


Cite this: *Nanoscale*, 2025, **17**, 7957

Gold nanorods as multidimensional optical nanomaterials: machine learning-enhanced quantitative fingerprinting of proteins for diagnostic applications†

Afsaneh Orouji,^{‡a} Mahdi Ghamsari,^{‡a} Samira Abbasi-Moayed,^b Mahmood Akbari,^c Malik Maaza^c and Mohammad Reza Hormozi-Nezhad  ^{*a}

The rapid and precise quantification and identification of proteins as key diagnostic biomarkers hold significant promise in allergy testing, disease diagnosis, clinical treatment, and proteomics. This is crucial because alterations in disease-associated genetic information during pathogenesis often result in changes in protein types and levels. Therefore, the design of portable, fast, user-friendly, and affordable sensing platforms rather than a single-sensor-per-analyte strategy for multiplex protein detection is quite consequential. In the present research, a robust multicolorimetric probe based on the inhibited etching of gold nanorods (AuNRs) allowing unambiguous high-performance visual and spectral quantification and identification of proteins in human urine samples was designed. Most recently, we discovered that *N*-bromosuccinimide (NBS) can quickly etch AuNRs with a distinct color change, allowing convenient and accurate visual recognition of all amino acids. Herein, further explorations revealed that the presence of proteins, as amino acids' polymers, reduces the effective concentration of NBS to different amounts and in turn prevents the etching of AuNRs to various degrees, thereby allowing precise quantification and identification of various proteins ranging from phosphatase (ACP), pepsin (Pep), hemoglobin (Hem), and transferrin (TRF) to immunoglobulin G (IgG), lysozyme (Lys), fibrinogen (Fib), and human serum albumin (HSA). The acquired dataset was statistically analyzed using linear discriminant analysis (LDA), partial least-squares regression (PLSR), and hierarchical cluster analysis (HCA) to accurately classify and identify individual proteins and their combinations at various levels. The multivariate regression models indicated that the colorimetric responses were linearly dependent on protein concentrations with low detection limits of around 1 ppm. Most importantly, the proposed multidimensional colorimetric probe was successfully utilized for protein discrimination in real urine samples. The diverse rainbow responses exhibited by the AuNRs in the proposed probe greatly enhance the accuracy of visual detection, making it a practical tool for straightforward protein monitoring in real samples.

Received 15th November 2024,

Accepted 25th February 2025

DOI: 10.1039/d4nr04797d

rsc.li/nanoscale

^aDepartment of Chemistry, Sharif University of Technology, Tehran, 111559516, Iran. E-mail: hormozi@sharif.edu

^bDepartment of Analytical Chemistry, Faculty of Chemistry, Kharazmi University, Tehran, 15719-14911, Iran

^cUNESCO-UNISA-ITALBS Africa Chair in Nanoscience & Nanotechnology (U2ACN2), College of Graduate Studies, University of South Africa (UNISA), Pretoria, South Africa

†Electronic supplementary information (ESI) available: Chemicals and materials; instrumentation and characterization; synthesis of AuNRs; preparation of the Britton–Robinson buffer; statistical analysis; absorbance spectra and TEM images of the synthesized AuNRs and the AuNRs after etching with NBS; UV-Vis absorption spectral variations of different-sized AuNRs; effect of NBS concentration on the AuNRs; effect of incubation time on the probe; variation in the absorption spectra of the probe at different protein concentration levels; 2D LDA score plots, HCA dendrograms, radar plots, and the jackknifed classification matrix; multivariate calibration using PLS regression. See DOI: <https://doi.org/10.1039/d4nr04797d>

‡These authors contributed equally.

Introduction

In the pathogenesis of diseases, alterations in disease-associated genetic information often lead to alterations in the types and quantities of proteins. These protein alterations within biological systems can enhance our understanding of disease biology.^{1,2} Multiplex detection of proteins simultaneously with high accuracy and sensitivity in complicated sample matrices is essential for allergy testing³ (a method to determine what substances a person is allergic to), disease diagnosis,⁴ clinical treatment,⁵ and proteomics.⁶ For diagnosing urinary system diseases, human urine is a valuable specimen due to its significant biological markers and its noninvasive collection in large quantities.^{7,8} Proteins in healthy human urine can vary from a few milligrams per 24 hours to a maximum of 150 mg

per 24 hours.^{9,10} Both the ultrafiltration mechanism of the kidneys and the natural elimination of the urogenital system influence the production of these proteins. Proteinuria is the medical name for excessive protein loss in urine caused by diseases affecting the kidneys and urogenital tract; it is characterized by protein levels exceeding 150 mg per 24 hours or 100 mg L⁻¹.¹¹ Different urinary system diseases are assessed using a variety of proteins, such as transferrin (TRF), human serum albumin (HSA), lysozyme (Lys), and immunoglobulin G (IgG), with microalbuminuria notably serving as a notable marker of nephropathy in diabetic patients.^{12–17} Therefore, developing robust analytical methods for detecting and differentiating a range of proteins is crucial for early clinical detection, monitoring disease progression, and predicting the development of various diseases.

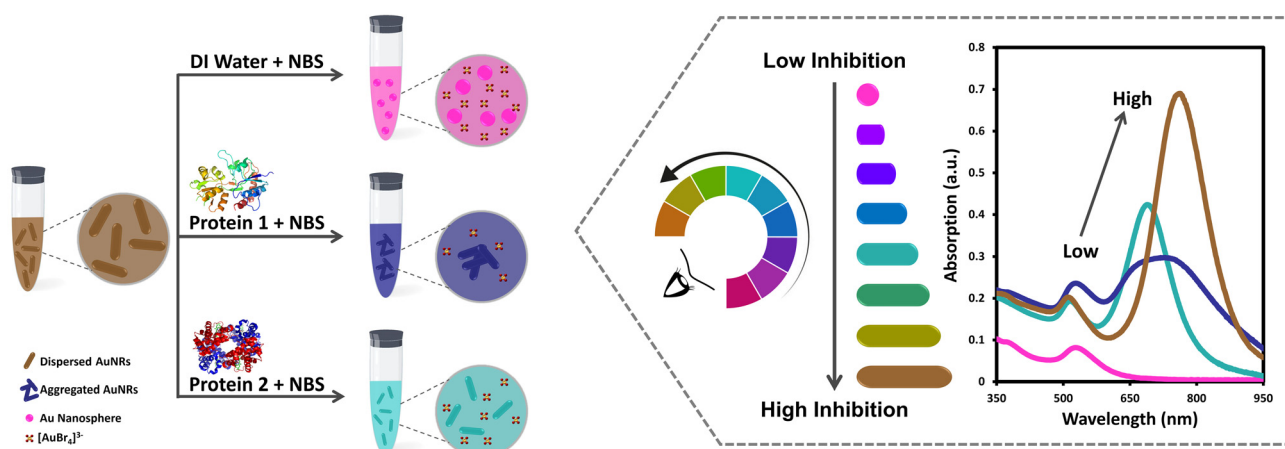
In recent decades, several analytical techniques have been developed for the detection of proteins, notably mass spectrometry (MS) with electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) and enzyme-linked immunosorbent assay (ELISA).^{18–20} The exceptional precision of MS is generally acknowledged in analyzing and characterizing macro-biomolecules, namely proteins.^{21,22} Nevertheless, the capacity of the method to identify a diverse array of chemicals is limited by the tedious process, complex technology, and costly equipment. On the other hand, the great specificity and sensitivity of ELISA—owing to the exact binding between antibodies and their target proteins—make it the most used approach for clinical protein quantification. Despite these advantages, ELISA is constrained by the high cost, unstable antibodies, lengthy procedures, large sample volume needs, and the inability to detect multiple analytes simultaneously. In addition to MS and ELISA, high-performance liquid chromatography (HPLC) and other chromatographic techniques are widely employed in clinical settings for protein analysis due to their high resolution, reproducibility, and ability to separate complex mixtures.^{23,24} However, these methods often require extensive sample preparation, expensive instrumentation, and lengthy analysis times, which limit their practicality for rapid diagnostics. Similarly, spectrofluorimetric techniques provide high sensitivity for protein detection, particularly in monitoring specific biomarkers through fluorescence labeling.²⁵ Despite their accuracy, these methods are constrained by the necessity of fluorescent tags, potential interference from biological matrices, and the requirement for sophisticated instrumentation.

