Article

Proteome speciation by mass spectrometry: characterization of composite protein mixtures in milk replacers

Marco Gaspari, Luca Chiesa, Annalisa Nicastri, Caterina Gabriele, Valeria Harper, Domenico Britti, Giovanni Cuda, and Antonio Procopio

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.6b02848 • Publication Date (Web): 28 Oct 2016

Proteome speciation by mass spectrometry: characterization of composite protein mixtures in milk replacers

Marco Gaspari,[a] Luca Chiesa,[b] Annalisa Nicastri,[a] Caterina Gabriele,[a] Valeria Harper,[d] Domenico Britti,[c] Giovanni Cuda,[a] Antonio Procopio[c]

a Research Center for Advanced Biochemistry and Molecular Biology, Department of Experimental and Clinical Medicine, ''Magna Græcia'' University of Catanzaro, Catanzaro, Italy

b Department of Veterinary Sciences and Public Health, University of Milan, Milan, Italy

c Department of Health Sciences, ''Magna Græcia'' University of Catanzaro, Catanzaro, Italy

d F.Q.C. Italia, Casalecchio di Reno, Italy

ABSTRACT: The ability of tandem mass spectrometry to determine the primary structure of proteolytic peptides, can be exploited to trace back the organisms from which the corresponding proteins were extracted. This information can be important when food products, such as protein powders, can be supplemented with lower-quality starting materials. In order to dissect the origin of proteinaceous material composing a given unknown mixture, a two-step database search strategy for bottom-up nanoLC-MS/MS data was implemented. A single nanoLC-MS/MS analysis was sufficient not only to determine the qualitative composition of the mixtures under examination, but also to assess the relative % composition of the various proteomes, if dedicated calibration curves were previously generated. The approach of two-step database search for qualitative analysis and proteome total ion current (pTIC) calculation for quantitative analysis was applied to several binary and ternary mixtures which mimic the composition of milk replacers typically used in calf feeding.

The application of mass spectrometry (MS) to food analysis is moving beyond the targeted determination of small molecules (generally exogenous, such as pesticides). MS-based "omics" approaches promise to allow the profiling of nearly all classes of analytes, such as proteins, lipids, glycans and metabolites in food matrixes with an unprecedented depth.[1–3] For what concerns protein analysis, mass spectrometry is well established as one of the key technologies for studying complex mixtures, such as entire "proteomes". In particular, the approach referred to as "bottom-up" proteomics has been widely adopted for the qualitative and quantitative analysis of proteomes and sub-proteomes since its introduction in the late nineties.[4–6] Bottom-up proteomics is based on the enzymatic digestion of a complex protein mixture and on the subsequent analysis of the generated protein fragments (peptides) by LC-MS/MS.[7] The introduction of mass spectrometers of increasing sensitivity and resolution has allowed the characterization of protein mixtures with broader proteome coverage. A recent achievement is the cataloguing of the entire set of genes expressed by a simple organism such as yeast in a single analysis.[8]

The identification of organisms based on the mass spectrometric analysis of their protein extracts has been widely used in microbiology.[9] MALDI-TOF has been adopted for the fast identification of pure bacterial isolates,[10] whereas LC-MS/MS has been successfully applied to multi-component mixtures.[11] Outside of the microbiology field, the possibility of identifying an organism by the profiling of its protein products, here named "proteome speciation" has been rarely explored. A few relevant examples of the application of mass spectrometry to proteome speciation in the field of food analysis are reported below.

Direct mass spectrometric analysis based on MALDI–TOF of protein digests was applied to the profiling of binary mixtures containing milk produced by various species (*Bos taurus*, *Ovis aries*, *Capra hircus*).[12,13] The authors reported successful detection of species-specific tryptic peptides down to a level of adulteration of 0.5-5%. Cordewener et al.[14] used LC-MS/MS to screen for adulteration in skimmed milk powder (SMP). In their proof-of principle study, they compared five SMP preparations of pure bovine origin to four preparations to which low amounts (5%) of plant proteins from either soy or pea were added. Tryptically digested proteins were analysed by LC-MS and principal component analysis (PCA) in order to discriminate between spiked SMPs and pure SMPs. Additional data-dependent LC-MS/MS analyses with inclusion lists, targeted at differential peptides, allowed the identification of plant-specific peptides present in adulterated SMP. In this work, no quantitative information on the extent of the adulteration could be extracted from the analysis. Besides, the approach relied on a complex workflow involving multiple LC-MS injections and the use of multivariate statistics, which usually demands for large training sets in order to gain statistical power.

