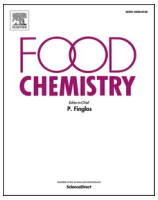
Food Chemistry 498 (2026) 147102

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem



Distinguishing plant and milk proteins and their interactions in hybrid cheese using confocal Raman microscopy with machine learning

Di Lu [a], Cushla McGoverin [d,e], Debashree Roy [a], Alejandra Acevedo-Fani [a], Harjinder Singh [a], Mark Waterland [b], Yi Zheng [c], Aiqian Ye [a,*]

[a] Riddet Institute, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand
[b] School of Food Technology and Natural Sciences, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand
[c] AgResearch Group, Bioeconomy Science Institute, Te Rourou, Massey University, Palmerston North 4442, New Zealand
[d] Department of Physics, University of Auckland, Auckland 1010, New Zealand
[e] Dodd-Walls Centre for Photonic and Quantum Technologies, Auckland 1010, New Zealand

ARTICLE INFO

Keywords:
Raman microscopy
Plant protein
Milk protein
Hybrid cheese
Machine learning
CatBoost

ABSTRACT

The increasing demand for plant-based alternatives to milk protein has led to the development of hybrid processed cheese analogues (HPCAs) combining plant proteins and casein. However, their complex microstructure and molecular interactions remain poorly understood. This study integrated confocal Raman spectroscopy with advanced machine learning for high-resolution spatial mapping and molecular characterization of HPCAs containing mung bean protein isolate (MPI) or hemp protein isolate (HPI) with casein. This integration helped distinguish between protein sources and elucidate structural changes. The addition of casein changed the HPI structure, promoting structural disorder, disulfide bond rearrangement, and a sharp decrease in the tyrosine doublet ratio from 4.5 in HPI100 to 1.2 in HPI50. Conversely, casein interaction with MPI led to microstructural segregation and changes of β-sheet content (from 53 % in MPI100 to 20 % in MPI30). This integrated method represents a powerful tool for analysing protein structure and interactions in complex food systems.

1. Introduction

The growing shift towards plant-based proteins as a sustainable alternative to animal proteins has attracted substantial interest in the food industry. Despite their potential, plant proteins often face challenges in matching the functional properties of animal proteins, such as solubility, foaming and emulsification (Zhang et al., 2024). These limitations hinder their ability to fully replace animal proteins in many food applications. To address these limitations, hybrid cheeses combining plant and milk proteins have been developed to explore new possibilities in food formulation. In our previous study, we compared various plant proteins in plant-based processed analogues, and among these, MPI and HPI showed the most distinct differences in structural and functional characteristics (Lu et al., 2025). Further, we combined MPI or HPI with rennet casein (RC) at varying ratios to produce hybrid processed cheese analogues (HPCAs). We found that the addition of plant proteins significantly altered the physical, textural and microstructural properties of the cheese analogues. Notably, the mechanisms underlying these changes differed between MPI/RC-based and HPI/RC-based cheeses, likely owing to differences in the size of protein aggregates and the presence of disulfide bonds, which influenced the structural and functional properties of the final product (Lu et al., 2026). However, the underlying molecular interactions between plant and milk proteins remain insufficiently understood. Therefore, building on our previous study, this study aimed to further investigate the structural and functional modifications in hybrid cheese systems through Raman spectroscopy.

Raman spectroscopy, an analytical technique based on the inelastic scattering of light, provides molecular insights into the composition and structure of materials. Although this technique has been used to analyse various food matrices, its application in cheese research has been relatively limited. Studies have investigated its use in detecting starch in spreadable cheese (Oliveira et al., 2016), monitoring protein and chemical changes during Cheddar cheese ageing (Dewantier et al., 2023; Holroyd et al., 2023), and predicting cheese quality traits (Stocco et al., 2024). Raman imaging, which integrates spectroscopy with spatial mapping, can be used to visualise the microstructure and distribution of cheese components (Chawanji et al., 2022; Holroyd & Nickless, 2024;

* Corresponding author.
E-mail address: a.m.ye@massey.ac.nz (A. Ye).

https://doi.org/10.1016/j.foodchem.2025.147102
Received 27 July 2025; Received in revised form 5 November 2025; Accepted 12 November 2025
Available online 14 November 2025

Smith, 2013). In addition, Nickless and Holroyd (2020) used Raman imaging and multivariate analysis to distinguish whey protein from casein. These studies highlight the advantages of multivariate techniques in differentiating components with similar Raman spectra, as computational algorithms can detect subtle spectral variations to distinguish microscopic domains.

However, the application of Raman imaging in complex cheeses, particularly those incorporating dairy and plant proteins, remains largely unexplored. No studies have specifically focused on identifying milk and plant proteins in hybrid cheese. The complexity of hybrid cheeses, along with their varying protein structures, fat contents and emulsifying agents, complicates the identification of individual components. A major obstacle is the spectral overlap between milk and plant proteins, which limits the effectiveness of conventional peak analysis. Moreover, the presence of emulsifying salts, water and fluorescence from plant proteins further complicates spectral interpretation.

This study aimed to investigate the interactions between casein and plant proteins in HPCAs through confocal Raman microscopy combined with machine learning (ML). We specifically focused on mapping the spatial distribution of components in both single-protein and dual-protein cheeses. In dual-protein cheeses, we distinguished casein from plant proteins, examined protein–protein and protein–lipid interactions, assessed changes in secondary protein structures in response to an increasing plant protein ratio and compared the effects of HPI and MPI when combined with casein. We hypothesised that 1) protein and lipid distributions in dual-protein cheeses might vary with the plant protein to casein ratio, influencing fat dispersion and protein–lipid interactions. 2) Casein and plant proteins may form large independent networks depending on the protein types. 3) Increasing the plant protein content in the hybrid cheese may alter secondary protein structures of both proteins, impacting cheese functionality. To verify our hypothesis, we used chemometric and ML techniques, including multivariate analysis and CatBoost models, to enhance spectral interpretation. This study highlights the potential of Raman spectroscopy in characterising the microstructure and protein interactions of hybrid cheeses, offering insights into the structural dynamics of mixed-protein systems.

2. Materials and methods

2.1. Materials

Rennet casein (RC; SureProtein™ 779) containing 82 % protein was obtained from Fonterra Co-operative Group Ltd., New Zealand. Mung bean protein isolate (85 %; MPI) was purchased from Bulk Powders New Zealand. Hemp protein isolate (70 %; HPI) and virgin coconut oil were purchased from Davis Trading Company, New Zealand. All chemicals used were of analytical grade and obtained from Sigma Chemical Co., unless otherwise specified.

2.2. Preparation of HPCAs

HPCAs were prepared by mixing MPI or HPI with RC at ratios of 100:0, 70:30, 50:50, 30:70 and 0:100 to achieve a total protein content of 18 % (wet weight). As reported by Lu et al. (2025), the formulation for HPCAs included 18 % protein, 20 % coconut oil, 0.5 % sodium chloride, 0.5 % calcium sulfate, 0.1 % potassium sorbate, 2.8 % trisodium citrate and water on a weight ratio basis, with pH adjusted to 5.4 using citric acid. HPCAs were prepared using a Rapid Visco Analyser (RVA-4; Newport Scientific Pty Ltd., Warriewood, NSW, Australia). HPCAs samples of 30 g were prepared in an aluminium cell fitted with a polycarbonate paddle. All dry ingredients and coconut oil were manually blended with water for 5 min and then hydrated for 1 h at room temperature in an aluminium cell. The mixture was processed using the following RVA profile: equilibration at 50 °C, followed by heating to 90 °C over 4 min, maintaining at 90 °C for 10 min. The sample was then cooled to 50 °C at a rate of 0.2 °C/s, mixing at 300 rpm from heating to cooling stages. The resulting samples were labelled MPI100, MPI70, MPI50, MPI30, HPI100, HPI50 and RC100, with the numbers indicating the plant protein ratios. Individually packed samples were stored at 4 °C and analysed using Raman spectroscopy within 1 week.

2.3. Instrumentation and parameters

A LabRAM HR Evolution Raman confocal microscope (HORIBA France SAS) was used to generate Raman images for HPCAs. This microscope is equipped with 532 and 785 nm excitation lasers and a thermoelectrically cooled charge-coupled device detector. Cheese samples were scanned using a 532 nm excitation wavelength with adjusted laser power to minimize fluorescence. For example, a laser power of 37.5 mW was used, corresponding to an estimated power density of approximately 190 mW/μm^2 for MPI–RC samples, while half of this power was applied to HPI–RC samples. A 100× oil immersion objective and a T1: 600 g mm^{-1} BLZ = 500 nm grating were used, giving a spectral resolution of 2–3 cm^{-1}.

The 532 nm excitation was selected to provide strong Raman scattering and high spatial resolution. Preliminary tests confirmed that the fluorescence background was manageable, and multiple spectra were collected from the same area to ensure that no significant spectral changes occurred during measurement.

2.4. Raman imaging

Cheese samples were stored at 4 °C and cut into thin layers (1–2 mm) using a razor blade was covered with a cover glass. Images were immediately acquired at 4 °C using a heating and cooling stage (THMS600 Linkam) with a 100× (1.30 NA) oil immersion lens and a 50 μm pinhole aperture. The images were acquired at 10–15 μm below the cut surface of the samples, and random regions in each sample were scanned. For low-resolution imaging, spectra were recorded in the range of 200–1800 cm^{-1} with a 0.5 s integration time and a 5 μm step with 41 and 21 increments in the x- and y-directions, respectively, to sample an area of 200 × 100 μm. For high-resolution imaging, spectra were recorded in the range of 200–1800 cm^{-1} with a 0.5 s integration time over a grid measuring 1.5 × 1.5 μm and 34 increments in both directions to sample an area of 50 × 50 μm. The same formulation was imaged with low-resolution mapping in three technical replicates and high-resolution mapping in two technical replicates. For each replicate, a single field of view was scanned, with an acquisition time of 0.5 s per point and a total imaging time of ~10 min.

