ACS APPLIED NANO MATERIALS

www.acsanm.org Article

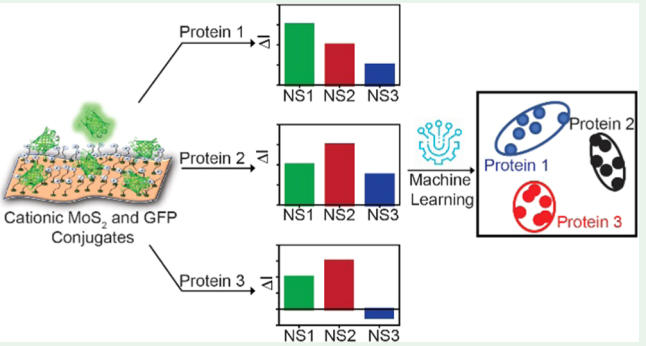
Machine Learning-Assisted Array-Based Detection of Proteins in Serum Using Functionalized MoS_2 Nanosheets and Green Fluorescent Protein Conjugates

Pradipta Behera, Krishna Kumar Singh, Subhendu Pandit, Diptarka Saha, Deepak Kumar Saini, and Mrinmoy De*

Cite This: *ACS Appl. Nano Mater.* 2021, 4, 3843−3851 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Abnormal concentrations of a specific protein or the presence of some biomarker proteins may indicate life-threatening diseases. Pattern-based detection of specific analytes using affinity-regulated receptors is one of the potential alternatives to specific antigen−antibody-based detection. In this report, we have schemed a sensor array by using various functionalized two-dimensional (2D)-MoS_2 nanosheets and green fluorescent protein (GFP) as the receptor and the signal transducer, respectively. Two-dimensional MoS_2 has been used as a promising candidate for recognition of the bioanalytes because of its high surface-to-volume ratio compared to those of other nanomaterials. Easy surface tunability of this material provides additional advantages to analyze the target of interest. The optimized 2D-MoS_2−GFP conjugates are able to discriminate 15 different proteins at 50 nM concentration with a detection limit of 1 nM. Moreover, proteins in the binary mixture and in the presence of serum were discriminated successfully. Ten different proteins in serum media at relevant concentrations were classified successfully with 100% jackknifed classification accuracy, which proves the potentiality of the above system. We have also implemented and discussed the implication of using different machine learning models on the pattern recognition problem associated with array-based sensing.



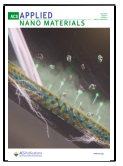
KEYWORDS: *two-dimensional MoS_2, array-based sensing, protein sensing, pattern recognition, machine learning*

INTRODUCTION

Proteins are biomacromolecules consisting of long-chain amino acid residues. They play a vital role in several biological functions such as cell signaling, molecular transportation, DNA transcription, the biological catalysis process, etc. Irregular protein concentration or the emergence of biomarker proteins is the signature step for various life-threatening diseases such as cancer,[1] influenza,[2] Ebola Virus Disease[3] (EVD), HIV,[4] coronavirus disease[5] (COVID-19), etc. Biomarkers are also useful for other disorders such as cardiovascular diseases, liver cirrhosis, neurodegeneration, idiopathic pulmonary fibrosis, etc.[6−9] For early diagnosis and treatment of these diseases, sensing of proteins is very essential. The most popular methods used in the last few decades to detect protein in bioanalytes are enzyme-linked immunosorbent assay[10] (ELISA), lateral flow immunoassay (LFIA), aptamer-based technologies, etc. The working principle of the ELISA protocol is based on the "lock and key approach", i.e., a specific antibody is capable of detecting a specific antigen. Other techniques such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel electrophoresis and mass spectrometry are also utilized as tools for protein sample analysis.[11] However, these techniques are time-consuming and the production cost is high, requiring a sophisticated setup. Moreover, the quantitative measurement does not furnish satisfactory results.

Pattern-based sensing relying on optical response offers an alternative way for the analysis of several analytes, including proteins,[12,13] bacteria,[14,15] diseased cell lines,[16,17] amino acids,[18] nitroexplosives,[19−21] neurotransmitters,[22−24] biothiols,[25−27] etc. This type of sensor has a simplified setup, containing receptor units, which interact with the analyte of interest. A signal transducer unit, which monitors the binding activity between the receptor and analytes, generates a pattern that can be recognized as specific to a certain analyte. In recent years, several receptors, such as fluorescent nanodots,[28,29]

Received: January 26, 2021
Accepted: March 19, 2021
Published: April 2, 2021



https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

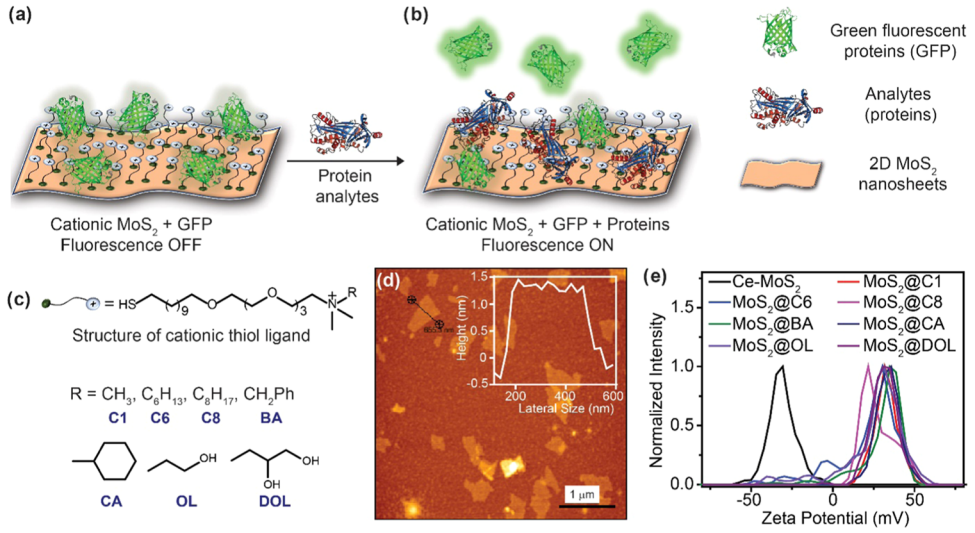


Figure 1. Schematic representation of sensing systems and characterization of sensor elements. (a) Design of the sensor array: the fluorescence of GFP is quenched by cationic MoS_2 and (b) displacement of GFP from the cationic MoS_2 surface, followed by fluorescence regeneration upon analyte addition. (c) Structure of cationic thiol ligands with different head groups at the nonthiol end. (d) Atomic force microscopy (AFM) image of Ce-MoS_2 with the height profile diagram. (e) ζ-Potential plot for Ce-MoS_2 with and without the thiol ligand functionalization.