Targeted LC-MS/MS analysis based on selected reaction monitoring (SRM) has been used for the detection of horse and pork meet in halal beef.[15] The method relied on the targeted, sensitive detection of characteristic marker peptides mainly

originating from haemoglobin, myoglobin and myosins in digested meat, using SRM. Species identification via specific peptide markers was also achieved in a recent SRM-based study.[16] In this case, myoglobin-derived marker peptides were assayed by triple quadrupole MS/MS in order to discriminate between beef, pork, horse and lamb meat. Though very sensitive (detection of adulterations at a level lower than 1% were reported in both publications), the SRM-based approach has some limitations. Being a targeted assay, it requires the development of a specific method (and, possibly, the purchase of isotopically labelled internal standards) for each adulteration to be screened. Thus, it lacks flexibility. Besides, targeted mass spectrometry methods based on SRM can be applied exclusively to the quantification of specific proteins, and not to the quantification of entire proteomes in protein mixtures generated by mixing protein extracts from different origin.

Most of the milk replacers used for veal calves show on their label incomplete indications regarding their composition. In particular the raw materials used in the formulation are listed by decreasing order of weight percentage without stating the quantities ("closed formula"). These labels also display the analytical levels of the different nutrients expressed in percentages. The overall protein content of the formulation, though, can be achieved either through the use of high quality proteins, such as dairy proteins (e.g., skimmed milk powder), or through the use of plant proteins (e.g., extracts from soy, pea, wheat, triticale). Dairy proteins are considered of higher quality than vegetable proteins, especially if used in the early stages of calf farming, because milk proteins are more digestible and allow better weight increase percentages.[17–19] The market value of formulations containing diary proteins is, therefore, often higher. Some feed companies, in order to increase the protein content of their products, make use of vegetable proteins in feed formulas, thus optimizing costs. Some large-scale distribution chains want to create "controlled diet" food-chains, i.e., food-chains in which the feed ration of the animals must follow certain requirements, such as, for instance, the use of milk replacers containing a minimum percentage of proteins and, specifically, a minimum content of dairy proteins. In this context, the development of analytical methods aimed at dissecting the proteome composition of a milk replacer formulation is of commercial interest, since it may help in selecting better-quality products and in detecting possible frauds.

In this work, a bottom-up LC-MS/MS proteomics approach was used to characterize, both qualitatively and quantitatively, protein mixtures resembling the composition of typical milk replacers. From a qualitative analysis point of view, the organisms from which the protein mixture was generated were traced back by a two-step data analysis strategy. From a quantitative analysis point of view, chromatographic ion signals from entire species-specific peptide populations were used for calculating % w/w proteome composition.

EXPERIMENTAL SECTION

Protein solubilisation and digestion. Protein extracts (Table 1) were kindly provided by the University of Milan. Protein powders were solubilized in 6 M urea, 100 mM tris buffer (pH 8) in order to obtain a protein concentration of 4.0 mg/mL. These mother solutions were either processed further as described below, or mixed in appropriate volumetric ratios in order to create 250 µL of either binary or ternary protein mixtures. Binary mixtures were used for generating calibration curves, and were prepared as follows: bovine/wheat, bovine/soybean, bovine/pea, soybean/wheat. Seven distinct binary mixtures were created for each couple: 95:5; 90:10; 80:20; 70:30; 60:40; 40:60; 20:80 (w/w of total proteins). Besides, two distinct ternary mixtures were prepared and processed in triplicates. The first ternary mixture consisted of a bovine/soybean/wheat mix 72/20/8 (w/w/w); the second mixture was composed of bovine/soybean/pea proteomes in a 80/10/10 (w/w/w) ratio. Protein solutions (250 µL) were subjected to reduction, alkylation and tryptic digestion as described below.

Proteins were reduced by adding 25 µL of aqueous 100 mM dithio-threitol (DTT) (1h at 37 °C with agitation) and subsequently alkylated by adding 30 µL of aqueous 200 mM iodoacetamide (1h at 37 °C with agitation in darkness); the alkylation reaction was quenched by adding an additional aliquot of 100 mM DTT (5 µL) and letting the reaction proceed for 20 min at 37 °C. Urea concentration was then reduced to 1.5 M by adding 690 µL of 20 mM tris buffer (pH 8). Finally, proteins were digested by 20 µg of proteomics-grade trypsin (VWR) (enzyme:substrate ratio of 1:50 w/w) overnight at 37 °C with agitation. Peptide mixtures were desalted by C_{18} StageTips[20] and subsequently injected for mass spectrometric analysis as described below.

Tryptic peptides (4 µg, corresponding to 4 µL of solution) were acidified to 0.5% TFA (v/v) and subsequently purified using StageTips C_{18} prepared using C_{18} disks (3M). StageTips were initially conditioned with: (i) 10 µL of a 1:1 mixture of 0.1% formic acid (v/v) and acetonitrile (solution E); (ii) 10 µL of 0.1% TFA (v/v). After loading the peptide mixture, StageTips were washed with: (i) 10 µL of 0.1% TFA (v/v); (ii) 10 µL of 0.1% formic acid (v/v) before being eluted by 8 µL of solution E. The eluate was diluted 10-fold in mobile phase A (see below). The resulting solution was analysed by nanoLC-MS/MS analysis.

