Chemical Engineering Journal 420 (2021) 129872

Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

Machine Learning-Enabled Repurposing and Design of Antifouling Polymer Brushes

Yonglan Liu [a,1], Dong Zhang [a,1], Yijing Tang [a], Yanxian Zhang [a], Xiong Gong [b], Shaowen Xie [c], Jie Zheng [a,*]

[a] Department of Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, OH 44325, USA
[b] Department of Polymer Engineering, The University of Akron, OH 44325, USA
[c] Hunan Key Laboratory of Biomedical Nanomaterials and Devices, College of Life Sciences and Chemistry, Hunan University of Technology, Zhuzhou 412007, China

ARTICLE INFO

Keywords:
Machine learning
Antifouling
Polymer brush
Protein adsorption
Artificial neural network
Supporting vector regression

ABSTRACT

Rational development of antifouling materials is of great importance for fundamental research and real-world applications. However, current experimental designs and computational modelings of antifouling materials still retain empirical flavor due to the data complexity of polymers and their associated structures/properties. In this work, we developed a data-driven, machine learning workflow, in combination with an in-house benchmark dataset of antifouling polymer brushes, to discover the potential antifouling property of existing polymer brushes using the descriptor-based artificial neural network (ANN) model and design the new antifouling polymer brushes using the group-based supporting vector regression (SVR) model. The resultant two machine learning models not only demonstrated their reliability, predictivity, and applicability, but also established the composition-structure–property relationships using both descriptors and functional groups. Finally, we synthesized different repurposed and newly designed polymer brushes, as predicted by ANN and SVR models, all of which exhibited excellent surface resistance to protein adsorption from undiluted human blood serum and plasma at optimal film thicknesses. Overall, our data-driven machine learning models can be used as an intelligent tool for determining, repurposing, and designing new superior antifouling materials beyond polymer brushes.

1. Introduction

Driven mostly by the Materials Genome Initiative, data-driven machine learning approaches have recently expanded their uses from traditional research domains (i.e., drug discovery [1–3], image recognition [4,5], disease diagnose [6,7], transportation analysis [8,9]) to materials informatics [10], because they have demonstrated the powerful data-mining ability to (i) extract the statistically significant composition/structure–property relationships from large materials datasets and (ii) accelerate materials design, discovery, and repurpose. These machine learning approaches allow to rapidly predict the materials properties of interest or design new materials with desirable properties, based purely on existing data rather than by "try-and-error" experimentation or computations/simulations. Further, deep learning from the reliable data sources can derive the previously undiscovered correlations between material properties in a qualitative or quantitative way. Most of machine learning studies in materials science have focused on inorganic or solid materials [11] (e.g., metals [12,13], ceramics [14,15], zeolites [16,17], metal–organic frameworks [18,19], nanoparticles [20,21], and semiconductors [22,23]) to evaluate and predict their specific properties (electronic/ionic conductivity [24,25], catalytic activity [26,27], photoluminescence [28,29], thermodynamics [30,31], and mechanical properties [32,33]). These studies have demonstrated their promising impacts on the acceleration of new materials design, the reduction of cost, time, efforts, and failure risks, and the improved understanding of the composition-structure–property relationship of a given class of materials. On the other hand, different from inorganic or solid materials with well-defined structures and reliable datasets, machine learning for soft materials with desirable functions and properties is still at infant stage. It still remains a great challenge to use machine learning to establish qualitative or quantitative rules for rationally assessing and designing functional soft materials (e.g. polymers [10], elastomers [34], and hydrogels [35]), largely because these soft materials have very diverse and flexible structures, making them difficult to

* Corresponding author.
E-mail address: zhengj@uakron.edu (J. Zheng).
[1] The authors contribute equally to this work.

https://doi.org/10.1016/j.cej.2021.129872
Received 11 February 2021; Received in revised form 3 April 2021; Accepted 12 April 2021
Available online 18 April 2021

generate consistent results for a given property. Thus, establishing a reliable dataset containing high-quality data for both structures and properties of soft materials is a first and important step for developing advanced machine learning approaches for materials design and assessment.

Antifouling materials/coatings are critical for many model systems and practical applications, including implanted devices [36], marine coatings/hulls [37], biosensors [38], membrane separation [39], and drug/gene delivery [40,41]. A wide variety of antifouling materials/coatings have been developed to resist the unwanted adsorption of proteins, cells, bacteria, and organisms to some extents. The resultant antifouling efficiency of a given surface is strongly depended on not only physicochemical characteristics of materials itself (e.g. carbon spacer lengths [42], molecular moieties [43], hydrophilic/hydrophobic ratio [44], electrostatic repulsion [44]), but also surface characteristics (e.g. grafting density[43,] [45], film thickness and roughness [46,47], chain conformation [48]). However, current design of antifouling coatings still retains empirical flavor and is driven by chemical/materials intuition, trial-and-error, and similar materials compatibility. Alternative to empirical experiments, while molecular simulations can provide atomic details of structural, dynamics, and free energy properties for the design and evaluation of antifouling materials/coatings [43,49–52], the "one-at-a-time" and computational expensive nature of molecular simulations only allows to study individual antifouling materials system, thus lacking a data-driven capacity for rapid prediction and design of a large amount of antifouling materials from a given database.

To date, very few machine learning studies of antifouling materials have been reported. Among them, most of them simply applied the multiple linear regression (MLR) method to establish a qualitative relationship between several selected physicochemical properties and antifouling results, e.g., between hydrogen bonds, branching ratios, and constitutional diversity of chalcone derivatives and antifouling activity against various micro- and macro-fouling species [53], between HOMO state, L-H energy, and molar refractivity of acylamino compounds and their antibacterial activity of *E. coli* [54], and between the polarizability indices, mass, and van der Waals volume of polysiloxane-based coatings and the fouling release activity of fouling organisms (bacteria, algae, and barnacles) [55]. Since these studies only used a single antifouling polymer system with some experimental variations to generate the very limited structure–property data, they were lack of a data-driven ability to derive a general structure–property prediction and assessment for other antifouling polymers. Until recently, three machine learning studies including ours have reported to use the same dataset of self-assembled monolayers (SAMs) from the Whitesides group [56] to construct different QSAR models of (i) descriptor-based linear (MLREM) and non-linear (BRANNGP and BRANNLP) models [57], (ii) descriptor-based artificial neural network (ANN) models [58], and (iii) functional group-based ANN models for quantitatively evaluating protein adsorption on both existing and newly designed SAMs [59]. Due to the simple synthesis, well-controlled surface structures, and a large diversity of surface chemistry of SAMs, these studies have demonstrated SAMs as an ideal model platform, enabling to collect reliable protein adsorption/resistance data in a uniform and controlled manner and establish different robust machine learning models for the design and assessment purposes.

Surprisingly, there is a remarkably lack of machine learning studies of antifouling polymeric coatings, not even mention to the experimental verification of computationally designed antifouling materials. Some major barriers still exist: (i) there are no any existing antifouling polymer databases available at least in the public domain, thus it requires the researchers to collect experimental results from literature to construct antifouling polymer databases; (ii) due to the structural complexity/diversity of polymers and their coatings, data inconsistence is additional concern, and even the same antifouling polymers could perform differently at different laboratories due to experimental conditions, chemical/materials purity, and different synthesis routes. To overcome these barriers, we are the first to construct an antifouling polymer brush dataset from literatures, then use this dataset to develop two distinct machine learning approaches for discovering the antifouling potentials of any existing polymer brush by the descriptor-based artificial neural network (ANN) model and designing the new antifouling polymer brushes by the group-based supporting vector regression (SVR) model, and finally conduct polymer synthesis and brush coatings to validate the antifouling performance of both repurposed and newly designed polymer brushes using protein assays. As a result, both models provide the sequence-structure–property relationship of antifouling polymer brushes at gross and fragmental levels. As guided by the both machine learning models, we synthesized three repurposed polymer brushes and three newly designed polymer brushes, all of which exhibited excellent surface resistance to protein adsorption from undiluted human blood serum and plasma, highly consistent with the predicted values from the models. This proof-of-concept study provides the starting point, from dataset construction of antifouling polymer brushes to development of machine learning models, for advancing the discovery of new antifouling materials and coatings by the extensive screening of the spatial, compositional, and interaction spaces of polymers.

2. Results and discussion

Overall, Fig. 1 outlines a four-step analytic workflow of this machine-learning study. (1) First, we started with data screening and collection of antifouling polymer brushes from literatures by carefully removing any redundant and inconsistent data (Fig. 1**a**). This is the first dataset for antifouling polymer brushes, which will be used as a reliable source to develop a machine-learning models for the structural-based screen and design of next-generation antifouling materials. (2) The dataset containing 28 polymer brushes with a wide variety of chemical structures allows to extensively explore the spatial, compositional, and interaction spaces of both polymers themselves and their coatings, in which molecular descriptors will be generated and identified to present component and structure features of antifouling polymer brushes (Fig. 1**b**). (3) To maximize the great potentials of dataset, we developed two different machine-learning models of (i) ANN model with gross-level, property-based descriptors for repurposing/discovering the existing polymer brushes with antifouling properties and (ii) SVR model with fragmental-level, group-based descriptors for designing new antifouling polymer brushes (Fig. 1**c**). The resulting models were able to determine the important molecular descriptors and functional groups critical for surface resistance to proteins. (4) Finally, as guided by model prediction, we synthesized four different polymer brushes for assessing their antifouling property against protein adsorption from undiluted human blood serum and plasma using surface plasma resonance (SPR) (Fig. 1**d**). Among them, the three brushes were selected from the existing literature (not from the dataset) for discovering their new antifouling property, while the other two brushes were newly designed with the predicted antifouling property. SPR results showed that these predicted polymer brushes exhibited excellent surface resistance to protein adsorption as evidenced by a small amount of adsorbed proteins (0.0–9.0 ng/cm^2) on the polymer brushes, confirming the model predictions. This work demonstrated data-driven, machine-learning models, built on a reliable dataset of antifouling polymer brushes, for screening, discovering, and designing antifouling materials/coatings and beyond.