2.5. Reference spectra of cheese ingredients

The reference Raman spectra of individual cheese ingredients, including hydrated RC, HPI, MPI, coconut oil, trisodium citrate (TSC), calcium sulfate and citric acid, were acquired using glass slides covered with aluminium foil. Spectra were collected with a 532 nm excitation laser (37.5 mW), a 50× objective lens, a 50 μm pinhole aperture and an acquisition time of 10 s. The presented reference spectra represent the average of at least three replicates ($n \geq 3$).

2.6. Raman imaging data processing using quasar

Raman imaging data were analysed using Quasar, an open-source Python library designed for ML and spectral data visualisation (Toplak et al., 2021). A representative workflow with a few examples of hierarchical clustering is shown in Fig. 1A. Raw data were pre-processed using rubber band correction (with the peak direction set to positive and the background action set to subtract), spike removal and Savitzky–Golay smoothing (window size, 13; polynomial order, 3) and normalised using a standard normal variate to enhance spectral quality and address fluorescence interference (Fig. 1A_a). Multiple analytical approaches were used to map the chemical composition of samples. To

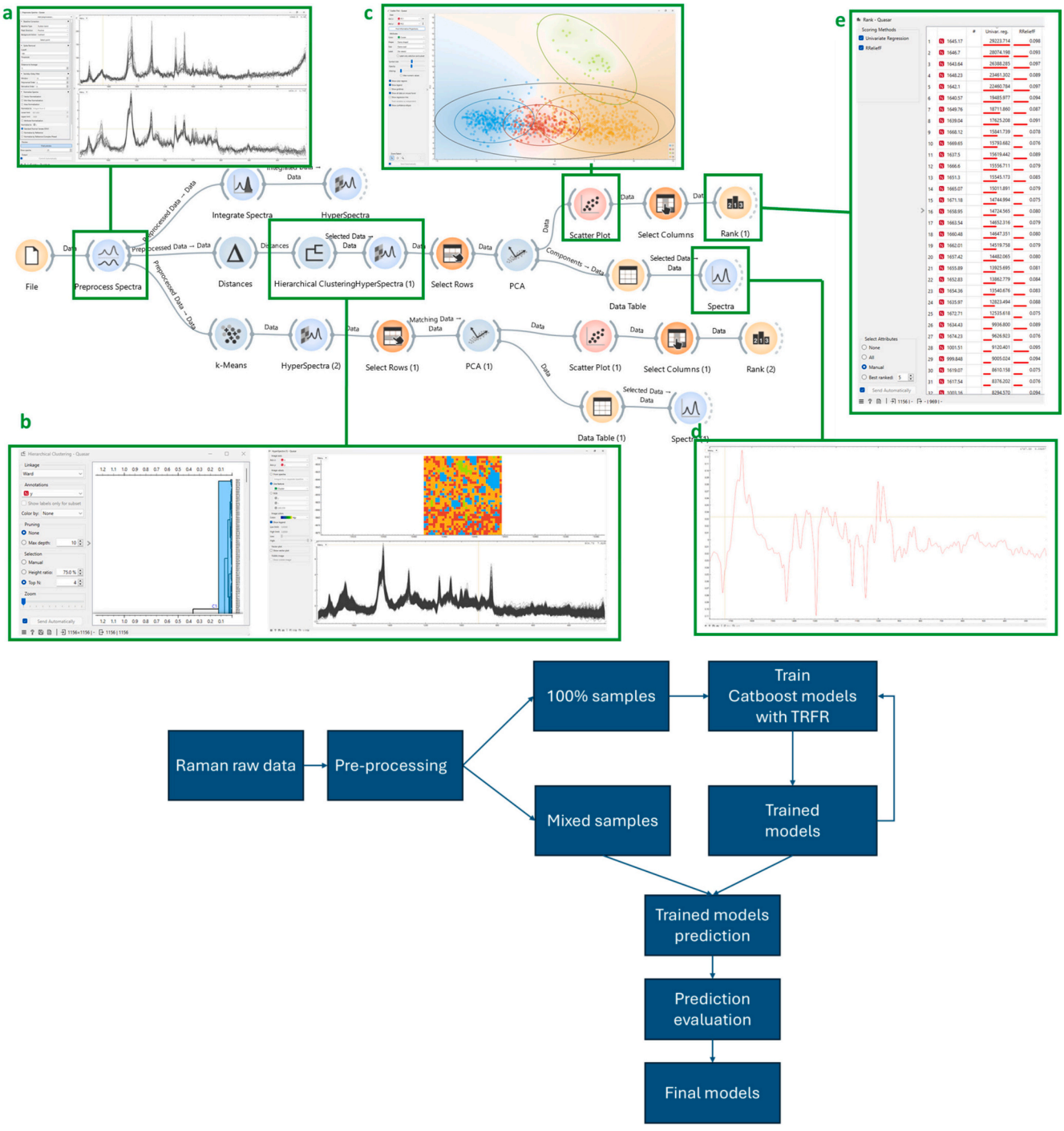


Fig. 1. Workflow for unsupervised analysis in Quasar (A). The 200–1800 cm^{-1} data were pre-processed with baseline correction, spike removal, Savitzky-Golay smoothing, and normalization (Box a). Hierarchical clustering grouped spectra into four clusters, with corresponding hyperspectral images (Box b). Principal Component Analysis (PCA) of PC1 and PC2 visualized cluster separation: PC1 separates the blue, red, and yellow clusters, while PC2 separates the green cluster (Box c). Loading plots from PC1 identified Raman shifts contributing to spectral variance (Box d), which were ranked by significance for feature selection (Box e). Workflow for supervised classification using CatBoost with Top-Ranked Feature Removal (TRFR) (B): pre-processed single-protein cheese samples trained the models, which were used to predict protein type in mixed cheese samples. Two predication evaluations were used to determine whether additional training cycles were required. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

visualise key chemical components, integrated peak mapping was performed using specific spectral regions: 1730–1750 cm^{-1} for lipids, 1600–1700 cm^{-1} for amide I, and 1200–1300 cm^{-1} for amide III, peak height of 1001 cm^{-1} for phenylalanine (Phe) and 834 cm^{-1} for TSC. In addition, two unsupervised clustering algorithms, hierarchical and k-means clustering, were used to classify spectral features and distinguish proteins in mixed systems. Hierarchical clustering using Ward's linkage generated a dendrogram (Fig. 1A_b), iteratively grouping similar data points, whereas k-means clustering divided data into a predefined number of clusters (K), iteratively adjusting cluster centroids until convergence. These approaches were applied to the spectral range of 200–1800 cm^{-1}, with 3–6 clusters being selected to efficiently differentiate between lipids, protein types and other cheese components. After initial clustering, the spectra in each cluster were carefully examined,

and the number of clusters was adjusted if necessary to ensure that lipids, protein types, and other cheese components were properly differentiated. Principal component analysis (PCA) was used to reduce dimensionality and visualise clusters on scatter plots (Fig. 1A_c). Loading plots were examined to identify Raman shifts contributing to spectral variance and interpret key spectral features differentiating the chemical compositions of samples (Fig. 1A_d). The first principal component (PC1) was used to capture protein-related variance. Subsequently, Raman shifts were ranked based on the absolute values of their PC1 loadings, highlighting the spectral regions most strongly associated with protein differentiation in the absence of class labels (Fig. 1A_e). (The high-resolution images of each component are shown in Supplementary materials).

Tyrosine doublet ratio (I_{850}/I_{830}) was analysed after being pre-processed in Quasar described above, including baseline correction and normalization. Spectral clusters corresponding to different proteins were identified using k-means clustering. The spectra within each cluster were then averaged, and the tyrosine doublet ratio was determined from the mean spectrum of each cluster. Results are presented as mean values (in triplicate) with standard deviations.

2.7. Raman imaging data processing using CatBoost and transfer learning

The data processed using Quasar were analysed using Python on the Google Colab platform. CatBoost ML models were trained using MPI100, HPI100 and RC100 datasets to classify protein types. Data related to fat content were removed based on their values at a Raman shift of 1744 cm^{-1}, with the threshold set at 60 % of the value range at 1744 cm^{-1}. The datasets were randomised and validated through 5-fold cross-validation before training with the whole dataset. Accuracy was used as the primary performance metric to evaluate the model. The CatBoost classifier was trained for 100 iterations with a tree depth of 4 and a learning rate of 0.1. The Logloss function was employed as the optimization objective. A fixed random seed (42) was applied to ensure model reproducibility. The trained models demonstrated the ability to distinguish MPI and HPI from RC.

To extend the applicability of the trained models, a transfer learning approach was implemented to enable the transfer of learned knowledge from models trained on single-protein cheese samples to mixed-protein cheese samples. This approach allowed the models to predict protein types in mixed-protein cheese samples, leveraging prior knowledge from pure protein systems. In addition, a feature selection method, top-ranked feature removal (TRFR), was integrated into the training process. TRFR iteratively removed the highest-ranked features identified by the models, compelling the models to learn from lower-ranked spectral features. This strategy aimed to enhance model robustness and improve generalisability in predicting protein composition in complex systems. The trained models were used to predict protein types in mixed-protein cheese samples, and the prediction results were used to evaluate the performance of the TRFR-trained models. The prediction evaluation was determined by two criteria: the similarity between the predicted and actual protein ratios and the similarity between the predicted plots and previous confocal laser scanning microscopy (CLSM) observation analysis (Lu et al., 2026).