Table 1. Protein extracts used for creating multi-species protein mixtures.

Sample	Protein content (%)
Milk powder	35.8
Wheat gluten	78.4
Soybean protein extract	65.9
Pea protein extract	22.8

Nano LC-MS/MS and database search. Chromatography was performed on an Easy LC 1000 nanoscale liquid chromatography (nanoLC) system (Thermo Fisher Scientific). The analytical nanoLC column was a pulled fused silica capillary, 75 µm i.d., in-house packed to a length of 12 cm with 3 µm C18 silica particles (Dr. Maisch GmbH). Peptide mixtures were loaded directly onto the analytical column. A binary gradient was used for peptide elution. Mobile phase A was 0.1% formic acid, 2% acetonitrile, whereas mobile phase B was 0.1% formic acid, 80% acetonitrile. Column equilibration (10 min) and sample loading (4 µL), both run at 500 nL/min, were

performed at 0% B. Gradient elution was achieved at 300 nL/min flow rate, and ramped from 8% B to 35% B in 60 min, and from 35% B to 100% B in additional 8 min; after 5 min at 100% B, mobile phase composition was finally brought to 0% B in 2 min. MS detection was performed on a quadrupole-orbitrap mass spectrometer Q-Exactive (Thermo Fisher Scientific) operating in positive ion mode, with nanoelectrospray (nESI) potential at 1800 V applied on the column front-end via a tee piece. Data-dependent acquisition was performed using a top-12 method with resolution (FWHM), AGC target and maximum injection time (ms) for full MS and MS/MS of, respectively, 70,000/17,500, 1e6/1e5, 50/60. Mass window for precursor ion isolation was 1.6 *m/z*, and normalized collision energy was 25. Dynamic exclusion was 5 s.

Data were processed using Proteome Discoverer 1.4 (Thermo Fisher Scientific), using Sequest as the search engine, and the appropriate database as indicated below. The following search parameters were used: MS tolerance 10 ppm; MS/MS tolerance 0.02 Da; variable modifications oxidised methionine, carbamylated N-terminus / lysine; fixed modifications carbamidomethyl cysteine; enzyme trypsin; max. missed cleavages 0 (first level search), 1 (second level search). For first level searches, high confidence peptides were selected using the "fixed value PSM validator" (first search) using the default Xcorr thresholds: 1.9 (+2 ions), 2.3 (3+ ions), 2.6 (4+ or higher); for second level searches, the standard node "target-decoy PSM validator" implemented in Proteome Discoverer was added to the workflow in order to achieve a maximum false discovery rate of 1%. When multiple organisms were searched, separate database searches were performed for each of the selected species.

For first level searches, the global Uniprot database (reviewed entries only, 439,942 sequences) was queried. In order for a species to be included in the second level search, it had to be identified with a minimum of two proteins, each one based on a minimum of two unique high-confidence peptides. The "protein grouping" feature was unselected during the evaluation of first level searches. Second level searches were targeted at specific databases, corresponding to the species selected during the first level search; depending on the sample being analysed (see Results and Discussion), two or more of the following databases were queried: (i) *Bos taurus* reference proteome (23,869 sequences); (ii) *Glycine max* complete proteome (64,639 sequences); (iii) *Oryza sativa* complete proteome (3086 sequences); (iv) *Pisum sativum* sequence database (1489 entries, unreviewed); (v) *Triticum æstivum* sequence database (5746 entries, unreviewed); (vi) *Hordeum vulgare* sequence database (1945 entries, unreviewed). All databases were accessed on Sep 2013. Confidence intervals for binary calibration curve equations were calculated in Prism v 7.0 (GraphPad Software, La Jolla, CA).

RESULTS AND DISCUSSION

Workflow. In order to perform "speciation" of a protein mixture, i.e., in order to detect the species of origin of all its proteinaceous components, a workflow combining a classical "bottom-up" approach (protein digestion and nanoLC-MS/MS analysis) with a two-step database search was developed (Figure 1). MS/MS data produced by the analysis of tryptic peptides were first searched on the entire Uniprot database. In order to reduce the enormous search space, thus shortening computer processing times, the search was restricted to proteins classified as "reviewed" by Uniprot. Additionally, maximum allowed missed cleavages was set to 0 in this first search. In order to further reduce search time, decoy database search was not performed during first level searches. Instead, standard Sequest Xcorr fixed value PSM validator was used to estimate putatively high confidence identifications.