Therefore, it is evident that developing portable, fast, user-friendly, and affordable sensing platforms, rather than a single-sensor-per-analyte approach, is required for detecting and differentiating various proteins and their complex mixtures. The prospective applications of probes in minimizing size while maximizing sensing performance have garnered significant attention in recent years.^{26–28} The integration of a multidimensional sensing probe with machine learning algorithms offers significant benefits by extracting multiple signals from a single sensing element, effectively creating a virtual sensor array capable of multiplex detection and discrimination

of various analytes. In the field of sensing, machine learning provides numerous advantages. It simplifies the processing of complex responses from multidimensional sources and can handle noisy, lower-resolution, or conflicting data, making it a particularly valuable analytical tool. Moreover, advanced multivariate analysis enables machine learning-powered sensors to precisely predict the concentration or identify a target by uncovering correlations between signals and variables.^{29–31} Therefore, combining machine learning techniques with the exceptional advantages of colorimetric sensors—such as portability and visual detection—presents a promising approach for developing an advanced platform for protein analysis.

Recently, plasmonic nanoparticles (NPs) have attracted considerable interest in the design of diverse colorimetric probes for accurate sensing of a series of analytes due to their visual signals, swiftness, low cost, and simplicity.^{29,32–35} Over the past decade, despite the existence of a variety of plasmonic NPs with different shapes and compositions, gold nanorods (AuNRs) have emerged as outstanding signal generators, and color labels in multicolorimetric probes have inspired researchers to utilize them in a wide range of applications.^{36,37} AuNRs have a rod-like, almost one-dimensional structure that creates two distinct localized surface plasmon resonance (LSPR) bands associated with longitudinal and transversal surface plasmon oscillations.^{36,38} The primary reason that AuNRs have such vivid and highly contrasted rainbow color variations is that the band locations depend highly on the aspect ratio of the nanorods; with a minor increase in the aspect ratio, the transversal band blue-shifts slightly and the longitudinal peak red-shifts significantly.^{39–41} Correlating this color variation with the identity and quantity of the analyte in the design of visual sensors is quite challenging. In this regard, etching of AuNRs with a suitable oxidizing agent that reacts with different targets can be a potential solution.^{41,42}

Most recently, the impressive capabilities of two cost-effective mild oxidizing agents, *N*-bromosuccinimide (NBS) and *N*-chlorosuccinimide (NCS), have been highlighted for their ability to rapidly control the etching of gold nanorods (AuNRs) at ambient temperature.⁴³ Hence, we aim to explore another potentiality of this strategy by developing a robust multicolorimetric probe that inhibits the etching of AuNRs for high-performance visual and spectral discrimination and quantification of proteins. In this strategy, NBS undergoes hydrolysis to form HBrO, which can quickly and softly oxidize AuNRs, resulting in distinct color changes that facilitate qualitative and semi-quantitative detection visible to the naked eye. The presence of various proteins, such as ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA, alters the effective concentration of NBS, thus preventing the etching of AuNRs to varying extents and generating unique colorimetric responses for each protein (Scheme 1). The colorimetric responses of the probe were investigated using linear discriminant analysis (LDA) and partial least-squares regression (PLSR) to discriminate proteins and establish the correlation between the concentration matrix and the independent variable matrix. Furthermore, hierarchical cluster analysis (HCA) was employed to accurately cluster



Scheme 1 Schematic illustration representing the principle of the proposed multicolorimetric probe for protein discrimination.

individual proteins and their combinations at various levels. Obviously, the multicolorimetric probe functions as a self-contained sensing unit that requires no internal or external adjustments, making it a practical platform for multiplex protein detection. The etching-based colorimetric probe exhibits vibrant rainbow color patterns and interesting spectral variations, enabling effective detection of proteins at low concentrations while maintaining accuracy even in complicated environments.

Experimental

Fabrication of the multidimensional colorimetric probe

To fabricate the proposed multidimensional colorimetric probe, etching of AuNRs with NBS at a pH of 7 was leveraged as the sensing element. To achieve this purpose, 50 μL of the protein sample were incubated with 50 μL of NBS solution in acetonitrile (5.0 mmol L^{-1}) for 10 minutes to allow the oxidation reaction to reach equilibrium. Following this incubation, the mixtures were added to a solution containing 100 μL of AuNRs and 100 μL of Britton–Robinson buffer (pH 7, 0.02 mol L^{-1}) and diluted to a final volume of 1.0 mL with Milli-Q water. After leaving the final mixture to remain undisturbed for 10 minutes at ambient temperature to ensure that equilibrium was achieved, absorption spectra (350–900 nm) and images of multicolor variations were obtained. This procedure was utilized to evaluate urinary protein samples at various concentrations ranging from 1.0 to 75.0 ppm.

Real sample analysis

To evaluate the applicability of the proposed multidimensional colorimetric probe for qualitative and quantitative detection of proteins in a real matrix, human urine was selected as the sample medium. During the experiment, a urine sample was obtained and centrifuged at 8000 rpm for 10 minutes to eliminate any possible particles from the matrix. The resulting supernatant was diluted 50-fold with Milli-Q water and used

for the experiment instead of deionized water. Four distinct protein samples—ACP, Pep, TRF, and Lys—were examined with three replicates conducted for each protein. Subsequently, to evaluate the practicality of the probe, the colorimetric responses from the urine protein analysis were input into the previously trained partial least squares regression (PLSR) models to predict the concentrations.

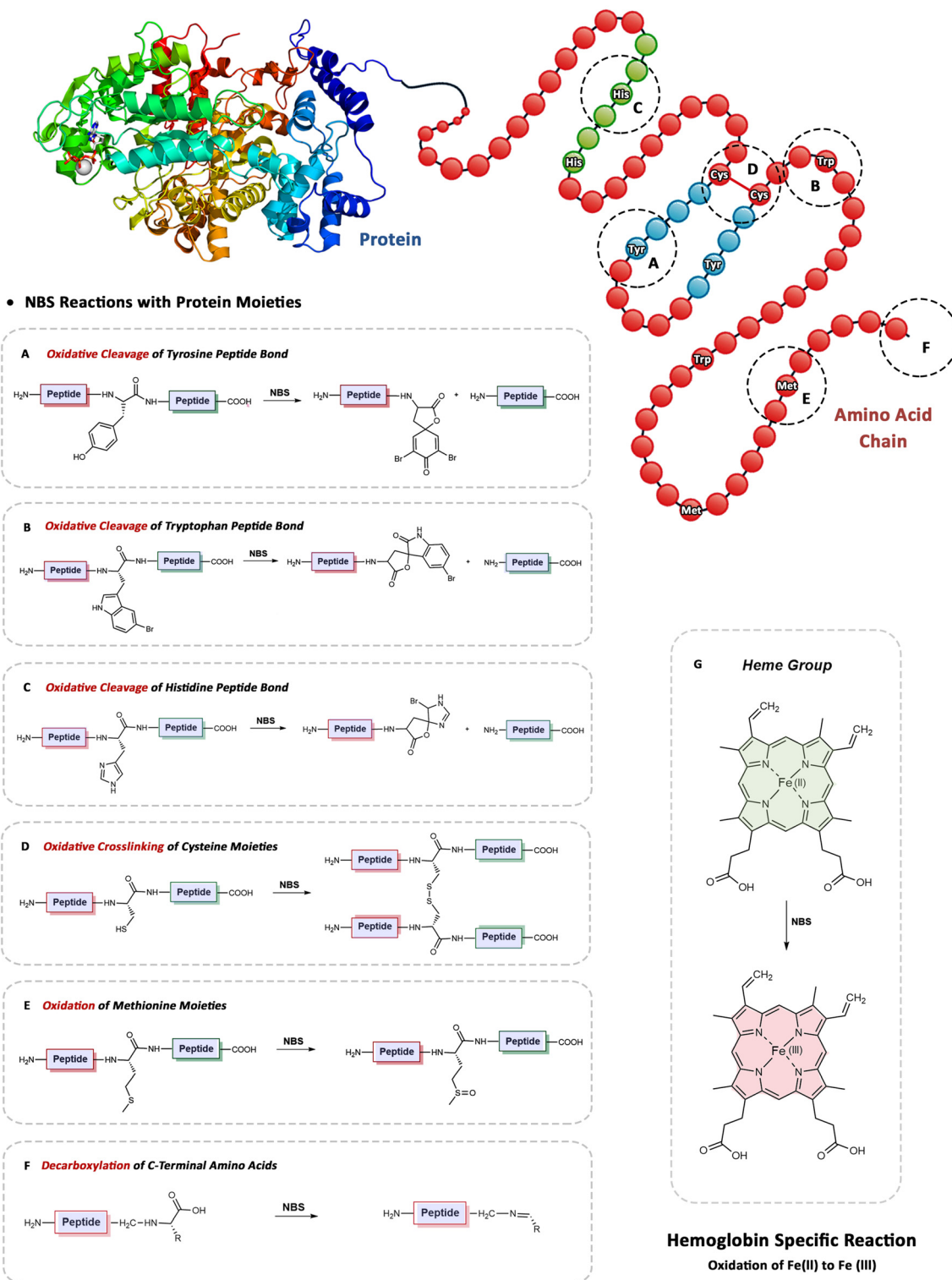
Results and discussion

Sensing strategy of the multidimensional colorimetric probe for protein detection

