# 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,]*

[a]Department of Chemistry, Sharif University of Technology, Tehran, 111559516, Iran

[b]Department of Analytical Chemistry, Faculty of Chemistry, Kharazmi University, Tehran, 15719-14911, Iran.

[c]UNESCO-UNISA-iTALBS Africa Chair in Nanoscience & Nanotechnology (U2ACN2), College of Graduate Studies, University of South Africa (UNISA), Pretoria, South Africa

† These authors contributed equally.

*Corresponding authors:

- Mohammad Reza Hormozi-Nezhad, email: hormozi@sharif.edu
ORCID iD: 0000-0002-7472-1850

## Table of Contents

## Chemicals and Materials

All of the chemicals were of high purity and used without any additional purification. The chemical hydrogen tetrachloroaurate ($HAuCl_4.3H_2O$) (99.5%), cetyltrimethylammonium bromide (CTAB), sodium borohydride ($NaBH_4$), ascorbic acid (AA), silver nitrate ($AgNO_3$), ortho-phosphoric acid (85%), acetic acid (glacial) (100%), boric acid, sodium borohydride, sodium hydroxide, and N-bromosuccinimide (NBS), were obtained from Sigma-Aldrich. Eight analytically pure proteins, including phosphatase ACP), pepsin (Pep), human hemoglobin (Hem), human transferrin (TRF), human immunoglobulin G (IgG), human lysozyme (Lys), human fibrinogen (Fib), and human serum albumin (HSA), were obtained from Sigma-Aldrich. All experiments were conducted using deionized (DI) water with the resistivity of 18.2 MΩ.

## Instrumentation and Characterization

The visible absorption spectra of the samples were recorded using an Agilent Carry 60 spectrophotometer and 1.0 cm disposable cuvettes. Images of color variations were captured using the manual mode of a Samsung Galaxy A71 smartphone from the top of a 96-well plate placed on an opaque light box with white illumination. A convex surface was created by filling each well to overflow, allowing for the acquisition of a uniformly clear picture of the solution (Fig. S1). Using a Zeiss EM900 (Germany) transmission electron microscope with an acceleration voltage of 200 kV, morphological characteristics of AuNRs were investigated through transmission electron microscopy (TEM).

## Synthesis of AuNRs

The synthesis of gold nanorods (AuNRs) was carried out using the previously described seed-mediated growth strategy [1]. Initially, two separate solutions were prepared, one called the growth solution and the other the seed solution. For the seed solution, 0.125 mL of HAuCl4 (0.01 mol $L^{-1}$) was added to 5.0 mL of CTAB (0.1 mol $L^{-1}$) under vigorous stirring. After the addition of 0.3 mL of a freshly prepared ice-cold NaBH4 solution (0.01 mol $L^{-1}$), a brownish-yellow solution containing gold nanoseeds was formed and stored at room temperature under continuous stirring and utilized within 2-5 hours. To prepare the growth solution, 2.5 mL of $HAuCl_4$ (0.01 mol $L^{-1}$) and 0.3 mL of $AgNO_3$ (0.01 mol $L^{-1}$) were sequentially introduced into 50.0 mL of CTAB (0.1 mol $L^{-1}$) while stirring vigorously. Subsequently, 0.3 mL of ascorbic acid (AA) with a concentration of 0.10 mol $L^{-1}$ was injected, and discoloration of the mixture as a result of the reduction of Au(III) to Au(I) was observed. After gently injecting 0.25 mL of the seed solution into the reaction mixture and halting the stirring after five seconds, the mixture

was left undisturbed at room temperature overnight, during which the solution gradually turned light brown, indicating the growth of Au nanoseeds into AuNRs. A large (~1 L) batch of AuNRs with an aspect ratio of 3.7 and a longitudinal LSPR peak at 760 nm was synthesized and stored in the fridge in darkness for use throughout the experiments. This ensured uniformity across all experimental parameters and eliminated the need for recalibration or adjustments to the machine learning model. All data were collected using this single batch, ensuring that the machine learning models were trained and validated under consistent conditions, which enhanced the robustness and reliability of the detection system. AuNRs were separated from unbound CTAB through centrifuging the mixture at 8000 rpm for 15 minutes; the supernatant was discarded, and the sediment was then dispersed in 15 mL of deionized water.

**Preparation of Britton-Robinson Buffer**

A Britton-Robinson buffer solution at pH 7.0 was prepared by adjusting the pH of a mixture containing 0.04 mol $L^{-1}$ boric acid ($H_3BO_3$), phosphoric acid ($H_3PO_4$), and acetic acid ($CH_3COOH$) using 0.2 mol $L^{-1}$ sodium hydroxide (NaOH) [2]. Additionally, all analyte stock solutions were prepared using deionized (DI) water to achieve the required concentration levels.

**Statistical Analysis**

The machine learning algorithms were run using the Origin Pro 2018 and MATLAB R2013a. To evaluate the qualitative performance of the probe, the obtained dataset matrix was analyzed using the widely recognized classification technique known as Linear Discriminant Analysis (LDA) along with Principal Component Analysis (PCA) to reduce the dimensionality of the dataset as required. The quantitative performance of the probe was assessed by applying the Partial Least Squares Regression (PLSR) technique to the acquired dataset matrices for each protein without any dimension reduction. Leave-one-out cross-validation was used to evaluate the models' predictive accuracy by iteratively training on all but one data point and testing on the excluded point. To represent the statistical significance of the classifications, 2D confidence ellipses were sketched around the class centroids, showing 95% confidence limits. The MVC1 toolbox a well-known toolbox in MATLAB R2013a was applied to perform the PLSR modeling and calculate analytical Figures of merit for multivariate calibration such as correlation coefficient ($R^2$), root-mean square error of calibration (RMSEC), cross-validation (RMSECV), prediction (RMSEP), sensitivity (SEN), analytical sensitivity (Anal. SEN), limit of detection (LOD), and limit of quantification (LOQ).

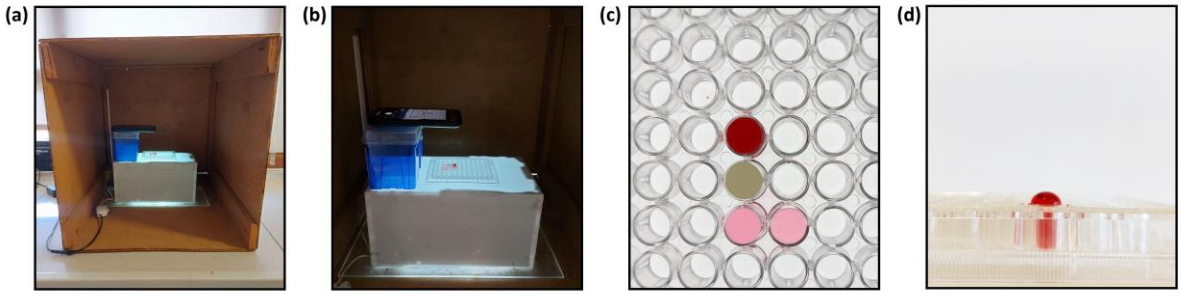


**Fig S1.** The experimental setup used for image capturing: (a) Constructed box used for placing the 96-well plate, (b) method for capturing color images, (c) captured color image ready to be cropped to a specific size, and (d) side view of the plate showing the convex surface of the solution after overfilling the well.

**Table S1.** Structural characteristics of the eight studied proteins.

| | Mw (kDa) | Total Amino Acid Count | Most Abundant Amino Acid | Oxidation and Cleavage Reactivity | | | | | Decarboxylation Reactivity | Specific Feature |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Cysteine Content | Methionine Content | Tryptophane Content | Tyrosine Content | Histidine Content | C-Terminal | |
| **Phosphatase (ACP)** | 94 | 855 | Alanine | 2 | 1 | 1 | 3 | 2-3 | Arginine | - |
| **Pepsin (Pep)** | 35 | 327 | Glutamic Acid | 1 | 1 | 1 | 3 | 4 | Glutamic acid | - |
| **Hemoglobin (Hem)** | 64.5 | 574 | Leucine | 2 | 2 | 1 | 6-7 | 10-15 | Histidine | four $Fe^{2+}$ ions |
| **Transferrin (TRF)** | 76 | 679 | Glutamic Acid | 2 | 1 | 1 | 2 | 6-8 | Lysine | two $Fe^{3+}$ ions |
| **Immunoglobin G (IgG)** | 150 | 1320 | Glutamic Acid | 10-12 | 2 | 1 | 10-12 | 7-9 | Lysine | - |
| **Lysozyme (Lys)** | 14 | 129 | Glutamic Acid | 2 | 2 | 1 | 5 | 6 | Leucine | - |
| **Fibrinogen (Fib)** | 340 | 610 | Glutamic Acid | 6 | 2 | 1 | 10-12 | 3-4 | Arginine | - |
| **Albumin (HAS)** | 66.5 | 585 | Glutamic Acid | 35 | 2-3 | 1 | 15 | 10-12 | Leucine | - |

*The reported values are referenced from UniProt: The Universal Protein Knowledgebase (2024).

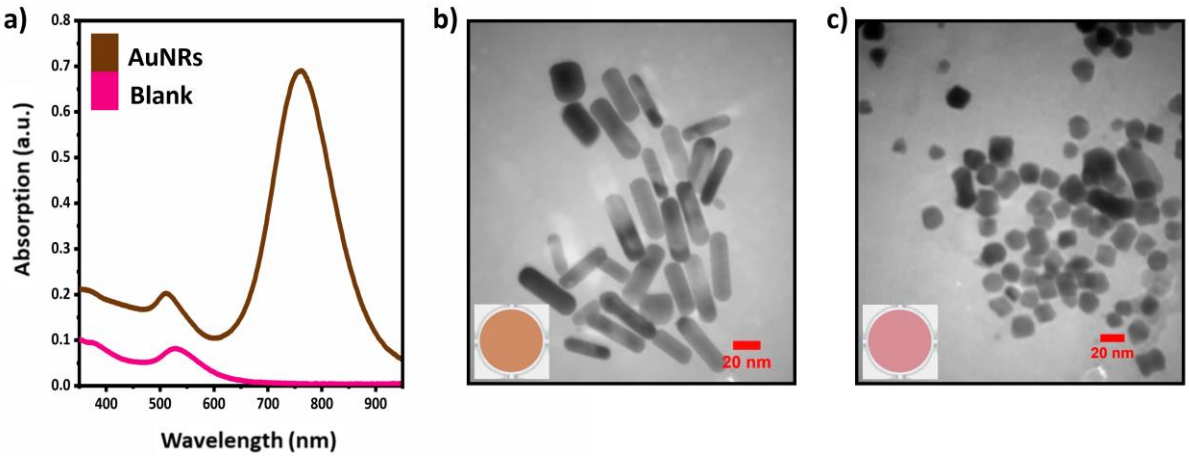


**Fig S2.** (a) Absorbance spectra of the synthesized AuNRs (brown line) and the AuNRs after etching by NBS (pink line). TEM images and corresponding images of (b) the synthesized AuNRs and (c) the AuNRs after etching by NBS.