fluorescent conjugated polymers,[30,31] metal nanoparticles,[32−34] nanoclusters,[35] etc., have been reported as a platform for sensors for the classification of a wide range of analytes. Atomically thin two-dimensional (2D) materials provide wide active surfaces, which help them play a vital role in molecular recognition. Previously, graphene oxide (GO) gained a worthy reputation in array-based sensing for its manifold surface properties.[36−38] Tuning the surface behavior through functionalization to accommodate the analyte of interest is the most important parameter in array-based sensing. Functionalization of GO was mainly achieved through the coupling reaction between the carboxylic acid group of GO and the amine group of external ligands. The above process involves the covalent attachment of ligands with GO surfaces through chemical reactions. Hence, the removal of different reagents and unreacted ligands from the system is a major task. Also, the colloidal stability of graphene oxide mainly relies on surface carboxylic acid groups, which provide hydrophilicity to the system. Once the carboxylic acid groups are converted to other functionalities, the stability may be reduced. For example, in the case of reduced graphene oxides, where all of the carboxylic groups are converted to alcohols, they possess less stability in aqueous solution as compared to graphene oxide.[39,40] These are the major bottlenecks for GO to be extensively used in array-based sensing platforms. These drawbacks can be successfully overcome by two-dimensional MoS_2 (2D-MoS_2), an inorganic version of graphene, although it is rarely explored in array-based sensing.[41,42]

Pattern recognition is central to array-based sensing. Each sensing event is a point in a multidimensional space where dimensionality is the number of sensor-array elements used. Traditionally, linear discriminant analysis (LDA) has been widely used because of its easy interpretability. Recently, it has been shown that nonlinear machine learning-based pattern recognition may improve detection efficiency. The nonlinear models might be better than the linear models at prediction, but it is very difficult to interpret the results and therefore harder to tweak according to specific requirements.

Here, we have used 2D-MoS_2 as a building block for receptor units. Two-dimensional nanosurfaces of MoS_2 provide extended sites for supramolecular interactions due to their high surface-to-volume ratio. This makes it a suitable candidate for biomolecular recognition. The most important factor is surface tunability through external modifiers. 2D-MoS_2 possesses plenty of sulfur defects on its nanosurfaces due to charge transfer of lithium to the layers, followed by sulfide leaching during exfoliation through ion intercalation methods.[43,44] These sulfur vacancies provide atomic-level interfaces and high surface free energy and can be attached with small molecules such as thiols with high affinities.[45,46] Hence, in this work, we have used surface functionalization as the main tool to achieve cross reactivity. A native receptor (2D-MoS_2) was functionalized with seven cationic thiol ligands, where nonthiol ends were diverged with appropriate side chains to entertain feasible biomolecular interactions. Rotello group used thiol-functionalized gold nanoparticles in combination with green fluorescent protein (GFP) to build a sensor array for discrimination of various bioanalytes.[47,48] Previously, functionalized 2D-MoS_2 proved as an alternative to gold nanoparticles in several applications such as enzyme inhibition,[49,50] antimicrobial activity,[51,52] etc. We assumed that this system can be extended in the case of array-based sensing. Negatively charged GFP was chosen as a signal transducer against cationic MoS_2. GFP can interact with cationic MoS_2, followed by formation of an electrostatic complex. The above process results in quenching of the fluorescence of GFP (Figure 1a). We have designed seven cationic MoS_2 and the sensor array comprised of seven electrostatic complexes of cationic MoS_2 and anionic GFP. We assumed, upon addition of the analyte, that the binding will be altered (Figure 1b). It provides an optical response pattern for analysis. Fifteen different proteins having wide variations in molecular weight and the isoelectric point were added to the established array and successfully discriminated at 50 nM with 100% jackknifed classification accuracy. The detection limit of the system was determined as 1 nM by considering macerozyme as the reference protein. The real success of this method mainly relies on detection of

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

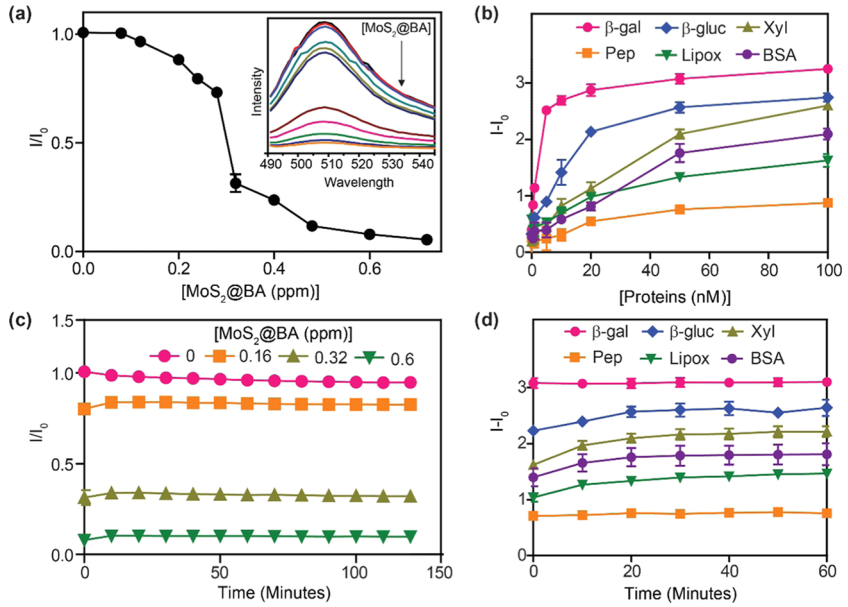


Figure 2. Quenching and regeneration of GFP fluorescence. (a) Fluorescence quenching of the GFP signal upon titration of MoS_2@BA at various concentrations. (b) Fluorescence enhancement of GFP upon decomplexation with MoS_2@BA (0.6 ppm) after addition of protein analytes at different concentrations ranging from 0 to 50 nM. (c) Time-dependent monitoring of GFP complexation and (d) decomplexation with MoS_2@BA (0.6 ppm).

proteins in mixtures as well as in biofluids. To establish the potency of the system, we have used this system in binary mixture proteins, as well as in serum. Ten different proteins in serum media were discriminated at 2.5 μM using four different receptors with 100% classification accuracy, which indicated the robustness of the sensor system. After obtaining high statistical throughput, additionally, here we have explored several machine learning models for pattern recognition in array-based sensing data and their accuracy in predicting an unknown test sample after training.

RESULTS AND DISCUSSION

Characterization and Optimization of the Sensor System. Chemically exfoliated 2D-MoS_2 (Ce-MoS_2) was prepared from its bulk counterpart using the butyl lithium intercalation method.[53] Atomic force microscopy was used for its characterization, which indicates that the bulk material was successfully exfoliated to atomically thick nanosheets (Figure 1d). The exfoliated nanosheets have a diameter range of 300−600 nm and a height ∼1.2 nm (Figure 1d). In recent years, Ce-MoS_2 has been used as a potential platform for biomolecular recognition.[54,55] However, the following are the disadvantages of Ce-MoS_2, which hinders their long-term usage in optical and electrochemical sensing. The major bottlenecks include[56] (1) Ce-MoS_2 being able to undergo agglomeration and settling down in the aqueous solution, (2) lacking dangling bonds or a pi-electron system to facilitate covalent attachment of external probes, and (3) degradation or oxidation of MoS_2 nanosheets upon exposure to moisture or oxygen. The abovementioned demerits can be overcome by surface functionalization, which can be easily achieved through sulfur defects. Exfoliated 2D-MoS_2 contains plenty of sulfur vacancies or defects during exfoliation, which can be easily compensated by sulfur-containing molecules through place-exchange reactions.[57] Seven different cationic thiol ligands were prepared with diverse functional groups at the nonthiol end (Figure 1c). The diversity in the nonthiol end was introduced to facilitate several biomolecular interactions, such as electrostatic (quaternary ammonium group), hydrophobic (C1, C6, C8, and CA), hydrophilic (OL and DOL), and aromatic (BA) interactions (Figure 1c). All of the ligands were synthesized using previously reported procedures. Functionalization of MoS_2 was achieved by stirring the synthesized thiol ligands with freshly exfoliated MoS_2 layers at room temperature for 48 h. Functionalization was confirmed through ζ-potential measurements (Figure 1e).