2.1. Dataset construction of antifouling polymer brushes from literature

To construct a reliable benchmarking dataset, we selected two different types of zwitterionic and hydrophilic polymer brushes, all of which demonstrated their antifouling property against protein adsorption from undiluted human serum or plasma, rather than from single or diluted protein solutions. Surprisingly, a very limited number of antifouling polymer brushes have been reported after challenging with

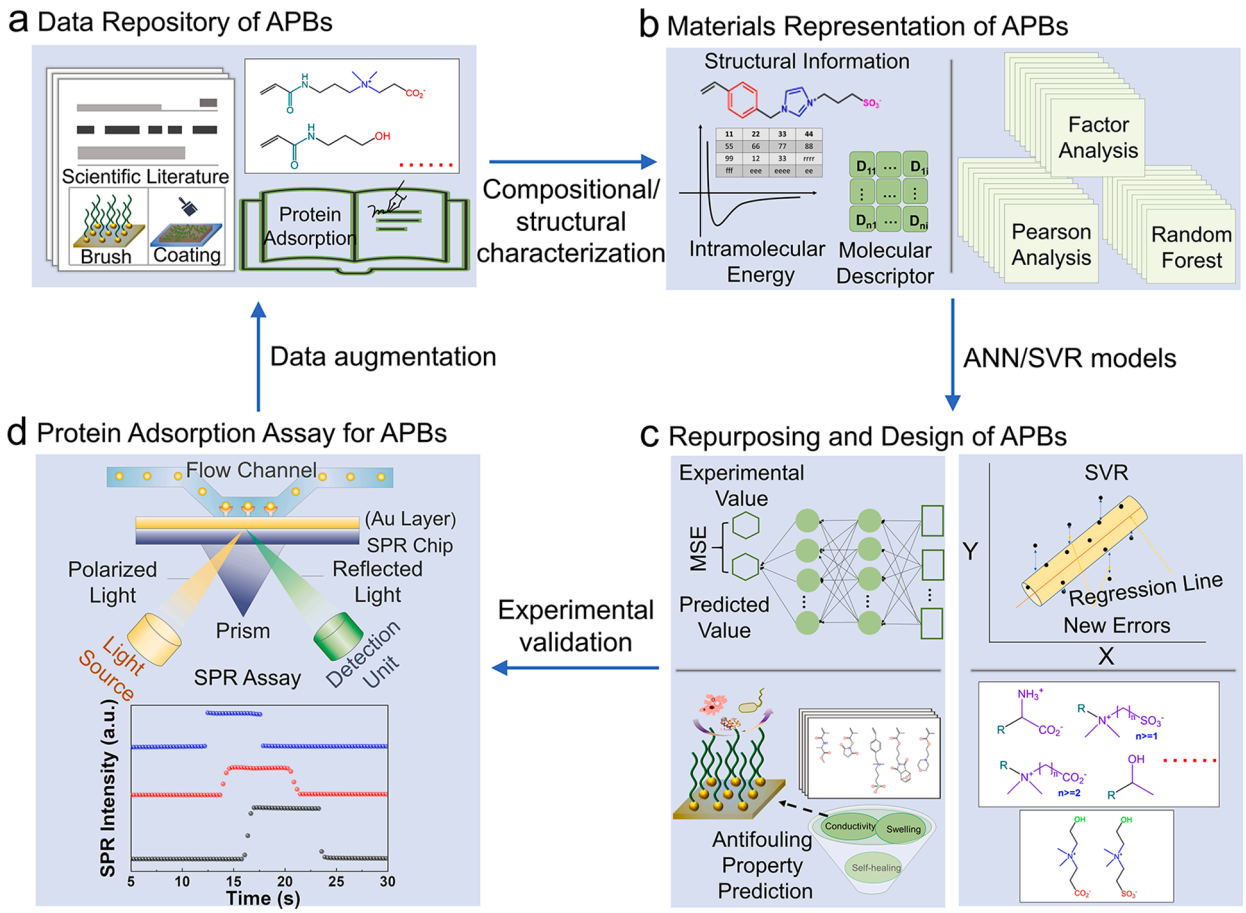


Fig. 1. Schematic workflow of four-step machine-learning models for the design and repurposing of antifouling polymer brushes. (a) Construction of data repository of antifouling polymer brushes (APBs) from literatures, (b) Compositional and structural representations of antifouling polymer brushes, (c) Development of artificial neural network (ANN) and supporting vector regression (SVR) models for repurposing and designing antifouling polymer brushes, and (d) experimental validation of computationally screened/designed antifouling polymer brushes by surface plasma resonance (SPR).

undiluted human serum or plasma. As a result, the in-house dataset contained 14 zwitterionic and 14 hydrophilic polymer brushes, whose monomer structures were summarized in Fig. 2. It can be seen that both zwitterionic and hydrophilic monomers possessed a wide combination of functional groups of acrylates, acrylamides, carboxylates, sulfonates, betaines, amino acids, hydroxyls, amides, olefins, peptoids, aromatics,

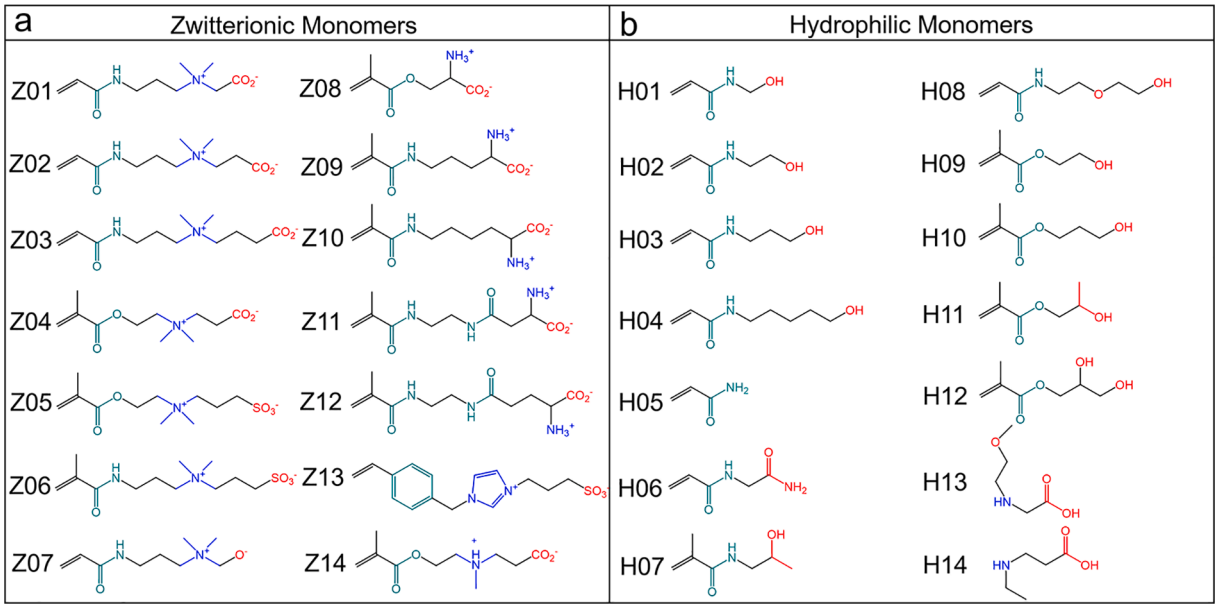


Fig. 2. A summary of monomer structures used for (a) 14 zwitterionic and (b) 14 hydrophilic polymer brushes collected from literature.

carbon space lengths, and a common double bond at the one end for polymerization, as well as different chain architectures from linear to branched structures. For each polymer brush, protein adsorption amounts (ng/cm^2) from undiluted human plasma or serum at different film thicknesses were also collected, leading to a total of 94 protein adsorption data points (Table S1), in which the brush thickness effect on protein adsorption was also considered as one of descriptors in machine learning models. Next, these 94 data points were used to construct gross-level, descriptor-based ANN models for discovering and repurposing the new antifouling property of existing polymer brushes, while 91 data points by removing the entry of H13-01, H14-01, and H14-02 were used to build fragmental-level, group-based SVR models for designing new antifouling polymer brushes.

2.2. Descriptor-Based ANN models for repurposing the new antifouling property of existing polymer brushes

For molecular descriptor-based ANN model, structural and compositional representations of antifouling polymer brushes were encoded with 104 molecular descriptors (constitutional, topological, connectivity indices, geometrical indices, functional groups, etc), 3 energy descriptors (charge-charge, charge-dipole, and dipole–dipole energies), and film thickness, leading to a total of 108 descriptors. To avoid the overfitting of the ANN model, we conducted the two sequential approaches to reduce the dimension of highly correlated descriptors. First, we applied Pearson analysis to quantify the independence between descriptors, where any descriptor with paired correlation coefficient of $>$ 0.6 was removed, finally reducing to 14 descriptors (film thickness, charge-charge energy, charge-dipole energy, Mi, RBF, O%, nR = Cs, nRCOOH, nRCONH2, nOHs, C-005, C-007, H-051, and ALOGP) with relatively high independence (Fig. 3**a**). Next, Random Forest was carried out to determine the importance of the resultant 14 descriptors for their contributions to protein adsorption ability on polymer brushes. Fig. 3**b** showed that only 8 out of 14 descriptors (film thickness, Mi, ALOGP, RBF, O%, charge-charge energy, C-005, and C-007) had an important score of $>$ 0.01 with most-contributing features. The resultant 8 descriptors were detailed in Table 1.

With the resultant 8 descriptors, we designed a five-layer artificial neural networks to train logarithmic protein adsorption amount (log (PAA)) as a function of the 8 descriptors, starting with the first input layer containing 8 descriptors, followed by the second, third, and fourth layers containing respective 64, 96, and 64 neurons, ending with an objective function of log(PAA) on polymer brushes. During the training procedure, Z03-01 and H07-01 (Table S1) were identified as outliers and thus were removed from the dataset, resulting in a total of 92 data points in the dataset. The dataset was randomly splited into a training set and a test set at the ratio of 7:3, in which training set was used to train the protein adsorption of polymer brushes in relation to the 8 descriptors in a ANN model, while test set was applied to evaluate the predictive ability of the ANN model. Fig. 3**c** shows the training epochs evolution of the convergence of mean absolute error (MAE) of log(PAA) for training set and test set, indicative of training performance of the ANN model. At a first glance, MAE curves for both training and test sets were converged rapidly at very early training of 35 epochs, indicating that the ANN model is well trained with minimal overfitting risks. Further, we statistically counted the distribution of absolute values of prediction errors of both training and test sets. As shown in Fig. 3**d**, all errors of training

Table 1
Description of 8 descriptors obtained from random forest and used in the ANN model.

No.	Descriptors	Detail Description	Type
1	Thickness	Film thickness	Experimental condition
2	Mi	Mean first ionization potential (scaled on Carbon atom)	Constitutional indices
3	ALOGP	Ghose-Crippen octanol–water partition coeff. (logP)	Molecular properties
4	RBF	Rotatable bond fraction	Constitutional indices
5	O%	Percentage of O atoms	Constitutional indices
6	charge-charge	Intramolecular charge-charge energy of the molecule	Calculated energy
7	C-005	CH3X	Atom-centered fragments
8	C-007	CH2X2	Atom-centered fragments