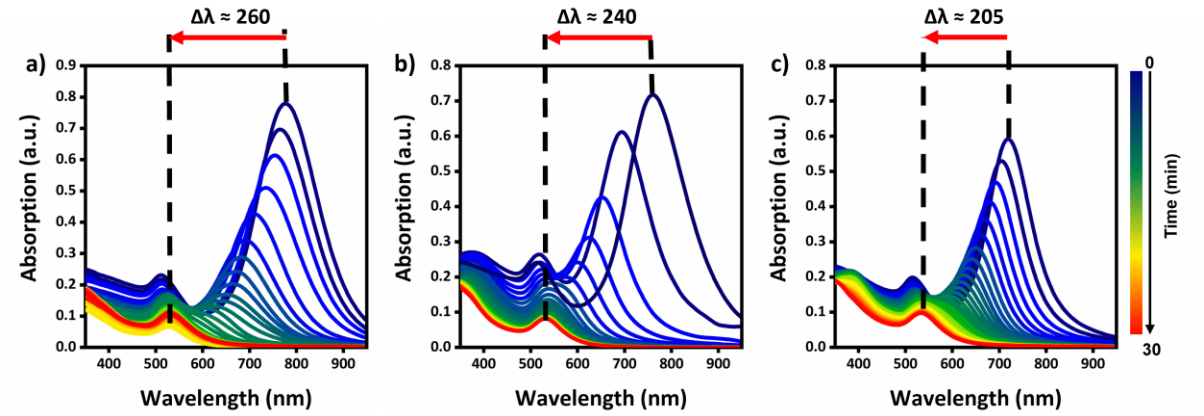


**Fig S3.** Time-dependent UV-Vis absorption spectra variations of differently-sized AuNRs with (a) 780 nm, (b) 760 nm, (c) 725 nm longitudinal peaks during etching process by NBS (etching duration: 30 min, Scan Rate: 2 scans/min).

**Table S2.** pH of the isoelectric point of eight proteins (i.e., ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA).

| Sample | pH of the isoelectric point |
|---|---|
| **ACP** | 4.7 |
| **Pep** | 1.0 |
| **Hem** | 7.5 |
| **TRF** | 5.6 |
| **IgG** | 7.4 |
| **Lys** | 9.4 |
| **Fib** | 5.8 |
| **HSA** | 4.7 |

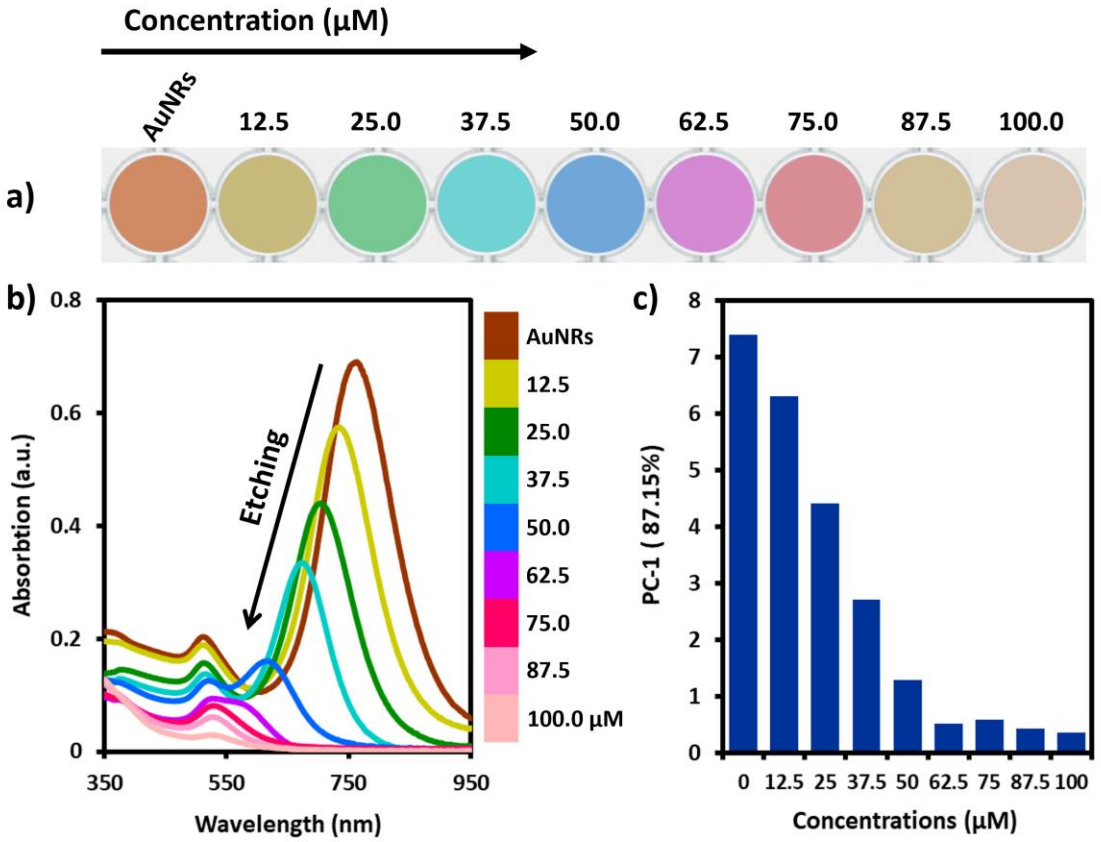


**Fig S4.** Effect of NBS concentration on the multidimensional colorimetric responses of AuNRs, (a) color images of the probe's solution, (b) the absorption spectra of the proposed probe, and (c) the variation of PC-1 for the corresponding etching absorbance response as a function of NBS concentration.

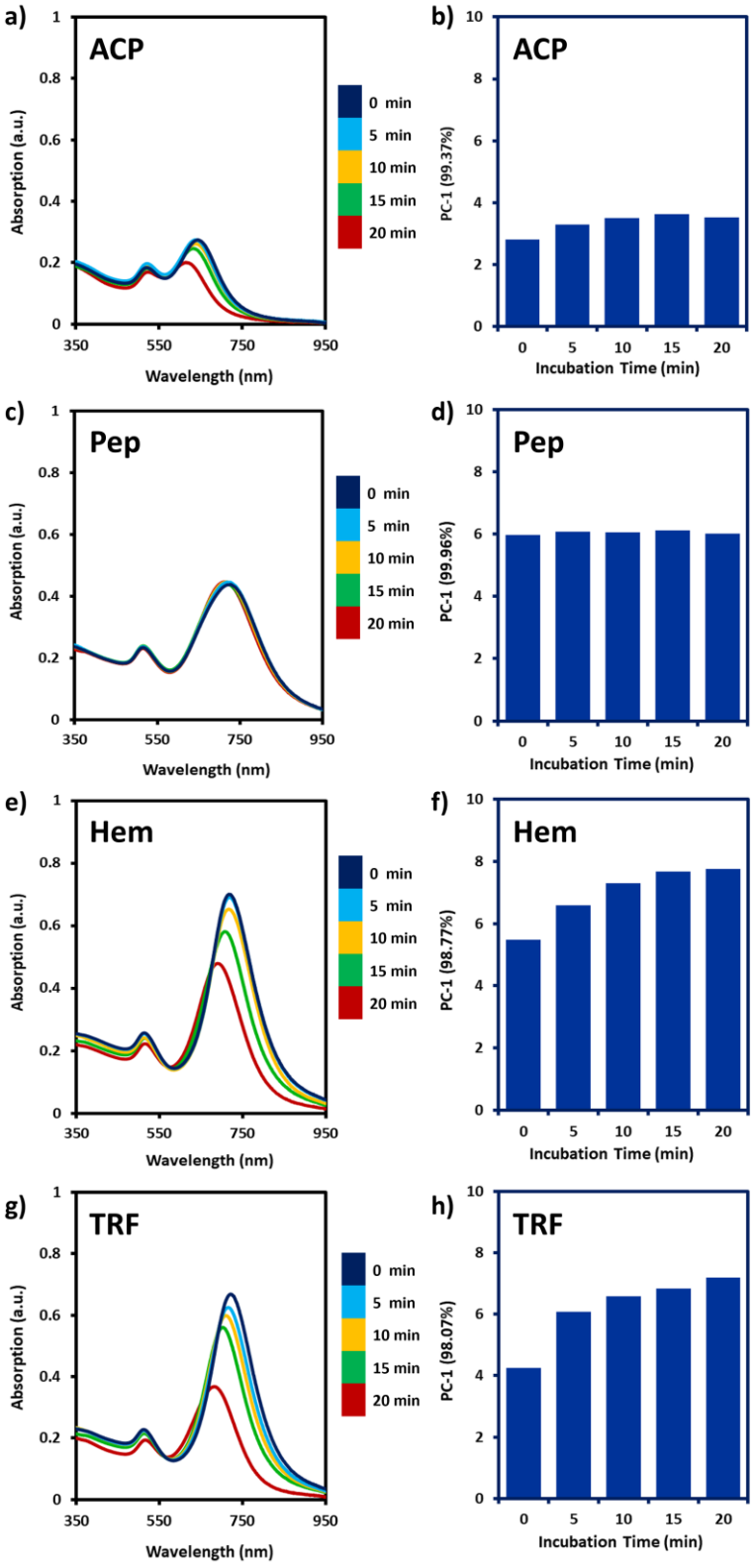


**Fig S5.** Effect of incubation time on the responses of a multidimensional colorimetric probe. Absorbance spectra and variation of PC-1 for corresponding inhibited etching absorbance response profiles of the probe in the presence of (a, b) ACP, (c, d) Pep, (e, f) Hem, and (g, h) TRF.

