ACS APPLIED MATERIALS & INTERFACES

www.acsami.org

Research Article

Machine Learning-Enabled Design and Prediction of Protein Resistance on Self-Assembled Monolayers and Beyond

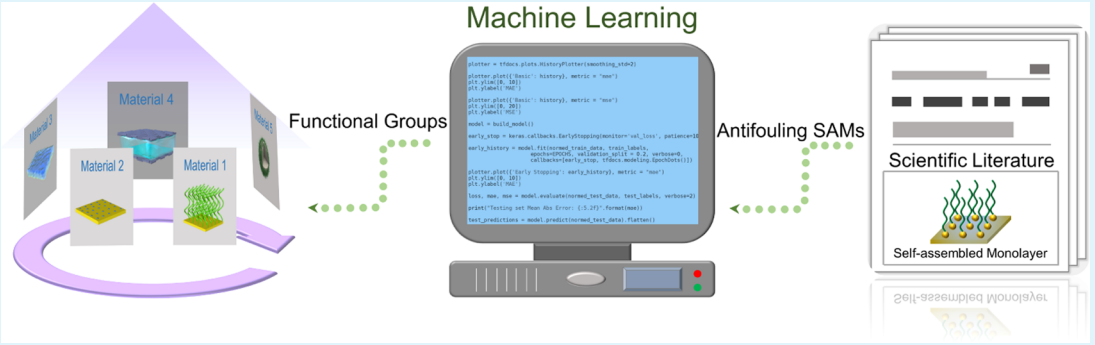
Yonglan Liu,[§] Dong Zhang,[§] Yijing Tang, Yanxian Zhang, Yung Chang, and Jie Zheng*

Cite This: *ACS Appl. Mater. Interfaces* 2021, 13, 11306–11319

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

Downloaded via UNIV OF ILLINOIS CHICAGO on December 7, 2021 at 21:22:29 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.



ABSTRACT: The rational design of highly antifouling materials is crucial for a wide range of fundamental research and practical applications. The immense variety and complexity of the intrinsic physicochemical properties of materials (i.e., chemical structure, hydrophobicity, charge distribution, and molecular weight) and their surface coating properties (i.e., packing density, film thickness and roughness, and chain conformation) make it challenging to rationally design antifouling materials and reveal their fundamental structure−property relationships. In this work, we developed a data-driven machine learning model, a combination of factor analysis of functional group (FAFG), Pearson analysis, random forest (RF) and artificial neural network (ANN) algorithms, and Bayesian statistics, to computationally extract structure/chemical/surface features in correlation with the antifouling activity of self-assembled monolayers (SAMs) from a self-construction data set. The resultant model demonstrates the robustness of Q^2_{CV} = 0.90 and $RMSE_{CV}$ = 0.21 and the predictive ability of Q^2_{ext} = 0.84 and $RMSE_{ext}$ = 0.28, determines key descriptors and functional groups important for the antifouling activity, and enables to design original antifouling SAMs using the predicted antifouling functional groups. Three computationally designed molecules were further coated onto the surfaces in different forms of SAMs and polymer brushes. The resultant coatings with negative fouling indexes exhibited strong surface resistance to protein adsorption from undiluted blood serum and plasma, validating the model predictions. The data-driven machine learning model demonstrates their design and predictive capacity for next-generation antifouling materials and surfaces, which hopefully help to accelerate the discovery and understanding of functional materials.

KEYWORDS: *antifouling, machine learning, QSAR, protein adsorption, self-assembled monolayer*

1. INTRODUCTION

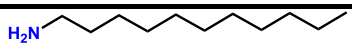
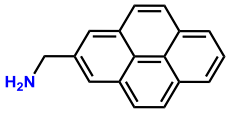
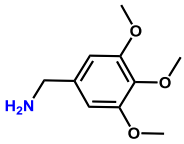
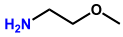
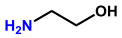
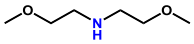
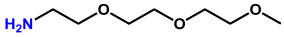
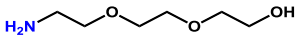
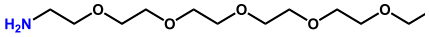
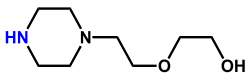
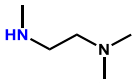
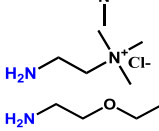
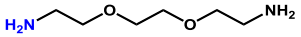
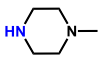
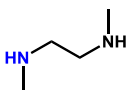
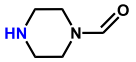
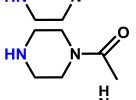
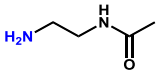
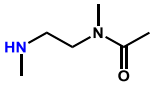
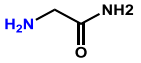
The design of functional materials for engineering and medicinal applications is still largely driven by the "trial-and-error" strategy, whose empirical design nature lacks control and prediction for the structure−function property of designed materials, thus making the entire design process expensive, time-consuming, and fraught with failures. With the advent of big data, artificial intelligence, and high-performance computing, the data-driven material discovery has become an emerging and powerful approach to accelerate the systematic design of original materials with predictable properties.[1] Historically, data-driven approaches have been well developed and widely applied to drug discovery,[2,3] disease diagnosis,[4] transportation analysis,[5] and financial risk assessment.[6] A general methodology of data-driven approaches still remains mostly unchanged, involving a four-step pipeline of data extraction, data enrichment, materials design/prediction, and experimental validation,[7] which has also been applied to data-driven material discovery for new functional materials (batteries,[8,9] catalysts,[10] nanoparticles,[11] and semiconductors[12,13]), as well as for predicting the specific property (conductivity,[14,15] catalytic activity,[10,16] photoluminescence,[17,18] thermodynamics,[19−21] and mechanical proper-

Received: January 11, 2021
Accepted: February 17, 2021
Published: February 26, 2021

© 2021 American Chemical Society

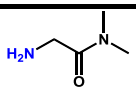
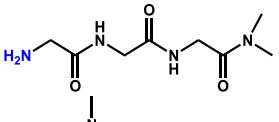
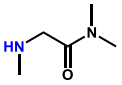
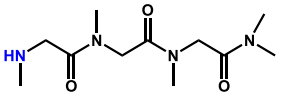
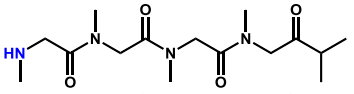
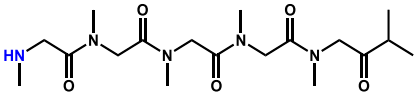
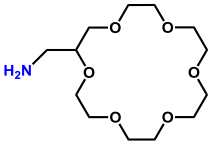
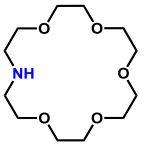
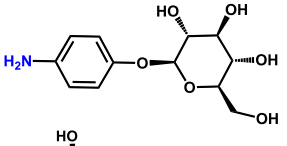
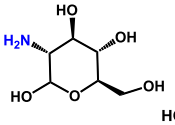
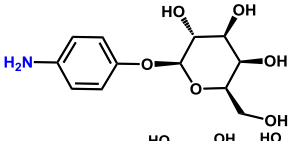
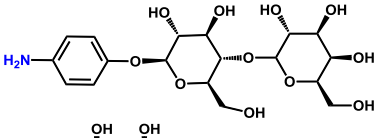
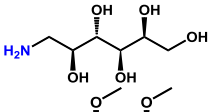
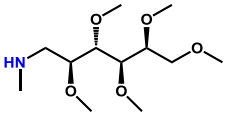
https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

Table 1. Relative Protein Adsorption Data (%ML) of Fibrinogen and Lysozyme on the 48 SAMs

Entry	Structure	Protein Adsorption	
		Fibrinogen (%ML)	Lysozyme (%ML)
01		100	100
02		96	86
03		106	130
04		75	105
05		53	15
06		47	14
07		50	35
08		1.5	1.1
09		1.6	1.0
10		0.3	0.5
11		0.4	1.0
12		3.7	27
13		40	14
14		25	3.9
15		37	18
16		16	1.0
17		8.5	7.8
18		54	80
19		39	12
20		38	6.5
21		40	5.4
22		12	3.8
23		58	30

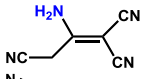
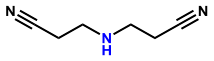
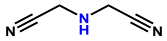
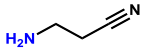
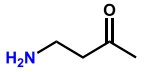
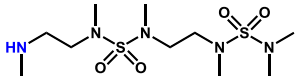
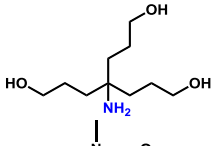
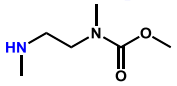
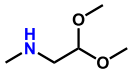
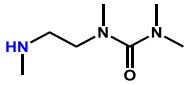
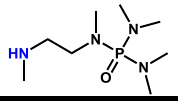
https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

Table 1. continued

Entry	Structure	Protein Adsorption	
		Fibrinogen (%ML)	Lysozyme (%ML)
24		33	15
25		22	8.8
26		9.1	2.4
27		2	1.1
28		1.7	1.5
29		1.3	1.0
30		59	9.6
31		11	6
32		67	83
33		55	25
34		49	30
35		32	9.3
36		68	20
37		2.9	6.1

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

Table 1. continued

Entry	Structure	Protein Adsorption	
		Fibrinogen (%ML)	Lysozyme (%ML)
38		58	28
39		58	40
40		73	92
41		72	69
42		74	71
43		80	100
44		57	73
45		58	43
46		36	19
47		25	11
48		4.3	0.3

ties[22,23]) of any given materials. Of note, these studies mainly focus on inorganic solid materials of metals,[24−26] ceramics,[27] zeolites,[28,29] and metal−organic frameworks (MOFs),[30,31] simply because of largely available data sets, well-characterized molecular structures, and mostly consistent chemical/physical/biological properties, which allows us to establish a reliable, data-driven model to encode the composition−structure−property relationships of original materials as accurate and comprehensive as possible. However, it remains a great challenge to develop a data-driven model for the rational design of original soft materials and coatings (e.g., polymers, elastomers, and hydrogels), mainly due to the lack of high-quality data at both structure and property levels.