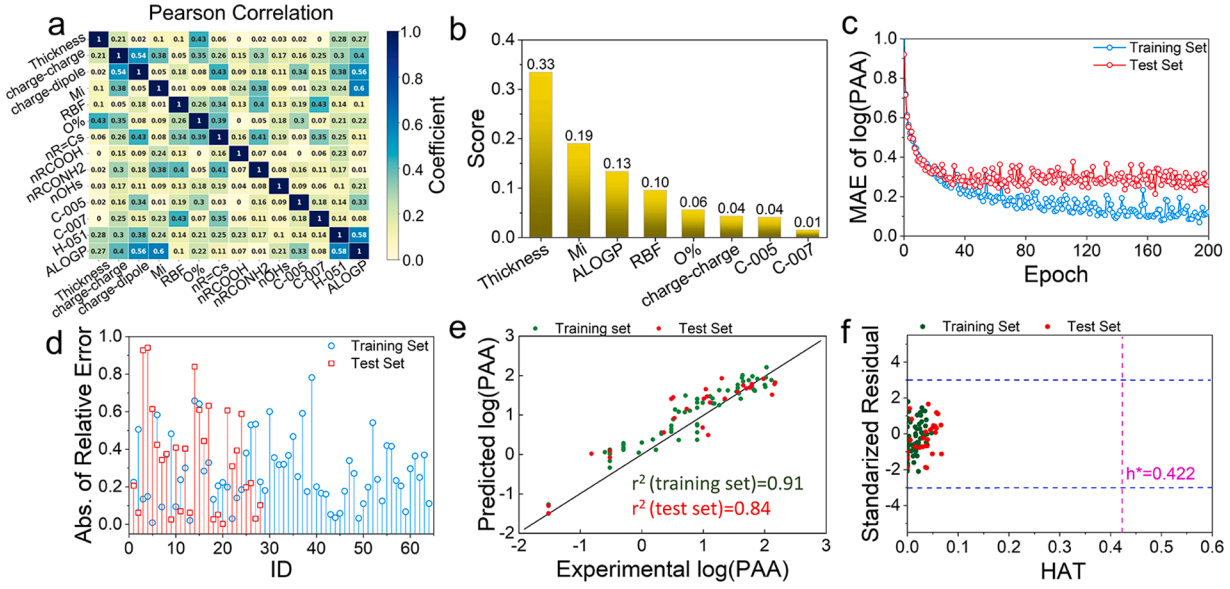


Fig. 3. Compositional and structural representations of antifouling polymer brushes for descriptor-based ANN model. (a) Correlation matrix between any two pair of decriptors from Pearson analysis. (b) Importance score ranking of 8 descriptors from random forest. (c) Training epochs evolution of MAE and (d) relative errors of predicted values with respect to experimental values of log(PAA) for both training and test sets. (e) Correlation between predicted and experimental log (PAA) values for both training and test sets by leave-five-out cross validation and external validation. (f) Applicability domain of training and test sets using William plot with h*=0.422.

and test sets were less than 1.0 and most of predictions errors were smaller than 0.5, confirming the good data fitting and training performance of the ANN model for both training and test sets.

Different approaches were also conducted to validate the ANN model from different angles. Leave-five-out cross validation and externel validation were performed to statistically evaluate the reliability and predictivity of the ANN model. As shown in Table 2, leave-five-out cross validation achieved $Q^2_{CV} = 0.86$ and $RMSE_{CV} = 0.11$ after five iterations on ramdonly selected training set, while external validation of test set produced $Q^2_{ext} = 0.81$ and $RMSE_{ext} = 0.20$, both indicating high reliability and predictivity of the ANN model. Consistently, a plot of predicted log(PAA) versus experimental ones for both training and test sets gave an r^2 of 0.91 and 0.84, respectively (Fig. 3**e**), again confirming the high predictive accuracy of the ANN model. Furthermore, Williams plot was used to measure the applicable domain (AD) of chemical structures for the ANN model in terms of leverage values against the standardized residuals. Fig. 3**f** showed that all data points in training and test sets had a h value smaller than $h^* = 0.422$ and standardized residuals of <3, indicating that (i) all molecules present inside AD of chemical structures and (ii) the predicted values for all molecules can be extrapolated using the ANN model.

To better understand the structure-antifouling property relationship of antifouling polymer brushes, we performed Bayesian statistics to quantify the contribution of 7 descriptors to protein adsorption capacity (i.e., fouling index) on polymer brushes. In Fig. 4, a positive fouling index indicates that a descriptor promotes protein adsorption on polymer brushes, while a negative index indicates otherwise for resisting protein adsorption. Among 7 descriptors, four descriptors of charge-charge energy, Mi, C-005, and C-007 exhibited strong protein resistance property as evidenced by very high negative fouling indexes of 98–100%, while the other three of RBF (78.7%), O% (73.6%), and ALOGP (97.9%) contributed significantly to protein adsorption. The co-presence of charge-charge energy and C-005 mainly encodes the zwitterionic groups of polymer brushes, which will induce strong ionic surface hydration that forms a physical barrier to prevent protein adsorption. Since Mi descriptor presents the first scaled ionization potential relative to carbon atom, it generally is existed in the polar/hydrophilic functional groups involving N and O atoms, which serve as hydrogen acceptors to form hydrogen bonds between polymer brushes and water molecules, thus contributing to protein resistance. C-007 (CH_2X_2) is strongly associated with the backbone of polymer chains that mainly function as mechanical support, rather than hydrophobic moiety, to retain polymer brush conformations for the enhanced protein resistance. In contrast, RBF and ALOGP descriptors usually involves hydrophobic properties (e.g., CH_3 group) that are known to associated with high protein adsorption. O% (percentage of oxygen atoms) is also identified as a significant descriptor for promoting protein adsorption, probably due to hydrogen bond donor effect. The fouling index of these descriptors can therefore be used as primary criteria to evaluate and design antifouling brushes. For instance, the zwitterionic, hydrogen bonding, and moderate mechanically strong descriptors enable to associate with water around them to form a tightly surface-bound water layer at polymer surface, which provides a strong barrier to prevent approaching proteins being adsorbed on the surface, consistent with the surface hydration theory [60].

Table 2
Cross- and external validations for descriptor-based ANN model and functional group-based SVR model.

Model	Validation	Q^2	RMSE
Descriptor-based ANN model	Leave-five-out cross validation	0.86	0.11
	External Validation	0.81	0.20
Functional group-based SVR model	Leave-five-out cross validation	0.96	0.19
	External Validation	0.79	0.45

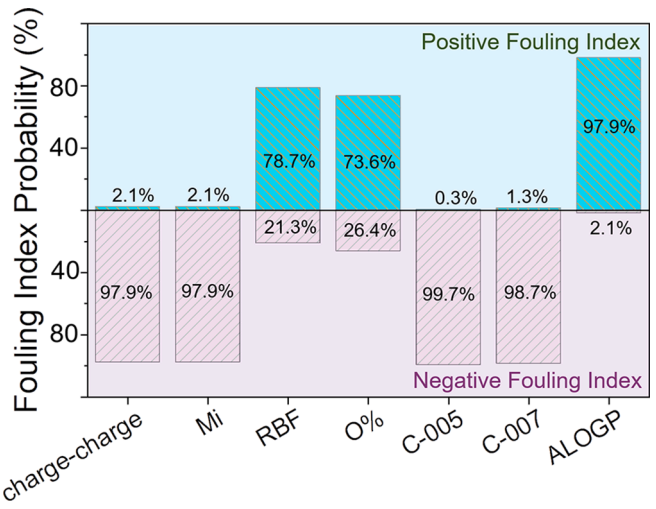


Fig. 4. Fouling index probability (%) of 7 descriptors used for predicting protein resistance and protein adsorption ability on polymer brushes in the ANN model.

2.3. Functional Group-Based SVR models for designing new antifouling polymer brushes

Complement to the ANN model capable of discovering the new antifouling property of existing polymer brushes, we also constructed a functional group-based SVR model for designing new antifouling polymer brushes. As a different machine learning approach, we extracted 29 important molecular descriptors to encode the most-contributing features of polymer brushes. Then, we implemented a factor analysis algorithm to convert 29 descriptors into the more physically meaningful functional groups. On the basis of the factor analysis of the eigenvalues and percentage of cumulative variance (%) on 29 descriptors, the first 8 functional groups (Fig. 5**a**) possessed an eigenvalue of > 1 (Fig. 5**b**), 93.44% structural variance of the 29 descriptors (**Table S2**), a communality value of > 0.5 (**Table S3**), and a rotated absolute regression coefficient of > 0.3 (**Table S4**). All of these characterizations indicate that the 8 functional groups are sufficient to interpret the compositional, structural, geometrical, and connectivity information of 29 descriptors. Next, we added film thickness to the 8 functional groups and reperformed a random forest to confirm the important contributions of these 9 factors to protein adsorption (Fig. 5**c**), which were further used in the SVR model.

To prevent underfitting and overfitting of functional group-based SVR model, we firstly applied a grid searching algorithm to determine the optimal hyperparameters of C and gamma. In the training process, mean average error (MAE) was calculated to monitor the convergence of the model. **Figure S1a** shows the loose grid searching on optimal C of 10 and gamma of 1.0 at global convergence, leading to the smallest MAE of 0.48. Further, a fine grid searching was also implemented to optimize C and gamma values by searching in the grid from 1.0 to 100 for C and from 0.1 to 10 for gamma (**Figure S1b**), resulting in the optimal C of 2.78 and gamma of 0.77 at the smallest MAE of 0.47. At the optimal C and gamma, the 8 functional groups and film thickness were then trained as inputs in response to log(PAA) of polymer brushes using SVR with RBF kernel. Similarly, the dataset was split randomly into training set and test set at the ratio of 7:3. Fig. 5**d** showed that (i) all absolute errors were less than 1.0 for both sets and (ii) 98% (training set) and 81% (test set) errors were less than 0.5, indicating a good training performance of the SVR model for describing the relationship between protein adsorption and functional groups. As shown in Table 2, the leave-five-out cross valididation on the training set yielded $Q^2_{CV} = 0.96$ and $RMSE_{CV} = 0.19$, while external validation on the test set gave rise to $Q^2_{ext} = 0.79$ and $RMSE_{ext} = 0.45$. Further, the correlation of predictions

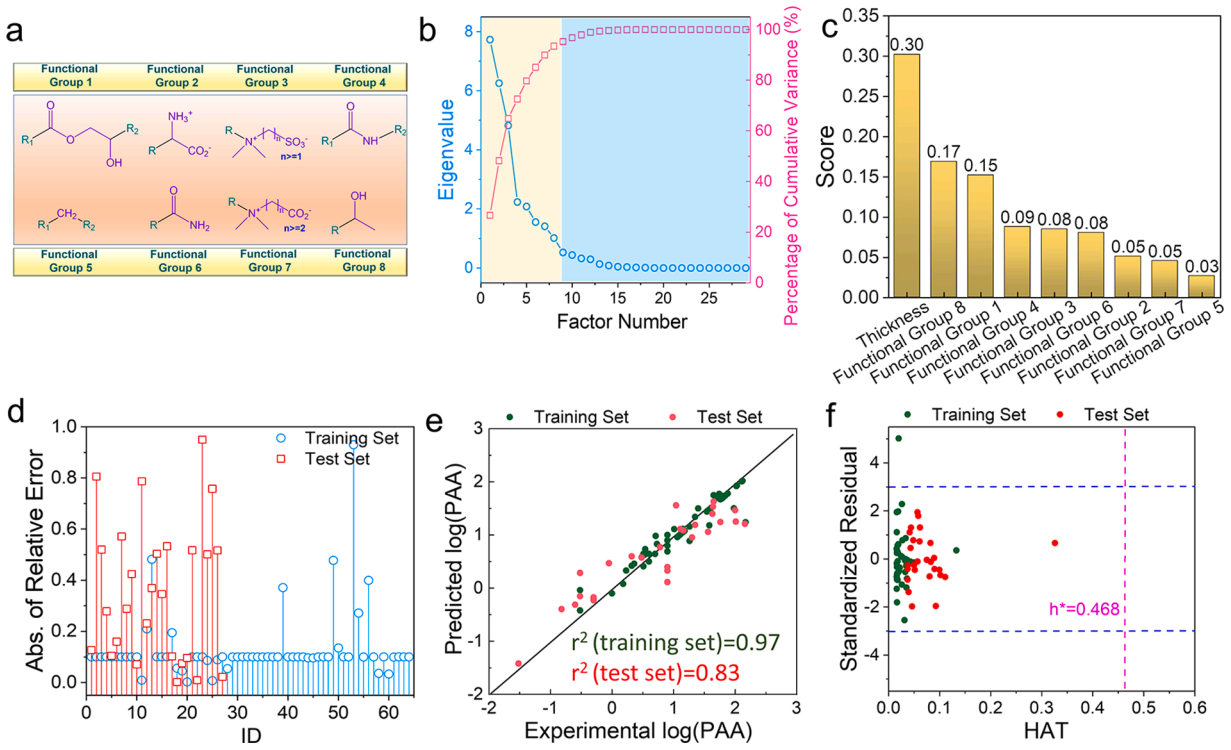


Fig. 5. Compositional and structural representation of antifouling polymer brushes for functional group-based SVR model. (a) Molecular structures of 8 specific functional groups from factor analysis. (b) Eigenvalues (blue) and percentage of cumulative variance (%) (pink) for 29 factors. (c) Importance score ranking of 8 functional groups and film thickness from random forest. (d) Absolute errors of predicted values with respect to experimental values of log(PAA) for both training and test sets. (e) Correlation between predicted and experimental log(PAA) values for both training and test sets by leave-five-out cross validation and external validation. (f) Applicability domain of training and test sets using William plot with h*=0.468.