CatBoost was chosen in this study for its efficiency and robustness in handling complex datasets. Unlike traditional gradient boosting algorithms, CatBoost employs ordered boosting, which reduces overfitting and target leakage. Moreover, its default parameters generally provide strong performance with little or no hyperparameter tuning. The workflow of the training process is shown in Fig. 1B.

2.8. Deconvolution of the amide I region

Deconvolution of the amide I region (1600–1700 cm^{-1}) was performed for quantitative analysis. The average spectrum from the protein cluster identified using k-means clustering was extracted using Quasar. The spectrum was subjected to baseline correction using a linear method between two selected points in the spectrum and smoothed using a 5-point Savitzky–Golay filter with a polynomial order of 2 to enhance the signal-to-noise ratio in Origin® (2021b). Peak fitting was performed using the second derivative method, followed by iterative Gaussian peak fitting. Peak assignment was performed using known Raman spectral data for secondary protein structures. Several iterations were performed until the fit converged, ensuring a final fitting curve with a corrected R^2 value of ≥ 0.99. The area under each peak was calculated relative to the total area and expressed as a percentage.

All results were expressed as the mean and standard deviation, as calculated from the three average spectra collected from three different images. Data were subjected to one-way analysis of variance using SPSS, followed by Tukey's comparison test. Differences were considered significant at a p-value of <0.05. The reported values for β-sheets and β-turns represent the cumulative contributions of multiple peaks.

3. Results and discussion

3.1. Overview of Raman spectra of components in HPCAs

This section briefly summarises the Raman spectra of protein and lipid components. Cheese exhibits several characteristic bands suitable for Raman imaging, particularly within the 200–1800 cm^{-1} region (Smith et al., 2017). Studies have identified Raman bands associated with proteins and lipids (Nickless & Holroyd, 2020; Smith et al., 2017; Wang et al., 2016) (Table 1).

In the 200–1800 cm^{-1} region, key protein-related bands provide insights into the identification and secondary structures of proteins. For instance, the phenylalanine ring breathing mode at 1001–1003 cm^{-1} serves as a marker for protein presence. The amide I band (1600–1700 cm^{-1}), primarily associated with C═O and N—H stretching vibrations, reflects protein secondary structures depending on hydrogen bonding and the conformation of polypeptides: α-helices at 1645–1657 cm^{-1}, random coil structures at 1650–1660 cm^{-1}, β-sheets at 1665–1680 cm^{-1} and β-turns at 1670–1699 cm^{-1}. The amide III band (1200–1300 cm^{-1}), which corresponds to C—N stretching and H—N—C bending vibrations, further distinguished protein secondary structures as follows: β-sheets at 1238–1245 cm^{-1}, random coils near 1250 cm^{-1} and α-helices at 1260–1300 cm^{-1}, which often overlapped with β-turns (Yada, Jackman & Smith, 1994; Wang et al., 2016; Zhao et al., 2004). These spectral features are essential for differentiating plant proteins from casein in complex food matrices. Herein, the reference Raman spectra of RC and HPI exhibited differences in respective amide I and amide III bands, with RC showing bands at 1663, 1257 and 1238 cm^{-1} and HPI showing bands at 1665, 1260 and 1240 cm^{-1} (Supplementary Fig. S1). These slight shifts suggest differences in the contents of secondary structures, particularly β-sheets and random coils, which may influence protein conformation and hydrogen interactions. In addition, the Raman spectrum of HPI showed bands at 528 and 542 cm^{-1}, indicating the presence of disulfide bonds. Similar findings have been reported in our previous study (Lu et al., 2026).

The reference Raman spectrum of coconut oil exhibited bands at 1746, 1655, 1440, 1298, 1125, 1082, 1062, 890 and 871 cm^{-1} (Supplementary Fig. S1). These bands correspond to specific molecular vibrations, including C═O stretching vibrations in triacylglycerol esters (1746 cm^{-1}); C═C stretching vibrations (1655 cm^{-1}); CH_2 scissoring associated (1440 cm^{-1}); CH_2 twisting attributed to phospholipids such as phosphatidylcholine (PC), phosphatidylinositol (PI) and phosphatidylserine (PS) (1298 cm^{-1}); C—C stretching vibrations in fatty acids, specifically saturated fatty acids (1082, 1062 and 1125 cm^{-1}); and phospholipid headgroup vibrations (890 and 871 cm^{-1}).

Table 1
Typical Raman shifts of Raman bands and general assignments in protein and lipid structure in the range of 200–1800 cm^{-1}.

Origin	Raman shifts (cm^{-1})	Assignment and Structural information
Cystine	496–513	S—S stretch
	514–525	*gauche–gauche–gauche* conformation,
	527–548	*gauche–gauche–trans* conformation
		trans–gauche–trans conformation
	630–670	*gauche* C—S of Met, Cys, Cys/2
	700–745	*trans* C—S of Met, Cys, Cys/2
Tyrosine	850/830	Fermi resonance between ring fundamental and overtone State of phenolic OH group (exposed or buried, hydrogen-bond donor or acceptor)
Tryptophan	760, 880, 1360	Indole ring Sharp intense band indicates buried residues; sensitive to environment polarity
Phenylalanine/tyrosine	622/644	Ring breathe I622/I644 indicates the ratio of Phe and Tyr
Phenylalanine	1003	Ring breathe Conformation insensitive; useful as an internal intensity standard
Aspartic and glutamic acids	1400–1430	C═O stretch of COO^- Ionized carboxyl groups
Aliphatic residues	1450, 1465	C–H bending Microenvironment, polarity
Tyrosine, tryptophan, phenylalanine residues	1610–1616	C═C stretch, aromatic ring vibrations
amide I	1650–1685	Amide *C*═O stretch, N—H wag
amide III	1235–1300	N—H in-plane bend, C—N stretch
Phospholipids	890, 871	phospholipid headgroup vibrations
	1298	CH_2 twisting attributed to phospholipids such as phosphatidylcholine (PC), phosphatidylinositol (PI), and phosphatidylserine (PS)
Fatty acids	1082, 1062	C—C stretching from fatty acids
	1125	C—C stretching from saturated fatty acids
	1440	CH_2 scissoring
Unsaturated fatty acids	1655	C═C stretching
Ester groups of lipids	1744	C═O carbonyl fats

Adapted from Nickless and Holroyd (2020), Yada, Jackman and Smith (1994), Smith et al. (2017), Wang et al. (2016), and Zhao et al. (2004).

3.2. Spatial distribution of components in both single- and dual-protein cheese systems

Initially, Raman images were generated using higher-spatial-resolution datasets, with a 1.5 μm step to cover an area of 50 × 50 μm. The spectral intensity at specific wavenumber ranges corresponding to key chemical components, such as fat (1730–1750 cm^{-1}), protein-related bands, including Phe ring breath (1001 cm^{-1}), the amide I (1600–1700 cm^{-1}), the amide III (1200–1300 cm^{-1}), and TSC (834 cm^{-1}). Integrated peak maps enabled the analysis of the spatial distribution and concentration of various chemical bonds, facilitating structural interpretation and interface characterization (Nickless & Holroyd, 2020). Fig. 2 presents representative Raman images of RC100, MPI50 and MPI100, demonstrating structural differences among these formulations. In RC100, fats were present as small, uniformly distributed droplets (approximately 1.5–5 μm), whereas proteins, observed from the Phe, amide I, and amide III peaks, formed a continuous and homogeneous network. All three spectral regions showed similar spatial distributions of protein, confirming consistent identification of protein-rich domains. Differences in colour intensity arise from variations in Raman scattering efficiency and vibrational origin. Thus, the observed intensity differences represent complementary contrast rather than inconsistency in protein mapping. After the incorporation of MPI into RC (MPI50), the fat droplet size increased to approximately 5–25 μm; however, in MPI100, fats coalesced into larger pools. Concurrently, plant protein formed aggregates, disrupting the continuous casein network in MPI50. These findings are consistent with those of CLSM reported in the previous study (Lu et al., 2026), demonstrating the structural modifications induced by plant protein addition. As the plant protein content increased, fat–protein interactions weakened, leading to greater microstructural segregation and larger fat droplet formation. Although casein in RC100 effectively stabilised fat droplets, the introduction of MPI disrupted the protein network, reducing the emulsification capacity of casein. This disruption is likely attributed to the higher molecular weight of plant proteins and the limited exposure of hydrophobic sites, which tend to remain buried within the compact molecular structures of plant proteins (Kim et al., 2020). The Raman images of other HPCAs formulated with different plant protein ratios (MPI30 and MPI70) and protein types (HPI50) are provided in Supplementary Fig. S2. These images confirmed that the selected feature peaks can effectively distinguish different components and reveal microstructural changes across formulations. In terms of protein type, the Phe-rich component in HPI exhibited smaller, more fragmented pixels, whereas MPI appeared as larger patches and aggregated regions (HPI50 vs MPI50).