Among different functional polymers, the development of highly bioinert and biocompatible antifouling materials is crucial for many scientific and technological applications including implantable devices, biosensors, membrane separation, antibacterial coatings, and regenerative medicine.[32,33] A wide variety of antifouling polymers have been developed to form biofouling-resistant surfaces against protein adsorption, cell adhesion, and/or bacterial attachment. However, the experimental design of antifouling surfaces that fulfill the criteria of low toxicity, broad-spectrum activity, and ease of production usually requires an empirical time/cost-expensive optimization procedure. Alternatively, the data-driven design of antifouling surfaces also encounters different road blockers. First, there are no existing antifouling databases available, thus it requires the researchers to collect experimental results from the literature to construct antifouling databases for different modelings. Furthermore, the antifouling performance of even the same materials, but coming from different laboratories, could be varied considerably due to different synthesis conditions and characterization methods. Next, molecular dynamics (MD) simulations allow us to reveal the specific interactions between any foulant and a given antifouling surface,[34−40] from which the antifouling performance of the surface can be readily determined by the extent of repulsive forces acting on the foulant, along with the examination of the structural effects of carbon spacer lengths,[34] surface terminal groups,[35] surface dipole orientations,[41] surface hydrophilicity,[42] and surface grafting density[36] on the antifouling performance of the surface. From a viewpoint of the computational design, it also remains challenging to discover special structural descriptors to encode antifouling materials that possess a wide variety of molecular weights, conformationally labile structures, and nonsystematic cross-linking. While MD simulations can provide atomic details of structural, dynamics, and free energy properties in relation to antifouling mechanisms, MD simulations only allow us to study individual antifouling material systems in a "one-at-a-time" manner, thus lacking a data-driven capacity for the rapid prediction of a large number of antifouling materials from a given database or for building a broad data set by examining enough materials, due to high computational cost. Apparently, these difficulties urge a strong need for constructing a reliable antifouling material database and a computational predictive model for better

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

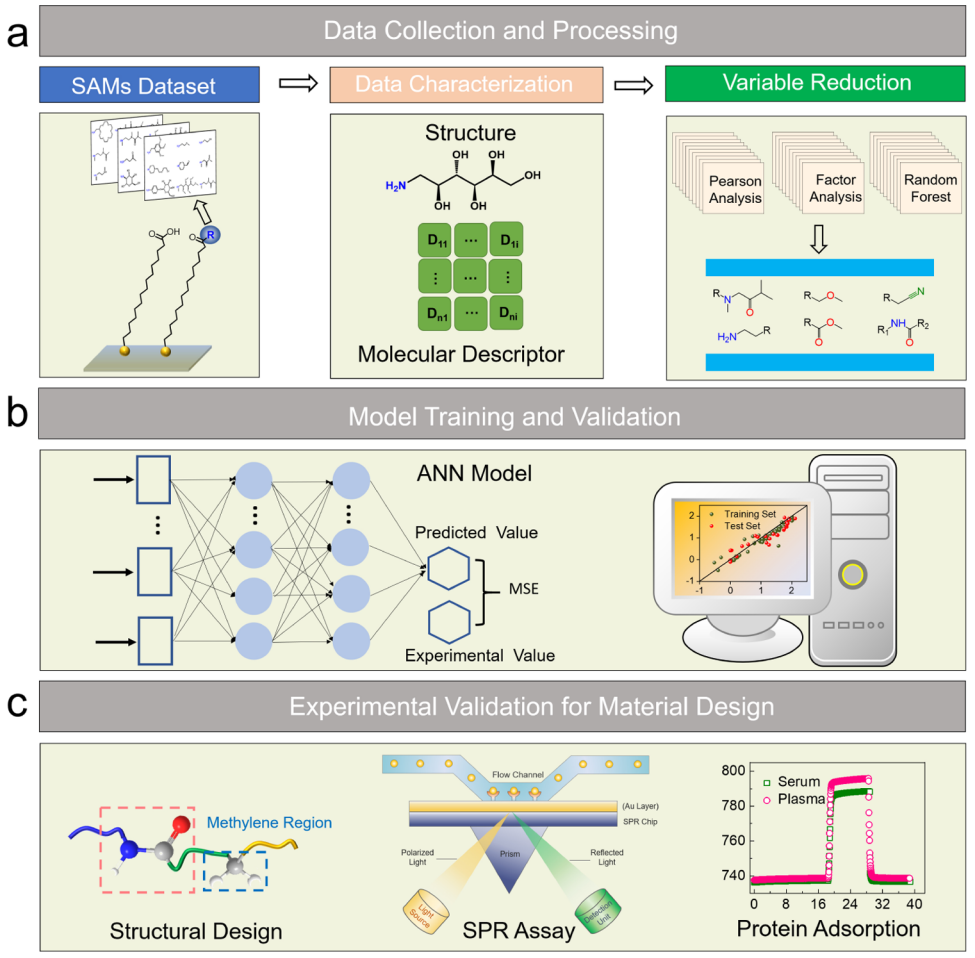


Figure 1. Schematic workflow of a three-step machine learning model, including (a) data collection and processing including data set construction of SAMs from the literature (left), data featuring by molecular descriptors (middle), and variable reduction by a combination of Pearson analysis, factor analysis, and random forest algorithm (right), (b) model training and validation by ANN, and (c) material prediction/design by SPR validation.

assessing the component−structure−property−performance relationship of antifouling surfaces.

The quantitative−structure−property/activity relationship (QSPR or QSAR)[43] assisted by machine learning algorithms has recently been reported to study the antifouling property of different materials and coatings. Almeida et al.[44] developed a linear regression QSAR model to evaluate the antifouling activity of chalcone derivatives against various micro- and macrofouling species. Descriptors that encode the hydrogen bonds, shape, branching ratio, and constitutional diversity of the molecule were found as influential factors for the antifouling activity. As guided by the QSAR model, several chalcone derivatives were synthesized to show a high inhibitory effect on some bacterial growth. Feng et al.[45] also built a QSAR model to evaluate the inhibition activity of acylamino compounds containing gramine groups against *Escherichia coli* growth. The model indicated that the antibacterial activity of these compounds can be greatly improved by reducing the HOMO energy and molar refractivity. Furthermore, QSAR modeling for antifouling coatings becomes even more complicated because apart from the physicochemical property of materials (i.e., chemical structure, hydrophobicity, charge distribution, and molecular weight), surface properties (i.e., packing density, film thickness and roughness, and chain conformation) also contribute to antifouling performance. Rasulev et al.[46] presented genetic algorithm-multiple linear regression-based (GA-MLR-based) QSAR models to predict the fouling-release activity of different marine fouling organisms (bacteria, algae, and barnacles) on amphiphilic polysiloxane-based polymer coatings. They found that several descriptors of polarizability indices, size, mass, and van der Waals volume were critically important for the fouling-release activity of polysiloxane-based coatings. Le et al.[47] applied both linear (MLREM) and non-linear (BRANNGP and BRANNLP) methods to quantitatively predict the protein adsorption on self-assembled monolayers (SAMs). The models revealed that the additional properties of size flexibility and the polarizability of SAMs are highly related to their antifouling ability. Specifically, SAMs with large, conformationally mobile, and polarizable functional groups (e.g., ethylene glycol group) are predicted to highly resist the protein adsorption. Recently, Kwaria et al.[48] recently used the different data sets of SAMs to construct a QSAR model, which can predict both the water contact angle and protein adsorption on SAMs and determine the importance of each structural parameter for the water contact angle and protein adsorption.