versus experimental log(PAA) showed r^2 values of 0.97 for a training set and 0.83 for a test set (Fig. 5**e**). The consensus predictions are normally close to the top performance of all individual models for both cross validation and external validations, indicating the model stability, reliability, and predictivity. William plot in Fig. 5**f** showed that almost all of data point (except one) fall within the boundaries of $h^* < 0.468$ and $-3 <$ standardized residual <3. However, there was a data in training set outside the applicability domain, which presented h_i value of 0.145 smaller than the threshold h^* value of 0.468. After removing this data point, the reliability and predictivity of the SVR model were not improved significantly, indicating that this data has no impact on the performance of the SVR model that is considered reliable.

Different from descriptors and their associated fouling indexes, our factor analysis and Bayesian statistics identify 8 functional groups and film thickness as critical factors for contributing to protein adsorption/resistance on any polymer brush (Fig. 6). These functional groups offer the much more straightforward chemical structures for a design purpose. Among them, functional groups of 2 (amino acid-based group), 3 (sulfobetaine group), 4 (secondary amide group), 7 (carboxybetaine group), 8 (isopropanol alcohol-like group) exhibited the much higher probability of negative fouling indexes of 74.5–100%, indicating the strong association with protein resistance property. Since these functional groups possess either zwitterionic or hydrogen bonding moieties, it is not surprisingly that zwitterionic polymers strongly interact with water molecules via ionic solvation, while hydrophilic polymers (e.g. polyacrylamides or polyacrylates) associate with water molecules via hydrogen bonds, both of which produce a strong surface hydration layer to resist protein adsorption. In contrast, the group 5 (methylene group) showed a high positive fouling index of 89.8% and thus is considered as a fouling-induced group, consistent with experimental observation that proteins favor to be adsorbed on hydrophobic surfaces. It is sort of unexpected results that both groups of 1 and 6 were identified as

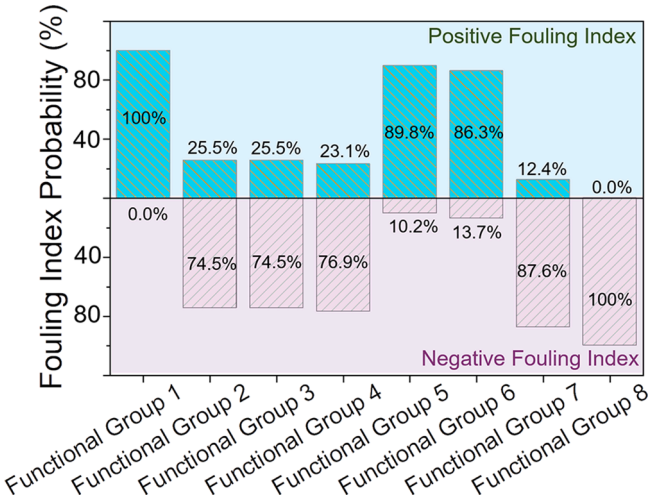


Fig. 6. Fouling index probability (%) of 8 functional groups for predicting protein resistance and protein adsorption ability on polymer brushes in the SVR model.

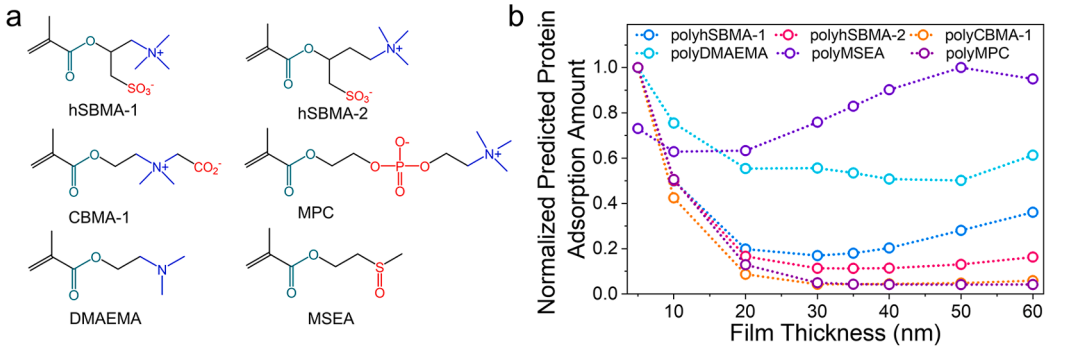
contributors to protein adsorption. Group 1 containing a higher ratio of oxygen atoms that make a positive contribution to protein adsorption, consistent with the prediction of O% in the descriptor-based ANN model. Group 6 is terminated with $–NH_2$ group that is readily charged in the acidic condition [61], which will induce protein adsorption on the charged surface. The functional group-based SVR model, combining with the predicted antifouling property of these functional groups, allows to design new superior antifouling brushes.

2.4. Experimental validation of protein adsorption on repurposed and new polymer brushes

Based on the predictive relationship between descriptors and protein resistance of polymer brushes, we used the ANN model to predict the thickness-dependent protein adsorption from undiluted human serum or plasma on 6 existing polymer brushes of polyhSBMA-1 [62], polyhSBMA-2 [62], polyCBMA-1 [63], polyMPC [64], polyDMAEMA [65], and polyMSEA [65], whose monomer structures were showed in Fig. 7**a**. At a first glance for Fig. 7**b** and **Table S5**, for any polymer brush, there existed an optimal film thickness to achieve the lowest protein adsorption, i.e., too low or too high film thicknesses induced large protein adsorption to some extents. Specifically in Fig. 7**c**, the four zwitterionic polyhSBMA-1, polyhSBMA-2, polyCBMA-1 and polyMPC brushes, which possess the higher negative fouling indexes as contributed by charge-charge energy, Mi, and $-CH_3$ of betaine group (C-005) descriptors, achieved the protein adsorption of 1.41, 2.89, 2.46, and 6.21 ng/cm [2], respectively, comparable to or better than a superlow fouling level (5 ng/cm^2). PolyDMAEMA brushes, belonging to a polyacrylamide family, showed a low fouling property of 11.05 ng/cm^2 at 50 nm, due to the presence of hydrogen bonds as reflected by RBF, O%, and ALOGP descriptors. As a counter example, polyMSEA brush with the higher positive fouling index of ALOGP and O% was predicted as a weaker antifouling surface, as evidenced by high protein adsorption of 143.55 ng/cm^2 even at an optimal film thickness of 10 nm, consistent with experimental results [65]. The ANN model indeed predicts the effect of film thickness of polymer brushes on protein adsorption particularly from undiluted blood plasma and serum. Consistently, collective experimental results have shown the optimal film thickness of polymer brushes to achieve the minimal protein adsorption at the low or superlow fouling levels (0.3–5 ng/cm^2), including poly(HEAA) brush of 10–40 nm [66], poly(AAEE) brush of 10–40 nm [67], poly(HEMA) brush of 20–30 nm [46], poly(VBIPS) brush of 27–33 nm [47], poly(CBMA) brush of 10–15 nm [68], poly(SBMA) brush of 5–12 nm [69], polyNAGA brush of 25–35 nm [52], polySerMA of 30–40 nm [70], polyOrnAA of 11–12 nm [71], and polyLysAA of 11–12 nm [71]. Both the ANN model and experimental data suggest that at optimal film thicknesses, intra- and intermolecular associations between polymer chains are optimized to achieve high surface hydration for effectively preventing protein adsorption from the surface.

Upon demonstrating the predictivity and accuracy of the descriptor-based ANN model for existing antifouling polymer brushes, we further applied the ANN model to screen and discover the potential antifouling property of 14 polymer brushes with other purposes and functions as selected from literature and summarized in **Table S6**. Among the 14 polymer brushes, the 7 polymer brushes have been studied for other properties including hydrophilicity of polyDHMA [72], polyHTMA [73], and polyNMEMA [74], boron removal property of polyHAEM [75], thermo-responsive property of polyMA-Ala-OMe [76], bio-sensitivity of polyNHSMA [77], and adhesive property of polyDMVSA [78], but none of them has been reported antifouling property before. From a synthesis viewpoint, we selected and synthesized 3 monomers of MA-Ala-OMe, NHSMA, and DMVSA (Fig. 8**a**), followed by polymerization and grafting them onto gold substrate using surface-initiated ATRP. The resultant polymer brushes were then challenged by undiluted human serum and plasma for their surface resistance to protein adsorption using SPR assays. SPR results in Fig. 8**b** showed that polyMA-Ala-OMe, polyNHSMA, and polyDMVSA brushes exhibited very low protein adsorption of 6.0/4.2, 9.0/8.9, and 6.8/6.2 ng/cm [2] from undiluted human serum/plasma, comparable to computationally predicted protein adsorption of 11.55, 5.19, and 7.02 ng/cm^2 on polyMA-Ala-OMe, polyNHSMA, and polyDMVSA, respectively. Clearly, the ANN model can indeed be used to screen existing polymer brushes and to discover their potential antifouling property if any.