It indicates that native MoS_2 possesses an average ζ-potential of −30.7 mV; however, after functionalization with seven cationic thiol ligands, the average ζ-potential values shifted to the range of +21.7 to +35.7 mV. The ζ-potential plots belonging to different cationic MoS_2 appeared at nearly similar regions, which means that all of the cationic MoS_2 possess a similar extent of cationic charge after functionalization (Figure S1). This also indicates the consistent functionalization of ligands using MoS_2 nanosheets. Surface functionalization was further confirmed using Fourier transformation infrared (FTIR) spectroscopy and X-ray photoelectron spectroscopy (XPS). FTIR spectra of MoS_2@DOL show characteristic bands corresponding to the C−H stretching frequency at 2840 and 2905 cm^{-1} and the O−H stretching frequency at 3315 cm^{-1} (Figure S2). Similar bands were also present in the DOL ligand at 2850 and 2920 cm^{-1} (aliphatic C−H bands) and 3410 cm^{-1} (O−H band). The S−H band at 2561 cm^{-1} was present for the DOL ligand, which disappeared after it was conjugated to Ce-MoS_2, i.e., in the case of MoS_2@DOL (Figure S2). This indicates that the −SH group was buried into the MoS_2 nanosurfaces after conjugation. In the XPS spectra, the relative ratio of peak intensity for Mo 3d/C 1s/O 1s was 1:0.1:0.12 for Ce-MoS_2 and 1:1.13:1.14 for MoS_2@DOL (Figure S4). The high relative intensities of C 1s and O 1s peaks in MoS_2@DOL as compared to those of Ce-MoS_2 (Figure S4) indicate successful grafting of thiol ligands in MoS_2 nanosurfaces. The thermogravimetric analysis of MoS_2@DOL further supports

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

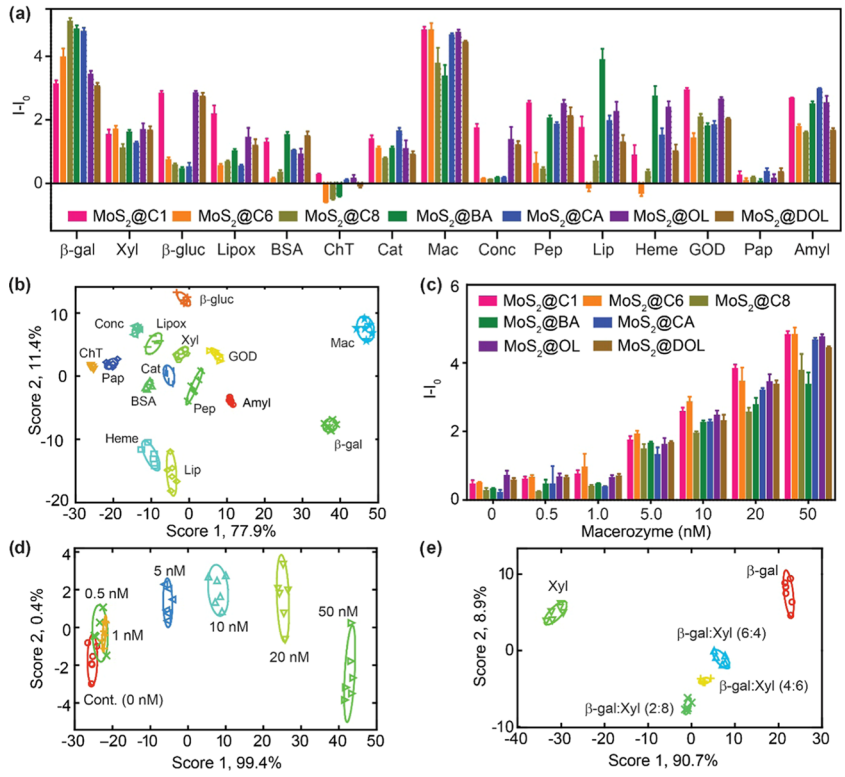


Figure 3. Discrimination of protein analytes in phosphate-buffered saline (PBS) (pH 7.4). (a) Fluorescence response pattern for 15 different proteins against seven cationic MoS_2−GFP conjugates and (b) their corresponding two-dimensional LDA score plots. (c) Fluorescence response pattern for the limit of detection study of macerozyme (Mac) at different concentrations ranging from 0 to 50 nM and (d) the corresponding canonical score plots. (e) Discrimination of a binary mixture of protein analytes (β-galactosidase and xylanase) at different concentration ratios.

the presence of attached thiol ligands (DOL ligands) and indicates the incorporation of ∼45% of the ligand by weight in the surface of Ce-MoS_2 (Figure S3).

Our sensing mechanism was based on a displacement assay and it proceeded with fluorescence quenching of the signal transducer with receptors. We considered GFP as the signal transducer. GFP, a biocompatible marker protein, possesses a β barrel shape with a negative charge (pI = 5.92)[58] and with maximum excitation and emission at 475 and 509 nm, respectively. GFP can form reversible electrostatic complexes with cationic receptors, such as gold nanoparticles,[59] fluorescence-conjugated polymers,[17] graphene oxide,[37] etc., and is eventually used in array-based sensing because of good displacement ability against these surfaces. However, the following criteria were not observed simultaneously by the abovementioned receptors as functionalized 2D-MoS_2 did. The advantages include (1) being synthesized easily through ion intercalation methods with low economic setups, (2) the atomically thin architecture possessing highly active surface areas for supramolecular adsorption, and (3) the surface properties being able to be easily tuned through functionalization. This indicates that 2D-MoS_2 can be considered as a novel receptor, as well as a quencher, in array-based sensing. In the quenching step, seven different cationic MoS_2 at increasing concentrations were titrated against 100 nM GFP. The quenching behavior of GFP with different quenchers was plotted as a function of the quencher concentration. All of the cationic receptors have firm binding abilities to GFP (Figures 2a and S5). Binding constant values were estimated using the Stern−Volmer equation, which indicates that MoS_2@C8 and MoS_2@DOL have the highest and the lowest binding affinities to GFP, respectively (Table S1). Ce-MoS_2 (negatively charged) has a very low binding affinity to GFP as compared to cationic functionalized MoS_2 (Figure S6). This indicates that the binding between cationic MoS_2 and anionic GFP was mainly due to electrostatic interactions. We also checked the nature of binding as a function of time with an interval of 10 min up to 2 h. The fluorescence intensity was consistent after 20 min (Figures 2c and S7). Thus, 20 min was fixed as the incubation time for the study of the quenching mechanism. The concentration of the quencher, i.e., cationic MoS_2 corresponding to >70% quenching of the GFP signal, was considered for further study. Complete quenching of the GFP signal by cationic MoS_2 was avoided because it needs a high concentration of the quenching material where the availability of free fluorophore is less for effective signal regeneration. Also, sometimes, the analyte can bind to GFP (instead of binding to the receptor) to further quench the signal.[60] Due to the abovementioned reasons, cationic MoS_2 corresponding to 100% quenching of the GFP signal was not considered. Our next target was to study the decomplexation or signal regeneration using protein as the analyte. Proteins can adsorb on the surface of cationic MoS_2 by both electrostatic and van der Waals' interactions. To estimate the working concentration of protein analytes, we selected two different cationic MoS_2 complexes (MoS_2@BA and MoS_2@CA) with GFP. After addition of six different proteins at different concentrations (1−100 nM) to both MoS_2@BA (Figure 2b) and MoS_2@CA (Figure S8a) complexes, we noticed that, in all cases, there is a considerable signal recovery at 50 nM (Figures 2b and S8a), which was consequentially chosen as the working concentration for protein sensing. The time-dependent fluorescence

