

Protein discrimination using erythrosin B-based GUMBOS in combination with UV–Vis spectroscopy and chemometrics

Ana M.O. Azevedo^a, Clara Sousa^{b,*}, Mi Chen^c, Caitlan E. Ayala^c, Rocío L. Pérez^{c,d}, João L. M. Santos^a, Isiah M. Warner^c, M. Lúcia M.F.S. Saraiva^{a,**}

^a LAQV, REQUIMTE, Departamento de Ciências Químicas, Laboratório de Química Aplicada, Faculdade de Farmácia, Universidade do Porto, Rua Jorge Viterbo Ferreira 228, 4050-313, Porto, Portugal

^b Universidade Católica Portuguesa, CBQF – Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Rua Diogo Botelho 1327, 4169-005, Porto, Portugal

^c Department of Chemistry, Louisiana State University, Baton Rouge, LA, 70803, United States

^d Department of Chemistry and Biochemistry, Georgia Southern University, Statesboro, GA, 30458, United States

ARTICLE INFO

Keywords:

Erythrosin B

GUMBOS

PLSDA

Protein discrimination

ABSTRACT

GUMBOS (Group of Uniform Materials Based on Organic Salts) have recently emerged as interesting materials for protein analysis due to their unique features and high tunability. In this regard, four novel erythrosin B (EB)-based GUMBOS were synthesized and their potential to discriminate among proteins with distinct properties (e.g., size, charge, and hydrophobicity) was assessed. These solid-phase materials were prepared using a single-step metathesis reaction between EB and various phosphonium and ammonium cations, namely tetrabutylphosphonium (P_{4444}^{+}), tributylhexadecylphosphonium (P_{44416}^{+}), tetrabutylammonium (N_{4444}^{+}), and benzyltrimethylhexadecylammonium (BDHA⁺). Subsequently, the effect of pH (3.0, 4.5, and 6.0) and reaction time (5, 10, and 15 min) on the discriminatory power of synthesized GUMBOS was evaluated. Absorption spectra resulting from the interaction between EB-based GUMBOS and proteins were analyzed using partial least squares discriminant analysis (PLSDA). Unlike time, the pH value was determined to have influence over GUMBOS discrimination potential. Correct protein assignments varied from 86.5% to 100.0%, and the best discriminatory results were observed for $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$ at pH 6.0. Additionally, these two GUMBOS allowed discrimination of protein mixtures containing different ratios of albumin and myoglobin, which appeared as individualized clusters in the PLSDA scores plots. Overall, this study showcases EB-based GUMBOS as simple synthetic targets to provide a label-free, cost-effective, rapid, and successful approach for discrimination of single proteins and their mixtures.

1. Introduction

Proteins have been used as biomarkers for early diagnosis of various diseases, customization of treatment plans, and prognostic stratification of patients. For example, elevated levels of serum myoglobin (Mb) and troponin I play an important role in the acute coronary syndrome

diagnosis and are predictive of a higher 5-year mortality risk [1,2]. High concentrations of biliary calprotectin and serum interleukin-8 are associated with disease severity in primary sclerosing cholangitis and transplant-free survival [3,4]. In fact, a combined analysis of multiple biomarkers has better accuracy for assessing diagnoses, providing disease progression predictions, or evaluating therapeutic responses when

Abbreviations: Alb, Albumin; BDHA, Benzyltrimethylhexadecylammonium; $CDCl_3$, Deuterated chloroform; α -ChT, α -Chymotrysin; DMSO, Dimethylsulfoxide; DSC, Differential scanning calorimetry; EB, Erythrosin B; ELISA, Enzyme-linked immunosorbent assay; FTIR, Fourier transform infrared; GUMBOS, Group of uniform materials based on organic salts; HCA, Hierarchical cluster analysis; HRMS, High-resolution mass spectrometry; IL, Ionic liquid; $K_{O/W}$, Octanol/water partition coefficient; LDA, Linear discriminant analysis; LV, Latent variable; Lyz, Lysozyme; Mb, Myoglobin; NMR, Nuclear magnetic resonance; N_{4444} , Tetrabutylammonium; PCA, Principal component analysis; PLSDA, Partial least squares discriminant analysis; 2D-PLSDA, Two-dimensional partial least squares discriminant analysis; P_{4444} , Tetrabutylphosphonium; P_{44416} , Tributylhexadecylphosphonium; RNase A, Ribonuclease A; SNV, Standard normal variate; UV–Vis, Ultraviolet–visible.

* Corresponding author.

** Corresponding author.

E-mail addresses: cssousa@ucp.pt (C. Sousa), lsaraiva@ff.up.pt (M.L.M.F.S. Saraiva).

<https://doi.org/10.1016/j.talanta.2021.123164>

Received 2 September 2021; Received in revised form 10 December 2021; Accepted 20 December 2021

Available online 23 December 2021

0039-9140/© 2021 Elsevier B.V. All rights reserved.

compared to analysis of individual proteins [5,6]. Moreover, a single biomarker can be associated with numerous diseases and various proteins can be indicative of the same medical condition [7,8].

Enzyme-linked immunosorbent assay (ELISA) is the most commonly used method for protein detection in biomedical research and clinical diagnosis owing to its high sensitivity and specificity [9,10]. However, the high cost, laborious and time-consuming processes, along with the need for specific antibodies have hindered its widespread application [11,12]. Mass spectrometry-based approaches offer several advantages over conventional immunoassay formats, including identification of proteoforms, as well as detection and quantification of protein biomarkers in a multiplexed manner. Nevertheless, this method requires expensive instrumentation and well-trained personnel [13,14]. Due to cost-effectiveness, simplicity, and the ability to obtain rapid responses, optical sensor arrays provide a reliable alternative to detect and discriminate structurally related proteins as well as complex protein mixtures [15,16]. In contrast to a “lock-and-key” approach, this sensing system comprises multiple receptors that interact differently with each protein or protein mixture and generate unique response patterns [17,18]. The acquired data can then be processed using chemometric tools such as hierarchical cluster analysis (HCA), linear discriminant analysis (LDA), principal component analysis (PCA), and partial least squares discriminant analysis (PLSDA) [19,20]. Despite the great potential of a sensor array strategy, its large-scale application has been hampered by need for a large number of sensing elements (sometimes even higher than the number of proteins), which increases the time required for data acquisition, processing, and analyses [16,21]. Furthermore, sensing materials often demand laborious chemical modifications to achieve the desired performance [22]. GUMBOS (Group of Uniform Materials Based on Organic Salts) have recently emerged as alternative materials for detection and discrimination of proteins because of their facile synthesis and high tunability [23,24]. The physicochemical properties of these solid-state materials can be easily modulated by varying the cation-anion combination similar to ionic liquids (ILs) chemistry, thus avoiding complicated synthetic procedures [25,26].

Erythrosin B (EB) is an anionic xanthene dye that has been used for various purposes, such as a colorant in food products [27], a vital stain [28], and a flaviviruses inhibitor [29]. It has also been applied as a sensor for oxygen [30], heavy metals and halide ions [31], as well as proteins [32]. To date, there are only a few spectrophotometric methods for protein determination using EB and they present several operational disadvantages, such as high temperature, long reaction time, and necessary additives [32–34]. In this context, there is a growing interest in the development of viable alternative strategies to both detect and discriminate multiple proteins.

In this study, we describe the synthesis of four novel EB-based GUMBOS and evaluate their ability to discriminate single and binary mixtures of proteins. Ultraviolet–visible (UV–Vis) spectra resulting from the binding of GUMBOS to proteins were later analyzed using PLSDA. Percentages of correct protein assignments were equal to or greater than 86.5% and mixtures with different ratios of albumin (Alb) and Mb appeared as well-defined clusters in two-dimensional partial least squares discriminant analysis (2D-PLSDA) scores plots. These findings support potential use of EB-derived GUMBOS as alternative sensing materials for protein identification.

2. Material and methods

2.1. Reagents and solvents

[Na]₂[EB] (2',4',5',7'-tetraiodofluorescein disodium salt), tetrabutylphosphonium bromide ([P₄₄₄₄][Br]), tetrabutylammonium bromide ([N₄₄₄₄][Br]), sodium phosphate dibasic (Na₂HPO₄), sodium phosphate monobasic (NaH₂PO₄), Alb (from human serum), α-chymotrysin (α-ChT, from bovine pancreas), lysozyme (Lyz, from chicken egg white), Mb (from equine heart), and ribonuclease A (RNAse A, from bovine

pancreas) were all purchased from Sigma-Aldrich (St. Louis, MO) and used as received. Tributylhexadecylphosphonium bromide ([P₄₄₄₁₆][Br]) was obtained from Tokyo Chemical Industries, Co. (TCI–America, Portland, OR) and benzyldimethylhexadecylammonium chloride ([BDHA][Cl]) from Alfa Aesar (Haverhill, MA). Citric acid, sodium hydroxide, and 1-octanol were purchased from Honeywell (Charlotte, NC). Dimethylsulfoxide (DMSO) was acquired from Merck (Darmstadt, Germany) and deuterated chloroform (CDCl₃, 99.8%) with 0.03% tetramethylsilane was obtained from Beantown Chemical (BTC, Hudson, NH).

A series of citrate buffers (0.1 mol L^{−1}) were prepared from citric acid and adjusted to pH 3.0, 4.5, or 6.0 by adding sodium hydroxide. Phosphate buffer (0.01 mol L^{−1}, pH 7.4) was used for the preparation of protein stock solutions (1 mg mL^{−1}) and obtained by mixing Na₂HPO₄ and NaH₂PO₄. All buffer solutions were prepared using high purity water (Milli-Q®, 18.2 MΩ cm). EB-based GUMBOS were dissolved in DMSO and protected from light degradation.

2.2. Instrumentation

Proton (¹H, 400 MHz) and carbon (¹³C, 125 MHz) nuclear magnetic resonance (NMR) spectra were recorded using a Bruker Avance III NanoBay 400 and a Bruker AV-500. NMR samples were prepared by dissolving GUMBOS in CDCl₃. Coupling constants (*J*) were reported in Hz and chemical shifts (*δ*) in ppm. Abbreviations used to explain multiplicities were as follows: singlet (s), doublet (d), triplet (t), and multiplet (m).

An Agilent 6230 B-TOF LC/MS was used for high-resolution mass spectrometry (HRMS) measurements at the Louisiana State University Mass Spectrometry facility. Differential scanning calorimetry (DSC) thermograms were acquired using a TA DSC Q200 instrument. Approximately 2–5 mg of GUMBOS were crimped in an aluminum pan and heated from 25 to 150 °C at a rate of 10 °C min^{−1} under a nitrogen purge (50 mL min^{−1}). An empty pan was used as a reference during DSC measurements.

Fourier transform infrared (FTIR) measurements were performed on a Bruker Tensor 27 spectrometer, equipped with a Pike Miracle ATR cell. Spectra were collected from 4000 to 650 cm^{−1} by averaging 32 scans at a resolution of 4 cm^{−1}. Opus 8.0 software was used for data collection and initial processing.