For a material design purpose, based on the predictive relationship between functional groups and corresponding protein resistance, we applied the SVR model to design the three molecules via a combination of different functional groups with negative fouling indexes. Specifically, we designed a methyl vinylcarbinol (MVC) monomer consisting of isopropanol-like (group 8) and vinyl groups, a 3-((2-hydroxyethyl)-dimethylammonio) propanoate (EDLP) monomer made of sulfobetaine (group 3), methylene (group 5), and hydroxyl groups, and a 2-((2-hydroxyethyl)-dimethylammonio) ethane-1-sulfonate (DAES) monomer made of carboxybetaine (group 7), methylene (group 5), and hydroxyl groups, respectively (Fig. 8**c**). Upon SI-ATRP to form polymer brushes on Au substrates, SPR spectra showed that the final adsorbed proteins from undiluted blood serum/plasma were 3.2/7.0 ng/cm^2 on polyMVC, 9.0/0.0 ng/cm^2 on PAA-EDLP and 0.0/4.0 ng/cm^2 on PAA-DAES (Fig. 8**d**),



c

Name	Predicted Optimal Film Thickness (nm)	Predicted Protein Adsorption Amount (ng/cm²)
polyhSBMA-1	30	1.41
polyhSBMA-2	35	2.89
polyCBMA-1	30	2.46
polyMPC	50	6.21
polyDMAEMA	50	11.05
polyMSEA	10	143.55

Fig. 7. Predicted dependence of film thickness of polymer brushes on protein adsorption by the descriptor-based ANN model. (a) Monomer structures used for synthesizing 6 antifouling polymer brushes (hSBMA-1, hSBMA-2, CBMA-1, MPC, DMAEMA, and MSEA). (b) Predicted protein adsorption at (b) different film thicknesses (nm) and (c) optimal film thickness (nm) of 6 polyhSBMA-1, polyhSBMA-2, polyCBMA-1, polyMPC, polyDMAEMA, and polyMSEA brushes.

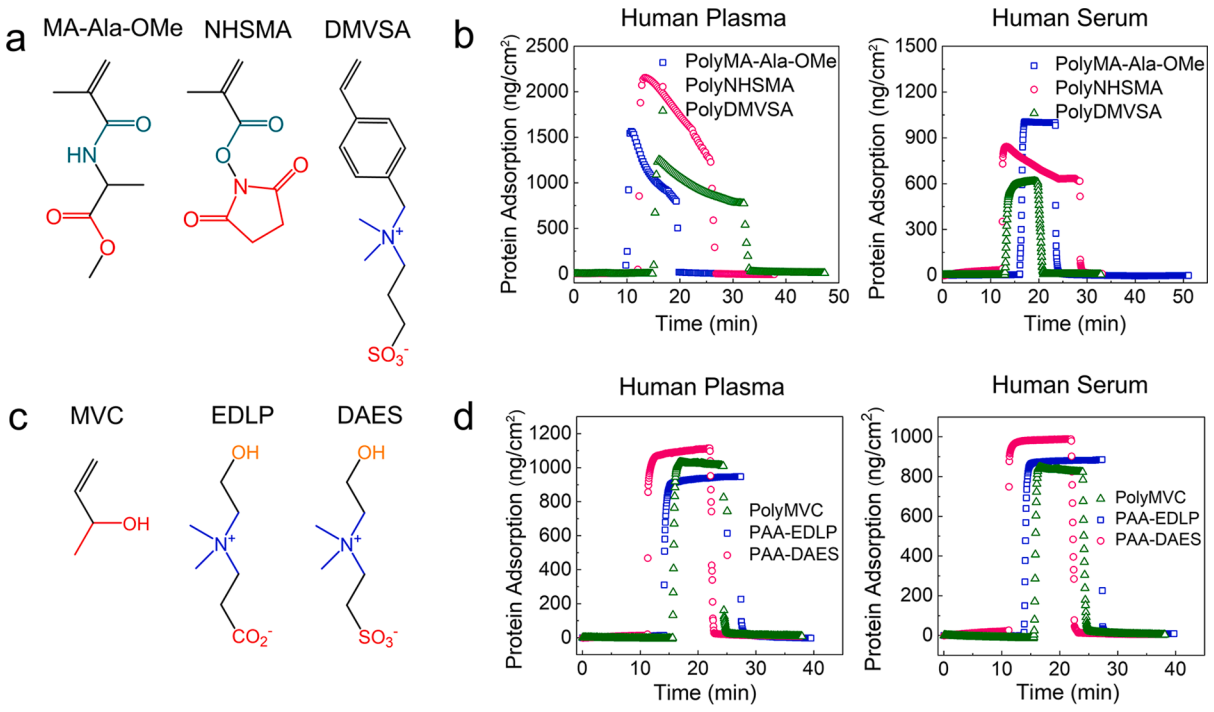


Fig. 8. SPR validation for the repurposing and design of antifouling polymer brushes by both ANN and SVR models. (a) Monomer structures of MA-Ala-OMe, NHSMA, and DMVSA by the ANN model. (b) Protein adsorption (ng/cm^2) from undiluted human serum and plasma on polyMA-Ala-OMe, polyDMVSA, and polyNHSMA brushes by SPR. (c) Monomer structures of MVC, EDLP, and DAES by the SVR model. (d) Protein adsorption from undiluted human serum and plasma on polyMVC, PAA-EDLP, and PAA-DAES brushes by SPR.

in comparison with the SVR-predicted protein adsorption of 3.8–4.48 ng/cm^2 on polyMVC, 3.3–3.9 ng/cm^2 on PAA-EDLP and 3.9–4.0 ng/cm^2 on PAA-DAES (Table S7). Both SPR and SVR results showed a consistent protein resistance capacity of polyMVC, PAA-EDLP and PAA-DAES, confirming the predictivity of the SVR model. It is not surprising to reveal the differences in protein adsorption amounts between machine-learning models and SPR results. Such differences could be attributed to some surface/experimental factors (e.g. surface roughness, grafting density, surface coverage, chain conformations, chain flexibility, etc) that are not considered in our model. We should also note that the key contribution of our machine-learning model does not necessarily intent to achieve 100% predictive accuracy, instead (i) to design new antifouling materials and coatings that will be validated by experiments and (ii) to provide specific antifouling functional groups for design purposes.

3. Conclusions

In this work, we first constructed a benchmark dataset of antifouling polymer brushes from literatures and then developed data-driven machine-learning models for determining, repurposing, and designing new antifouling polymer brushes. Specifically, the descriptor-based ANN model was developed to repurpose and discover the antifouling property of existing polymer brushes with other desirable properties, while the functional group-based SVR model was established to design new antifouling brushes. Both machine-learning models (i) achieved high reliability and predictivity of $Q^2_{CV} = 0.86$ and $Q^2_{ext} = 0.81$ (ANN) and $Q^2_{CV} = 0.96$ and $Q^2_{ext} = 0.79$ (SVR), (ii) identified the structural/component-dependent descriptors (charge-charge energy, Mi, C-005, and C-007) and functional groups (amino acid-based group 2, sulfobetaine group 3, secondary amide group 4, caboxybetaine group 5, and isopropanol-like group 8) that contribute to the enhanced antifouling performance, and (iii) predicted the dependence of film thickness of polymer brushes on protein adsorption. By employing both machine learning models, we synthesized three ANN-repurposed polyMA-Ala-OMe, polyNHSMA, and polyDMVSA brushes and three SVR-designed polyMVC, PAA-EDLP, and PAA-DAES brushes for discovering their new antifouling property by SPR. The resultant polymer brushes can achieve very low protein adsorption of 6.0/4.2 ng/cm^2 on polyMA-Ala-OMe, 9.0/8.9 ng/cm^2 on polyNHSMA, 6.8/6.2 ng/cm^2 on polyDMVSA, 3.2/7.0 ng/cm^2 on polyMVC, 9.0/0.0 ng/cm^2 on PAA-EDLP, and 0.0/4.0 ng/cm^2 on PAA-DAES from undiluted human serum/plasma, respectively, thus validating the model prediction and design. This work provides a data-driven computational workflow, built on a small benchmark dataset, for exploring a wide range of compositional and structural spaces to repurpose and design new antifouling materials and coatings beyond polymer brushes.

It is worthy of adding additional remarks that the more relevant and reliable parameters are incorporated into the machine-learning models, the higher accurate prediction is expected. For antifouling materials/coatings, many intrinsic properties (e.g. carbon spacer lengths, molecular moieties, hydrophilic/hydrophobic ratio, electrostatic interactions), surface properties (e.g., grafting density, film thickness and roughness, chain conformation, molecular weight), and experimental conditions (e.g., temperature, pH, salt conditions) are contributed to antifouling properties of surface coatings to some extents. Here, we mainly focus on (i) the structural-based design and (ii) the fundamental structure–property relationship of antifouling materials/coatings for several reasons: (1) Intrinsic physicochemical properties are the most important structural information for any material design. More importantly, these intrinsic properties can be qualified by different molecular descriptors by modern computational chemistry (e.g., simplest atom types, functional groups, fragment counts, topological and geometrical features); (2) Surface properties are also very critical for antifouling performance of surface coatings. Among them, film thickness (coating thickness) has been well demonstrated its importance and effects on antifouling property and this property has also been often reported in literature [46,60]. Other surface coating properties, e.g., surface roughness, grafting molecular weight, chain conformation, were only reported by very few studies or systems, thus we can not use these incomplete properties to construct a consistent dataset for machine-

learning; (3) Similarly, since experimental conditions are varied greatly for different antifouling systems, they are not suitable to derive the fundamental structure–property relationship of given materials. Thus, we apply "the best we can do with the available data we have" strategy to extract all possible composition-structure-performance information from the limited existing data for achieving the state-of-art materials design.

4. Materials and methods

4.1. Machine learning section

4.1.1. Dataset

We collected 94 protein adsorption data from undiluted human serum or plasma and the corresponding film thickness (**Table S1**) of 14 zwitterionic-based and 14 hydrophilic-based polymer brushes (Fig. 2) from different literatures. The dataset containing all the 94 protein adsorption data were used to constructed descriptor-based ANN model. The dataset possessing 91 protein adsorption data after removing data points of entry H13-01, H14-01 and H14-02 (**Table S1**) was used to build functional group-based SVR model.

4.1.2. Molecule drawing and descriptor calculation

The molecular structures of all polymer brushes were sketched using ChembioDraw software. The geometrical optimizations of these drawn molecular structures were carried out using Chem3D via Merck Molecular Fore Field (MMFF94) with steepest descent algorithm. Three energy descriptors (intramolecular charge-charge, charge-dipole, and dipole–dipole energies) for these compounds were computed in the geometrical optimization process. The molecular descriptors for the 28 structures were calculated using alvaDesc 1.0.8 software (https://www.alvascience.com/alvadesc).

4.1.3. Variable reduction for machine-learning models

To reduce the high dimensional data for machine-learning models, three-step procedures of variable reduction were performed to reduce the number of variables for both ANN and SVR models. To select descriptors for ANN model, molecular descriptors from alvaDesc software with zero variance or near zero variance were firstly removed, giving rise to 108 variables (104 molecular descriptors, three energy descriptors, and film thickness). Followed by Pearson analysis, variables with absolute correlation coefficient>0.6 were removed, leading to 14 variables. Then, random forest algorithm was implemented to find the variables with key contribution to surface resistance/adsorption ability of polymer brushes, resulting in 8 variables (6 molecular descriptors, 1 energy descriptor, and film thickness) for ANN model.

To obtain specific functional groups for SVR model, firstly, we statistically analyzed the correlation between a pool of functional group-based and atom-centered molecular descriptors and removed redundant descriptors by taking care of multicollinearity between the descriptors, causing 29 molecular descriptors. Then, factor analysis [79] was carried out to generate 8 explicit functional groups which can summarize the entire cluster of 29 resultant molecular descriptors using MATLAB software. Further, random forest algorithm was performed to screen 9 variables (8 functional groups and film thickness) which highly correlated with the antifouling property/protein adsorption ability of polymer brushes.