Furthermore, Raman imaging enabled the identification and spatial mapping of TSC, which was detected within the casein matrix in RC100 but not within the MPI matrix in MPI100. This difference may occur because mung bean protein lacks the strong calcium-binding sites found in casein, which limits its interaction with TSC. Incorporating TSC resulted in a shift in its characteristic bands from 844 and 955 cm^{-1} (as observed in the reference spectrum of TSC, Supplementary Fig. S1) to 834 cm^{-1} and 969 cm^{-1}, respectively, in all cheese matrix. This shift is likely indicative of calcium citrate formation, as citrate molecules sequester Ca^{2+} ions to inhibit the formation of large casein-coated fat globules, thereby enhancing cheese stability through an emulsifying effect. The other spectral shifts observed in MPI–RC formulations, in which calcium sulfate was included, further support the likelihood of calcium citrate formation. These findings are consistent with those of previous studies, as pure calcium citrate exhibits characteristic bands at 836 and 970 cm^{-1}, which are similar to those detected in the hybrid cheese matrix (Campos et al., 2022; Smith et al., 2017). Other ingredients present at concentrations below 1 % in the cheese formulations were not detected on Raman images. This phenomenon may be attributed to subsampling, that is, these ingredients were absent in the analysed regions, or their concentrations were too low to be detected using the selected instrument parameters (Smith et al., 2017).

3.3. Distinguish plant proteins and caseins in dual-protein cheese systems

To distinguish plant proteins from casein in mixed-protein cheese formulations, the Raman spectra of MPI, HPI and RC were initially recorded in single-protein cheese systems. Briefly, the average spectra of MPI, HPI, RC and coconut oil were extracted from MPI100, HPI100 and RC100 samples, respectively, by selecting the lipid and amide I regions with high intensity (Fig. 3). As the Raman spectrum of coconut oil showed no significant differences among RC100, MPI100 and HPI100, the spectrum of coconut oil in RC100 was used as a representative. This spectrum provides insights into protein structures in a single-protein cheese system and differences between these structures. Although changes in spectral regions corresponding to proteins and fats were observed before and after processing (compared with their reference spectra), discussing the effects of processing on structural changes in proteins and fats is beyond the scope of this study.

The Raman spectrum of RC100 (Fig. 3A, green) exhibited lipid-related peaks at 1738, 1440, 1297, 1123 and 1060 cm^{-1} and protein-related peaks at 1648, 1614, 1251, 1001, 850 and 830 cm^{-1}. These

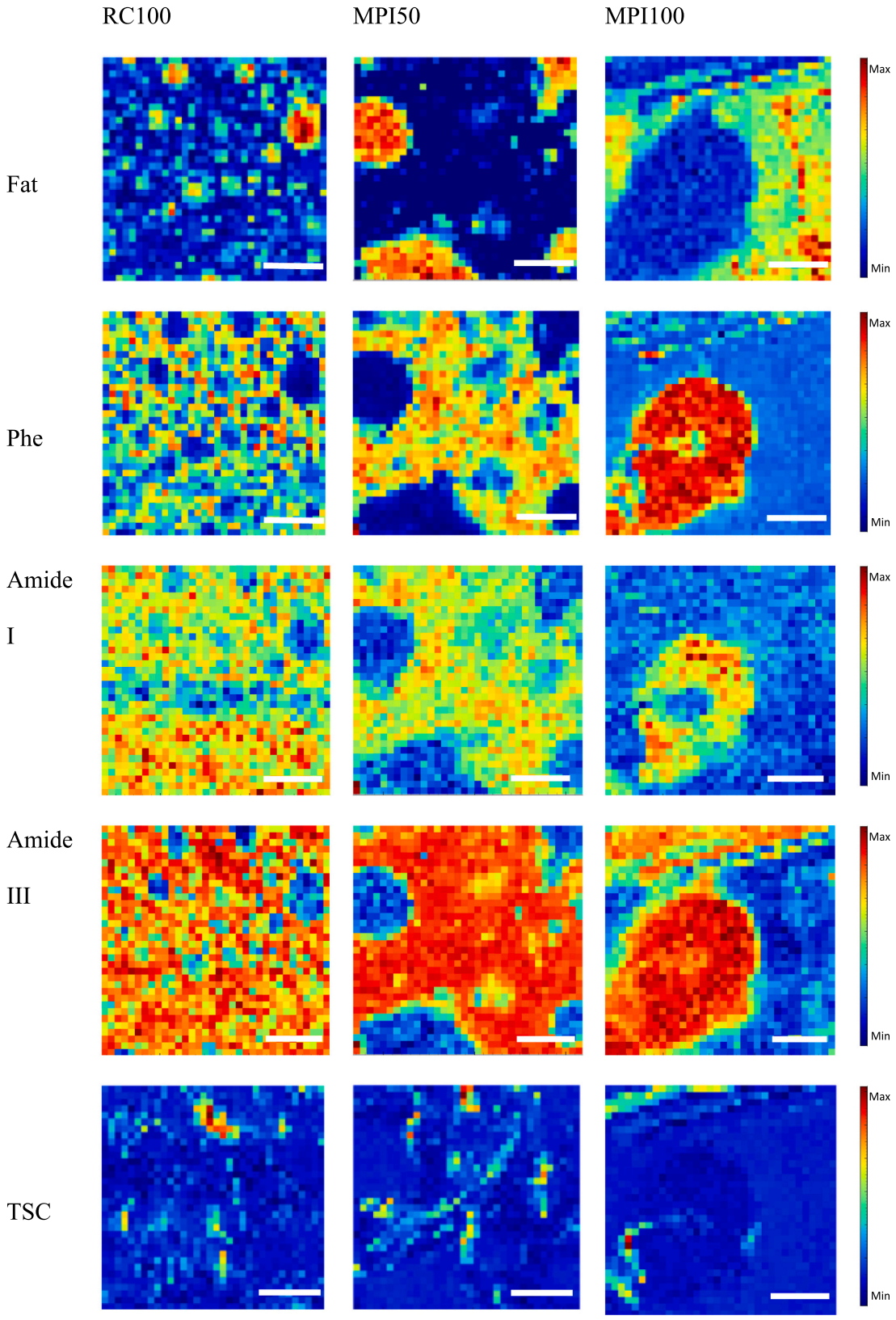


Fig. 2. Raman images of processed cheeses formulated with 100 % rennet casein (RC100), 50 % rennet casein and 50 % mung bean protein (MPI50), and 100 % mung bean protein (MPI100). Distribution of fat (1730–1750 cm^{-1}), phenylalanine (Phe, 1001 cm^{-1}), amide I (1600–1700 cm^{-1}), amide III (1200–1300 cm^{-1}), and trisodium citrate (TSC, 834 cm^{-1}) is shown, with red indicating high and dark blue low concentration. The scale bars are each 15 μm in length. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

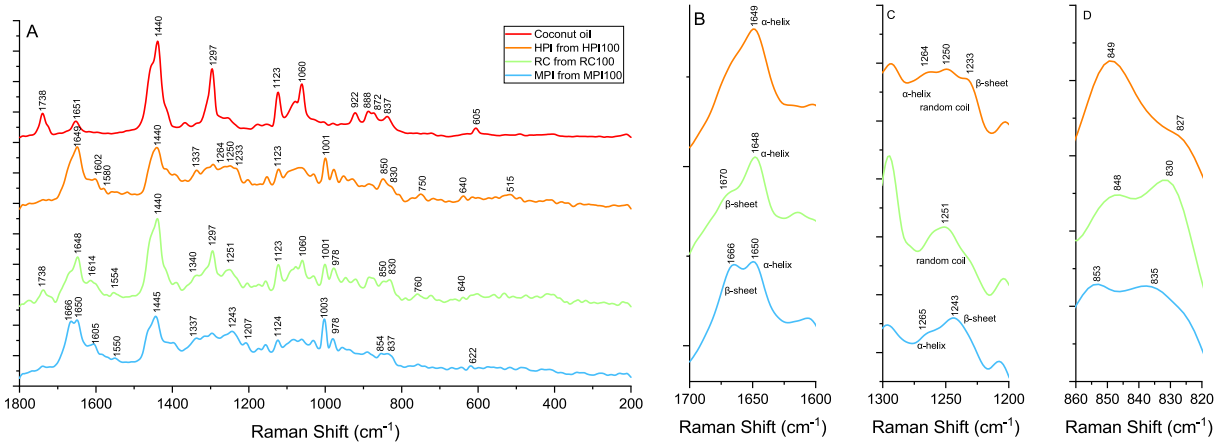


Fig. 3. Average Raman spectra of mung bean protein, hemp protein, casein protein, and coconut oil extracted from cheese formulated with 100 % mung bean protein (MPI100), 100 % hemp protein (HPI100), and 100 % rennet casein (RC100) by selecting the lipid and amide I regions with high intensity in the range of 200–1800 cm^{-1} (A), amide I region (1600–1700 cm^{-1}) (B), amide III region (1200–1300 cm^{-1}) (C), and tyrosine residues region (820–850 cm^{-1}) (D).

findings indicate co-localization of protein and lipid signals, suggesting possible protein-lipid interactions via the hydrophobic effect within the matrix. However, these co-localizations were not detected in HPI100 and MPI100, as there were no peaks at 1738 cm^{-1} for C═O stretching vibrations in triacylglycerols and 1297 cm^{-1} for CH_2 twisting in phospholipids.

