Analytica Chimica Acta 1201 (2022) 339616

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

# A method to identify and quantify the complete peptide composition in protein hydrolysates

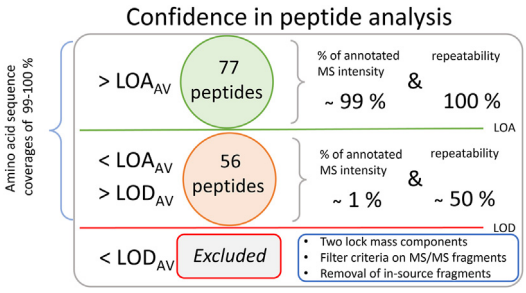
Gijs J.C. Vreeke , Wouter Lubbers , Jean-Paul Vincken , Peter A. Wierenga *

*Laboratory of Food Chemistry, Wageningen University & Research, P.O. 17, 6708 AA, Wageningen, the Netherlands*

## HIGHLIGHTS

- Automated untargeted method to analyse the peptide composition of food hydrolysates.
- Guidelines to evaluate the completeness and reproducibility of peptide annotations.
- Absolute and label-free peptide quantification with UV absorbance.
- Amino acid sequence coverages for single and mixed hydrolysates were 99–100%.
- Concentration based sequence coverage only decreased on average 4% after mixing.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 29 October 2021
Received in revised form
19 January 2022
Accepted 14 February 2022
Available online 17 February 2022

*Keywords:*
Food peptides
Peptidomics
Digestion
UNIFI
Quantitative proteomics
Peptide release kinetics

## ABSTRACT

Automated approaches from proteomics are used to characterise peptides for food applications and in protein digests. Peptide annotations and confidence in these annotations are then based on the fragment spectra. Low reproducibility in repeat analyses has been reported even for annotations with high confidence. When analysing protein hydrolysates (in food) it is important to determine criteria that yield highly reproducible annotations. This study provides a structured approach to determine these criteria. Tryptic hydrolysates of α-lactalbumin, β-lactoglobulin and β-casein were analysed manually and automatically, using an UPLC-PDA-MS method for untargeted identification and absolute label-free quantification of peptides. A lock mass with two components was introduced resulting in an average mass error of 1 ppm. Processing filters were set to ensure reliable annotations based on MS/MS fragmentation, while maintaining maximum amount of information. Peptides in the individual hydrolysates with an MS intensity above the limit of annotation represented 99% of total MS intensity and were 100% consistently annotated between four replicates. Amino acid and peptide sequence coverages for the individual protein hydrolysates were 99–100% and 89–95%, respectively. Mixing the hydrolysates resulted in a loss of 11% of the peptide annotations above the LOA and lower reproducibility (97%) for the remaining annotations, as well as more co-eluting peptides. Calculated concentrations of co-eluting peptides in mixed hydrolysates varied 37 ± 21% from the value for single hydrolysates. The proposed approach allows complete description of peptide composition with highly repeatable annotations and quantification of peptides even in mixed hydrolysates.

© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author.
*E-mail address:* peter.wierenga@wur.nl (P.A. Wierenga).

https://doi.org/10.1016/j.aca.2022.339616
0003-2670/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

Identification of peptides present in enzymatic protein hydrolysates using UPLC-MS is essential in a large variety of scientific disciplines and industrial research [1]. In recent years, this is done often with data processing software originating from proteomics. In traditional proteomics, the goal is to identify the proteins that were in the original sample based on unique peptides identified. This does not necessarily require identification of all the peptides that originate from that protein. In contrast to proteomics, the goal for food applications and digestion studies, is to identify the formed peptides when the proteins in the original sample are known. In such studies, often the presence of peptides is compared in a set of different samples. It is important to know the level of confidence in the presence of individual peptides as reported after automated annotation. It is also important to know how many of the total hydrolysate is included in the analysis. Therefore, the completeness of the analysis should also be evaluated using different parameters than used in proteomics. For instance, a parameter should be used to identify if any peptides were lost during sample preparation (i.e. check for mass balance).

In this study, we test and optimize a method for automated identification and absolute label-free quantification of peptides using a non-targeted UPLC-PDA-MS approach. The aim was to propose a structured approach for data processing and reporting on completeness of peptide analysis.