4.1.4. Training and validation of machine-learning models

In general, artificial neural network (ANN), as a predictive model, is designed by several layered structures, each layer containing input/output neurons being functioned and weighted by the linear/nonlinear optimization algorithms (e.g., Relu, tanh, sigmoid, or softmax). Additionally, for each layer, a bias term is added to the weighted sum of the input neurons for adjusting the output neurons. Throughout the iteration process, the input/output neurons will be optimized and weighted to achieve the desirable outcome. In this work, the training of a five-layer ANN model was implemented in keras and TensorFlow modulus of Python 3.6. The dataset was randomly divided into training set (70%) and test set (30%). The training set was used to training the model based on the designed neural network, while the test set was applied to validate the predictivity of the model. This ANN model consists of an input layer containing information of the resulting 8 variables (6 molecular descriptors, 1 energy descriptor, and film thickness) and an output layer which is a prediction of antifouling property of polymer brushes. Three "dense" hidden layers with 64, 96, and 64 neurons were constructed to connect the input and the output layers through every node in the previous layer weighted connecting with every node in the next layer. The activation functions applied in these layers are rectified linear unit (Relu), hyperbolic tangent (tanh), and rectified linear unit (Relu) functions, respectively. The inputs (X_i) of the resulting 8 descriptors (6 molecular descriptors, 1 energy descriptor, and film thickness) are neurons, followed by the weighted sum of inputs along with the bias term $x_1w_1 + x_2w_2 + \cdots + x_8w_8 + bias$. The summation of resulting outputs was subject to the activation function of Relu in the second layer, then the weighted outputs of the second layer were fed into activation function of tanh in the third layer, and the weighted outputs of the third layer were functioned using activation function of Relu in the fourth layer. The predicted log(PAA) were obtained through weighted summation of outputs of the fourth layer with the addition of bias term. Before training of the model, the weights between neurons of previous layer and neurons of next layer were randomly assigned using He normal method. During the training process, the weights between each node of previous layer and each node of next layer were optimized. Mean absolute errors (MAE) between the experimental log(PAA) and the predicted log(PAA) for training and test sets were used to evaluate the convergence of the model. To this end, the deep-learning methods including automatic learning-rate reduction and early stopping were implemented to avoid overfitting. When MAE of both training set and test set reaches to their convergence, the model was regarded as the optimal model.

The training of a SVR model used random model initiation and the dataset was randomly split into training set and test set at the ratio of 7:3. The training of functional group-based SVR model was carried out through two-step grid searching on hyper parameters of C and gamma using RBF kernel. Loose grid searching was firstly carried out from 10^{-6} to 10^{6} for both C and gamma, resulting in optimal C of 10 and gamma of 1.0. Then, fine grid searching was performed from 1.0 to 100 for C and from 0.1 to 10 for gamma, giving rise to optimal C of 2.78 and gamma of 0.77. In the training process, MAE between the experimental log(PAA) and the predicted log(PAA) for training and test sets was used to estimate the convergence of the model.

In this work, we calculated chemical applicability domain using leverage approach based on HAT values of the descriptors' matrix diagonal in a regression model to evaluate the robustness of the two machine-learning models. *X* outliers were characterized by a HAT threshold, while *Y* outliers were determined by the cutoff values of three times standardized residual. The cutoff of leverage values is $3*(k+1)/n$, where k is the number of independent variables and n is the size of dataset. Testing samples within the cutoff value of leverage and within 3 times standardized residuals were considered in the applicability domain of the machine-learning models.

4.1.5. Evaluation of machine-learning models

The reliability and predictive ability of the ANN and SVR models were evaluated using squared correlation coefficient (Q^2_{CV}), 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)):

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

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

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

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

Where y_i^{obs} and $y_i^{pred_{cv}}$ in equation (1) and (2) are (log(PAA)) of experimental and predicted values of training samples and $\bar{y}^{obs}$ is the average value of log(PAA) from experiment for training set, respectively, y_i^{obs} and $y_i^{pred_{ext}}$ in equation (3) and (4) are experimental and predicted log(PAA) values of testing samples, respectively and $\bar{y}^{obs}$ is the average value of log (RAA) from experiment for test set.

5. Experimental section

5.1. Materials

Methacryloyl chloride, dichloromethane, L-alanine methyl ester hydrochloride, triethylamine, sodium bicarbonate ($NaHCO_3$), sodium chloride (NaCl), anhydrous magnesium sulfate ($MgSO_4$), potassium carbonate, acetonitrile (MeCN), dimethylformamide (DMF), methyl vinylcarbinol (MVC), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), 4-dimethylaminopyridine (DMAP), 4-vinlybenzyl chloride, acrylic acid, diethyl ether, ethanol, 2-(dimethylamino)ethanol, 1, 3-propane sultone, SPR chip (Au), sodium bromoalkylsulfonate, ethyl acetate, *n*-hexane, Me_6TREN, cuprous bromide (CuBr), 2-morpholinoethyl methacrylate and phosphate buffer saline (PBS) were purchased from Sigma-Aldrich Co. Ltd. 100% human blood plasma and serum were obtained from BioChemed Service (Winchester, VA). Water was purified by a Millipore water purification system with a minimum resistivity of 18.0 MΩ cm. All other chemicals or biological culture were used as purchased without any purification.

5.2. Synthesis of N-methacryloyl-alanine methyl ester (MA-Ala-OMe)

MA-Ala-OMe was synthesized and purified by according to the previous literature [76]. Briefly, methacryloyl chloride (3.15 g) was added to a mixture of L-alanine methyl ester hydrochloride (4 g) and triethylamine (6 mL) in dichloromethane (150 mL) at 0 °C. After stirring at room temperature overnight, the resulting mixture was washed with 1 M HCl (100 mL), a saturated $NaHCO_3$ solution (100 mL) and a saturated NaCl solution (100 mL). The organic layer was dried over anhydrous $MgSO_4$, filtered, and then concentrated by rotary evaporation. The obtained MA-Ala-OMe was purified using column chromatography (ethyl acetate/*n*-hexane = 1/5) with a yield of about 65%. ^{1}H NMR (D_2O; 300 MHz): –CH_3, 1.44 ~ 1.98, 6H; –$COOCH_3$, 3.55 ~ 3.77, 3H; –CH-, 4.50 ~ 4.65, 1H; =CH_2, 5.38 ~ 5.75, 2H; –CONH-, 6.30 ~ 6.36, 1H)

5.3. Synthesis of 3-(dimethyl-(4-vinylbenzyl) ammonio) propyl sulfonate (DMVSA)

Potassium carbonate (27.6 g, 0.20 mol), 4-vinlybenzyl chloride (15.3 g, 0.10 mol) and dimethylamine solution (10.0 mL, 0.90 g/ml) were dissolved in 100 machine learning of ethanol in a flask. After degassed by nitrogen flow, the flask was heated to 50 °C for 24 h with magnetic stirring. The crude product, obtained by suction filtration and rotary evaporation in turn, was purified by column chromatography and distillation in vacuum to get a transparent liquid. The transparent liquid (6.0 g, 37.03 mmol) and 1, 3-propane sultone (4.5 g, 36.88 mmol) were dissolved in 160 machine learning of dry acetonitrile in a flask. Then, the flask was heated to 50 °C for 48 h under magnetic stirring. A white precipitate, DMVSA monomer, was obtained by suction filtration and dry in a vacuum oven at room temperature. ^{1}H NMR (D_2O; 300 MHz): –CH_2-, 2.25 ~ 3.30, 6H; –NCH_3, 2.90 ~ 2.96, 6H; –CH2-, 4.40 ~ 4.43, 2H; CH_2 = CH–, 5.36 ~ 6.80; -C_6H_4, 7.56 ~ 8.00, 4H.

5.4. Synthesis of 3-((2-hydroxyethyl)dimethylammonio)propanoate (EDLP)

EDLP was synthesized by coupling equivalent amounts of acrylic acid and 2-(dimethylamino)ethanol under vigorous stirring. In brief, 2-(dimethylamino)ethanol was added dropwise into acrylic acid solution using a simple ice-water bath. After approximately 30 min, the viscous mixture was further reacted at room temperature for 3 h. The resultant was crushed and washed with a small portion of MeCN and dried under reduced pressure (white powder; yielding: ~90%). ^{1}H NMR (D_2O; 300 MHz): (–CH_2-, 2.50 ~ 2.55, 2H; –NCH_3, 3.00, 6H; –CH_2-, 3.33 ~ 3.35, 2H; –CH_2-, 3.45 ~ 3.50, 2H; –CH_2-, 3.86 ~ 3.90, 2H; –OH, 4.24 ~ 4.30, 1H).

5.5. Synthesis of 2-((2-hydroxyethyl)dimethylammonio)ethane-1-sulfonate (DAES)

DAES was synthesized by a substitution reaction. Briefly, sodium bromoalkylsulfonate (25.0 mmol) was added into 100 mL of DMF in a 250 mL flask. After heated to 70 °C for 1 h, the quadruple molar-ratio 2-(dimethylamino)ethanol was added to the mixture and stirred for another 48 h. The resultant white powders were obtained by filtering the suspension, followed by dimethylformamide and diethyl ether washing. ^{1}H NMR (D_2O; 300 MHz): –NCH_3, 3.22, 6H; –CH_2, 3.45–3.48, 2H; –CH_2, 3.55–3.57, 2H; –CH_2, 3.79–3.82, 2H; –OCH_2, 4.06–4.09, 2H.

5.6. Grafting polymer brushes onto SPR chip

Surface plasma resonance (SPR) chip was rinsed with ethanol, acetone and water sequentially, later treated under UV ozone for 20 min, washed by DI-water and finally air-dried. Subsequently, an initiator self-assembled monolayer (SAM) was anchored onto the SPR chips by soaking SPR chips into 1 mM ω-mercaptoundecyl bromoisobutyrate (initiator) ethanol solution at room temperature overnight. Secondly, one tube containing monomer (0.6 ~ 1.2 g), Me_6TREN (40 μL), and degassed methanol: water (1:1, v%) solution was transferred to the second tube containing SPR gold chip coated with immobilized initiators and CuBr (20 mg), undergoing SI-ATRP reaction at room temperature. After the controlled reaction time, the reaction was stopped by exposing to air. In order to remove unreacted monomers or unbounded polymer, the chips were soaked in PBS buffer overnight. Polymer brushes with various thicknesses were controlled by tuning the polymerization time (6 ~ 24 h). Since SPR chip (Au/silicate layers) is quite tedious to etch directly, the corresponding grafted polymer brushes are hard to obtain. Therefore, to characterize the molecular weight and polymerization distribution index (PDI) for each polymer brush, we also added 0.1 mM dissociative sulfhydryl modified initiators (ω-mercaptoundecyl bromoisobutyrate) into solution to obtain free single polymer chains. The free polymer chains were expected to possess similar molecular weight and PDI performances with polymer brushes after undergoing a same polymerization environment.