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

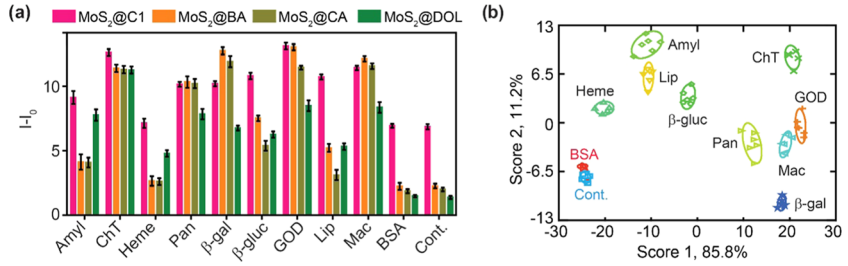


Figure 4. Discrimination of protein analytes at 2.5 μM in the presence of fetal bovine serum. (a) Fluorescence response pattern observed for different proteins against four cationic MoS_2 and (b) their corresponding two-dimensional canonical score plots.

turn-on response shows consistency in signal regeneration after 30 min (Figures 2d and S8b).

Discrimination of Protein Analytes. After optimization of the quenching and the analyte incubation time, we added the protein analyte to predesigned sets of seven different MoS_2−GFP complexes. The proteins that cover a wide range of molecular weight (23.6−540.0 kDa) and pI (2.8−8.7) values were chosen (Table S2). Each protein analyte was repeated six times against each cationic MoS_2−GFP complex. The fluorescence response pattern (Figure 3a) was generated from the $I-I_0$ value, where I and I_0 are the fluorescence intensities of GFP after and before the addition of the protein analytes, respectively. From the fluorescence response pattern, it was clear that β-galactosidase and macerozyme had a superior ability to displace GFP from the cationic MoS_2 nanoenvironment. A total of 15 different proteins were classified using the array containing seven different sensor systems. A fluorescence response matrix with 630 data sets (15 proteins × 7 quenchers × 6 repeats) was subjected to the statistical analysis called linear discriminant analysis (LDA).[61] Seven different scores, viz., score 1 = 77.9, score 2 = 11.4, score 3 = 7.3, score 4 = 2, score 5 = 0.6, score 6 = 0.5, and score 7 = 0.3, were obtained from classical discriminant analysis. Out of these, two major scores (score 1 and score 2) were plotted to obtain two-dimensional LDA score plots (Figure 3b), where all of the proteins could be discriminated with 100% jackknifed classification accuracy. To study the sensitivity of our sensor array, a detection limit study was performed by taking the macerozyme at different concentrations ranging from 0.5 to 50 nM, and the response was plotted (Figure 3c). From the LDA (Figure 3d) score plot, it is evident that significant classification of ellipses was achievable up to 1 nM concentration of macerozyme, and this was considered as the detection limit of the sensor system. The effectiveness of our sensor system toward a binary mixture of protein was verified by considering two different proteins, i.e., β-galactosidase and xylanase, in the ratios 6:4, 4:6, 2:8, which resulted in 100% jackknifed classification accuracy in the LDA score plot (Figure 3e).

The successful detection of a mixture of proteins indicates that the system can be used for detection of protein samples in the presence of complex biological mixtures. For the purpose of real-sample analysis, we selected fetal bovine serum (FBS) as the biological medium. The optimization of the sensor array for protein sensing in serum was different from that in buffer media because serum contains >20 000 different proteins with high concentrations of serum albumins.[62,63] Due to the higher optical density of serum, 500 nM GFP and almost 5 times higher concentrations of cationic MoS_2 (as compared to the previous array) were used against protein sensing in the serum. Four different cationic MoS_2, i.e., MoS_2@C1, MoS_2@BA, MoS_2@CA, and MoS_2@DOL, were taken as receptors and titrated against GFP. MoS_2@CA exhibits the highest binding affinity to GFP among all of the receptors (Figure S9). The serum was spiked with different proteins at a concentration of 2.5 μM and added to the optimized sensor array, and the response was plotted (Figure 4a). Nine different proteins were classified successfully from the control (only serum) and BSA in the LDA score plot (Figure 4b). BSA shares a common region in the LDA plot with the control, which was expected because of the high concentration of the same proteins present in FBS. The efficacy of the sensor array was validated through unknown analysis where 37 out of 40 samples were correctly identified, which corresponds to 92.5% accuracy. The above observation indicates that the optimized array can discriminate imbalances of different proteins, such as hemoglobin, lipase, glucose oxidase, and β-galactosidase, which are present in serum at 273, 14.7, 6.25, and 20 μM concentrations, respectively. Pancreatin, a mixture of different enzymes, such as lipase, amylase, and trypsin, was well-separated in the score plot from its individual compositions (Figure 4b). This indicates that the sensor array is suitable for the analysis of the mixture of proteins in serum. Pancreatin was also well classified in the array conducted in the PBS medium (Figure S10).

Implementation in Various Machine Learning Techniques. The data sets obtained were subjected to different machine learning techniques to obtain the suitable analysis method. The array sensing problem can be viewed as a pattern recognition problem in machine learning, in particular, this consists of a multiclass classification problem. We can formally define this in the following way. We have training data with n = 90 observations, seven predictor variables $X_1,\ldots, X_7$ (which are seven differently functionalized MoS_2), and the target Y (which are the proteins). The target here is a factor variable with 15 levels (15 different proteins).

Performance Metric. The performance measure we have used is accuracy. If the estimated class for ith observation is $\hat{Y}_i$ and the actual class it belongs to is Y_i, then the overall accuracy of the applied method is measured as

$$\theta = \frac{1}{n}\sum_i I(Y_i = \hat{Y}_i)$$

where $I(x = j)$ is the indicator function, which takes the value 1 if $x = j$ and 0 otherwise.

The accuracy thus defined varies within (0,1), with higher values indicating better performance.