### 1.1. Peptide identification with mass spectrometry

To characterise the peptide composition in a hydrolysate, an untargeted approach is required. In this approach, all *m/z* signals detected in the mass spectra should be included in the analysis and then converted to a list of identified peptides. The steps in this process are to (1) separate the signals from the noise, and (2) to attribute peptide sequences to the signals. The dilemma in separating the signals from the noise is that with a high noise threshold peptides with a low intensity are not identified. With a low noise threshold non peptide related MS signals are also included in the analysis. In the process of attributing the included *m/z* signals to peptides, multiple challenges occur:

- One peptide in the sample can result in multiple *m/z* peaks in the spectrum: Peptides typically (i) occur in different charge states, (ii) can be present as adducts, or (iii) can be present as in-source fragments [2].
- To link the *m/z* values to peptide sequences a list of potential peptide masses could be generated based on the primary amino acid sequence of the substrate and protease specificity. This requires choices on whether to include (i) peptides that do not fall within the protease specificity, (ii) peptides with missed-cleavages, (iii) peptides with modified AA residues or (iv) peptides that originate from protein impurities.
- In addition, to link the *m/z* value to a peptide sequence, a certain mass error should be taken into account. The number of matches is highly dependent on the mass error [3,4]. If the mass error is set too strictly, peptides may not be included in the final list. If the mass error is set too widely, there is a chance of incorrect identification of the *m/z* value.
- The last challenge is that in some cases, multiple peptide sequences can be matched to an *m/z* signal within the mass accuracy. This is for instance the case for isobaric peptide sequences. To decide what is the correct peptide that should be assigned to a *m/z* value, often the fragmentation spectra are used. The MS/MS fragments are decisive to confirm the positive identification of a peptide.

### 1.2. Key parameters for peptide identification

Despite the dilemmas listed above, many people publish lists of peptides annotated in complex mixtures. To come to the list of peptides, several approaches are used in practice to deal with (1) mass accuracy and (2) MS/MS fragmentation. The mass error used in peptide identification is often reported without explanation how the set value was chosen. In some cases, the choice is made based on the type and settings of the mass spectrometer [5], or based on the observed distribution of mass deviations [6].

To confirm the identity of the peptide, fragmentation spectra need to be analysed. A choice is made on how many of the possible fragments need to be identified to confirm the identification of the peptide. Although this choice is crucial, there is no general consensus on the (absolute or relative) number of fragments that is required for confirmation. Clearly, the number of required MS/MS fragments for confirmation depends on the number of options within the mass error. To distinguish between tryptic peptides originating from a single substrate does not require as many identified fragments as for *de novo* sequencing of peptides [7]. Many standardised algorithms are used in the field of proteomics to automatically identify peptides based on MS/MS fragments as for instance MASCOT [8], SEQUEST [9], or Andromeda [10] with (incorporated) scoring functions [11,12]. In literature, different studies using the same algorithm often do not apply the same threshold scores [13]. The score of a peptide annotation is often linked to a certain confidence level. However, even for annotations above the threshold score, still (only) 32−45% of the identified peptides were repeatably annotated in all replicate injections [14,15]. The question is how one could define a parameter to describe the confidence in the repeatability of the annotation, without the need to analyse multiple replicates.

### 1.3. Peptide quantification

To quantify compounds in mass spectrometry, typically the MS intensity of the ions is used. This intensity is known to vary because of ion-suppression, matrix effects, variation in charge states and day-to-day differences in absolute intensity [16−18]. Ideally, in the targeted MS approach, the MS intensity is corrected for these variations by using isotopically labelled standards, preferably with correction based on a standard addition to a reference sample [19]. In the untargeted approach, it is impossible to have isotopically labelled standards for each peptide, since beforehand it is not known which peptide are present. To avoid the need for (isotopically labelled) standards, Butré et al. have developed in recent years an approach for absolute label-free quantification of peptides based on UV absorbance [20]. The approach uses the predicted molar extinction coefficient of each peptide based on Kuipers et al. to convert UV peak areas to absolute peptide concentrations [21]. This quantification method was successfully applied in the past to for example determine differences in peptide release kinetics by bovine, human and porcine trypsin [22] and to quantify complex peptide mixtures with size-exclusion high-performance liquid chromatography [23]. In complex mixtures, UV peaks of eluting peptides are not always baseline separated. In some cases the individual UV peaks cannot be separately integrated, so that one UV peak should be divided over multiple peptides. It was previously suggested that this could be done using the ratio of MS intensities of the co-eluting peptides. Considering that peptides with similar retention times have more or less similar chemical properties, it was considered that ionisation efficiencies would be comparable as well [20].