The abovementioned QSAR/QSPR studies, despite very few to date, mostly focus on the evaluation of the antifouling property of any given materials and surfaces, instead of the design of original antifouling materials and surfaces. Addition-

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

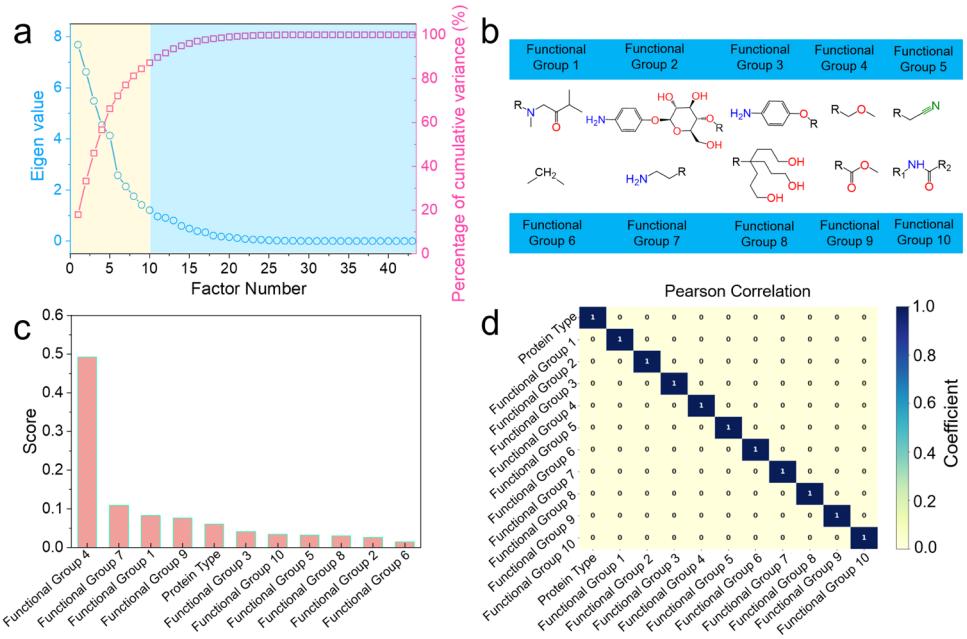


Figure 2. Data set selection and factor determination for a machine learning model. (a) Eigenvalues (blue) and percentage of cumulative variance (%) (red) for 43 factors and (b) molecular structures of 10 explicit functional groups from factor analysis. (c) Importance score ranking of 11 variables (10 functional groups and protein type) from the random forest algorithm. (d) Correlation coefficient matrix between every two variables of 11 variables from Pearson analysis.

ally, the key descriptors in these models (e.g., N-066, N-067, C-002, C-006, AlogP, etc.) lack the physical or chemical basis, making the design of original antifouling materials not straightforward. To address the above limitations, we first assembled the protein adsorption data on self-assembled monolayers (SAMs) from the literature.[49] Table 1 summarizes the compositions of the 48 SAMs, all of which contain well-characterized end points of the protein adsorption amount and cover a wide variety of chemical structures (e.g., hydrophilic groups of hydroxyl, amide, acrylate, ether, and ethylene glycol[50−53] and zwitterionic groups of carboxylic, sulfonate, sulfate, and quaternary/tertiary/secondary/primary ammonium[54−56]). The molecular-level surface uniformity and high structural diversity of SAMs make them an ideal model for generating protein adsorption data in a consistent manner, which can be used as a reliable source to develop machine learning models for studying and understanding the adsorption of proteins on surfaces. Figure 1 shows a three-step process of our machine learning model with a four-layer artificial neural network (ANN), including data collection and processing from experiments (Figure 1a), model training and validation by ANN (Figure 1b), and material design validation by experiments (Figure 1c). The resulting models were able to determine the important molecular descriptors for surface resistance to proteins, provide the specific functional groups important for antifouling materials, and design molecules with the predicted antifouling properties as both positive and negative fouling indexes. Finally, we synthesized the three antifouling molecules, coated them onto surfaces in different forms of SAMs and polymer brushes, and tested their protein resistance property by surface plasma resonance (SPR). SPR results showed that SAMs and polymer brushes with negative fouling indexes exhibited excellent surface resistance to protein adsorption from undiluted human blood serum and plasma, while surface coatings with positive fouling indexes had high protein adsorption, consistent with the model predictions. This work demonstrates a data-driven machine learning model for the functional group-based design of antifouling surfaces.

2. RESULTS AND DISCUSSION

2.1. Data Set Selection and Factor Determination for Machine Learning Models. Table 1 collects a total of 48 self-assembled monolayers (SAMs), each containing two different protein adsorption data (fibrinogen and lysozyme). The protein resistance capacity (%ML) of any SAM is determined and normalized by the protein adsorption amount on a hydrophobic SAM of $-CONH(CH_2)_{10}CH_3$ as a reference. Preliminary analysis has shown that entries 02, 14, 43, and 48 possessed significant noise and weak reliability in the machine learning model, thus they were excluded from the data set. Finally, there were a total of 88 protein adsorption data points (44 SAMs × 2 protein adsorption data = 88) for subsequent data featuring using molecular descriptors by alvaDesc software and data training using ANN. The initial structural and property analysis of SAMs led to 105 descriptors, which were then reduced to 43 descriptors by removing some redundant or over-correlated descriptors using a criterion of the paired correlation coefficient of >0.85. Next, Figure 2a shows the factor analysis of the eigenvalues and percentage of cumulative variance (%) on 43 descriptors. It can be seen that the first 10 functional group factors (Figure 2b) possessed an eigenvalue of >1, a communality value of >0.5 (Table S1), a rotated regression coefficient of >0.3 (Table S2), and 87.3% structural variance of the 43 descriptors (Table S3), all indicating that these 10 functional groups are sufficient for encoding the compositional, structural, geometrical, and connectivity information of 43 descriptors. Next, we applied the random forest algorithm to determine the importance of 10 functional groups and the protein type for their contributions to the protein adsorption on SAMs. Figure 2c shows a decreasing important score of 11 variables (10 functional groups and protein type), all of which have an acceptable and

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

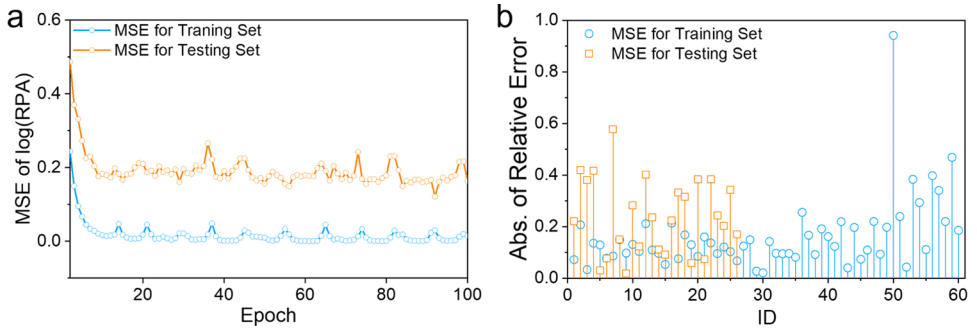


Figure 3. Training performance of the ANN model. (a) Training epoch evolution of MSE and (b) relative errors of predicted values with respect to the experimental values for both training and test sets in the ANN model.

important score of >0.01, while Pearson analysis in Figure 2d shows the high independence of 11 variables, both confirming that 11 resulting variables are qualified for a machine learning model.

2.2. Training and Validation of Machine Learning Models. We have applied both ANN and supporting vector regression (SVR) models with radial basis function (RBF) and polynomial kernels to (i) train and test the antifouling property of SAMs in relation to the functional groups and (ii) compare their reliability and predictivity of different machine learning models. However, both internal and external cross-validation have shown that the ANN model (Q^2_{CV} = 0.9 and Q^2_{ext} = 0.84) exhibits the better reliability and predictivity than SVR-RBF (Q^2_{CV} = 0.86 and Q^2_{ext} = 0.71) and SVR-Polynomial (Q^2_{CV} = 0.93 and Q^2_{ext} = 0.29) models. More importantly, the SVR-Polynomial model showed the overfitting behavior, as evidenced by a much higher value of Q^2_{CV} (0.93) than that of Q^2_{ext} (0.29) (data not shown). In this work, we used a four-layer ANN model for the prediction of protein adsorption on SAMs, starting with the first input layer containing the data feature of 11 variables (10 functional group factors and protein type), followed by the second and third hidden layers containing 96 and 48 neurons, ending with an objective function of the logarithmic relative protein adsorption amount, log(RPA), on SAMs. A binary function was used to present the protein type (1 for fibrinogen and 0 for lysozyme). During the ANN training process, two protein adsorption data of lysozyme on SAMs (entries 16 and 30) were identified as outliers, thus they were removed from the data set. For the remaining 86 protein adsorption data, 70% of the data set (60 data) was used for training, while the remaining 30% (26 data) was used for the testing of the trained model. The training performance of the ANN model was assessed by the convergence of mean squared errors (MSEs) of log(RPA) for both training and test sets. Figure 3a shows that both MSE values for training and validation stopped at a very early stage of 30 epochs, while Figure 3b further shows that all errors (100%) for the test set were less than 0.6, while 59 errors (98%) for the training set were less than 0.5. We further randomly distributed the whole data to 70% of the training set and 30% of the test set and similar behaviors of MSE values for the validation and testing sets were observed, confirming the satisfactory distribution of the data. These results indicate the good fitting and training performance of the ANN model for both validation and testing sets.

Next, we applied both leave-five-out cross-validation and external validation to assess the reliability and predictivity of the ANN model, respectively. In Table 2, the leave-five-out cross-validation conducted the five iterative and random tests on the training set, yielding Q^2_{CV} = 0.90 and $RMSE_{CV}$ = 0.21. The high value of Q^2_{CV} and the small value of $RMSE_{CV}$ confirmed that the constructed model was stable. Meanwhile, external validation on the test set gave Q^2_{ext} = 0.84 and $RMSE_{ext}$ = 0.28, both of which were also close to those for cross-validation on the training set. This indicates the good prediction power of the ANN model without significant overfitting. In addition, the low values of both $RMSE_{CV}$ and $RMSE_{ext}$ also indicated that the model was uncertain. Consistently, a plot of the predicted log(RPA) versus the experimental ones for both training and test sets gave r^2 values of 0.92 and 0.87, respectively (Figure 4a), indicative of the high predictive accuracy of the model.