Here, we provide a brief description of the four methods, i.e., multinomial logistic regression (MLR), linear discriminant analysis (LDA), K-nearest neighbor (KNN), and the neural

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

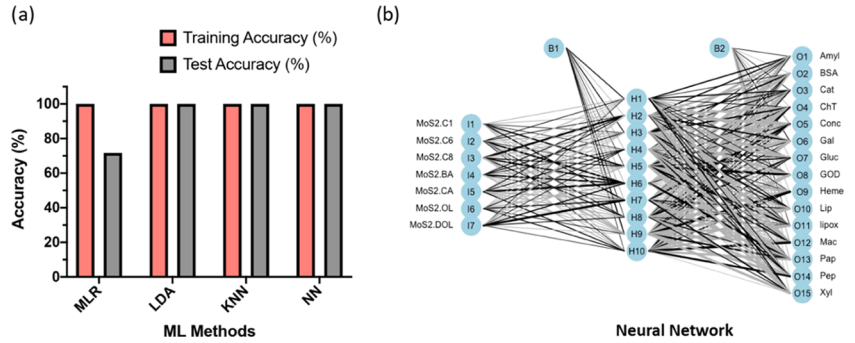


Figure 5. Different ML methods and their performance in analyzing array-based sensing data. (a) Comparison of different ML methods in terms of training accuracy and test accuracy. (b) A simple neural network architecture with one hidden layer and 10 nodes used to analyze the array-based sensing data.

network (NN). In the case of MLR, a binary logistic regression (suppose the classes are 1 and 0) models the logit of the conditional probability of observing 1 given a set of predictors $(x_1,..., x_n)$ as a linear function of the predictors. Denoting $\mathbf{X} = (X_1,..., X_p)$, $\mathbf{x} = (x_1,..., x_n)$

$$\log \frac{p}{1-p} = \beta_0 + \sum_p \beta_p x_p \Leftrightarrow p = \frac{e^{\beta_0 + \sum_p \beta_p x_p}}{1 + e^{\beta_0 + \sum_p \beta_p x_p}}$$

Here, $p = P(Y = 1|\mathbf{X} = \mathbf{x})$

Since our problem is a multiclass (classes 1, 2, ..., K) one, this can be generalized by modeling

$$P(c|\mathbf{X} = \mathbf{x}) = \frac{e^{\beta_{0c} + \sum_p \beta_{pc} x_p}}{\sum_j e^{\beta_{0j} + \sum_p \beta_{pj} x_p}} \quad \forall \quad 1 \le c \le K$$

We can use these probabilities to find which class has the highest probability, including certain observations, given its predictor values.

$$\hat{Y} = \arg \max_{1 \le c \le K} P(Y = c|\mathbf{X} = \mathbf{x})$$

As we can note, this still is a linear model and is thus very simple in nature, which helps in retaining the interpretability of the resulting outcomes.

In the LDA approach, we model the behavior of the predictors in each class and then employ Bayes' theorem to flip these to get the class probability, given the predictor values. More formally, we assume

$$\mathbf{X}|Y = c \sim N\left(\mu_c, \sum\right) \quad \forall \quad 1 \le c \le K$$

We use the training data to estimate the parameters μ_1, μ_2,..., μ_K, $\sum$

Then, we estimate the class probabilities as

$$P(Y = c|\mathbf{X} = \mathbf{x}) = \frac{P(\mathbf{X} = \mathbf{x}|\, Y = c)}{\sum_j P(\mathbf{X} = \mathbf{x}|Y = j)} = \frac{\delta_c(\mathbf{x})}{\sum_j \delta_j(\mathbf{x})} \qquad (1)$$

$$\hat{Y} = \arg \max_{1 \le c \le K} P(Y = c|\mathbf{X} = \mathbf{x})$$

Here, $\delta_c(\mathbf{x})$ is the Gaussian density.

$$\delta_c(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\sum|^{1/2}} e^{-1/2(\mathbf{x}-\mu_c)^T \sum^{-1} (\mathbf{x}-\mu_c)}$$

LDA is an extremely popular method since, while it preserves the probabilistic interpretations of the logistic regression, it actually produces more stable estimates when the classes are well-separated and the n is small (such as in our case).

KNN is a nonparametric method where the class of any observation is estimated based on the k observations closest to it.

$$P(Y = c|\mathbf{X} = \mathbf{x}) = \frac{1}{K} \sum_{i \in \mathcal{N}_k} I(Y_i = c)$$

$$\hat{Y} = \arg \max_{1 \le c \le K} P(Y = c|\mathbf{X} = \mathbf{x})$$

where $\mathcal{N}_k$ denotes the k points closest to $\mathbf{x}$.

We make no assumptions about the shape of the decision boundary and hence KNN is expected to outperform the previous methods when the assumptions therein are not met; however, this comes at the cost of interpretability as the contribution of the individual variable.

A feed-forward NN or a multilayer perceptron (MLP) is simply a model that employs several layers of units (or nodes/perceptrons/neurons) connected through nonlinear activation functions to estimate the data generating process. In particular, for a one-layer NN, the estimated model has the following form

$$\hat{y} = \sigma_2(\sigma_1(x^T w_1 + b_1)^T w_2 + b_2)$$

Here, w_1, w_2, b_1, and b_2 are parameters to be estimated. σ_1 is a typically nonlinear activation function, such as ReLU or sigmoid. For a classification task, such as ours, σ_2 this is taken as a softmax function. The complexity of the model can be further increased by the increasing number of layers and/or nodes in each layer. A NN is a powerful tool to model a complex data generating function. However, this has two major setbacks: it lacks the interpretability, inference of simpler parametric models and this is not suitable for small data sets such as ours. That being said, we include this in our study for comparison (Figure 5a,b).

We run all four of these models on our dataset using the statistical software R, and the results are summarized in Figure 5a. As we can see in Figure 5a, other than the logistic regression, all other models perform ideally. The train and test performances match up to 100%. However, among these, we do prefer the LDA due to its simplicity and explainability, which gives 95% test accuracy (57 out of 60 samples identified correctly). In fact, we can determine how each variable combines to produce the classification boundaries using eq 1. The other two methods lack this and hence would be unable to

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

produce theoretically robust feature importance or uncertainty quantifications.

The proposed sensor array can be an easy and robust method for the discrimination of bioanalytes such as proteins and biomarkers. The interaction between cationic MoS_2 and GFP mimics the protein−protein interactions and is hence successfully implemented for detection of proteins in the buffer as well as in serum media. Earlier, native MoS_2 nanosheets were used in the sensor array for protein discrimination with a detection limit of 500 nM.[42] In the current work, the sensitivity of the system was found to be 1 nM. This indicates the potential recognition ability of the functionalized materials. A similar setup was also utilized earlier for gold nanoparticles.[59] As compared to gold nanoparticles, 2D-MoS_2 possesses highly active surfaces and hence the optimized array is able to discriminate 15 different proteins (50 nM) in the buffer. In the presence of serum media, using only four receptors, 10 proteins (2.5 μM) were successfully classified, which exhibits higher sensitivity compared to functionalized nanoparticles. Previously, graphene oxide, a well-known receptor for biomolecular recognition, was used in array-based sensing against GFP as one of the signal transducers.[37] Although the sensitivity of the system was similar to that of this work, the ability of the sensor array in a biological medium was not examined. In another study, graphene oxide was used as a quencher against fluorescent aptamers,[38] where the working concentration in serum medium was maintained at 5 μM for detection of five different proteins; however, the cationic MoS_2-based sensor array could discriminate 10 different proteins at 2.5 μM. Along with this, charge tunability through surface modification was very tedious. Hence, in the sensor array, graphene oxide was fixed as the receptor unit, and the cross-reactivity was achieved by changing the signal transducers. But for molecular recognition, modifying the surface according to the analyte on demand is very crucial. Such a thing can be easily achieved through 2D-MoS_2. In addition, implementation of the machine learning technique for unknown sample analysis helped to achieve better accuracy than that of manual statistical analysis, which was earlier not used in 2D nanosensor arrays. In a nutshell, a combination of the 2D nanosurface and easy functionalization of MoS_2 made it an efficient sensor in pattern-based recognition with the assistance of machine learning techniques.