### 1.4. Parameters to evaluate the completeness of analysis

When peptides are studied in food sciences, mostly a list or table of annotations is reported, e.g. Ref. [24] without any parameters describing the completeness of the analysis. In some of these cases, a plot is provided in which the identified peptides are mapped against the sequences of the initial protein substrates, e.g. Ref. [25]. These plots may aid the reader in evaluating the completeness, but do not give a value that describes the completeness. In other cases [26,27], the protein sequence coverage, known from the field of proteomics [28], is reported. This parameter describes how many amino acids from the parental protein sequence were identified in at least one of the peptides. This parameter is purely based on unique amino acids, and therefore renamed to amino acid sequence coverage by Butré et al. [20]. They further introduced the quantitative parameters "peptide sequence coverage" and the "molar sequence coverage", to describe the completeness of the peptide identification and of the peptide quantification respectively [20].

To test the reproducibility and completeness of automated annotation, three single protein hydrolysates were analysed. A set of criteria was developed to optimize completeness and validity of annotations. In addition, based on replicate analyses an objective parameter was defined to distinguish the annotations with high reproducibility and low reproducibility.

## 2. Materials and methods

### 2.1. Protein isolates, protease and chemicals

α-lactalbumin (α-LA) was obtained from Davisco Foods International. Inc. (Le Sueur, MN, USA). The α-LA was treated with ethylenediaminetetraacetic acid (EDTA) to remove the calcium ions attached to protein, as described by Deng et al. [29]. β-lactoglobulin (β-LG, L0130), β-casein (β-cas, C6905), bovine trypsin (EC 3.4.21.4, T1426) and aprotinin from bovine lung (A6279) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Leucine enkephalin (Leu-enk, L9133), and Insulin (XI5500) were obtained from Sigma-Aldrich (St. Louis, MO, USA) Angiotensin was obtained from Alfa Aesar (Karlsruhe, Germany). The bovine trypsin had a protein content of 80% of which 100% was bovine trypsin, based on $UV_{214nm}$ area analysis with UPLC-PDA-MS. According to the supplier, the trypsin activity was ≥10,000 BAEE units $mg^{-1}$ protein. The bovine trypsin was treated with N-tosyl-L-phenylalanyl chloromethyl ketone (TPCK) to inactivate any chymotrypsin activity (≤0.1% BTEE units $mg^{-1}$ protein). The aprotinin solution contained 2.3 mg $mL^{-1}$ aprotinin based on previous UPLC-MS results [30]. All other chemicals were purchased of analytical grade and purchased from Sigma or Merck.

### 2.2. Enzymatic hydrolysis of proteins

α-LA, β-LG and β-cas were each dissolved in 10 mL Millipore water at 1% [weight powder/volume]. The pH was adjusted to pH 8.0 and the solutions were equilibrated for 0.5 h at 37 °C. Trypsin was dissolved (10 mg powder $mL^{-1}$) in Millipore water and added to the equilibrated solutions at an enzyme to substrate ratio of 1:100 [w/w]. Protein hydrolysis was performed in duplicate for 2 h in a pH-stat (Metrohm, Herisau, Switzerland) with a 0.2 M NaOH solution to keep the pH constant. The volume of added NaOH was used to calculate the degree of hydrolysis (DH) with equation (1) [31],

$$DH_{stat}[\%] = V_b \times N_b \times \frac{1}{\alpha} \times \frac{1}{m_p} \times \frac{1}{h_{tot}} \times 100\,\% \quad (1)$$

where $V_b$ [mL] is the volume of added NaOH; $N_b$ [mol $L^{-1}$] is the normality of NaOH; α is the average degree of dissociation of the α-NH group ($1/\alpha = 1.3$ at 37 °C and pH 8 [32]; $m_p$ [g] is the amount of protein in solution; $h_{tot}$ [mmol $g^{-1}$] is the number of peptide bonds per gram of protein.

Trypsin was inactivated by addition of 15 μL aprotinin/mL hydrolysate, afterwards the samples were stored at −20 °C.