Absorption studies were performed, at room temperature, on a Jasco V-660 spectrophotometer, equipped with a 1 cm path length quartz cuvette. Spectra were collected in the range of 250–800 nm using a slit width of 5 nm. The whole spectral data were subjected to PLSDA in order to discriminate aqueous protein samples.

2.3. Synthesis and structural characterization of GUMBOS

EB-derived GUMBOS were synthesized using a simple metathesis reaction [23]. Sodium cations of EB dye were exchanged with several bulky organic cations (P₄₄₄₄⁺, P₄₄₄₁₆⁺, N₄₄₄₄⁺, and BDHA⁺). Briefly, both cation and anion (2:1.1 molar ratio) were dissolved in phosphate buffer pH 7.4 and stirred for 24 h in the dark at room temperature (Fig. 1). GUMBOS precipitated after the allotted reaction time, were washed several times with water to eliminate sodium halide by-products, and freeze-dried. Ultimately, GUMBOS formation was confirmed using NMR and FTIR spectroscopy as well as HRMS. Structural characterization data can be found in Supplementary Material (Figs. S1–S9 and Table S1).

2.4. Octanol/water partitioning

Following synthesis of EB-based GUMBOS, octanol/water partition coefficient (*K*_{O/W}) values were determined using a modified shake-flask method [35]. Before experiments, 1-octanol and water were mutually saturated by stirring for 24 h, and then phases were separated. Each

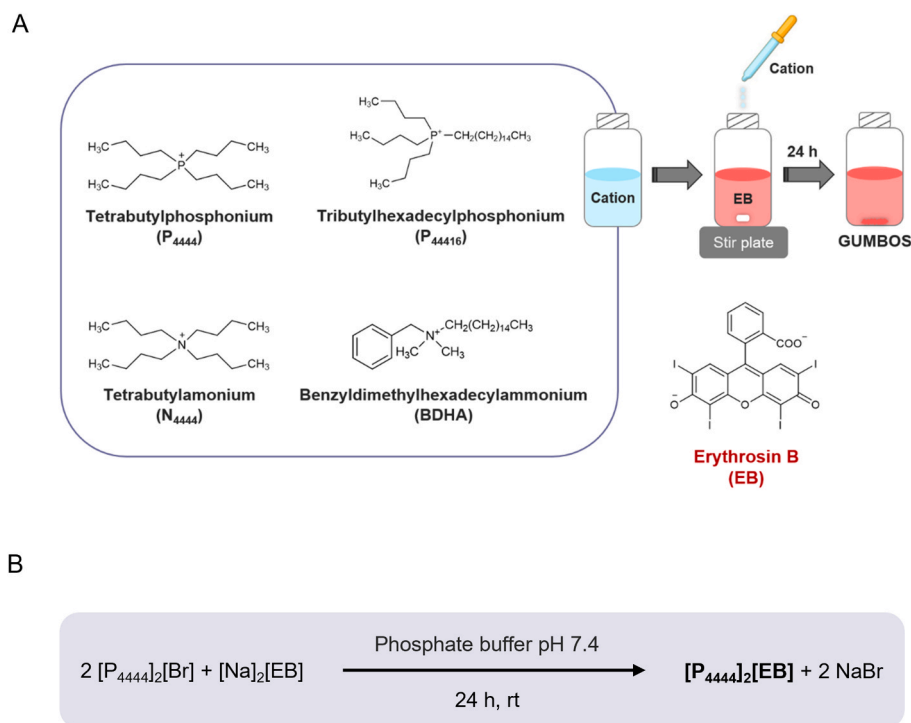


Fig. 1. (A) Chemical structures of the cations used in the synthesis of EB-based GUMBOS. (B) Synthetic diagram of $[P_{4444}]_2[EB]$ GUMBOS.

compound was dissolved in 5 mL of 1-octanol saturated with water (C_i), to which an equal volume of water-saturated with octanol was added. After stirring the mixture for 24 h at room temperature, the absorbance of upper organic layer was measured at 540 nm. GUMBOS concentration (C_o) was calculated from the respective calibration curve, which was previously established using 1-octanol standard solutions (range of 2–8 $\mu\text{mol L}^{-1}$). The aqueous concentration (C_w) was obtained by subtracting GUMBOS concentration in the octanol layer (C_o) from its initial concentration (C_i). Finally, $K_{O/W}$ was calculated using the equation:

$$K_{O/W} = C_o/C_w \quad (1)$$

Three independent experiments were performed for each compound.

2.5. Protein discrimination using the synthesized GUMBOS

Discrimination capabilities of EB-based GUMBOS ($[P_{4444}]_2[EB]$, $[P_{44416}]_2[EB]$, $[N_{4444}]_2[EB]$, and $[BDHA]_2[EB]$) were tested using five proteins, e.g., RNase A, Lyz, Alb, Mb, and α -ChT. The assays were performed at three different pH values (3.0, 4.5, and 6.0), and absorption spectra were acquired after 5, 10, and 15 min. GUMBOS and protein concentrations were held constant at 1 $\mu\text{mol L}^{-1}$ and 20 $\mu\text{g mL}^{-1}$, respectively.

The ability of $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$ to discriminate between mixtures of Alb and Mb at different proportions (20% Alb + 80% Mb, 40% Alb + 60% Mb, 60% Alb + 40% Mb, and 80% Alb + 20% Mb) was also examined. Studies were conducted only with GUMBOS mentioned above at pH 6.0 and after a reaction period of 5 min since these were the conditions that led to best discrimination results. Total protein concentration was maintained at 4 $\mu\text{g mL}^{-1}$. For all assays, seven replicate samples were obtained from independent experiments on two different days.

2.6. Data analysis

UV–Vis spectra were preprocessed using standard normal variate (SNV) [36], and then mean-centered before PLSDA application. PLSDA

is a supervised model based on the PLS2 algorithm, which requires previous knowledge of the data set [37,38]. In PLSDA, to each known sample (x_i , GUMBOS spectra in the presence of a single protein) is assigned a vector of zeros with value one at the position corresponding to its class (y_i , in this study to each protein). The structure of the PLSDA model is described by Eqs. (2) and (3). Model loadings (P and Q) and corresponding scores (T and U) are obtained by sequentially extracting the components or latent variables (LVs) from matrices X (spectra) and Y (matrix codifying the proteins).

$$X = TP^t + E \quad (2)$$

$$Y = UQ^t + F \quad (3)$$

The algorithm correlates the scores of each block (T and U), yielding an internal regression matrix. This internal regression can be transformed on a regression matrix B. In this case, the regression matrix is composed of three vectors: one regression vector corresponding to each protein. E and F are the residual matrices and depend on the number of LVs selected. Predictions for new samples are obtained by multiplying a new spectrum (x_{new}) by the regression matrix B:

$$y_{\text{new}} = x_{\text{new}} B \quad (4)$$

The prediction ($y_{\text{new}} = [y_{\text{new},1}, y_{\text{new},2}, \dots, y_{\text{new},n}]$) is then converted into a class assignment (protein) from which confusion matrices are obtained. During the development of PLSDA model, 70% of the samples are randomly selected to obtain the regression matrix B and 30% are used as new samples for prediction. This procedure is repeated 100 times per PLSDA model and confusion matrices present the mean values of correct protein assignments for the prediction samples.

Chemometric data analysis was performed in MATLAB version 9.5 Release 2018b (MathWorks, Natick, MA) and PLS Toolbox version 8.7 (2019) for MATLAB (Eigenvector Research, Manson, WA).

2.7. Rationale of the study

Four EB-based materials ($[P_{4444}]_2[EB]$, $[P_{44416}]_2[EB]$, $[N_{4444}]_2[EB]$, and $[BDHA]_2[EB]$) were synthesized for the first time and structurally

characterized using several traditional techniques as previously described in section 2.3. In order to confirm if they could be categorized as GUMBOS, their melting points were determined by DSC. Relative hydrophobicities were measured as $K_{O/W}$ to better understand GUMBOS-protein binding.

Absorption spectra were used to evaluate the potential of synthesized materials to discriminate among five proteins with different features at three pH values (3.0, 4.5, and 6.0) and reaction times (5, 10, and 15 min). Acidic pH conditions were studied as they are known to promote the formation of complexes between EB dye and proteins [32–34]. For each EB-derived GUMBOS and experimental condition, a single PLSDA model was developed, in which a total of 36 PLSDA models: 4 GUMBOS $\times$ 3 pH values $\times$ 3 reaction times were combined, and the corresponding confusion matrix was generated (further details in section 2.6). The highest percentages of correct predictions obtained from confusion matrices were observed at pH 6.0 and 5 min of reaction for $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$. These two GUMBOS were subsequently used to discriminate between four binary mixtures of Alb and Mb under the conditions previously established (pH 6.0, 5 min).

3. Results

3.1. Structural characterization of EB-based GUMBOS

Structures of the resulting GUMBOS were determined by NMR (1H and ^{13}C), FTIR, and HRMS spectra analysis. A 2:1 stoichiometric ratio of cation to anion was confirmed through integration of proton NMR signals (Figs. S1, S3, S5, and S7). It is worth noting that FTIR peaks attributable to starting materials were also observed in the spectra of final products (Fig. S9). Analysis of HRMS results further evidenced successful synthesis of EB-based GUMBOS. In the positive ion mode, peaks with m/z values of 259.2557, 427.4432, 242.2841, and 360.3630 were assigned to P_{4444} , P_{44416} , N_{4444} , and BDHA cations. The negative ion mode HRMS spectra showed presence of EB anion in all synthesized compounds. Moreover, experimental results were in good agreement with expected m/z values (Table S1).

3.2. UV-vis properties of EB-derived GUMBOS

Absorption spectra of synthesized compounds were measured at a concentration of $1 \mu\text{mol L}^{-1}$ in DMSO. The UV-Vis profile of all EB-based GUMBOS is similar to the parent dye (Fig. 2) and shows an intense band at 540 nm, a shoulder at approximately 500 nm, corresponding to $S_0 \rightarrow S_1$ transitions, as well as a lower intensity band near 330 nm associated with transitions $S_0 \rightarrow S_2$ [39,40]. Differences in absorbance values of EB-based GUMBOS are most likely the result of cation variations.

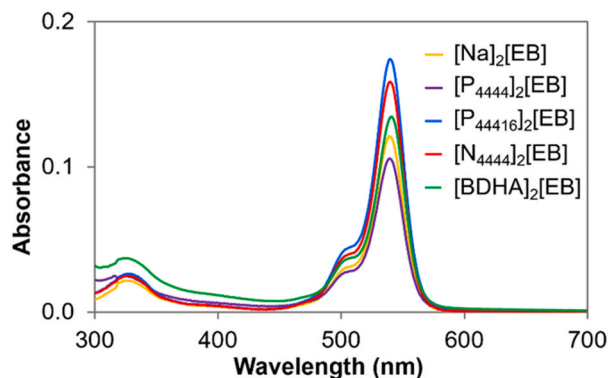


Fig. 2. UV-Vis spectra of $[Na]_2[EB]$ and EB-based GUMBOS ($1 \mu\text{mol L}^{-1}$) in DMSO.