As previously stated, the quantification and discrimination of proteins were achieved through the inhibition of AuNR etching using NBS. Indeed, proteins and peptides can interact with NBS in various ways depending on their unique structures, including the amino acid content and sequence.^{43–55} These interactions encompass, but are not limited to, oxidation and cleavage of peptide bonds—particularly tryptophan, histidine, and tyrosine residues—oxidation of cysteine thiol groups to disulfides, and oxidation of methionine sulfides to sulfoxides. Additionally, certain proteins possess structural features that facilitate their distinct identification in this mechanism. For example, hemoglobin contains four Fe^{2+} ions, which can be readily oxidized using NBS, distinguishing it from other proteins. Furthermore, as previously reported, amino acids can undergo NBS-induced decarboxylation reactions. In the proteins we investigated, the C-terminal amino acids are most likely involved in such decarboxylation processes. In Table S1,[†] we present the structural characteristics of the proteins studied, including the type of C-terminal amino acid, the content of oxidizable and cleavable amino acids (such as tryptophan, tyrosine, histidine, cysteine, and methionine), the total number of amino acids in each protein, and the most abundant amino acid. These features collectively explain the varying capacities of each protein to consume NBS to different extents, thereby accounting for the observed high discrimi-

nation power of the proposed sensor. Scheme 2 illustrates key reactions between NBS and proteins that are expected to play a role in the proposed sensing strategy. So, according to the variety of amino acid residues in target proteins, the degree of

oxidation of proteins using NBS varies, which causes different etching patterns in the AuNRs. As a result, the incubation of NBS with proteins can be utilized as an effective method to determine the correlation between the identity and concen-



Scheme 2 Key reactions between NBS and proteins that contribute to the proposed sensing strategy.

trations of these proteins and the degree of AuNR etching. Thus, distinct colorimetric patterns were generated by varying protein concentrations, facilitating their quantification and discrimination using both visual observation and machine learning techniques.

To achieve this, UV-Vis spectroscopy and TEM were used to characterize the synthesized AuNRs, which had an average aspect ratio of 3.7, demonstrated excellent monodispersity, and exhibited two characteristic LSPR peaks at 514 nm (transversal) and 760 nm (longitudinal) (Fig. S2a and S2b†). Upon etching the AuNRs, a shift towards shorter wavelengths in the longitudinal LSPR peak was observed, accompanied by a color change from brown to pink (Fig. S2†). This process also resulted in the appearance of an absorption peak at 525 nm, indicating the formation of gold nanospheres.⁴³ Although the study focused on AuNRs with an aspect ratio of 3.7, the methodology can be adapted to AuNRs with different aspect ratios, provided that experimental conditions remain consistent within each set of experiments. To explore this adaptability, the spectral variations of three AuNRs with different longitudinal LSPR wavelengths during the etching process were recorded (Fig. S3†). It was found that larger aspect ratios produced more pronounced plasmonic shifts during etching, which is crucial for enhancing color tonality and extending the detection range. For example, AuNRs with shorter LSPR wavelengths (*e.g.*, 725 nm) exhibited subtler color variations and blue shifts, leading to a narrower response range. In summary, the use of a single large batch of AuNRs with a fixed aspect ratio of 3.7 minimized variability and improved the reliability and robustness of the detection system. While the methodology is adaptable to AuNRs with different aspect ratios, larger AuNRs offer a broader response range and enhanced color tonality.

The TEM images depicted in Fig. S2c† demonstrate the formation of nanospheres during the etching process, resulting in vibrant multicolor variations. When proteins were incubated with NBS, distinct multicolor variations were observed, along with a reduction in the blue shift in the spectrum of the AuNRs (Fig. 1). At a concentration of 10.0 ppm, the degree of protein oxidation using NBS varied based on the specific amino acid residues in the target proteins, leading to different etching patterns in the AuNRs (Fig. 1). Additionally, aggregation of the AuNRs was observed only upon exposure to Pep, as indicated by the red shift and broadening of the AuNRs' longitudinal peak (Fig. 1). As shown in Table S2,† this aggregation likely occurs due to the different isoelectric point pH values of Pep (*i.e.*, 1.0) compared to those of the other proteins, as it carries a negative charge at pH levels above 1.0, leading to significant aggregation of the AuNRs rather than etching due to the positive surface charge of the AuNRs. Consequently, the distinct spectral patterns of Pep facilitated its differentiation from the other proteins, with each protein exhibiting a unique absorption pattern, enabling effective discrimination.

Optimization of the experimental conditions

To enhance the performance and sensitivity of the designed multidimensional colorimetric probe for protein discrimi-

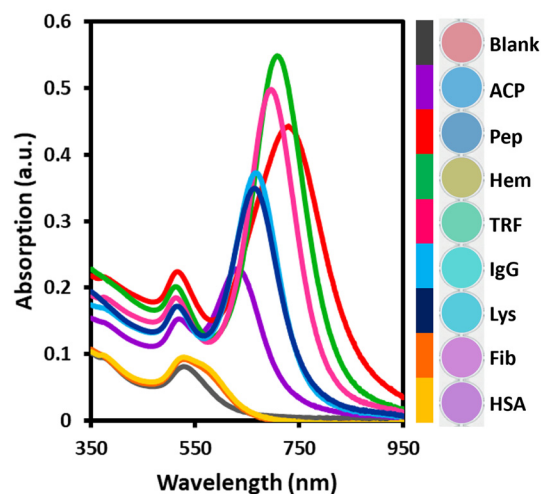


Fig. 1 Absorption spectra and the corresponding images of the etching of the AuNR probe as a blank and in the presence of protein samples (10.0 ppm).

nation, two key parameters controlling the degree of AuNR etching—NBS concentration and the incubation time between NBS and each of the eight proteins—were optimized. To visualize the spectral changes, a bar plot was generated using principal component analysis (PCA) to illustrate variations in the first principal component (PC-1) as a function of the assessed parameter. PCA, a technique employed to reduce the dimensionality of multivariate datasets, was used to transform the number of variables (specifically wavelengths) into a smaller set of principal components (PCs) that effectively represent the entire spectra.

Based on our previous report, the ideal pH for fast and regulated etching is 7, while providing a complete range of colors to be observed in the development of a visual multicolor probe.⁴³ This process is initiated with the etching of brown-colored AuNRs and ends with the formation of red-colored Au nanospheres and colorless Au(I) during the process. Hence, during the optimization process, the pH was adjusted to 7, and all subsequent analyses were performed accordingly.

First of all, Fig. S4† illustrates that increasing the concentration of NBS significantly enhanced the etching of AuNRs, resulting in a greater blue shift of the longitudinal LSPR peak, shifting it from ~ 760 to ~ 525 nm. In fact, $75 \mu\text{mol L}^{-1}$ of NBS completely etched the AuNRs into red-colored Au nanospheres, exhibiting a single LSPR peak at ~ 525 nm. Accordingly, $75.0 \mu\text{mol L}^{-1}$ was determined to be the optimized concentration for NBS.

Incubation time is another important factor that significantly impacts the performance of the proposed probe. Preliminary experiments showed that the etching process is inhibited to varying degrees when NBS is incubated with proteins. Specifically, aqueous solutions of each protein were incubated with NBS for a duration ranging from 0 to 20 minutes, after which the resulting mixture was introduced to AuNRs at pH 7. In the absence of proteins, the etching

process remains unhindered, forming red-colored Au nano-spheres. However, in the presence of proteins, the etching of AuNRs is inhibited to varying degrees, depending on the incubation time, until equilibrium is reached (Fig. S5 and S6†). A fixed time of 10 minutes was selected to ensure the repeatability of the method, as most proteins reach equilibrium within this timeframe, and distinct behaviors between the proteins were observed during this period (Fig. S7†).