Table 2. Internal Cross-Validation and External Validation for the ANN Model

validation method	number of samples	Q^2	RMSE
leave-five-out cross-validation	60	0.90	0.21
external validation	26	0.84	0.28

In parallel, to analyze the accuracy and robustness of models, we assessed whether the predefined applicability domain could accurately identify query samples from the test set using the *k*-nearest neighbor (*k*NN) approach.[57] Figure 4b shows that the applicability domain increased dramatically as the *k* value increased from 1 to 4 but remained almost unchanged after *k* = 5. When the *k* value (1−4) is too small, more samples would be out of the applicability domain, leading to the diminution of the data diversity. When the *k* value is too large, very few samples will be excluded to achieve the purpose of optimizing models. Taking these factors into consideration, the optimal *k* value was set to 8. The models with *k* = 8 performed best, proving the prediction reliability of models.

2.3. Structure-Dependent Design and Protein Resistance Property of SAMs. To better understand the structure−antifouling property relationship of SAMs, we performed Bayesian statistics to quantify the contribution of 10 functional groups and the protein type to protein adsorption capacity (i.e., fouling index) on SAMs. In Figure 5, a positive fouling index indicates that the functional group promotes protein adsorption on SAMs, while a negative index indicates a better antifouling property. The probability distribution of the fouling index of functional groups presents the extent of importance for protein adsorption on SAMs. It can be seen that among 10 functional groups, functional groups 1 and 4 had almost 100% of the negative fouling index, suggesting that SAMs containing these groups would have the

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

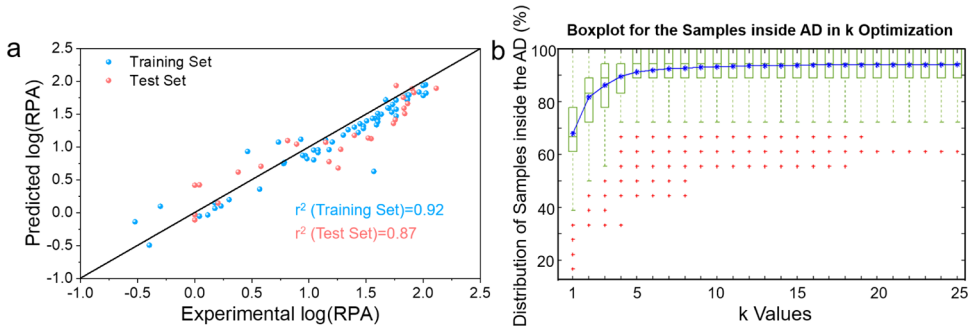


Figure 4. Validation of the predictivity and reliability of the ANN model. (a) Plot of the predicted versus experimental log(RPA) for both training and test sets from leave-five-out cross-validation and external validation. (b) Applicability domain analysis of 26 testing samples using the *k*NN approach.

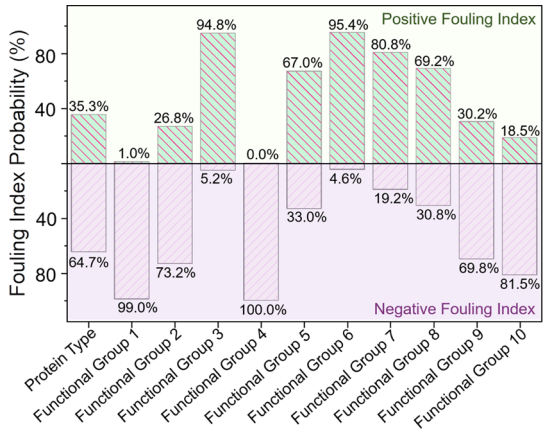
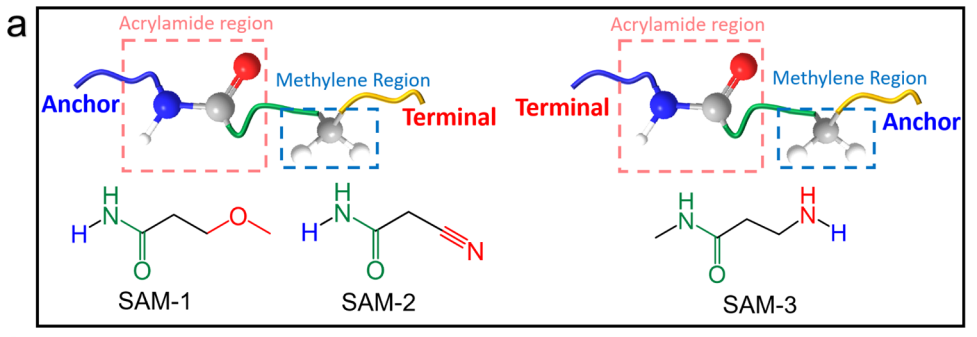


Figure 5. Fouling index probability (%) to describe the protein resistance and protein adsorption of 10 functional groups and the protein type.

strongest surface resistance to proteins. Functional group 1 containing tertiary amine ($-NR_3$) and carbonyl ($=O$) groups as hydrogen bond acceptors has a similar chemical structure to acrylamide, while functional group 4 has a sole hydrogen bond acceptor of the ether ($-O-$) group. The presence of a large number of hydrogen bond acceptors in SAMs enables the formation of a strong surface hydration layer on SAMs, which in turn offers a physical barrier to prevent protein adsorption. Consistently, SAM entries 26, 27, 28, and 29 or entries 8, 10, and 37 containing either group 1 or group 4 exhibited the low protein adsorption of 0.3−9.1%ML for fibrinogen and 0.5−6.1%ML for lysozyme (Table 1). Next, additional functional group 10 also had the higher negative fouling index of 81.5%, which was consistent with the experimental observation that SAM entries 21 and 25 containing group 10 exhibited high surface resistance to nonspecific protein adsorption, as evidenced by 22−40%ML for fibrinogen and 5.4−8.8%ML for lysozyme. The other two functional groups 2 and 9 had smaller negative fouling indexes of 73.2 and 69.8%, respectively, leading to the weaker surface resistance ability as compared to functional groups 1, 4, and 10. These three groups are hydrophilic and contain both hydrogen bond acceptors and donors of O−, OH−, NH−, and COO−, again indicating that the hydrogen bonds formed at SAM/water interfaces are critical factors for protein resistance.[58,59] Thus, the protein resistance of SAMs is attributed to their ability to form a strong surface hydration layer via hydrogen bonding groups. In contrast, groups 3, 5, 6, 7, and 8 possessed dominant hydrophobic motifs of the aromatic ring, methyl chains, and



b

SAM Entry	Predicted Protein Adsorption Amount (%ML)	
	Fibrinogen	Lysozyme
SAM-1	12.64	8.47
SAM-2	71.75	47.31
SAM-3	7.52	4.53

Figure 6. (a) Molecular design of antifouling SAMs with different terminal groups and anchors for synthesis. (b) Predicted protein adsorption amount (%ML) of fibrinogen and lysozyme on the three designed antifouling SAMs.

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

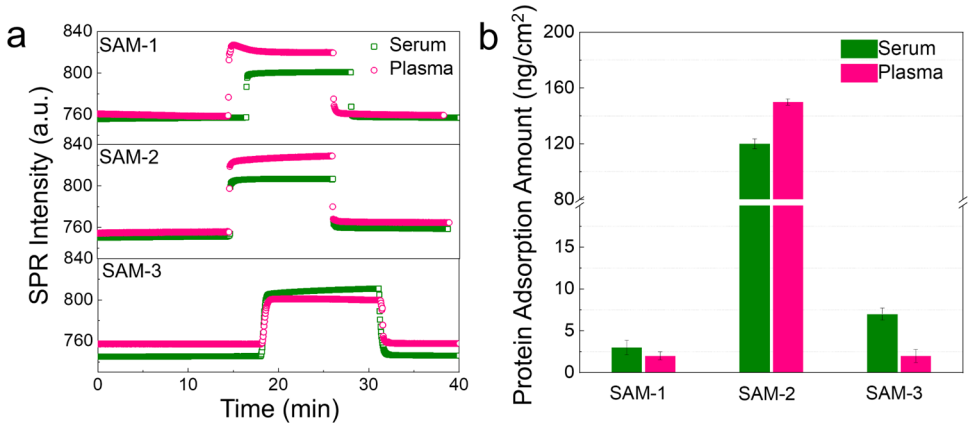


Figure 7. Experimental validation of the protein resistance property of the three designed SAMs. (a) SPR sensorgrams of protein adsorption from undiluted human blood plasma on the three SAMs. (b) Final protein adsorption amount on the three SAMs, whose data are extracted from SPR sensorgrams in (a).

highly charged cyano groups, which contributed to the higher positive fouling indexes of 67.0−95.4%. This is not surprising because hydrophobic and charged surfaces (e.g., entries 32, 34, 35, 38, 39, 40, 41, and 44) always promote nonspecific protein adsorption.[60,61] Apart from the intrinsic antifouling property of functional groups alone, the surface properties of packing, grafting density, smoothness, and thickness also contribute to the overall antifouling property. In addition, fibrinogen has a higher positive fouling index than lysozyme, suggesting that fibrinogen is relatively easier to be adsorbed on the studied SAMs than lysozyme. This could be due to different sizes and pI between fibrinogen (340 kDa, pI = 5.5) and lysozyme (15 kDa, pI = 10.9). The large and negatively charged fibrinogen at physiological pH has a higher probability to be adsorbed on the SAMs we examined in the data set.