3.3. Melting point and hydrophobicity of the synthesized materials

Similar to ILs, the physicochemical properties of GUMBOS can be fine-tuned through use of different ion exchange. In order to demonstrate this effect, the melting point and hydrophobicity of synthesized GUMBOS were determined. DSC thermograms (Fig. S10) show a broad endothermic peak for all compounds, which can be associated with either dehydration of adsorbed water or the amorphous property of GUMBOS [41,42]. Variation in melting points of synthesized compounds (from approximately 55.1 to 81.4 °C) can be explained by differences in size and symmetry of cations. In fact, the reduced melting points of GUMBOS as compared to $[Na]_2[EB]$ (m.p. = 315 °C, Sigma-Aldrich) can be attributed to replacement of Na^+ by relatively large phosphonium and ammonium cations [43]. Furthermore, the higher melting points of $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$ in relation to the other GUMBOS studied can be justified by symmetry of these cations, which could lead to a more efficient crystal packing and stronger ionic interactions [44,45]. The melting point of $[P_{44416}]_2[EB]$ was not determined by DSC since it remained a sticky solid at room temperature.

The relative hydrophobicity of GUMBOS was assessed through determination of octanol/water partition coefficients. Table S2 lists $\log K_{O/W}$ values for all synthesized compounds. Data analysis shows that hydrophobicity follows the order of $[N_{4444}]_2[EB] < [P_{4444}]_2[EB] < [P_{44416}]_2[EB] < [BDHA]_2[EB]$; thus demonstrating the cation's influence on GUMBOS hydrophobicity [23]. Moreover, and in line with previous reports [46,47], the elongation of alkyl side chain from four (P_{4444}^+) to sixteen (P_{44416}^+) carbons leads to an increase in $\log K_{O/W}$.

3.4. Protein discrimination using the synthesized GUMBOS

In order to evaluate the discrimination potential of EB-based GUMBOS, five proteins with distinct molecular weights (13.7–66.5 kDa), isoelectric points (4.7–11.4), and abundance levels were chosen as target analytes (Table S3). For example, Alb constitutes about 50% of the total protein content in plasma, while cardiac Mb is released after myocardial infarction [48,49]. UV-Vis spectra of the four synthesized materials in the presence of each protein were then recorded at three pH values (3.0, 4.5, and 6.0) and time points (5, 10, and 15 min) (Fig. S11). Spectral data were subjected to PLSDA to find differences among protein samples. Thirty-six PLSDA models were developed (Figs. 3–6), and the corresponding confusion matrices were generated. The total percentages of correct protein identifications obtained with the four GUMBOS at three pH values and reaction times are summarized in Table 1.

3.4.1. $[P_{4444}]_2[EB]$

For $[P_{4444}]_2[EB]$, the scores plots corresponding to the first three LVs of PLSDA regression models obtained at pH 3.0 for 5, 10, and 15 min of reaction time are presented in Fig. 3 (panels IA–IC), which also globally encompass approximately 95% of spectral variability. The scores plots showed five clusters, each corresponding to a single protein. Alb, Mb, and RNase A clusters were perfectly separated, while some degree of overlap was found between α -ChT and Lyz clusters. It should be noted that the obtained results are quite similar for all time points. The first LV (LV1) encompasses around 60% of spectral variability being mostly responsible for discrimination between Alb and Mb. LV2 mainly accounts for discrimination between RNase A and Lyz + α -ChT (32–35% of spectral variability), whereas Lyz and α -ChT were mainly discriminated in the third LV. Table S4 presents confusion matrices obtained from the corresponding PLSDA regression models (at pH 3.0 for 5, 10, and 15 min of reaction). Global percentages of correct discrimination of proteins were calculated by sum of diagonals at nearly 90% for the three reaction times, which can be considered quite satisfactory. Alb and Mb were correctly predicted for the three allotted times and were determined to be in good agreement with the clusters observed in scores plots. With respect to RNase A, the great majority were correctly predicted

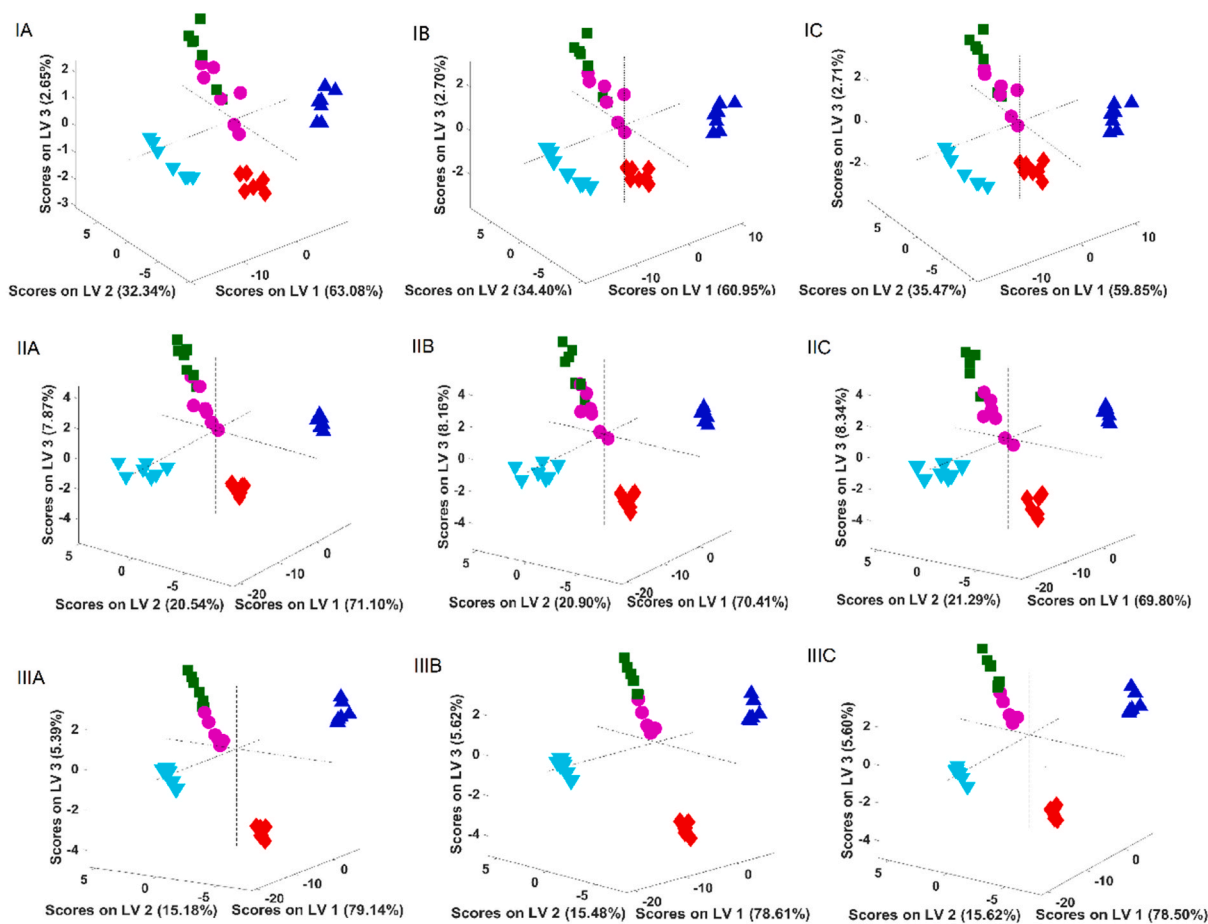


Fig. 3. Scores plots corresponding to the first three LVs of PLSDA regression models obtained using UV–Vis spectra at pH 3.0 (IA–IC), pH 4.5 (IIA–IIC), and pH 6.0 (IIIA–IIIC) for protein discrimination. $[P_{4444}]_2[EB]$ GUMBOS + RNase A (♦), Lyz (■), Alb (▲), Mb (▼), and α -ChT (●). Models A, B, and C correspond to 5, 10, and 15 min of reaction.

(95.5%–5 min, 96.0%–10 min, 98.5%–15 min); although a few were predicted as α -ChT (4.5%–5 min, 4.0%–10 min, 1.5%–15 min). Around 80% of Lyz spectra (79.5%–5 min, 80.5%–10 min, 77.0%–15 min) were correctly predicted, with the remaining 20% misidentified solely as α -ChT spectra. Concerning the latter, the percentage of correct identifications varied from 56.0% (10 min) to 65.5% (15 min). After 5 and 10 min of reaction, α -ChT spectra were misidentified only as Lyz spectra (35.5% and 44.0%, respectively), whereas upon 15 min 0.5% were misidentified as being RNase A and 34.0% as Lyz spectra.

Regarding pH 4.5 and 6.0, the scores plots (panels IIA–IIC and IIIA–IIIC from Fig. 3) revealed similar clusters to those observed at pH 3.0, with the first three LVs encompassing around 98% (LV1 approximately 70%, LV2 approximately 20%, LV3 approximately 8%) and 99% (LV1 approximately 79%, LV2 approximately 15%, LV3 approximately 5%) of spectral variability, respectively. Protein discrimination was also observed among the same LVs (LV1–Alb/Mb, LV2–RNase A/Lyz + α -ChT, LV3–Lyz/ α -ChT). These findings were consistently noticed for the three reaction times. Regarding the confusion matrices obtained from PLSDA models at pH 4.5 and 6.0 (Tables S5 and S6, respectively), a very satisfactory percentage of correct discrimination of proteins was also achieved (98.4%–5 min, 99.6%–10 min, 100.0%–15 min at pH 4.5 and 100.0%–5, 10, 15 min at pH 6.0). Similar to what was observed for pH 3.0, it appears that reaction time did not influence protein discrimination. On the contrary, pH value impacted the results. Correct protein discrimination was nearly 90% for pH 3.0, 99% for pH 4.5, and 100% for pH 6.0 (Table 1). Regarding misidentifications at pH 4.5, all were related to α -ChT spectra and were erroneously identified as Lyz spectra.