Due to the unique structure of polyzwitterions, PAA-EDLP and PAA-DAES brushes were synthesized by post-modification of the synthetic PAA (poly(acrylic acid)) brushes. For instance, after preparing the PAA brushes on SPR chips, the SPR substrates were washed and further immersed into 10 mL of aqueous solution containing EDLP (or DAES; 2.0 g) and EDC/DMAP (1.5 g/1.8 g). The active pair groups of hydroxyl

and carboxyl would undergo a condensation reaction, thus the target structural polymer brushes would be prepared (**Figure S2**).

5.7. Protein adsorption by SPR measurement

A customized SPR sensor based on wavelength interrogation was used to determine protein adsorption performance on polymer brushes with different thicknesses. The solution (undiluted human protein solution or PBS buffer) was flowed through 4 channels under the pressure of peristaltic pump. Specifically, the SPR chip coated with polymer brushes was connected to the surface of prism. A normal baseline signal was measured by flowing PBS solution through the detector of sensor. Each protein solution, i.e. 100% human blood plasma or serum, was independently flowed through channels for ~ 10 min, and SPR wavelength would shift if any protein adsorption happened. Usually, due to the nonspecific adhesion of the protein, the wavelength pattern would generate an ascending "step". After that, the protein solution was replaced by PBS solution again to remove the unbound or slightly bounded protein on the prism surface and maintained > 10 min. The flow rate of solutions was controlled at 0.05 mL·min^{-1}. The wavelength shift, mainly derives from the difference value between the PBS baselines obtained before and after corresponding protein solution flow. It should be pointed out that a 1 nm SPR wavelength shift at 750 nm corresponds to ~15 ng·cm^{-2} protein adsorption according to the previous reported strategy [80].

5.8. Gel permeation chromatography

We also performed parallel synthesis and characterization to qualify the molecular weight and PDI of polymer in solution. The polymers produced from the solution ATRP method is often used to roughly mimic the polymers grafted on the surface via the SI-ATRP method under the same condition. Gel permeation chromatography (GPC) analysis was characterized on a Tosoh EcoSEC HLC-8320 GPC using chromatographic-level THF as an eluent. Dissociative polymers extracted from SI-ATRP solutions were dialyzed into pure water for at least 3 days to remove unreacted monomer and cupric salts. The obtained polymers were subsequently dissolved into chromatographic-level THF and further filtered by using a 0.2 μm PTFE syringe filter. GPC traces showed the low distribution of PDI of 1.02 and molecular weight of 6452 g/mol for polyDMVSA and PDI of 1.04 and molecular weight of 4692 g/mol for polyMVC, indicating good control of the final product molecular characteristics (**Table S8**).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

J.Z. thanks financial supports from NSF (1806138 and 1825122).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cej.2021.129872.

References

[1] S. Ekins, A.C. Puhl, K.M. Zorn, T.R. Lane, D.P. Russo, J.J. Klein, A.J. Hickey, A. M. Clark, Exploiting machine learning for end-to-end drug discovery and development, Nat. Mater. 18 (5) (2019) 435–441.
[2] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, Applications of machine learning in drug discovery and development, Nat. Rev. Drug Discovery 18 (6) (2019) 463–477.
[3] A. Yosipof, R.C. Guedes, A.T. García-Sosa, Data mining and machine learning models for predicting drug likeness and their disease or organ category, Front. Chem. 6 (2018) 162.
[4] K. Santosh, S. Antani, D.S. Guru, N. Dey, Medical Imaging: Artificial Intelligence, Image Recognition, and Machine Learning Techniques, CRC Press, 2019.
[5] L. Li, H. Ruan, C. Liu, Y. Li, Y. Shuang, A. Alù, C.-W. Qiu, T.J. Cui, Machine-learning reprogrammable metasurface imager, Nat. Commun. 10 (1) (2019) 1082.
[6] R. Alizadehsani, M. Roshanzamir, M. Abdar, A. Beykikhoshk, A. Khosravi, M. Panahiazar, A. Koohestani, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, A database for using machine learning and data mining techniques for coronary artery disease diagnosis, Sci. Data 6 (1) (2019) 1–13.
[7] A. Hazra, S.K. Mandal, A. Gupta, A. Mukherjee, A. Mukherjee, Heart disease diagnosis and prediction using machine learning and data mining techniques: a review, Adv. Comput. Sci. Technol. 10 (7) (2017) 2137–2159.
[8] G. Tardioli, R. Kerrigan, M. Oates, O.D. James, D. Finn, Data driven approaches for prediction of building energy consumption at urban level, Energy Procedia 78 (2015) 3378–3383.
[9] C. Fan, D. Yan, F. Xiao, A. Li, J. An, X. Kang, Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches, Build. Simul. 14 (1) (2021) 3–24.
[10] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, NPJ Comput. Mater. 3 (1) (2017) 54.
[11] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science, NPJ Comput. Mater. 5 (1) (2019) 83.
[12] H. Zhao, C.I. Ezeh, W. Ren, W. Li, C.H. Pang, C. Zheng, X. Gao, T. Wu, Integration of machine learning approaches for accelerated discovery of transition-metal dichalcogenides as Hg0 sensing materials, Appl. Energy 254 (2019), 113651.
[13] V. Fung, G. Hu, P. Ganesh, B.G. Sumpter, Machine learned features from density of states for accurate adsorption energy prediction, Nat. Commun. 12 (1) (2021) 88.
[14] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T. J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, NPJ Comput. Mater. 6 (1) (2020) 1–9.
[15] P.V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning, Nat. Commun. 9 (1) (2018) 1668.
[16] M. Moliner, Y. Román-Leshkov, A. Corma, Machine learning applied to zeolite synthesis: the missing link for realizing high-throughput discovery, Acc. Chem. Res. 52 (10) (2019) 2971–2980.
[17] S. Lin, Y. Wang, Y. Zhao, L.R. Pericchi, A.J. Hernández-Maldonado, Z. Chen, Machine-learning-assisted screening of pure-silica zeolites for effective removal of linear siloxanes and derivatives, J. Mater. Chem. A 8 (6) (2020) 3228–3237.
[18] R. Batra, C. Chen, T.G. Evans, K.S. Walton, R. Ramprasad, Prediction of water stability of metal–organic frameworks using machine learning, Nat. Machine Intell. 2 (11) (2020) 704–710.
[19] P.Z. Moghadam, S.M. Rogge, A. Li, C.-M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D.A. Gomez-Gualdron, V. Van Speybroeck, Structure-mechanical stability relations of metal-organic frameworks via machine learning, Matter 1 (1) (2019) 219–234.
[20] L. Yao, Z. Ou, B. Luo, C. Xu, Q. Chen, Machine learning to reveal nanoparticle dynamics from liquid-phase tem videos, ACS Cent. Sci. 6 (8) (2020) 1421–1430.
[21] J. He, C. He, C. Zheng, Q. Wang, J. Ye, Plasmonic nanoparticle simulations and inverse design using machine learning, Nanoscale 11 (37) (2019) 17444–17459.
[22] K. Tanaka, K. Hachiya, W. Zhang, K. Matsuda, Y. Miyauchi, Machine-Learning Analysis to Predict the Exciton Valley Polarization Landscape of 2D Semiconductors, ACS Nano 13 (11) (2019) 12687–12693.
[23] A. Mannodi-Kanakkithodi, M.Y. Toriyama, F.G. Sen, M.J. Davis, R.F. Klie, M. K. Chan, Machine-learned impurity level prediction for semiconductors: the example of Cd-based chalcogenides, NPJ Comput. Mater. 6 (1) (2020) 1–14.
[24] S. Wu, Y. Kondo, M.-A. Kakimoto, B. Yang, H. Yamada, I. Kuwajima, G. Lambard, K. Hongo, Y. Xu, J. Shiomi, Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm, NPJ Comput. Mater. 5 (1) (2019) 1–11.
[25] B. Mortazavi, E.V. Podryabinkin, S. Roche, T. Rabczuk, X. Zhuang, A.V. Shapeev, Machine-learning interatomic potentials enable first-principles multiscale modeling of lattice thermal conductivity in graphene/borophene heterostructures, Mater. Horiz. 7 (9) (2020) 2359–2367.
[26] J.R. Kitchin, Machine learning in catalysis, Nat. Catal. 1 (4) (2018) 230–232.
[27] Z. Yang, W. Gao, Q. Jiang, A machine learning scheme for the catalytic activity of alloys with intrinsic descriptors, J Mater Chem A 8 (34) (2020) 17507–17515.
[28] K.A. Brown, S. Brittman, N. Maccaferri, D. Jariwala, U. Celano, Machine learning in nanoscience: big data at small scales, Nano Lett. 20 (1) (2019) 2–10.
[29] Y. Zhuo, A.M. Tehrani, A.O. Oliynyk, A.C. Duke, J. Brgoch, Identifying an efficient, thermally robust inorganic phosphor host via machine learning, Nat. Commun. 9 (1) (2018) 1–10.
[30] V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, NPJ Comput. Mater. 4 (1) (2018) 1–14.
[31] Q. Wang, J. Ding, L. Zhang, E. Podryabinkin, A. Shapeev, E. Ma, Predicting the propensity for thermally activated β events in metallic glasses via interpretable machine learning, NPJ Comput. Mater. 6 (1) (2020) 1–12.
[32] C. Wang, H. Fu, L. Jiang, D. Xue, J. Xie, A property-oriented design strategy for high performance copper alloys via machine learning, NPJ Comput. Mater. 5 (1) (2019) 1–8.