Based on the predictive relationship between the functional groups and protein resistance of SAMs, we applied an inverse molecular design strategy to design the three antifouling molecules via a combination of different functional groups with negative fouling indexes. Briefly, the designed antifouling molecule consists of a common backbone of methylene (group 6) and acrylamide (group 10) and different terminal groups of 1−9. A number of studies have demonstrated that acrylamide-based polymers can achieve excellent antifouling performance to resist protein adsorption, cell/bacteria adhesion, and macrofoulant attachment.[35,39,60,62,63] A primary or secondary amine group is included for the substitution reaction by replacing one hydrogen atom attached to a nitrogen atom with the acyl chain.[49] The inclusion of the hydrophobic methylene group is to increase the mechanical property of coating materials. From a synthesis viewpoint, two design strategies were applied to use either acrylamide or methylene regions as surface anchor sites to connect with the acyl chain (Figure 6a). Before the experimental tests, we computationally assessed the protein resistance property of the three designed SAMs. As shown in Figure 6b, SAM-2 containing group 5 promoted the adsorption of both fibrinogen (71.75%ML) and lysozyme (47.31%ML). This again confirms that group 5 with a positive fouling index contributes to protein adsorption. In contrast, the other two designs of SAM-1 and SAM-3 showed a high surface resistance to both fibrinogen and lysozyme adsorption. Among them, SAM-3 presented the best and predicted protein resistance property, due to the presence of multiple hydrogen bonding groups, allowing the formation of a massive hydrogen bond network with interfacial water for preventing protein adsorption. More importantly, since all of the designed SAMs have the same functional groups in the backbone, our modeling indicates that different adsorbed amounts of proteins are largely attributed to terminal groups in SAMs.[64]

2.4. Protein Adsorption on the Designed Antifouling SAMs. Upon demonstration of the predicted protein adsorption property of the designed SAMs, we synthesized the three molecules capped with S atoms and then formed SAMs on the gold substrate via S−Au anchoring chemistry. Next, the antifouling property (i.e., protein adsorption amount) of the resultant SAMs was evaluated by undiluted blood plasma and serum containing ∼500 proteins by SPR. Figure 7a shows three typical SPR sensorgrams of protein adsorption on SAMs. As an example of SAM-1 with the excellent predicted protein resistance property, upon flowing freshly prepared undiluted blood plasma through independent SPR channels coated with SAM-1, a very large amount of proteins was immediately adsorbed on SAM-1. However, when the plasma solution was switched to the phosphate-buffered saline (PBS) solution, the initially adsorbed proteins on the SAM-1 surface were almost completely washed away by PBS, leading to protein adsorption of ∼2 and ∼3 ng/cm^2 from undiluted blood plasma and serum. This indicates that initially adsorbed proteins are very loosely bound to SAM-1. For comparison, final adsorbed proteins from undiluted blood serum/plasma on the three SAMs were ∼3/∼2 ng/cm^2 on SAM-1, ∼120/∼150 ng/cm^2 on SAM-2, and ∼7/∼4 ng/cm^2 on SAM-3 (Figure 7b), whose protein resistance capacity is in the decreasing order of SAM-1 > SAM-3 > SAM-2, consistent with modeling prediction. There are two major differences between SAM-1 and SAM-3. First, while the amide group is presented in the three newly designed SAMs (SAM-1, SAM-2, and SAM-3), the amide group was used as an anchor group for SAM-1 and SAM-2 but as a terminal group for SAM-3. Second, while both SAM-1 and SAM-3 achieved excellent surface resistance to protein adsorption of ∼3/∼2 and ∼7/∼4 ng/cm^2 from undiluted blood serum/plasma, the ether group in SAM-1 and the amide group in SAM-3 contributed differently to their respective protein resistant ability.

Different from other indirect, descriptor-based material designs, we further expanded our functional group-based design strategy to design, from a synthesis viewpoint, other types of acrylamide derivatives of *N*-(2-cyanoethyl)acrylamide

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

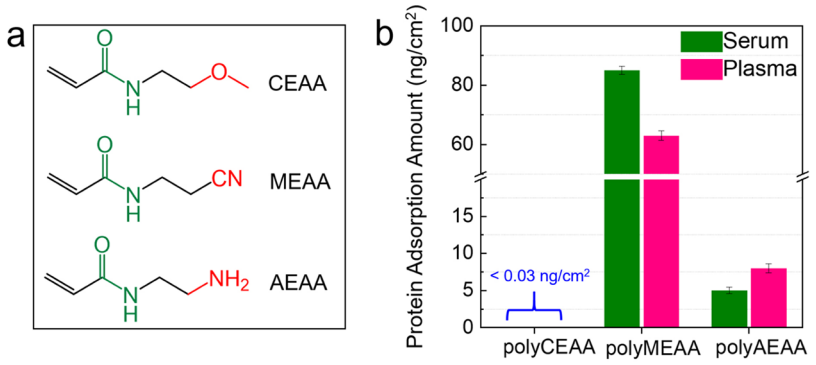


Figure 8. Experimental validation of the protein resistance property of the three designed acrylamide derivatives and polymer brushes. (a) Machine learning-enabled design of the three monomeric acrylamide derivatives of CEAA, MEAA, and AEAA. (b) Protein adsorption from undiluted human serum and plasma on the three polyCEAA, polyMEAA, and polyAEAA brushes by SPR.

(CEAA), *N*-(2-methoxyethyl)acrylamide (MEAA), and *N*-(2-aminoethyl)acrylamide (AEAA) (Figure 8a), followed by the polymerization and grafting of them onto a gold substrate to form polymer brushes using a surface-initiated atom transfer radical polymerization (ATRP) method.[65,66] The three resultant polymer brushes were tested for their protein resistance capacity by undiluted blood plasma and serum. As shown in Figure 8b, the polyCEAA brush involving group 4 of methyl-alkoxy ether with a higher negative fouling index achieved extremely high surface resistance to protein adsorption, as evidenced by the undetectable amount (<0.03 ng/cm^2) of the adsorbed protein on polyCEAA from undiluted human serum/plasma. PolyAEAA brush with functional group 7 as the terminal group, instead of as the anchor site, enabled to produce a strong surface hydration layer through their rich hydrogen bond donors, thus leading to very low protein adsorption of ∼5 ng/cm^2 from human serum and ∼8 ng/cm^2 from human plasma. As a negative design, the polyMEAA brush containing group 5 of cyano with a higher positive fouling index exhibited a weaker surface resistance to protein adsorption of ∼85 ng/cm^2 from human serum and ∼63 ng/cm^2 from human plasma. In addition to antifouling SAMs and polymer brushes, the machine learning-enabled functional group-based design strategy could be further applied to other surface coatings of hydrogels and elastomers.

3. CONCLUSIONS

In this work, we complied protein adsorption data on SAMs from the literature and developed a data-driven machine learning model to assess the antifouling performance of any given and designed SAMs. The resultant model not only achieved excellent robustness of Q^2_{CV} = 0.90 and $RMSE_{CV}$ = 0.21 and predictive ability of Q^2_{ext} = 0.84 and $RMSE_{CV}$ = 0.21 of the model but also identified both critical molecular descriptors and functional groups important for the rational design of antifouling materials. Among them, the functional groups 1 (tertiary amine derivative), 4 (methyl-alkoxy ether), 9 (carboxy), and 10 (amide) were predicted to be the most promising groups for antifouling materials. Based on the model prediction above, we designed and synthesized three different SAMs and three polyacrylamide-based brushes containing functional groups with both positive and negative fouling indexes. As expected, among these designed SAMs and polymer brushes, two antifouling SAMs (SAM-1 and SAM-3) and two antifouling brushes (polyCEAA and polyAEAA) exhibited an excellent protein resistance of ≤8 ng/cm^2 from undiluted human serum/plasma measured by SPR, consistent with model prediction. An alternative to experimental or molecular simulation methods, this work demonstrates a data-driven machine learning method to not only gain quantitative perspectives on the informative material correlations between functional groups and the antifouling property of SAMs but also offer specific functional group-based design guidance for effective antifouling materials. Of note, it is clear for any data-driven method that the acquirement and expansion of high-quality data are equally important for the development of efficient machine learning algorithms and workflows, which are expected to greatly accelerate the material design process by examining wide compositional and structural spaces efficiently. On the other hand, it is not always necessary to mine massive data from open-access literature and databases, instead, it is more innovative and challenging to extract all possible information from the limited existing data to conduct state-of-art material designs via a "the best we can do with the limited data we have" strategy.

4. MATERIALS AND METHODS

4.1. Machine Learning Section. *4.1.1. Data Set.* In this work, we applied the data set published by Whitesides' group[49] for the adsorption of lysozyme and fibrinogen on 48 mixed self-assembled monolayers (SAMs) with 30 min exposure to present the machine learning study (Table 1). The protein adsorption (%ML) is referred to the amount of protein adsorbed on each mixed SAM(−COR) surface as the percentage of the protein adsorption amount on a referenced monolayer SAM of $-CONH(CH_2)_{10}CH_3$.[49] We used a binary function to present the protein type (1 for lysozyme and 0 for fibrinogen). To reduce the noise and ensure the reliability of the machine learning model, entries 02, 14, 43, and 48 were removed from the data set.

4.1.2. Molecule Drawing and Molecular Descriptor Calculation. The molecular structures of all SAMs were drawn using ChemDraw software. Geometrical optimization of these compounds was carried out using the Chem3D package through the Merck molecular force field (MMFF94) with the steepest descent algorithm. A pool of one-dimensional (1D) molecular descriptors containing constitutional indices, ring descriptors, functional group counts, atom-centered fragments, and molecular properties was selected and computed to characterize the SAMs in the data set using alvaDesc 1.0.14 software (https://www.alvascience.com/alvadesc). Based on the univariate statistics and correlation analysis, descriptors with missing values, constant values, near-constant values, and paired correlation coefficients of >0.95 were removed for accuracy, ultimately leading to 105 descriptors for model construction.