3.4.2. $[P_{44416}]_2[EB]$

For $[P_{44416}]_2[EB]$ at pH 3.0, the scores plots of PLSDA regression models (first three LVs encompassing about 99% of spectral variability) presented five individualized clusters, each containing a single protein, for all times studied (Fig. 4, panels IA–IC). In contrast to data observed for $[P_{4444}]_2[EB]$ GUMBOS, no overlapping clusters were visible to the naked eye. The clusters obtained for Alb, Mb, and RNase A were more scattered, which could lead to some misidentifications. LV1 mainly contributed to discrimination between Alb and Mb, while LV2 and LV3 seemed to be a cause for the remaining discriminations. As seen in Table 1, the percentages of correct protein discrimination range from 95.0% (15 min) to 97.8% (5 min). Similar to what was observed for $[P_{4444}]_2[EB]$, Alb and Mb were correctly predicted consistently (Table S7), whereas some RNase A was predicted as α -ChT (1.5%–5 min, 15.0%–10 min, 9.0%–15 min). It should also be noted that, when using the $[P_{44416}]_2[EB]$ GUMBOS, the percentage of RNase A incorrectly predicted after 10 and 15 min was higher than that for $[P_{4444}]_2[EB]$. This result was consistent with the widest scatter of clusters. Regarding Lyz spectra, they were all correctly predicted upon 10 min, being the wrong predictions related to α -ChT spectra (3.0%–5 min, 7.5%–15 min). In fact, α -ChT showed the worst predictions (93.5%–5 min, 92.5%–10 min, 91.5%–15 min), being misidentified as Lyz.

Concerning pH 4.5, the scores plots obtained for all time points (first three LVs encompassing around 98% of spectral variability) showed one well-individualized cluster belonging to Mb, two somehow contiguous clusters corresponding to Alb and RNase A as well as a fourth cluster containing Lyz and α -ChT proteins (Fig. 4, panels IIA–IIC). At pH 6.0, the scores plots (first three LVs accounting for 98% of spectral variability)

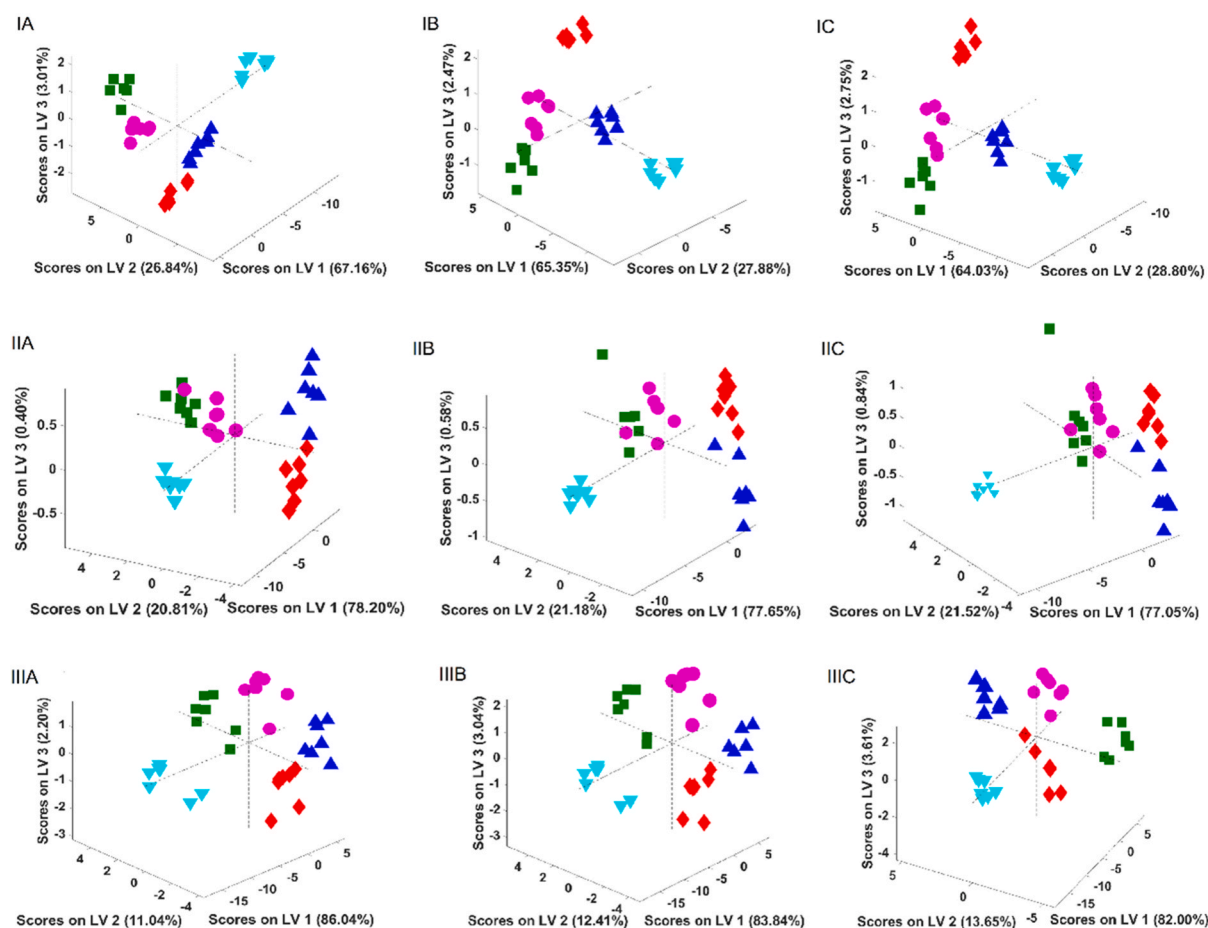


Fig. 4. Scores plots corresponding to the first three LVs of PLSDA regression models obtained using UV–Vis spectra at pH 3.0 (IA–IC), pH 4.5 (IIA–IIC), and pH 6.0 (IIIA–IIIC) for protein discrimination. $[N_{4416}]_2[EB]$ GUMBOS + RNase A (♦), Lyz (■), Alb (▲), Mb (▼), and α -ChT (●). Models (A) 5, (B) 10, and (C) 15 min of reaction.

were quite similar to those achieved at pH 3.0 (Fig. 4, panels IIIA–IIIC). In these cases, no visible overlapping was observed, but clusters remained scattered. Overall, the percentage of correct protein identifications varied from 91.9% (10 min) to 94.2% (5 min) at pH 4.5, and from 94.5% (5 min) to 97.7% (15 min) at pH 6.0 (Tables S8 and S9, respectively). The lower percentages observed at pH 4.5, as compared with the corresponding values at pH 3.0 and 6.0, were in agreement with the overlapping clusters observed in PLSDA scores plots (Fig. 4, panels IIA–IIC). When regarding each protein individually, with both pH values and the three reaction times, only Mb spectra had a consistently correct prediction of 100.0%. The great majority of Alb spectra were also correctly predicted (97.0%–5 min, 84.0%–10 min, 86.0%–15 min at pH 4.5, and 92.5%–5 min, 98.0%–10 min, 98.5%–15 min at pH 6.0), while the remaining were erroneously identified as RNase A or RNase A and α -ChT spectra. Additionally, RNase A was also correctly predicted in general for both pH values. It was, however, misidentified as Alb (5.5%–5 min), Lyz (1.5%–15 min), and Lyz + Alb (2.5% + 1.5%–10 min) at pH 4.5, or as Alb (1.0%–5 min and 0.5%–15 min) at pH 6.0. Lyz spectra were 100.0% correctly predicted at 5 and 10 min for pH 4.5 and were misidentified only as α -ChT spectra after 15 min of reaction (3.0% of wrong predictions). With respect to pH 6.0, Lyz was globally well predicted (91.5%–5 min, 94.5%–10 min, 98.0%–15 min), being misidentified only as RNase A at 5 and 10 min as well as RNase A (1.0%) and α -ChT (1.0%) at 15 min. Overall, α -ChT spectra were the worst predicted (87.5%–5 min, 76.5%–10 min, 85.0%–15 min at pH 4.5, and 89.5%–5 min, 88.5%–10 min, 92.5%–15 min at pH 6.0), being misidentified with several proteins depending on the pH and reaction period.

3.4.3. $[N_{4444}]_2[EB]$

The scores plots obtained from PLSDA models (three LVs encompassing around 97% of spectral variability for the three reaction times) presented in Fig. 5 (panels IA–IC) showed three well-individualized clusters. Two of them contain spectra of a single protein (Alb or Mb) and the third one spectra of RNase A, Lyz, and α -ChT. It should also be noted that spectra of these three proteins appeared in the scores plots contiguously, with minimal cluster overlapping. The worst discrimination was observed at 5 and 10 min with responses to Lyz and α -ChT analytes. LV1 seemed to be the principal cause for the discrimination of Alb and Mb, while other proteins fell into the third cluster and were partially discriminated in LV2 and LV3. No relevant differences were observed in the scores plots obtained for the three reaction periods, which was in accordance with the percentage of correct predictions acquired through the confusion matrices (around 90% for the times studied, Table 1). Alb, Mb, and RNase A were almost always correctly predicted (100.0%), with exception of RNase A at 15 min (0.5% misidentified as α -ChT). In contrast, Lyz and α -ChT spectra presented lower percentages of correct identifications and were misidentified as each other. The percentages of Lyz misidentifications were 17.0% at 5 min, 18.5% at 10 min, 21.5% at 15 min, while in relation to α -ChT were 30.0% at 5 min, 31.0% at 10 min, and 36.0% at 15 min (Table S10).

With respect to both pH 4.5 (Fig. 5, panels IIA–IIC) and pH 6.0 (Fig. 5, panels IIIA–IIIC), the scores maps showed more clearly defined clusters than those obtained at pH 3.0. For the $[N_{4444}]_2[EB]$ GUMBOS, time did not seem to be an influential factor for protein discrimination, which is in contrast to what was observed for pH value. Spectral

