration). The primary data was collected from the open literature using the main literature search engines, e.g., Scopus, Google Scholar, and ISI Web of Science, and using several combinations of relevant keywords, e.g., (protein adsorption)-AND((contact angle)OR(surface tension)). The initial collection of data was followed by critical analysis of the results and augmentation of the published data, e.g., contact angles for identical surfaces, whenever possible.

BAD 2.0 comprises 865 protein adsorption records. In several instances, the pair of values of (protein concentration in solution) vs (protein concentration on the surface) had to be estimated from the original contributions through graphical interpolation of data points in the original figures. 78.5% of the data collected in BAD 2.0 are the result of adsorption isotherms data, and 21.5% from single adsorption experiments. Only adsorption data from solutions containing a single protein are reported in BAD 2.0. The relevant bibliography used to build BAD 2.0 is presented Supporting Information SI 1, and a complete data set used to build BAD 2.0 is presented in tabular format in Supporting Information SI 2.

The architecture of BAD 2.0 comprises several linked tables that are displayed on web pages. While the focus of the data explored in this report is presented in the core BAD table, the structure also features additional tables describing the BAD protein properties and the output values obtained by a molecular surface calculator (ProMS). This structure reduces database-wide redundancy while also providing structure for the expansion of BAD. The tables are designed to be quickly filterable, sortable, and searchable to facilitate targeted data extraction. The database was created using MySQL and additional Python scripts were developed to help with the maintenance and scalability of the database and to facilitate the validation of new data.

Protein Adsorption Variables Reported in BAD 2.0. The physicochemical data in BAD 2.0 are organized along three classes of descriptors, i.e., for the adsorbed protein, for the adsorbing surface, and for the protein-carrier fluid.

Protein Descriptors. BAD 2.0 contains data regarding the adsorption of 20 proteins (all with their respective PDB IDs, and their proportion in the total data, indicated in parentheses): albumin, human (1AO6), or

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

bovine (3V03): 50.41%; fibrinogen (3GHG): 12.83%; immunoglobulin G (1IGT): 10.75%; lysozyme (2LYZ): 9.25%; myoglobin (1MBO): 3.74%; α-lactalbumin (1HML): 2.89%; fibronectin (3M7P): 2.2%; insulin (4INS): 1.85%; β-lactoglobulin (3BLG): 1.39%; Cry1Ac protoxin (4ARX): 1.27%; glucose oxidase (1CF3): 0.92%; α-2-macroglobulin (4ACQ): 0.69%; immunoglobulin M (6KXS): 0.69%; α-amylase (1PIF): 0.46%; lactoferrin (1BOL): 0.46%; α-chymotrypsin (2CHA): 0.23%; cytochrome c (1HRC): 0.12%; hemoglobin (1BUW): 0.12%. As opposed to the previous embodiment of BAD,[8] all proteins are associated now with their PDB structures.

BAD 2.0 records protein primary descriptors, as determined from their respective PDB ID, or in a few instances, from alternative sources: number of residues, secondary structure motifs (α helices, β sheets, and S−S bonds). Using the protein sequence in FASTA format,[22] BAD 2.0 also records secondary descriptors: molecular weight, isoelectric point (calculated by averaging the pKa of the amino acids in the primary structure), and protein rigidity calculated as a ratio between the sum of the residues contributing to α helices and β sheets plus two times the number of SS bonds, divided by the total number of residues.

Finally, *tertiary descriptors*, i.e., total area, total hydrophobicity and hydrophilicity, and hydrophobic and hydrophilic areas, were calculated by probing the protein molecular surface using the protein PDB structures and a methodology reported elsewhere.[23,24] This methodology is also capable of the calculation of the total positive and negative charges and positive and negative areas on protein surface. Briefly, to calculate these descriptors, the protein PDB structure was probed by a virtual ball with a radius of 20 Å (representing the threshold after which the properties on the protein molecular surface do not vary much[23]) using an upgraded version of Conolly's algorithm.[25,26] At the point of contact between the protein molecular surface and the probe, the identity of the residue, or the atom, will translate in a (p,x,y,z) data point, where p stands for the physicochemical property on the surface of the contacted residue or atom, and (x,y,z) stands for the spatial position. The program is set to calculate the charges at the pH of the fluid environment, which translates to a particular ionization of a residue, and then derived from quantum mechanics calculations, as reported before.[23] The hydrophobicity/hydrophilicity on the protein molecular surface is calculated using an experimentally measured hydrophobicity scale of residues in pentapeptides.[27] These residue-based hydrophobicities were used to determine the atom-based ones, following a procedure reported elsewhere.[23] The values of the descriptors of the proteins in BAD 2.0 are presented in Supporting Information SI 2.

The computing methodologies were implemented in a software application, ProMS (standing for Protein Molecular Surface Program), which provides an easy-to-use graphical user interface (GUI) with the option for the user to select the input parameters. ProMS also has a built-in visualization mode and the option to continuously run many molecules independently. ProMS, which is written in Python3.9, uses Pandas and Tkinter modules for data manipulation and generation of the GUI, respectively. Object-oriented programming (OOP) principles were employed to ensure easy scaling and feature addition. ProMS also makes use of two external programs: Connolly's msRoll executable,[26] and PyMOL,[28] an open-source molecular visualization system. The main advantage of ProMS over its predecessor PSPC[23,24] is its ability to streamline several tasks into a single application. The program also includes a "batch mode", to automatically run multiple molecules consecutively. ProMS runtime is limited by the execution speed of Connolly's msRoll program. The generation of the molecular surface is linear in the number of atoms in the molecule and quadratic in the probe radius. ProMS was developed as an open-source plugin which extends Schrodinger's PyMOL, thus allowing facile distribution. A detailed description of ProMS is presented in Supporting Information SI 3.

The calculation of shape-relevant descriptors of the protein molecular surfaces, i.e., solvent excluded volume (V) and molecular area (A), as measured by a probe with a radius of 20 Å (close to an actual nanotopography of an experimentally flat adsorbing surface), allowed the calculation of protein sphericity as follows:

$$\text{sphericity} = \frac{\pi^{1/3}\cdot(6V)^{2/3}}{A} \quad (1)$$

The sphericity of the proteins allowed the estimation of the density of a protein monolayer (assuming that proteins are rigid objects), using correlation derived from published work[29] correlating the sphericity of objects with the porosity of a layer comprising these objects (presented in Supporting Information SI 2).

Cluster analysis was applied to the classification of proteins, according to their shape properties, and according to their tertiary descriptors, using the joining (tree clustering) method, amalgamation according to single linkage, and Euclidian distances, as implemented in the software package Statistica (from TIBCO Software, Inc.).

Descriptors of the Adsorbing Surface. Presently, BAD 2.0 reports protein adsorption on four types of surfaces, in the following proportions: polymers 36.88%; oxides 30.98% (of which the majority is silica, with 21.85% of the whole data, 41.27% of which is modified silica); metals and semiconductors 24.86% (of which the large majority is silicon, with 17.34%, 88% of which was functionalized silicon); and finally self-assembled monolayers 7.28%. The only descriptor of the adsorbing surfaces is hydrophobicity, usually reported as contact angle (for water) or more rarely as surface tension. In a number of instances, and only for common types of surfaces, e.g., silica, the literature did not report the surface hydrophobicity, but it did describe the nature of the surface to an extent that allows the assigning of contact angle (or surface tension) from other reports with the same authors or assumed equal to an average of values reported elsewhere for the same surface. BAD 2.0 reports contact angles as an average of the advancing and receding contact angles,[30] when available. Finally, where applicable, BAD 2.0 also reports details of surface chemistry, e.g., for SAMs, polymers, and functionalized surfaces.

Other descriptors of the adsorbing surface relevant to protein adsorption are surface charging, positive or negative, measured as the ζ potential, surface topography, and local mechanical properties. However, collecting the data regarding these descriptors raised considerable difficulties, e.g., scarcity of reports and lack of standardization of reporting. For instance, fewer than 1% of the reports state the ζ potential of the adsorbing surface. Furthermore, while the literature regarding protein adsorption on surfaces with distinct micro- or nanotopographies is considerable and increasing, BAD 2.0 does not record this descriptor of the adsorbing surface because of the difficult comparison with protein adsorption on flat surfaces. Consequentially, all of the data in BAD 2.0 refer only to microscopically flat surfaces, including several instances where the size of the topographic features is considerably larger than the size of the adsorbed protein, e.g., microbeads. Finally, the protein adsorption data on softer surfaces, often intimately linked with the molecular topographies, e.g., polymer brushes, was recorded but not used for advanced statistical analysis. For all of these reasons, ζ-potential, surface topography, and local mechanical properties are not recorded in BAD 2.0, but can be accessed, if available, via a DOI link to the respective reports.

Descriptors of the Protein-Carrier Fluid. BAD 2.0 reports various buffer solutions with distinct compositions used for protein adsorption experiments. The descriptors recorded are pH, ionic strength, and temperature (where available). If the ionic strength was not explicitly reported, but the fluid composition was, the ionic strength was calculated using a calculator available online (https://www.liverpool.ac.uk/pfg/Research/Tools/BuffferCalc/Buffer.html).[31] For experiments citing "room temperature", the value was assumed 22 °C. Phosphate buffers with no added NaCl are designated by PB, and those with added NaCl as PBS.

Protein Concentration on the Surface. The source of the values of the protein concentration on the adsorbing surface, expressed in mg/m^2, comprised (i) adsorption isotherms data; (ii) adsorption kinetics experiments; and (iii) single adsorption experiments. Nineteen methods were used for the quantification of the protein mass adsorbed: ellipsometry 25.84%; radiolabeling 23.87%; quartz micro balance (QCM) 16.8%; Lowry method 12.86%; capillary gravimetry 3.01%; surface plasmon resonance (SPR) 2.9%; UV absorption 2.32%; sedimentation field-flow fractionation (SdFFF) 2.2%; attenuated total

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

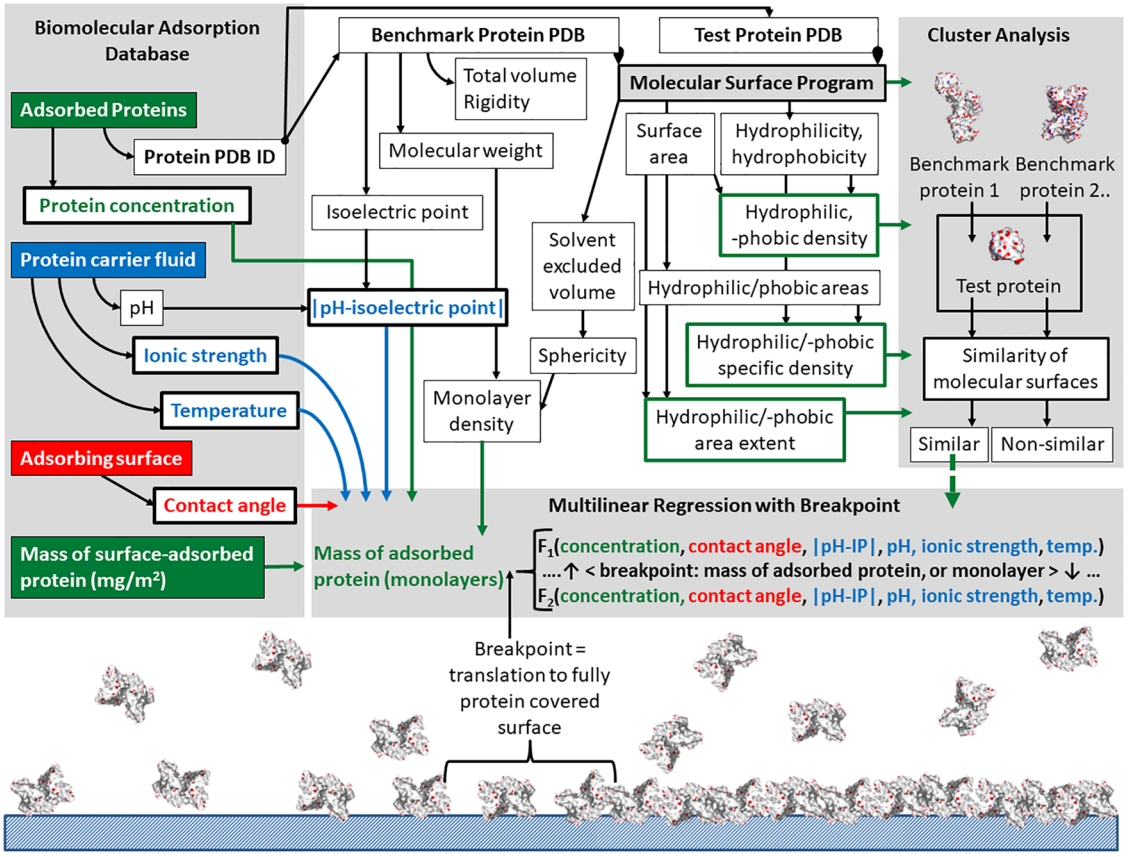


Figure 1. Conceptual design of the study of protein adsorption. The data in the Biomolecular Adsorption Database (left), together with data from the quantification of protein properties, including on their molecular surfaces (middle top), are inputs to the regression analysis leading to semiempirical correlations for representative proteins (middle bottom). These correlations can be tentatively applied to estimate protein adsorption for other proteins if their molecular surfaces are similar (assessed by cluster analysis, right). The scheme at the bottom is a simplified representation of the two stages of protein adsorption with a breakpoint threshold for the transition from the adsorbing surface-controlled to the protein surface-controlled process.

reflectance-Fourier transform infrared (ATR FTIR) 1.62%; total internal reflection fluorescence (TIRF) 1.51%; neutron reflectivity 1.16%; whispering gallery mode (WGM) 0.93%; X-ray reflectometry 0.93%; retention time high-performance liquid chromatography (RT HPLC) 0.81%; optical waveguide light spectroscopy (OWLS) 0.7%; size exclusion (SE) HPLC 0.58%; streaming potential 0.46%; and florescence spectroscopy 0.12%.