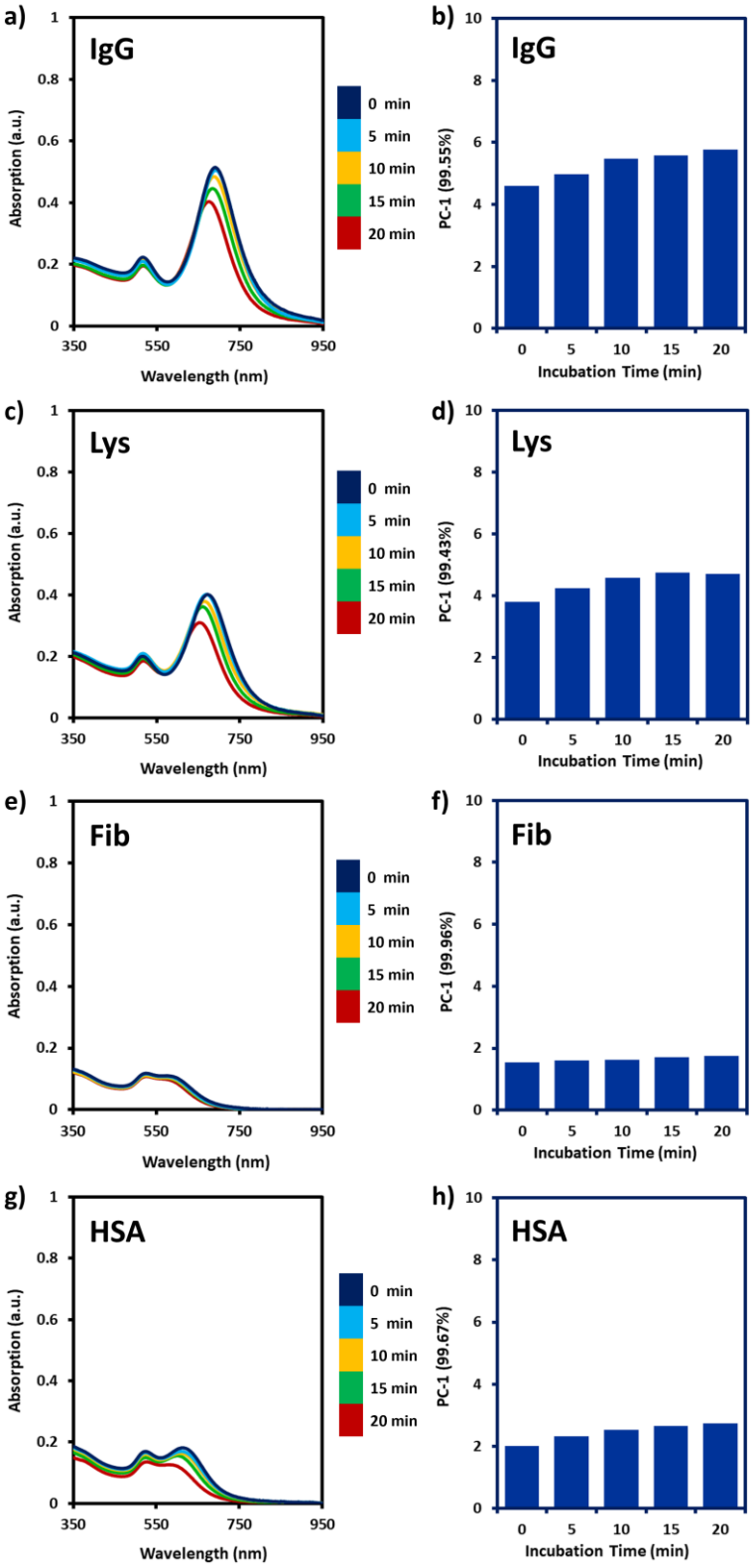


**Fig S6.** Effect of incubation time on the responses of a multidimensional colorimetric probe. Absorbance spectra and variation of PC-1 for corresponding inhibited etching absorbance response profiles of the probe in the presence of (a, b) IgG, (c, d) Lys, (e, f) Fib, and (g, h) HSA.

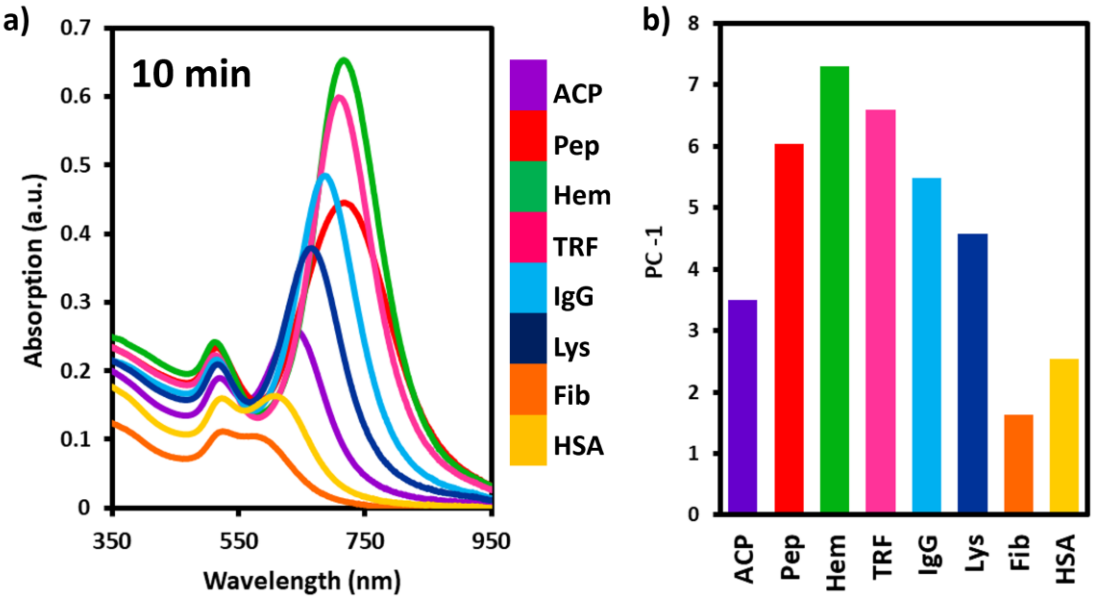


**Fig S7.** Response profile of all proteins with 10 min incubation time. (a) Absorbance spectra and (b) variation of PC-1 for corresponding inhibited etching absorbance response profiles of the multidimensional colorimetric probe.

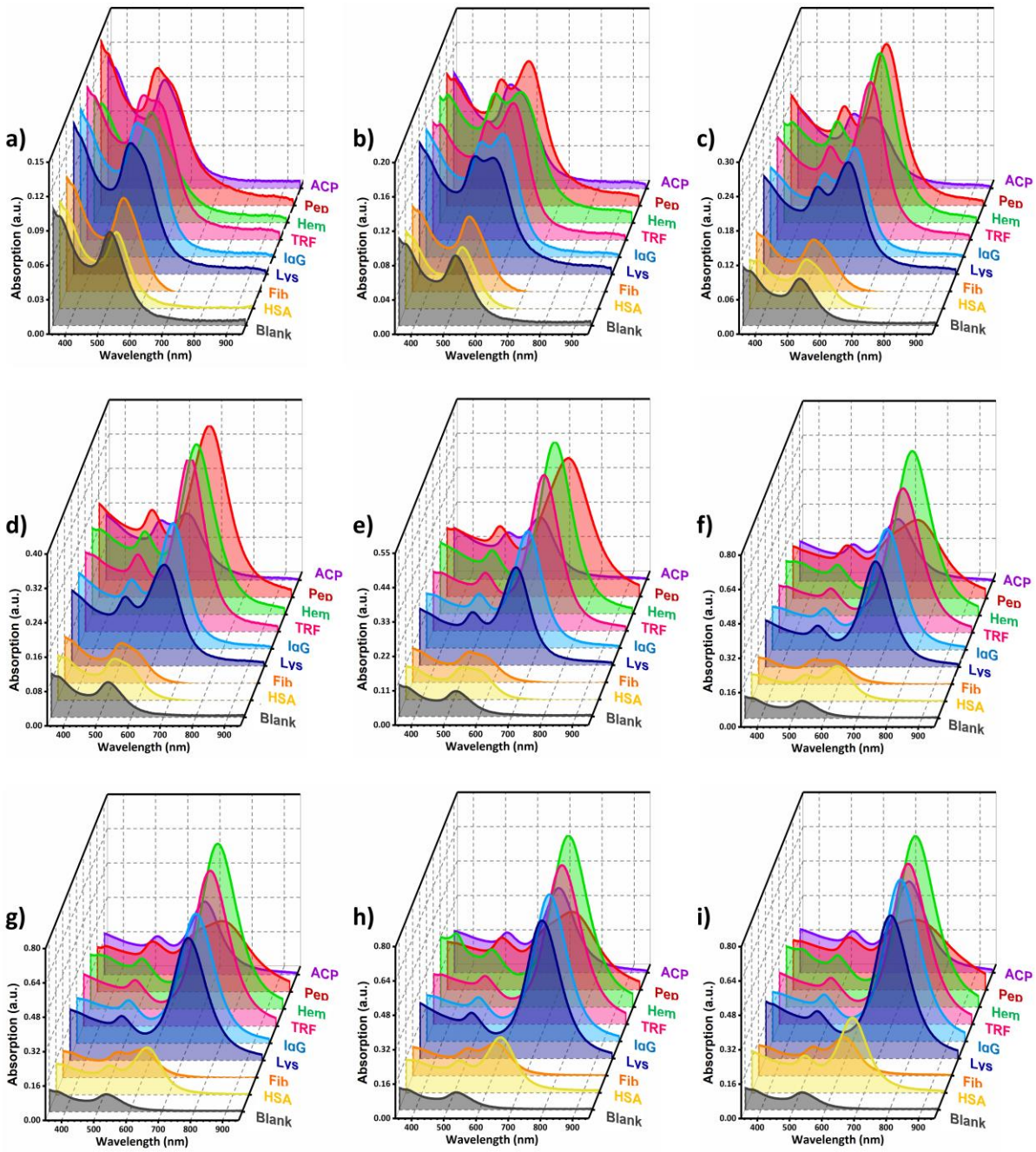


**Fig S8.** Absorption spectra variation of the multidimensional colorimetric probe in the presence of (a) 1.0, (b) 2.5, (c) 5.0, (d) 7.5, (e) 10.0, (f) 17.5, (g) 20.0, (h) 22.5, and (i) 25.0 ppm of different proteins.