Several protein-related Raman peaks provide valuable insights for comparing different proteins in cheese. In the amide I region (Fig. 3B), HPI exhibited a peak at 1649 cm^{-1}, which is primarily attributed to α-helical structures. In the amide III region (Fig. 3C), HPI showed distinct peaks at 1264 (α-helices), 1250 (random coils) and 1233 (β-sheets) cm^{-1}, suggesting a mixed secondary structure composition. In the amide I region, RC showed a peak at 1648 cm^{-1}, which is also indicative of α-helical structures; however, this peak may be influenced by coconut oil, given its proximity to the 1651 cm^{-1} band associated with lipids (Smith et al., 2017). In addition, RC showed a small shoulder at 1670 cm^{-1}, which is typically attributed to β-sheet formation, and a broad peak at 1251 cm^{-1} in the amide III region, which is associated with random coils. On the contrary, MPI exhibited a dominant α-helical signature with peaks at 1650 and 1265 cm^{-1} alongside β-sheet-related peaks at 1666 and 1243 cm^{-1}. These spectral differences suggest variations in protein structures and potential interactions within the cheese matrix. The presence of both α-helices and β-sheets in MPI indicates a more structured protein conformation than that in HPI or RC, whereas the broad peak at 1251 cm^{-1} in RC suggests increased protein unfolding, a more unordered structure or reduced secondary structure stability.

The tyrosine (Tyr) doublet bands at 830 and 850 cm^{-1} result from the Fermi resonance and ring vibrations of the para-substituted benzene ring (Yada et al., 1994). The intensity ratio, I_{850}/I_{830} monitors the hydrogen bonding and accessibility of the phenolic hydroxyl (OH) groups, indicating whether the Tyr residues are exposed or buried. If the I_{850}/I_{830} ratio is >1, Tyr residues are most likely located on the protein network surface. The phenolic OH groups are involved in moderate-to-weak hydrogen interactions (acting as both proton acceptors and donors). If the I_{850}/I_{830} ratio is <1, Tyr residues are likely to be fully or partially buried inside the protein structure. The phenolic OH groups function as proton donors in strong hydrogen interactions (Sadat et al., 2022).

As shown in Fig. 3D, RC had an I_{850}/I_{830} ratio of <1 (0.76) (data in Supplementary Table S1), which suggests that Tyr residues are buried within the protein network. This finding may be attributed to the strong casein–casein interactions forming a dense, cross-linked matrix in the cheese. This matrix buries many Tyr residues and promotes strong hydrogen interactions, potentially influencing the firmness and elasticity of the cheese. On the contrary, the I_{850}/I_{830} ratio of HPI was >1 (4.5), which indicates that Tyr residues are mostly exposed in HPI. This finding suggests weaker protein–protein interactions, which allow Tyr residues to be more accessible for forming hydrogen bonds with water. These interactions contribute to a softer texture in cheese, a finding consistent with those reported by Lu et al. (2026). The I_{850}/I_{830} ratio of MPI (1.15) was between those of RC and HPI, indicating the exposure of some Tyr residues. The reported ratios were derived from pre-processed, averaged spectra and should be interpreted cautiously, as they mainly illustrate relative trends in Tyr residue exposure among protein systems.

Given the large dataset, both supervised (CatBoost) and unsupervised (hierarchical clustering and k-means) models were used to differentiate mung bean protein and hemp protein from casein in dual-protein cheese systems. Fig. 4 presents the Raman images of MPI50 and HPI50 along with their corresponding average spectral data. The CatBoost model was trained using labelled data from RC100, MPI100 and HPI100, with predictions displayed as follows: RC in green, MPI or HPI in blue and lipids in red in MPI50 and HPI50 samples. Results showed that the model effectively applied learned patterns to predict protein distribution in mixed-protein cheese systems, even as spectral characteristics changed. CatBoost provides the advantage of feature ranking using its tree-based algorithm, which identifies the relative importance of each Raman shift in the predictive performance of the model (Supplementary Fig. S3). This ranking helps understand the key spectral regions that contribute most to the differentiation of protein sources, offering insights into the underlying chemical differences. The model revealed that the wavenumber region near 1440–1460 cm^{-1} was most significant for distinguishing RC from MPI, whereas the regions at 1440–1460, 1280 and 1367 cm^{-1} were most significant for differentiating RC from HPI. These Raman shifts correspond to vibrational modes associated with CH_2/CH_3 bending (1440–1460 cm^{-1}), amide III and C—N stretching (near 1280 cm^{-1}) and CH deformation or bending (near 1367 cm^{-1}), which reflect differences in the protein backbone structure, side-chain composition and lipid or hydrophobic group interactions between RC and plant proteins (Shoko et al., 2024; Smith et al., 2017; Yada et al., 1994; Zhang et al., 2023). In particular, casein, rich in hydrophobic residues, exhibits a side-chain environment and protein–lipid association distinct from those of globular plant proteins, leading to stronger CH_2/CH_3 and amide III band intensities. The 1440–1460 cm^{-1} region likely serves as a proxy for lipid-associated domains, consistent with the stronger lipid co-localization observed in RC-containing systems.

The Raman images generated using the CatBoost model served as a reference for comparing with those obtained from hierarchical and k-means clustering. For simplicity, the clustering results are interpreted as follows: red represents fats, green represents RC-dominated regions,

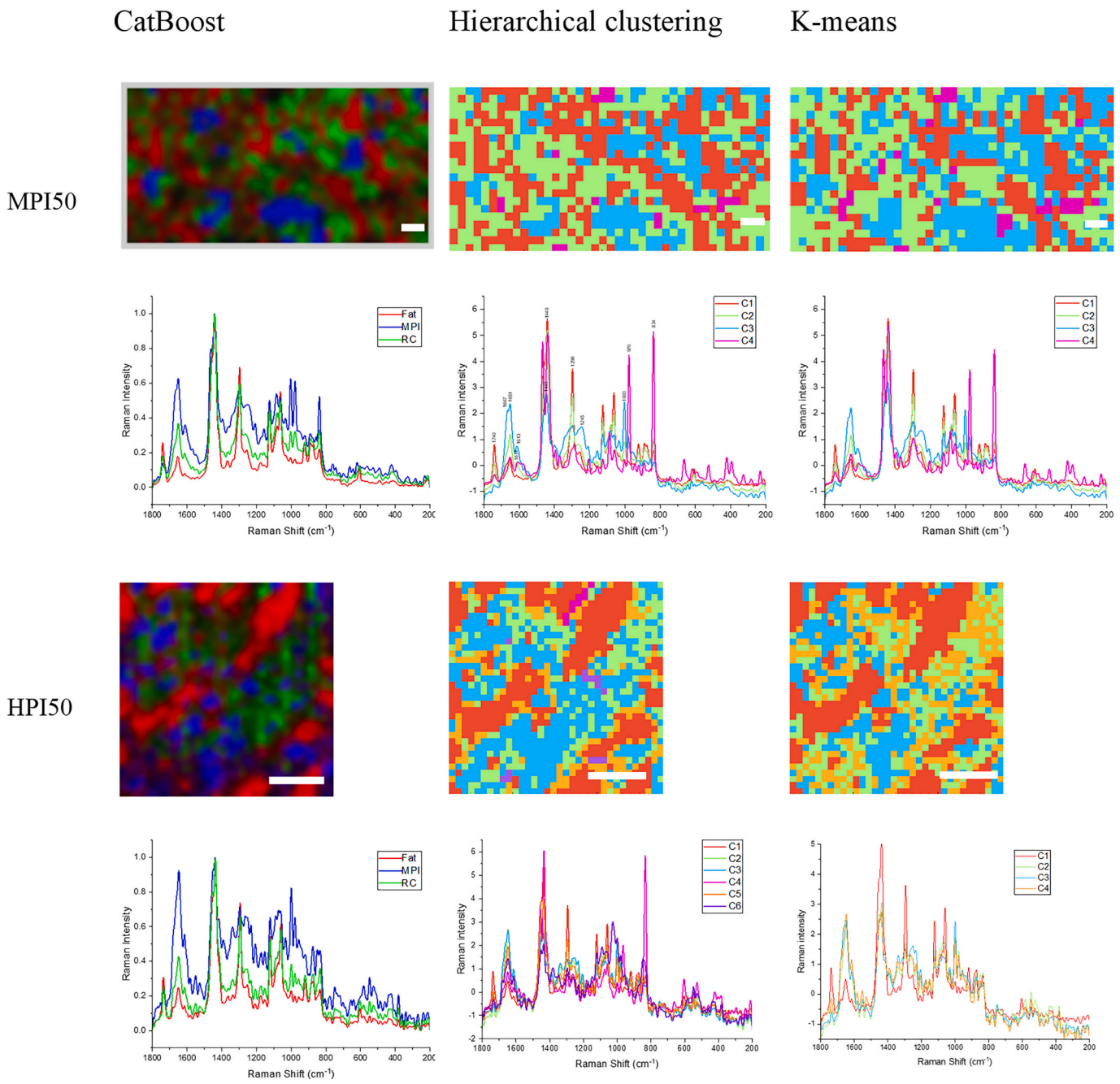


Fig. 4. Raman images of MPI50 and HPI50 cheeses (top) and their average spectra data (bottom) using the CatBoost (left), hierarchical clustering (middle), and k-means models (right). The CatBoost model was trained using single-protein cheese samples – RC100, MPI100, and HPI100 – with predicted RC shown in green, mung bean/hemp protein in blue, and fat in red in MPI50 and HPI50. MPI50 was imaged at low resolution (5 × 5 μm grid, 200 × 100 μm area) and HPI50 at high resolution (1.5 × 1.5 μm grid, 50 × 50 μm area). Colors indicate clusters from hierarchical clustering or k-means. Scale bars are 15 μm in length. MPI50: cheese formulated with 50 % mung bean protein and 50 % rennet casein. HPI50: cheese formulated with 50 % hemp protein and 50 % rennet casein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

blue represents MPI/HPI-dominated regions, pink represents TSC and orange represents the interface between fats and proteins. Of both unsupervised methods, k-means clustering identified plant proteins in a manner more consistent with CatBoost predictions. These findings align with previously reported CLSM and transmission electron microscopy findings (Lu et al., 2026), such as the circular structures observed in MPI particles. Supplementary Fig. S4 provides additional Raman images of HPCAs. The results show that these analysis models are consistent and can distinguish components in hybrid cheese with different formulations. Notably, in HPI50, hierarchical clustering required six clusters to distinguish HPI from RC, whereas k-means clustering achieved this differentiation using only four clusters. Both unsupervised models detected additional components, such as TSC. In addition, they successfully identified interfaces between fats and proteins in the high-resolution images of HPI50.