Protein Concentration in Solution. The protein concentration in the carrier fluid, expressed in mg/ml, needed to be estimated according to the type of protein adsorption experiments: (i) for adsorption isotherms, the near-equilibrium concentration in solution after adsorption, (ii) for adsorption kinetics experiments, the bulk protein concentration, and (iii) for single adsorption experiments, the initial protein concentration.

Other Information. BAD 2.0 also provides links to additional information about the proteins, buffers, surfaces, and methods of measurement, as well as the DOI of the articles from which the data was collected, through a link to previous work.[23]

Machine Learning. *Feature Exploration.* We used machine learning principles to explore the importance of specific variables in the prediction of protein adsorption. The analysis was performed with Python in Jupyter Notebook, a popular tool used in the data science community due to its ability to combine coding snippets, markdown text, and data visualization. To manipulate, analyze, and visualize the protein adsorption data set, various Python software packages were used, notably pandas, NumPy, Scikit-learn, and matplotlib.

The complete compiled data set has over 50 features, where each feature is represented by a column. The idea of this feature exploration step was to determine the variables most important to protein adsorption and to eliminate the less important or redundant ones. From the main data set, two smaller curated data sets were created. The hydrophobic set had 111 entries, and the hydrophilic set had 71. Furthermore, feature subsets were created. These subsets were groups of explanatory variables that could be used to predict surface adsorption. For each subset, the dependent variable, or in other words, the value being predicted, was the surface concentration of the adsorbed protein.

For each subset of features, an analysis was performed with both the hydrophobic and hydrophilic data sets. First, the analysis started by normalizing the variables. The variables were then fit to a ridge regression model. This model was chosen because it is commonly used in feature selection pipelines. Additionally, this machine learning model inherently considers the role of each input variable and its regularization term prevents overfitting. Once the model was fit, the values of the parameter coefficients, which represent the relative importance of each input, were obtained and plotted.

Model Fitting & Performance. The feature analysis provided insight into the relative importance of the independent variables within each subset. However, this analysis did not provide any quantitative metric as to how well these variables predicted protein adsorption. For that reason, we needed to measure the performance of the model. For each subset, linear regression, ridge regression, and random forest regressor models were fitted (3 models, 2 data sets so 6 total models per group of variables). For each model, the data set was shuffled and then split into training and testing sets via a 5-fold cross-validation step. This was important to reduce the model bias and variance. Individual model performance was then evaluated based on the averaged coefficient of determination (R^2) across the five folds.

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

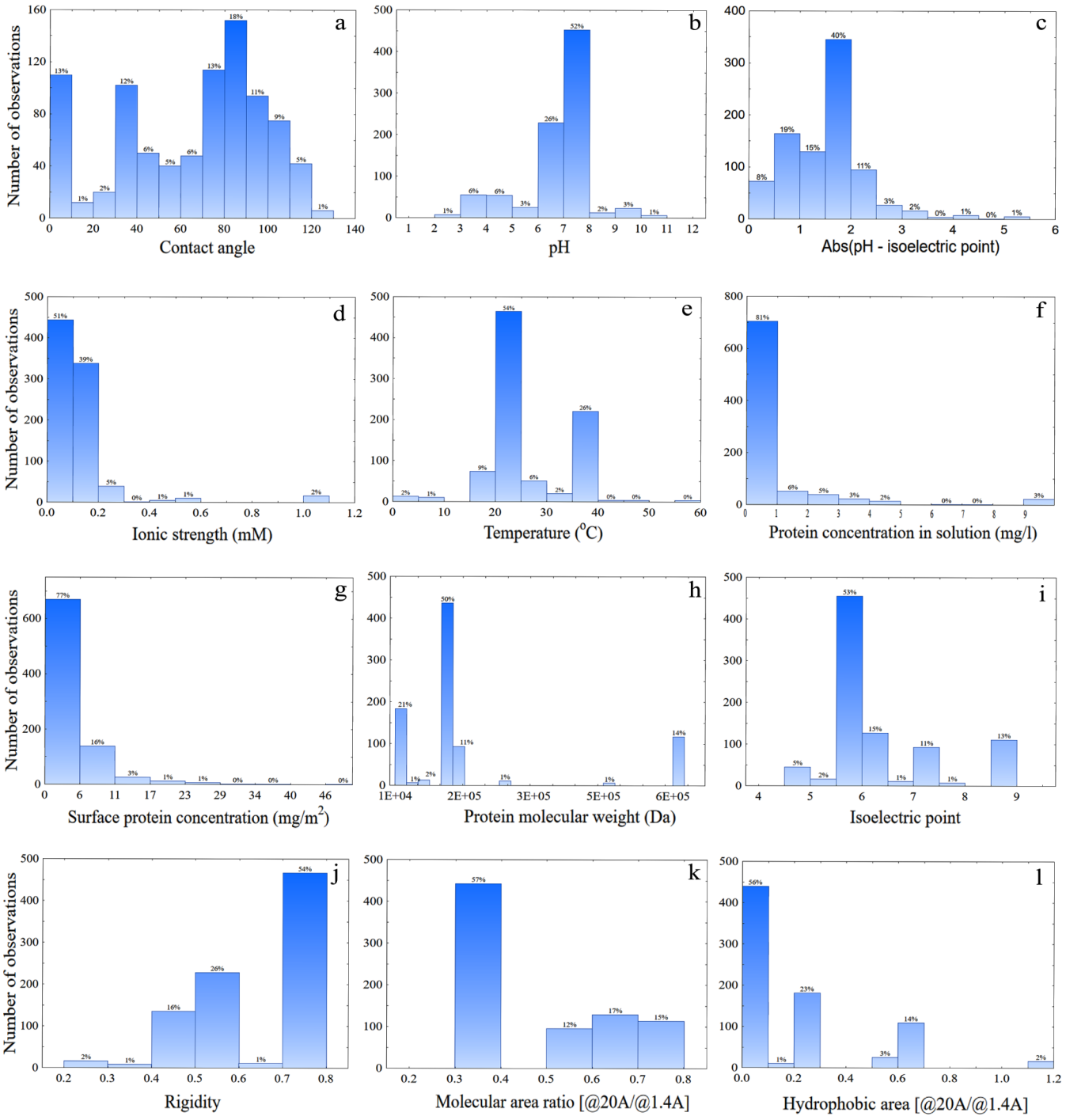


Figure 2. Distribution of the protein adsorption descriptors in BAD 2.0: (a) adsorbing surface hydrophobicity, (b) pH of the carrier fluid, (c) difference between the pH and the protein isoelectric point (absolute values), (d) ionic strength of the carrier fluid, (e) temperature of the fluid, (f) concentration in solution, (g) concentration on surface, (h) protein molecular weight, (i) isoelectric point of the protein, (j) protein rigidity (ratio of residues in α helices, β sheets, and S−S bonds and the total residues), (k) ratio of molecular areas quantified using probes with radii of 1.4 and 20 Å, respectively, and (l) ratio hydrophobic areas using 1.4 and 20 Å probes.

Regression Analysis. The data in BAD was used to derive correlations linking the output variable, i.e., protein concentration on the surface, with the input variables, i.e., protein concentration in solution, adsorbing surface descriptors (contact angle), carrier fluid descriptors (temperature, pH, ionic strength), and protein descriptors (primary, secondary, or tertiary descriptors). These relationships were derived using least-squares regression analysis with Levenberg−Marquardt and Gauss−Newton algorithms, applied to piecewise multilinear regression with breakpoint, as implemented in the software package Statistica (from TIBCO Software, Inc.). In most instances, 500 iterations were sufficient to reach the lowest possible loss with the convergence criterion set to 10^{-6} (the optimization stops when the changes in the parameters from iteration to iteration are no more than the convergence criterion).

The piecewise multilinear regression with breakpoint has the form of

$$\text{surfconc}^{-1}_{\text{surfconc}<\text{bkpt}} = B_0 + \sum_{i=1}^{n} a_i x_i, \text{ for surfconc} < \text{bkpt} \quad (2)$$

and

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

$$\text{surfconc}^{-1}_{\text{surfconc}>\text{bkpt}} = B'_0 + \sum_{i=1}^{n} a'_i x_i, \text{ for surfconc} > \text{bkpt} \quad (3)$$

where surfconc^{-1} is the inverse of the protein concentration on the surface, surfconc (mg/m^2); B_0 and B'_0 are constants for before and after breakpoint; x_i are the input variables, with a maximum $i = 5$ for the maximum number of variables used for regression analysis (concentration in solution represented as its inverse, i.e., solconc^{-1}, contact angle, ionic strength, absolute value of the difference between solution pH and protein isoelectric point, and temperature); and bkpt (with the same units as the output variable, i.e., surfconc^{-1}) is the breakpoint calculated by the automatic procedure for regression. Equations (2) and (3) represent a linearization of the Langmuir isotherm for protein adsorption, with the breakpoint marking the transition between a regime described by eqs (2) and (3), respectively. While very good statistical correlations were obtained using surface concentration expressed in mg/m^2, an alternative regression analysis was performed using surface concentration expressed in fractions (or numbers) of monolayers, thus allowing the representation of the breakpoint in terms of monolayers. Another "normalization" of the data was performed by representing the protein concentration in solution in terms of moles/volume.

First, the multilinear regression analysis was applied to the four "benchmark" proteins, which have the highest representation in BAD 2.0: (i) albumin (PDB ID: 1BJ5) representing 51.17% of the BAD data; (ii) fibrinogen (PDB ID: 1M1J), 11.71%; (iii) immunoglobulin G (PDB ID: 1IGT), 10.77%; and (iv) lysozyme (PDB ID: 2LYZ), 9.25%, all totaling 82.9%. These sets were further reduced as follows: (i) select only data for temperatures between 20 and 30 °C; (ii) data for concentration in solution below 5 mg/mL; and (iii) data measured by the major analytical methods, i.e., radiolabeling, ellipsometry, and QCM. The selected data used for regression analysis is presented in Supporting Information SI 2, for adsorption on hydrophilic and hydrophobic surfaces, respectively. The semiempirical correlations were implemented in a Matlab estimation program (presented in Supporting Information SI 4, and the actual MATLAB program can be accessed at www.bionanoinfo.com/resources).

The regressed correlations of the benchmark proteins were then applied to the experimental conditions reported for two other ("test") proteins, i.e., α-lactalbumin (1A4 V, 2.81%) and β-lactoglobulin (3BLG, 1.41%), for which the extent of the data was insufficient to derive similar relationships, but covered both hydrophilic and hydrophobic surfaces (insulin, 1HLS, at 1.87% of BAD data was excluded as covering only one hydrophobic surface, with the same contact angle). Finally, the errors between the estimations and the actual reported experimental results were correlated with the similarity between the benchmark proteins, i.e., albumin, fibrinogen, immunoglobulin G, lysozyme, and test proteins, as resulting from the clustering analysis using hydrophilicity-related intensive descriptors, i.e., density, specific density, and extent of hydrophilic area, for comparisons of protein adsorption on hydrophilic surfaces and hydrophobicity-related descriptors for hydrophobic adsorbing surfaces, respectively.