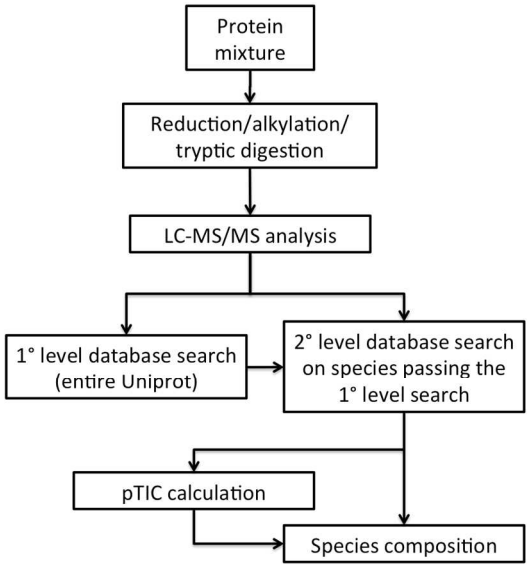

Figure 1. Analytical workflow.

The inspection of results produced by the first search allowed generate a list of candidate species to be queried in the second level, more specific searches.

Table 2. Qualitative analysis of protein extracts (as listed in Table 1).

Sample	Species found in first search*	Species found in second search
Milk powder	*Bos taurus* (16, 105)	not required
Pea protein extract	*Pisum sativum* (67, 186)	not required
Soy protein extract	*Glycine max* (64, 214)	not required
Wheat gluten	*Triticum aestivum* (40, 87) *Oryza sativa*** (2, 8) *Hordeum vulgare*** (2, 4)	*T. aestivum* 100%

(*) Species detected with a minimum of 2 unique proteins containing at least 2 high-confidence unique peptides; in parenthesis, the number of unique proteins and unique associated peptides is indicated. (specific proteins and total unique peptides in parenthesis). (**) False positives.

Table 3. Analysis of two model ternary mixtures.

Sample (true % values)	Replicate	Species detected in first search*	Second search and pTIC calculation	% composition based on binary curves	% composition (average values)
Bos taurus 72% *Glycine max* 20% *T. aestivum* 8%	1	*Bos taurus* (130) *Glycine max* (80) *T. aestivum* (39)	*Bos taurus* 86.3% *Glycine max* 8.6% *T. aestivum* 5.1%	*Bos taurus* 76% *Glycine max* 14% *T. aestivum* 10%	*Bos taurus* 74% (CV=2%) *Glycine max* 15% (CV=10%) *T. aestivum* 10% (CV=6%)
	2	*Bos taurus* (121) *Glycine max* (74) *T. aestivum* (44)	*Bos taurus* 83.9% *Glycine max* 10.3% *T. aestivum* 5.9%	*Bos taurus* 74% *Glycine max* 15% *T. aestivum* 11%	
	3	*Bos taurus* (82) *Glycine max* (52) *T. aestivum* (29)	*Bos taurus* 82.7% *Glycine max* 12.0% *T. aestivum* 5.3%	*Bos taurus* 73% *Glycine max* 17% *T. aestivum* 10%	
Bos taurus 80% *Glycine max* 10% *P. sativum* 10%	1	*Bos taurus* (104) *Glycine max* (59) *P. sativum* (64)	*Bos taurus* 79.3% *Glycine max* 6.0% *P. sativum* 14.7%	*Bos taurus* 78% *Glycine max* 11% *P. sativum* 11%	*Bos taurus* 79% (CV=2%) *Glycine max* 11% (CV=9%) *P. sativum* 10% (CV=6%)
	2	*Bos taurus* (136) *Glycine max* (85) *P. sativum* (81)	*Bos taurus* 80.1% *Glycine max* 5.8% *P. sativum* 14.1%	*Bos taurus* 80% *Glycine max* 10% *P. sativum* 10%	
	3	*Bos taurus* (130) *Glycine max* (70) *P. sativum* (84)	*Bos taurus* 79.3% *Glycine max* 7.5% *P. sativum* 13.2%	*Bos taurus* 78% *Glycine max* 12% *P. sativum* 10%	

(*) Species detected with a minimum of 2 unique proteins containing at least 2 high-confidence unique peptides (unique peptides in parenthesis). % composition for *Bos Taurus, Glycine max*, *Triticum aestivum* (*T. aestivum)* and *Pisum sativum (P. sativum)* were estimated using binary calibration curves.

In order to generate such list, proteins were not grouped according to the principle of parsimony. In this way, unique peptides were univocally assigned to only a single protein hit, thus eliminating the ambiguity possibly caused by either protein isoforms or homologous proteins from several species sharing the same unique peptides. This ultimately led to characterizing each unique peptide identified in the first level search as "species-specific" (one peptide from a single protein, from a single gene). Peptides classified as "unique" by Proteome Discoverer after protein ungrouping were, indeed, uniquely associated to a particular species (*Bos taurus*, *Glycine max*, etc).