### 2.3. Sample preparation

The protein hydrolysates were incubated for 2 h in 10 mM dithiothreitol (DTT) and 50 mM Tris-HCl buffer at pH 8.0 at a protein concentration of 0.5% [w/v], to reduce disulphide bonds. After incubation, the individual protein hydrolysates were mixed in mixtures of two substrates (α-LA + β-LG, α-LA + β-cas, β-LG + β-cas) and a mixture of three substrates (α-LA + β-LG + β-cas). The individual hydrolysates were diluted [1:2] with 0.15% TFA [v/v] in MQ and the mixtures of two substrates were diluted [2:1] with 0.3% TFA [v/v] in MQ water. 2 μL of 5% [v/v] TFA was added to 100 μL of the mixture with three substrates. The final molar protein concentrations are shown in (Table 1). Afterwards, the samples were centrifuged (10 min, 14,000×*g*, 20 °C) and the supernatants were injected in four replicates on the UPLC-MS.

### 2.4. Reverse phase ultra-high performance liquid chromatography (RP-UPLC)

The hydrolysates were analysed on a Waters H-class Acquity UPLC system (Milford, MA, USA). Peptide separation was done using a BEH C18 column (1.7 μm, 2.1 × 150 mm, Waters) that was coupled to a Waters Acquity UPLC PDA detector. The mobile phase consisted of a gradient of two solutions. Eluent A was UPLC-Grade water with 1% [v/v] acetonitrile (ACN) and 0.1% [v/v] trifluoroacetic acid (TFA) and eluent B was ACN with 0.1% [v/v] TFA. 4 μL of the supernatant was injected into the column thermostated at 30 °C. The peptides were separated using the following elution profile: 0–2 min isocratic on 3% B; 2–10 min linear gradient from 3 to 22% B; 10–16 min linear gradient 22–30% B; 16–21 min linear gradient 30–100% B; 21–26 min isocratic on 100% B; 26–28 min linear gradient 100-3% B and 28–32 min isocratic on 3% B. During the first 2 min of isocratic elution (1.3 column volumes), the flow was directed to the waste to protect the MS and avoid any influence of remaining salt or unbound material on the MS or UV signals. The flow rate was set on 350 μL $min^{-1}$. Detection was performed using a PDA, which scanned the absorbance at the fixed wavelength of 214 nm at 1.2 nm resolution and 40 datapoints $s^{-1}$.

### 2.5. Electron spray ionisation time of flight mass spectrometry (ESI-Q-TOF-MS)

Mass spectra were obtained by an online Waters Synapt G2-Si high definition mass spectrometer coupled to the RP-UPLC, equipped with a z-spray electrospray ionisation source, a hybrid quadrupole and an orthogonal time-of-flight analyser. The capillary voltage was set to 3 kV with the source operation in positive ion mode and the source temperature at 150 °C. The sample cone was operated at 35 V and nitrogen was used as desolvation gas (500 °C, 800 L $h^{-1}$) and cone gas (200 L $h^{-1}$). Full scan MS and MS/MS data were acquired between 200 and 3000 *m/z* with a scan time of 0.3 s in resolution mode (V-mode) using an MSe method. MSe is a data-independent approach, were all precursor ions present in the MS at a given time were fragmented simultaneously. The trap collision energy was set at 4 V in single MS mode and ramped from 20 to 45 V in MS/MS mode. Prior to the analysis, the system was calibrated using sodium iodide, which was accepted when the average mass error on the calibrant peaks was below 2 ppm. Online lock

**Table 1**
Characteristics of the protein material as used as starting material for the hydrolysis.

| Protein and Uniprot code[a] | N-factor[a] [g of protein/g N] | Protein content [w/w] | Purity[b] [%] | Protein loss with hydrolysis[c] [w/w] | Injected hydrolysate concentration [μM] | Molecular weight[a] | $\varepsilon_{214}$ [L/Mol/cm] | #AA[a] | #CS | #Possible specific peptides | $DH_{Stat}$ [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| α-LA (P00711) | 6.25 | 93% | 90% | 7.8% | 86 μM | 14,186 | 300,395 | 123 | 13 | 105 | 5.6 ± 0.1% |
| β-LG (P02754) | 6.29 | 96% | 100% | 7.8% | 78 μM | 18,367 (A[e])<br>18,281 (B[e]) | 293,410 (A[e])<br>293,362 (B[e]) | 162 | 18 | 136 | 7.7 ± 0.2% |
| β-cas (P02666) | 6.39 | 90% | 90% | 7.8% | 50 μM | 23,983[d] | 423,992 | 209 | 15 | 190 | 6.2 ± 0.1% |
| Trypsin (P00760) | 5.97 | 80% | 100% | | – | – | – | – | – | | |