4.1.3. Variable Reduction for the Machine Learning Model. To reduce the high-dimensional data, we carried out four consecutive

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

procedures to reduce the number of variables. First, we statistically analyzed the correlation between the descriptors for the 105 molecular descriptors and removed redundant descriptors by taking care of multicollinearity between the descriptors. After removing the descriptors with the correlation coefficient greater than the benchmark value of 0.85, 43 descriptors were obtained. Then, factor analysis[67] was carried out to generate a smaller number of explicit functional groups, which can summarize the entire cluster of 43 resultant descriptors using IBM SPSS Statistics 26 software. The obtained functional groups are linearly correlated with the 43 descriptors and can describe the covarying variables by a latent structure. In this study, the principal component method was used to extract the functional groups by clustering the highly intercorrelated descriptors. The factor coefficients involved in a matrix are regression coefficients, which can quantify the relationship between the descriptors and the inferred functional groups. Rotation was performed to make the factors interpretable since it can maximize the loadings of some descriptors for the specific functional groups but minimize the loadings of these descriptors for other factors. To obtain independent functional groups with no correlation with each other, the varimax rotation algorithm with Kaiser normalization was used to rotate the factors to generate interpretable functional groups. A value of 0.3 was used as the cutoff for the absolute value of regression coefficients to select descriptors with high correlation with the interpretable functional group. In this way, descriptors with squared correlated coefficient values greater than 0.3 with the factors can be used for interpreting the factor as the functional group. Further, the random forest algorithm was performed to screen variables (functional groups and protein type) that were highly correlated with the antifouling property/protein adsorption ability of SAMs. Finally, Pearson analysis was performed to ensure the independence of these variables.

4.1.4. Training and Validation of the Machine Learning Model. In this work, we designed a four-layer sequential artificial neural network (ANN) using keras and TensorFlow modulus implemented in Python 3.6. The training of a machine learning model used random model initiation and the data set was randomly split into a training set and a test set with a ratio of 7/3 to ensure the repeatability of the predictive performance of the trained machine learning model. The training set was used to train the model based on the designed neural network, while the test set was applied to validate the predictivity and robustness of the model. This ANN model consists of an input layer that contains the variable information of the data. The dimension of the input layer was first flattened. We next added two dense hidden layers with 96 and 48 neurons. The activation function applied for the input layer and the two dense hidden layers were rectified linear unit (relu), hyperbolic tangent (tanh), and rectified linear unit (relu) functions, respectively. An output layer was used to present a prediction of the property of interest. Two dense hidden layers connect the input and the output layers through every node in the previous layer weighted connecting with every node in the next layer. Each dense layer enables to manage its own weight matrix and a vector of bias terms. Before starting the training of the model, the weights between the neurons of the previous layer and the neurons of the next layer were randomly assigned using the He normal method. During the training process, the weights between each neuron of the previous layer and each node of the next layer were consensus optimized. To obtain the optimal performance in the designed neural network, the error of the cost function was applied for optimizing the model evaluated by mean squared errors (MSEs) between the experimental log(RPA) and the predicted log(RPA) of the training set and the test set. For the training process, leave-five-out cross-validation and external validation were performed to estimate the convergence of MSE. To this end, deep-learning methods (e.g., automatic learning-rate reduction and early stopping) were carried out to avoid overfitting. When the MSE of both training and validation reaches the minimum cross-entropy, the model at this state was regarded as the optimal model.

The *k*-nearest neighbors (*k*NN) algorithm[57] was performed to evaluate the applicability domain (AD) for testing samples. The principle of *k*NN applied for evaluating AD basically reflects whether a testing sample is suitable to predict the property of interest by the constructed machine learning model. The threshold for the training samples was first defined. Then, the distance of a testing sample is considered from its *k* closest data points from the training samples in the chemical space. If the testing data was outside the threshold, it will be regarded as an outlier and vice versa. A lower distance reflects a higher similarity between training samples and testing samples, while a higher distance corresponds to a higher mismatch in the structural space between the training and testing samples. Herein, in the process of performing *k*NN, the *k* value was continuously optimized. At the *k* value with least outliers from the testing sample, *k* is the optimal one to evaluate AD.

4.1.5. Evaluation of the Machine Learning Model. The robustness and predictive ability of the ANN model were evaluated using the squared correlation coefficient (Q^2_{CV}) and root-mean-square error ($RMSE_{CV}$) from leave-five-out cross-validation as well as squared correlation coefficient (Q^2_{ext}) and root-mean-square error ($RMSE_{ext}$) from external validation. The following equations are used to quantify Q^2_{CV} (eq 1), $RMSE_{CV}$ (eq 2), Q^2_{ext} (eq 3), and $RMSE_{ext}$ (eq 4)

$$Q^2_{CV} = 1 - \frac{\sum_{i=1}^{j} (y_i^{obs} - y_i^{pred_{cv}})^2}{\sum_{i=1}^{j} (y_i^{obs} - y^{obs})^2} \quad (1)$$

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{j} (y_i^{obs} - y_i^{pred_{cv}})^2}{j}} \quad (2)$$

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{j} (y_i^{obs} - y_i^{pred_{ext}})^2}{\sum_{i=1}^{j} (y_i^{obs} - y^{obs})^2} \quad (3)$$

$$RMSE_{ext} = \sqrt{\frac{\sum_{i=1}^{j} (y_i^{obs} - y_i^{pred_{ext}})^2}{j}} \quad (4)$$

where y_i^{obs} and $y_{i^{cv}}^{pred}$ in eqs 1 and 2 are the experimental and predicted log(RPA) values of the training samples, respectively, $\overline{y}^{obs}$ is the average value of log(RPA) from the experiment for the training set, y_i^{obs} and $y_{i^{ext}}^{pred}$ in eqs 3 and 4 are the experimental and predicted log(RPA) values of the testing samples, respectively, and $\overline{y}^{obs}$ is the average value of log(RPA) from the experiment for the test set.

5. EXPERIMENTAL SECTION

5.1. Materials. Trifluoroacetic anhydride, 16-mercaptohexadecanoic acid, 3-methoxypropanamide, acryloyl chloride, triethylamine, sodium bicarbonate, cyanoacetamide, and 3-amino-propionitrile were purchased from Sigma-Aldrich. *N*-methyl-2-pyrrolidone (NMP, anhydrous), 3-amino-*N*-methylpropanamide hydrochloride, *N*-(2-aminoethyl)acrylamide, and 1-amino-2-methoxyethane were obtained from VWR. All other chemicals introduced in this project were used without prepurification.

5.2. Preparation of Targeted SAMs. The preparation protocol of SAMs was followed as the previous work reported by Prof. Whitesides's group.[49,68] In brief, the SPR chips were cut into 1.5 cm × 2 cm pieces, washed with absolute ethanol, and exposed into a 3 nM 16-mercaptohexadecanoic acid solution (solvent: ethanol) at room temperature overnight. The modified SPR chips were named "SPR-1". Subsequently, precleaned SPR-1 substrates colonized with carboxylic acid were immersed into a 10 mL solution of 0.2 M triethylamine and 0.1 M trifluoroacetic anhydride in anhydrous dimethylformamide (DMF) for at least 30 min. The obtained chips were referred to as "SPR-2" and rinsed thoroughly with CH_2Cl_2. After undergoing the abovementioned two steps, the grafted interchain anhydride was generated. The precleaned SPR-2 substrates were immersed into a 10 mL solution containing 10 mM alkylamine (solvent: anhydrous 1-methyl-2-pyrrolidinone) for 1 h. The resultant chips referred to as SAM-1−3 chips (cyanoacetamide 3-amino-*N*-methylpropanamide

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

hydrochloride) were rinsed with ethanol and dried with a nitrogen stream for standby.

5.3. Synthesis of the Predicted Monomers. According to the previous studies, two predicted yet uncommercial monomers, i.e., *N*-(2-cyanoethyl)acrylamide (CEAA) and *N*-(2-methoxyethyl)-acrylamide (MEAA) were synthesized by coupling acryloyl chloride and corresponding amine derivatives (Figure S1). For example, acryloyl chloride (0.04 mol) in 100 mL of anhydrous CH_2Cl_2 was added dropwise to the mixture containing anhydrous CH_2Cl_2 (100 mL), triethylamine (20 mL), and equimolar 1-amino-2-methoxy-ethane (0.04 mol) at ice−salt baths. After the feeding was completed, the substitution reaction stopped after additional 6 h at 25 °C. The unreacted acryloyl chloride was consumed by adding 100 mL of pure water. In addition, the oil/water layers were washed with a $NaHCO_3$ solution and water twice, separated using a separatory funnel, and fully dried by sodium sulfate (anhydrous) overnight. The organic layer was then concentrated and evaporated to obtain the transparent liquid products (yield: 67, 65, and 74%, respectively). ^{1}H-NMR (DMSO-d_6 or $CDCl_3$, 300 MHz): (CEAA: $-CH_2-$, 3.04−3.67, 4H; $-CH_3$, 3.28, 3H; $-CH_2$ = CH−, 5.74−6.48, 3H; −NH−, 8.41, 1H), (MEAA: $-CH_2-$, 3.17−3.33, 4H; $-CH_2$ = CH−, 5.74−6.52, 3H; −NH−, 8.42, 1H).