RESULTS AND DISCUSSION

The conceptual plan of the present study, schematically shown in Figure 1, comprises three modules: (i) building the Biomolecular Adsorption Database (BAD 2.0), (ii) finding correlations describing protein adsorption for the most representative proteins present in the database, and (iii) exploring the possibility of using these correlations for proteins for which the lack of experimental data does not allow the derivation of similar correlations. The last two study modules were supported by quantification of the properties of the adsorbed proteins.

Distribution of the Descriptors of Protein Adsorption in BAD 2.0. BAD 2.0. effectively represents an ad hoc, crowd-sourced, very large-scale multifactorial experiment collected from the open literature.

The distribution of the major classes of parameters modulating protein adsorption, i.e., (i) descriptors of the adsorbing surface; (ii) descriptors of the protein-carrier fluid; and (iii) protein descriptors, as represented in BAD 2.0., is as follows.

Surfaces. The inspection of the distribution of the adsorbing surface contact angles (Figure 2a) suggests that BAD 2.0. comprises two distinct populations: a larger data set regarding hydrophobic surfaces, with a nearly normal distribution, centered around 90°, and a slightly smaller hydrophilic bimodal data set, with peaks around 0° and around 30°. A more detailed characterization of the adsorbing surfaces would benefit both the comprehensiveness of the database and a more precise derivation relationships predicting the amount of the adsorbed protein.

In principle, a more detailed characterization can be provided by surface analysis, such as XPS and ATR FTIR, which would quantitatively provide information regarding surface chemistry. However, this desideratum cannot be achieved because the studies regarding protein adsorption that use XPS analysis for surface characterization represent a minority of the reports in the literature, i.e., less than 6% (ATR FTIR information is even more rare), and more importantly, they do not also report other important parameters. A similar argument applies for reporting of the zeta potential of the adsorbing surfaces, which would provide information regarding its charging prior to contact with the carrier fluid. In addition, the works reporting zeta potential of the adsorbing surface also represent a small minority, i.e., below 5%, presenting the same drawback of incomplete coverage of parameters relevant to protein adsorption. However, the charge of the dry adsorbing surface will, in any case, change in contact with the buffers with various pH and ionic strengths. While the information regarding surface chemistry and charging did not reach a level that would make it useful for statistical analysis, the respective reports are highlighted in the bibliography used for the assembly of BAD 2.0 (Supporting Information SI 1.).

Fluid Media. The distribution of pH of the buffers used for protein adsorption experiments follows a near-normal distribution (Figure 2b), with most pH data centered around physiological pH (7.4) but ranging from as low as 2 to as high as 11. Furthermore, the difference between the pH of the carrier fluid and the isoelectric point of the adsorbed protein (Figure 2c) suggests that most experiments are run with proteins not too far away from their neutral electrical state, that is, when the pH is equal to the isoelectric point of the protein. The ionic strength of the buffers used in protein adsorption experiments (Figure 2d) is prevalently at the low-end values, although there are few instances where it reaches 1 mM. Finally, the protein solution temperatures range from 0 to 60 °C (Figure 2e), with most data around room temperature but also a sizable minority at higher temperatures.

Protein Concentrations. The distributions of the protein concentration in solution and on the adsorbing surface are presented in Figure 2f,g, respectively. Much of the data in the BAD 2.0 reflects protein adsorption experiments at low concentrations in solution (87% below 2 mg/mL) and, consequently, a low concentration on the surface (77% below 6 mg/mL).

Proteins. The molecular weights and the isoelectric points of the proteins in BAD 2.0. are presented in Figure 2h,i, respectively. While the range of the molecular weights is large, i.e., from ~16 kDa for α-lactalbumin to ~650 kDa for fibrinogen,

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

their distribution is considerably uneven due to the over-representation of albumin (~133 kDa), fibronectin (~631 kDa), and a cluster of smaller proteins, e.g., lysozyme, insulin, and hemoglobin. The distribution of the isoelectric point is more even, with over-representation of albumin (IP = 5.47), IgG (IP = 6.57), and lysozyme (IP = 8.34).

The proteins for which the adsorption on surfaces was reported in BAD 2.0 are extremely varied in their properties, as can be easily observed from the visualization of their molecular surfaces (presented in Figure 3), with molecular surfaces mapped by a probe with a radius of 20 Å. The visualization of molecular surfaces and the quantification of the structural and physicochemical properties used ProMS, an in-house developed program (a description of ProMS is presented in Supporting Information SI 3). While interfacing a protein with an ideally flat surface is equivalent to probing the molecular surface by a probe of infinite radius, most surfaces present rugosities equivalent to a smaller probe radius. The upper threshold for the probe radius of 20 Å was chosen for two reasons. First, the flat surfaces on which proteins adsorb still present various degrees of rugosity. For instance, glass surfaces treated for reduction of roughness still present 2 nm features[32] and freshly cleaved mica presents features around 25 Å.[33] Second, it was observed[23] that the physicochemical properties quantified on the molecular surfaces of proteins do not vary significantly beyond a probe radius of 20 Å.

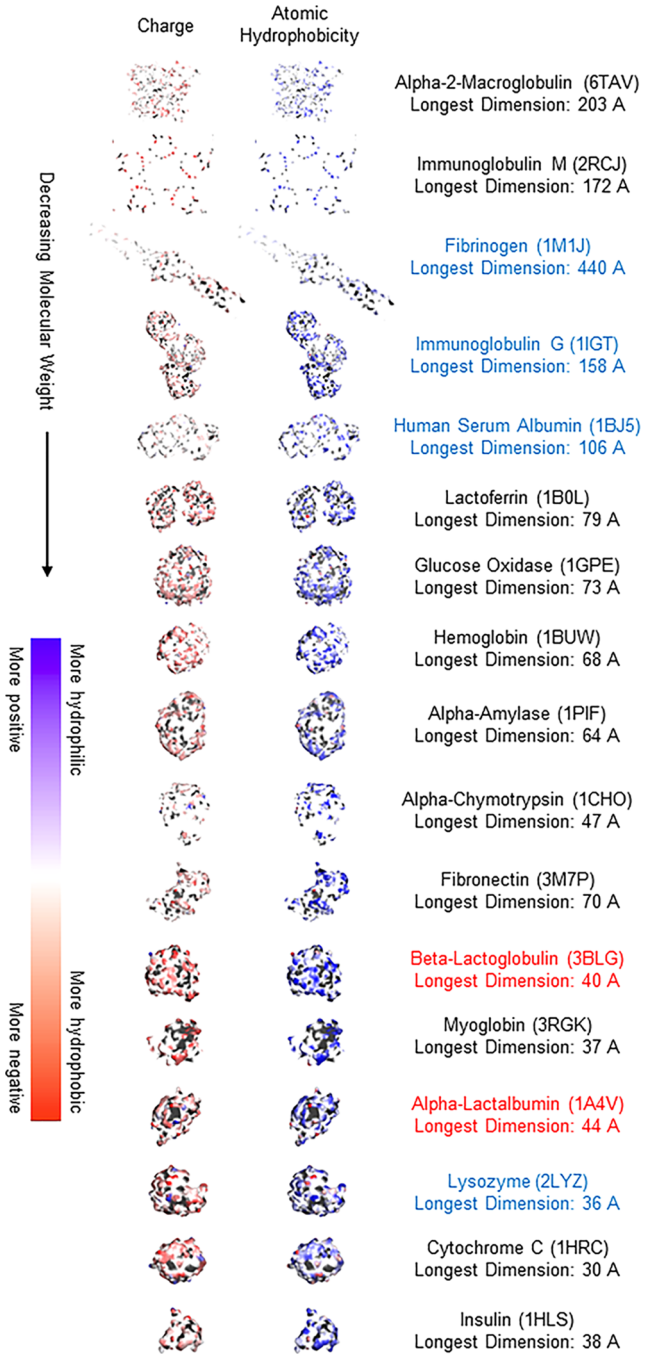


Figure 3. Molecular surfaces of the proteins present in BAD 2.0 (not to scale) probed with a 20 Å radius. Benchmark proteins are indicated in blue, and the test proteins are indicated in red.

Further analysis of the parameters relevant to protein adsorption was obtained by analyzing the secondary and tertiary protein descriptors. For instance, a high ratio of the residues in α helices, β sheets, and connected by S−S bonds, reported to the total residues, suggests a higher rigidity of the protein. The inspection of the distribution of this ratio (Figure 2j) reveals that 54% of proteins in BAD 2.0 are more rigid, with a smaller (42%) cluster of more flexible proteins. The distribution of the ratios of molecular surfaces, mapped with probes with a large radius, e.g., 20 Å, versus those mapped by smaller probes, e.g., 1.4 Å, equivalent to the dimensions of the water molecule (molecular surfaces presented in Figure SI 7), can be a measure of the compactness of the protein. The distribution of these ratios (Figure 2k) suggests that most of the proteins in BAD 2.0 (57%) present complex geometries with considerable concave inner volumes, difficult to be in contact with the adsorbing surfaces, whereas a separate cluster exists with relatively convex geometries, e.g., spherical, ellipsoidal. Finally, the reduction of the exposure of the hydrophobic core to probes with larger areas, such as the adsorbing surface, could indicate a reduced propensity for adsorption on hydrophobic surfaces. The distribution of these ratios (Figure 2l) indicates that 56% of the proteins present less than 10% of the hydrophobic core accessible to small molecules, e.g., water, than flat adsorbing surfaces, with smaller clusters presenting a lesser reduction of the exposure of the hydrophobic core.

Comparison of Measurement Methods for Protein Adsorption. Research related to protein adsorption has evolved continuously in the last decades, resulting in a large body of experimental reports (the selected bibliography used for building BAD 2.0 is presented in Supporting Information SI 1), including complete books reporting the works presented at periodic conferences dedicated exclusively to protein interactions with surfaces.[34−36] To a large extent, this progress has benefitted from the steady evolution of measurement methods, comprehensively reviewed recently.[37] The large variety of measurement methods for protein adsorption is not, however, distributed evenly. Out of the 19 methods identified by scanning the literature on protein adsorption, four stand out as being used more frequently. Indeed, ellipsometry, QCM, radiolabeling, and the Lowry method represent approximately 80% of the total measurements of protein adsorption in the last four decades. Moreover, the popularity of using these methods varied over time (Figure SI 8). The Lowry method was primarily used until the late 1980s, followed by radiolabeling until the 2000s. Conversely, the popularity of ellipsometry looks constant, and the QCM is increasingly used for protein adsorption measurements, including for the assessment of convoluted interactions