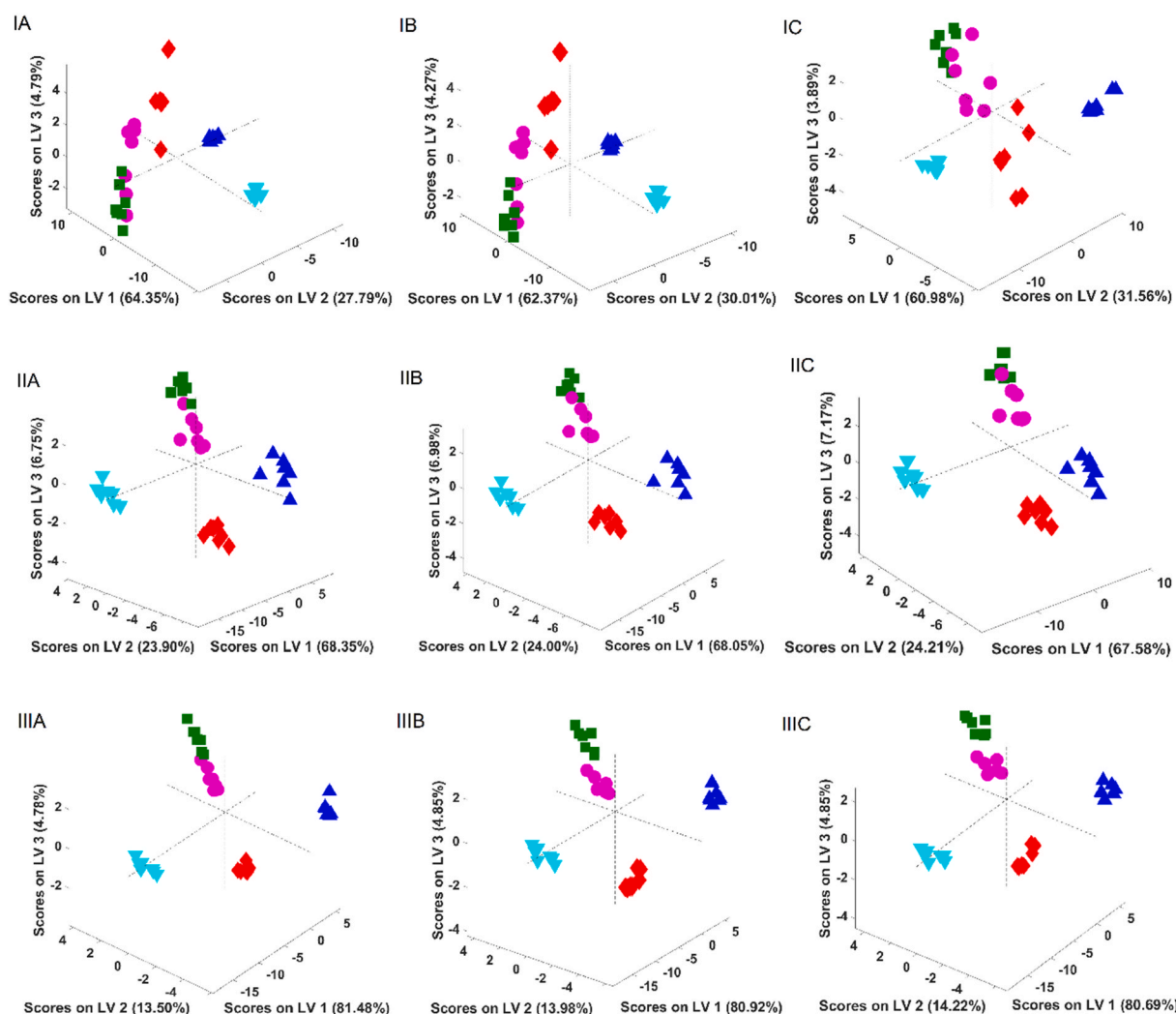


Fig. 5. Scores plots corresponding to the first three LVs of PLSDA regression models obtained using UV–Vis spectra at pH 3.0 (IA–IC), pH 4.5 (IIA–IIC), and pH 6.0 (IIIA–IIIC) for protein discrimination. $[N_{4444}]_2[EB]$ GUMBOS + RNase A (♦), Lyz (■), Alb (▲), Mb (▼), and α -ChT (●). Models A, B, and C correspond to 5, 10, and 15 min of reaction.

variability captured on the first three LVs was around 99% for all reaction times and both pH values. The percentages of correct discrimination of proteins were consistent with such findings, at approximately 93–95% for pH 4.5 and 100% for pH 6.0 (Table S11–pH 4.5 and S12–pH 6.0). At pH 4.5, Alb and Mb spectra were correctly identified as well as RNase A spectra at 15 min. After 5 and 10 min, however, RNase A was predicted as α -ChT (0.5%–5 min and 1.0%–10 min). Similar to what was observed for pH 3.0, Lyz and α -ChT were misidentified as each other (Lyz as α -ChT: 11.5%–5 min, 18.0%–10 min, 15.0%–15 min, and α -ChT as Lyz: 12.0%–5 min, 15.5%–10 min, 17.5%–15 min). At pH 6.0, the only misidentification was related to 0.5% of α -ChT spectra, which were predicted as Lyz upon 10 min of reaction.

3.4.4. $[BDHA]_2[EB]$

The scores plots obtained for protein discrimination using $[BDHA]_2[EB]$ were similar to those acquired for the $[P_{4444}]_2[EB]$ GUMBOS, but with a slightly lower overlap between Lyz and α -ChT spectra (Fig. 6, panels IA–IC). Five well-individualized clusters (each one corresponding to a single protein) were discriminated among LV1 (Alb from Mb from RNase A + Lyz + α -ChT), LV2 (RNase A from Lyz + α -ChT), and LV3 (Lyz from α -ChT). Spectral variability captured on the first three LVs was around 99% for all times studied. The total percentages of correct protein assignments (>99%) were in agreement with

the well-individualized clusters obtained for the three reaction periods (Table S13). Alb, Mb, RNase A, and Lyz spectra were 100% correctly predicted, with erroneous identifications related to a few α -ChT spectra identified as Lyz (2.5%–5 min, 1.5%–10 min, 2.0%–15 min) (Table S13).

Regarding pH 4.5 and 6.0, the scores maps presented similar clusters to those observed at pH 3.0, with 98% of spectral variability captured on the first three LVs for all reaction times (panels IIA–IIC and IIIA–IIIC from Fig. 6). The percentages of correct identifications of proteins were nearly 93 and 99% for pH 4.5 and 6.0, respectively (Tables S14 and S15). As noted previously, GUMBOS-protein reaction times do not seem to affect these percentages, but pH value does. At pH 4.5 (Table S14), only Mb spectra were correctly identified for the three times studied. Regarding Alb spectra, 97.5%–5 min, 91.0%–10 min, and 88.5%–15 min were correctly predicted, as these responses were misidentified as α -ChT (2.5%–5 min, 6.0%–10 min, 8.0%–15 min), RNase A (2.5%–10 min, 1.0%–15 min), and Lyz spectra (1.0%–10 min, 1.5%–15 min). RNase A spectra correct identifications varied from 86.5% (15 min) to 88.0% (10 min), as it was misidentified as α -ChT (11.0%–5 min, 11.5%–10 min, 12.5%–15 min), Mb (1.5%–5 min), Alb (0.5%–10 min), and Lyz spectra (1.0%–15 min). With respect to Lyz spectra, the incorrect identifications were solely predicted as α -ChT spectra (6.5%–5 min, 3.5%–10 min, 2.0%–15 min). Similarly, α -ChT spectra were misidentified as Lyz (9.0%–5 min, 9.5%–10 min, 7.5%–15 min) and as RNase A spectra

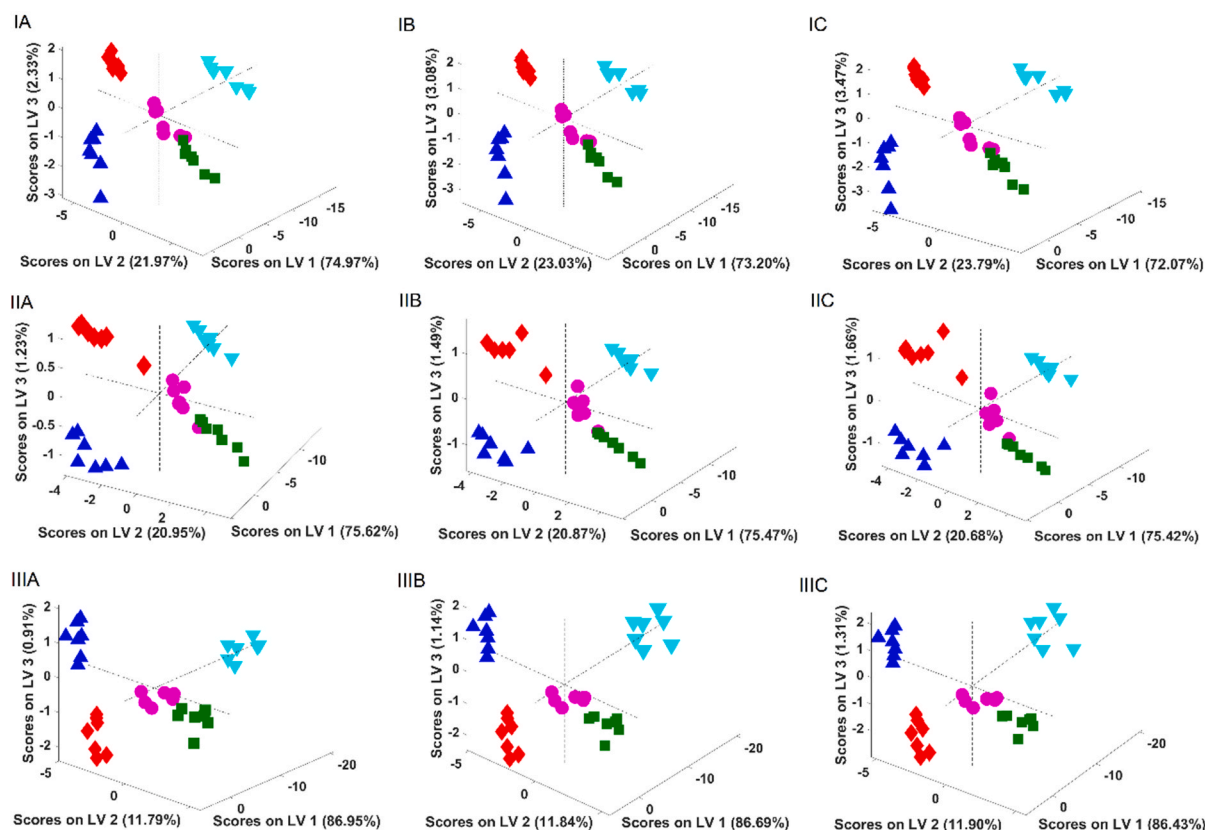


Fig. 6. Scores plots corresponding to the first three LVs of PLSDA regression models obtained using UV–Vis spectra at pH 3.0 (IA–IC), pH 4.5 (IIA–IIC), and pH 6.0 (IIIA–IIIC) for protein discrimination. [BDHA]₂[EB] GUMBOS + RNase A (♦), Lyz (■), Alb (▲), Mb (▼), and α-ChT (●). Models (A) 5, (B) 10, and (C) 15 min of reaction.

Table 1

Total percentages of correct predictions (%) obtained from the confusion matrices of PLSDA models (4 LVs), which were developed for protein discrimination using the synthesized EB-based GUMBOS.

pH	Time (min)	Total percentage of correct protein predictions			
		[P ₄₄₄₄] ₂ [EB]	[P ₄₄₄₁₆] ₂ [EB]	[N ₄₄₄₄] ₂ [EB]	[BDHA] ₂ [EB]
3.0	5	87.9	97.8	90.6	99.5
	10	86.5	95.5	90.1	99.7
	15	88.2	95.0	88.4	99.6
4.5	5	98.4	94.2	95.2	93.9
	10	99.6	91.9	93.1	93.2
	15	100.0	93.3	93.5	93.0
6.0	5	100.0	94.5	100.0	98.3
	10	100.0	96.2	99.9	98.7
	15	100.0	97.7	100.0	98.8

(0.5%) at 15 min. At pH 6.0, higher percentages of correct protein assignments were observed (Table S15), with Alb, Mb, and RNase A correctly predicted in a consistent fashion. In contrast, Lyz was relatively poorly predicted as RNase A (4.5%–5 min, 2.5%–10 min, 2.5%–15 min) and α-ChT (4.0%–5 min, 1.5%–10 min, 1.5%–15 min). α-ChT spectra were correctly predicted consistently at 5 min and poorly predicted solely as Lyz spectra (2.5%–10 min, 2.0%–15 min).