Colorimetric responses of the multidimensional probe

The multidimensional colorimetric probe was exposed to eight proteins (*i.e.*, ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA) at various concentrations (ranging from 1.0 to 75.0 ppm) under optimal conditions (Fig. 2) to obtain its spectral responses and

multicolor variations. A series of vivid, high-contrast rainbow colors were generated, providing a distinct pattern for the precise identification and visual analysis of each protein. The variations in color were influenced by both increasing protein concentrations and the identity of proteins, which affected the degree of AuNR etching. By leveraging the human eye's sensitivity to color differences, these rainbow-like multicolor variations were utilized to facilitate the semi-quantitative determination of protein levels without the need for expensive or complex equipment. Additionally, the absorption profiles of the etched AuNRs in the presence of different proteins at various concentrations were monitored and compared (Fig. S8 and S9†). The absorbance spectra and color images captured from the multidimensional colorimetric probe at protein con-

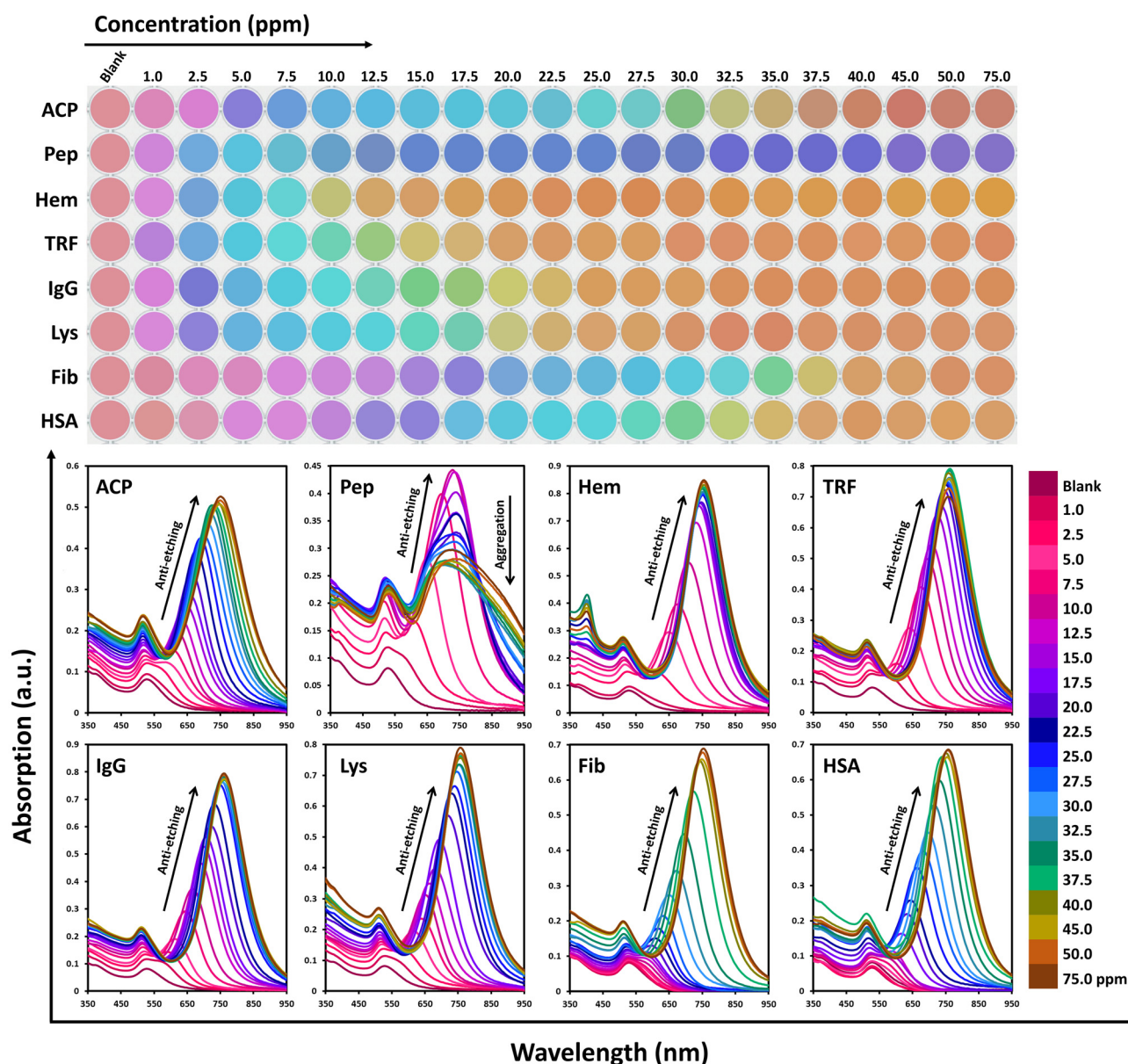


Fig. 2 Color variation patterns and spectral variation responses of the colorimetric multidimensional probe in the presence of eight proteins.

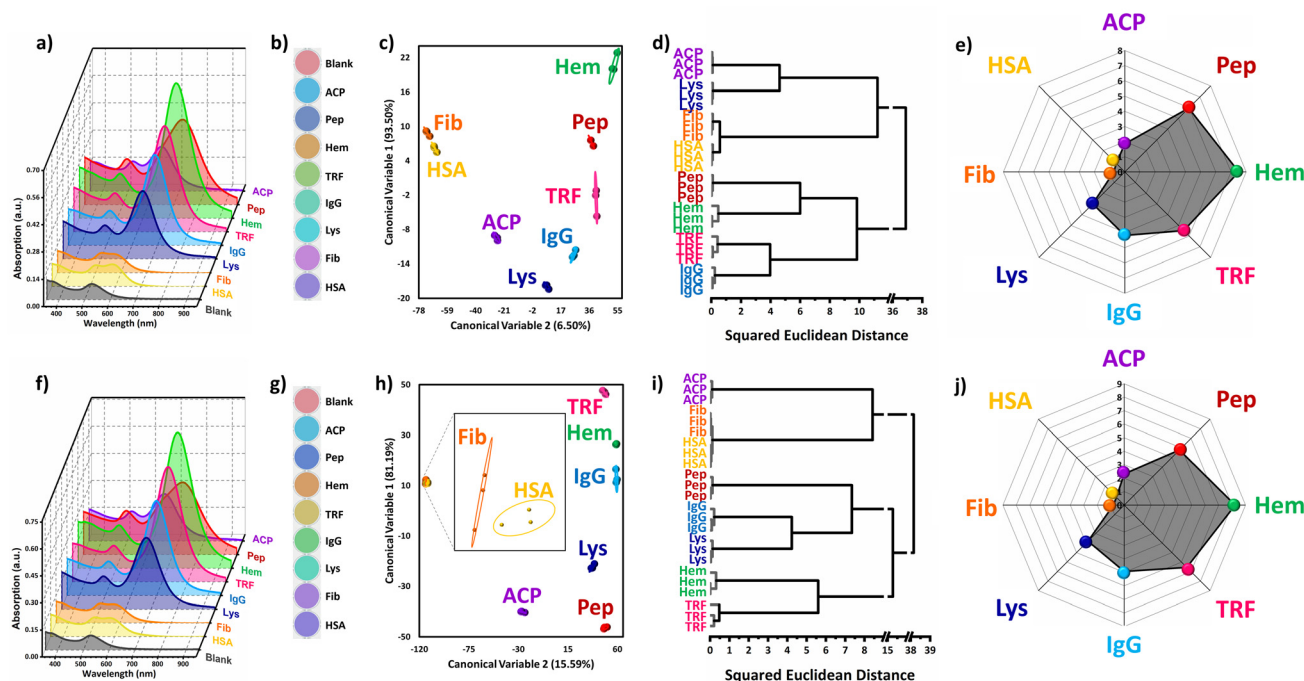


Fig. 3 Unique response patterns of the colorimetric multidimensional probe observed in (a and f) the UV–Vis spectra, (b and g) the images, (c and h) the 2D LDA score plots discriminating different proteins, (d and i) the HCA dendrograms obtained using the Ward method, and (e and j) the Radar plot fingerprints in the presence of 12.5 and 15.0 ppm of proteins, respectively. The upper row displays the data pattern for a concentration of 12.5 ppm, whereas the lower row displays the data pattern for a concentration of 15.0 ppm of eight proteins.

centrations of 12.5 and 15.0 ppm are shown in Fig. 3a, b and f, g. In the presence of proteins, the LSPR bands of AuNRs consistently blue-shifted and decreased in intensity due to the etching of AuNRs and the conversion of Au^0 to Au(I) .⁴³ The results also indicated that the identity and concentration of proteins influenced the magnitude of the spectral shifting. As shown in Fig. 2, increasing Pep concentration not only inhibited etching but also promoted the aggregation of AuNRs. Furthermore, increasing Hem concentration led to the appearance of a peak at around ~ 400 nm, attributed to the inherent absorption of Hem in the solution.

Identification and discrimination of proteins

The combination of a distinctive multidimensional colorimetric probe and a machine learning-based pattern recognition algorithm provides significant advantages in improving the ability of the probe to distinguish proteins. To evaluate the capacity of the probe to discriminate among all eight proteins, the spectral responses acquired for each concentration level were employed to train 20 unique LDA models.