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

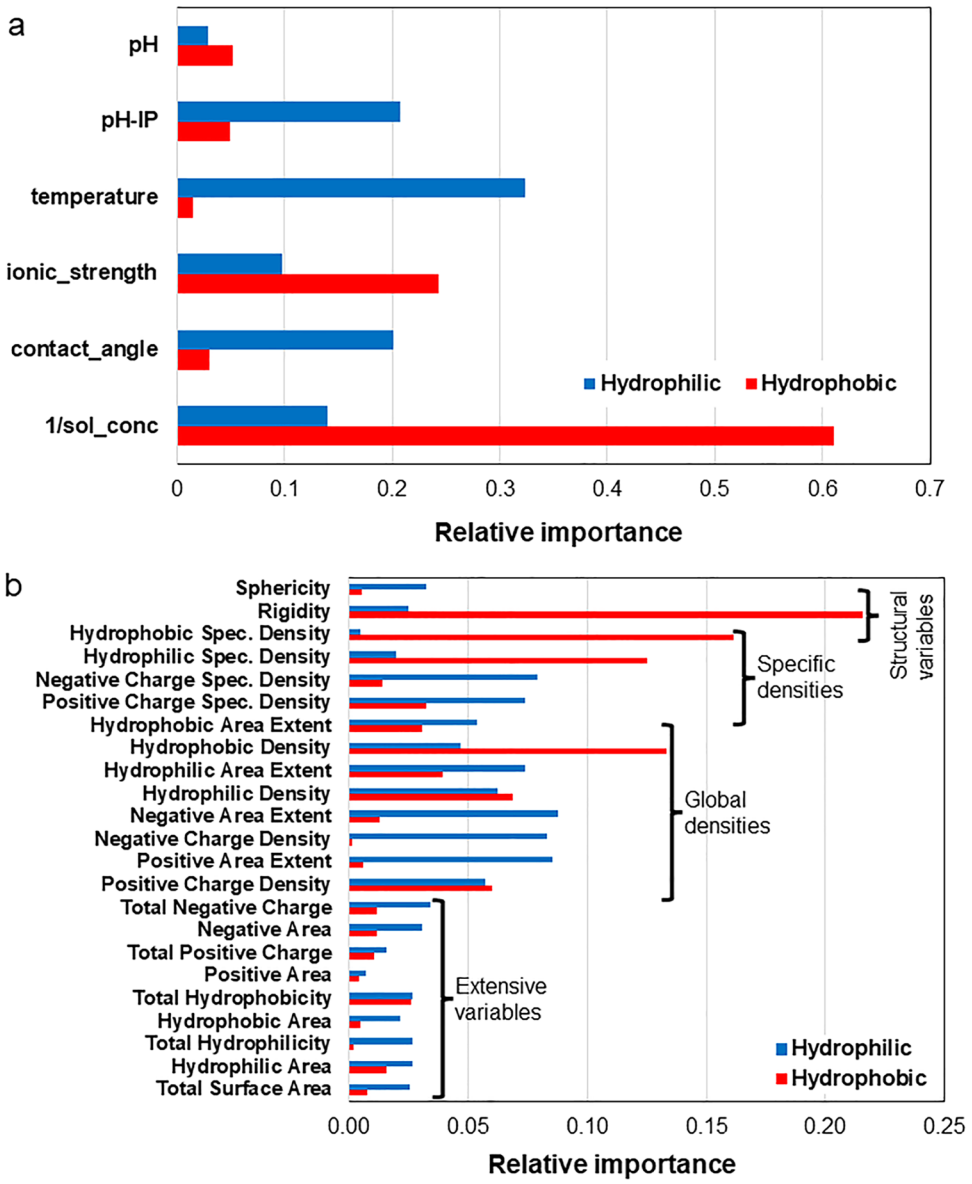


Figure 4. Relative importance of (a) the operational parameters and (b) the physicochemical properties of the proteins quantified on their molecular surface, as derived by machine learning (Ridge Regression algorithm).

between the complex adsorbing surface, e.g., polymers, and complex proteins, e.g., mechano-enzymes.[38]

A preliminary comparison between the concentration of the protein in solution and its subsequent concentration on the adsorbing surfaces (Figure SI 9) showed that (i) overall, protein concentration on the surface follows a logarithmic-like relationship with the protein concentration in solution (excluding some outlier measurements measured with ellipsometry and radiolabeling, caused by few studies comprehensively investigating the full range of protein concentration or contact angle) and that (ii) radiolabeling data has the largest range of mapping this relationship for concentration in solution. A more in-depth comparison between the data measured by ellipsometry and QCM (Figure SI 10) indicated that the ellipsometry data appear to consist of two populations, i.e., higher and lower concentrations on the surface, for similar concentrations in solution, whereas QCM data do not appear to present this bimodal distribution. This bimodality could be the result of the bimodality of the contact angles of the adsorbing surfaces explored by ellipsometry (Figure SI 11), with higher hydrophobic surfaces inducing a higher surface concentration of the protein for similar concentrations of protein in solutions (QCM data do not present this bimodality). Consequently, this preliminary analysis suggests that there are no obvious structural biases between protein adsorption data obtained by ellipsometry, and by QSM, respectively and consequently, the regression analysis used data collected by both methods.

Machine Learning Exploration of Protein Adsorption Data. Inspection of the BAD demonstrates the large variability, the nonuniform distribution, and the large number of parameters modulating protein adsorption on solid surfaces. The rapid development of artificial intelligence methodology suggests that this large variability can be successfully addressed through a fully empirical, unstructured approach describing this overcomplex process. Indeed, machine learning,[39−43] including artificial neuronal networks,[8,41,44,45] has already been used to describe and predict protein adsorption on solid surfaces. To this end, machine learning algorithms were applied to the data

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

contained in BAD, first to determine the relative importance of the operational and of protein variables on their molecular surface, respectively, on the mass of adsorbed protein (Figure 4a,b, respectively).

The *operational* variables modulating protein adsorption (Figure 4a) comprise (i) the protein concentration in solution (represented as the inverse value of monolayers), (ii) the contact angle of the adsorbing surface, (iii) ionic strength, (iv) temperature, (v) pH of the carrier fluid, and (vi) the difference between the solution pH and the isoelectric point (IP) of the protein. Notably, the pH is contained in two separate variables, both with different physical meaning, and therefore having two different coefficients. First, although potentially an important modulator for protein adsorption, especially on hydrophilic surfaces, the zeta potential of the adsorbing surface is very rarely reported in the relevant literature. However, the zeta potential of the surface will change in contact with the carrier fluid, following a near-linear decrease with the increase of the pH of the fluid interfacing the surface.[46] Second, the difference between the pH and the isoelectric point of the protein will report the overall charging of the protein, not only on the molecular surface (which was quantified separately). If this difference is 0 the positive and negative charges of the protein are in equilibrium and the protein is electrostatically neutral.

The inspection of the importance of variables showed that their impact is considerably different for the adsorption on hydrophobic surfaces than that on hydrophilic ones. Indeed, for protein adsorption on hydrophobic surfaces, the importance of the contact angle (rather unexpectedly) appeared to be negligible. In contrast, protein adsorption on hydrophilic surfaces appeared to be strongly modulated by pH, pH-IP, and the contact angle of the adsorbing surface. For both surfaces, the protein concentration in solution and the fluid temperature appear to be strong modulators, albeit more so for adsorption on hydrophilic surfaces, whereas ionic strength does not appear to be important.

The *protein*-derived predictors modulating protein adsorption comprise the total surface area, rigidity, sphericity, hydrophobic area (from which the hydrophilic area can be derived by substitution), total hydrophobicity (from which hydrophobic density and hydrophobic specific density can be calculated), negative area, and total negative charge (presented in Supporting Information Figure SI 12). An inspection of the relative importance of these predictors (presented in Figure 4b) revealed that the *extensive* predictors, that is, variables depending on the molecular weight of the protein, presented substantially lower values than the *intensive* predictors, that is, variables that do not depend on the size of the protein. Focusing on the importance of intensive variables, it can be observed that protein adsorption on *hydrophilic* surfaces is more dependent on *global densities*, such as area extents, i.e., ratios of the area of positive and negative charges, respectively, and hydrophilic and hydrophobic areas, respectively, reported to the total molecular area of the protein, and property densities, i.e., total positive charges, negative charges, total hydrophilicity, and total hydrophobicity, reported to the total area of the protein. Conversely, protein adsorption on *hydrophobic* surfaces is modulated by *specific densities*, i.e., total positive charges, negative charges, total hydrophilicity, and total hydrophobicity, reported to their respective areas on the molecular surface of the protein.

The considerable difference in the relative importance of the operational and protein-derived predictors for protein adsorption on hydrophobic and hydrophilic surfaces suggests that the mechanism of protein adsorption on these two types of surfaces presents considerable differences, suggesting separate analysis. However, while machine learning analysis provided important insights into protein adsorption, it could not provide reasonably predictive correlations (i.e., low correlation coefficients, detailed in Supporting Information Table SI 1). Indeed, the predictive power, and thus the benefit, of machine learning decreases rapidly with the distance away from the training data,[47] due to the lack of *explicit* relationships between input parameters of the protein, of the adsorbing surface, and of the carrier fluid, and the output—the amount of adsorbed protein. This decrease in predicting power is amplified by the uneven distribution of the data points in the training set, as is the case for very different proteins present in BAD, as well as by the nonlinear relationships between some of its input and output parameters. Consequently, the relative importance of operational and protein-based predictors was solely used to inform the derivations of semiempirical models of protein adsorption, separately for hydrophilic and hydrophobic surfaces.

Semiempirical Model of Protein Adsorption. Studies on protein adsorption on solid surfaces have naturally translated in many and various attempts of modeling this process,[5,37,48−50] ranging from "first-principles" approaches, such as molecular simulations,[51−54] and thermodynamic-based models,[55] to more empirical approaches, such as those using neural networks[8] and machine learning.[41] However, and despite these diverse efforts, presently no model is capable of predicting, with adequate accuracy, the amount of adsorbed mass of a *specific* protein, on a *specific* surface, and from a *specific* carrier fluid, even if all information regarding these three classes of parameters is available. The wide variety of applications involving protein adsorption makes the impact of this lack of adequate prediction to be felt differently. For instance, biomaterials research is more concerned with the cascade of biomolecular events induced by the interaction between the protein and the artificial material,[56] rather than the amount of adsorbed protein. However, in many other instances, e.g., chromatography,[57,58] protein analysis,[7] single-use technology for biopharmaceutical industry,[59] and microfluidics devices,[8,60] the amount of the adsorbed protein is one of the key performance criteria. Consequently, the design and optimization of surfaces involved in these applications would considerably benefit from generic correlations *explicitly* linking the descriptors of proteins, surfaces, and fluids with the amount of protein adsorption.

Many mathematical models attempted to describe this relationship as isotherms, e.g., Langmuir,[61−65] Freundlich,[61,62,65] Langmuir−Freundlich,[61,64,65] Temkin,[61,64] Dubinin−Radushkevich,[65] and Random Sequential Adsorption (RSA).[20,66] Langmuir isotherm takes the form of:

$$\mathrm{conc}_{\mathrm{surf}} = \frac{q_{\mathrm{m}\cdot\mathrm{conc}_{\mathrm{sol}}}}{K_{\mathrm{d}} + \mathrm{conc}_{\mathrm{sol}}} \tag{4}$$

where $\mathrm{conc}_{\mathrm{surf}}$ is the protein concentration on the surface; $\mathrm{conc}_{\mathrm{sol}}$ is the protein concentration in solution; q_{m} is the maximum binding capacity of the adsorbent for the protein; and K_{d} is the dissociation constant.

It was argued,[67] correctly, that the Langmuir isotherm does not accurately represent the fundamentals of protein adsorption. Indeed, a brief survey of the original assumptions of the ideal Langmuir model,[67] i.e., (i) homogeneity of the adsorption sites, (ii) each adsorption site binding an individual solute molecule,

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

Table 1. Coefficients of the Multilinear Correlations with Breakpoint for the Linearized Langmuir Isotherms for the Benchmark Proteins[a]

protein	B_0	k_1' (conc sol)$^{-1}$	a_1 contact angle	a_2 ionic strength	a_3 temperature	a_4 (pH-IP)	a_5 pH	R^2 (%)
				Hydrophilic Surfaces				
lysozyme	−0.506237	0.740126	−0.086222	0.132627	1.425999	4.941447	−1.00530	91.15
	−1.20630	133.2088	−0.175375	−0.134474	3.719678	6.890774	−5.75643	
	range:	0.01:1.00	9.00:74.55	0.15:0.20	20.00:25.00	−6.20:−3.60	4.80:7.40	
albumin	−6.47604	0.211580	−0.004829	187.8419	0.106741	5.228697	−4.96953	95.99
	32.46490	0.351126	−4.77638	−18.4424	−16.6753	−115.897	88.81736	
	range:	0.0001:1.0000	6.50:57.50	0.000:0.154	20.00:27.00	−1.40:2.50	3.50:7.40	
IgG	0.104388	0.048554	0.000597	−1.11977	−0.003789	0.100538	0.084305	98.58
	0.102567	0.049500	0.106360	0.100057	−0.016660	0.100284	0.105311	
	range:	0.002:2.000	6.00:75.00	0.01:0.15	22.00:25.00	0.1	7.00	
fibrinogen	0.249413	0.003534	−0.003671	0.130772	0.056064	0.311794	−0.13949	99.11
	−0.715564	0.009579	−0.011530	0.030090	1.003885	2.676016	−3.33370	
	range:	0.0012:1.0000	6.50:74.55	0.010:0.154	20.00:23.00	−2.30:1.60	3.50:7.40	
				Hydrophobic Surfaces				
lysozyme	0.273546	0.035005	−0.002069	−0.137797	−0.022993	−0.226959	0.215835	91.12
	0.025988	0.011826	−0.000490	0.373250	0.101763	0.381571	−0.24773	
	range:	0.01:1.00	74.55:125.00	0.05:0.17	20.00:25.00	−6.20:−3.60	4.80:7.40	
albumin	0.292736	0.059097	0.001455	−0.044077	−0.062216	−0.321946	0.335993	93.58
	0.328881	0.308488	−0.203793	−1.98833	0.364164	0.663358	1.682032	
	range:	0.005:1.000	70.00:116.00	0.019:1.100	20.00:27.00	−0.50:2.50	4.37:7.40	
IgG	−0.015438	0.026139	−0.005622	0.006890	0.054147	1.164398	−0.02935	96.28
	0.271965	0.007303	−0.109978	60.42560	0.254227	−0.421891	0.513625	
	range:	0.0010:2.0000	75.00:96.00	0.0050:0.1750	20.00:25.00	−2.90:1.10	4.00:8.00	
fibrinogen	0.273546	0.035005	−0.002069	−0.137797	−0.022993	−0.226959	0.215835	91.12
	0.025988	0.011826	−0.000490	0.373250	0.101763	0.381571	−0.24773	
	range:	0.0050:1.0000	88.00:125.00	0.0370:1.1000	20.00:22.00	−3.05:1.6	2.75:7.40	

[a]**Part 1:** Correlations for concentration in solution (mg/mL), surface contact angle (°), ionic strength (mM), temperature (°C), difference between pH and isoelectric point, and pH.