3.5. Discrimination of binary mixtures containing Alb and Mb

Among the synthesized GUMBOS, [P₄₄₄₄]₂[EB] and [N₄₄₄₄]₂[EB] were those that showed better ability to discriminate proteins at pH 6.0. Moreover, Alb and Mb were the proteins that consistently exhibited higher percentages of correct assignments. With this in mind, the

[P₄₄₄₄]₂[EB] and [N₄₄₄₄]₂[EB] GUMBOS were used to discriminate between four protein mixtures of Alb and Mb (20% Alb + 80% Mb, 40% Alb + 60% Mb, 60% Alb + 40% Mb, and 80% Alb + 20% Mb). Scores plots of the developed PLSDA models for [P₄₄₄₄]₂[EB] and [N₄₄₄₄]₂[EB] are provided in Fig. 7. In both cases, discrimination of protein mixtures clearly occurs in LV1, which encompassed almost 90% of spectral variability and where more well-defined clusters were found. GUMBOS spectra in the presence of a single protein (100% Alb or Mb) appeared on opposite sides of the scores map. It is interesting to note that on the positive side of the LV1 axis, where 100% Alb samples are located, all samples containing more than 50% of Alb were also observed. In contrast, spectra with less than 50% of Alb and more than 50% of Mb appeared on the negative side of LV1, which contains 100% Mb spectra. Separation of spectra along LV1 in two groups (>50% Alb and <50% Mb or < 50% Alb and >50% Mb) was clear evidence that both GUMBOS possess the ability for discrimination of protein mixtures.

4. Discussion

In this work, it was demonstrated that EB-based GUMBOS in combination with UV–Vis spectroscopy and PLSDA can successfully discriminate not only single proteins, but also binary mixtures of Alb and Mb. This study also investigated whether pH and reaction time affect GUMBOS discrimination performance.

Although numerous methods have been developed to discriminate proteins, most of them are laborious and time-consuming. Synthesis of sensing materials often involves multi-step reactions and purification procedures, which has limited their widespread use and has motivated the search for viable alternatives [50–52]. As a result of simplicity of their design and synthetic process, along with the possibility to tune their properties, GUMBOS appear to be promising candidates for

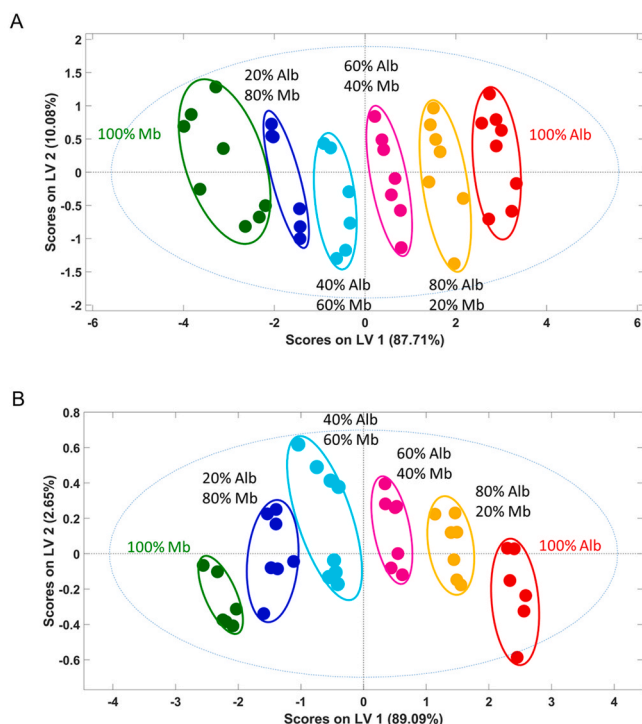


Fig. 7. Scores plots corresponding to the first two LVs of PLSDA regression models developed using the UV–Vis spectra of (A) $[P_{4444}]_2[EB]$ and (B) $[N_{4444}]_2[EB]$ after incubation with binary mixtures of Alb + Mb.

discrimination of proteins [23,24,26]. In this work, it was shown that melting point and hydrophobicity of EB-based GUMBOS as well as their ability to discriminate proteins can be modulated through simple cation exchange (Na^+ by P_{4444}^+ , P_{44416}^+ , N_{4444}^+ , and $BDHA^+$). A single by-product is formed during this reaction, which is easily removed simply by washing with water.

It is worth noting that, because of its relatively large, planar, and rigid structure, EB can strongly bind to hydrophobic protein surface residues. Although van der Waals and π -stacking interactions can also occur, the primary driving forces are hydrophobic interactions [53–55]. EB-based GUMBOS were found to display varying degrees of hydrophobicity that were attributed to cation exchange and have led to changes in protein binding affinity. The absorbance response patterns were then used to discriminate each protein by applying PLSDA models. Chemometric analyses revealed that $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$ show higher discriminatory power than $[P_{44416}]_2[EB]$ and $[BDHA]_2[EB]$ GUMBOS. Moreover, similarities in their ability to discriminate proteins can be explained using the chemical resemblance between phosphorus and nitrogen atoms and shorter alkyl side chain length, which reduces steric hindrance and improves GUMBOS accessibility to proteins [56].

Results demonstrated that the size, surface charge, and hydrophobic properties of proteins may also affect the extent of interaction with sensing materials, thus corroborating previous studies [23,50,57]. Overall, Alb and Mb were determined to have the highest percentages of correct predictions as seen from the confusion matrices of PLSDA models (Tables S4–S15). These findings can be explained by the fourfold difference in molecular weight between the two proteins (66.5 vs 16.7 kDa), opposite surface charges at pH 7.4, and great variation in their hydrophobicity [58]. In contrast, Lyz and α -ChT were the most misclassified and appeared as overlapping clusters in the scores plots. The same tendency was reported by Hewitt and Wilson [59].

The impact of pH and reaction time on the discriminatory power of EB-based GUMBOS was further examined. After 5, 10, and 15 min, the PLSDA scores plots did not show major differences in clustering patterns

(Figs. 3–6), which suggests a rapid binding between GUMBOS and proteins. Moreover, the percentages of correct predictions were similar for all time points (Table 1), and thus, a period of 5 min was selected. In contrast, Soedjak reported that binding of EB dye to proteins is slow at room temperature and requires about 2 h to be completed (24 times more than the corresponding GUMBOS) [32]. That study also showed that an acidic environment tends to promote formation of EB-protein complexes, which was later confirmed by other studies [33,34]. In an attempt to determine if the same procedures apply to EB-derived GUMBOS, experiments were conducted at pH 3.0, 4.5, and 6.0. As seen in Table 1, the percentages of correct assignments varied from 86.5 to 100.0% depending on the pH of citrate buffer and GUMBOS chemical structure. The best results were obtained at pH 6.0 and can be assigned to the dianionic form of EB in the GUMBOS ($pK_{COOH} = 2.35$ and $pK_{OH} = 3.79$) [60]. Overall, it was found that time does not affect the discriminatory power of synthesized GUMBOS, while pH does.

Discrimination of protein binary mixtures is far more challenging as the two proteins can interact with each other and compete for binding to the sensing material [50]. Both $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$ GUMBOS were able to discriminate between mixtures containing different proportions of Alb and Mb. 2D-PLSDA scores plots showed clearly defined and separate clusters of mixtures and allowed discrimination of pure and mixed proteins, similar to what has been reported in the literature for sensor arrays and other chemometric models [24,61,62]. However, in this work, an assembly of sensors was not required, with protein discrimination being achieved by use of single GUMBOS.

Finally, it is important to highlight the ability of chemometric tools to extract useful analytical information from complex datasets. Both unsupervised (PCA and HCA) and supervised (LDA) models have been used to analyze the response patterns generated from GUMBOS-based sensor arrays [23,24]. In this work, PLSDA models were developed to discriminate not only single but also binary mixtures of proteins by using the entire spectral wavelength range of data collected. The first LV seemed to be the main factor responsible for discrimination of Alb from Mb and then from RNase A + Lyz + α -ChT. Such results were obtained for all EB-based GUMBOS, independent of the reaction time and pH, and can be attributed to the nature of proteins as previously discussed.

5. Conclusions

In this study, four novel EB-based GUMBOS were synthesized and used to discriminate single and binary protein mixtures. The pH and reaction time were also investigated for their effect on discrimination performance of synthesized materials. It was found that changes in pH value (3.0–6.0) affected discrimination ability of EB-based GUMBOS, while the reaction period (5–15 min) did not seem to affect the results. The percentages of correct protein assignments ranged from 86.5 to 100.0%, being the best results achieved with $[P_{4444}]_2[EB]$ and $[N_{4444}]_2[EB]$ at pH 6.0 and 5 min of reaction time. Under these conditions, it was possible to discriminate between protein mixtures containing four different ratios of Alb and Mb since they appeared as well individualized clusters in the 2D-PLSDA scores plots. Although the underlying mechanism still necessitates further clarification, it is reasonable to infer that the results are derived from changes in the absorption spectra of GUMBOS due to protein binding. Thus, the obtained results demonstrate the potential of using these and other GUMBOS in combination with UV–Vis spectroscopy and chemometric tools to discriminate individual proteins and their mixtures. Moreover, this GUMBOS approach offers many advantages as compared to traditional analytical methods, including simplicity, rapidity, and cost-effectiveness. This alternative strategy is label-free, requires only 5 min of reaction time, and avoids the use of expensive instrumentation and complicated synthesis of probes. As part of future studies, it is planned to extend the application of this approach to more complex proteins and matrices (e. g., urine and blood).

Credit author statement

Ana M.O. Azevedo: Conceptualization, Investigation, Methodology, Writing – original draft and Editing. Clara Sousa: Formal analysis, Writing – original draft and Editing. Mi Chen: Investigation and Methodology. Caitlan E. Ayala and Rocío L. Pérez: Investigation and Writing-Reviewing. João L.M. Santos, Isiah M. Warner, and M. Lúcia M.F.S. Saraiva: Conceptualization, Supervision, and Writing-Reviewing.

Declaration of competing interest

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

Acknowledgments

This work received financial support from PT national funds (FCT/MCTES, Fundação para a Ciência e Tecnologia and Ministério da Ciência, Tecnologia e Ensino Superior) through the project UIDB/50006/2020. It was also supported by the European Union (FEDER funds through the Operational Competitiveness Program (COMPETE) POCI-01-0145-FEDER-030163 – Project Tailored NanoGUMBOS: The green key to wound infections chemsensing).