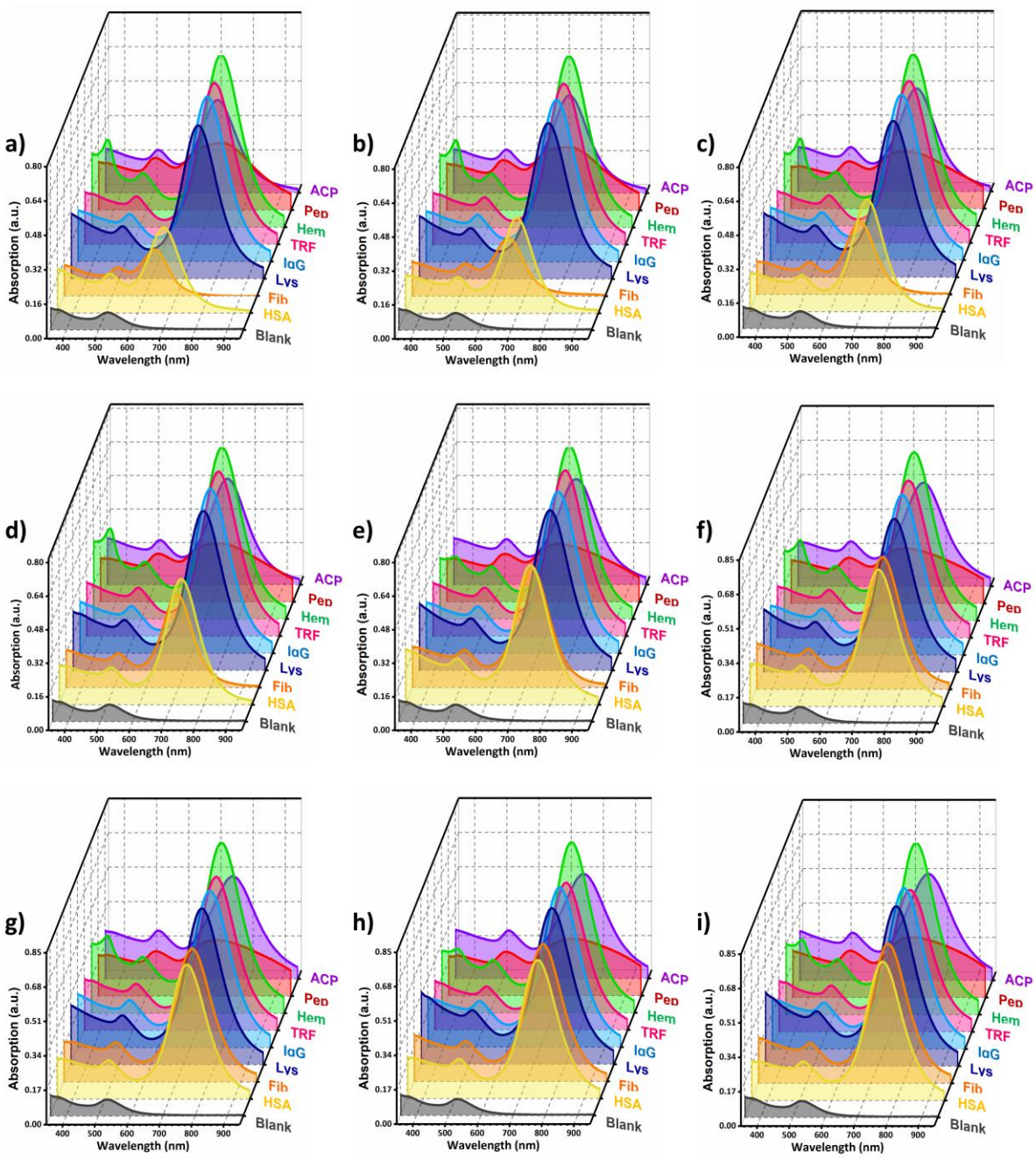


**Fig S9.** Absorption spectra variation of the multidimensional colorimetric probe in the presence of (a) 27.5, (b) 30.0, (c) 32.5, (d) 35.0, (e) 37.5, (f) 40.0, (g) 45.0, (h) 50.0, and (i) 75.0 ppm of different proteins.

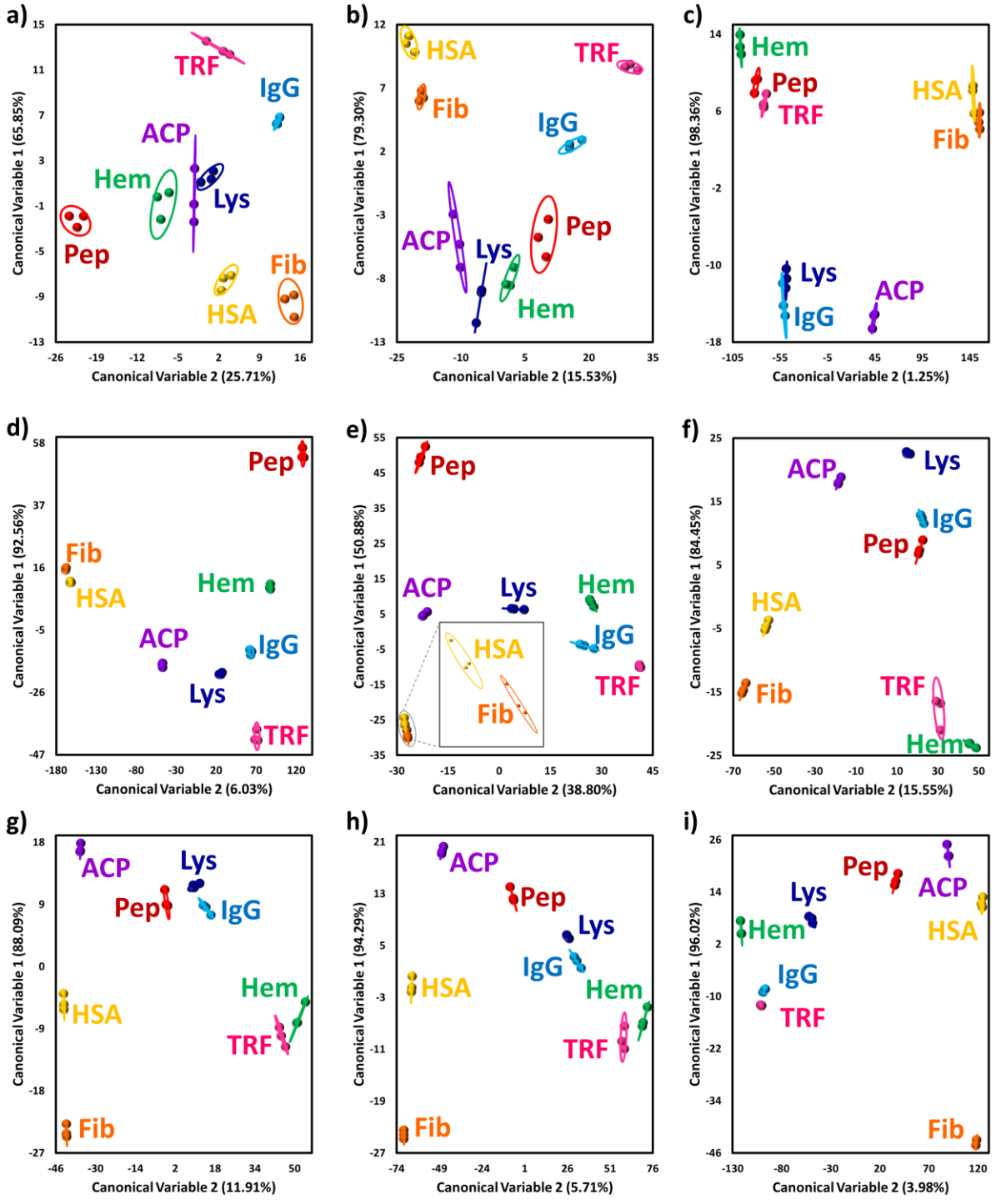


**Fig S10.** 2D LDA score plot discriminating different proteins at concentrations of (a) 1.0, (b) 2.5, (c) 5.0, (d) 7.5, (e) 10.0, (f) 17.5, (g) 20.0, (h) 22.5, and (i) 25.0 ppm.