CONCLUSIONS

In summary, we have developed a sensor array comprising two-dimensional cationic MoS_2 as cross-reactivity receptors and GFP as signal transducers. The versatility in binding affinity toward GFP and analytes was achieved through functionalization with cationic thiol ligands bearing variable nonthiol head groups. The array consisting of seven different receptors could successfully discriminate between 15 protein analytes at 50 nM concentration with a detection limit of 1 nM. The system was also extended to detect protein analytes in a binary mixture and in a serum sample. The high response from negatively charged proteins was attributed to the electrostatic nature of the receptors (cationic MoS_2), and this advantage can be successfully exploited for the detection of other negatively charged analytes, viz., bacteria, cell lines, cell lysates, etc. Additionally, we have shown that for a small array-based sensing dataset, both simple and complex machine learning can be equally effective in pattern recognition. It has been well-reported in the literature that incorporating complex ML analysis into large high-dimensional data sets generated by array-based sensing platforms in real-world scenarios results in significant advantages over statistical analysis and leads to smart and adaptable sensor systems.[64] For a typical 6-replicate data, a linear method such as LDA suffices, but if the dataset becomes larger in the case of real-world application scenarios, neural networks are likely to perform better in handling nuances in a large dataset.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsanm.1c00244.

Quenching and the regeneration plot of GFP; fluorescence response pattern; ζ-potential plot; tables containing binding constants; Mahalanobis distances; protein properties and dataset for analysis of analytes and an unknown sample (PDF)

R code used for machine learning (TXT)

AUTHOR INFORMATION

Corresponding Author

Mrinmoy De − *Department of Organic Chemistry, Indian Institute of Science, Bangalore 560012, India;* orcid.org/0000-0001-8394-9059; Email: md@iisc.ac.in

Authors

Pradipta Behera − *Department of Organic Chemistry, Indian Institute of Science, Bangalore 560012, India*

Krishna Kumar Singh − *Vascular Biology Center, Augusta University, Augusta, Georgia 30912, United States; Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore 560012, India*

Subhendu Pandit − *Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0002-4542-2069

Diptarka Saha − *Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

Deepak Kumar Saini − *Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore 560012, India;* orcid.org/0000-0001-6671-7256

Complete contact information is available at: https://pubs.acs.org/10.1021/acsanm.1c00244

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank BRNS (37(2)/14/07/2017-BRNS) and SERB (CVD/2020/000855) for financial support. P.B. thanks IISc and CSIR for the doctoral fellowship. Financial support as a seed grant from the Indian Institute of Science for the Digital Health project under the Institute of Eminence funds provided by the Government of India, is gratefully acknowledged.

REFERENCES

(1) Chatterjee, S. K.; Zetter, B. R. Cancer Biomarkers: Knowing the Present and Predicting the Future. *Future Oncol.* **2005**, *1*, 37−50.