Ana M.O. Azevedo thanks FCT and FSE (Fundo Social Europeu) through Programa Operacional Regional do Norte for her PhD grant with reference number SFRH/BD/118566/2016. The use of Shared Instrumentation Facility at Louisiana State University is also acknowledged. Isiah M. Warner acknowledges support from the National Science Foundation under grant # CHE-1905105. The opinions, findings, and conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.talanta.2021.123164>.

References

- [1] Z. Jaffery, R. Nowak, N. Khoury, G. Tokarski, D.E. Lanfear, G. Jacobsen, J. McCord, Myoglobin and troponin I elevation predict 5-year mortality in patients with undifferentiated chest pain in the emergency department, *Am. Heart J.* 156 (5) (2008) 939–945, <https://doi.org/10.1016/j.ahj.2008.06.020>.
- [2] V. Kumar, J.R. Brent, M. Shorie, H. Kaur, G. Chadha, A.G. Thomas, E.A. Lewis, A. P. Rooney, L. Nguyen, X.L. Zhong, M.G. Burke, S.J. Haigh, A. Walton, P. D. McNaughter, A.A. Tedstone, N. Savjani, C.A. Muryn, P. O'Brien, A.K. Ganguli, D. J. Lewis, P. Sabherwal, Nanostructured aptamer-functionalized black phosphorus sensing platform for label-free detection of myoglobin, a cardiovascular disease biomarker, *ACS Appl. Mater. Interfaces* 8 (35) (2016) 22860–22868, <https://doi.org/10.1021/acsami.6b06488>.
- [3] M. Vesterhus, A. Holm, J.R. Hov, S. Nygard, E. Schruppf, E. Melum, L. W. Thorbjornsen, V. Paulsen, K. Lundin, I. Dale, O.H. Gilja, S.J.L.B. Zweers, M. Vatn, F.G. Schaap, P.L.M. Jansen, T. Ueland, H. Rosjo, B. Moum, C.Y. Ponsioen, K.M. Boberg, M. Farkkila, T.H. Karlsen, F. Lund-Johansen, Novel serum and bile protein markers predict primary sclerosing cholangitis disease severity and prognosis, *J. Hepatol.* 66 (6) (2017) 1214–1222, <https://doi.org/10.1016/j.jhep.2017.01.019>.
- [4] M. Vesterhus, T.H. Karlsen, Emerging therapies in primary sclerosing cholangitis: pathophysiological basis and clinical opportunities, *J. Gastroenterol.* 55 (6) (2020) 588–614, <https://doi.org/10.1007/s00535-020-01681-z>.
- [5] T. Xu, Y.X. Fang, A.A. Rong, J.H. Wang, Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy, *BMC Med. Res. Methodol.* 15 (2015) 7, <https://doi.org/10.1186/s12874-015-0085-z>.
- [6] M.J. Pena, H. Mischak, H.J.L. Heerspink, Proteomics for prediction of disease progression and response to therapy in diabetic kidney disease, *Diabetologia* 59 (9) (2016) 1819–1831, <https://doi.org/10.1007/s00125-016-4001-9>.
- [7] K. Uryu, X.H. Chen, D. Martinez, K.D. Browne, V.E. Johnson, D.J. Graham, V.M. Y. Lee, J.Q. Trojanowski, D.H. Smith, Multiple proteins implicated in neurodegenerative diseases accumulate in axons after brain trauma in humans, *Exp. Neurol.* 208 (2) (2007) 185–192, <https://doi.org/10.1016/j.expneurol.2007.06.018>.
- [8] S. Spindel, K.E. Sapsford, Evaluation of optical detection platforms for multiplexed detection of proteins and the need for point-of-care biosensors for clinical use, *Sensors* 14 (12) (2014) 22313–22341, <https://doi.org/10.3390/s141222313>.
- [9] J.F. Rusling, Multiplexed electrochemical protein detection and translation to personalized cancer diagnostics, *Anal. Chem.* 85 (11) (2013) 5304–5310, <https://doi.org/10.1021/ac401058v>.
- [10] A. Van Gool, F. Corrales, M. Colovic, D. Krstic, B. Oliver-Martos, E. Martinez-Caceres, I. Jakasa, G. Gajski, V. Brun, K. Kyriacou, I. Burzynska-Pedziwiatr, L. A. Wozniak, S. Nierkens, C.P. Garcia, J. Katrlík, Z. Bojic-Trbojevic, J. Vacek, A. Llorente, F. Antohé, V. Suica, G. Suarez, R. T'Kindt, P. Martin, D. Penque, I. L. Martins, E. Bodoki, B.C. Jacob, E. Aydinoglu, S. Timur, J. Allinson, C. Sutton, T. Luidar, S. Wittfooth, M. Sammar, Analytical techniques for multiplex analysis of protein biomarkers, *Expert Rev. Proteomics* 17 (4) (2020) 257–273, <https://doi.org/10.1080/14789450.2020.1763174>.
- [11] J. Stahl-Zeng, V. Lange, R. Ossola, K. Eckhardt, W. Krek, R. Aebersold, B. Dorn, High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites, *Mol. Cell. Proteomics* 6 (10) (2007) 1809–1817, <https://doi.org/10.1074/mcp.M700132-MCP200>.
- [12] J.V. Dzimianski, N. Lorig-Roach, S.M. O'Rourke, D.L. Alexander, J.M. Kimmey, R. M. DuBois, Rapid and sensitive detection of SARS-CoV-2 antibodies by biolayer interferometry, *Sci. Rep.* 10 (1) (2020) 12, <https://doi.org/10.1038/s41598-020-78895-x>.
- [13] D.P. Donnelly, C.M. Rawlins, C.J. DeHart, L. Fornelli, L.F. Schachner, Z.Q. Lin, J. L. Lippens, K.C. Aluri, R. Sarin, B.F. Chen, C. Lantz, W. Jung, K.R. Johnson, A. Koller, J.J. Wolff, I.D.G. Campuzano, J.R. Auclair, A.R. Ivanov, J.P. Whitelegge, L. Pasa-Tolic, J. Chamot-Rooke, P.O. Danis, L.M. Smith, Y.O. Tsybin, J.A. Loo, Y. Ge, N.L. Kelleher, J.N. Agar, Best practices and benchmarks for intact protein analysis for top-down mass spectrometry, *Nat. Methods* 16 (7) (2019) 587–594, <https://doi.org/10.1038/s41592-019-0457-0>.
- [14] N.P.M. Smit, L.R. Ruhaak, F.P.H.T.M. Romijn, M.M. Pieterse, Y.E.M. van der Burgt, C.M. Cobbaert, The time has come for quantitative protein mass spectrometry tests that target unmet clinical needs, *J. Am. Soc. Mass Spectrom.* 32 (3) (2021) 636–647, <https://doi.org/10.1021/jasms.0c00379>.
- [15] Z. Li, J.R. Askim, K.S. Suslick, The optoelectronic nose: colorimetric and fluorometric sensor arrays, *Chem. Rev.* 119 (1) (2019) 231–292, <https://doi.org/10.1021/acs.chemrev.8b00226>.
- [16] J.M. Fan, L. Qi, H.F. Han, L.P. Ding, Array-based discriminative optical biosensors for identifying multiple proteins in aqueous solution and biofluids, *Front. Chem.* 8 (2020) 18, <https://doi.org/10.3389/fchem.2020.572234>.
- [17] C.J. Hou, J.L. Dong, G.P. Zhang, Y. Lei, M. Yang, Y.C. Zhang, Z. Liu, S.Y. Zhang, D. Q. Huo, Colorimetric artificial tongue for protein identification, *Biosens. Bioelectron.* 26 (10) (2011) 3981–3986, <https://doi.org/10.1016/j.bios.2010.11.025>.
- [18] A. Bigdeli, F. Ghasemi, H. Golmohammadi, S. Abbasi-Moayed, M.A.F. Nejad, N. Fahimi-Kashani, S. Jafarinejad, M. Shahrabadian, M.R. Hormozi-Nezhad, Nanoparticle-based optical sensor arrays, *Nanoscale* 9 (43) (2017) 16546–16563, <https://doi.org/10.1039/c7nr03311g>.
- [19] A.P. Umali, E.V. Anslyn, A general approach to differential sensing using synthetic molecular receptors, *Curr. Opin. Chem. Biol.* 14 (6) (2010) 685–692, <https://doi.org/10.1016/j.cbpa.2010.07.022>.
- [20] J.E. Nielsen, K.S. Pedersen, K. Vestergaard, R.G. Maltesen, G. Christiansen, S. Lundbye-Christensen, T. Moos, S.R. Kristensen, S. Pedersen, Novel blood-derived extracellular vesicle-based biomarkers in alzheimer's disease identified by proximity extension assay, *Biomedicine* 8 (7) (2020) 21, <https://doi.org/10.3390/biomedicine8070199>.
- [21] M. De, S. Rana, H. Akpınar, O.R. Miranda, R.R. Arvizio, U.H.F. Bunz, V.M. Rotello, Sensing of proteins in human serum using conjugates of nanoparticles and green fluorescent protein, *Nat. Chem.* 1 (6) (2009) 461–465, <https://doi.org/10.1038/nchem.334>.
- [22] H.Y. Xi, X. Li, Q.Y. Liu, Z.B. Chen, Cationic polymer-based plasmonic sensor array that discriminates proteins, *Analyst* 143 (22) (2018) 5578–5582, <https://doi.org/10.1039/c8an01360h>.
- [23] W.I.S. Galpotheniya, F.R. Fronczek, M. Cong, N. Bhattarai, N. Siraj, I.M. Warner, Tunable GUMBOS-based sensor array for label-free detection and discrimination of proteins, *J. Mater. Chem. B* 4 (8) (2016) 1414–1422, <https://doi.org/10.1039/c5tb02038g>.
- [24] R.L. Perez, M. Cong, S.R. Vaughan, C.E. Ayala, W.I.S. Galpotheniya, J.K. Mathaga, I.M. Warner, Protein discrimination using a fluorescence-based sensor array of thiocarbocyanine-GUMBOS, *ACS Sens.* 5 (8) (2020) 2422–2429, <https://doi.org/10.1021/acssensors.0c00484>.
- [25] I.M. Warner, B. El-Zahab, N. Siraj, Perspectives on moving ionic liquid chemistry into the solid phase, *Anal. Chem.* 86 (15) (2014) 7184–7191, <https://doi.org/10.1021/ac501529m>.
- [26] A.M.O. Azevedo, J.L.M. Santos, I.M. Warner, M.L.M.F.S. Saraiva, GUMBOS and nanoGUMBOS in chemical and biological analysis: a review, *Anal. Chim. Acta* 1133 (2020) 180–198, <https://doi.org/10.1016/j.aca.2020.06.028>.
- [27] S. Lehto, M. Buchwitz, A. Klimm, R. Strassburger, C. Bechtold, F. Ulberth, Comparison of food colour regulations in the EU and the US: a review of current provisions, *Food Addit. Contam. Part a-Chemistry Analysis Control Exposure & Risk Assessment* 34 (3) (2017) 335–355, <https://doi.org/10.1080/19440049.2016.1274431>.
- [28] J.D. Franke, A.L. Braverman, A.M. Cunningham, E.E. Eberhard, G.A. Perry, Erythrosin B: a versatile colorimetric and fluorescent vital dye for bacteria, *Biotechniques* 68 (1) (2020) 7–13, <https://doi.org/10.2144/btn-2019-0066>.
- [29] Z. Li, S. Sakamuru, R.L. Huang, M. Brecher, C.A. Koetznner, J. Zhang, H.Y. Chen, C. F. Qin, Q.Y. Zhang, J. Zhou, L.D. Kramer, M.H. Xia, H.M. Li, Erythrosin B is a