Qualitative analysis. An example of data output, relative to the analysis of pure protein standards, is illustrated in Table 2. When MS/MS data obtained from the analysis of the pure bovine milk powder protein digest were processed through first level database search, Proteome Discoverer identified 16 proteins by two or more high-confidence unique peptides after protein ungrouping. The sum of all species-specific peptides identified in the bovine protein digest was 105. For the purpose of qualitative analysis, no additional search was required for pure bovine, soybean and pea protein extracts, since first level search returned a single species (the expected one), characterized by hundreds of unique, species specific identifications. In case of the wheat protein extract, three species (2 false positives) were singled out in the first search. The second search, performed on expanded, species-specific databases (see Experimental section for details) and using false discovery rate estimations, left the only true positive species (wheat) identified by 40 proteins containing 2 or more high confidence peptides (160 peptides in total), whereas no single species-specific proteins were identified for rice and barley. The few peptide sequences belonging to *Oryza sativa* or *Hordeum vulgare* proteins, identified as species-specific in the first search were thus either: (i) false positive identifications; (ii) sequences which did not match any "reviewed" wheat protein entry and did match some homologous rice/barley proteins classified as "reviewed" in the Uniprot database, but were ultimately assigned to wheat proteins classified as "unreviewed" in the second level search.

Concerning binary mixtures (n=28, Supporting information), in several occasions first level search was sufficient to determine the qualitative composition of the binary mixture. Results from pure standards as well as multi-species mixtures (see Supporting information, Tables S1-S4 for binary mixtures and Table 3 for ternary mixtures), gave the correct speciation of the protein mixture, except for one case (bovine/wheat 95:5 mixture). Besides, in a few cases concerning protein mixtures containing wheat proteins, *Oryza sativa* (rice) and *Hordeum vulgare* (barley) were identified in first level searches. Nevertheless, the second level search always assigned the correct number and identity of the species from which proteins were derived. The fact that no *Triticum aestivum* was identified in first level search of the milk powder:wheat gluten 95:5 (protein w/w) mixture is due to the very conservative nature of the filter which has been chosen (positive identification of a minimum of two specific proteins assigned with a minimum of two unique peptides).