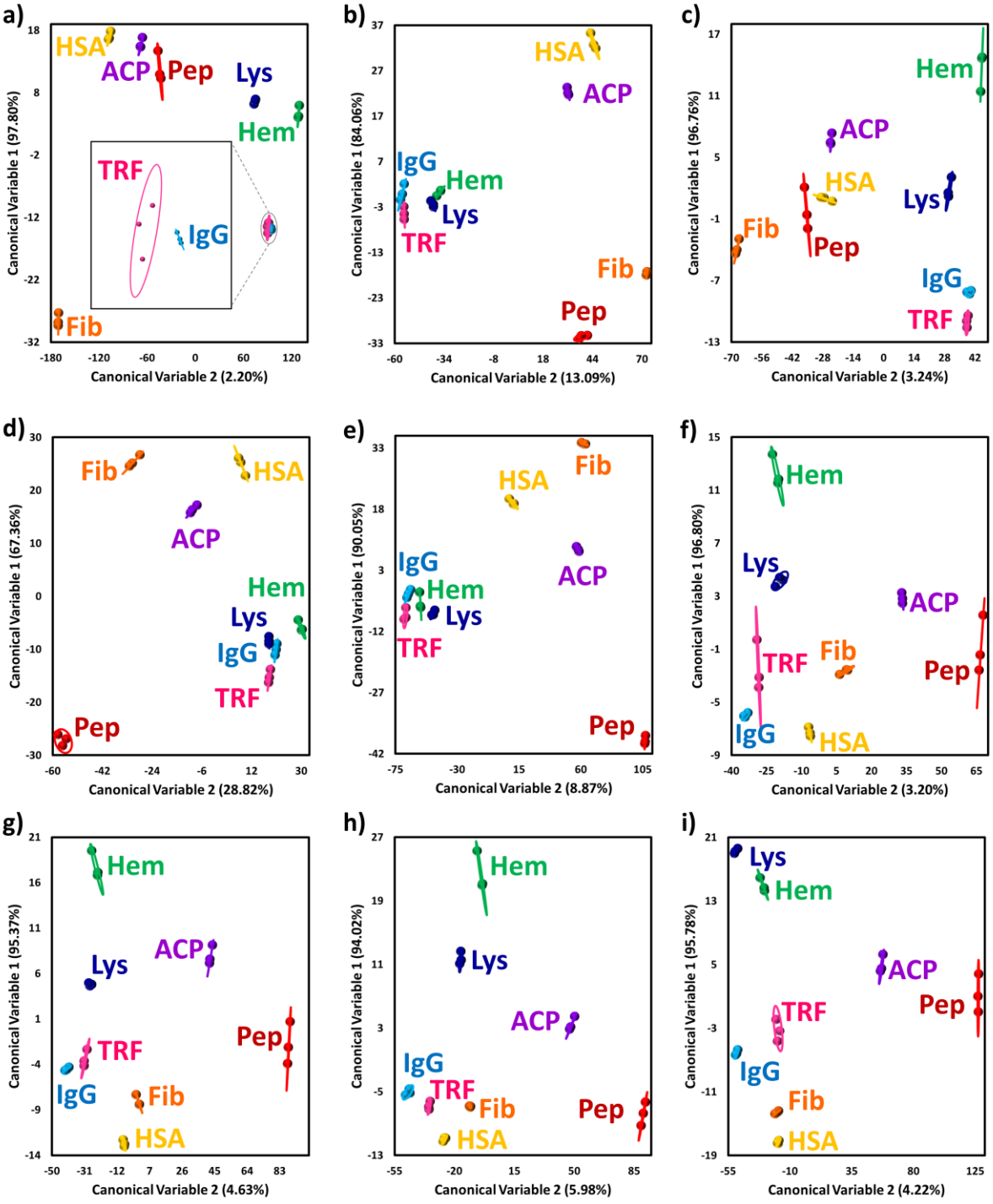


**Fig S11.** 2D LDA score plot discriminating different proteins at concentrations of (a) 27.5, (b) 30.0, (c) 32.5, (d) 35.0, (e) 37.5, (f) 40.0, (g) 45.0, (h) 50.0, and (i) 75.0 ppm.

**Table S3.** Jackknifed classification matrix for the discrimination of eight protein samples (i.e., ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA) at 10.0 ppm concentration.

| Nominal Class \ Predicted Class | ACP | Pep | Hem | TRF | IgG | Lys | Fib | HSA | Total | Sensitivity | Specificity | Precision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACP | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| Pep | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| Hem | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| TRF | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| IgG | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| Lys | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 100.00% | | | |
| Fib | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 100.00% | | | |
| HSA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | | | |
| Total | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 24 | 100 | 100 | 100 |
| | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 100.00% | | | |

**Table S4.** Jackknifed classification matrix for the discrimination of eight protein samples (i.e., ACP, Pep, Hem, TRF, IgG, Lys, Fib, and HSA) at 15.0 ppm concentration.

| Nominal Class \ Predicted Class | ACP | Pep | Hem | TRF | IgG | Lys | Fib | HSA | Total | Sensitivity | Specificity | Precision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACP | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| Pep | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| Hem | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| TRF | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| IgG | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| Lys | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 100.00% | | | |
| Fib | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 100.00% | | | |
| HSA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | | | |
| Total | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 24 | 100 | 100 | 100 |
| | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 12.50% | 100.00% | | | |

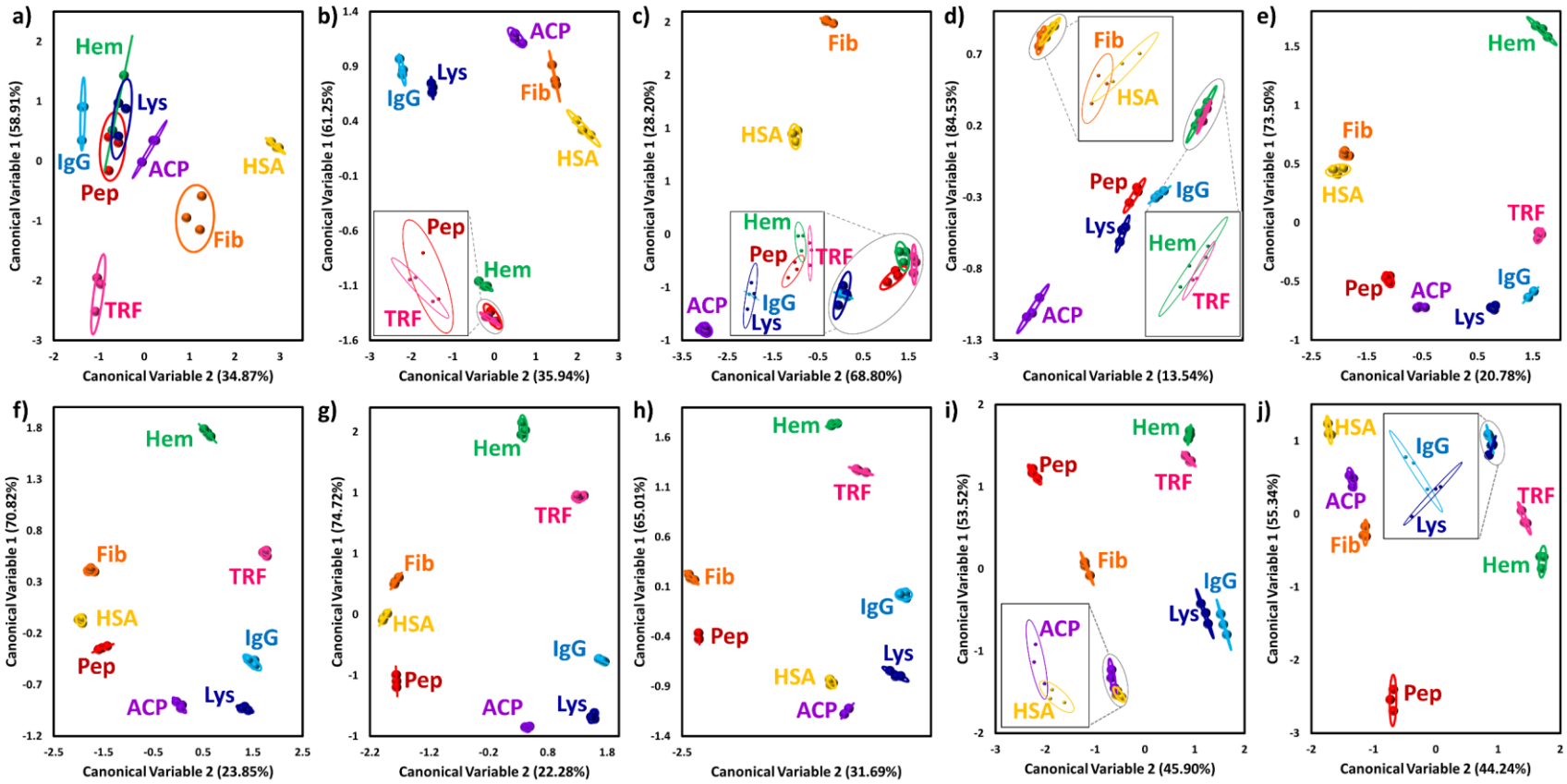


**Fig S12.** 2D LDA score plot discriminating different proteins based on CIELAB parameters at concentrations of (a) 1, (b) 2.5, (c) 5, (d) 7.5, (e) 10, (f) 12.5, (g) 15, (h) 17.5, (i) 20.0, and (j) 22.5 ppm.