[33] Y.-J. Hu, G. Zhao, M. Zhang, B. Bin, T. Del Rose, Q. Zhao, Q. Zu, Y. Chen, X. Sun, M. de Jong, Predicting densities and elastic moduli of SiO 2-based glasses by machine learning, NPJ Comput. Mater. 6 (1) (2020) 1–13.
[34] K.D. Saharuddin, M.H.M. Ariff, I. Bahiuddin, S.A. Mazlan, S.A.A. Aziz, N. Nazmi, A. Y.A. Fatah, K. Mohmad, Constitutive models for predicting field-dependent viscoelastic behavior of magnetorheological elastomer using machine learning, Smart Mater. Struct. 29 (8) (2020), 087001.
[35] F. Li, J. Han, T. Cao, W. Lam, B. Fan, W. Tang, S. Chen, K.L. Fok, L. Li, Design of self-assembly dipeptide hydrogels and machine learning via their chemical features, Proc. Natl. Acad. Sci. 116 (23) (2019) 11259–11264.
[36] Q. Zeng, Y. Zhu, B. Yu, Y. Sun, X. Ding, C. Xu, Y.-W. Wu, Z. Tang, F.-J. Xu, Antimicrobial and antifouling polymeric agents for surface functionalization of medical implants, Biomacromolecules 19 (7) (2018) 2805–2811.
[37] A.G. Nurioglu, A.C.C. Esteves, Non-toxic, non-biocide-release antifouling coatings based on molecular structure design for marine applications, J. Mater. Chem. B 3 (32) (2015) 6547–6570.
[38] V.C. Ruiz-Valdepeñas Montiel, J.R. Sempionatto, B. Esteban-Fernández de Ávila, A. Whitworth, S. Campuzano, J.M. Pingarrón, J. Wang, Delayed sensor activation based on transient coatings: Biofouling protection in complex biofluids, J. Am. Chem. Soc. 140 (43) (2018) 14050–14053.
[39] E.R. Kenawy, S.D. Worley, R. Broughton, The chemistry and applications of antimicrobial polymers: A state-of-the-art review, Biomacromolecules 8 (5) (2007) 1359–1384.
[40] S. Peng, H. Wang, W. Zhao, Y. Xin, Y. Liu, X. Yu, M. Zhan, S. Shen, L. Lu, Zwitterionic polysulfamide drug nanogels with microwave augmented tumor accumulation and on-demand drug release for enhanced cancer therapy, Adv. Funct. Mater. 30 (23) (2020) 2001832.
[41] S. Jung, S. Park, D. Choi, J. Hong, Efficient drug delivery carrier surface without unwanted adsorption using sulfobetaine zwitterion, Adv. Mater. Interfaces 7 (22) (2020) 2001433.
[42] Y. Liu, D. Zhang, B. Ren, X. Gong, A. Liu, Y. Chang, Y. He, J. Zheng, Computational investigation of antifouling property of polyacrylamide brushes, Langmuir 36 (11) (2020) 2757–2766.
[43] Y. Liu, D. Zhang, B. Ren, X. Gong, L. Xu, Z.-Q. Feng, Y. Chang, Y. He, J. Zheng, Molecular simulations and understanding of antifouling Zwitterionic polymer brushes, J. Mater. Chem. B 8 (17) (2020) 3814–3828.
[44] Y. Xiang, R.G. Xu, Y.S. Leng, Molecular simulations of the hydration behavior of a zwitterion brush array and its antifouling property in an aqueous environment, Langmuir 34 (6) (2018) 2245–2257.
[45] D.L. Cheung, K.H.A. Lau, Atomistic study of Zwitterionic peptoid antifouling brushes, Langmuir 35 (5) (2019) 1483–1494.
[46] C. Zhao, L. Li, Q. Wang, Q. Yu, J. Zheng, Effect of film thickness on the antifouling performance of poly(hydroxy-functional methacrylates) grafted surfaces, Langmuir 27 (8) (2011) 4906–4913.
[47] H. Chen, J. Yang, S. Xiao, R. Hu, S.M. Bhaway, B.D. Vogt, M. Zhang, Q. Chen, J. Ma, Y. Chang, L. Li, J. Zheng, Salt-responsive polyzwitterionic materials for surface regeneration between switchable fouling and antifouling properties, Acta Biomater. 40 (2016) 62–69.
[48] S. Chen, J. Zheng, L. Li, S. Jiang, Strong resistance of phosphorylcholine self-assembled monolayers to protein adsorption: insights into nonfouling properties of zwitterionic materials, J. Am. Chem. Soc. 127 (41) (2005) 14473–14478.
[49] J. Zheng, L. Li, H.-K. Tsao, Y.-J. Sheng, S. Chen, S. Jiang, Strong repulsive forces between protein and oligo (ethylene glycol) self-assembled monolayers: a molecular simulation study, Biophys. J. 89 (1) (2005) 158–166.
[50] Y. He, Y. Chang, J.C. Hower, J. Zheng, S. Chen, S. Jiang, Origin of repulsive force and structure/dynamics of interfacial water in OEG–protein interactions: a molecular simulation study, PCCP 10 (36) (2008) 5539–5544.
[51] C. Li, C. Liu, M. Li, X. Xu, S. Li, W. Qi, R. Su, J. Yu, Structures and antifouling properties of self-assembled zwitterionic peptide monolayers: effects of peptide charge distributions and divalent cations, Biomacromolecules 21 (6) (2020) 2087–2095.
[52] F. Yang, Y. Liu, Y. Zhang, B. Ren, J. Xu, J. Zheng, Synthesis and characterization of ultralow fouling poly(N-acryloyl-glycinamide) brushes, Langmuir 33 (49) (2017) 13964–13972.
[53] J.R. Almeida, J. Moreira, D. Pereira, S. Pereira, J. Antunes, A. Palmeira, V. Vasconcelos, M. Pinto, M. Correia-da-Silva, H. Cidade, Potential of synthetic chalcone derivatives to prevent marine biofouling, Sci. Total Environ. 643 (2018) 98–106.
[54] K. Feng, X. Li, L. Yu, Synthesis, antibacterial activity, and application in the antifouling marine coatings of novel acylamino compounds containing gramine groups, Prog. Org. Coat. 118 (2018) 141–147.
[55] B. Rasulev, F. Jabeen, S. Stafslien, B.J. Chisholm, J. Bahr, M. Ossowski, P. Boudjouk, Polymer coating materials and their fouling release activity: a cheminformatics approach to predict properties, ACS Appl. Mater. Interfaces 9 (2) (2017) 1781–1792.
[56] E. Ostuni, R.G. Chapman, R.E. Holmlin, S. Takayama, G.M. Whitesides, A survey of structure-property relationships of surfaces that resist the adsorption of protein, Langmuir 17 (18) (2001) 5605–5620.
[57] T.C. Le, M. Penna, D.A. Winkler, I. Yarovsky, Quantitative design rules for protein-resistant surface coatings using machine learning, Sci. Rep. 9 (1) (2019) 265.
[58] R.J. Kwaria, E.A.Q. Mondarte, H. Tahara, R. Chang, T. Hayashi, Data-driven prediction of protein adsorption on self-assembled monolayers toward material screening and design, ACS Biomater. Sci. Eng. 6 (9) (2020) 4949–4956.
[59] Y. Liu, D. Zhang, Y. Tang, Y. Zhang, Y. Chang, J. Zheng, Machine Learning-Enabled Design and Prediction of Protein Resistance on Self-Assembled Monolayers and Beyond, ACS Appl. Mater. Interfaces 13 (9) (2021) 11306–11319.
[60] S. Chen, L. Li, C. Zhao, J. Zheng, Surface hydration: Principles and applications toward low-fouling/nonfouling biomaterials, Polymer 51 (23) (2010) 5283–5293.
[61] M. Sternberg, D. Cornelius, N. Eberts, F. Schwencle, J. Chiang, M. Kare, Glycinamide hydrochloride: A compound with common salt flavor, Biol. Behav. Aspects Salt Intake (1980) 83–97.
[62] E. Schönemann, J. Koc, N. Aldred, A.S. Clare, A. Laschewsky, A. Rosenhahn, E. Wischerhoff, Synthesis of novel sulfobetaine polymers with differing dipole orientations in their side chains, and their effects on the antifouling properties, Macromol. Rapid Commun. 41 (1) (2020) 1900447.
[63] Z. Zhang, H. Vaisocherova, G. Cheng, W. Yang, H. Xue, S.Y. Jiang, Nonfouling behavior of polycarboxybetaine-grafted surfaces: structural and environmental effects, Biomacromolecules 9 (10) (2008) 2686–2692.
[64] A.M. Alswieleh, N. Cheng, I. Canton, B. Ustbas, X. Xue, V. Ladmiral, S.J. Xia, R. E. Ducker, O. El Zubir, M.L. Cartron, C.N. Hunter, G.J. Leggett, S.P. Armes, Zwitterionic Poly(amino acid methacrylate) Brushes, J. Am. Chem. Soc. 136 (26) (2014) 9404–9413.
[65] L.A. Navarro, A.E. Enciso, K. Matyjaszewski, S. Zauscher, Enzymatically degassed surface-initiated atom transfer radical polymerization with real-time monitoring, J. Am. Chem. Soc. 141 (7) (2019) 3100–3109.
[66] C. Zhao, J. Zheng, Synthesis and characterization of poly(N-hydroxyethylacrylamide) for long-term antifouling ability, Biomacromolecules 12 (11) (2011) 4071–4079.
[67] H. Chen, M. Zhang, J. Yang, C. Zhao, R. Hu, Q. Chen, Y. Chang, J. Zheng, Synthesis and characterization of antifouling poly(N-acryloylaminoethoxyethanol) with ultralow protein adsorption and cell attachment, Langmuir 30 (34) (2014) 10398–10409.
[68] Z. Zhang, M. Zhang, S. Chen, T.A. Horbett, B.D. Ratner, S. Jiang, Blood compatibility of surfaces with superlow protein adsorption, Biomaterials 29 (32) (2008) 4285–4291.
[69] C. Zhao, J. Zhao, X. Li, J. Wu, S. Chen, Q. Chen, Q. Wang, X. Gong, L. Li, J. Zheng, Probing structure-antifouling activity relationships of polyacrylamides and polyacrylates, Biomaterials 34 (20) (2013) 4714–4724.
[70] Q. Liu, A. Singh, L. Liu, Amino acid-based zwitterionic poly (serine methacrylate) as an antifouling material, Biomacromolecules 14 (1) (2013) 226–231.
[71] W. Li, Q. Liu, L. Liu, Antifouling gold surfaces grafted with aspartic acid and glutamic acid based zwitterionic polymer brushes, Langmuir 30 (42) (2014) 12619–12626.
[72] M. Kobayashi, R. Matsuno, H. Otsuka, A. Takahara, Precise surface structure control of inorganic solid and metal oxide nanoparticles through surface-initiated radical polymerization, Sci. Technol. Adv. Mater. 7 (7) (2006) 617–628.
[73] X. Dai, Y. He, Y. Wei, B. Gong, Preparation of hydrophilic polymer-grafted polystyrene beads for hydrophilic interaction chromatography via surface-initiated atom transfer radical polymerization, J. Sep. Sci. 34 (22) (2011) 3115–3122.
[74] X. Chen, S. Armes, S. Greaves, J. Watts, Synthesis of hydrophilic polymer-grafted ultrafine inorganic oxide particles in protic media at ambient temperature via atom transfer radical polymerization: use of an electrostatically adsorbed polyelectrolytic macroinitiator, Langmuir 20 (3) (2004) 587–595.
[75] X. Du, J. Meng, R. Xu, Q. Shi, Y. Zhang, Polyol-grafted polysulfone membranes for boron removal: Effects of the ligand structure, J. Membrane Sci. 476 (2015) 205–215.
[76] Y. Shen, G. Li, Y. Ma, D. Yu, J. Sun, Z. Li, Smart surfaces based on thermo-responsive polymer brushes prepared from L-alanine derivatives for cell capture and release, Soft Matter. 11 (38) (2015) 7502–7506.
[77] K. Takasu, K. Kushiro, K. Hayashi, Y. Iwasaki, S. Inoue, E. Tamechika, M. Takai, Polymer brush biointerfaces for highly sensitive biosensors that preserve the structure and function of immobilized proteins, Sens. Actuators B 216 (2015) 428–433.
[78] J. Zhang, J. Yuan, Y. Yuan, X. Zang, J. Shen, S. Lin, Platelet adhesive resistance of segmented polyurethane film surface-grafted with vinyl benzyl sulfo monomer of ammonium zwitterions, Biomaterials 24 (23) (2003) 4223–4231.
[79] R.A. Johnson, D.W. Wichern. Applied multivariate statistical analysis. Prentice hall Upper Saddle River, NJ: 2002; Vol. 5.
[80] D. Zhang, B. Ren, Y. Zhang, Y. Liu, H. Chen, S. Xiao, Y. Chang, J. Yang, J. Zheng, Micro- and macroscopically structured zwitterionic polymers with ultralow fouling property, J. Colloid Interface Sci. 578 (2020) 242–253.