The main challenge in performing LDA is that the number of samples (rows) in the dataset matrix must be equal to or greater than the number of variables (columns).⁵⁶ In this investigation, absorbance values recorded at various wavelengths were used as independent variables, while proteins at different concentrations served as samples, resulting in a smaller number of samples than variables. Therefore, PCA was initially performed to reduce the dimensionality of the train-

ing sets for each case. The LDA models were then trained using the first three PCs, which captured the most significant variance within each dataset. The LDA models showed remarkable discrimination of the eight protein clusters, achieving 100% accuracy across all 20 concentration levels (Fig. S10 and S11†). Furthermore, as shown in Fig. 3c and h, the 2D LDA plots demonstrated the excellent discriminatory power of the multidimensional colorimetric probe for the eight proteins at two concentration levels (12.5 and 15.0 ppm as samples). The exceptional performance of the models was further confirmed by the corresponding jackknife tables (Tables S3 and S4†), which indicated that both sensitivity and selectivity for the eight proteins at these two concentration levels (12.5 and 15.0 ppm as samples) were 100.0%, with no misclassifications. Next, eight distinct two-dimensional (2D) LDAs were conducted to evaluate the sensitivity of the probe to specific proteins at varying concentrations, producing well-clustered 2D score plots for each protein with no misclassification (Fig. 4).

Additionally, CIELAB has recently demonstrated promising potential in representing the features in the spectra of AuNRs for both qualitative and quantitative modeling.³⁹ To assess the applicability of this approach in our proposed sensor, CIELAB parameters were extracted from the color images presented in Fig. 2. The resulting dataset was then used to train an LDA model for classification. As shown in Fig. S12 and S13,† the LDA models based on CIELAB parameters achieved 100% accuracy in discriminating proteins only at specific concentrations: 7.5, 10.0, 12.5, 15.0, 17.5, 20.0, 25.0, 32.5, 35.0, and

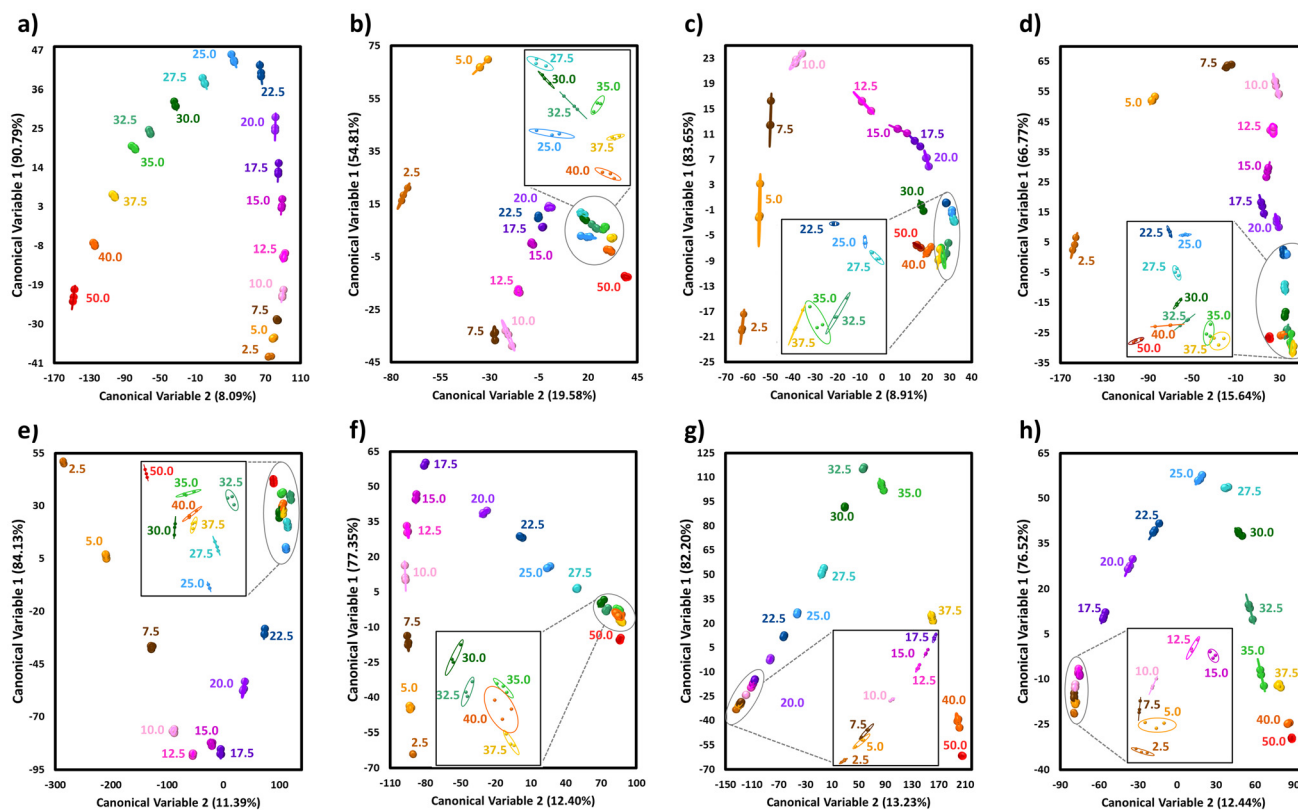


Fig. 4 2D score plots for (a) ACP, (b) Pep, (c) Hem, (d) TRF, (e) IgG, (f) Lys, (g) Fib, and (h) HSA at different concentrations.

45.0 ppm. At other concentrations, the accuracy dropped significantly. In contrast, as shown in Fig. S10 and S11,[†] LDA score plots based on the absorption spectra accurately identified proteins at all 20 concentration levels.

To further evaluate the potential of the probe in discriminating the eight proteins, variations in PC-1 were utilized to perform hierarchical cluster analysis (HCA), a commonly used chemometric technique associated with unsupervised learning. HCA is typically conducted on the sample space to identify clusters within the data. As shown in Fig. S14 and S15,[†] the HCA dendrograms for discrimination of all eight proteins at each concentration level clearly demonstrate that the multidimensional colorimetric probe accurately clustered all three replicates of the protein samples without any misclassification. Furthermore, as illustrated in Fig. 3d and i, the diverse protein samples at two concentration levels (12.5 and 15.0 ppm) were successfully clustered by HCA without any misclassification.

Another approach for representing the obtained response profiles is through radar plots, a practical graphical method for displaying multivariate data in 2D, where three or more quantitative variables are plotted on axes originating from the same point. Radar plots can also be used to represent the sample space, highlighting the variable with the maximum variance between samples. The radar plots shown in Fig. S16 and S17[†] were obtained by using the variations in PC-1 for all proteins at different concentration levels. These unique patterns effectively represent complex numerical data, illustrating

the unique behavior of each protein in comparison with others. As demonstrated in Fig. 3e and j, the radar plots revealed distinct patterns for the eight proteins at two concentration levels (as two sample concentrations).

Regression analysis for quantification of proteins

In addition to identification, the quantification of the target analytes is essential. To achieve this, partial least squares (PLS) regression was applied to evaluate the quantitative capabilities of the multidimensional probe across all eight proteins by analyzing their concentration-dependent spectral responses.

The potential of the probe for quantitative analysis was confirmed by the high correlation observed between the predicted concentrations and the actual measurements (Fig. 5). The multivariate regression models indicated that the colorimetric spectral responses from the probe were linearly dependent on protein concentrations within the ranges of 0.8–40.0, 1.4–32.5, 1.2–22.5, 1.5–22.5, 1.0–27.5, 0.8–30.0, 1.4–45.0, and 1.0–40.0 ppm, with detection limits of 0.3, 0.5, 0.4, 0.5, 0.3, 0.3, 0.5, and 0.3 ppm for ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA, respectively (Table 1). A series of analytical figures of merit, including accuracy, precision, sensitivity, and response range, were calculated and are presented in Table 1. The robustness of the regression models was further demonstrated by the high R^2 values ($R^2 > 0.99$) and the low root-mean-square error (RMSE) values.

Additionally, the CIELAB dataset obtained in the previous section was used to train PLSR models for each protein. A com-