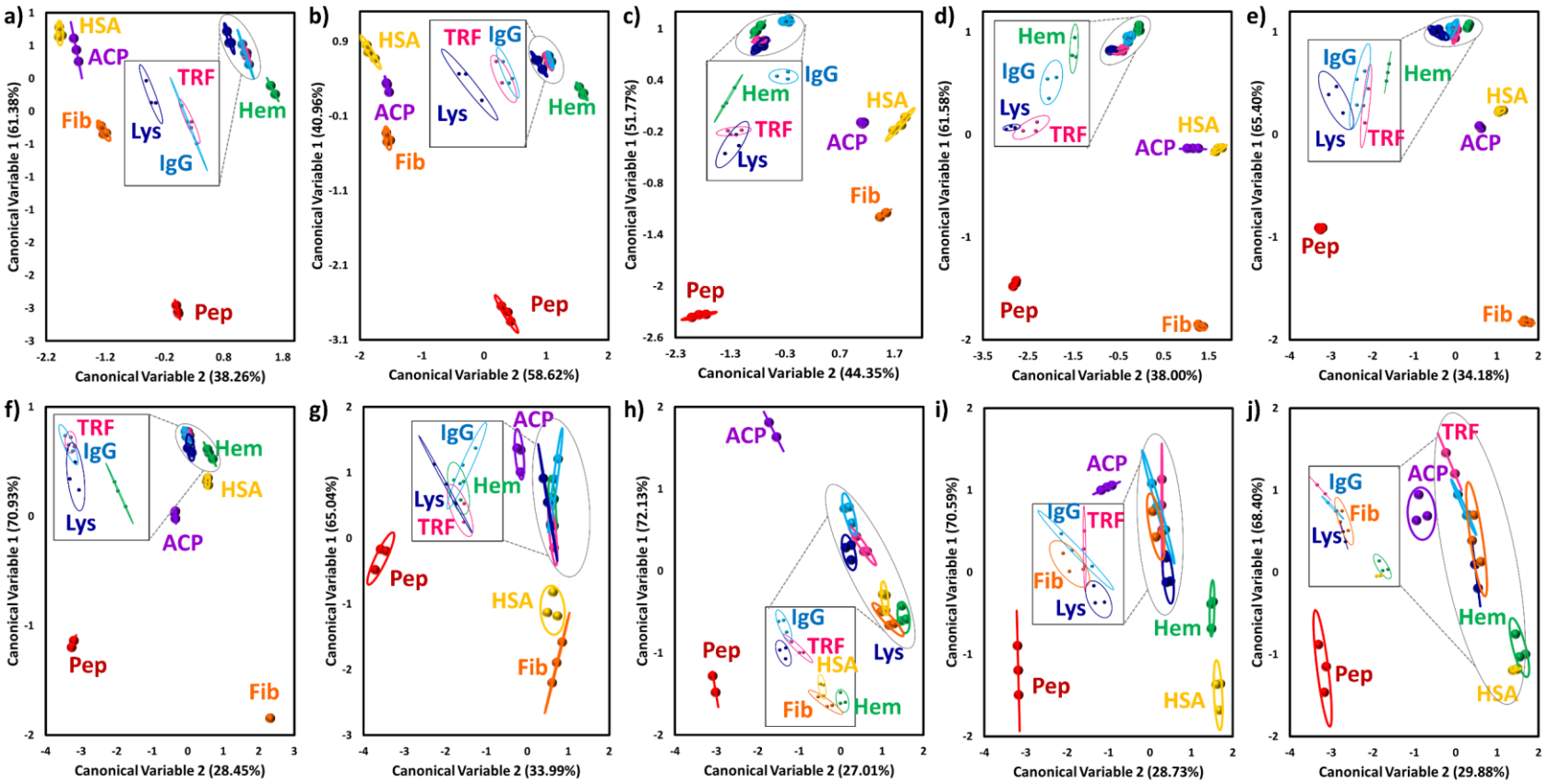


**Fig S13.** 2D LDA score plot discriminating different proteins based on CIELAB parameters at concentrations of (a) 25.0, (b) 27.5, (c) 30.0, (d) 32.5, (e) 35.0, (f) 37, (g) 40.0, (h) 45.0, (i) 50.0, and (j) 75.0 ppm.

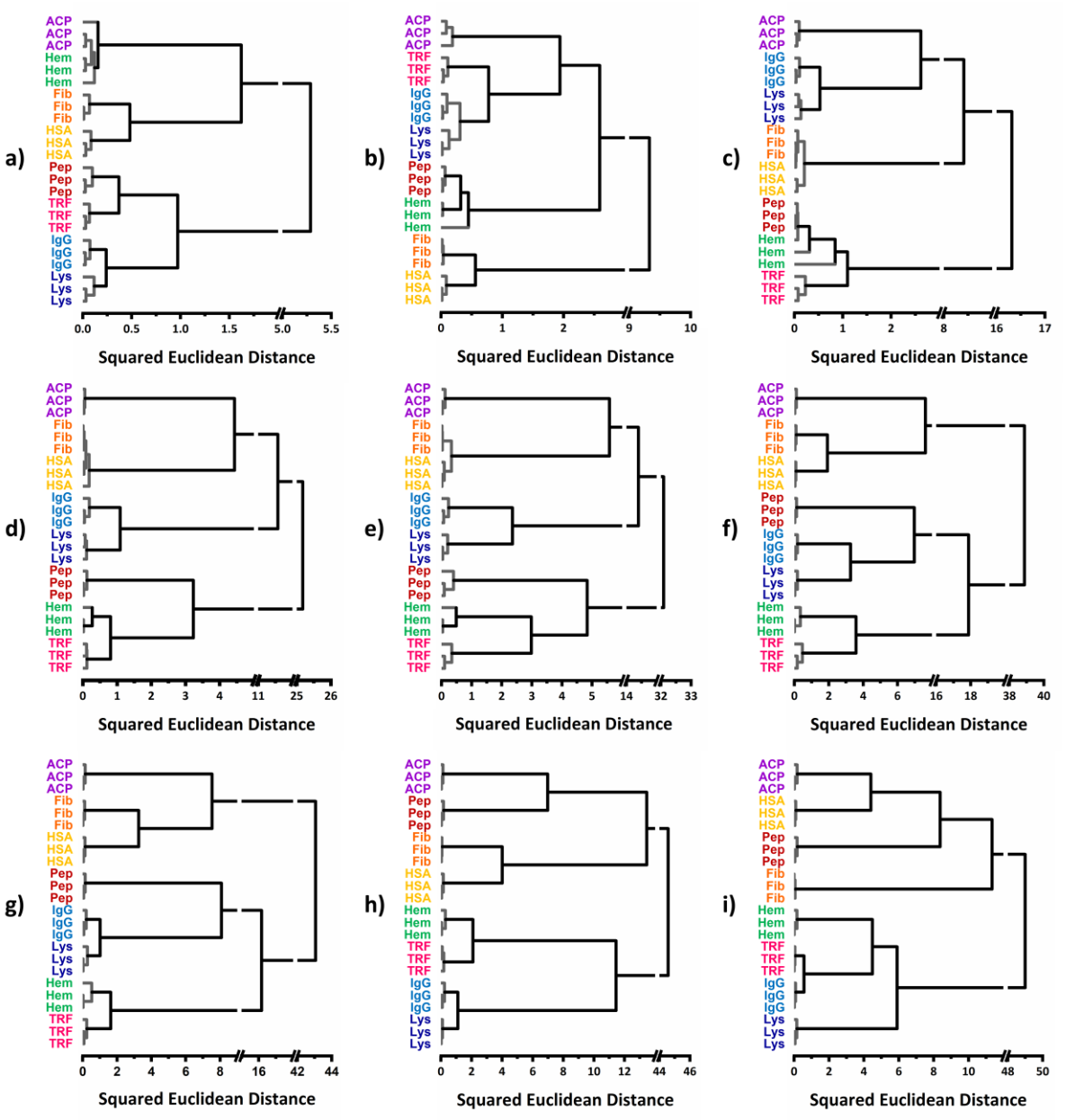


**Fig S14.** HCA dendrogram with Ward method for different proteins at concentrations of (a) 1.0, (b) 2.5, (c) 5.0, (d) 7.5, (e) 10.0, (f) 17.5, (g) 20.0, (h) 22.5, and (i) 25.0 ppm. No confusion in the classification of proteins was observed even in three replicates of the experiments.

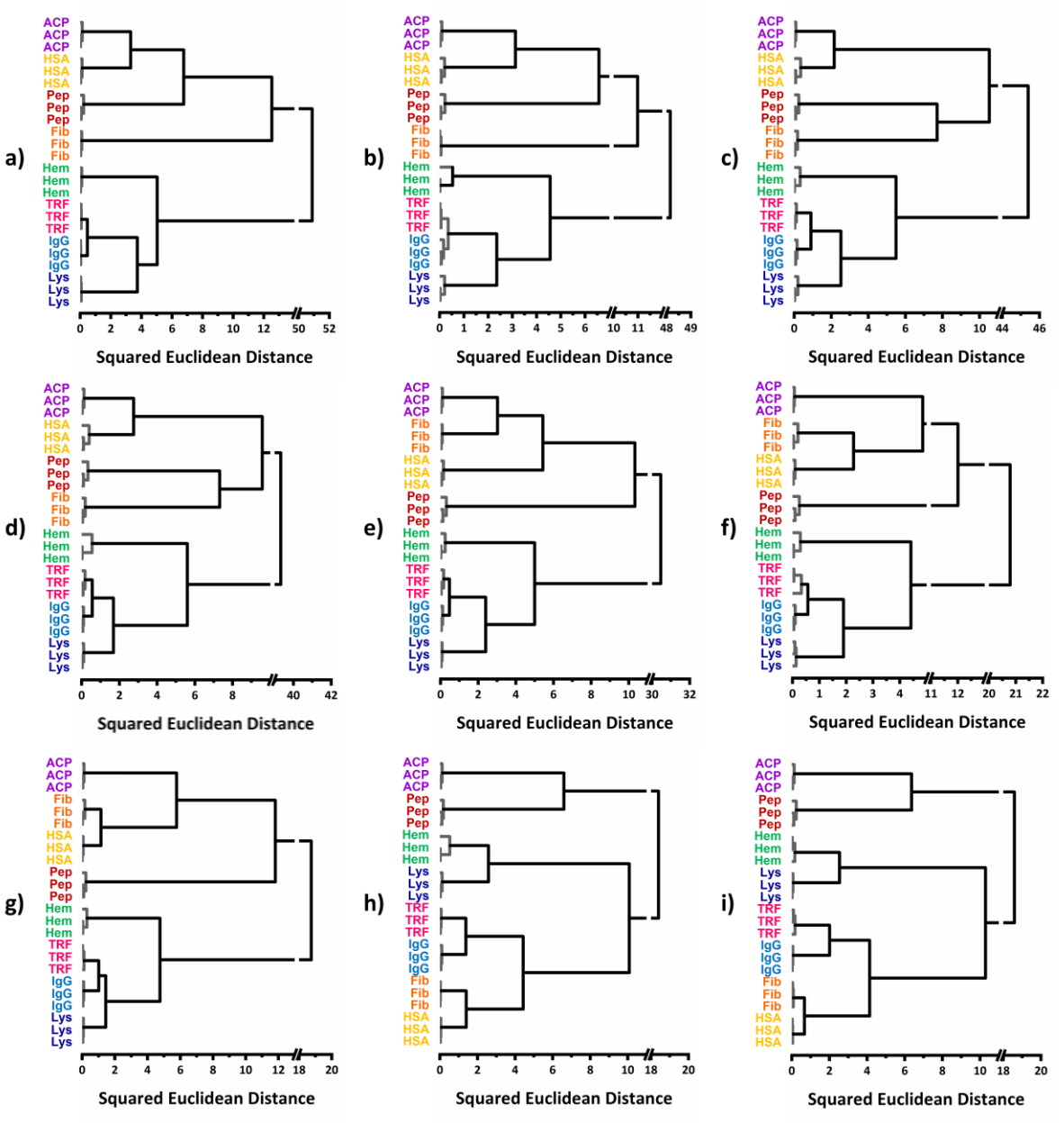


**Fig S15.** HCA dendrogram with Ward method for different proteins at concentrations of (a) 27.5, (b) 30.0, (c) 32.5, (d) 35.0, (e) 37.5, (f) 40.0, (g) 45.0, (h) 50.0, and (i) 75.0 ppm. No confusion in the classification of proteins was observed even in three replicates of the experiments.

