Machine Learning-Assisted Array-Based Detection of Proteins in Serum Using Functionalized MoS2 Nanosheets and Green Fluorescent Protein Conjugates

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

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)-MoS2 nanosheets and green fluorescent protein (GFP) as the receptor and the signal transducer, respectively. Two-dimensional MoS2 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-MoS2−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 MoS2, 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 cancer1, influenza,2 Ebola Virus Disease3 (EVD), HIV,4 coronavirus disease (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 assay10 (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 analysis11 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 proteins12,13 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
fluorescent conjugated polymers, metal nanoparticles, 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. 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. 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 MoS2, an inorganic version of graphene, although it is rarely explored in array-based sensing.

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-MoS2 as a building block for receptor units. Two-dimensional nanosurfaces of MoS2 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-MoS2 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. 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. Hence, in this work, we have used surface functionalization as the main tool to achieve cross reactivity. A native receptor (2D-MoS2) 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. Previously, functionalized 2D-MoS2 proved as an alternative to gold nanoparticles in several applications such as enzyme inhibition, antimicrobial activity, 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 MoS2. GFP can interact with cationic MoS2, 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 MoS2 and the sensor array comprised of seven electrostatic complexes of cationic MoS2 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...
unknown test sample after training. array-based sensing data and their accuracy in predicting a several machine learning models for pattern recognition in statistical throughput, additionally, here we have explored the robustness of the sensor system. After obtaining high receptors with 100% classiifiication accuracy, which indicated

53 Atomic force microscopy was used for prepared from its bulk counterpart using the butyl lithium (0.6 ppm).

Ce-MoS2 being able to undergo agglomeration and settling of Ce-MoS2, which hinders their long-term usage in optical and MoS2 has been used as a potential platform for biomolecular recognition. However, the following are the disadvantages MoS2 nanosheets. Surface functionalization was further confirmed using Fourier transform-infrared (FTIR) spectroscopy and X-ray photoelectron spectroscopy (XPS). FTIR spectra of MoS2@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$^{-1}$ (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-MoS2, i.e., in the case of MoS2@DOL (Figure S2). This indicates that the –SH group was buried into the MoS2 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-MoS2 and 1:1.13:1.14 for MoS2@DOL (Figure S4). The high relative intensities of C 1s and O 1s peaks in MoS2@DOL as compared to those of Ce-MoS2 (Figure S4) indicate successful grafting of thiol ligands in MoS2 nanosurfaces. The thermogravimetric analysis of MoS2@DOL further supports the characterization and Optimization of the Sensor System. Chemically exfoliated 2D-MoS2 (Ce-MoS2) was prepared from its bulk counterpart using the butyl lithium intercalation method. 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-MoS2 has been used as a potential platform for biomolecular recognition. However, the following are the disadvantages of Ce-MoS2, which hinders their long-term usage in optical and electrochemical sensing. The major bottlenecks include (1) Ce-MoS2 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 MoS2 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-MoS2 contains plenty of sulfur vacancies or defects during exfoliation, which can be easily compensated by sulfur-containing molecules through place-exchange reactions. 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 MoS2 was achieved by stirring the synthesized thiol ligands with freshly exfoliated MoS2 layers at room temperature for 48 h. Functionalization was confirmed through $\zeta$-potential measurements (Figure 1e).

It indicates that native MoS2 possesses an average $\zeta$-potential of ~30.7 mV; however, after functionalization with seven cationic thiol ligands, the average $\zeta$-potential values shifted to the range of +21.7 to +35.7 mV. The $\zeta$-potential plots belonging to different cationic MoS2 appeared at nearly similar regions, which means that all of the cationic MoS2 possess a similar extent of cationic charge after functionalization (Figure S1). This also indicates the consistent functionalization of ligands using MoS2 nanosheets. Surface functionalization was further confirmed using Fourier transform-infrared (FTIR) spectroscopy and X-ray photoelectron spectroscopy (XPS). FTIR spectra of MoS2@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-MoS2, i.e., in the case of MoS2@DOL (Figure S2). This indicates that the –SH group was buried into the MoS2 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-MoS2 and 1:1.13:1.14 for MoS2@DOL (Figure S4). The high relative intensities of C 1s and O 1s peaks in MoS2@DOL as compared to those of Ce-MoS2 (Figure S4) indicate successful grafting of thiol ligands in MoS2 nanosurfaces.

Figure 2. Quenching and regeneration of GFP fluorescence. (a) Fluorescence quenching of the GFP signal upon titration of MoS2@BA at various concentrations. (b) Fluorescence enhancement of GFP upon decolouration with MoS2@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) decolouration with MoS2@BA (0.6 ppm).
the presence of attached thiol ligands (DOL ligands) and indicates the incorporation of ~45% of the ligand by weight in the surface of Ce-MoS2 (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-MoS2 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-MoS2 can be considered as a novel receptor, as well as a quencher, in array-based sensing. Proteins can adsorb on the surface of cationic MoS2 by both electrostatic and van der Waals’ interactions. To estimate the working concentration of protein analytes, we selected two different cationic MoS2 complexes (MoS2@BA and MoS2@CA) with GFP. After addition of six different proteins at different concentrations (1–100 nM) to both MoS2@BA (Figure 2b) and MoS2@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