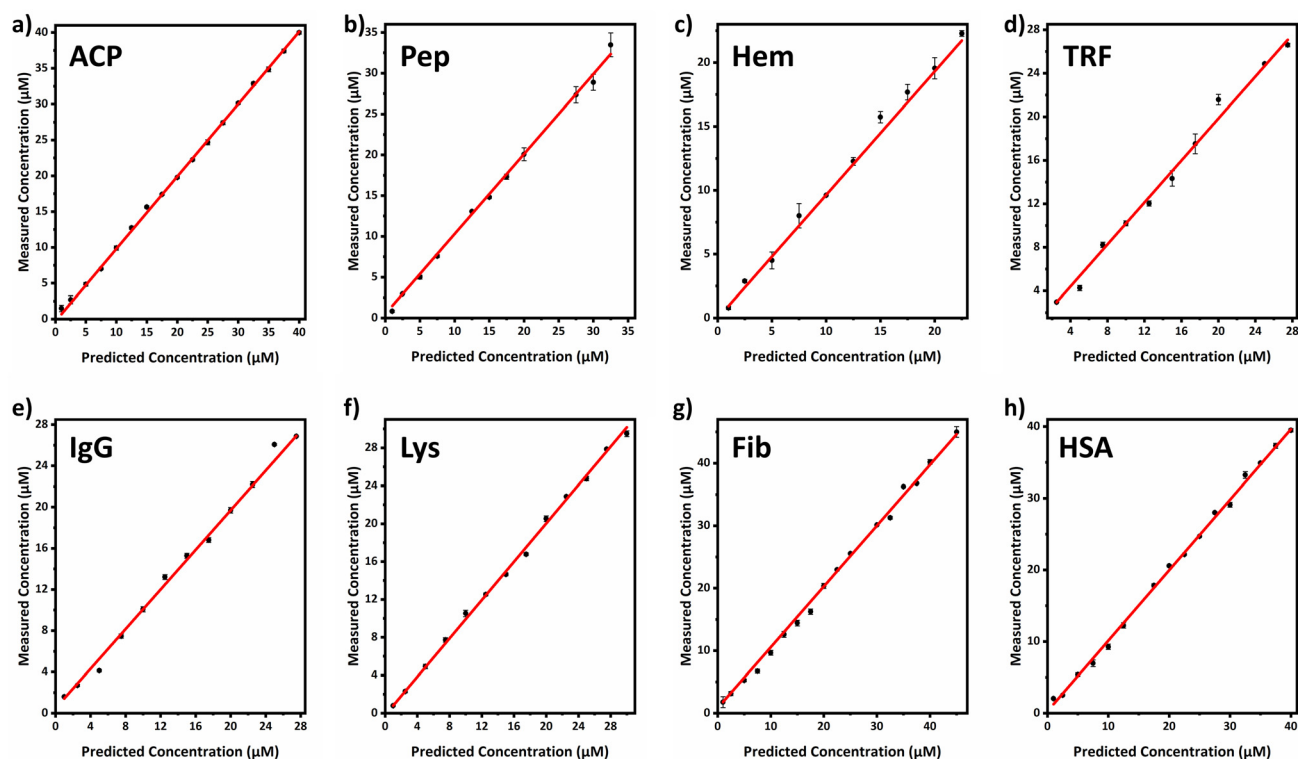


Fig. 5 Predicted vs. measured concentration plots as multivariate calibration using PLS regression for (a) ACP, (b) Pep, (c) Hem, (d) TRF, (e) IgG, (f) Lys, (g) Fib, and (h) HSA in their entire concentration ratio range.

Table 1 Analytical figures of merit of PLS regression on individual proteins (*i.e.*, including ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA)

| Sample | Opt. LVs | RMSEP | REP% | R^2 | SEN | Anal. SEN | LOD (ppm) | LOQ (ppm) | Linear range (ppm) |
|--------|----------|-------|------|--------|-------|-----------|-----------|-----------|--------------------|
| ACP | 4 | 0.4 | 1.8 | 0.9995 | 0.108 | 14.21 | 0.3 | 0.8 | 0.8–40.0 |
| Pep | 5 | 0.8 | 4.1 | 0.9981 | 0.044 | 16.21 | 0.5 | 1.4 | 1.4–32.5 |
| Hem | 6 | 0.6 | 5.2 | 0.9964 | 0.054 | 21.45 | 0.4 | 1.2 | 1.2–22.5 |
| TRF | 5 | 0.8 | 5.7 | 0.9948 | 0.123 | 60.23 | 0.5 | 1.5 | 1.5–27.5 |
| IgG | 4 | 0.6 | 4.3 | 0.9975 | 0.268 | 42.28 | 0.3 | 1.0 | 1.0–27.5 |
| Lys | 4 | 0.4 | 2.8 | 0.9990 | 0.181 | 30.42 | 0.3 | 0.8 | 0.8–30.0 |
| Fib | 6 | 1.0 | 4.7 | 0.9971 | 0.027 | 21.06 | 0.5 | 1.4 | 1.4–45.0 |
| HSA | 6 | 0.6 | 3.1 | 0.9988 | 0.058 | 22.06 | 0.3 | 1.0 | 1.0–40.0 |

parison between the PLSR models derived from the absorption spectra and those based on the CIELAB dataset revealed that the latter exhibited significantly narrower linear ranges. Specifically, the linear ranges for ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA were limited to 10.0–25.0, 2.5–25.0, 1.0–10.0, 2.5–15.0, 5.0–20.0, 5.0–20.0, 5.0–20.0, and 7.5–20.0 ppm, respectively (Fig. S18†). These findings confirm that absorption signals provide a more reliable and robust approach for quantitative analysis, enabling a broader and more comprehensive evaluation of protein interactions across a wider dynamic range.

Mixture analysis

One of the remarkable advantages of the proposed multidimensional colorimetric probe is its capability for visual and spectral discrimination between different ratios of protein mix-

tures. This task is considerably more challenging than discrimination based on pure proteins and holds significant promise for medical diagnostics. To demonstrate the effectiveness of this multidimensional sensing strategy, the spectral responses of binary protein mixtures, such as HSA/Lys and HSA/TRF, as well as ternary protein mixtures (HSA/Lys/TRF), at varying component concentration ratios (with a total protein concentration of 15.0 ppm) were achieved (Fig. S19†). The spectral response matrix was analyzed using PCA and the first 20 PCs were utilized as variables to train three distinct LDA models. This approach confirmed the discrimination power of the sensor with 100% cross-validation accuracy achieved for all models. The 2D score plots shown in Fig. 6a–c illustrate the precise classification of all mixture samples. The corresponding jackknife tables further validate the exceptional performance of the model without any misclassifications and

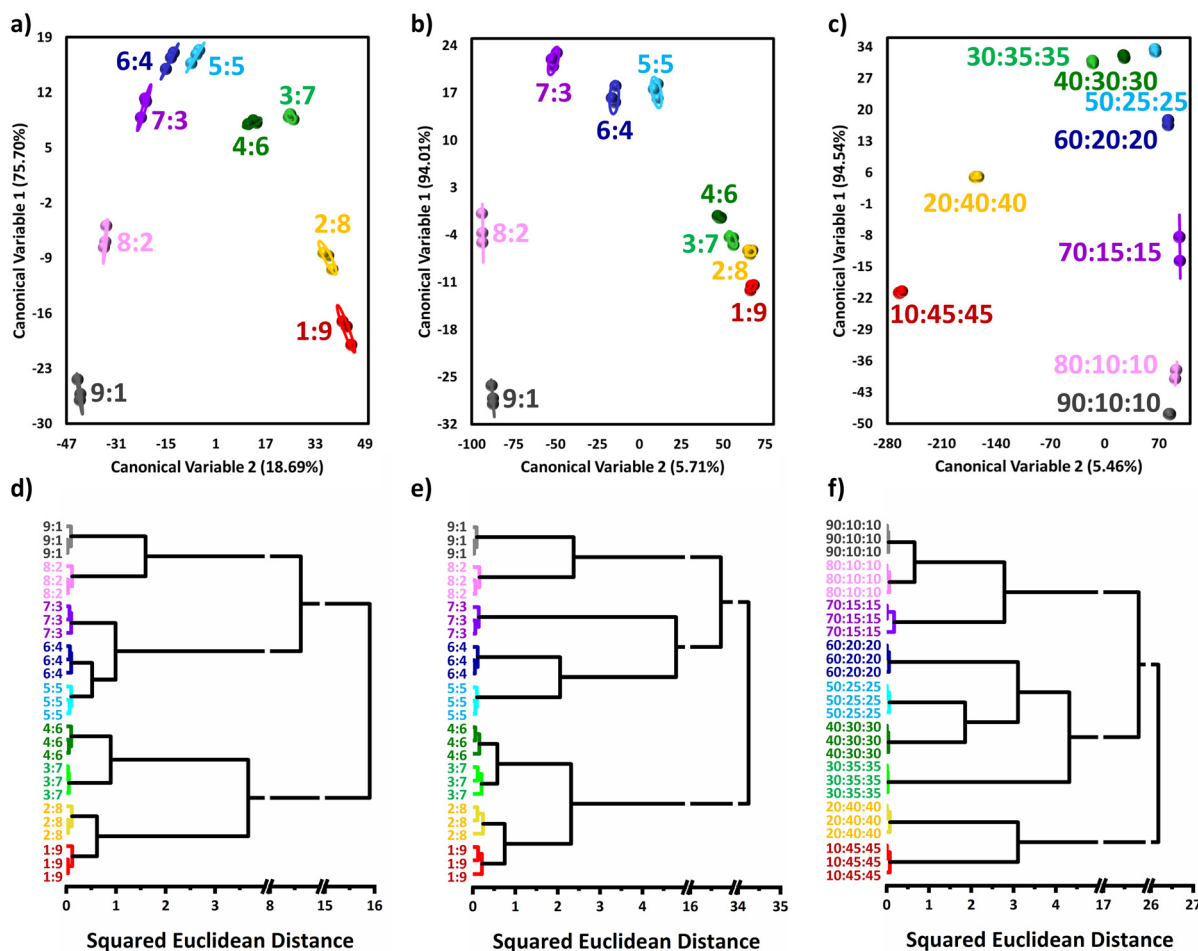


Fig. 6 2D LDA score plot and HCA dendrograms obtained using the Ward method for discriminating the different concentration ratios of binary protein mixtures: (a and d) HSA/Lys and (b and e) HSA/TRF, and ternary protein mixtures: (c and f) HSA/Lys/TRF. The total concentration for mixtures is 15.0 ppm.

both sensitivity and selectivity reach 100.0% for the binary protein mixtures of HSA/Lys and HSA/TRF and the ternary mixtures of HSA/Lys/TRF (Tables S5–S7†). Additionally, the HCA dendrograms for the binary and ternary protein mixtures presented in Fig. 5d–f clearly demonstrate that the probe can accurately cluster different protein mixture ratios without any misclassification. Moreover, the radar plots for the binary and ternary mixtures reveal unique fingerprint patterns, allowing for visual discrimination without the need for statistical analysis (Fig. S20†).