5.4. Preparation of Polymer Brushes on SPR Chips. The typical protocol of polymer brushes on SPR chips was according to our previous work via a surface-initiated (SI) ATRP method.[65,66] The precleaned SPR chips were first immersed into a solution containing 1 mM ω-mercaptoundecyl bromoisobutyrate (solvent: ethanol) at 25 °C overnight. Next, a tube containing the target monomer (0.5 g), Me_6TREN (30 mg), and degassed methanol were transferred to another tube with SPR chips grafted with immobilized initiators and CuBr (15 mg) under the protection of high-purity nitrogen. The SI-ATRP reaction was immediately initiated and was stopped by exposing to air after 12 h. The dissociative polymer chains and unreacted monomers were removed by soaking chips into a PBS solution overnight.

5.5. Protein Adsorption by SPR Measurement. The detailed protocol was followed as our previous reported literature.[65] Generally, a homemade SPR sensor based on wavelength interrogation was utilized to determine protein adsorption on target SAM- or polymer brush-grafted substrates. The undiluted human protein (plasma and serum) solution and PBS buffer flowed through four channels under the pressure of a peristaltic pump. To obtain the absolute adsorption value, a 1 nm SPR wavelength shift was regarded as ∼15 ng/cm^2 protein adsorption.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.1c00642.

Three tables of loadings and communalities for the 10 rotated functional groups; regression coefficients of molecular descriptors; summary of the percentage of variance; and one figure of the synthesis procedure of two monomers of CEAA and MEAA (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jie Zheng − *Department of Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Ohio 44325, United States;* orcid.org/0000-0003-1547-3612; Email: zhengj@uakron.edu

Authors

Yonglan Liu − *Department of Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Ohio 44325, United States*

Dong Zhang − *Department of Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Ohio 44325, United States;* orcid.org/0000-0001-7002-7661

Yijing Tang − *Department of Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Ohio 44325, United States*

Yanxian Zhang − *Department of Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Ohio 44325, United States*

Yung Chang − *Department of Chemical Engineering, R&D Center for Membrane Technology, Chung Yuan Christian University, Taoyuan 32023, Taiwan;* orcid.org/0000-0003-1419-4478

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.1c00642

Author Contributions

§Y.L. and D.Z. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.Z. is thankful for financial supports from NSF grants (1806138 and 1825122)

■ REFERENCES

(1) Wang, A. Y.-T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brgoch, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chem. Mater.* **2020**, *32*, 4954−4965.

(2) Ekins, S.; Puhl, A. C.; Zorn, K. M.; Lane, T. R.; Russo, D. P.; Klein, J. J.; Hickey, A. J.; Clark, A. M. Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **2019**, *18*, 435−441.

(3) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discovery* **2019**, *18*, 463−477.

(4) Kavakiotis, I.; Tsave, O.; Salifoglou, A.; Maglaveras, N.; Vlahavas, I.; Chouvarda, I. Machine Learning and Data Mining Methods in Diabetes Research. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 104−116.

(5) Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable Sustainable Energy Rev.* **2018**, *82*, 1027−1047.

(6) Galindo, J.; Tamayo, P. Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. *Comput. Econ.* **2000**, *15*, 107−143.

(7) Cole, J. M. A Design-to-Device Pipeline for Data-Driven Materials Discovery. *Acc. Chem. Res.* **2020**, *53*, 599−610.

(8) Sendek, A. D.; Cheon, G.; Pasta, M.; Reed, E. J. Quantifying the Search for Solid Li-Ion Electrolyte Materials by Anion: A Data-Driven Perspective. *J. Phys. Chem. C* **2020**, *124*, 8067−8079.

(9) Fujimura, K.; Seko, A.; Koyama, Y.; Kuwabara, A.; Kishida, I.; Shitara, K.; Fisher, C. A. J.; Moriwake, H.; Tanaka, I. Accelerated Materials Design of Lithium Superionic Conductors Based on First-Principles Calculations and Machine Learning Algorithms. *Adv. Energy Mater.* **2013**, *3*, 980−985.

(10) Kitchin, J. R. Machine learning in catalysis. *Nat. Catal.* **2018**, *1*, 230−232.

(11) Sun, B.; Fernandez, M.; Barnard, A. S. Machine learning for silver nanoparticle electron transfer property prediction. *J. Chem. Inf. Model.* **2017**, *57*, 2413−2423.

(12) Kunkel, C.; Schober, C.; Margraf, J. T.; Reuter, K.; Oberhofer, H. Finding the right bricks for molecular legos: A data mining

approach to organic semiconductor design. *Chem. Mater.* **2019**, *31*, 969−978.
(13) Khakifirooz, M.; Chien, C. F.; Chen, Y.-J. Bayesian inference for mining semiconductor manufacturing big data for yield enhancement and smart production to empower industry 4.0. *Appl. Soft Comput.* **2018**, *68*, 990−999.
(14) Chen, L.; Tran, H.; Batra, R.; Kim, C.; Ramprasad, R. Machine learning models for the lattice thermal conductivity prediction of inorganic materials. *Comput. Mater. Sci.* **2019**, *170*, No. 109155.
(15) Araya, S. N.; Ghezzehei, T. A. Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. *Water Resour. Res.* **2019**, *55*, 5715−5737.
(16) Yang, W.; Fidelis, T. T.; Sun, W.-H. Machine Learning in Catalysis, From Proposal to Practicing. *ACS Omega* **2020**, *5*, 83−88.
(17) Brown, K. A.; Brittman, S.; Maccaferri, N.; Jariwala, D.; Celano, U. Machine Learning in Nanoscience: Big Data at Small Scales. *Nano Lett.* **2020**, *20*, 2−10.
(18) Lai, S.; Zhao, M.; Qiao, J.; Molokeev, M. S.; Xia, Z. Data-driven photoluminescence tuning in Eu2+-doped phosphors. *J. Phys. Chem. Lett.* **2020**, *11*, 5680−5685.
(19) Stanev, V.; Oses, C.; Kusne, A. G.; Rodriguez, E.; Paglione, J.; Curtarolo, S.; Takeuchi, I. Machine learning modeling of superconducting critical temperature. *npj Comput. Mater.* **2018**, *4*, 1−14.
(20) Babanezhad, M.; Nakhjiri, A. T.; Marjani, A.; Rezakazemi, M.; Shirazian, S. Evaluation of product of two sigmoidal membership functions (psigmf) as an ANFIS membership function for prediction of nanofluid temperature. *Sci. Rep.* **2020**, *10*, No. 22337.
(21) Zhang, H.; Hippalgaonkar, K.; Buonassisi, T.; Løvvik, O. M.; Sagvolden, E.; Ding, D. Machine learning for novel thermal-materials discovery: early successes, opportunities, and challenges. *ES Energy Environ.* **2019**, *2*, 1−8.
(22) Wang, C.; Fu, H.; Jiang, L.; Xue, D.; Xie, J. A property-oriented design strategy for high performance copper alloys via machine learning. *npj Comput. Mater.* **2019**, *5*, No. 87.
(23) Hu, Y.-J.; Zhao, G.; Zhang, M.; Bin, B.; Del Rose, T.; Zhao, Q.; Zu, Q.; Chen, Y.; Sun, X.; de Jong, M.; et al. Predicting densities and elastic moduli of SiO 2-based glasses by machine learning. *npj Comput. Mater.* **2020**, *6*, No. 22.
(24) Zhao, H.; Ezeh, C. I.; Ren, W.; Li, W.; Pang, C. H.; Zheng, C.; Gao, X.; Wu, T. Integration of machine learning approaches for accelerated discovery of transition-metal dichalcogenides as Hg0 sensing materials. *Appl. Energy* **2019**, *254*, No. 113651.
(25) Toyao, T.; Suzuki, K.; Kikuchi, S.; Takakusagi, S.; Shimizu, K.-i.; Takigawa, I. Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys. *J. Phys. Chem. C* **2018**, *122*, 8315−8326.
(26) Wu, L.; Xiao, Y.; Ghosh, M.; Zhou, Q.; Hao, Q. Machine Learning Prediction for Bandgaps of Inorganic Materials. *ES Mater. Manuf.* **2020**, *9*, 34−39.
(27) Kaufmann, K.; Maryanovsky, D.; Mellor, W. M.; Zhu, C.; Rosengarten, A. S.; Harrington, T. J.; Oses, C.; Toher, C.; Curtarolo, S.; Vecchio, K. S. Discovery of high-entropy ceramics via machine learning. *npj Comput. Mater.* **2020**, *6*, No. 42.
(28) Moliner, M.; Román-Leshkov, Y.; Corma, A. Machine Learning Applied to Zeolite Synthesis: The Missing Link for Realizing High-Throughput Discovery. *Acc. Chem. Res.* **2019**, *52*, 2971−2980.
(29) Lin, S.; Wang, Y.; Zhao, Y.; Pericchi, L. R.; Hernández-Maldonado, A. J.; Chen, Z. Machine-learning-assisted screening of pure-silica zeolites for effective removal of linear siloxanes and derivatives. *J. Mater. Chem. A* **2020**, *8*, 3228−3237.
(30) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19*, 640−645.
(31) He, Y.; Cubuk, E. D.; Allendorf, M. D.; Reed, E. J. Metallic metal−organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *J. Phys. Chem. Lett.* **2018**, *9*, 4562−4569.
(32) Dalsin, J. L.; Messersmith, P. B. Bioinspired antifouling polymers. *Mater. Today* **2005**, *8*, 38−46.
(33) Banerjee, I.; Pangule, R. C.; Kane, R. S. Antifouling coatings: recent developments in the design of surfaces that prevent fouling by proteins, bacteria, and marine organisms. *Adv. Mater.* **2011**, *23*, 690−718.
(34) Liu, Y.; Zhang, D.; Ren, B.; Gong, X.; Liu, A.; Chang, Y.; He, Y.; Zheng, J. Computational Investigation of Antifouling Property of Polyacrylamide Brushes. *Langmuir* **2020**, *36*, 2757−2766.
(35) Liu, Y.; Zhang, D.; Ren, B.; Gong, X.; Xu, L.; Feng, Z.-Q.; Chang, Y.; He, Y.; Zheng, J. Molecular simulations and understanding of antifouling zwitterionic polymer brushes. *J. Mater. Chem. B* **2020**, *8*, 3814−3828.
(36) Cheung, D. L.; Lau, K. H. A. Atomistic Study of Zwitterionic Peptoid Antifouling Brushes. *Langmuir* **2019**, *35*, 1483−1494.
(37) Liu, Z.-Y.; Jiang, Q.; Jin, Z.; Sun, Z.; Ma, W.; Wang, Y. Understanding the Antifouling Mechanism of Zwitterionic Monomer-Grafted Polyvinylidene Difluoride Membranes: A Comparative Experimental and Molecular Dynamics Simulation Study. *ACS Appl. Mater. Interfaces* **2019**, *11*, 14408−14417.
(38) Zheng, J.; Li, L.; Tsao, H.-K.; Sheng, Y.-J.; Chen, S.; Jiang, S. Strong repulsive forces between protein and oligo (ethylene glycol) self-assembled monolayers: a molecular simulation study. *Biophys. J.* **2005**, *89*, 158−166.
(39) Yang, F.; Liu, Y.; Zhang, Y.; Ren, B.; Xu, J.; Zheng, J. Synthesis and characterization of ultralow fouling poly (n-acryloyl-glycinamide) brushes. *Langmuir* **2017**, *33*, 13964−13972.
(40) He, Y.; Chang, Y.; Hower, J. C.; Zheng, J.; Chen, S.; Jiang, S. Origin of repulsive force and structure/dynamics of interfacial water in OEG−protein interactions: a molecular simulation study. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5539−5544.
(41) Chen, S.; Zheng, J.; Li, L.; Jiang, S. Strong Resistance of Phosphorylcholine Self-Assembled Monolayers to Protein Adsorption: Insights into Nonfouling Properties of Zwitterionic Materials. *J. Am. Chem. Soc.* **2005**, *127*, 14473−14478.
(42) Xiang, Y.; Xu, R. G.; Leng, Y. S. Molecular Simulations of the Hydration Behavior of a Zwitterion Brush Array and Its Antifouling Property in an Aqueous Environment. *Langmuir* **2018**, *34*, 2245−2257.
(43) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49*, 3525−3564.
(44) Almeida, J. R.; Moreira, J.; Pereira, D.; Pereira, S.; Antunes, J.; Palmeira, A.; Vasconcelos, V.; Pinto, M.; Correia-da-Silva, M.; Cidade, H. Potential of synthetic chalcone derivatives to prevent marine biofouling. *Sci. Total Environ.* **2018**, *643*, 98−106.
(45) Feng, K.; Li, X.; Yu, L. Synthesis, antibacterial activity, and application in the antifouling marine coatings of novel acylamino compounds containing gramine groups. *Prog. Org. Coat.* **2018**, *118*, 141−147.
(46) Rasulev, B.; Jabeen, F.; Stafslien, S.; Chisholm, B. J.; Bahr, J.; Ossowski, M.; Boudjouk, P. Polymer Coating Materials and Their Fouling Release Activity: A Cheminformatics Approach to Predict Properties. *ACS Appl. Mater. Interfaces* **2017**, *9*, 1781−1792.
(47) 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*, No. 265.
(48) 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*, 4949−4956.
(49) Ostuni, E.; Chapman, R. G.; Holmlin, R. E.; Takayama, S.; Whitesides, G. M. A survey of structure-property relationships of surfaces that resist the adsorption of protein. *Langmuir* **2001**, *17*, 5605−5620.