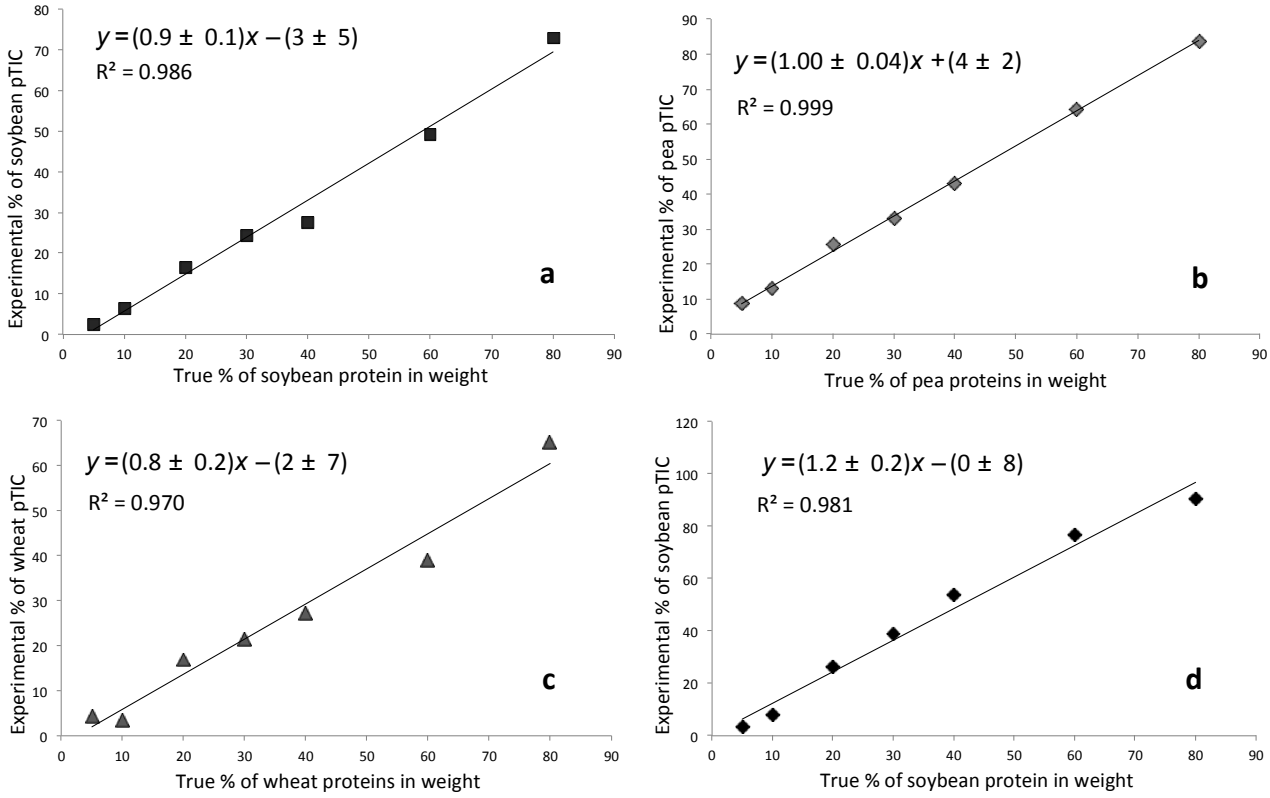

Figure 2. Calibration curves for: (a) bovine:soybean protein mix (w/w), the experimental pTIC of soybean peptides is expressed as a function of % of soybean proteins (w/w); (b) bovine:pea protein mix (w/w); (c) bovine:wheat protein mix (w/w); (d) wheat-soybean protein mix (w/w). Uncertainties of slope and intercept are expressed as 95% CI.

Quantitative data analysis. The second objective of this work was to determine as accurately as possible the relative abundances of the bovine and the vegetable proteomes composing binary model mixtures under investigation. Inferring protein abundance from the chromatographic peak areas of associated tryptic peptides is a well-established practise in proteomics.[21] Since several factors can influence recovery and signal intensity of a single peptide, this label-free approach (like similar efforts based on spectral counting) fails to be as accurate as alternative methods relying on isotopic labelling, especially for low abundance proteins.

Nevertheless, when the concept of direct proportionality between peptide peak areas and protein relative abundance is transferred to the level of whole proteomes, the measurement can be much more precise and accurate, because it relies on tens to hundreds of distinct measurements (peak areas of all species-specific peptides which were identified in a given nanoLC-MS/MS run). The sum of all peak areas of peptides unique to a particular organism was here called pTIC (proteome total ion current). Figure 2 shows how a direct proportionality between the relative amount (w/w) of a given proteome and its relative pTIC exists. In fact, calibration curves were generated by plotting on the x-axis true w/w proteomic fraction of component "B" (e.g., w/w fraction of proteins of soybean origin in Figure 2a) *versus* pTIC relative abundance of component "B", expressed as $pTIC_B / (pTIC_A + pTIC_B)*100$. Calibration curves indicate a good correlation between pTIC relative abundance and the corresponding proteome relative abundance.

In a single nanoLC-MS/MS analysis, it was thus possible to determine the species from which a binary protein mixture was generated, and estimate the relative % composition of the two proteomes. To the authors' knowledge, this is the first report demonstrating the possibility of extracting such comprehensive information from a protein powder formulation by using any analytical technique.

In order to assess the capability of the described method to characterize multi-species protein mixtures (n>2) from a qualitative, and, possibly, quantitative point of view, two ternary mixtures were prepared from pure standards, and analysed in technical triplicates (starting from the reduction and alkylation step). Table 3 shows that, in all cases, first level searches detected all components of the mixtures without any false positive identification. Second level searches and pTIC calculations were performed as described for the binary mixtures. Since the major component of both mixtures was from bovine origin (situation which is generally encountered in commercial milk replacer formulations), in order to have a quantitative estimation on the respective w/w proteome fractions from soybean, wheat and pea, % pTICs were interpolated with the respective binary calibration curves (Figures 2a, 2b, 2c)

Finally, the w/w fraction of proteins from bovine origin (X_B) was calculated as follows: $X_B = 1 - X_S - X_Z$, where X_S was the w/w fraction of proteins from soybean origin, and X_Z was the w/w fraction of proteins from either wheat (mixture 1) or pea (mixture 2) origin. X_B, X_S and X_Z were expressed as % values in Table 3. Triplicate analyses of the ternary mixtures confirmed the precision of the method. In fact, coefficients of

variation (CVs) for relative % composition reported in Table 3 ranged from 2 to 10% with an average CV of 6%.