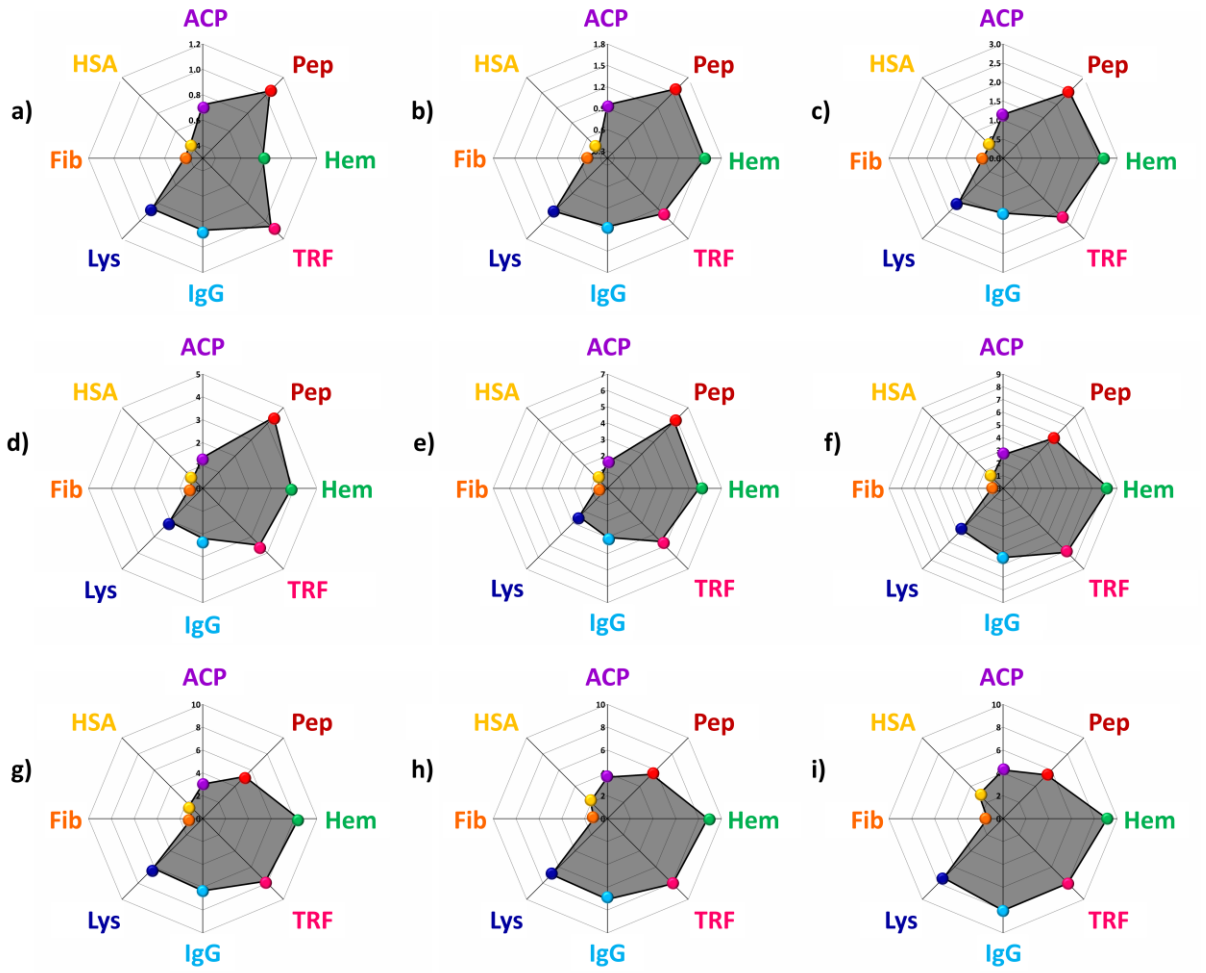


**Fig S16.** Radar plot fingerprints in the presence of different proteins at concentrations of (a) 1.0, (b) 2.5, (c) 5.0, (d) 7.5, (e) 10.0, (f) 17.5, (g) 20.0, (h) 22.5, and (i) 25.0 ppm. The first principal components derived from PCA were the response signals utilized to illustrate the radar plot.

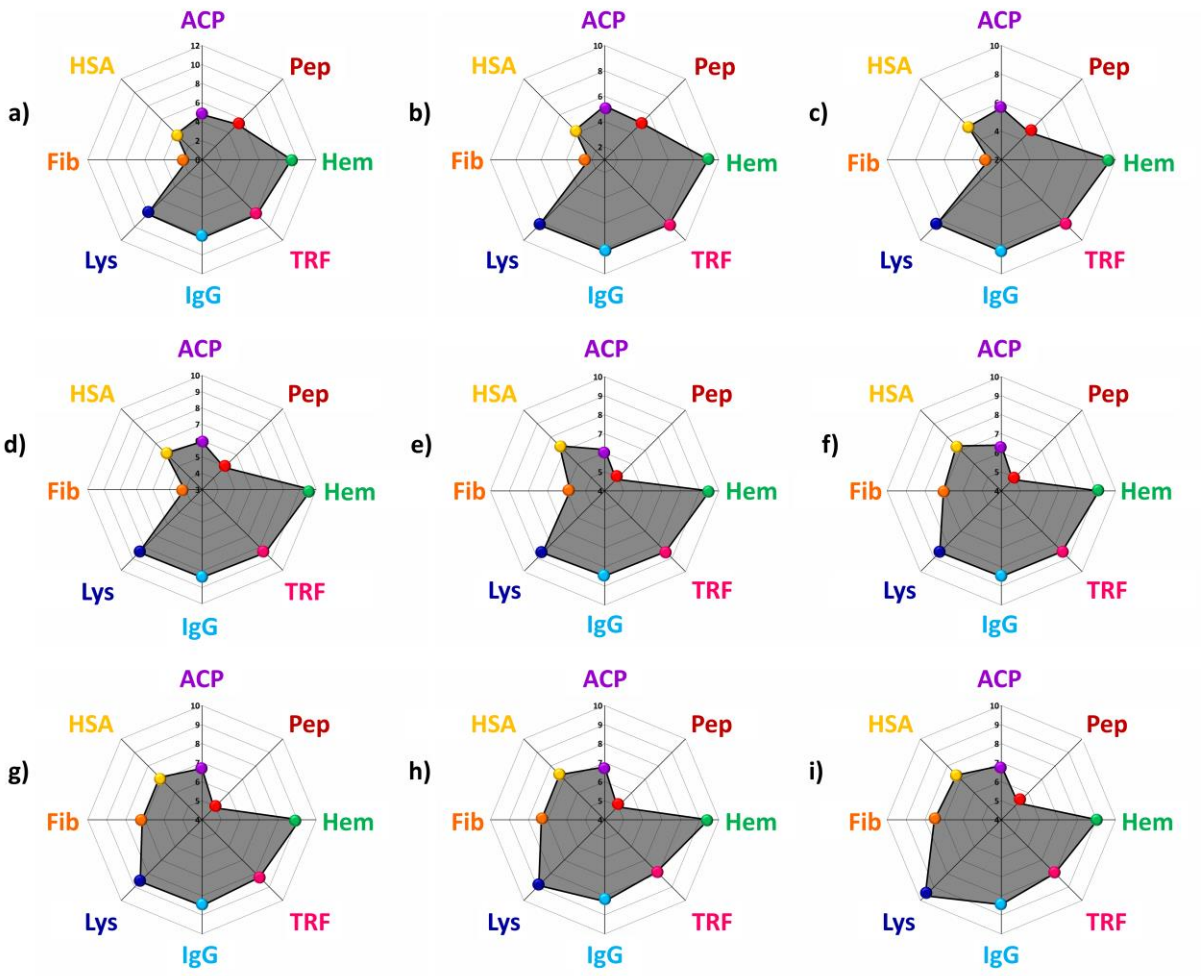


**Fig S17.** Radar plot fingerprints in the presence of different proteins at concentrations of (a) 27.5, (b) 30.0, (c) 32.5, (d) 35.0, (e) 37.5, (f) 40.0, (g) 45.0, (h) 50.0, and (i) 75.0 ppm. The first principal components derived from PCA were the response signals utilized to illustrate the radar plot.

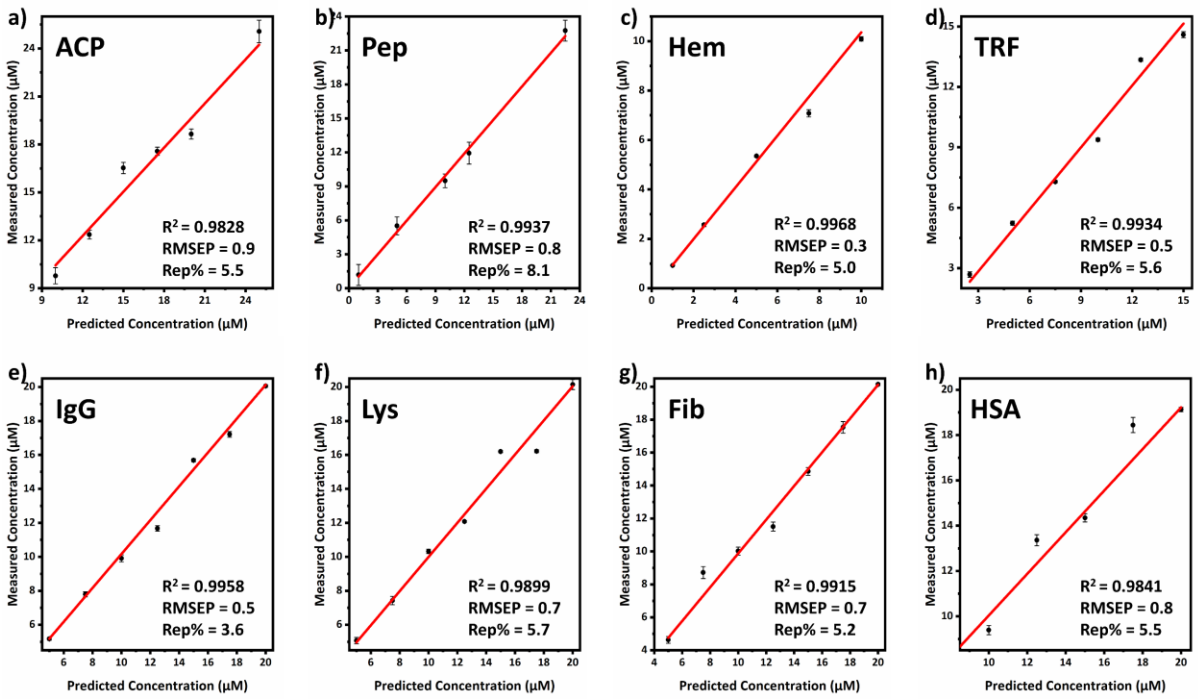


**Fig S18.** Predicted vs. measured concentration plots as multivariate calibration by PLS-regression based on CIELAB parameters for (a) ACP, (b) Pep, (c) Hem, (d) TRF, (e) IgG, (f) Lys, (g) Fib, and (h) HSA in their entire concentration ratio range.