(iii) reversibility of adsorption, and (iv) adsorption behavior not being impacted by the interactions between adsorbing molecules on the surface, indicates many deviations from the process of biomolecular adsorption on solid surfaces. Despite this fundamental criticism, many reports[61,62,64,65] found that the Langmuir isotherm model fits well with experimental data for protein adsorption, as well as for DNA hybridization on microarrays.[68]

Protein adsorption, and biomolecular adsorption in general, can be viewed as a two-stage process: (i) the adsorption of protein molecules on bare, hydrophobic, or hydrophilic solid surface; and when this stage is complete, (ii) the adsorption of protein molecules on a protein-covered surface, which is, in many instances, more hydrophilic than the initial basal solid surface. The translation from the first phase to the second phase is expected to occur gradually, which is in line with the gradual coverage of the adsorbing surface. This gradual translation aside, the different nature of the surface on which proteins adsorb is expected to translate into different adsorption mechanisms. The most obvious of these dissimilarities is the different slope of the relationship between the adsorbed amount of protein and the protein concentration in solution. However, the differences would also extend to the modulation of protein adsorption by other parameters of the carrier fluid, protein, and even basal surfaces (the latter, by inducing conformational changes in the proteins adsorbed in the first layer).

This two-stage model framework also resolves, or mitigates, many differences between the ideality of the Langmuir isotherm model and the reality of the protein adsorption process, especially at low protein concentrations in solution. First, the Langmuir isotherm-like processes assume the presence of homogeneous surfaces, whereas the progress of protein adsorption leads to increasing heterogeneous ones.[67] However, the proposed two-stage model assumes two extreme, Langmuir isotherm-like stages: protein adsorption on the initial homogeneous solid surface and later the adsorption on another homogeneous surface when a protein monolayer was formed (with a transition between these two stages). Second, while the Langmuir isotherm processes assume that each adsorption site binds an individual molecule,[67] the two-stage protein adsorption inherently assumes the creation of multilayers. Third, protein adsorption, especially on hydrophobic surfaces, is more biased toward irreversibility than a process described by a Langmuir isotherm,[67] but the reversibility of protein adsorption increases once the first monolayer is formed (and it is more relevant for the initial adsorption on hydrophilic surfaces). Finally, a process described strictly by a Langmuir isotherm assumes that no interaction occurs between solutes on the surfaces,[67] but proteins do interact, thus influencing the process of adsorption. However, the data used here for the derivation of predicting relationships uses low concentrations on the surface, e.g., on hydrophilic surfaces or surfaces with protein monolayers, or at low concentrations in solutions, which cover most reported cases in the literature, where this protein−protein interactions is less important.

An additional advantage of the Langmuir isotherm is its simplicity and easy linearization, which facilitate multivariable regression analysis. For the purposes of multilinear regression analysis with breakpoint, eq 4 can be rewritten as

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

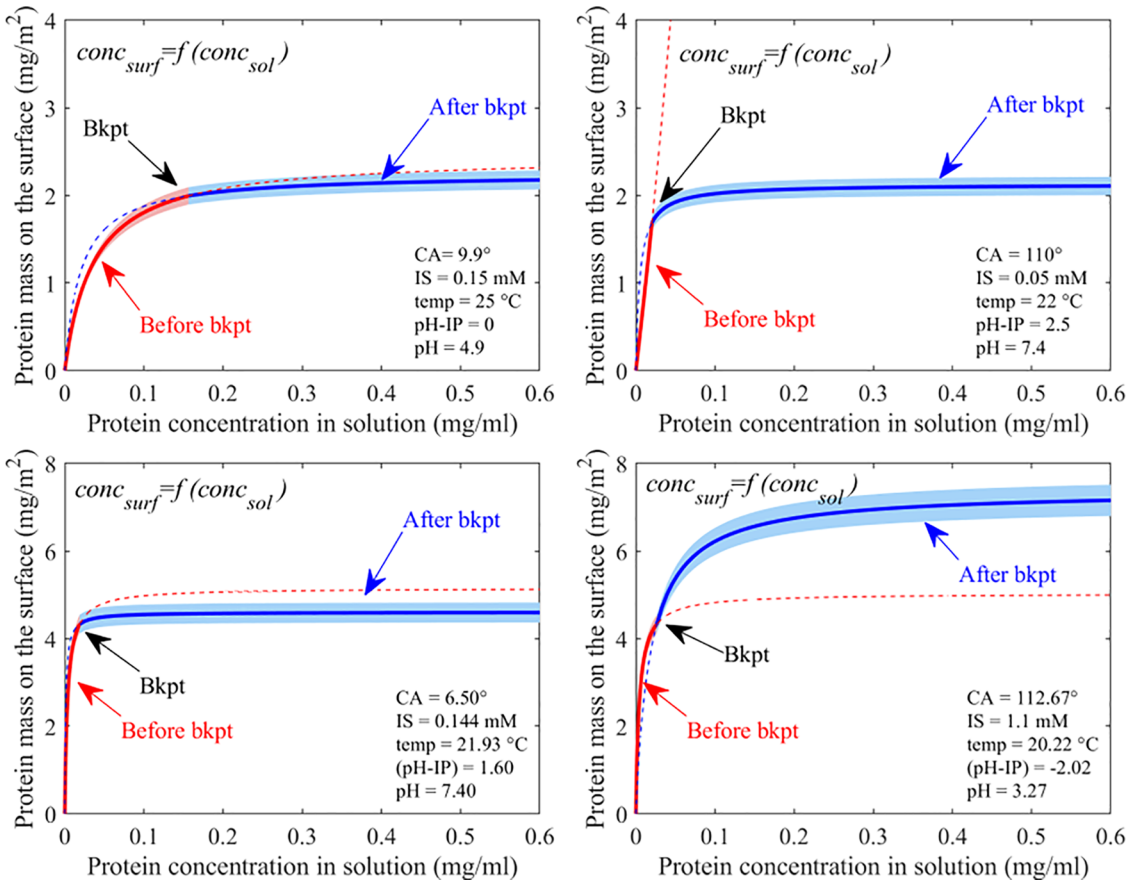


Figure 5. Examples of the output of the prediction of protein adsorption for albumin (1BJ5, top row) and fibrinogen (1M1J, bottom row), on hydrophilic (left column) and hydrophobic surfaces (right column) for various sets of technological process parameters.

$$\mathrm{conc}_{\mathrm{surf}} = \frac{q_{\mathrm{m}\cdot\mathrm{conc}_{\mathrm{sol}}}}{K_{\mathrm{d}} + \mathrm{conc}_{\mathrm{sol}}} = \frac{q_{\mathrm{m}}/K_{\mathrm{d}}\cdot\mathrm{conc}_{\mathrm{sol}}}{1 + K_{\mathrm{d}}^{-1}\cdot\mathrm{conc}_{\mathrm{sol}}} = \frac{K_{1\cdot\mathrm{conc}_{\mathrm{sol}}}}{1 + K_2\cdot\mathrm{conc}_{\mathrm{sol}}} \quad (5)$$

Equation 5 can be presented in a linearized form as

$$\mathrm{conc}'_{\mathrm{surf}} = k_1\cdot\mathrm{conc}'_{\mathrm{sol}} + k_2 \quad (6)$$

where $\mathrm{conc}'_{\mathrm{surf}} = (\mathrm{conc}_{\mathrm{surf}})^{-1}$; $\mathrm{conc}'_{\mathrm{sol}} = (\mathrm{conc}_{\mathrm{sol}})^{-1}$; $k_1 = (K_1)^{-1} = \frac{q_{\mathrm{m}}}{K_{\mathrm{d}}}$; and $k_2 = \frac{1}{q_{\mathrm{m}}}$.

Regression Analysis of the Adsorption of Benchmark Proteins. Assuming a two-stage process of protein adsorption, and accounting for more operational parameters that modulate it, eq 6 can be rewritten as a multilinear relationship with breakpoint (eqs 7 and 8):

$$\mathrm{conc}'_{\mathrm{surf}} = k_{11}\cdot\mathrm{conc}'_{\mathrm{sol}} + a_{11}\cdot\mathrm{CA} + a_{21}\cdot\mathrm{IS} + a_{31}\cdot\mathrm{temp} + a_{41}\cdot\mathrm{pH} + a_{51}\cdot(\mathrm{pH\text{-}IP}) + B_{01} \text{ for } \mathrm{conc}'_{\mathrm{surf}} < \text{breakpoint} \quad (7)$$

and

$$\mathrm{conc}'_{\mathrm{surf}} = k_{12}\cdot\mathrm{conc}'_{\mathrm{sol}} + a_{12}\cdot\mathrm{CA} + a_{22}\cdot\mathrm{IS} + a_{32}\cdot\mathrm{temp} + a_{42}\cdot\mathrm{pH} + a_{52}\cdot(\mathrm{pH\text{-}IP}) + B_{02} \text{ for } \mathrm{conc}'_{\mathrm{surf}} > \text{breakpoint} \quad (8)$$

where CA is the contact angle of the basal surface; IS is the ionic strength of the carrier fluid; temp is its temperature; and (pH-IP) is the difference between the fluid pH and the isoelectric point of the protein.

To increase the generality of this mathematical formalism to various proteins with various sizes and shapes, the concentration on the surface can be expressed nondimensionally as number of monolayers using the ProMS-estimated sphericity. Also, the protein concentration in solution can be expressed as number of moles/unit volume. Given the proposed model above and because the concentration on the surface is expressed in monolayers, the breakpoint is set to one monolayer, regardless of the protein adsorbed.

The multilinear regression analysis with breakpoint using linearized Langmuir isotherm formalism led to good-to-excellent regression coefficients for the four benchmark proteins for which enough data points could be collected from BAD, i.e., lysozyme, albumin, IgG, and fibrinogen, but only after the data were split in reports of protein adsorption on hydrophilic and hydrophobic surfaces, respectively. The results are presented in Table 1 (and in Supporting Information Table SI. 2. and Table SI 3.).

The implementation of these correlations in a MATLAB-based estimation program (described in Supporting Information SI 4) allows for the straightforward estimation of the adsorbed amount of protein as a function of the input variables, i.e., surface contact angle, ionic strength, pH, and difference between the pH and protein isoelectric point (examples are presented in Figure 5).