PCA was performed to the protein-specific clusters in MPI50 identified using k-means clustering (i.e. fats, TSC and interface clusters were removed) to distinguish RC from MPI. PC1 successfully mapped their spatial distribution (Fig. 5A). The PC1 loadings plot showed that positive loadings were dominated by lipid-associated bands, including 1741, 1440, 1297, 1123 and 1063 cm^{-1}, whereas negative loadings were dominated by protein-associated bands, including 1668, 1650,

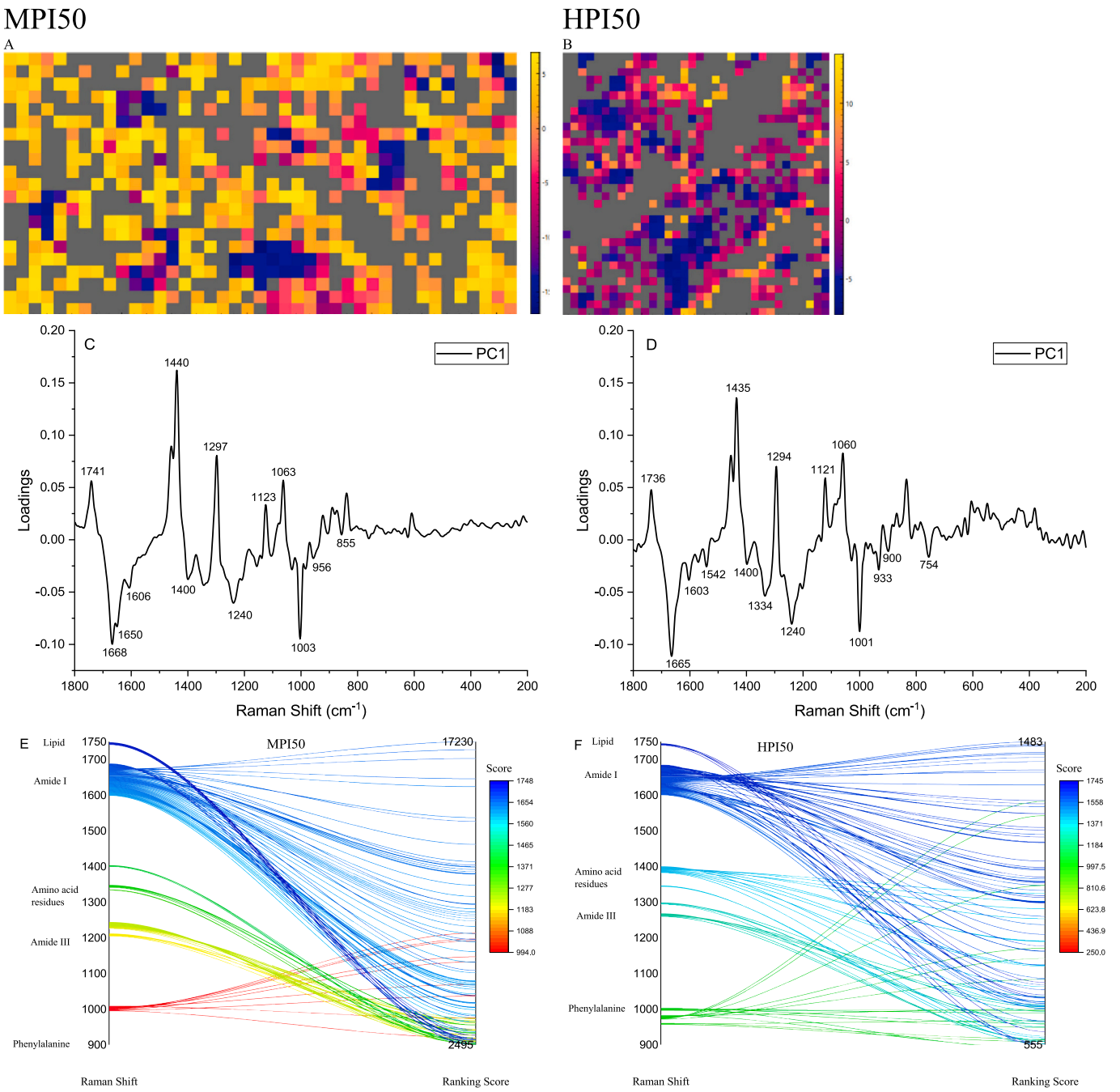


Fig. 5. Raman images of MPI50 and HPI50 based on PC1 scores for the 200–1800 cm^{-1} region (A & B, yellow = high, dark blue = low plant protein concentration), with corresponding loadings plots (C & D). Top 100 Raman shifts distinguishing RC from MPI or HPI are ranked (E & F), with lines connecting each shift (left axis) to its importance score (right axis, colour-coded). MPI50 was imaged at low resolution (5 × 5 μm grid, 200 × 100 μm), HPI50 at high resolution (1.5 × 1.5 μm grid, 50 × 50 μm). Key bands include Amide *I* (1600–1700 cm^{-1}), Amide III (~1250–1300 cm^{-1}), and bands associated with lipid (1745 cm^{-1}). MPI50: cheese formulated with 50 % mung bean protein and 50 % rennet casein. HPI50: cheese formulated with 50 % hemp protein and 50 % rennet casein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1606,1400, 1240, 1003, 956 and 855 cm^{-1} (Fig. 5C). Ranking variables by their absolute values confirmed that the amide I region (1667–1674 cm^{-1}), which is associated with β-sheets, was most influential in differentiating MPI from RC. Additional contributors included lipid-related peaks, Phe bands and amide III differences (Fig. 5E). These findings are consistent with the spectral characteristics observed in single-protein cheese systems (RC100 and MPI100), highlights that the higher content of β-sheets content in MPI than in RC is a key factor influencing the structural and functional properties of the hybrid cheese matrix.

A similar trend was observed in the case of HPI50 (Fig. 5B), wherein both PC1 loadings and importance ranking scores highlighted amide I as a key spectral region contributing to the differentiation between RC and HPI. In this region, spectral variations were observed near 1660 (random coils) and 1664–1672 cm^{-1} (β-sheets) (Fig. 5D and F). Additional contributors identified from PC1 loadings included peaks at 1603 (Phe), 1542 (amide II), 1334 (possibly CH bending or side-chain vibrations) (Shoko et al., 2024), 933 (β-sheet stretching), 900 (tentatively assigned to C—C stretching or ring vibrations) (Freire et al., 2017), and 754 cm^{-1} (possibly Cystine C—S stretching) (Yada et al., 1994). These

spectral differences effectively reflected the spatial distributions of HPI and RC in HPI50.

β-Sheets (1664–1672 cm^{-1}) were prominent in both HPI50 and MPI50, indicating that both plant proteins influenced the secondary structure of casein, but the extent of β-sheet contribution varied. The amide II peak (1542 cm^{-1}) appeared significant in HPI50 but was less pronounced in MPI50, suggesting differences in hydrogen bonding and backbone flexibility between the two plant proteins. The presence of additional peaks in HPI50 (932, 900 and 755 cm^{-1}) indicated that interactions between HPI and RC led to more structural heterogeneity than those between MPI and RC, with distinct RC and MPI regions showing clearer spectral separation. By comparing the spectral features of mixed-protein cheeses with those of single-protein cheeses, the models successfully differentiated between plant and milk proteins while providing valuable structural insights.

3.4. Protein–protein and protein–lipid interactions

To examine protein–protein and protein–lipid interactions in dual-protein cheese systems, the average Raman spectra of RC, HPI and MPI in single-protein cheeses (RC100, HPI100 and MPI100, respectively) were compared with their counterparts in dual-protein cheeses (HPI50 and MPI50). The spectra were extracted using k-means clustering to assess structural changes upon mixing.

As shown in Fig. 6A, the Raman spectrum of RC in MPI50 exhibited slight shifts in amide I and amide III bands (from 1670, 1648 and 1251 cm^{-1} in RC100 to 1678, 1651 and 1260 cm^{-1} in MPI50, respectively). These spectral shifts suggest an increase in β-sheet content and potential α-helix stabilisation. On the contrary, the spectrum of RC in HPI50 showed multiple peaks in the amide III region (1265, 1251 and 1230 cm^{-1}), indicating a more heterogeneous secondary structure. These findings suggest that compared with mung bean protein, hemp protein induces greater structural heterogeneity in casein. The subsequent analysis of Tyr residues supported these structural differences. In RC100, Tyr residues were largely buried owing to strong casein–casein interactions indicated by an I_{850}/I_{830} ratio of <1 (0.76). Conversely, RC in HPI50 exhibited a slightly higher I_{850}/I_{830} ratio (~0.95) (Supplementary Table S1), suggesting higher exposure of Tyr residues owing to weakened casein–casein interactions following the incorporation of HPI. Notably, the spectrum of RC in MPI50 lacked the peak at 850 cm^{-1} but showed a sharp peak at 837 cm^{-1}. These changes indicated that more Tyr residues were buried after the incorporation of MPI, likely reflecting reduced interactions between RC and MPI. These findings are consistent with earlier observations that mung bean protein forms larger aggregates (Lu et al., 2025; Lu et al., 2026). These aggregates create gaps between casein micelles and modify the Tyr environment in RC. The spectrum of RC in RC100 exhibited a peak at 530 cm^{-1}, which is attributed to the disulfide bonds in α_{S2}-casein. However, in MPI50, no distinct disulfide peak was observed, suggesting reduced cross-linking in the presence of MPI. Conversely, in HPI50, RC showed peaks at 526 and 549 cm^{-1}, indicating potential disulfide bond rearrangements.