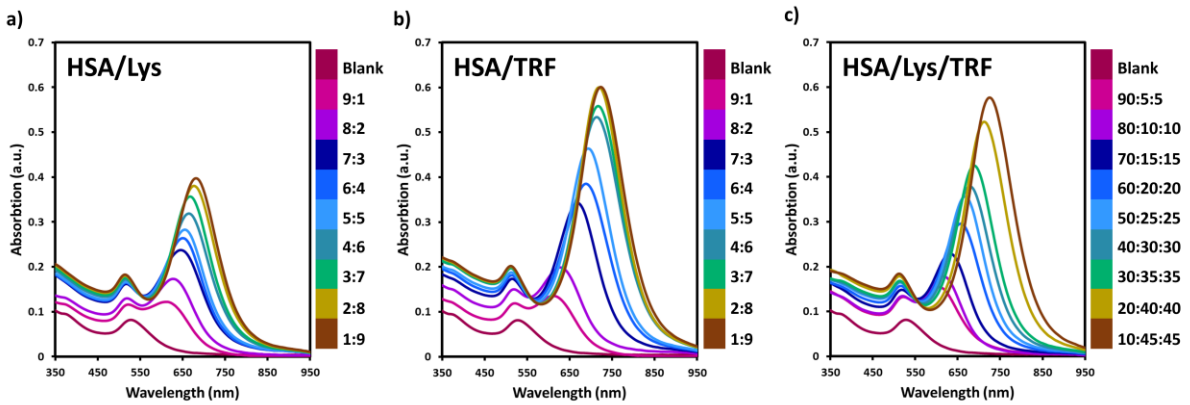


**Fig S19.** UV-Vis spectra variation of the multidimensional colorimetric probe in the presence of binary protein mixtures (a) HSA/Lys, (b) HSA/TRF, and ternary protein mixtures (c) HSA/Lys/TRF in their entire concentration ratio range (total concentration is 15.0 ppm).

**Table S5.** Jackknifed classification matrix for the discrimination of individual (HSA/Lys) binary protein mixture samples.

| Nominal Class | Predicted Class: 9:1 | 8:2 | 7:3 | 6:4 | 5:5 | 4:6 | 3:7 | 2:8 | 1:9 | Total | Sensitivity | Specificity | Precision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9:1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 8:2 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 7:3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 6:4 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 5:5 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 4:6 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 3:7 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 100.00% | | | |
| 2:8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 100.00% | | | |
| 1:9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | | | |
| Total | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 27 | 100 | 100 | 100 |
| | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 100.00% | | | |

**Table S6.** Jackknifed classification matrix for the discrimination of individual (HSA/TRF) binary protein mixture samples.

| Nominal Class | Predicted Class: 9:1 | 8:2 | 7:3 | 6:4 | 5:5 | 4:6 | 3:7 | 2:8 | 1:9 | Total | Sensitivity | Specificity | Precision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9:1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 8:2 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 7:3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 6:4 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 5:5 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 4:6 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| 3:7 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 100.00% | | | |
| 2:8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 100.00% | | | |
| 1:9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 100 | 100 | 100 |
| | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | | | |
| Total | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 27 | 100 | 100 | 100 |
| | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 100.00% | | | |

**Table S7.** Jackknifed classification matrix for the discrimination of individual (HSA/Lys/TRF) ternary protein mixture samples.

| | | ***Predicted Class*** | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ***90:5:5*** | ***80:10:10*** | ***70:15:15*** | ***60:20:20*** | ***50:25:25*** | ***40:30:30*** | ***30:35:35*** | ***20:40:40*** | ***10:45:45*** | ***Total*** | ***Sensitivity*** | ***Specificity*** | ***Precision*** |
| ***Nominal Class*** | ***90:5:5*** | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***80:10:10*** | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***70:15:15*** | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***60:20:20*** | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***50:25:25*** | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***40:30:30*** | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***30:35:35*** | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 0.00% | 100.00% | | | |
| | ***20:40:40*** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 0.00% | 100.00% | | | |
| | ***10:45:45*** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 100 | 100 | 100 |
| | | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% | 100.00% | | | |
| | ***Total*** | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 27 | 100 | 100 | 100 |
| | | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 11.11% | 100.00% | | | |

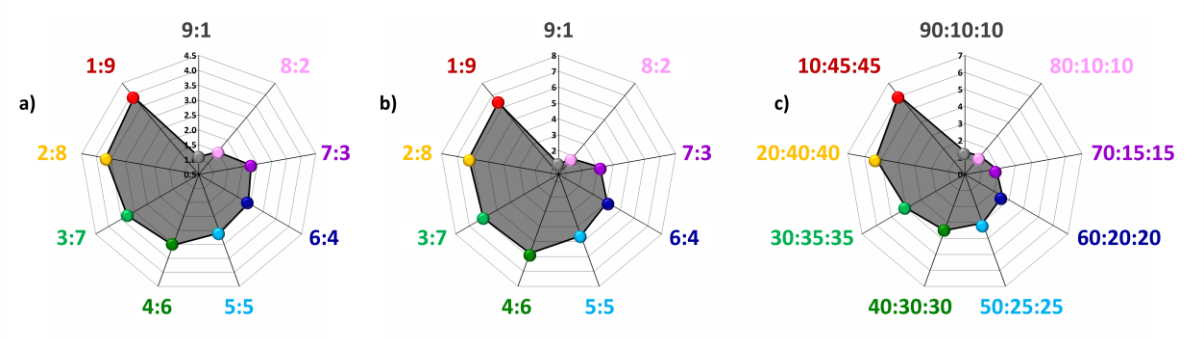


**Fig S20.** Radar plot fingerprints in the presence of binary protein mixtures (a) HSA/Lys, (b) HSA/TRF, and ternary protein mixtures (c) HSA/Lys/TRF in their entire concentration ratio range.

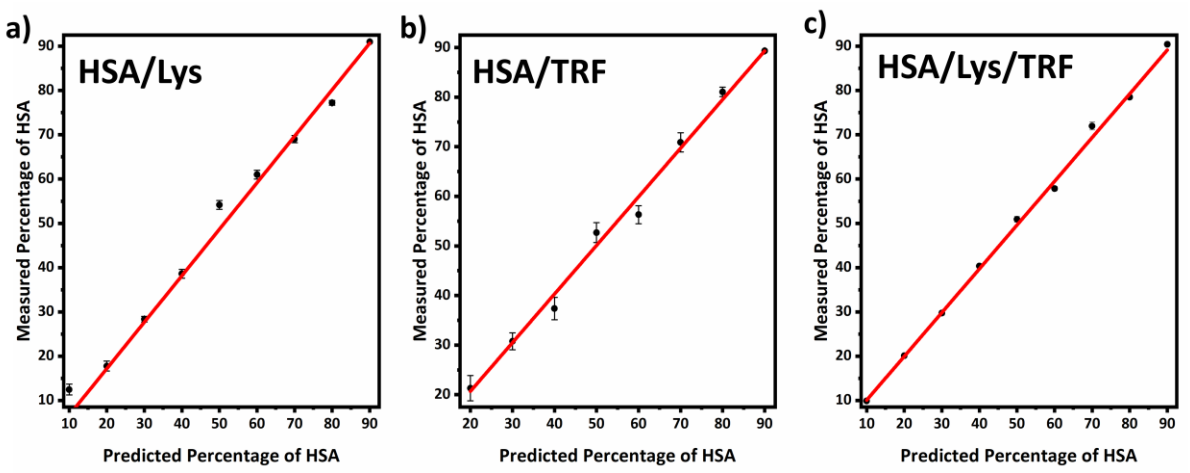


**Fig S21.** Predicted vs. measured concentration plots as multivariate calibration by PLS-regression for binary protein mixtures (a) HSA/Lys, (b) HSA/TRF, and ternary protein mixtures (c) HSA/Lys/TRF.

**Table S8.** LDA posterior probability outcomes for identifying of protein unknown samples in human urine. (All 12 samples were given as a test set to the pre-trained LDA model.

| ***Alleged*** | ***Post probabilities*** | | | | | | | | ***Allocated*** |
|---|---|---|---|---|---|---|---|---|---|
| | ***ACP*** | ***Pep*** | ***Hem*** | ***TRF*** | ***IgG*** | ***Lys*** | ***Fib*** | ***HSA*** | |
| ACP-Real | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ACP |
| ACP-Real | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ACP |
| ACP-Real | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ACP |
| Pep-Real | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Pep |
| Pep-Real | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Pep |
| Pep-Real | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Pep |
| TRF-Real | 0 | 0 | 8.96E-20 | 1 | 2.80E-123 | 0 | 0 | 0 | TRF |
| TRF-Real | 0 | 0 | 2.71E-40 | 1 | 1.20E-108 | 0 | 0 | 0 | TRF |
| TRF-Real | 0 | 0 | 7.23E-60 | 1 | 1.77E-106 | 0 | 0 | 0 | TRF |
| Lys-Real | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | Lys |
| Lys-Real | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | Lys |
| Lys-Real | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | Lys |

# References

1. M. R. Hormozi-Nezhad, H. Robatjazi and M. Jalali-Heravi, *Analytica chimica acta*, 2013, **779**, 14-21.
2. H. T. S. Britton and R. A. Robinson, *Journal of the Chemical Society (Resumed)*, 1931, 1456-1462.