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

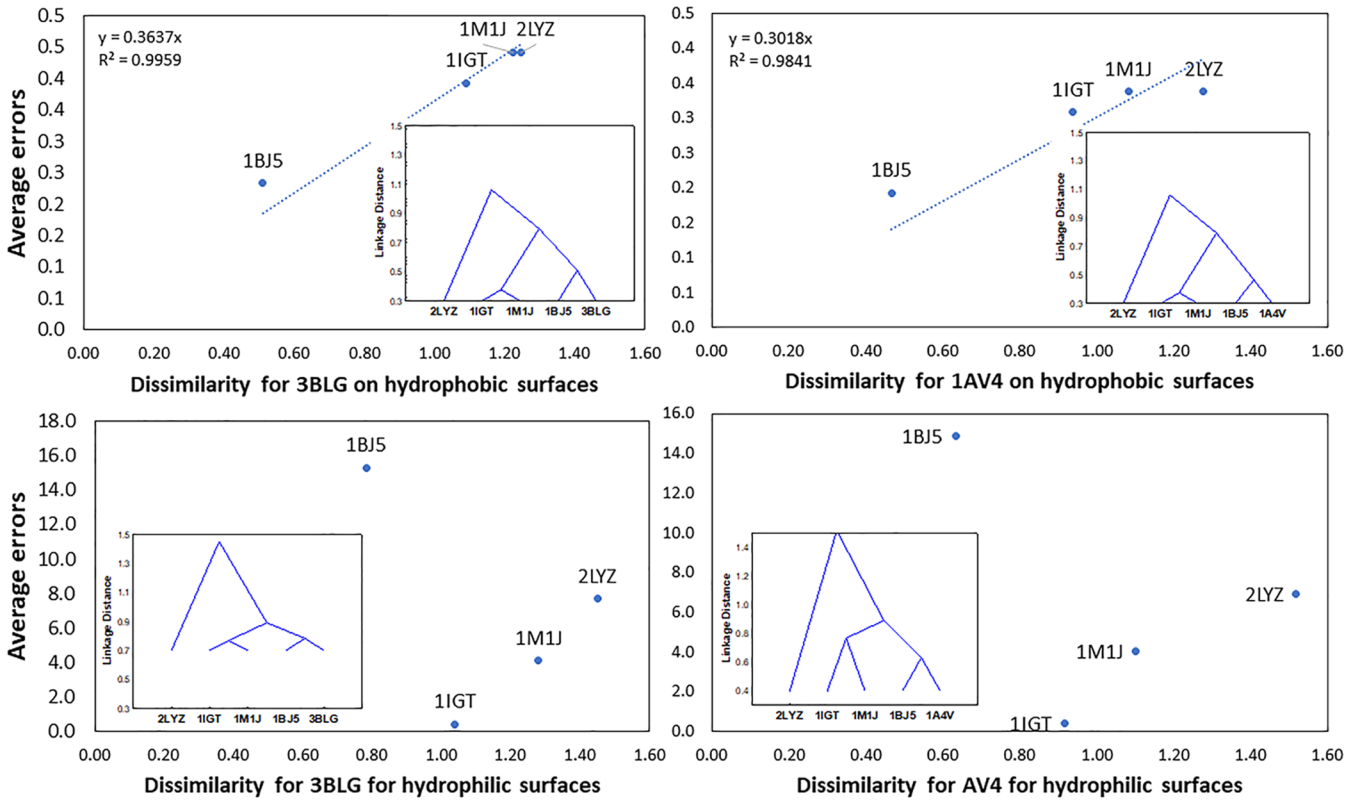


Figure 6. Relative errors of estimation of adsorbed mass for two proteins, i.e., β-lactoglobulin (3BLG, left column) and α-lactalbumin (1A4V, right column) as functions of the dissimilarity between them and each of the four benchmark proteins, i.e., lysozyme (2LYZ), albumin (1BJ5), fibrinogen (1M1J), and antibody (1IGT). The dissimilarity between pairs of proteins was calculated as the distance resulting from cluster analysis using similarity parameters quantified on the protein molecular surface for protein adsorption on hydrophobic surfaces (top row) and hydrophilic surfaces (bottom row). The bottom insets represent the clustering tree for each of the five protein groups, i.e., one test protein vs four benchmark proteins. The top insets represent the empirical linear correlation between the relative errors and dissimilarity.

Comparison of the Semiempirical Relationships for Protein Adsorption with Data for Test Proteins. Two proteins, i.e., β-lactoglobulin (3BLG) and α-lactalbumin (1A4V), were used to explore the limits of using the correlations obtained for the benchmark proteins for other proteins for which the adsorption data were less comprehensive. These relationships were applied for the same conditions reported for the test proteins, to estimate the adsorbed protein amount on hydrophilic and hydrophobic surfaces, respectively.

As expected, most predictions using the correlations obtained for benchmark proteins resulted in large errors when they were used for test proteins. However, for each test protein, a reasonably accurate prediction did exist when using at least one benchmark protein (results presented in Supporting Information SI 2.).

To validate the hypothesis that similar proteins have similar adsorption behavior, the similarity rules by which proteins can be classified must be first established. Given the observed dissimilarity of the impact of physicochemical properties of the proteins on adsorption on hydrophilic and hydrophobic surfaces, respectively, machine learning analysis suggested the similarity parameters for protein adsorption on hydrophilic and hydrophobic surfaces, respectively. Machine learning demonstrated that protein adsorption is modulated by *intensive* parameters, that is those that are not correlated with the molecular weight of the protein. Furthermore, this analysis (results presented in Supporting Information SI 2) revealed that for adsorption on hydrophilic surfaces, the similarity between various proteins is determined by *global densities* (total positive and negative charges, and hydrophilicity), and by *areas extents* (positive, negative, hydrophilic, and hydrophobic areas). Finally, for hydrophobic surfaces, the similarity between various proteins is determined by a combination of *specific densities* (positive charge, hydrophilic, hydrophilic) and sphericity.

Cluster analysis of the proteins using their structural and physicochemical properties manifested on their molecular surfaces, using parameters relevant to protein adsorption on hydrophobic and hydrophilic surfaces, respectively, led to the quantification of the distances between the cases as a measure of the dissimilarities between proteins.

Separately, the semiempirical correlations estimating the adsorbed protein (expressed in monolayers) derived for the four benchmark proteins were applied to the two test proteins for which there are enough data in BAD (results presented in Supporting Information SI 2). These extrapolations allowed the quantification of errors when using a particular correlation derived for one of the four benchmark proteins applied to the operational conditions reported for one of the two test proteins.

For the adsorption on hydrophobic surfaces, the average relative errors range, depending on the benchmark protein-based estimation, from 19 to 34% for α-lactalbumin (1A4V), and from 23 to 44% for β-lactoglobulin (3BLG). For both test proteins, albumin (1BJ5) provides the closest prediction of adsorbed protein. Importantly, these errors were in a near-perfect linear relationship with the cluster analysis-calculated dissimilarity (Figure 6, top row).

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

The relative errors related to the adsorption on hydrophilic surfaces (Figure 6, bottom row) are considerably higher, ranging from 41 to 1527% for β-lactoglobulin (3BLG) and from 43 to 1491% for α-lactalbumin (1A4V). For both test proteins, immunoglobulin (1IGT) provides the closest prediction of adsorbed protein. Importantly, the highest errors result from the application of albumin-derived semiempirical relationships. In fact, when comparing the relative errors with the dissimilarity between the test and the benchmark proteins, albumin is a clear outliner, with the other three following a quasi-linear relationship between errors and the dissimilarity between proteins, as derived from their properties on their molecular surfaces.

The differences in the errors of the predictions of protein adsorption on hydrophobic and hydrophilic surfaces, respectively, using relationships derived for the benchmark proteins applied to the test proteins can be better understood considering the following:

i. The adsorption on hydrophobic surfaces is governed more by *short-range* hydrophobic−hydrophobic interactions. Consequently, the process is, to a larger extent, modulated by the localized distributions of properties on the molecular surfaces, in particular, specific densities of hydrophobicity and hydrophilicity. This short-range character of the parameters modulating protein adsorption on hydrophobic surfaces, leads to relatively higher success in extrapolating the predictions of the benchmark proteins, based on molecular surface-derived parameters, applied to the test proteins.

ii. Conversely, adsorption on hydrophilic surfaces is governed more by *long-range* electrostatic interactions. Consequently, the process is to a larger extent modulated by the global distributions of properties on the molecular surfaces, in particular densities of properties and the relative area extents for these properties (charges, both positive and negative, and hydrophilicity and hydrophobicity). This long-range character of the parameters modulating protein adsorption on hydrophilic surfaces results in lower importance of molecular surfaces and, consequently, considerably higher errors when extrapolating the relationships obtained for the benchmark proteins. Furthermore, the closest prediction occurs for immunoglobulin (1IGT), which presents the highest rigidity of all studied proteins, and thus the most stable molecular surface, and the exceedingly high relative errors occur for albumin (1BJ5), which presents the lowest rigidity of all studied proteins.

The differences between BAD 1.0 and BAD 2.0, as well as the differences between the previously reported[8] treatment of data and the present work, are synthetically presented in Supporting Information SI 5.

Perspectives and Further Work. The correlations connecting the input variables of protein adsorption, i.e., protein concentration in solution, contact angle of the adsorbing surface, ionic strength and temperature of the carrying fluid, and the difference between the pH of the fluid and the isoelectric point of the protein, with the output variable, i.e., protein mass adsorbed on the surface, suggest several general conclusions, perspectives, and perhaps further essential work:

- In general, there are no apparent, major biases between protein adsorption measurements performed by the three main analytical methods, i.e., radiolabeling, ellipsometry, and QCM. Consequently, the measurements using these three methods can be reasonably compared to build a more general framework of the process of protein adsorption on solid surfaces.
- Protein adsorption can be thought of as progressing in two stages. Notwithstanding the gradual transition between them, the first stage consists of the adsorption of protein molecules on the bare solid surface. This stage ends when the absorbed protein molecules effectively block further access to the basal surface. The second phase consists of the adsorption of proteins on top of a protein-covered surface.
- While it was correctly argued[67] that the Langmuir isotherm is not appropriate for describing protein adsorption, the splitting of protein adsorption into two stages decreases the dissimilarity between the processes described by the Langmuir isotherm and the actual protein adsorption.
- When enough experimental data exist covering all input variables of protein adsorption, correlations can be found based on the linear relationship between the inverse concentration of the protein on the surface on one side, and the inverse of the concentration in solution, surface contact angle, ionic strength and temperature of the solution, and the difference between solution pH and isoelectric point of the protein, on the other. However, no correlation could be found covering the full range of surface tension, from hydrophilic to hydrophobic, suggesting that the mechanisms involved are markedly different, and therefore the protein adsorption on these surfaces must be treated separately.
- These relationships, obtained by multilinear regression with breakpoint of linearized Langmuir isotherms (describing the correlation between protein concentration on the surface, and in solution, with the constants in the Langmuir isotherm assumed to be linear functions of the rest of the input variable), present excellent fits with the experimental data for the benchmark proteins.
- The correlations derived from expressing the protein concentration in monolayers, and the protein concentration in solution in moles/volume, can be used to estimate, with reasonable accuracy, the equivalent protein adsorption of a test protein, if the benchmark protein used for extrapolation presents a reasonable similarity of the properties on its molecular surface. More specifically, this estimation of protein adsorption on hydrophilic and hydrophobic solid surfaces, respectively, will hold only if the hydrophilic- and hydrophobic-related descriptors, i.e., hydrophilic/hydrophobic density, specific density, and the ratio of the respective area on molecular surface, of the benchmark and test protein, have a reasonable high degree of similarity.
- The semiempirical correlations presented here, for proteins representing more than 80% of the reported data in the literature, offer good estimations for flat solid surfaces, especially for low concentrations of proteins on the surface and in solutions. These correlations could also be used tentatively for initial estimations for other proteins for which data is sparse or not available, provided that the protein structure is known and if the properties on the respective molecular surface are similar to those for other proteins for which the original semiempirical correlations were derived. However, for more complex surfaces, e.g., polymer brushes, complex proteins, e.g.,

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

mechano-enzymes with inherently changing conformations, or for higher protein concentrations, e.g., leading to protein cooperativity, clustering, and aggregation, these semiempirical correlations have limited, if any value. In these instances, direct experimental measurements are necessary, for which high-throughput screening approaches, e.g., microarray technology coupled with microfluidics, are available. Moreover, semiempirical correlations have very little value for the adsorption of proteins with solid objects with similar sizes, e.g., nanoparticles. However, in these instances, which show an increased interest presently, the progress in molecular modeling and computer power suggest that predictive simulations can offer reliable predictions from first principles.[16]

The derivation of semiempirical relationships predicting protein adsorption for the benchmark proteins, covering approximately 80% of the published literature so far, as well as the demonstration of the methodology to use for proteins for which no or sparse data exist but for which molecular structures have been published, can be used in various ways. Without being exhaustive, three examples are as follows:

Single-Use Technology. One difficulty in biomanufacturing for the pharmaceutical industry is the need for very precise control of the operational parameters, such as temperature and concentration of reactants in the reactor. However, while large reactors are economically justified, the precisely even spatial distribution of the operational parameters in each microvolume is impossible to achieve in large reactors due to the gradient of shear stress of large propellers and consequently the uneven spatial distribution of temperature and concentration of reactants. The solution to this conundrum is single-use technology, which proposes the use of a large number of orders of magnitude smaller reactors, which allows a much finer control of operational parameters within the reaction volume.[69] To be economically feasible, this technology requires that the mini-reactors be made of cheap, disposable materials, such as polymers. This technologically elegant solution comes, however, at the price of using materials for which the interaction between the manufactured materials and the extremely valuable chemical species in the reaction volume, especially proteins, is unknown. Moreover, these proteins could adsorb on the surfaces of mini-reactors, which have inherently a larger surface/volume ratio, instead of being collected for final processing. To this end, the semiempirical relationships predicting protein adsorption, or the data presented in BAD 2.0, can help to narrow down the choices of materials amenable for the fabrication of mini-reactors for the fine pharmaceutical industry.