On comparing the Raman spectra of MPI in MPI100 and MPI50 (Fig. 6B), we found that the amide I region exhibited no significant peak shifts; however, a decrease in β-sheet intensity was observed after the incorporation of RC. No significant differences were observed in the amide III region of MPI between MPI100 and MPI50, indicating the predominance of β-sheet structures. These findings suggest that MPI retains its aggregated structure and that its interaction with casein may lead to microstructural segregation instead of conformational changes. The I_{850}/I_{830} ratio slightly increased from 1.15 in MPI100 to 1.43 in MPI50. This change suggests that the incorporation of RC increases the exposure of Tyr residues in MPI, slightly weakening protein–protein interactions. An increase in lipid intensity was observed in the spectrum of MPI in MPI50, indicating an increase in MPI–lipid co-localizations.

The Raman spectra of HPI in HPI100 and HPI50 (Fig. 6C) exhibited a peak at 1649 cm^{-1} (α-helix). In HPI50, an additional shoulder was observed at 1661 cm^{-1}, which indicated an increase in the content of random coils upon the incorporation of RC. These findings were further supported by the spectral changes in the amide III region of HPI50, wherein higher intensity at 1251 cm^{-1} (random coil) suggested structural destabilisation. These findings indicate that casein disrupts the structural integrity of hemp protein, promoting structural disorder. The I_{850}/I_{830} ratio decreased from 4.5 in HPI100 to 1.20 in HPI50, suggesting a significant reduction in Tyr residue exposure in the presence of casein. These findings indicate that casein–HPI interactions lead to structural rearrangements, increasing intramolecular interactions and reducing the accessibility of water to Tyr residues. Furthermore, shifts in disulfide bond–related peaks in HPI50 (from 515 and 535 cm^{-1} in HPI100 to 507, 526 and 549 cm^{-1} in HPI50) suggested disulfide bond rearrangements, likely influenced by casein–HPI interactions. An increase in lipid intensity was observed in the spectrum of HPI in HPI50, indicating increased HPI–lipid co-localizations.

Altogether, Raman spectroscopy revealed distinct structural interactions between RC and MPI as well as RC and HPI in dual-protein cheese systems. The incorporation of RC does not significantly alter the structure of MPI but rather promotes its microstructural segregation. In contrast, RC disrupts the structural integrity of HPI, inducing structural disorder and disulfide bond rearrangements while reducing the exposure of Tyr residues. For both plant proteins, the inclusion of RC increases their co-localization with lipids.

3.5. Changes in protein secondary structures in MPI–RC cheese systems

Given that the amide I region was identified as an important spectral feature for distinguishing MPI and HPI from RC and relevant spectral changes were observed in dual-protein cheese systems, quantitative analysis was performed to elucidate changes in protein secondary structures in this region through deconvolution.

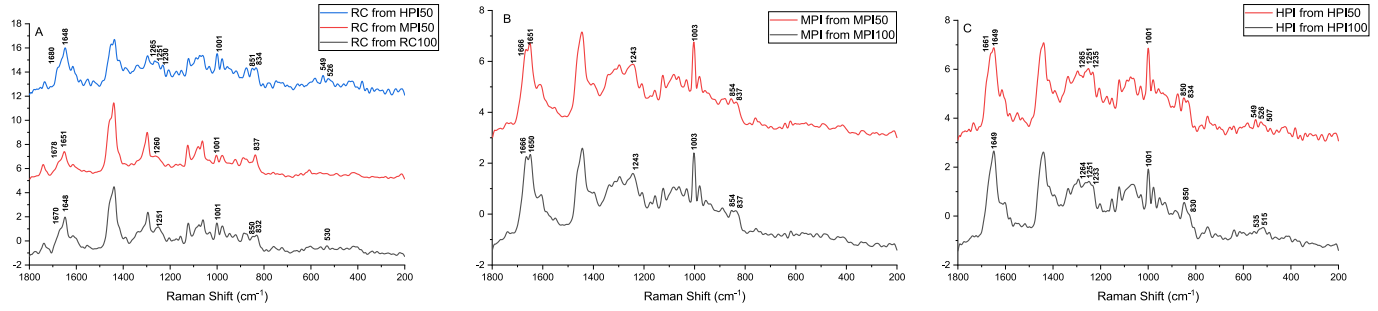


Fig. 6. Average spectra of RC from RC100, MPI50, and HPI50 (A), MPI from MPI100 and MPI50 (B), and HPI from HPI100 and HPI50 (C). Spectra were extracted using k-means clustering to evaluate structural changes upon mixing. RC100: cheese formulated with 100 % rennet casein. MPI100: cheese formulated with 100 % mung bean protein. MPI50: cheese formulated with 50 % mung bean protein and 50 % rennet casein. HPI100: cheese formulated with 100 % hemp protein. HPI50: cheese formulated with 50 % hemp protein and 50 % rennet casein.

Table 2 shows the quantitative analysis of protein secondary structure changes in HPCAs containing MPI and RC at various ratios. Fig. S5 presents the peak fitting of the amide I band, providing additional information for the quantification. In MPI100, α-helix (38 %) and β-sheet (53 %) structures were predominant. However, as the proportion of casein increased, a shift from β-sheet (20 % in MPI30) to β-turn (9 % in MPI100 and 41 % in MPI30) was detected, suggesting alterations in hydrogen interactions (Sim et al., 2021). An increase in the proportion of MPI influenced the structural conformation of RC. A higher proportion of MPI was associated with higher contents of α-helices and β-sheets and a lower content of random coils. These structural shifts suggest stabilisation of the protein network, which may contribute to enhanced rigidity and impact textural attributes such as firmness and elasticity. In addition, the aggregated intermolecular β-sheet structures near 1620 cm^{-1} in both MPI100 and RC100 (Supplementary Fig. S5) indicated a denser microstructure and higher firmness (Ong et al., 2020), which aligns with the findings of our previous study. However, it is important to consider the differences between MPI–lipid and RC–lipid interactions. RC exhibits more frequent co-localizations with coconut oil, which likely obscures the amide band at approximately 1670 cm^{-1} (Smith et al., 2015). Consequently, the observed secondary structure of RC is more influenced by the intensity of co-localized lipid features in the Raman spectra than that of MPI. As cheese samples are a complex food matrix, the results are intended to highlight relative trends and indicative structural differences.

3.6. Comparison of RC–MPI and RC–HPI interactions in dual-protein cheese systems

Raman spectroscopy highlighted distinct structural interactions between RC and MPI and between RC and HPI in hybrid cheese systems. These interactions may influence the texture and functionality of cheese. As reported previously (Lu et al., 2026), MPI-based HPCAs exhibited greater stretchability and higher G' values and hardness, but lower resilience, cohesiveness and meltability compared to HPI-based counterparts. The Raman spectra provide a possible explanation for these behaviours, indicating that MPI promotes pronounced protein domain separation and higher β-sheet content, consistent with stronger protein aggregation. Consequently, MPI tends to form a more segregated and rigid protein microstructure. This interpretation is in line with the rheological results, where elevated G′ values and lower tan δ reflect a more elastic and less deformable network, ultimately leading to reduced meltability and stretchability. Furthermore, the altered Tyr interactions and weakened disulfide cross-linking in MPI–RC cheese samples indicated reduced interconnectivity between casein and MPI. In contrast, HPI–RC mixtures displayed indication of disulfide bond rearrangement and random coil enrichment, implying a more flexible and heterogeneous network. These structural features are consistent with the SDS-PAGE and rheological behaviour observed in our earlier study (Lu et al., 2026), where under non-reducing conditions, a decrease in α_{s1}-casein and α_{s2}-casein band intensities was observed in HPI–RC systems but not in MPI–RC systems, implying potential disulfide-mediated interactions between α_{s2}-casein and HPI. The HPI–RC systems also showed thermally responsive network reorganization in rheological tests, supporting the presence of a dynamic protein network that likely contributes to enhanced meltability compared with the more rigid MPI–RC systems. Overall, the effects of RC–HPI interactions on cheese functionality may differ from those of MPI–RC interactions.

4. Conclusions

This study shows that integrating confocal Raman microscopy with both supervised and unsupervised ML models can effectively map the distribution of key components and differentiate between plant proteins and casein in hybrid cheese systems. PCA showed that variations in the amide I and lipid-associated bands were key spectral features for distinguishing casein from plant proteins. Protein–protein and protein–lipid interactions were investigated by comparing the structures of single- and dual-protein cheese systems. In single-protein cheese systems, RC showed more frequent interactions with lipids than MPI and HPI. In hybrid cheeses, an increase in RC content slightly reduced β-sheet content in MPI–RC cheeses and increased random coil content in HPI–RC cheeses as evidenced by changes in both amide I and amide III bands. Deconvolution of the amide I region in MPI–RC cheeses prepared at various ratios validated these findings quantitatively, showing a decrease in β-sheet content but an increase in β-turn content in MPI. Moreover, MPI–RC and HPI–RC interactions had different effects. In particular, the incorporation of HPI induced greater structural disorder and disulfide bond rearrangements. By contrast, the incorporation of MPI enhanced protein domain separation owing to its strong aggregation tendency and high β-sheet content. Although these findings are specific to MPI and HPI in hybrid cheese matrices, the methodology—confocal Raman microscopy combined with ML analysis—can be applied to other plant proteins and food systems. The spatial mapping and spectral analysis approach provides high-resolution insights into protein and lipid distribution, but the specific structural effects and interpretations are expected to vary depending on the protein type and food matrix. This technique offers high-resolution insights that can be validated and complemented by infrared (IR) spectroscopy–based techniques in future research.