- potent and broad-spectrum orthosteric inhibitor of the flavivirus NS2B-NS3 protease, *Antivir. Res.* 150 (2018) 217–225, <https://doi.org/10.1016/j.antiviral.2017.12.018>.
- [30] S.K. Lam, M.A. Chan, D. Lo, Characterization of phosphorescence oxygen sensor based on erythrosin B in sol-gel silica in wide pressure and temperature ranges, *Sensor. Actuator. B Chem.* 73 (2–3) (2001) 135–141, [https://doi.org/10.1016/S0925-4005\(00\)00695-X](https://doi.org/10.1016/S0925-4005(00)00695-X).
- [31] M.L. Wang, G.W. Meng, Q. Huang, Iodine-based fluorescent and colorimetric sensing for Ag^+ , Hg^{2+} , Fe^{3+} , and further for halide ions in aqueous solution, *RSC Adv.* 4 (16) (2014) 8055–8058, <https://doi.org/10.1039/c3ra47928e>.
- [32] H.S. Soedjak, Colorimetric micromethod for protein determination with erythrosin B, *Anal. Biochem.* 220 (1) (1994) 142–148, <https://doi.org/10.1006/abio.1994.1310>.
- [33] J. Ise, E. Kaneko, A new spectrophotometric method for determination of urinary protein using Erythrosin B, *Chem. Lett.* 35 (8) (2006) 922–923, <https://doi.org/10.1246/cl.2006.922>.
- [34] E. Kaneko, H. Yasuda, A. Higurashi, H. Yoshimura, Spot test of urinary protein using Erythrosin B and a membrane film, *Analyst* 135 (7) (2010) 1564–1568, <https://doi.org/10.1039/c0an00107d>.
- [35] OECD, Guideline No. 107, Partition Coefficient (N-octanol/water), Shake Flask Method, 1995.
- [36] T. Naes, T. Isaksson, T. Fearn, T. Davies, A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002.
- [37] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, *Anal. Chim. Acta* 185 (1986) 1–17, [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- [38] B.K. Alsberg, D.B. Kell, R. Goodacre, Variable selection in discriminant partial least-squares analysis, *Anal. Chem.* 70 (19) (1998) 4126–4133, <https://doi.org/10.1021/ac980506o>.
- [39] G.R. Fleming, A.W.E. Knight, J.M. Morris, R.J.S. Morrison, G.W. Robinson, Picosecond fluorescence studies of xanthene dyes, *J. Am. Chem. Soc.* 99 (13) (1977) 4306–4311, <https://doi.org/10.1021/ja00455a017>.
- [40] E. Montero, M.A. Garcia, M.A. Villegas, J. Llopis, Spectral pH dependence of erythrosin B in sol-gel silica coatings and buffered solutions, *Bol. Soc. Espanola Ceram. Vidr.* 47 (1) (2008) 1–6, <https://doi.org/10.3989/cyv.2008.v47.i1.205>.
- [41] M. Tobyn, J. Brown, A.B. Dennis, M. Fakes, Q. Gao, J. Gamble, Y.Z. Khimyak, G. McGeorge, C. Patel, W. Sinclair, P. Timmins, S. Yin, Amorphous drug-PVP dispersions: application of theoretical, thermal and spectroscopic analytical techniques to the study of a molecule with intermolecular bonds in both the crystalline and pure amorphous state, *J. Pharmaceut. Sci.* 98 (9) (2009) 3456–3468, <https://doi.org/10.1002/jps.21738>.
- [42] M. Chen, R.L. Perez, P. Du, N. Bhattarai, K.C. McDonough, S. Ravula, R. Kumar, J. M. Mathis, I.M. Warner, Tumor-targeting NIRF NanoGUMBOS with cyclodextrin-enhanced chemo/photothermal antitumor activities, *ACS Appl. Mater. Interfaces* 11 (31) (2019) 27548–27557, <https://doi.org/10.1021/acsami.9b08047>.
- [43] P. Wasserscheid, T. Welton, *Ionic Liquids in Synthesis*, second ed., Wiley-VCH, Weinheim, 2008.
- [44] I. Lopez-Martin, E. Burello, P.N. Davey, K.R. Seddon, G. Rothenberg, Anion and cation effects on imidazolium salt melting points: a descriptor modelling study, *ChemPhysChem* 8 (5) (2007) 690–695, <https://doi.org/10.1002/cphc.200600637>.
- [45] H. Zhang, M.T. Li, B.L. Yang, Design, synthesis, and analysis of thermophysical properties for imidazolium-based geminal dicationic ionic liquids, *J. Phys. Chem. C* 122 (5) (2018) 2467–2474, <https://doi.org/10.1021/acs.jpcc.7b09315>.
- [46] M.G. Freire, C.M.S.S. Neves, P.J. Carvalho, R.L. Gardas, A.M. Fernandes, I. M. Marrucho, L.M.N.B.F. Santos, J.A.P. Coutinho, Mutual Solubilities of water and hydrophobic ionic liquids, *J. Phys. Chem. B* 111 (45) (2007) 13082–13089, <https://doi.org/10.1021/jp076271e>.
- [47] E. Yamamoto, S. Yamaguchi, T. Nagamune, Protein refolding by N-alkylpyridinium and N-alkyl-N-methylpyrrolidinium ionic liquids, *Appl. Biochem. Biotechnol.* 164 (6) (2011) 957–967, <https://doi.org/10.1007/s12010-011-9187-1>.
- [48] N.L. Anderson, N.G. Anderson, The human plasma proteome - history, character, and diagnostic prospects, *Mol. Cell. Proteomics* 1 (11) (2002) 845–867, <https://doi.org/10.1074/mcp.R200007-MCP200>.
- [49] H.J. Issaq, Z. Xiao, T.D. Veenstra, Serum and plasma proteomics, *Chem. Rev.* 107 (8) (2007) 3601–3620, <https://doi.org/10.1021/cr068287r>.
- [50] H.C. Zhou, L. Baldini, J. Hong, A.J. Wilson, A.D. Hamilton, Pattern recognition of proteins based on an array of functionalized porphyrins, *J. Am. Chem. Soc.* 128 (7) (2006) 2421–2425, <https://doi.org/10.1021/ja056833c>.
- [51] Z. Fang, L.P. Liu, Y. Wang, D.M. Xi, S.S. Zhang, Unambiguous discrimination of multiple protein biomarkers by nanopore sensing with double-stranded DNA-based probes, *Anal. Chem.* 92 (2) (2020) 1730–1737, <https://doi.org/10.1021/acs.analchem.9b02965>.
- [52] W. Zhou, J.Z. Hou, Y.X. Li, H.P. Zhou, H. Huang, L. Zhang, M.A.H. Nawaz, C. Yu, Protein discrimination based on DNA induced perylene probe self-assembly, *Talanta* 224 (2021) 10, <https://doi.org/10.1016/j.talanta.2020.121897>.
- [53] K. Egbaria, M. Friedman, Adsorption of fluorescein dyes on albumin microspheres, *Pharmaceut. Res.* 9 (5) (1992) 629–635, <https://doi.org/10.1023/a:1015897909739>.
- [54] L. Ganesan, E. Margolles-Clark, Y. Song, P. Buchwald, The food colorant erythrosine is a promiscuous protein-protein interaction inhibitor, *Biochem. Pharmacol.* 81 (6) (2011) 810–818, <https://doi.org/10.1016/j.bcp.2010.12.020>.
- [55] L. Ganesan, P. Buchwald, The promiscuous protein binding ability of erythrosine B studied by metachromasy (metachromasia), *J. Mol. Recogn.* 26 (4) (2013) 181–189, <https://doi.org/10.1002/jmr.2263>.
- [56] S.R. Vaughan, N.C. Speller, P. Chhotaray, K.S. McCarter, N. Siraj, R.L. Perez, Y. Li, I.M. Warner, Class specific discrimination of volatile organic compounds using a quartz crystal microbalance based multisensor array, *Talanta* 188 (2018) 423–428, <https://doi.org/10.1016/j.talanta.2018.05.097>.
- [57] H.Y. Xi, W.W. He, Q.Y. Liu, Z.B. Chen, Protein discrimination using a colorimetric sensor array based on gold nanoparticle aggregation induced by cationic polymer, *ACS Sustain. Chem. Eng.* 6 (8) (2018) 10751–10757, <https://doi.org/10.1021/acssuschemeng.8b02063>.
- [58] S. Ranjan, W.K. Chung, M. Zhu, D. Robbins, S.M. Cranner, Implementation of an experimental and computational tool set to study protein-mAb interactions, *Biotechnol. Prog.* 35 (4) (2019) 12, <https://doi.org/10.1002/btpr.2825>.
- [59] S.H. Hewitt, A.J. Wilson, Protein sensing and discrimination using highly functionalised ruthenium(II) tris(bipyridyl) protein surface mimetics in an array format, *Chem. Commun.* 53 (91) (2017) 12278–12281, <https://doi.org/10.1039/c7cc06175g>.
- [60] V.R. Batistela, D.S. Pelloso, F.D. de Souza, W.F. da Costa, S.M.D. Santin, V.R. de Souza, W. Caetano, H.P.M. de Oliveira, I.S. Scarminio, N. Hioka, pK_a determinations of xanthene derivatives in aqueous solutions by multivariate analysis applied to UV-Vis spectrophotometric data, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 79 (5) (2011) 889–897, <https://doi.org/10.1016/j.saa.2011.03.027>.
- [61] X.C. Wei, Y.X. Wang, Y.X. Zhao, Z.B. Chen, Colorimetric sensor array for protein discrimination based on different DNA chain length-dependent gold nanoparticles aggregation, *Biosens. Bioelectron.* 97 (2017) 332–337, <https://doi.org/10.1016/j.bios.2017.06.020>.
- [62] P. Yan, X.Z. Li, Y.H. Dong, B.Y. Li, Y.Y. Wu, A pH-based sensor array for the detection and identification of proteins using CdSe/ZnS quantum dots as an indicator, *Analyst* 144 (9) (2019) 2891–2897, <https://doi.org/10.1039/c8an02285b>.