Design of Lab-on-a-Chip Devices. A similar and perhaps a more critical problem exists for microfluidics devices for diagnostics, especially those for personalized use. In these devices, biofluids containing dissolved proteins are moved along microfluidic channels toward a reaction chamber where the analytes of interest are detected, and their concentrations are measured. In many instances, these devices are disposable, almost always made of polymeric materials. The single use of these devices means that the fluids pass the channels only once, and therefore there is the distinct possibility that the microfluidic device operates as a minichromatographic column, depleting the concentration of the analytes before they reach the detection chamber. Additionally, because of their small size, these devices exhibit very high surface/volume ratios. Furthermore, the parasitic adsorption of proteins on the walls of microfluidic channels will alter both their geometrical design as well as the hydrodynamics in microconfined spaces, due to changes of the surface tension of the channel surfaces.[9] Again, these effects can be estimated using the semiempirical relationships predicting protein adsorption or the data presented in BAD 2.0.

Calibration of Initial Readings of Equipment Measuring Protein Adsorption. Analysis of the data in BAD 2.0 shows that, in general, there are no sizable differences between the measurements of protein adsorption with various experimental methodologies, of which, presently the most used are ellipsometry and QCM. There are however a sizable number of instances in BAD where the protein adsorption data differ substantially for identical or close operational parameters and the same proteins, using either the same or a different measurement methodology. Furthermore, although both ellipsometry and QCM are now well-established methodologies, both are susceptible to artifacts. Consequently, the use of semiempirical relationships can be used to establish credible ranges of protein adsorption data or even calibrate first measurements with new equipment.

CONCLUSIONS

The sheer complexity of protein adsorption has frustrated, so far, any reasonably precise prediction of adsorbed protein from first principles of the amount of protein adsorbed. The updating and enlargement of the Biomolecular Adsorption Database (BAD 2.0) have allowed the derivation of predictive correlations for the adsorption on solid surfaces of four benchmark proteins, i.e., lysozyme, albumin, IgG, and fibrinogen, which represent the vast majority of the reports in the open literature. Machine learning-based analysis shows that protein adsorption on hydrophobic and hydrophilic surfaces is modulated by different sets of operational, structural, and physicochemical parameters quantified on the protein molecular surface. Separate correlations based on Langmuir isotherm models were derived for hydrophilic and hydrophobic surfaces, with protein monolayer being the breakpoint representing the transition between protein adsorption on solid surfaces and subsequent protein adsorption on a protein-covered surface. These correlations can be used to predict the amount of adsorbed protein as a function of the operating parameters, i.e., the contact angle of the adsorbing surface, pH, ionic strength, and temperature of the carrier fluid, and the difference between the pH of the buffer and the isoelectric point of the protein. These adsorbing surface-specific modulators were then rationally used to estimate the similarity between proteins. It was found that, when applying the semiempirical relationships derived for benchmark proteins on hydrophobic surfaces to other proteins with known structure, the errors of this extrapolation are in a linear relationship with the dissimilarity of the benchmark and test proteins. The present work can be used for the first tentative estimation of the operational parameters, e.g., carrier fluid properties, and surface properties, for protein adsorption involved in various industrial applications, such as diagnostic devices, biomanufacturing, medical devices, pharmaceuticals, or food industry. Furthermore, should enough data for a particular set of protein, surface, and carrier fluid be available, the semiempirical conceptual methodology of concatenated Langmuir isotherm-like processes could be used to derive new predictive correlations for protein adsorption.

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.4c06759.

Description of the parameters in the database, the actual BAD 2.0 in table format; calculation of properties of proteins, including on their molecular surface; calculation of monolayer density; data used for regression on hydrophilic and hydrophobic surfaces; calculations for using relationships derived for benchmark proteins for test proteins; calculation of distances between protein properties cf. cluster analysis (Euclidian distances); and comparison of the Euclidian distances with the errors of extrapolation (XLSX)

Bibliography of reports used to build BAD 2.0 (SI 1); Biomolecular Adsorption Database (SI 2); description of the molecular surface calculator (SI 3); description of the protein adsorption predictor (SI 4); BAD 2.0 vs. BAD 1.0 (SI 5); ProMS graphical user interface (Figure SI 1); protein ubiquitin (1UBQ) molecular surface (Figure SI 2); input file for batch mode (Figure SI 3); protein adsorption predictor graphical user interface (Figure SI 4); screenshot of Biomolecular Adsorption Database portal (Figure SI 5); entity relationship diagram for BAD 2.0 (Figure SI 6); molecular surfaces of the proteins present in BAD 2.0 (Figure SI 7); relative distribution of reports of protein adsorption measurement methods vs. time (Figure SI 8); overall distribution of surface, and solution concentration of proteins (Figures SI 9 and SI 10); distribution protein adsorption vs. contact angle for ellipsometry and QCM (Figure SI 11); relative importance determined by ML of the protein variables (Figure SI 12); correlation coefficients (Table S1); correlations for concentration in solution, surface contact angle, ionic strength, difference between pH and isoelectric point, and pH (Table S2); and correlations for concentration in solution, surface contact angle, ionic strength, and pH (Table S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Dan V. Nicolau − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada; Swinburne University of Technology, Hawthorn, Vic 3122, Australia;* orcid.org/0000-0002-9956-0600; Email: dan.nicolau@mcgill.ca

Authors

Matthew Cho − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada*

Zahra Mahmoodi − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada*

Prasad Shetty − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada*

Lauren R. Harrison − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada*

Maru Arias Montecillo − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada*

Ayyappasamy Sudalaiyadum Perumal − *Faculty of Engineering, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada*

Gerardin Solana − *Swinburne University of Technology, Hawthorn, Vic 3122, Australia*

Dan V. Nicolau, Jr. − *Swinburne University of Technology, Hawthorn, Vic 3122, Australia; Faculty of Life Sciences & Medicine, School of Immunology & Microbial Sciences, Peter Gorer Department of Immunobiology, King's College London, London SE1 1UL, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.4c06759

Author Contributions

‖M.C., Z.M., and P.S. contributed equally. P.S., M.A.M., L.R.H., G.S., and D.V.N. collected data for the database. M.C., L.R.H., and D.V.N. developed the quantification and visualization package for protein molecular surface. M.C., Z.M., and D.V.N. calculated the molecular surface parameters. M.C. and D.V.N. created the architecture of the database. G.S., Z.M., D.V.N. Jr., and D.V.N. run regression correlations. Z.M. and D.V.N. prepared the prediction tool for benchmark proteins. D.V.N. Jr., A.S.P., and D.V.N. performed cluster analysis. D.V.N. Jr. and D.V.N. wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research was sponsored by grants from Natural Sciences and Engineering Research Council of Canada (NSERC), grant nos. RGPIN-2016-05019 and RGPIN-2022-04053; and Defense Advanced Research Projects Agency (DARPA), grant no. AFRL-IF-RS-TR-2003-304.

■ REFERENCES

(1) Zimmerberg, J.; Kozlov, M. M. How Proteins Produce Cellular Membrane Curvature. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (1), 9−19.

(2) Speziale, P.; Pietrocola, G.; Foster, T. J.; Geoghegan, J. A. Protein-based Biofilm Matrices in Staphylococci. *Front. Cell. Infect. Microbiol.* **2014**, *4*, No. 171.

(3) Silva-Bermudez, P.; Rodil, S. E. An Overview of Protein Adsorption on Metal Oxide Coatings for Biomedical Implants. *Surf. Coat. Technol.* **2013**, *233*, 147−158.

(4) Sutherland, D. W., Jr.; Blanks, Z. D.; Zhang, X.; Charest, J. L. Relationship Between Central Venous Catheter Protein Adsorption and Water Infused Surface Protection Mechanisms. *Artif. Organs* **2018**, *42* (11), E369−E379.

(5) Pinholt, C.; Hartvig, R. A.; Medlicott, N. J.; Jorgensen, L. The Importance of Interfaces in Protein Drug Delivery - Why is Protein Adsorption of Interest in Pharmaceutical Formulations? *Expert Opin. Drug Delivery* **2011**, *8* (7), 949−964.

(6) Rebeski, D. E.; Winger, E. M.; Shin, Y. K.; Lelenta, M.; Robinson, M. M.; Varecka, R.; Crowther, J. R. Identification of Unacceptable Background Caused by Non-specific Protein Adsorption to the Plastic Surface of 96-well Immunoassay Plates Using a Standardized Enzyme-Linked Immunosorbent Assay Procedure. *J. Immunol. Methods* **1999**, *226* (1−2), 85−92.

(7) Wegner, G. J.; Wark, A. W.; Lee, H. J.; Codner, E.; Saeki, T.; Fang, S.; Corn, R. M. Real-time Surface Plasmon Resonance Imaging Measurements for Multiplexed Determination of Protein Adsorption/