CRediT authorship contribution statement

Di Lu: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cushla McGoverin:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Debashree Roy:** Writing – review & editing, Supervision. **Alejandra Acevedo-Fani:** Writing – review & editing, Supervision. **Harjinder Singh:** Writing – review & editing, Supervision. **Mark Waterland:** Writing – review & editing. **Yi Zheng:** Writing – original draft, Methodology, Formal analysis. **Aiqian Ye:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Table 2
Relative amount of protein secondary structures in MPI–RC cheese samples determined by peak fitting of the amide I in Raman Spectra.

	MPI secondary structure change (%)				RC secondary structure change (%)			
	α-Helix	Random coil	β-Sheet	β-Turn	α-Helix	Random coil	β-Sheet	β-Turn
MPI100	38 ± 10	0	53 ± 8^{a}	9 ± 3^{a}	–	–	–	–
MPI70	41 ± 12	0	50 ± 14^{a}	8 ± 3^{a}	54 ± 9	0^{a}	37 ± 5^{a}	8 ± 4
MPI50	49 ± 5	0	19 ± 8^{b}	32 ± 6^{a}	45 ± 16	22 ± 12^{b}	28 ± 10^{ab}	5 ± 4
MPI30	38 ± 17	0	20 ± 9^{b}	41 ± 16^{b}	36 ± 13	37 ± 8^{b}	16 ± 6^{b}	11 ± 10
RC100	–	–	–	–	35 ± 8	44 ± 1^{b}	11 ± 4^{b}	10 ± 4

Values within a column not having common superscripts differ significantly ($p < 0.05$). RC100: cheese formulated with 100 % rennet casein. MPI100: cheese formulated with 100 % mung bean protein. MPI70: cheese formulated with 70 % mung bean protein and 30 % rennet casein. MPI50: cheese formulated with 50 % mung bean protein and 50 % rennet casein. MPI30: cheese formulated with 30 % mung bean protein and 70 % rennet casein.

Funding

This research work was supported by the Riddet Institute Centre of Research Excellence (CoRE), funded by the New Zealand Tertiary Education Commission.

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

The authors thank Dr. Matthew Perrott for facilitating the sample preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2025.147102.

Data availability

Data will be made available on request.

References

Campos, N., Paula, I., Perrone, Í., De Oliveira, L. F., Doriguetto, A., Wolfschoon, A., & Stephani, R. (2022). Characterization of the white dots defect ("PIPS") in "Doce de leite". *European Food Research and Technology, 249*, 1–11. https://doi.org/10.1007/s00217-022-04149-x

Chawanji, A., Holroyd, S. E., & Nickless, E. (2022). Raman confocal microscopy to assess changes in cheddar cheese during maturation [Article]. *International Journal of Dairy Technology, 75*(4), 882–891. https://doi.org/10.1111/1471-0307.12897

Dewantier, G. R., Torley, P. J., & Blanch, E. W. (2023). Identifying chemical differences in cheddar cheese based on maturity level and manufacturer using vibrational spectroscopy and chemometrics. *Molecules, 28*(24), 8051. https://www.mdpi.com/1420-3049/28/24/8051.

Freire, P. D. T., Lima, J., Jr., Mendes Filho, J., Melo, F., & Barboza, F. (2017). In M. Khan (Ed.), *Raman spectroscopy and applications*. IntechOpen. https://doi.org/10.5772/65480.

Holroyd, S. E., & Nickless, E. (2024). Understanding mozzarella cheese microstructure with Raman confocal imaging. *International Dairy Journal, 152*, Article 105893. https://doi.org/10.1016/j.idairyj.2024.105893

Holroyd, S. E., Nickless, E., & Watkinson, P. (2023). Raman and mid-infrared spectroscopy to assess changes in Cheddar cheese with maturation. *International Journal of Dairy Technology, 76*(2), 408–417. https://doi.org/10.1111/1471-0307.12929

Kim, W., Wang, Y., & Selomulya, C. (2020). Dairy and plant proteins as natural food emulsifiers. *Trends in Food Science & Technology, 105*, 261–272. https://doi.org/10.1016/j.tifs.2020.09.012

Lu, D., Roy, D., Acevedo-Fani, A., Singh, H., Waterland, M., & Ye, A. (2026). Physical properties and microstructure of hybrid processed cheeses formulated with plant protein and milk protein ingredients. *Food Hydrocolloids, 170*, Article 111688. https://doi.org/10.1016/j.foodhyd.2025.111688

Lu, D., Roy, D., Acevedo-Fani, A., Singh, H., & Ye, A. (2025). Investigation of various plant protein ingredients for processed cheese analogues: Physical properties and microstructure compared with milk proteins. *International Journal of Food Science and Technology, 60*(1). https://doi.org/10.1093/ijfood/vvae018

Nickless, E., & Holroyd, S. (2020). Raman imaging of protein in a model cheese system. *Journal of Spectral Imaging, 9*(1), Article a9. https://doi.org/10.1255/jsi.2020.a9

Oliveira, K., Callegaro, L., Stephani, R., Almeida, M., & De Oliveira, L. F. (2016). Analysis of spreadable cheese by Raman spectroscopy and chemometric tools. *Food Chemistry, 194*, 441–446. https://doi.org/10.1016/j.foodchem.2015.08.039

Ong, L., Pax, A. P., Ong, A., Vongsvivut, J., Tobin, M. J., Kentish, S. E., & Gras, S. L. (2020). The effect of pH on the fat and protein within cream cheese and their influence on textural and rheological properties. *Food Chemistry, 332*, Article 127327. https://doi.org/10.1016/j.foodchem.2020.127327

Sadat, A., Corradini, M. G., & Joye, I. J. (2022). Vibrational and fluorescence spectroscopy to study gluten and zein interactions in complex dough systems. *Current Research in Food Science, 5*, 479–490. https://doi.org/10.1016/j.crfs.2022.02.009

Shoko, P. T., Blanch, E. W., Torley, P. J., & Pillidge, C. (2024). Raman spectroscopy for the differentiation of muscles and tissues in meat using chicken as a model system. *Journal of Raman Spectroscopy, 55*(11), 1146–1155. https://doi.org/10.1002/jrs.6725

Sim, S. Y. J., Srv, A., Chiang, J. H., & Henry, C. J. (2021). Plant proteins for future foods: A roadmap. *Foods, 10*(8), 1967. https://www.mdpi.com/2304-8158/10/8/1967.

Smith, G. P. S., Holroyd, S. E., Reid, D. C. W., & Gordon, K. C. (2017). Raman imaging processed cheese and its components. *Journal of Raman Spectroscopy, 48*(3), 374–383. https://doi.org/10.1002/jrs.5054

Smith, G. P. S., McGoverin, C. M., Fraser, S. J., & Gordon, K. C. (2015). Raman imaging of drug delivery systems. *Advanced Drug Delivery Reviews, 89*, 21–41. https://doi.org/10.1016/j.addr.2015.01.005

Smith, J. R. (2013). *Assessment of structure and component mobility within mozzarella cheese*. Doctoral dissertation. Massey University.

Stocco, G., Gómez-Mascaraque, L. G., Deshwal, G. K., Sanchez, J. C., Molle, A., Pizzamiglio, V., … Cipolat-Gotet, C. (2024). Exploring the use of NIR and Raman spectroscopy for the prediction of quality traits in PDO cheeses. *Frontiers in Nutrition, 11*, Article 1327301. https://doi.org/10.3389/fnut.2024.1327301

Toplak, M., Read, S., Sandt, C., & Borondics, F. (2021). Quasar: Easy Machine Learning for Biospectroscopy. *Cells, 10*, 2300. https://doi.org/10.3390/cells10092300

Wang, C.-H., Huang, C.-C., Lin, L.-L., & Chen, W. (2016). The effect of disulfide bonds on protein folding, unfolding, and misfolding investigated by FT–Raman spectroscopy. *Journal of Raman Spectroscopy, 47*(8), 940–947. https://doi.org/10.1002/jrs.4935

Yada, R. Y., Jackman, R. L., & Smith, J. L. (Eds.). (1994). *Protein structure-function relationships in foods*. Blackie Academic & Professional.

Zhang, D., Jiang, K., Luo, H., Zhao, X., Yu, P., & Gan, Y. (2024). Replacing animal proteins with plant proteins: Is this a way to improve quality and functional properties of hybrid cheeses and cheese analogs? *Comprehensive Reviews in Food Science and Food Safety, 23*(1), Article e13262. https://doi.org/10.1111/1541-4337.13262

Zhang, Z. Y., Jiang, M. Q., & Xiong, H. M. (2023). Optimized identification of cheese products based on Raman spectroscopy and an extreme learning machine. *New Journal of Chemistry, 47*(14), 6889–6894. https://doi.org/10.1039/d2nj06350f

Zhao, Y., Ma, C. Y., Yuen, S. N., & Phillips, D. L. (2004). Study of succinylated food proteins by Raman spectroscopy. *Journal of Agricultural and Food Chemistry, 52*(7), 1815–1823. https://doi.org/10.1021/jf030577a