Most importantly, the proposed strategy is particularly effective for determining the HSA percentage in mixture samples containing HSA/Lys and HSA/TRF binary mixtures, as well as the HSA/Lys/TRF ternary mixtures with varying ratios. Changing the ratio in mixture samples from 9:1 to 1:9 for binary mixtures and from 90:5:5 to 10:45:45 for ternary mixtures results in a strong correlation with the inhibition of AuNR etching (Fig. S21†). Consequently, the dataset matrix comprising binary and ternary protein mixtures was used to train three separate PLS-1 models based on the HSA concen-

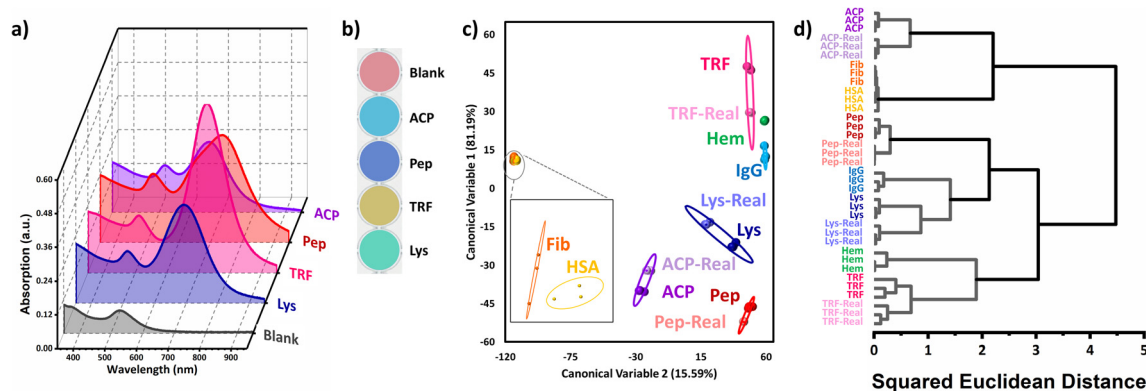
tration percentage. Three robust regression models were developed and are documented in Table 2, demonstrating exceptional performance in terms of sensitivity, accuracy, precision, and response range. The models exhibited high accuracy, as indicated by large R^2 ($R^2 > 0.99$) and low root-mean square error (RMSE) values. Ultimately, the results clearly indicate that the proposed multidimensional colorimetric probe is highly suitable for accurately analyzing proteins in a quantitative manner.

Performance of the multidimensional colorimetric probe in real samples

The proposed multidimensional colorimetric probe was ultimately utilized for the identification and quantification of proteins in real urine samples. For this purpose, urine samples were collected and spiked with four distinct proteins—ACP, Pep, TRF, and Lys—each at a concentration of 15.0 ppm. The response profiles of the proteins within the urine matrix illustrated in Fig. 7a and b were integrated into the overall spectral matrix and processed using PCA to reduce dimensionality. The

Table 2 Analytical figures of merit of PLS regression on binary protein mixtures (HSA/Lys and HSA/Lys) and ternary protein mixtures (HSA/Lys/TRF) based on the HSA percentage

| Sample | Opt. LVs | RMSEP | REP% | R^2 | SEN | Anal. SEN | LOD (%) | LOQ (%) | Linear range (%) |
|-------------|----------|-------|------|--------|-------|-----------|---------|---------|------------------|
| HSA/Lys | 4 | 2.3 | 4.6 | 0.9959 | 0.008 | 9.90 | 1.5 | 4.6 | 4.6–90 |
| HSA/TRF | 6 | 2.5 | 4.6 | 0.9940 | 0.004 | 7.18 | 1.8 | 5.4 | 5.4–90 |
| HSA/Lys/TRF | 6 | 1.2 | 2.4 | 0.9989 | 0.005 | 7.87 | 0.9 | 2.7 | 2.7–90 |

**Fig. 7** (a) Spectral variation, (b) corresponding image responses, (c) 2D LDA score plot and (d) HCA dendrograms obtained using the Ward method of the colorimetric multidimensional probe in the presence of proteins in real urine samples.

first 20 PCs were subsequently used as test variables in a pre-trained LDA model, as shown in Fig. 3h. Table S7† presents the successful classification of unknown protein samples into their respective groups based on posterior probabilities, which indicates the likelihood of each protein—ACP, Pep, TRF, and Lys—belonging to a specific class. The assignment of each protein to its target class was determined by calculating the minimum distance between the sample and the centroid of the corresponding class (Fig. 7c). As summarized in Table S8,† all unknown protein samples were accurately classified and assigned to their appropriate groups, demonstrating the efficacy of the probe in distinguishing proteins in biological samples. Furthermore, as depicted in Fig. 7d, variations in PC-1 of the unknown protein samples were introduced into the pre-trained HCA model (Fig. 3i). The resulting dendrogram clearly indicated that the probe was able to accurately categorize ACP, Pep, TRF, and Lys, along with their replicates, into their respective target groups without misclassification. Following the successful discrimination of proteins in urine samples, the concentrations of the unknown protein samples were determined by inputting the absorption responses into a pre-trained PLSR model without dimensionality reduction. The multivariate calibration MVC-1 toolbox in MATLAB was used to incorporate both the training and test sets, automatically training the model to predict the protein concentrations in urine samples. The recovery and relative standard deviation (RSD) values reported in Table 3 validate the reliability of the probe for quantifying individual proteins in human urine. Collectively, the classification and regression results confirm the robustness of the proposed colorimetric probe for both the

Table 3 Results of determining unknown proteins in human urine samples

| Sample | Spiked (ppm) | Found (ppm) | Recovery (%) | RSD ($n = 3$, %) |
|----------|--------------|-------------|--------------|--------------------|
| ACP-real | 15.0 | 16.2 | 108.2 | 1.1 |
| Pep-real | 15.0 | 15.5 | 103.6 | 0.1 |
| TRF-real | 15.0 | 16.1 | 107.3 | 2.8 |
| Lys-real | 15.0 | 15.1 | 100.8 | 1.3 |

quantification and identification of protein samples in human urine, underscoring its potential as a practical tool for protein monitoring in biological samples.

Conclusion

In summary, a machine learning-assisted multidimensional colorimetric probe was successfully developed to distinguish proteins, drawing inspiration from various color and spectral changes of AuNRs resulting from etching with NBS. The etching process of AuNRs was variably inhibited when incubated with different protein samples (*i.e.*, ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA), leading to distinct and precise color and spectral responses for each protein, which resulted in high-resolution colorimetric data. This innovative approach showed considerable promise in accurately identifying and quantifying individual proteins, as well as binary protein mixtures (HSA/Lys and HSA/TRF) and ternary protein mixtures (HSA/Lys/TRF) through spectral analysis. The obtained

response profiles were further analyzed using LDA models to precisely determine protein identities. Additionally, the PLSR algorithm proved to be highly effective for quantifying individual protein amounts and determining the percentage of HSA in the mixtures. The outstanding figures of merit achieved by these models clearly demonstrated the exceptional sensitivity, accuracy, and precision of the proposed strategy. Furthermore, the multidimensional colorimetric probe was successfully employed to differentiate and quantify proteins in real human urine samples, showing no significant interference from other substances present in the samples. With its simplicity, low cost, and user-friendliness, this method is positioned as a portable platform for on-site monitoring of proteins in real samples.

Data availability

The corresponding author can provide the datasets upon request.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

The authors express their sincere gratitude to the Sharif University of Technology and the South African National Research Foundation (NRF) for their support of this work. Financial support from the Iran National Science Foundation (INSF) under Grant No. 99015661 is also gratefully acknowledged.