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319

(50) Yang, J. C.; Zhao, C.; Hsieh, I. F.; Subramanian, S.; Liu, L. Y.; Cheng, G.; Li, L. Y.; Cheng, S. Z. D.; Zheng, J. Strong resistance of poly (ethylene glycol) based L-tyrosine polyurethanes to protein adsorption and cell adhesion. *Polym. Int.* **2012**, *61*, 616−621.
(51) Chen, S. F.; Li, L. Y.; Zhao, C.; Zheng, J. Surface hydration: Principles and applications toward low-fouling/nonfouling biomaterials. *Polymer* **2010**, *51*, 5283−5293.
(52) Wu, J.; Chen, S. F. Investigation of the Hydration of Nonfouling Material Poly(ethylene glycol) by Low-Field Nuclear Magnetic Resonance. *Langmuir* **2012**, *28*, 2137−2144.
(53) Meyers, S. R.; Grinstaff, M. W. Biocompatible and Bioactive Surface Modifications for Prolonged In Vivo Efficacy. *Chem. Rev.* **2012**, *112*, 1615−1632.
(54) Ladd, J.; Zhang, Z.; Chen, S.; Hower, J. C.; Jiang, S. Zwitterionic polymers exhibiting high resistance to nonspecific protein adsorption from human serum and plasma. *Biomacromolecules* **2008**, *9*, 1357−1361.
(55) Cheng, G.; Zhang, Z.; Chen, S. F.; Bryers, J. D.; Jiang, S. Y. Inhibition of bacterial adhesion and biofilm formation on zwitterionic surfaces. *Biomaterials* **2007**, *28*, 4192−4199.
(56) Yang, J. T.; Chen, H.; Xiao, S. W.; Shen, M. X.; Chen, F.; Fan, P.; Zhong, M. Q.; Zheng, J. Salt-Responsive Zwitterionic Polymer Brushes with Tunable Friction and Antifouling Properties. *Langmuir* **2015**, *31*, 9125−9133.
(57) Sahigara, F.; Ballabio, D.; Todeschini, R.; Consonni, V. Defining a novel k-nearest neighbours approach to assess the applicability domain of a QSAR model for reliable predictions. *J. Cheminf.* **2013**, *5*, No. 27.
(58) Chen, S.; Li, L.; Zhao, C.; Zheng, J. Surface hydration: Principles and applications toward low-fouling/nonfouling biomaterials. *Polymer* **2010**, *51*, 5283−5293.
(59) Zhao, C.; Li, L.; Guo, M.; Zheng, J. Functional polymer thin films designed for antifouling materials and biosensors. *Chem. Pap.* **2012**, *66*, 323−339.
(60) Chen, H.; Zhao, C.; Zhang, M.; Chen, Q.; Ma, J.; Zheng, J. Molecular understanding and structural-based design of polyacrylamides and polyacrylates as antifouling materials. *Langmuir* **2016**, *32*, 3315−3330.
(61) Zhang, Y.; Liu, Y.; Ren, B.; Zhang, D.; Xie, S.; Chang, Y.; Yang, J.; Wu, J.; Xu, L.; Zheng, J. Fundamentals and applications of zwitterionic antifouling polymers. *J. Phys. D: Appl. Phys.* **2019**, *52*, No. 403001.
(62) Yang, J.; Zhang, M.; Chen, H.; Chang, Y.; Chen, Z.; Zheng, J. Probing the structural dependence of carbon space lengths of poly(n-hydroxyalkyl acrylamide)-based brushes on antifouling performance. *Biomacromolecules* **2014**, *15*, 2982−2991.
(63) Liu, Y.; Zhang, Y.; Ren, B.; Sun, Y.; He, Y.; Cheng, F.; Xu, J.; Zheng, J. Molecular dynamics simulation of the effect of carbon space lengths on the antifouling properties of hydroxyalkyl acrylamides. *Langmuir* **2019**, *35*, 3576−3584.
(64) Herrwerth, S.; Eck, W.; Reinhardt, S.; Grunze, M. Factors that Determine the Protein Resistance of Oligoether Self-Assembled Monolayers− Internal Hydrophilicity, Terminal Hydrophilicity, and Lateral Packing Density. *J. Am. Chem. Soc.* **2003**, *125*, 9359−9366.
(65) Zhang, D.; Ren, B.; Zhang, Y.; Liu, Y.; Chen, H.; Xiao, S.; Chang, Y.; Yang, J.; Zheng, J. Micro- and macroscopically structured zwitterionic polymers with ultralow fouling property. *J. Colloid Interface Sci.* **2020**, *578*, 242−253.
(66) Chen, H.; Yang, J.; Xiao, S.; Hu, R.; Bhaway, S. M.; Vogt, B. D.; Zhang, M.; Chen, Q.; Ma, J.; Chang, Y.; Li, L.; Zheng, J. Salt-responsive polyzwitterionic materials for surface regeneration between switchable fouling and antifouling properties. *Acta Biomater.* **2016**, *40*, 62−69.
(67) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall, Inc, 1988.
(68) Yan, L.; Marzolin, C.; Terfort, A.; Whitesides, G. M. Formation and reaction of interchain carboxylic anhydride groups on self-assembled monolayers on gold. *Langmuir* **1997**, *13*, 6704−6412.

https://dx.doi.org/10.1021/acsami.1c00642
ACS Appl. Mater. Interfaces 2021, 13, 11306−11319