(2) Peng, F.; He, J.; Loo, J. F. C.; Kong, S. K.; Li, B.; Gu, D. Identification of Serum Micrornas as Diagnostic Biomarkers for Influenza H7N9 Infection. *Virol. Rep.* **2017**, *7*, 1−8.
(3) Speranza, E.; Bixler, S. L.; Altamura, L. A.; Arnold, C. E.; Pratt, W. D.; Taylor-Howell, C.; Burrows, C.; Aguilar, W.; Rossi, F.; Shamblin, J. D.; Wollen, S. E.; Zelko, J. M.; Minogue, T.; Nagle, E.; Palacios, G.; Goff, A. J.; Connor, J. H. A Conserved Transcriptional Response to Intranasal Ebola Virus Exposure in Nonhuman Primates Prior to Onset of Fever. *Sci. Transl. Med.* **2018**, *10*, No. eaaq1016.
(4) Nixon, D. E.; Landay, A. L. Biomarkers of Immune Dysfunction in Hiv. *Curr. Opin. HIV AIDS* **2010**, *5*, 498−503.
(5) Malik, P.; Patel, U.; Mehta, D.; Patel, N.; Kelkar, R.; Akrmah, M.; Gabrilove, J. L.; Sacks, H. Biomarkers and Outcomes of Covid-19 Hospitalisations: Systematic Review and Meta-Analysis. *BMJ* **2020**, No. bmjebm-2020.
(6) Vasan, R. S. Biomarkers of Cardiovascular Disease. *Circulation* **2006**, *113*, 2335−2362.
(7) Mölleken, C.; Sitek, B.; Henkel, C.; Poschmann, G.; Sipos, B.; Wiese, S.; Warscheid, B.; Broelsch, C.; Reiser, M.; Friedman, S. L.; Tornøe, I.; Schlosser, A.; Klöppel, G.; Schmiegel, W.; Meyer, H. E.; Holmskov, U.; Stühler, K. Detection of Novel Biomarkers of Liver Cirrhosis by Proteomic Analysis. *Hepatology* **2009**, *49*, 1257−1266.
(8) Drakopanagiotakis, F.; Wujak, L.; Wygrecka, M.; Markart, P. Biomarkers in Idiopathic Pulmonary Fibrosis. *Matrix Biol.* **2018**, *68*−*69*, 404−421.
(9) Rachakonda, V.; Pan, T. H.; Le, W. D. Biomarkers of Neurodegenerative Disorders: How Good Are They? *Cell Res.* **2004**, *14*, 349−358.
(10) Haab, B. B. Applications of Antibody Array Platforms. *Curr. Opin. Biotechnol.* **2006**, *17*, 415−421.
(11) Chan, D. W.; Wang, Y. Y.; Rosenzweig, J.; Zhang, Z.; Li, J. Proteomics and Bioinformatics Approaches for Identification of Serum Biomarkers to Detect Breast Cancer. *Clin. Chem.* **2002**, *48*, 1296−1304.
(12) Miranda, O. R.; You, C.-C.; Phillips, R.; Kim, I.-B.; Ghosh, P. S.; Bunz, U. H. F.; Rotello, V. M. Array-Based Sensing of Proteins Using Conjugated Polymers. *J. Am. Chem. Soc.* **2007**, *129*, 9856−9857.
(13) Yan, P.; Li, X.; Dong, Y.; Li, B.; Wu, Y. A Ph-Based Sensor Array for the Detection and Identification of Proteins Using Cdse/Zns Quantum Dots as an Indicator. *Analyst* **2019**, *144*, 2891−2897.
(14) Han, J. S.; Cheng, H. R.; Wang, B. H.; Braun, M. S.; Fan, X. B.; Bender, M.; Huang, W.; Domhan, C.; Mier, W.; Lindner, T.; Seehafer, K.; Wink, M.; Bunz, U. H. F. A Polymer/Peptide Complex-Based Sensor Array That Discriminates Bacteria in Urine. *Angew. Chem., Int. Ed.* **2017**, *56*, 15246−15251.
(15) Tao, Y.; Ran, X.; Ren, J. S.; Qu, X. G. Array-Based Sensing of Proteins and Bacteria by Using Multiple Luminescent Nanodots as Fluorescent Probes. *Small* **2014**, *10*, 3667−3671.
(16) Das Saha, N.; Sasmal, R.; Meethal, S. K.; Vats, S.; Gopinathan, P. V.; Jash, O.; Manjithaya, R.; Gagey-Eilstein, N.; Agasti, S. S. Multichannel DNA Sensor Array Fingerprints Cell States and Identifies Pharmacological Effectors of Catabolic Processes. *ACS Sens.* **2019**, *4*, 3124−3132.
(17) Rana, S.; Elci, S. G.; Mout, R.; Singla, A. K.; Yazdani, M.; Bender, M.; Bajaj, A.; Saha, K.; Bunz, U. H. F.; Jirik, F. R.; Rotello, V. M. Ratiometric Array of Conjugated Polymers-Fluorescent Protein Provides a Robust Mammalian Cell Sensor. *J. Am. Chem. Soc.* **2016**, *138*, 4522−4529.
(18) Buryak, A.; Severin, K. A. Chemosensor Array for the Colorimetric Identification of 20 Natural Amino Acids. *J. Am. Chem. Soc.* **2005**, *127*, 3700−3701.
(19) Behera, P.; Mohanty, A.; De, M. Functionalized Fluorescent Nanodots for Discrimination of Nitroaromatic Compounds. *ACS Appl. Nano Mater.* **2020**, *3*, 2846−2856.
(20) Huang, W.; Bender, M.; Seehafer, K.; Wacker, I.; Schröder, R. R.; Bunz, U. H. F. A. Tetraphenylethene-Based Polymer Array Discriminates Nitroarenes. *Macromolecules* **2018**, *51*, 1345−1350.
(21) Ghasemi, F.; Hormozi-Nezhad, M. R. Determination and Identification of Nitroaromatic Explosives by a Double-Emitter Sensor Array. *Talanta* **2019**, *201*, 230−236.
(22) Abbasi-Moayed, S.; Bigdeli, A.; Hormozi-Nezhad, M. R. Application of Nayf4:Yb/Er/Tm Ucnps in Array-Based Sensing of Neurotransmitters: From a Single Particle to a Multichannel Sensor Array. *ACS Appl. Mater. Interfaces* **2020**, *12*, 52976−52982.
(23) Jafarinejad, S.; Bigdeli, A.; Ghazi-Khansari, M.; Sasanpour, P.; Hormozi-Nezhad, M. R. Identification of Catecholamine Neurotransmitters Using a Fluorescent Electronic Tongue. *ACS Chem. Neurosci.* **2020**, *11*, 25−33.
(24) Abbasi-Moayed, S.; Hormozi-Nezhad, M. R.; Maaza, M. A Multichannel Single-Well Sensor Array for Rapid and Visual Discrimination of Catecholamine Neurotransmitters. *Sens. Actuators, B* **2019**, *296*, No. 126691.
(25) Wu, Y.; Liu, X.; Wu, Q.; Yi, J.; Zhang, G. Carbon Nanodots-Based Fluorescent Turn-on Sensor Array for Biothiols. *Anal. Chem.* **2017**, *89*, 7084−7089.
(26) Abbasi-Moayed, S.; Golmohammadi, H.; Bigdeli, A.; Hormozi-Nezhad, M. R. A Rainbow Ratiometric Fluorescent Sensor Array on Bacterial Nanocellulose for Visual Discrimination of Biothiols. *Analyst* **2018**, *143*, 3415−3424.
(27) Liu, Y.; Duan, Y.; Gill, A. D.; Perez, L.; Jiang, Q.; Hooley, R. J.; Zhong, W. Metal-Assisted Selective Recognition of Biothiols by a Synthetic Receptor Array. *Chem. Commun.* **2018**, *54*, 13147−13150.
(28) Pandit, S.; Banerjee, T.; Srivastava, I.; Nie, S.; Pan, D. Machine Learning-Assisted Array-Based Biomolecular Sensing Using Surface-Functionalized Carbon Dots. *ACS Sens.* **2019**, *4*, 2730−2737.
(29) Pandit, S.; Behera, P.; Sahoo, J.; De, M. In Situ Synthesis of Amino Acid Functionalized Carbon Dots with Tunable Properties and Their Biological Applications. *ACS Appl. Bio Mater.* **2019**, *2*, 3393−3403.
(30) Wu, Y.; Tan, Y.; Wu, J.; Chen, S.; Chen, Y. Z.; Zhou, X.; Jiang, Y.; Tan, C. Fluorescence Array-Based Sensing of Metal Ions Using Conjugated Polyelectrolytes. *ACS Appl. Mater. Interfaces* **2015**, *7*, 6882−6888.
(31) Bunz, U. H.; Zhang, H.; Wang, B.; Seehafer, K. Sensor Array Based Determination of Edman Degradated Amino Acids Using Poly(P-Phenyleneethynylene)S. *Chem. - Eur. J.* **2020**, *26*, 7779−7782.
(32) Bigdeli, A.; Ghasemi, F.; Golmohammadi, H.; Abbasi-Moayed, S.; Nejad, M. A. F.; Fahimi-Kashani, N.; Jafarinejad, S.; Shahrajabian, M.; Hormozi-Nezhad, M. R. Nanoparticle-Based Optical Sensor Arrays. *Nanoscale* **2017**, *9*, 16546−16563.
(33) Miranda, O. R.; Creran, B.; Rotello, V. M. Array-Based Sensing with Nanoparticles: 'Chemical Noses' for Sensing Biomolecules and Cell Surfaces. *Curr. Opin. Chem. Biol.* **2010**, *14*, 728−736.
(34) Agasti, S. S.; Rana, S.; Park, M. H.; Kim, C. K.; You, C. C.; Rotello, V. M. Nanoparticles for Detection and Diagnosis. *Adv. Drug Delivery Rev.* **2010**, *62*, 316−328.
(35) Yang, H.; Lu, F.; Sun, Y.; Yuan, Z.; Lu, C. Fluorescent Gold Nanocluster-Based Sensor Array for Nitrophenol Isomer Discrimination Via an Integration of Host−Guest Interaction and Inner Filter Effect. *Anal. Chem.* **2018**, *90*, 12846−12853.
(36) Behera, P.; De, M. Nano-Graphene Oxide Based Multichannel Sensor Arrays Towards Sensing of Protein Mixtures. *Chem. - Asian J.* **2019**, *14*, 553−560.
(37) Chou, S. S.; De, M.; Luo, J. Y.; Rotello, V. M.; Huang, J. X.; Dravid, V. P. Nanoscale Graphene Oxide (nGO) as Artificial Receptors: Implications for Biomolecular Interactions and Sensing. *J. Am. Chem. Soc.* **2012**, *134*, 16725−16733.
(38) Pei, H.; Li, J.; Lv, M.; Wang, J. Y.; Gao, J. M.; Lu, J. X.; Li, Y. P.; Huang, Q.; Hu, J.; Fan, C. H.; Graphene-Based, A. Sensor Array for High-Precision and Adaptive Target Identification with Ensemble Aptamers. *J. Am. Chem. Soc.* **2012**, *134*, 13843−13849.
(39) Qi, Y.; Xia, T.; Li, Y.; Duan, L.; Chen, W. Colloidal Stability of Reduced Graphene Oxide Materials Prepared Using Different Reducing Agents. *Environ. Sci. Nano* **2016**, *3*, 1062−1071.