Desorption Kinetics and Surface Enzymatic Reactions on Peptide Microarrays. *Anal. Chem.* **2004**, *76* (19), 5677−5684.
(8) Vasina, E. N.; Paszek, E.; Nicolau, D. V., Jr; Nicolau, D. V. The BAD project: Data Mining, Database and Prediction of Protein Adsorption on Surfaces. *Lab Chip* **2009**, *9* (7), 891−900.
(9) Nicolau, D. V., Jr; Fulga, F.; Nicolau, D. V. Impact of Protein Adsorption on the Geometry of Microfluidics Devices. *Biomed. Microdevices* **2003**, *5* (3), 227−233.
(10) Weiner, H.; Glökler, J.; Hultschig, C.; Büssow, K.; Walter, G. Protein, Antibody and Small Molecule Microarrays. In *Microarray Technology and Its Applications*; Müller, U. R.; Nicolau, D. V., Eds. Biological and Medical Physics, Biomedical Engineering; Springer: Berlin, Heidelberg, 2005; pp 279−295.
(11) Chen, D.; Ganesh, S.; Wang, W.; Amiji, M. Plasma Protein Adsorption and Biological Identity of Systemically Administered Nanoparticles. *Nanomedicine* **2017**, *12* (17), 2113−2135.
(12) Corbo, C.; Molinaro, R.; Parodi, A.; Toledano Furman, N. E.; Salvatore, F.; Tasciotti, E. The Impact of Nanoparticle Protein Corona on Cytotoxicity, Immunotoxicity and Target Drug Delivery. *Nanomedicine* **2016**, *11* (1), 81−100.
(13) Durán, N.; Silveira, C. P.; Durán, M.; Martinez, D. S. T. Silver Nanoparticle Protein Corona and Toxicity: A Mini-review. *J. Nanobiotechnol.* **2015**, *13* (1), No. 55.
(14) Kim, J. Systematic Approach to Characterize the Dynamics of Protein Adsorption on the Surface of Biomaterials using Proteomics. *Colloids Surf., B* **2020**, *188*, No. 110756.
(15) Nga, N. K.; Thanh Tam, L. T.; Ha, N. T.; Hung Viet, P.; Huy, T. Q. Enhanced Biomineralization and Protein Adsorption Capacity of 3D Chitosan/Hydroxyapatite Biomimetic Scaffolds Applied for Bone-Tissue Engineering. *RSC Adv.* **2020**, *10* (70), 43045−43057.
(16) Latour, R. A. Fundamental Principles of the Thermodynamics and Kinetics of Protein Adsorption to Material Surfaces. *Colloids Surf., B* **2020**, *191*, No. 110992.
(17) Lathrop, R. H. The Protein Threading Problem with Sequence Amino Acid Interaction Preferences is NP-complete. *Protein Eng., Des. Sel.* **1994**, *7* (9), 1059−1068.
(18) Nienhaus, K.; Xue, Y.; Shang, L.; Nienhaus, G. U. Protein Adsorption onto Nanomaterials Engineered for Theranostic Applications. *Nanotechnology* **2022**, *33* (26), No. 262001.
(19) Adamczyk, Z. Protein adsorption: A Quest for a Universal Mechanism. *Curr. Opin. Colloid Interface Sci.* **2019**, *41*, 50−65.
(20) Armstrong, M. J.; Rodriguez, J. B.; Dahl, P.; Salamon, P.; Hess, H.; Katira, P. Power Law Behavior in Protein Desorption Kinetics Originating from Sequential Binding and Unbinding. *Langmuir* **2020**, *36* (45), 13527−13534.
(21) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.
(22) Pearson, W. R.; Lipman, D. J. Improved Tools for Biological Sequence Comparison. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85* (8), 2444−2448.
(23) Nicolau, D. V.; Paszek, E.; Fulga, F.; Nicolau, D. V. Mapping Hydrophobicity on the Protein Molecular Surface at Atom-level Resolution. *PLoS One* **2014**, *9* (12), No. e114042.
(24) Nicolau, D. V.; Paszek, E.; Fulga, F.; Nicolau, D. V., Jr Protein Molecular Surface Mapped at Different Geometrical Resolutions. *PLoS One* **2013**, *8* (3), No. e58896.
(25) Connolly, M. L. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221* (4612), 709−713.
(26) Connolly, M. L. The Molecular Surface Package. *J. Mol. Graphics* **1993**, *11* (2), 139−141.
(27) White, S. H.; Wimley, W. C. Hydrophobic Interactions of Peptides with Membrane Interfaces. *Biochim. Biophys. Acta, Rev. Biomembr.* **1998**, *1376* (3), 339−352.
(28) *The PyMOL Molecular Graphics System*; Schrödinger, LLC., 2021. https://pymol.org/ (accessed April 29, 2024).
(29) Zou, R. P.; Yu, A. B. Evaluation of the Packing Characteristics of Mono-sized Non-spherical Particles. *Powder Technol.* **1996**, *88* (1), 71−79.
(30) Decker, E. L.; Frank, B.; Suo, Y.; Garoff, S. Physics of Contact Angle Measurement. *Colloids Surf., A* **1999**, *156* (1−3), 177−189.
(31) Beynon, R.; Easterby, J. *Buffer Solutions*; Taylor & Francis, 2003.
(32) Chada, N.; Sigdel, K. P.; Gari, R. R. S.; Matin, T. R.; Randall, L. L.; King, G. M. Glass is a Viable Substrate for Precision Force Microscopy of Membrane Proteins. *Sci. Rep.* **2015**, *5* (1), No. 12550.
(33) Liu, Z.; Li, Z.; Zhou, H.; Wei, G.; Song, Y.; Wang, L. Observation of the Mica Surface by Atomic Force Microscopy. *Micron* **2005**, *36* (6), 525−531.
(34) Horbett, T. A.; Brash, J. L. Proteins at Interfaces: Current Issues and Future Prospects. In *Proteins at Interfaces*; American Chemical Society, 1987; ACS Symposium Series. Vol. *343*, pp 1−33.
(35) Horbett, T. A.; Brash, J. L. Proteins at Interfaces II: Fundamentals and Applications.*ACS Symposium Series No. 602*, 1995.
(36) Norde, W.; Horbett, T. A.; Brash, J. L. Proteins at interfaces III: Introductory Overview. *ACS Symp. Ser.* **2012**, *1120*, 1−34.
(37) Cho, D. H.; Xie, T.; Truong, J.; Stoner, A. C.; Hahm, J. I. Recent Advances towards Single Biomolecule Level Understanding of Protein Adsorption Phenomena Unique to Nanoscale Polymer Surfaces with Chemical Variations. *Nano Res.* **2020**, *13* (5), 1295−1317.
(38) Hanson, K. L.; Fulga, F.; Dobroiu, S.; Solana, G.; Kaspar, O.; Tokarova, V.; Nicolau, D. V. Polymer Surface Properties Control the Function of Heavy Meromyosin in Dynamic Nanodevices. *Biosens. Bioelectron.* **2017**, *93*, 305−314.
(39) Ouassil, N.; Pinals, R. L.; Del Bonis-O'Donnell, J. T.; Wang, J. W.; Landry, M. P. Supervised Learning Model Predicts Protein Adsorption to Carbon Nanotubes. *Sci. Adv.* **2022**, *8* (1), No. eabm0898.
(40) Duan, Y.; Coreas, R.; Liu, Y.; Bitounis, D.; Zhang, Z.; Parviz, D.; Strano, M.; Demokritou, P.; Zhong, W. Prediction of Protein Corona on Nanomaterials by Machine Learning using Novel Descriptors. *NanoImpact* **2020**, *17*, No. 100207.
(41) Kwaria, R. J.; Mondarte, E. A. Q.; Tahara, H.; Chang, R.; Hayashi, T. Data-Driven Prediction of Protein Adsorption on Self-Assembled Monolayers toward Material Screening and Design. *ACS Biomater. Sci. Eng.* **2020**, *6* (9), 4949−4956.
(42) Le, T. C.; Penna, M.; Winkler, D. A.; Yarovsky, I. Quantitative Design Rules for Protein-Resistant Surface Coatings using Machine Learning. *Sci. Rep.* **2019**, *9* (1), No. 265.
(43) Liu, Y.; Zhang, D.; Tang, Y.; Zhang, Y.; Chang, Y.; Zheng, J. Machine Learning-Enabled Design and Prediction of Protein Resistance on Self-Assembled Monolayers and Beyond. *ACS Appl. Mater. Interfaces* **2021**, *13* (9), 11306−11319.
(44) Liu, Y.; Zhang, D.; Tang, Y.; Zhang, Y.; Gong, X.; Xie, S.; Zheng, J. Machine Learning-Enabled Repurposing and Design of Antifouling Polymer Brushes. *Chem. Eng. J.* **2021**, *420*, No. 129872.
(45) Smith, J. R.; Kholodovych, V.; Knight, D.; Kohn, J.; Welsh, W. J. Predicting Fibrinogen Adsorption to Polymeric Surfaces in Silico: A combined method approach. *Polymer* **2005**, *46* (12), 4296−4306.
(46) Ferraris, S.; Cazzola, M.; Peretti, V.; Stella, B.; Spriano, S. Zeta Potential Measurements on Solid Surfaces for in Vitro Biomaterials Testing: Surface Charge, Reactivity upon Contact with Fluids and Protein Absorption. *Front. Bioeng. Biotechnol* **2018**, *6* (MAY), No. 60.
(47) Carbone, M. R. When not to Use Machine learning: A Perspective on Potential and Limitations. *MRS Bull.* **2022**, *47* (9), 968−974.
(48) Garland, A.; Shen, L.; Zhu, X. Mobile Precursor Mediated Protein Adsorption on Solid Surfaces. *Prog. Surf. Sci.* **2012**, *87* (1−4), 1−22.
(49) Wang, K.; Zhou, C.; Hong, Y.; Zhang, X. A Review of Protein Adsorption on Bioceramics. *Interface Focus* **2012**, *2* (3), 259−277.
(50) Rabe, M.; Verdes, D.; Seeger, S. Understanding Protein Adsorption Phenomena at Solid Surfaces. *Adv. Colloid Interface Sci.* **2011**, *162* (1−2), 87−106.
(51) Quan, X.; Liu, J.; Zhou, J. Multiscale Modeling and Simulations of Protein Adsorption: Progresses and Perspectives. *Curr. Opin. Colloid Interface Sci.* **2019**, *41*, 74−85.
(52) Raffaini, G.; Ganazzoli, F. Protein Adsorption on Biomaterial and Nanomaterial Surfaces: A Molecular Modeling Approach to Study Noncovalent Interactions. *J. Appl. Biomater. Biomech.* **2010**, *8* (3), 135−145.

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

(53) Penna, M.; Yarovsky, I. Nanoscale: In silico Classification of Ligand Functionalised Surfaces for Protein Adsorption Resistance. *Nanoscale* **2020**, *12* (13), 7240−7255.
(54) Latour, R. A. Perspectives on the Simulation of Protein-Surface Interactions using Empirical Force Field Methods. *Colloids Surf., B* **2014**, *124*, 25−37.
(55) Kim, J. Mathematical Modeling Approaches to Describe the Dynamics of Protein Adsorption at Solid Interfaces. *Colloids Surf., B* **2018**, *162*, 370−379.
(56) Brash, J. L.; Horbett, T. A.; Latour, R. A.; Tengvall, P. The Blood Compatibility Challenge. Part 2: Protein Adsorption Phenomena Governing Blood Reactivity. *Acta Biomaterialia* **2019**, *94*, 11−24.
(57) Chung, W. K.; Freed, A. S.; Holstein, M. A.; McCallum, S. A.; Cramer, S. M. Evaluation of Protein Adsorption and Preferred Binding Regions in Multimodal Chromatography using NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (39), 16811−16816.
(58) Stone, M. C.; Carta, G. Protein Adsorption and Transport in Agarose and Dextran-grafted Agarose Media for Ion Exchange Chromatography. *J. Chromatogr. A* **2007**, *1146* (2), 202−215.
(59) Langer, E. S.; Rader, R. A. Single-use Technologies in Biopharmaceutical Manufacturing: A 10-year Review of Trends and the Future. *Eng. Life Sci.* **2014**, *14* (3), 238−243.
(60) Tsougeni, K.; Petrou, P. S.; Papageorgiou, D. P.; Kakabakos, S. E.; Tserepi, A.; Gogolides, E. Controlled Protein Adsorption on Microfluidic Channels with Engineered Roughness and Wettability. *Sens. Actuators, B* **2012**, *161* (1), 216−222.
(61) Sharma, S.; Agarwal, G. P. Interactions of Proteins with Immobilized Metal Ions: A Comparative Analysis using Various Isotherm Models. *Anal. Biochem.* **2001**, *288* (2), 126−140.
(62) Smith, S. C.; Ahmed, F.; Gutierrez, K. M.; Frigi Rodrigues, D. A Comparative Study of Lysozyme Adsorption with Graphene, Graphene oxide, and Single-walled Carbon Manotubes: Potential Environmental Applications. *Chem. Eng. J.* **2014**, *240*, 147−154.
(63) Mollerup, J. M. A Review of the Thermodynamics of Protein Association to Ligands, Protein Adsorption, and Adsorption Isotherms. *Chem. Eng. Technol.* **2008**, *31* (6), 864−874.
(64) Tan, S.; Saito, K.; Hearn, M. T. W. Isothermal Modelling of Protein Adsorption to Thermo-Responsive Polymer Grafted Sepharose Fast Flow Sorbents. *J. Sep. Sci.* **2021**, *44* (9), 1884−1892.
(65) Mohsen-Nia, M.; Massah Bidgoli, M.; Behrashi, M.; Mohsen Nia, A. Human Serum Protein Adsorption onto Synthesis Nano-hydroxyapatite. *Protein J.* **2012**, *31* (2), 150−157.
(66) Katira, P.; Agarwal, A.; Hess, H. A Random Sequential Adsorption Model for Protein Adsorption to Surfaces Functionalized with Poly(ethylene oxide). *Adv. Mater.* **2009**, *21* (16), 1599−1604.
(67) Latour, R. A. The Langmuir Isotherm: A Commonly Applied but Misleading Approach for the Analysis of Protein Adsorption Behavior. *J. Biomed. Mater. Res., Part A* **2015**, *103* (3), 949−958.
(68) Halperin, A.; Buhot, A.; Zhulina, E. B. On the Hybridization Isotherms of DNA Microarrays: The Langmuir Model and its Extensions. *J. Phys.: Condens. Matter* **2006**, *18* (18), S463−S490.
(69) Eibl, R.; Eibl, D. *Single-use Technology In Biopharmaceutical Manufacture*; John Wiley & Sons, 2019.

https://doi.org/10.1021/acsami.4c06759
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX