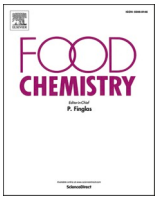
Food Chemistry 440 (2024) 138206

Contents lists available at ScienceDirect

Food Chemistry

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



Rapid and green discrimination of bovine milk according to fat content, thermal treatment, brand and manufacturer via colloidal fingerprinting

Stefano Giordani [a], Nicholas Kassouf [a], Alessandro Zappi [a], Andrea Zattoni [a,b], Barbara Roda [a,b], Dora Melucci [a], Valentina Marassi [a,b,*]

[a] Department of Chemistry "Giacomo Ciamician", University of Bologna, 40126 Bologna, Italy
[b] byFlow srl, 40129 Bologna, Italy

ARTICLE INFO

Keywords:
Field Flow Fractionation (FFF) multidetection - chemometrics
Green analytical methods
Milk proteins
Food analysis
Heat treatment, fat content, manufacturer, sub-typologies
Colloidal fingerprinting

ABSTRACT

Addressing food safety and detecting food fraud while fulfilling greenness requisites for analysis is a challenging but necessary task. The use of sustainable techniques, with limited pretreatment, non-toxic chemicals, high throughput results, is recommended. A combination of Field Flow Fractionation (FFF), working in saline carrier and with minimal preprocessing, and chemometrics was for the first time applied to bovine milk grouping. A set of 47 bovine milk samples was analyzed: a single analysis yielded a characteristic multidimensional colloidal dataset, that once processed with multivariate tools allowed simultaneously for different discriminations: fat content, thermal treatment, brand and manufacturing plant. The analytical methodology is fast, green, simple, and inexpensive and could offer great help in the field of quality control and frauds identification. This work represents also the first attempt to identify milk sub-typologies based on colloidal profiles, and the most complete study concerning multivariate analysis of FFF fingerprint.

1. Introduction

Food authentication and origin discrimination are some of the most important tasks in the food industry. In the EU, clearly labelled geographical origin is required for the globalization of food trade and to fulfill transparency requirements from consumers (Law 2002). However, regulations are unfortunately unable to prevent food fraud (Ballin 2010) and consequently the availability of analytical methods suited to the task is of the outmost importance. Food analysis is crucial to ensure consumer safety, assess authenticity, and detect frauds. However, the high volume of analyses and the often-complex protocols, requiring sample extraction and organic solvents, open to the search of more sustainable techniques, that are nonetheless able to give satisfactory results, fast answers, and high-quality data. The main foods subjected to fraud are the ones of animal origin such as meat, honey, milk, dairy products, fish and seafood (Cubero-Leon, Peñalver, and Maquet 2014). In particular milk is one of the staples of Western diet and is a food product vastly commercialized as drink or processed as dairy product. According to Eurostat, raw milk production in the EU was 161 million tons in 2021, 96% of which was cows' milk making its authentication a key topic in the food industry (Nascimento et al. 2017).

From a chemical point of view, milk is a colloidal suspension whose main components can be divided into (1) lipids, (2) proteins, (3) carbohydrates (mainly lactose) and (4) microconstituents (vitamins, mineral salts), and whose relative composition is extremely variable, especially between different milk types (Fox, Mcsweeney, and Paul, 1998). Milk proteins can be divided into two large groups: caseins and whey proteins. Whey proteins are a collection of various globular proteins, in particular globulins and albumins. (Madureira et al. 2007). Caseins are complex protein aggregates forming micelles with diameters ranging between 50 and 600 nm (average diameter of 150 nm) (Hristov et al. 2016). These proteins are characterized by a not well-defined secondary structure (due to proline richness), and by the presence of phosphate groups residuals (mainly phospho-serine and phospho-threonine). Fat content can vary from below 3% to more than 6%, is composed mainly of triglycerides (>98% of the fat bulk) and it is present in the milk as milk fat globules (MFG) with diameters ranging from < 1 to about 10 μm (average size is 1 μm) (Jensen 2002). The density of fat is lower than the surrounding aqueous serum leading to separation of unprocessed milk into a phase enriched in MFG (i.e., cream) and a phase largely depleted of such globules (i.e., skim milk) (Huppertz, Uniacke-Lowe, and Kelly 2020). This phenomenon, accelerated by the

* Corresponding author at: Department of Chemistry "Giacomo Ciamician", University of Bologna, 40126 Bologna, Italy.
E-mail address: valentina.marassi@unibo.it (V. Marassi).

https://doi.org/10.1016/j.foodchem.2023.138206
Received 30 September 2023; Received in revised form 20 November 2023; Accepted 11 December 2023
Available online 13 December 2023

application of a centrifugal force, is exploited to obtain skimmed (<0.5% fat) and semi skimmed milk (1.7%-2.0% fat). Bovine milk undergoes various types of heat treatment to eliminate the content of pathogenic microorganisms for humans and to reduce the endogenous microflora, in order to increase the shelf life of the product. Treatments such as pasteurization are essentially carried out on all types of milk and are mild treatments. Others, such as Ultra High Temperature (UHT), are defined as strong treatments as they cause denaturation and aggregation of the whey proteins. Caseins, having a greater resistance to temperature (De Kruif and Holt 2003), are less affected. Treatment can cause a variation in the global shape of the micelle and or the formation of serum proteins-casein aggregates (Hennetier et al. 2020; Krishna et al. 2021). A mild alternative to UHT treatment to increase milk shelf life is microfiltration (Lie-Piang et al. 2021). The latter reduces the content of pathogenic microorganisms by filtrating milk trough membranes of micro-sed pores able to retain bacteria. However, to guarantee the total absence of pathogenic bacteria the MF has to be combined with a heat treatment (at least a pasteurization step). Milk adulteration and authentication studies usually exploit single target testing approaches which have the drawback of providing limited and specific information on the sample's status (Azad and Ahmed 2016). As highlighted in a recent work, improvement in this field are being made through multi-targeted and fingerprinting approaches which, by exploiting chemometric tools, provide more information while reducing the tediousness of the workflow (Zappi et al. 2023). NMR, GC, HPLC and genomics/transcriptomics are the most exploited techniques to the untargeted authentication of milk and dairy products (Qin et al. 2022; Esteki, Shahsavari, and Simal-Gandara 2020). However, these approaches require expensive instrumentation (NMR, transcriptomics), suffer from interferences (genomics), require intensive sample pretreatment (GC, HPLC) and overall can be extremely complex in terms of data treatment (e.g., identification of certain markers from MS analysis) (Qin et al. 2022).

In physical terms, milk is a colloid. Origin, treatment, milk type, all contribute to a different colloidal composition. Colloidal nano systems can be successfully separated and analyzed by Field Flow Fractionation (FFF) multidetection, an analytical platform that can be customized by means of different online detection. Asymmetrical Flow Field Flow Fractionation (AF4), along with its miniaturized version Hollow Fiber Flow FFF (HF5), is the most common and exploited FFF variant (Plavchak et al. 2021; Wang et al. 2022). Sample components, differing by hydrodynamic size and/or other physical properties, are differentially eluted. The coupling with diode array (DAD), differential refractive index (dRI), fluorescence (FLD) and multi angle light scattering (MALS) detectors allows a size, morphological, and spectroscopical characterization of the separated species. Compared to other separation devices, AF4 platforms are characterized by a wide operating range, provide an extremely gentle separation, and are characterized by high flexibility of mobile phase and sample injectable (often requiring little to no pretreatment). These features allowed AF4 techniques to be exploited to separate and characterize (mass, size and spectroscopic properties) several samples such as nanoparticles (Caputo et al. 2021), bioparticles (Marassi et al. 2022; Nilsson 2013; Ventouri et al. 2022) and highly complex matrices such as biological fluids (Zhang and Lyden 2019; Mangal et al. 2018; Giordani et al. 2023) and food (Correia and Loeschner 2018; Krebs et al. 2019; Marassi et al. 2021).

Applying multivariate analysis on the information-rich outputs provided by separative-multidetection platforms is still widely unexplored on FFF data since, when FFF is involved, most of the times multivariate analysis is not directly performed on FFF results but rather on downstream dataset obtained applying other analytical techniques to the analytes firstly separated by FFF (Austin et al. 2000; Kim et al. 2018). Such methodologies, although very powerful, are extremely time consuming and complex, since most of the times more than one additional analytical technique (e.g., uHPLC, SDS-page) is required. The need of using multiple analytical techniques to obtain the dataset necessary to the multivariate elaboration implies also higher costs and reduces the greenness of the study, especially when approaches based on organic solvents (e.g., HPLC) are required.

A simpler, faster, cheaper, and more environmentally friendly alternative, still widely unexplored in FFF, involves the direct use of the signals over time collected by the detectors during the analysis as fingerprints to distinguish different species. Down this line, an approach exploiting Principal Components Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA) on FFF-derived datasets only, allowed to gather information of regional provenience of tomato sauces (Zappi et al. 2022) and distinguish between viable and non-viable cells of the same strain.

Due to its colloidal nature, milk and its derivates have been studied with AF4 (Lie-Piang et al. 2021; Guyomarc'h et al. 2010; Hennetier et al. 2020). The works in the literature have only focused on the dimensional and morphological characterization of whey and caseins proteins (Velazquez-Dominguez et al. 2023; Nogueira et al. 2023) and none of them focused on FFF colloidal fingerprint.

In the framework of researching for analytical methods able to both fulfill greenness requirements and address the most challenging health concerns in the food industry, we explored the use of FFF colloidal fingerprinting-chemometrics towards food classification and grouping. As a case study, a set of bovine milks differing based on fat content (Whole, Semi Skimmed, Skimmed), thermal treatment (UHT, NON-UHT), brand and manufacturing plant have been analyzed with an AF4-UV-FLD-MALS platform after a limited sample preparation. The developed method did not require organic solvents, working in a saline solution mimicking milk ionic strength, pH, and calcium concentration. Statistical elaboration performed on datasets constituted by selected portions of the whole signals, acting as fingerprint, gathered an unprecedented series of results extremely important in the field of quality control. In particular, it was possible to discriminate samples according to thermal treatment, fat content, and manufacturing plant. The results also highlight the importance of associating multivariate analysis to data directly provided by FFF platforms as an additional detection tool. This represents a major breakthrough in the field of milk analysis since fat content and processing method are the two major aspects of milk subject to fraud, and literature is lacking simple and inexpensive untargeted methods especially to investigate the latter (Qin et al. 2022). Our semi-automatic method (sample pretreatment needed, but automated analyses) allowed to distinguish a milk sample basing on fat and thermal treatment with a single, short (<1h) untargeted analysis requiring very minimal pre-treatment. On the contrary, at present, milk fat determination is evaluated with a non-automatic low throughput method (Gerber Method). Additionally, to the best of our knowledge no method exists to rapidly and simply discriminate UHT and non-UHT milk samples. Last, this methodology can be promptly translated to different food matrices to enable both performing and sustainable food analysis.

2. Materials and methods

2.1. Samples and chemicals

Sodium Chloride (NaCl, Cat. No. S9888, ACS reagent, ≥99.0%), Calcium Chloride (CaCl2, Cat. No. 223506, ACS reagent, ≥99%), Bovine Serum Albumin (A7030, ≥98%), β-Lactoglobulin from bovine milk (No. L3908, ≥90%) β-Casein from bovine milk (No. C6905, ≥98%), Casein from bovine milk (No. C7078), Ethylendiaminetetraacetic acid disodium salt dihydrate (No. E4884, ACS reagent, 99.0–101.0%) were obtained by Sigma Aldrich. Sodium Azide (NaN3, No. 71289, ≥99.5%) was obtained by Fluka.

Forty-seven bovine milk samples were purchased after their distribution in supermarket chains for this study. Samples differ from each other based on fat content (whole, semi skimmed, skimmed), thermal treatment (UHT, NON-UHT), brand and manufacturing plant. We would like to stress out the difference between brand and manufacturing plant

since it is common for manufacturers (i.e. companies processing milk) to produce milk for more than one brand which then retails it under their own name. For detailed information on each milk sample see the Supporting Information (SI) section, Table S1.

2.2. AF4 analysis

The AF4-UV-FLD-MALS analyses were performed by using a 1100 Series HPLC system (Agilent Technologies, Palo Alto, CA), connected to a module to control AF4 flow rates and operations (Eclipse 3, Wyatt Technology Europe, Dernbach, Germany). The ChemStation version B.04.02 (Agilent Technologies, Santa Clara, CA, USA) data system for Agilent instrumentation was used to set and control the instrumentation and method parameters.

The AF4 channel was 152 mm long, 16 mm wide, and 350 μm thick. The membrane was made by regenerated cellulose with 5 kDa cut-off (Microdyn-Nadir, Wiesbaden, Germany). Carrier solutions were degassed using an online vacuum degasser (Agilent, 1100 series, Agilent Technologies). Online detection of the eluted species was performed with a series of detectors here reported based on their order of coupling: an Agilent 1100 DAD UV/Vis spectrophotometer, a MALS detector operating at 658 nm wavelength model DAWN HELEOS (Wyatt Technology Corporation, Santa Barbara, CA, USA), a differential refractometer model Optilab rEX (Wyatt Technology Corporation, Santa Barbara, CA, USA), and an Agilent 1200 spectrofluorometer working using a 280 nm absorption wavelength and a 340 nm emission wavelength. Data collection was performed by the software ChemStation version B.04.02 and ASTRA® software version 6.1.7 (Wyatt Technology Corporation, Santa Barbara, CA, USA).

The aqueous carrier composition used (45 mM NaCl, 10 mM $CaCl_2$, 0.02%NaN_3, pH = 6.7) was chosen to have an ionic strength, pH, and calcium concentration similar to that of milk serum to avoid any changes in the size distribution during the analysis (Gaucheron 2005). The fat content of the milk samples was partially reduced, following a procedure already reported in literature (Lie-Piang et al. 2021), by sample centrifugation at 2000 g for 20 min at room temperature in a Beckman Coulter Avanti J-E-centrifuge (Brea, California, U.S.A) using a J-17 rotor to avoid blocking the capillary tubes of the AF4 system. To reduce sample viscosity, the centrifugated sample was diluted 100 folds with mobile phase immediately before AF4 analysis. Compared to analysis of undiluted sample, separation profiles resulted identical while reproducibility improved due to the use of an injection volume farther from the injector lower limit. The detector flow rate was 0.60 mL min^{-1}, the injection flow rate was 0.20 mL min^{-1} and the Focus-injection step was 5 min long with a 1.0 mL min^{-1} crossflow rate. During the elution step of the separation method, the crossflow decayed exponentially from 1.0 mL min^{-1} to 0.12 mL min^{-1} over 12 min (Slope = 3) and remained constant for 35 min before field release. For each analysis the injection volume was 50 μL. Each analysis was repeated in triplicate. Gyration radius (RG) was calculated using 0.1850 mL g-1 as dn/dc. Theorical Rh calculations were performed with the ISIS simulation software version 1.2.0 (Superon).

The greenness of the developed method was evaluated based on the AGREE software (Pena-Pereira, Wojnowski, and Tobiszewski 2020) which calculates a coefficient ranging from 0.00 to 1.00 based on increasing greenness.

2.3. Chemometric analyses

Principal components analysis (PCA) was used as explorative analysis for multivariate data, to understand potential in grouping and find outliers (Bro and Smilde 2014). Through a linear combination of the original variables, PCA converts them into new variables, called principal components (PCs), orthogonal to each other (Arrizabalaga-Larrañaga et al. 2021). The output yielded two plots: loadings plot, to evaluate variables role in objects discrimination, and scores plot, to evaluate samples clustering based on PCs.

Classification analyses were carried out by linear discriminant analysis (LDA) and partial-least squares discriminant analysis (PLS-DA). LDA allows the maximization of class separability in the model. A new object projected onto the model can be assigned to one of the known classes. The plot showing the result of the object's projection is called discriminant plot. PLS-DA, instead, is based on PLS regression. In PLS-DA, the response variable is not a numeric one, as in PLS, but, by splitting a category variable, two or more dummy variables (based on the number of classes) are obtained, whose values are 1 when an object pertains to the corresponding class or 0 otherwise. Therefore, the response for each object is a vector of 0 and 1 indicating the pertaining class.

For both classification analyses, a cross-validation (CV) method was applied to evaluate the performances of the models.

For all computations, the R software v.4.1.0 (R Core Team, Vienna, Austria) was used. The AF4 data used as starting variables to generate the PCs consisted in specifically selected portions of the full profiles obtained by the DAD and FLD detector and the peak areas calculated from said profiles. Before analysis, the AF4 profiles have been mean centered. This process consists of subtracting the mean value from each variable data; the resulting values are adjusted to offset the gap between low and high values originated by different separative injections. In the elaborations performed the profiles obtained by different injections of the sample have been considered as different objects in the PCA/LDA matrices.

3. Results and discussion

3.1. Fractograms interpretation

All FFF analyses were conducted with an optimized separation method described in the Experimental Section. Recovery was ≥ 94% for all samples and standards. Repeatability and reproducibility were assessed both on retention times and on signal intensity by performing three independent replicates (both intra- and inter-day) for each milk sample and protein standard (BSA, β-Lactoglobulin) used to develop the methods. The profiles exhibited a maximum of 0.5% and 1% deviation in terms of retention time and signal intensity, respectively. LOD (three-sigma) and LOQ (ten-sigma) for both whey proteins and caseins were assessed. For BSA, this was done via injection of standard solutions.. Injections of standard caseins solutions were also performed, but they exhibited bad recovery due to difficulties in their solubilization in the carrier solution without the presence of the other milk components. The corresponding LoD and LoQ values ere then extrapolated from signal/noise ratios of casein peak from a non-UHT milk profile of known casein content. The quantitation signals chosen correspond to the 280 nm absorption and the intrinsic fluorescence of proteins. Whey protein quantitation had a LOD and LOQ of 0.3/0.6 μg and 0.1/0.2 μg from absorption and fluorescence signals, respectively. Casein's LOD and LOQ resulted to be slightly higher with both detectors, and were calculated as 0.7/0.14 and 0.4/0.8 μg respectively for absorption and fluorescence. Overall, the developed method adhered to the standards that a validated method must satisfy according to the harmonized guideline ICHQ2R1 and ISO/TS 21,362 (Mildner et al. 2021; ISO 2018) and can be considered as a standardized method. Furthermore, the calculated coefficient obtained from the AGREE software for greenness assessment of analytical methods was 0.67, a positive value: for example, an HPLC-UV analysis with simple pretreatment would carry a coefficient of about 0.4. Inputs and outputs of the calculation are reported in Fig. S1.

To optimize the statistical results, it is important to pay attention when defining the appropriate set of variables to discriminate the objects of interest. This aspect becomes extremely important while working with sets of data such as fractograms since they may contain a lot of non-informative variables. In this case study, all the fractogram points associated to the Focus-Inject experimental phase are not informative

since they, by default, are associated to flat signals for all the samples (no analyte is reaching the detectors). Within this context, at least an approximate identification of the fractogram peaks is helpful to find the most informative fractogram regions to subject to multivariate analysis.

In the initial part of the work, the focus was mainly on FLD fractograms since the signal provided is protein specific and less affected by pressure and flows fluctuations, which may alter PCA results. Fig. 1A reports two representative FLD signals of UHT and NON-UHT milk samples; the first shows three peaks (namely 1, 2, and 3), while the second just two (1, 2′). The overlap of the FLD and UV signals for the UHT sample (Fig. 1B) shows a different relative emission and absorption intensity between Peaks 2 and 3: if one considers the intensity ratio in UV absorption, peak 3 has an intensity that is four or five times higher than peak 2, where the two values are similar for what concerns intensity of emission. This means that peak 2 and 3 have a different behavior, with peak 2 being richer in emitting species, and therefore they differ in composition.

Peak 1 is the only one that can be found in both fractograms. By comparing its elution time with the ones of standard whey proteins resuspended in mobile phase (Fig. 1C), Peak 1 was identified as associated to whey proteins. Calculated values of Rh and RG for Peak 2, 2′ and 3 were found coherent with the ones typical of caseins (Fig. 1D).

Peak 2 is observed only for UHT milks: consequently, it is safe to assume that it is originated by the process itself. Based on its size (R_h ~ 16 nm), a first identification hypothesis would be that of caseins smaller than the average value (R_h ~ 75 nm). However, Fig. 1B highlighted a difference in composition between Peak 2 and 3. Therefore it is likely that Peak 2 is at least partially composed of whey protein aggregates generated by the thermal treatment. This hypothesis is further supported by the fact that the area/intensity of Peak 1 (free whey protein) is always higher for no UHT milks compared to UHT ones (Fig. 1A).

To further investigate on this topic, a series of other experiments were performed.

Fig. 2 reports the result of the injection of the same volume of the UHT milk sample spiked with an amount of Casein from bovine milk corresponding to the 100% of its theoretical content. While we observed basically no rise in Peak 1, we observed a significant rise in Peak 2 and 3 confirming that smaller caseins may elute in correspondence of Peak 2. Following what was reported in another work (Lie-Piang et al. 2021) we also spiked another aliquot of the same UHT sample with a 50 mM EDTA solution. Since caseins stability is affected by the presence of Ca^{2+}, the chelating agent should cause precipitation of the caseins (thus disappearance of the corresponding signal) by calcium sequestration. The

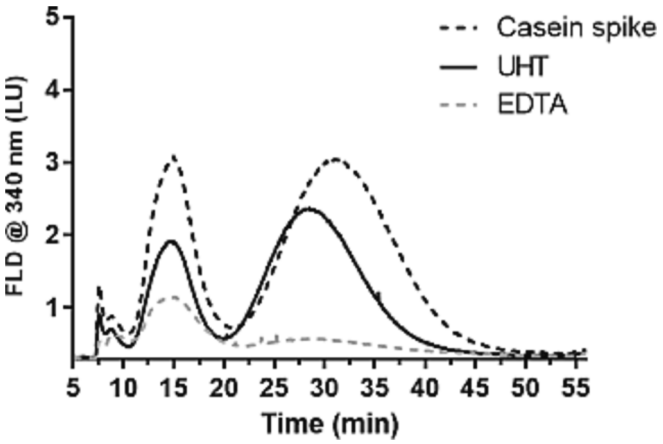


Fig. 2. Overlay of the FLD fractograms of a UHT milk sample and of the samples resulting from its corresponding spike with a 100% theorical amount of caseins or with an 50mM EDTA solution.

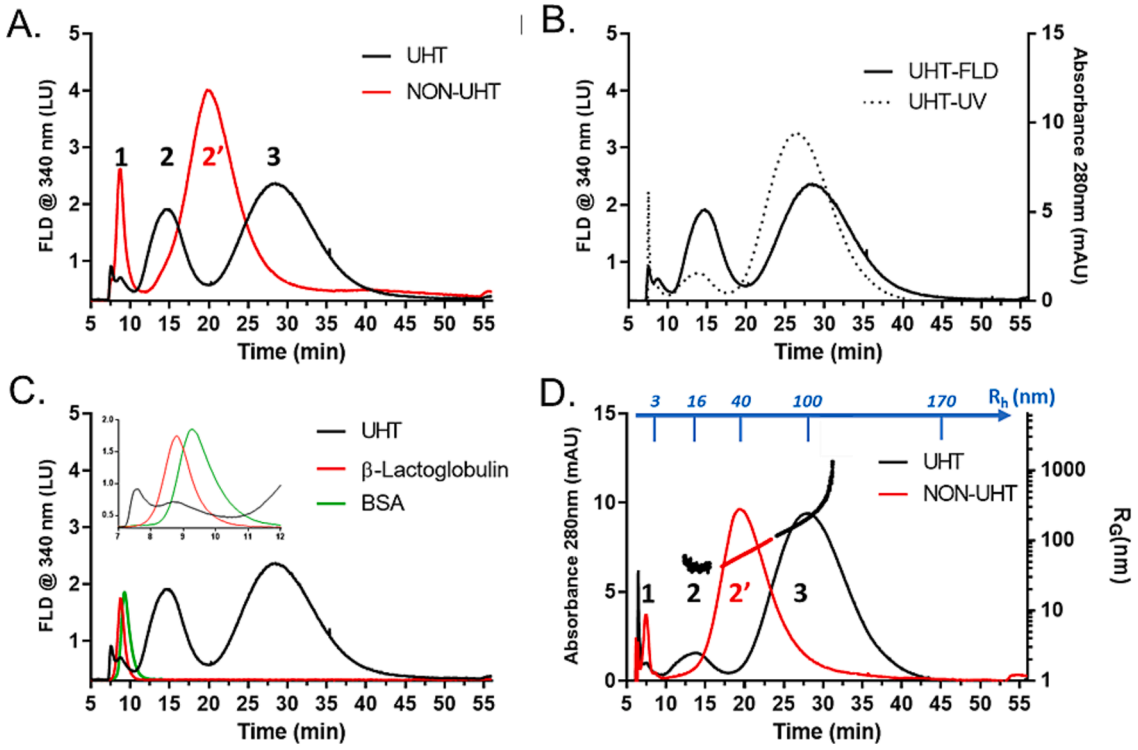


Fig. 1. A). Representative FLD profiles of UHT and NON-UHT milk samples. B). Overlap between the UV (280 nm) and FLD fractograms of the same UHT milk sample. C) Comparison between the retention times of two standard proteins with the ones observed for a UHT milk sample. Inset: zoom of the fractograms overlay to better visualize the differences in retention times for the peaks eluting between 7 and 12 min. D) Representative UV (280 nm) profiles of UHT and NON-UHT milk samples overlapped with the corresponding RG values calculated from the LS signal (logarithmic scale). Rh values calculated trough separative theory are also reported.

results of the experiment are still reported in Fig. 2; compared to the unspiked signals a loss of 37% and of 82% (integration area) respectively for Peak 2 and 3 were observed from the EDTA-spiked sample fractogram. This result is consistent with the simultaneous presence of caseins and protein aggregates in correspondence of Peak 2.

Further experiments (not shown) were carried out spiking the same UHT milk sample with "β-Casein from bovine milk" (No. C6905, ≥98%, Sigma Aldrich) which however resulted in the precipitation of the colloidal content of the sample during the centrifugation phase. It appears that rising in β-casein content increases the system reticulation thus promoting precipitation.

Based on these results we distinguished 3 main portions of the FLD fractograms namely X (7.75–10 min), Y (10–17 min), and Z (17–56 min) carrying different chemical information (Fig. 3A). In particular, X was associated to the whey protein content while Z to the casein content. Portion Y, instead, contains the information carried by the peak associated solely to the UHT thermal treatment. Its content is associated both to whey protein aggregates and to smaller or broken casein micelles. As already mentioned, working with the highly stable and protein specific FLD signal possesses a series of advantages. However, it is not possible to discriminate the differences in composition of non-protein-based components such as residuals fat micelles. Since our sample set is comprised of whole, semi skimmed, and skimmed milk samples a fat-based discrimination criterion may be helpful.

Within this context, a critical analysis of the less specific 280 nm UV profiles allowed to distinguish a fourth discriminatory area (Fig. 3B).

3.2. Principal component analysis

Our approach prior to every PCA analysis of the milk colloidal fingerprint can be summarized in 3 main points: (a) Definition of the set of samples to be discriminated; (b) definition of the best samples property (also called filter) to discriminate the set of samples (i.e., Thermal Treatment) and the a priori classes on which the sample can be classified based on such property (i.e. UHT and NON-UHT). The a priori classes for each sample properties are reported in Table S2; (c) definition of an optimized dataset to be analyzed with PCA. The dataset definition aimed at selecting only the portion of the colloidal fingerprints (the starting variables) of the sample which have more discriminatory power based on the samples property of study.

At first, all 47 samples have been analyzed to group according to thermal treatment. Since proteins are the milk components most affected by such treatments, the highly protein-specific FLD signal has been exploited for the study. Five different PCA have been computed using the full profile of different portions of the FLD fractograms based on the labelling reported in Fig. 2: X, Y, Z, X + Y, X + Y + Z. The best results were obtained by PCA-X + Y (Fig. 4). The results denote an excellent discrimination of samples on the basis of heat treatment (regardless of the other properties). Hotelling T2-based computation was applied to the scores plot to calculate the ellipses, representing the 95% confidence intervals for each class.

PC1 carrying 77.1% of the explained variance seems to be associated to the intensity of the thermal treatments since UHT samples are mainly characterized by positive PC1 values while NON-UHT samples are characterized by negative PC1 values. Moreover, considering the loadings (Fig. 4B), UHT samples are mainly characterized by the X peak, associated whey proteins, while the NON-UHT show a higher contribution of the Y-peak species. The fact that the best PCA result was obtained considering the X + Y FDL fingerprint is also coherent with the FFF results and literature: the X and Y portion of the fractograms are associated to whey proteins and their aggregates, the milk components most affected by thermal treatments. The PCA scores plot was also visualized based on the other category variables (not shown): fat content, manufacturing plant, and brand; none of these highlighted particular and clear clustering, indicating thermal treatment as the most discriminating variable for this dataset.

Since a first clustering was obtained, the data analysis focused on the two separated milk sets (UHT and NON-UHT milk sets) obtained from the first PCA analysis. In this case, fat content was chosen as the second discriminating filter.

The same approach (using the FLD signal) did not show any interesting grouping for such datasets. This is justified since the FLD colloidal fingerprint should not carry information on the fat content. Based on what was already discussed, our elaboration then focused on section W of the UV 280-nm absorbance profile (Fig. 3B). For the UHT dataset two PCA were performed. The first one (PCA-W), similarly to the previous approaches, used as starting variables all the absorbance values of section W of the fractograms for each sample (results not shown). The second one (PCA-W-A), which overall provided the best results, used as starting variables the areas of 8 segments obtained from the division of W in 4-min intervals. The first 7 areas carry information on the abundance of colloidal subpopulations with Rh comprised between 65 and 225 nm, while the last one is associated to the remaining species which are retained by the systems and elute during field release (see Fig. S2).

The results for the UHT set highlight a good discrimination of samples based on the fat content (Fig. 5A). Based on how the three subclasses of samples (Skimmed, Semi Skimmed and Whole milk) distribute themselves along the PC1 axis, it is possible to observe that PC1 has the

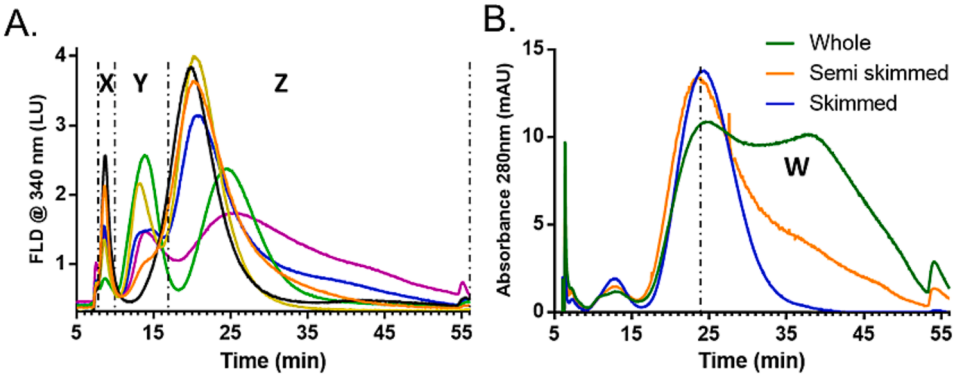


Fig. 3. A) Overlay of FLD fractograms of different milk samples. The three portions (X, Y, Z) of the signals carrying different chemical information are highlighted as X, Y, Z. B) Overlay of UV (280 nm) fractograms representative of a whole, a semi skimmed, and a skimmed milk sample. The portion (W) which presumably carries information on the residual fat content of the samples is highlighted. In particular, after 24 min significant signal intensity differences between samples can be observed as well as the formation of an additional peak after 28 min. The intensity of the latter is increasingly higher for whole milk samples compared to semi skimmed ones, moreover it is mostly flat for the defatted samples. Based on what already discussed, we supposed that this portion of the UV signals, namely W (24–56 min), may carry information on the residual fat content represented by small micelles which remained/formed in the systems after the initial centrifugation of the samples.

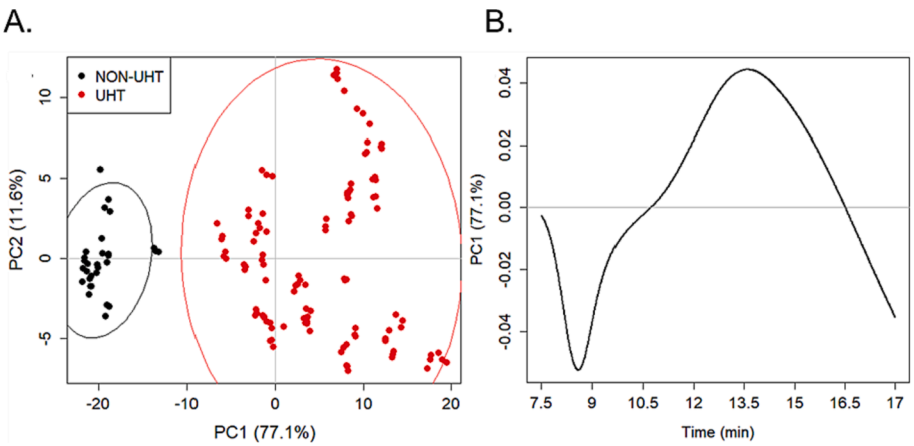


Fig. 4. A) Scores plot and loadings plot for PCA-X + Y performed on the complete milk dataset (47 samples); Circles represent the 95% confidence intervals, based on Hotelling T2 computation, of each thermal-treatment class. B) Loadings plot for PC1.

same direction of the fat-content variable. Low values of PC1 are thus associated to high fat content while high values are associated to small or none fat content.

The fact that better discrimination of the samples was obtained using areas as starting variables instead of an absorbance profile can also be explained: fat content is a property referred to an amount and integration areas intrinsically carry information of the amount of species. Based on the clustering provided by the PCA-W-A on the simplified UHT and NON-UHT datasets it was possible to further split the datasets in 6 small datasets differing for thermal treatment and/or fat content.

On such datasets, a series of PCA analyses was performed to further improve the clustering of the samples according to the leftover discriminating variables still unexplored (Manufacturing plant and Brand). Overall, the best results were obtained using Manufacturing plant as filter and the X + Y portion of the FLD fractograms as starting variables. Since some of these final datasets contain a very small number of samples, the discussion will be focused on the two with the biggest number of samples: UHT semi skimmed (UHT-SS, 29 samples, 3 replicates) and NON-UHT semi skimmed (NON-UHT-SS, 4 samples, 3 replicates).

Fig. 5 C,E represent the scores plot deriving from the PCA performed on the dataset of UHT-SS milk samples using as filters Manufacturing plant and Brand. The circled regions in both plots highlight how samples commercialized under different brands, but produced in the same manufacturing plans, are grouped together.

This is particularly evident when looking at the samples produced in the plant P3 (circled in Fig. 5 C,E). In this case, all samples are located only in a single quadrant of the plot and clearly appear as a uniform group when labeled as Manufacturing plant instead of Brand.

The same elaboration was applied to the NON-UHT dataset (Fig. 5B) stemming analogous results. A discrimination based on fat content could also be observed along PC1 (carrying 70% of explained variance); also in this case, higher values of PC1 are associated with low fat content.

Moreover, it is possible to find macro groups of samples, not necessarily uniform in terms of plant, but clearly distanced from each other. A possible interpretation to these results is based on the hypothesis that different plants carry out different UHT treatments (in terms of temperature, process time and heat conduction system) which may have an impact on the protein profile. This hypothesis is also supported by the fact that the PCA providing the best grouping uses as initial dataset the X + Y FLD portion of the colloidal fingerprint which is the most affected by thermal treatments. Based on this interpretation, it is possible to hypothesize that plants whose objects are near each other may use similar UHT treatment. To obtain confirmation of such results the details of the UHT procedure of each plant should be known; unfortunately, it was not possible to obtain this information.

The same statistical analysis on the NON-UHT-SS subset provided results similar to UHT-SS. A good clustering based on manufacturing plant can be observed (Fig. 5D), that is lost using the Brand filter (Fig. 5F). Since no UHT treatment was performed for such samples, the rationale of the grouping can be explained with differences in pasteurization treatment. Overall, these promising results suggest the possibility to develop models able to also distinguish the manufacturing plant in which commercial bovine milk is produced.

3.3. Classification analysis

After exploring the datasets using PCA, classifications (Fig. 6) were also performed on the two biggest datasets previously analyzed, in order to confirm the hypotheses formulated and evaluate the possible applicability of the models. PLS-DA was applied to the full dataset and the X + Y peaks (as in Fig. 4) using the thermal-treatment class as independent variable. LDA was instead used on the UHT-milk dataset using the 8 areas previously described as descriptors and the fat-content class for the discrimination (as in Fig. 5A).

Fig. 6A shows the scores plot obtained from PLS-DA on thermal-treatment dataset. The red dots, associated to UHT milk objects, are placed at positive values of Factor 1, while NON-UHT milk objects (black dots) can be found in the negative region of Factor 1. PLS-DA cross-validation confirms the good discrimination between the two classes observed in the scores plot. Most of UHT and NON-UHT samples (45 in total) were, indeed, classified in the correct class with recalculated Y higher than 0.8 for the corresponding class.

Two samples (the ones outside the 95% confidence interval, highlighted by blue circles in Fig. 6A), instead, had a recalculated Y close to 0.5 for both classes, indicating that it was not possible to assign them to any class. Overall, the model showed good predictive abilities, having Non Error Rate (NER) of 95.7%. Fig. 6B shows, instead, the discriminant plot obtained by processing the UHT dataset to discriminate samples based on fat content by LDA. In this case only one object (highlighted by the blue circle) is badly classified and the NER% reaches 99.1%, confirming the good classification obtained by the model.

4. Conclusions

For the first time, a simultaneous discrimination of different milk samples based on the colloidal fingerprints obtained through an AF4 multidetection platform has been reported. A dataset was created able to achieve different discriminations within a single run, such as thermal treatment, fat content and manufacturing plant; for the first time, also milk sub-typologies could be identified. The mathematical models created based on such dataset could help to identify product fraudulent

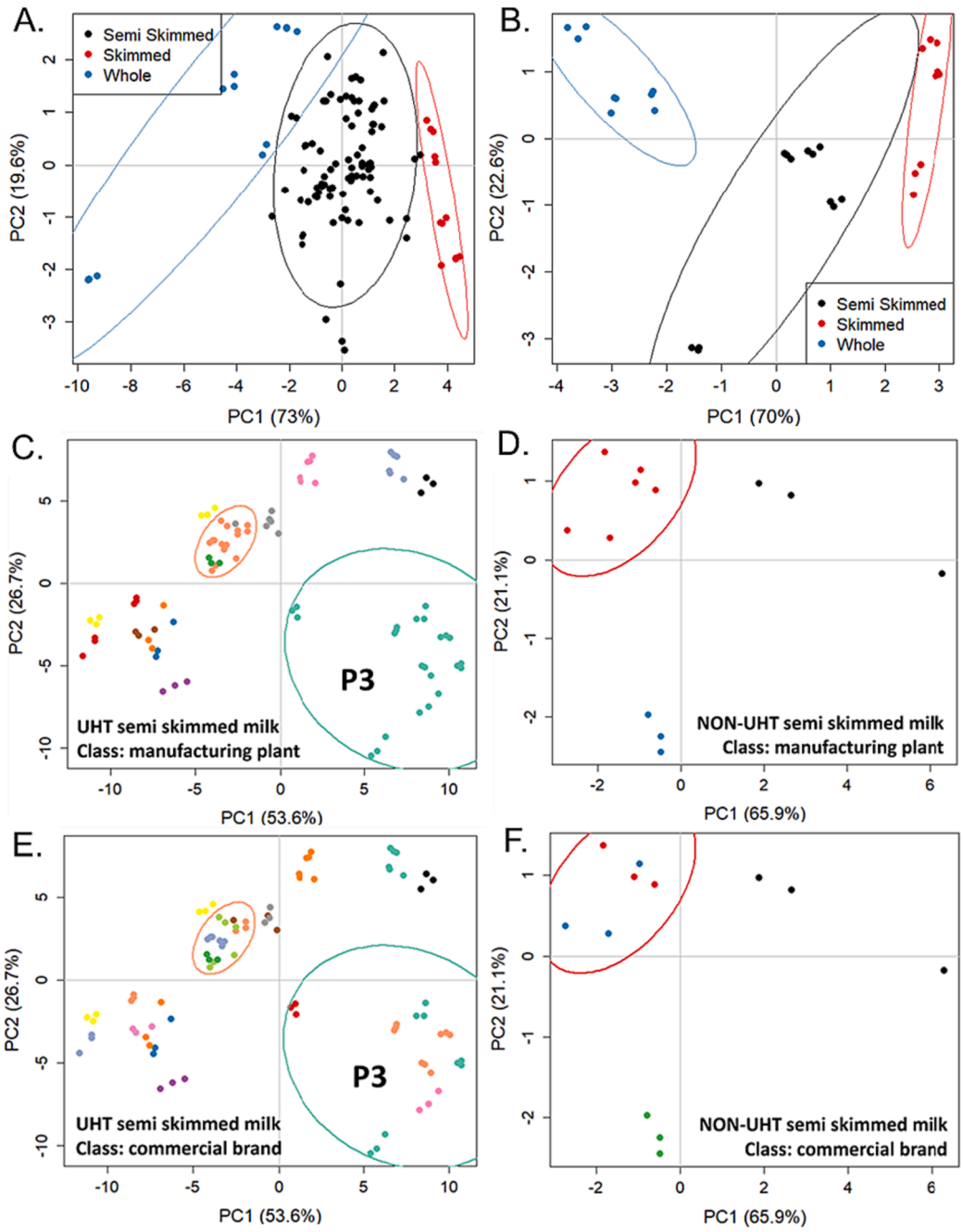


Fig. 5. Scores plot for PCA-W-A (A, B) and PCA-X + Y (C-F) datasets of UHT and NON-UHT milk. A,B: scores plot of UHT (A) and non-UHT (B) dataset and discrimination between whole, semi skimmed, skimmed. Circles represent the 95% confidence intervals, based on Hotelling T2 computation, of each fat-content class. C-F Scores plot for on the UHT (C,E) and NON-UHT (D,F) semi skimmed milk sets. The plot differs based on the category variable used, points with the same color corresponds to objects with the same: Manufacturing plant (C-D) or brand (E-F). Hotelling T2 ellipses highlight areas in which clustering between different brands is explained by the same manufacturing plant.

labelling (ex. UHT-milk labeled as fresh). However, since the models were created starting from commercial samples, we cannot exclude the possibility of some samples being subject to fraud. Within this context our future aim will be development of an analogous approach on samples of verified origin and treatment. The developed AF4 separation method is fast, requires small amounts of sample (0.5 μL of milk per injection), and minimal sample preparation, with no chemical treatment and in saline conditions. The multivariate methodologies used allowed a rapid visualization of the results and can represent a starting point to improve the overall processes of quality control and fraud identification in the food industry. Colloidal fingerprinting was proposed as a performing, green and fast approach to source data for sample grouping and chemometric analysis. The approach can easily be translated to other food samples presenting a colloidal content (ex. wine, tomato sauce, beer, vinegar) and possibly to the detection of other kinds of milk food frauds after the definition of a proper dataset and offers a breakthrough in reducing analysis impact and time while retaining high quality and reliable results. Finally, this study further highlights the necessity of increasing the implementation of multivariate analysis to the results directly provided by FFF platforms.

5. Author statement

The authors declare that the work described has not been published

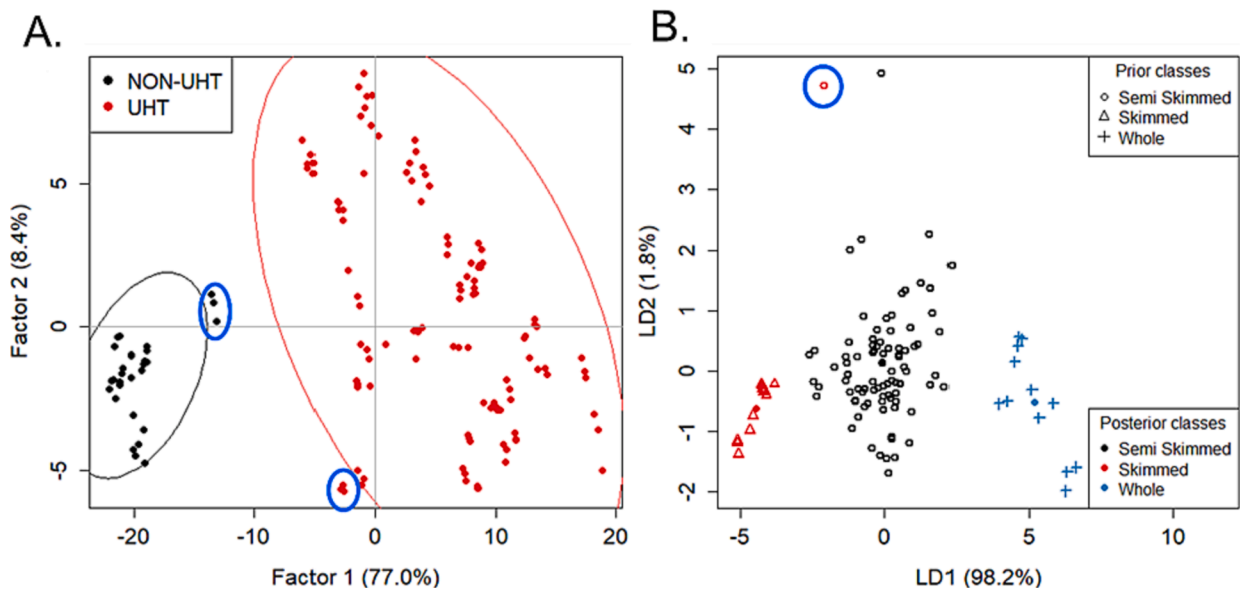


Fig. 6. A) Scores plot obtained through PLS-DA on the complete sample set using the portion X + Y of the FLD fractograms as variables. B) Discriminant plot obtained through LDA on the UHT milk sample set, using as variables the 7 areas obtained through integration of the 4 min intervals in which the W portion on the 280 nm UV fractograms were divided. Blue circles indicate the objects wrongly classified by the models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

previously, it is not under consideration for publication elsewhere, its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright holder.

CRediT authorship contribution statement

Stefano Giordani: Conceptualization, Data curation, Writing – original draft, Investigation. **Nicholas Kassouf:** Conceptualization, Data curation, Writing – original draft, Investigation. **Alessandro Zappi:** Conceptualization, Data curation, Writing – original draft. **Andrea Zattoni:** Funding acquisition, Writing – review & editing. **Barbara Roda:** Funding acquisition, Writing – review & editing. **Dora Melucci:** Funding acquisition, Writing – review & editing. **Valentina Marassi:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Valentina Marassi, Barbara Roda, Andrea Zattoni are associates of the spinoff company byFlow srl. The company mission includes know-how transfer, development, and application of novel technologies and methodologies for the analysis and characterization of samples of nano-biotechnological interest

Data availability

Data will be made available on request.

Appendix A. Supplementary material

Sample list, a priori classes, greenness evaluation and W-A dataset.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2023.138206.

References

Arrizabalaga-Larrañaga, A., Epigmenio-Chamú, S., Santos, F. J., & Moyano, E. (2021). Determination of banned dyes in red spices by ultra-high-performance liquid chromatography-atmospheric pressure ionization-tandem mass spectrometry. *Analytica Chimica Acta, 1164*, Article 338519.

Austin, M. A., Rodriguez, B. L., McKnight, B., McNeely, M. J., Edwards, K. L., David Curb, J., & Sharp, D. S. (2000). Low-density lipoprotein particle size, triglycerides, and high-density lipoprotein cholesterol as risk factors for coronary heart disease in older Japanese-American men. *The American journal of cardiology, 86*, 412–446.

Azad, T., & Ahmed, S. (2016). Common milk adulteration and their detection techniques. *International Journal of Food Contamination, 3*, 1–9.

Ballin, N. Z. (2010). Authentication of meat and meat products. *Meat science, 86*, 577–587.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods, 6*, 2812–2831.

Caputo, F., Dora Mehn, J. D., Clogston, M. R., Prina-Mello, A., Borgos, S. E., Gioria, S., & Calzolai, L. (2021). 'Asymmetric-flow field-flow fractionation for measuring particle size, drug loading and (in) stability of nanopharmaceuticals. The joint view of European Union Nanomedicine Characterization Laboratory and National Cancer Institute-Nanotechnology Characterization Laboratory. *Journal of Chromatography A, 1635*, Article 461767.

Correia, M., & Loeschner, K. (2018). Detection of nanoplastics in food by asymmetric flow field-flow fractionation coupled to multi-angle light scattering: Possibilities, challenges and analytical limitations. *Analytical and Bioanalytical Chemistry, 410*, 5603–5615.

Cubero-Leon, E., Peñalver, R., & Maquet, A. (2014). Review on metabolomics for food authentication. *Food Research International, 60*, 95–107.

De Kruif, C. G., Holt, C. (2003). Casein micelle structure, functions and interactions. In: *Advanced dairy chemistry—1 proteins: part a/part b* (Springer).

Esteki, M., Shahsavari, Z., & Simal-Gandara, J. (2020). Gas chromatographic fingerprinting coupled to chemometrics for food authentication. *Food Reviews International, 36*, 384–427.

Fox, Patrick F, Paul LH Mcsweeney, and LH Paul. 1998. 'Dairy chemistry and biochemistry'.

Gaucheron, F. (2005). The minerals of milk. *Reproduction Nutrition Development, 45*, 473–483.

Giordani, S., Marassi, V., Placci, A., Zattoni, A., Roda, B., & Reschiglian, P. (2023). Field-Flow Fractionation in Molecular Biology and Biotechnology. *Molecules, 28*, 6201.

Guyomarc'h, Fanny, Frédéric Violleau, Olivier Surel, and Marie-Hélène Famelart. 2010. 'Characterization of heat-induced changes in skim milk using asymmetrical flow field-flow fractionation coupled with multiangle laser light scattering', *Journal of Agricultural and Food Chemistry*, 58: 12592-601.

Hennetier, Marie, Amira Halabi, Agnès Burel, Amélie Deglaire, Thomas Croguennec, Audrey Romelard, Alain Baniel, and Frédéric Violleau. 2020. "FIELD-FLOW FRACTIONATION TECHNIQUES FOR MILK PROTEIN CHARACTERIZATION." In *20th International Symposium on Field-and Flow-Based Separations*, np.

Hristov, P., Mitkov, I., Sirakova, D., Mehandgiiski, I., & Radoslavov, G. (2016). Measurement of casein micelle size in raw dairy cattle milk by dynamic light scattering. *Milk Proteins—From Structure to Biological Properties and Health Aspects*, 19–32.

Huppertz, T, T Uniacke-Lowe, and AL Kelly. 2020. 'Physical chemistry of milk fat globules', *Advanced Dairy Chemistry, Volume 2: Lipids*: 133-67.

ISO, TS. 2018. '21362: 2018 (E)', Nanotechnologies-Analysis of nano-Objects using asymmetrical-flow and centrifugal field flow fractionation.

Jensen, R. G. (2002). The composition of bovine milk lipids: January 1995 to December 2000. *Journal of dairy science, 85*, 295–350.

Kim, S. H., Yang, J. S., Lee, J. C., Lee, J.-Y., Lee, J.-Y., Kim, E., & Moon, M. H. (2018). Lipidomic alterations in lipoproteins of patients with mild cognitive impairment and Alzheimer's disease by asymmetrical flow field-flow fractionation and nanoflow ultrahigh performance liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A, 1568*, 91–100.

Krebs, G., Müller, M., Becker, T., & Gastl, M. (2019). Characterization of the macromolecular and sensory profile of non-alcoholic beers produced with various methods. *Food Research International, 116*, 508–517.

Krishna, T. C., Najda, A., Bains, A., Tosif, M. M., Papliński, R., Kapłan, M., & Chawla, P. (2021). Influence of ultra-heat treatment on properties of milk proteins. *Polymers, 13*, 3164.

Law, General Food. 2002. 'Regulation (EC) No. 178/2002 of the European Parliament and of the Council of 28 January 2002, laying down the general principles and requirements of food law, establishing the European Food Safety Authority, and laying down procedures in matters of food safety', *OJ L*, 31.

Lie-Piang, A., Leeman, M., Castro, A., Börjesson, E., & Nilsson, L. (2021). Investigating the effect of powder manufacturing and reconstitution on casein micelles using asymmetric flow field-flow fractionation (AF4) and transmission electron microscopy. *Food Research International, 139*, Article 109939.

Madureira, A. R., Pereira, C. I., Gomes, A. MP., Pintado, M. E., & Xavier Malcata, F. (2007). Bovine whey proteins–Overview on their main biological properties. *Food Research International, 40*, 1197–1211.

Mangal, V., Stenzler, B. R., Poulain, A. J., & Guéguen, C. (2018). Aerobic and anaerobic bacterial mercury uptake is driven by algal organic matter composition and molecular weight. *Environmental Science & Technology, 53*, 157–165.

Marassi, V., Macis, M., Giordani, S., Ferrazzano, L., Tolomelli, A., Roda, B., … Cabri, W. (2022). Application of Af4-Multidetection to Liraglutide in Its Formulation: Preserving and Representing Native Aggregation. *Molecules, 27*, 5485.

Marassi, V., Marangon, M., Zattoni, A., Vincenzi, S., Versari, A., Reschiglian, P., … Curioni, A. (2021). Characterization of red wine native colloids by asymmetrical flow field-flow fractionation with online multidetection. *Food Hydrocolloids, 110*, Article 106204.

Mildner, R., Hak, S., Parot, J., Astrid Hyldbakk, S. E., Borgos, D. S., Johann, C., & Caputo, F. (2021). Improved multidetector asymmetrical-flow field-flow fractionation method for particle sizing and concentration measurements of lipid-based nanocarriers for RNA delivery. *European Journal of Pharmaceutics and Biopharmaceutics, 163*, 252–265.

Nascimento, C. F., Santos, P. M., Pereira-Filho, E. R., & Rocha, F. RP. (2017). Recent advances on determination of milk adulterants. *Food Chemistry, 221*, 1232–1244.

Nilsson, L. (2013). Separation and characterization of food macromolecules using field-flow fractionation: A review. *Food Hydrocolloids, 30*, 1–11.

Nogueira, M. H., Scudeler, L. A., Humblot, L., Doumert, B., Hennetier, M., Violleau, F., … Peixoto, P. P. S. (2023). Assessment of structures in phosphocaseinate dispersions by A4F, NMR and SAXS: The impact of demineralization and heat treatment on viscosity. *Food Hydrocolloids, 137*, Article 108366.

Pena-Pereira, F., Wojnowski, W., & Tobiszewski, M. (2020). AGREE—Analytical GREEnness Metric Approach and Software. *Analytical Chemistry, 92*, 10076–10082.

Plavchak, C. L., Smith, W. C., Bria, C. RM., Kim Ratanathanawongs, S., & Williams. (2021). New advances and applications in field-flow fractionation. *Annual Review of Analytical Chemistry, 14*, 257–279.

Qin, C., Lei Liu, Y.u., Wang, T. L., Zhu, M., Gan, B., Xie, J., … Chen, Y.i. (2022). Advancement of omics techniques for chemical profile analysis and authentication of milk. *Trends in Food Science & Technology, 127*, 114–128.

Velazquez-Dominguez, A., Hennetier, M., Abdallah, M., Hiolle, M., Violleau, F., Delaplace, G., & Peixoto, P. P. D. S. (2023). Influence of enzymatic cross-linking on the apparent viscosity and molecular characteristics of casein micelles at neutral and acidic pH. *Food Hydrocolloids, 139*, Article 108552.

Ventouri, I. K., Loeber, S., Somsen, G. W., Schoenmakers, P. J., & Astefanei, A. (2022). Field-flow fractionation for molecular-interaction studies of labile and complex systems: A critical review. *Analytica Chimica Acta, 1193*, Article 339396.

Wang, J., Giordani, S., Marassi, V., Roda, B., Reschiglian, P., & Zattoni, A. (2022). Quality control and purification of ready-to-use conjugated gold nanoparticles to ensure effectiveness in biosensing. *Frontiers in Sensors, 3*, 1087115.

Zappi, A., Marassi, V., Giordani, S., Kassouf, N., Roda, B., Zattoni, A., … Melucci, D. (2023). Extracting information and enhancing the quality of separation data: A review on chemometrics-assisted analysis of volatile, soluble and colloidal samples. *Chemosensors, 11*, 45.

Zappi, A., Marassi, V., Kassouf, N., Giordani, S., Pasqualucci, G., Garbini, D., … Melucci, D. (2022). A Green Analytical Method Combined with Chemometrics for Traceability of Tomato Sauce Based on Colloidal and Volatile Fingerprinting. *Molecules, 27*, 5507.

Zhang, H., & Lyden, D. (2019). Asymmetric-flow field-flow fractionation technology for exomere and small extracellular vesicle separation and characterization. *NatureProtocols, 14*, 1027–1053.