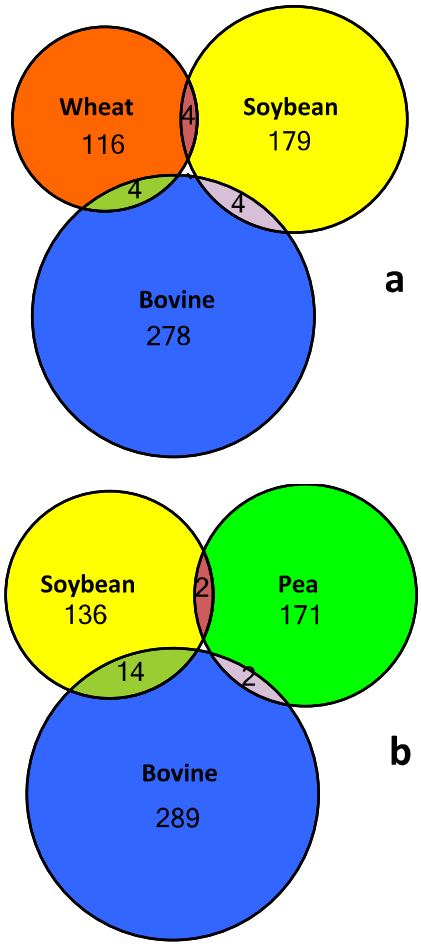

Figure 3. Venn diagrams relative to the analysis of ternary mixtures. (a) bovine/soybean/wheat mixture 72/20/8 (w/w/w of total proteins); (b) bovine/soybean/pea mixture 80/10/10 (w/w/w of total proteins). The numbers refer to peptide identifications obtained from the second search.

Concerning quantitation accuracy, the calculated w/w abundance of bovine proteins present in the ternary mixtures was not in accordance with the true value by just 2 and 1 percentage points, respectively, (absolute error) when calibration curves were used. On the other hand, absolute errors referred to the same measured amounts were 12 and 1 percentage points, respectively, when abundances were estimated without calibration curves. Similar results were obtained for the binary mixtures previously considered. In particular, in the range 60-90% (w/w) of bovine proteins, which is the range of highest interest for milk replacer analysis, results from three types of binary mixtures gave an average error of 6 percentage points (range 3-12%) in determining the fraction of proteins from bovine origin, when the estimation was made relying on pTICs alone (Table S5). On the other hand, when the same pTIC values were interpolated using the corresponding best-fit curves, the average error was reduced to 2 percentage points, with an overall range of 0-6% (Table S5).

By assuming that an acceptable achievement for a quality control analysis of milk replacers would be the determination of the relative abundance of bovine proteins within 5 percentage points, it is clear that such accuracy could be currently met only by interpolating pTIC values with standard calibration curves. On the contrary, by relying on pTICs alone, it would be just possible to make an estimation of proteome's relative abundances.

Table 4. Analysis of commercially available products.

Sample	Protein ingredients*	Proteome speciation
1	Milk powder (50%), whey powder, pea flour	*Bos taurus* (99%) *Pisum sativum* (1%)
2	Milk powder , whey powder, wheat flour	*Bos Taurus* (97%) *T. aestivum* (3%)
3	Milk powder ,wheat flour	*Bos taurus* (88%) *T. aestivum* (12%)
4	Whey powder, wheat protein concentrate, yeast extract, soy protein concentrate	*Bos Taurus* (81%) *T. aestivum* (9%) *Glycine Max* (9%) *S. cerevisiae* (1%)***
5	Whey powder, wheat gluten, soy protein concentrate, pea flour	*Bos Taurus* (78%) *T. aestivum* (18%) *Glycine Max* (2%) *Pisum sativum* (2%)
6	Milk powder , wheat gluten	*Bos Taurus* (78%) *T. aestivum* (16%) *Glycine Max* (3%)** *Vicia faba* (2%)**, ***
7	Whey powder, wheat gluten, soy protein concentrate, pea flour	*Bos Taurus* (68%) *T. aestivum* (19%) *Glycine Max* (11%) *Pisum sativum* (2%)
8	Milk powder, whey powder, wheat gluten, soy protein concentrate	*Bos Taurus* (66%) *Glycine Max* (19%) *T. aestivum* (15%)
9	Whey powder, wheat gluten, pea flour	*Bos Taurus* (64%) *T. aestivum* (34%) *Pisum sativum* (2%)
10	Whey powder, wheat protein concentrate, potato protein concentrate, soy protein concentrate	*Bos Taurus* (57%) *Solanum tuberosum**** (18%) *T. aestivum* (13%) *Glycine Max* (12%)

* as declared by the manufacturer, sorted by decreasing order of weight precentage; ** not declared by the manufacturer; *** determined without the use of a calibration curve.

Another interesting feature of this method is displayed in Figure 3. The Figure shows a Venn diagram, which classifies peptide identifications in both ternary mixtures previously discussed. Peptides are displayed as features, which are either unique (species-specific) or shared between two different species. As it can be seen, the majority of peptides identified in

the second level searches is, indeed, species-specific. Since the w/w fraction of the species is calculated on the basis of species-specific peptides, we may conclude that, at least for the protein mixtures under investigation, the determination of the relative % fraction of the respective proteomes from bovine, pea, soybean and wheat will be feasible, because most of the full scan MS signal comes from species-specific peptides. It has to be remarked that the great majority of milk replacer formulations is primarily based on these four different protein sources, though a preliminary study on commercially available products (Table 4) has detected other protein sources as well (*Solanum tuberosum, Vicia faba*).