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851

(40) Azizighannad, S.; Mitra, S. Stepwise Reduction of Graphene Oxide (Go) and Its Effects on Chemical and Colloidal Properties. *Sci. Rep.* **2018**, *8*, No. 10083.
(41) Zhang, F.; Lu, C. W.; Wang, M.; Yu, X. S.; Wei, W. L.; Xia, Z. N. A Chiral Sensor Array for Peptidoglycan Biosynthesis Monitoring Based on Mos2 Nanosheet-Supported Host-Guest Recognitions. *ACS Sens.* **2018**, *3*, 304−312.
(42) Hizir, M. S.; Robertson, N. M.; Balcioglu, M.; Alp, E.; Rana, M.; Yigit, M. V. Universal Sensor Array for Highly Selective System Identification Using Two-Dimensional Nanoparticles. *Chem. Sci.* **2017**, *8*, 5735−5745.
(43) Eda, G.; Yamaguchi, H.; Voiry, D.; Fujita, T.; Chen, M.; Chhowalla, M. Photoluminescence from Chemically Exfoliated MoS_2. *Nano Lett.* **2011**, *11*, 5111−5116.
(44) Eda, G.; Fujita, T.; Yamaguchi, H.; Voiry, D.; Chen, M. W.; Chhowalla, M. Coherent Atomic and Electronic Heterostructures of Single-Layer MoS_2. *ACS Nano* **2012**, *6*, 7311−7317.
(45) Yu, Z.; Pan, Y.; Shen, Y.; Wang, Z.; Ong, Z.-Y.; Xu, T.; Xin, R.; Pan, L.; Wang, B.; Sun, L.; Wang, J.; Zhang, G.; Zhang, Y. W.; Shi, Y.; Wang, X. Towards Intrinsic Charge Transport in Monolayer Molybdenum Disulfide by Defect and Interface Engineering. *Nat. Commun.* **2014**, *5*, No. 5290.
(46) Makarova, M.; Okawa, Y.; Aono, M. Selective Adsorption of Thiol Molecules at Sulfur Vacancies on Mos2(0001), Followed by Vacancy Repair Via S−C Dissociation. *J. Phys. Chem. C* **2012**, *116*, 22411−22416.
(47) Rana, S.; Singla, A. K.; Bajaj, A.; Elci, S. G.; Miranda, O. R.; Mout, R.; Yan, B.; Jirik, F. R.; Rotello, V. M. Array-Based Sensing of Metastatic Cells and Tissues Using Nanoparticle−Fluorescent Protein Conjugates. *ACS Nano* **2012**, *6*, 8233−8240.
(48) Bajaj, A.; Rana, S.; Miranda, O. R.; Yawe, J. C.; Jerry, D. J.; Bunz, U. H. F.; Rotello, V. M. Cell Surface-Based Differentiation of Cell Types and Cancer States Using a Gold Nanoparticle-GFP Based Sensing Array. *Chem. Sci.* **2010**, *1*, 134−138.
(49) Karunakaran, S.; Pandit, S.; De, M. Functionalized Two-Dimensional MoS_2 with Tunable Charges for Selective Enzyme Inhibition. *ACS Omega* **2018**, *3*, 17532−17539.
(50) Ali, S. R.; Pandit, S.; De, M. 2D-MoS_2-Based β-Lactamase Inhibitor for Combination Therapy against Drug-Resistant Bacteria. *ACS Appl. Bio Mater.* **2018**, *1*, 967−974.
(51) Karunakaran, S.; Pandit, S.; Basu, B.; De, M. Simultaneous Exfoliation and Functionalization of 2H-MoS_2 by Thiolated Surfactants: Applications in Enhanced Antibacterial Activity. *J. Am. Chem. Soc.* **2018**, *140*, 12634−12644.
(52) Pandit, S.; Karunakaran, S.; Boda, S. K.; Basu, B.; De, M. High Antibacterial Activity of Functionalized Chemically Exfoliated MoS_2. *ACS Appl. Mater. Interfaces* **2016**, *8*, 31567−31573.
(53) Fan, X. B.; Xu, P. T.; Zhou, D. K.; Sun, Y. F.; Li, Y. G. C.; Nguyen, M. A. T.; Terrones, M.; Mallouk, T. E. Fast and Efficient Preparation of Exfoliated 2H MoS_2 Nanosheets by Sonication-Assisted Lithium Intercalation and Infrared Laser-Induced 1T to 2H Phase Reversion. *Nano Lett.* **2015**, *15*, 5956−5960.
(54) Kalantar-zadeh, K.; Ou, J. Z.; Daeneke, T.; Strano, M. S.; Pumera, M.; Gras, S. L. Two-Dimensional Transition Metal Dichalcogenides in Biosystems. *Adv. Funct. Mater.* **2015**, *25*, 5086−5099.
(55) Kalantar-zadeh, K.; Ou, J. Z. Biosensors Based on Two-Dimensional MoS_2. *ACS Sens.* **2016**, *1*, 5−16.
(56) Hu, H.; Zavabeti, A.; Quan, H.; Zhu, W.; Wei, H.; Chen, D.; Ou, J. Z. Recent Advances in Two-Dimensional Transition Metal Dichalcogenides for Biological Sensing. *Biosens. Bioelectron.* **2019**, *142*, No. 111573.
(57) Chou, S. S.; De, M.; Kim, J.; Byun, S.; Dykstra, C.; Yu, J.; Huang, J. X.; Dravid, V. P. Ligand Conjugation of Chemically Exfoliated Mos2. *J. Am. Chem. Soc.* **2013**, *135*, 4584−4587.
(58) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.
(59) De, M.; Rana, S.; Akpinar, H.; Miranda, O. R.; Arvizo, R. R.; Bunz, U. H. F.; Rotello, V. M. Sensing of Proteins in Human Serum Using Conjugates of Nanoparticles and Green Fluorescent Protein. *Nat. Chem.* **2009**, *1*, 461−465.
(60) You, C.-C.; Miranda, O. R.; Gider, B.; Ghosh, P. S.; Kim, I.-B.; Erdogan, B.; Krovi, S. A.; Bunz, U. H. F.; Rotello, V. M. Detection and Identification of Proteins Using Nanoparticle−Fluorescent Polymer 'Chemical Nose' Sensors. *Nat. Nanotechnol.* **2007**, *2*, 318−323.
(61) Jurs, P. C.; Bakken, G. A.; McClelland, H. E. Computational Methods for the Analysis of Chemical Sensor Array Data from Volatile Analytes. *Chem. Rev.* **2000**, *100*, 2649−2678.
(62) Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. Toward a Human Blood Serum Proteome—Analysis by Multidimensional Separation Coupled with Mass Spectrometry. *Mol. Cell. Proteom.* **2002**, *1*, 947−955.
(63) Pieper, R.; Gatlin, C. L.; Makusky, A. J.; Russo, P. S.; Schatz, C. R.; Miller, S. S.; Su, Q.; McGrath, A. M.; Estock, M. A.; Parmar, P. P.; Zhao, M.; Huang, S. T.; Zhou, J.; Wang, F.; Esquer-Blasco, R.; Anderson, N. L.; Taylor, J.; Steiner, S. The Human Serum Proteome: Display of Nearly 3700 Chromatographically Separated Protein Spots on Two-Dimensional Electrophoresis Gels and Identification of 325 Distinct Proteins. *Proteomics* **2003**, *3*, 1345−1364.
(64) Ha, N.; Xu, K.; Ren, G.; Mitchell, A.; Ou, J. Z. Machine Learning-Enabled Smart Sensor Systems. *Adv. Intell. Syst.* **2020**, *2*, No. 2000063.

https://doi.org/10.1021/acsanm.1c00244
ACS Appl. Nano Mater. 2021, 4, 3843−3851