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 MoS2−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.
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 MoS₂−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 MoS₂−GFP complex. The fluorescence response pattern (Figure 3a) was generated from the I−I₀ value, where I and I₀ 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 MoS₂ 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). High discrimination 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. Due to the higher optical density of serum, 500 nM GFP and almost 5 times higher concentrations of cationic MoS₂ (as compared to the previous array) were used against protein sensing in the serum. Four different cationic MoS₂, i.e., MoS₂@C1, MoS₂@BA, MoS₂@CA, and MoS₂@DOL, were taken as receptors and titrated against GFP. MoS₂@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, 62.5, 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 X₁,…, X₇ (which are seven differently functionalized MoS₂), 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 Ŷᵢ and the actual class it belongs to is Yᵢ, 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
This can be generalized by modeling $\mu_k = \beta_0^k + \sum_{j} \beta_j x_j$ as a linear function of the predictors. Denoting $X = (X_1, \ldots, X_p)$, $x = (x_1, \ldots, x_p)$

$$
\log \frac{p}{1-p} = \beta_0^k + \sum_{j} \beta_j x_j \Rightarrow p = \frac{e^{\beta_0^k + \sum_{j} \beta_j x_j}}{1 + e^{\beta_0^k + \sum_{j} \beta_j x_j}}
$$

Here, $p = P(Y = 1|X = x)$

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

$$
P(c|X = x) = \frac{e^{\beta_0^c + \sum_{j} \beta_j x_j}}{\sum_{c} e^{\beta_0^c + \sum_{j} \beta_j x_j}} \forall 1 \leq c \leq K
$$

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

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

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 its predictor values. More formally, we assume

$$
\mathbf{X} Y = \mathbf{c} \sim \mathcal{N}(\mu_c, \sum) \forall 1 \leq c \leq K
$$

We use the training data to estimate the parameters $\mu_1$, $\mu_2$, ..., $\mu_K$, $\sum$.

Then, we estimate the class probabilities as

$$
P(Y = c|X = x) = \frac{p(X = x|Y = c)}{\sum_p p(X = x|Y = c)} = \frac{\delta_c(x)}{\sum \delta_j(x)}
$$

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

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

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

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.

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

$$
\hat{Y} = \arg \max_{1 \leq i \leq K} P(Y = i|X = x)
$$

where $N_k$ denotes the $k$ points closest to $x$.

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, $\sigma_1$, $\sigma_2$, $w_1$, $b_1$, and $b_2$ are parameters to be estimated. $\sigma_1$ is a typically nonlinear activation function, such as ReLU or sigmoid. For a classification task, such as ours, $\sigma_2$ 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
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. 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. 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. 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, 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. 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**

Pradiptra 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.
(46) Azzizighannad, S.; Mitra, S. Stepwise Reduction of Graphene Oxide (Go) and Its Effects on Chemical and Colloidal Properties. *Sci. Rep.* 2018, 8, No. 10083.
(47) 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.
(48) 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.
(49) Eda, G.; Yamaguchi, H.; Voiry, D.; Fujita, T.; Chen, M.; Chhowalla, M. Photoluminescence from Chemically Exfoliated MoS2. *Nano Lett.* 2011, 11, 5111–5116.
(50) Eda, G.; Fujita, T.; Yamaguchi, H.; Voiry, D.; Chen, M. W.; Chhowalla, M. Coherent Atomic and Electronic Heterostructures of Single-Layer MoS2. *ACS Nano* 2012, 6, 7311–7317.
(51) 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.
(52) 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.
(53) 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.
(54) 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.
(55) Karunakaran, S.; Pandit, S.; De, M. Functionalized Two-Dimensional MoS2 with Tunable Charges for Selective Enzyme Inhibition. *ACS Omega* 2018, 3, 17532–17539.
(56) Ali, S. R.; Pandit, S.; De, M. 2D-MoS2-Based β-Lactamase Inhibitor for Combination Therapy against Drug-Resistant Bacteria. *ACS Appl. Bio Mater.* 2018, 1, 967–974.
(57) Karunakaran, S.; Pandit, S.; Basu, B.; De, M. Simultaneous Exfoliation and Functionalization of 2H-MoS2 by Thiolated Surfactants: Applications in Enhanced Antibacterial Activity. *J. Am. Chem. Soc.* 2018, 140, 12634–12644.
(58) Pandit, S.; Karunakaran, S.; Boda, S. K.; Basu, B.; De, M. High Antibacterial Activity of Functionalized Chemically Exfoliated MoS2. *ACS Appl. Mater. Interfaces* 2016, 8, 31567–31573.
(59) 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 MoS2 Nanosheets by Sonication-Assisted Lithium Intercalation and Infrared Laser-Induced 1T to 2H Phase Reversion. *Nano Lett.* 2015, 15, 5956–5960.
(60) 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.
(61) Kalantar-zadeh, K.; Ou, J. Z. Biosensors Based on Two-Dimensional MoS2. *ACS Sens.* 2016, 1, 5–16.
(62) 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.
(63) 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.
(64) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* 1998, 67, 509–544.
(65) De, M.; Rana, S.; Ákpinar, H.; Miranda, O. R.; Arvizu, 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.
(66) You, C.-C.; Miranda, O. R.; Gider, B.; Ghosh, P. S.; Kim, L.-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.
(67) 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.
(68) 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.
(69) 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.
(70) Ha, N.; Xu, K.; Ren, G.; Mitchell, A.; Ou, J. Z. Machine Learning-Enabled Smart Sensor Systems. *Adv. Intell. Syst.* 2020, 2, No. 2000063.