As it can be assessed by looking at Table 4, the proposed method allowed to detect all ingredients present in ten commercially available milk replacers. In just one case (sample 6) two minor ingredients were not declared by the manufacturer. Proteome composition in w/w % was obtained by interpolation with binary calibration curves. Though the true value was not available in this case, there is a high correlation between what declared by the manufacturer (listed by decreasing order of weight percentage) and what found by the proteomic analysis. Some discrepancies could be due to the fact that ingredients may differ in protein content; thus, for example, soy protein concentrate might be present at lower abundance than wheat gluten in milk replacer 8. Nevertheless, the higher protein content of the former justifies the fact that proteomic analysis found *Glycine max* as the sample's second most abundant protein source.

CONCLUSION

Thanks to two-step database searches, pTIC calculations and dedicated calibration curves, it was possible to determine the proteome composition of protein mixtures generated by combining protein extracts from several species. This method is being applied to the characterization of milk replacer formulations in order to assess one important aspect of their nutritional quality. To the author's knowledge, there is no currently available alternative analytical method able to provide similar qualitative and quantitative information for this specific application.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Data relative to the calibration curves for binary mixtures (DOC)

AUTHOR INFORMATION

Corresponding Author

* Email: Gaspari@unicz.it; phone +39 0961 3694337.

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

The authors would like to acknowledge Coop Italia (Casalecchio di Reno, BO, Italy) for very helpful feedbacks and financial support.

REFERENCES

(1) Herrero, M.; Simó, C.; García-Cañas, V.; Ibáñez, E.; Cifuentes, A. *Mass Spectrom. Rev.* **2012**, *31* (1), 49–69.
(2) Aiello, D.; De Luca, D.; Gionfriddo, E.; Naccarato, A.; Napoli, A.; Romano, E.; Russo, A.; Sindona, G.; Tagarelli, A. *Eur. J. Mass Spectrom. (Chichester, Eng).* **2011**, *17* (1), 1–31.
(3) Cunsolo, V.; Muccilli, V.; Saletti, R.; Foti, S. *J. Mass Spectrom.* **2014**, *49* (9), 768–784.
(4) Yates, J. R.; Carmack, E.; Hays, L.; Link, A. J.; Eng, J. K. *Methods Mol. Biol.* **1999**, *112*, 553–569.
(5) Pandey, A.; Mann, M. *Nature* **2000**, *405* (6788), 837–846.
(6) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17* (10), 994–999.
(7) Yates, J. R. *J. Am. Chem. Soc.* **2013**, *135* (5), 1629–1640.
(8) Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *Nat. Protoc.* **2015**, *10* (5), 701–714.
(9) Intelicato-Young, J.; Fox, A. *J. Microbiol. Methods* **2013**, *92* (3), 381–386.
(10) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20* (4), 157–171.
(11) Karlsson, R.; Davidson, M.; Svensson-Stadler, L.; Karlsson, A.; Olesen, K.; Carlsohn, E.; Moore, E. R. B. *J. Proteome Res.* **2012**, *11* (5), 2710–2720.
(12) Calvano, C. D.; De Ceglie, C.; Monopoli, A.; Zambonin, C. G. *J. Mass Spectrom.* **2012**, *47* (9), 1141–1149.
(13) Cuollo, M.; Caira, S.; Fierro, O.; Pinto, G.; Picariello, G.; Addeo, F. *Rapid Commun. Mass Spectrom.* **2010**, *24* (11), 1687–1696.
(14) Cordewener, J. H. G.; Luykx, D. M. A. M.; Frankhuizen, R.; Bremer, M. G. E. G.; Hooijerink, H.; America, A. H. P. *J. Sep. Sci.* **2009**, *32* (8), 1216–1223.
(15) von Bargen, C.; Dojahn, J.; Waidelich, D.; Humpf, H. U.; Brockmeyer, J. *J. Agric. Food Chem.* **2013**, *61* (49), 11986–11994.
(16) Watson, A. D.; Gunning, Y.; Rigby, N. M.; Philo, M.; Kemsley, E. K. *Anal. Chem.* **2015**, *87* (20), 10315–10322.
(17) Akinyele, I. O.; Harshbarger, K. E. *J. Dairy Sci.* **1983**, *66* (4), 825–832.
(18) Seegraber, F. J.; Morrill, J. L. *J. Dairy Sci.* **1986**, *69* (2), 460–469.
(19) Han, S. W.; Chee, K. M.; Cho, S. J. *Food Chem.* **2015**, *172*, 766–769.
(20) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2* (8), 1896–1906.
(21) Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A. *Proteomics* **2011**, *11* (4), 535–553.

Table of Contents artwork here