References

- 1 Y. Jiang, M. Shi, Y. Liu, S. Wan, C. Cui, L. Zhang and W. Tan, *Angew. Chem., Int. Ed.*, 2017, **56**, 11916–11920.
- 2 X. Rong, L. Xiang, Y. Li, H. Yang, W. Chen, L. Li, D. Liang and X. Zhou, *Front. Aging Neurosci.*, 2020, **12**, 248.
- 3 M. Milana, E. D. van Asselt and I. H. van der Fels-Klerx, *Compr. Rev. Food Sci. Food Saf.*, 2025, **24**, e70123.
- 4 H. Di, Z. Mi, Y. Sun, X. Liu, X. Liu, A. Li, Y. Jiang, H. Gao, P. Rong and D. Liu, *Theranostics*, 2020, **10**, 9303.
- 5 H. Li and A. J. Steckl, *Anal. Chem.*, 2018, **91**, 352–371.
- 6 J. Xue, L. Yang, Y. Jia, H. Wang, N. Zhang, X. Ren, H. Ma, Q. Wei and H. Ju, *ACS Sens.*, 2019, **4**, 2825–2831.
- 7 R. Pieper, C. L. Gatlin, A. M. McGrath, A. J. Makusky, M. Mondal, M. Seonarain, E. Field, C. R. Schatz, M. A. Estock and N. Ahmed, *Proteomics*, 2004, **4**, 1159–1174.
- 8 J. Adachi, C. Kumar, Y. Zhang, J. V. Olsen and M. Mann, *Genome Biol.*, 2006, **7**, 1–16.
- 9 J. M. González-Buitrago, L. Ferreira and I. Lorenzo, *Clin. Chim. Acta*, 2007, **375**, 49–56.
- 10 C. S. Spahr, M. T. Davis, M. D. McGinley, J. H. Robinson, E. J. Bures, J. Beierle, J. Mort, P. L. Courchesne, K. Chen and R. C. Wahl, *Proteomics*, 2001, **1**, 93–107.
- 11 E. Franklin, *J. Clin. Invest.*, 1959, **38**, 2159–2167.
- 12 C. Cheung, C. Cockram, V. Yeung and R. Swaminathan, *Clin. Chem.*, 1989, **35**, 1672–1674.
- 13 G. D'amico and C. Bazzi, *Kidney Int.*, 2003, **63**, 809–825.
- 14 W. Guder and W. Hofmann, *Clin. Biochem.*, 1993, **26**, 277–282.
- 15 J. Harrison, R. Parker and K. De Silva, *J. Clin. Pathol.*, 1973, **26**, 278–284.
- 16 P. A. Peterson and I. Berggård, *Eur. J. Clin. Invest.*, 1971, **1**, 255–264.
- 17 G. Viberti, R. Jarrett and H. Keen, *Lancet*, 1982, **320**, 611.
- 18 A. Dobo and I. A. Kaltashov, *Anal. Chem.*, 2001, **73**, 4763–4773.
- 19 R. Lemaire, M. Wisztorski, A. Desmons, J. Tabet, R. Day, M. Salzert and I. Fournier, *Anal. Chem.*, 2006, **78**, 7145–7153.
- 20 C. K. Dixit, S. K. Vashist, F. T. O'Neill, B. O'Reilly, B. D. MacCraith and R. O'Kennedy, *Anal. Chem.*, 2010, **82**, 7049–7052.
- 21 A. Ambrosi, F. Airo and A. Merkoçi, *Anal. Chem.*, 2010, **82**, 1151–1156.
- 22 C.-P. Jia, X.-Q. Zhong, B. Hua, M.-Y. Liu, F.-X. Jing, X.-H. Lou, S.-H. Yao, J.-Q. Xiang, Q.-H. Jin and J.-L. Zhao, *Biosens. Bioelectron.*, 2009, **24**, 2836–2841.
- 23 S. Aitekenov, A. Gaipov and R. Bukasov, *Talanta*, 2021, **223**, 121718.
- 24 G. Seipke, H. Müllner and U. Grau, *Angew. Chem., Int. Ed. Engl.*, 1986, **25**, 535–552.
- 25 S. Saraswat, B. Snyder and D. Isailovic, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2012, **902**, 70–77.
- 26 W. Sun, Y. Lu, J. Mao, N. Chang, J. Yang and Y. Liu, *Anal. Chem.*, 2015, **87**, 3354–3359.
- 27 A. Hierlemann and R. Gutierrez-Osuna, *Chem. Rev.*, 2008, **108**, 563–613.
- 28 C. Jiang, H. Huang, X. Kang, L. Yang, Z. Xi, H. Sun, M. D. Pluth and L. Yi, *Chem. Soc. Rev.*, 2021, **50**, 7436–7495.
- 29 X. Ma, S. He, B. Qiu, F. Luo, L. Guo and Z. Lin, *ACS Sens.*, 2019, **4**, 782–791.
- 30 A. Orouji, F. Ghasemi and M. R. Hormozi-Nezhad, *Anal. Chem.*, 2023, **95**, 10110–10118.
- 31 C. Li, P. Wu and X. Hou, *Nanoscale*, 2016, **8**, 4291–4298.
- 32 J. Sun, Y. Lu, L. He, J. Pang, F. Yang and Y. Liu, *TrAC, Trends Anal. Chem.*, 2020, **122**, 115754.
- 33 J. Mao, Y. Lu, N. Chang, J. Yang, S. Zhang and Y. Liu, *Biosens. Bioelectron.*, 2016, **86**, 56–61.
- 34 J. Mao, Y. Lu, N. Chang, J. Yang, J. Yang, S. Zhang and Y. Liu, *Analyst*, 2016, **141**, 4014–4017.
- 35 H. Liu, L. Ma, S. Xu, W. Hua and J. Ouyang, *J. Mater. Chem. B*, 2014, **2**, 3531–3537.
- 36 H. Chen, L. Shao, Q. Li and J. Wang, *Chem. Soc. Rev.*, 2013, **42**, 2679–2724.

- 37 J. Zheng, X. Cheng, H. Zhang, X. Bai, R. Ai, L. Shao and J. Wang, *Chem. Rev.*, 2021, **121**, 13342–13453.
- 38 K. Park, S. Biswas, S. Kanel, D. Nepal and R. A. Vaia, *J. Phys. Chem. C*, 2014, **118**, 5918–5926.
- 39 Y. Zheng, M. Xiao, S. Jiang, F. Ding and J. Wang, *Nanoscale*, 2013, **5**, 788–795.
- 40 A. Orouji, S. Abbasi-Moayed, F. Ghasemi and M. R. Hormozi-Nezhad, *Sens. Actuators, B*, 2022, **358**, 131479.
- 41 H. Rao, X. Xue, H. Wang and Z. Xue, *J. Mater. Chem. C*, 2019, **7**, 4610–4621.
- 42 K. Kermanshahian, A. Yadegar and H. Ghourchian, *Coord. Chem. Rev.*, 2021, **442**, 213934.
- 43 M. Ghamsari, A. Orouji and M. R. Hormozi-Nezhad, *Anal. Chem.*, 2023, **95**, 15985–15993.
- 44 L. Ramachandran and B. Witkop, *Methods Enzymol.*, 1967, **11**, 283–299.
- 45 M. Kumara, N. L. Gowda, K. Mantelingu and K. K. S. Rangappa, *J. Mol. Catal. A: Chem.*, 2009, **309**, 172–177.
- 46 N. Konigsberg, G. Stevenson and J. M. Luck, *J. Biol. Chem.*, 1960, **235**, 1341–1345.
- 47 G. Gopalakrishnan and J. L. Hogg, *J. Org. Chem.*, 1985, **50**, 1206–1212.
- 48 A. Martínez-Ramírez, J. L. Bella and J. C. Stockert, *Micron*, 2002, **33**, 399–402.
- 49 Y. Shechter, Y. Burstein and A. Patchornik, *Biochemistry*, 1975, **14**, 4497–4503.
- 50 W. E. Savage and A. Fontana, *Methods Enzymol.*, Academic Press, 1977, vol. 47, pp. 453–459.
- 51 T. Spande and B. Witkop, *Methods Enzymol.*, 1967, **11**, 528–532.
- 52 S. P. Manly and K. S. Matthews, *J. Biol. Chem.*, 1997, **254**, 3341–3347.
- 53 B. J. Smith, *Protein Protocols Handbook*, Springer, 1996, pp. 375–380.
- 54 J. Roeser, R. Bischoff, A. P. Bruins and H. P. Permentier, *Anal. Bioanal. Chem.*, 2010, **397**, 3441–3455.
- 55 G. L. Schmir and L. A. Cohen, *J. Am. Chem. Soc.*, 1961, **83**, 723–728.
- 56 M. Ghamsari, N. Fahimi-Kashani and M. R. Hormozi-Nezhad, *ACS Appl. Mater. Interfaces*, 2023, **15**, 26081–26092.