α-LA: α-Lactalbumin, β-cas: β-casein; β-LG: β-Lactoglobulin.
[a] From Uniprot (http://www.uniprot.org).
[b] Reported previously in Ref. [30].
[c] This protein loss factor corrects for (1) sampling during hydrolysis (2) Addition of trypsin inhibitor. after hydrolysis and (3) pH adjustment in the pH-stat.
[d] The molecular weight of β-casein takes into account five phosphorylated serine residues, as identified with RP-UPLC-MS.
[e] The A and B indicate the genetic variant of β-lactoglobulin.

mass data were acquired as a separate trace using Waters LockSpray at a set lockspray capillary voltage of 3.0 kV and at a sample infusion rate of 20 μL $min^{-1}$. Three peptides were evaluated as lock mass components: Leucine-enkephaline, $[M+H]^+$: 556.276575 *m/z*, Angiotensin II, $[M+2H]^{2+}$: 523.774534 *m/z* $[M+H]^+$: 1046.541791 *m/z* and Insulin ($[M+3H]^{3+}$: 1910.876843 *m/z*. The optimised lock mass solution contained 0.4 μM Leu-Enk and 0.7 μM insulin dissolved in 50% [v/v] methanol containing 3% [v/v] acetic acid and 0.4% [v/v] diethylamine in UPLC grade water.

### 2.6. The development of data processing

The data processing method was developed by following the proposed steps (Fig. 1). To set up the method in UNIFI, the α-LA hydrolysate was processed with the default method, without lock mass, without filters on MS/MS fragmentation or in-source fragments. The lock mass was optimised by analysis of the β-LG hydrolysate using different (combinations of) lock mass compounds. A concentration series of the α-LA hydrolysate was analysed with the optimised lock mass compounds to determine the LOD and LOA. The data, of the individual α-LA, β-LG and β-cas hydrolysates were obtained with the optimised double lock mass, and processed manually (1 replicate/protein) and automatically (4 replicates/protein). The mixtures of the proteins were processed only automatically (4 replicates/protein).

### 2.7. Peptide identification manually

Analysis of the mass spectrometry data of the individual protein hydrolysates was done manually in MassLynx software version 4.2. Manual annotation was performed similarly as in previous studies of our group [20,30]. The *m/z* signals in the spectra were linked to possible peptides, based on the primary amino acid sequence of the substrate of interest. The peptides from other proteins than the main protein were not considered in manual annotation. Two AA modifications were taken into account: Methionine oxidation (+16 Da) for α-LA and serine phosphorylation (+80 Da per phosphoserine) for β-cas. The maximum allowed mass error was 100 ppm. To confirm the tentative annotations, the MS/MS spectra were used to identify b- and y-fragments. Mass spectrum deconvolution was used to extract the intact protein mass with the MaxEnt function in MassLynx.

### 2.8. Automated peptide identification

Automated peptide annotation was performed using the peptide mapping package in UNIFI software version 1.8. The amino acid sequences of α-LA, β-LG (variant A and B) and β-cas were inserted and processed with trypsin as enzyme specificity on the semi-digest option. The semi-digest option included peptides that matched trypsin's specificity on at least the N- or C-terminal side. The included variable modifications were oxidation of the methionine (up to 1 per peptide) and serine phosphorylation (up to 5 per peptide). First, a default peak processing method was used with the default UNIFI settings without lock mass correction. In the default method, all signals were processed with a minimum signal intensity of 1000 detector counts in both the MS and MS/MS chromatograms. The maximum acceptable mass error was set at 100 ppm in the MS and 20 ppm in the MS/MS. For the final peak processing method, the minimum signal intensity was changed to 250 detector counts for the MS and 75 detector counts for the MS/MS, which corresponded with 3x the noise in the MS and MS/MS spectra. The maximum acceptable mass error was decreased from 100 ppm to 10 ppm. After peak processing, the match between *m/z* signals and a peptide sequence (peptide-spectrum match) was done with the algorithm incorporated in UNIFI. The algorithm returned for all precursor ions a potential peptide annotation (when possible). In case of multiple tentative annotations within the mass error, the peptide was matched to the annotation with

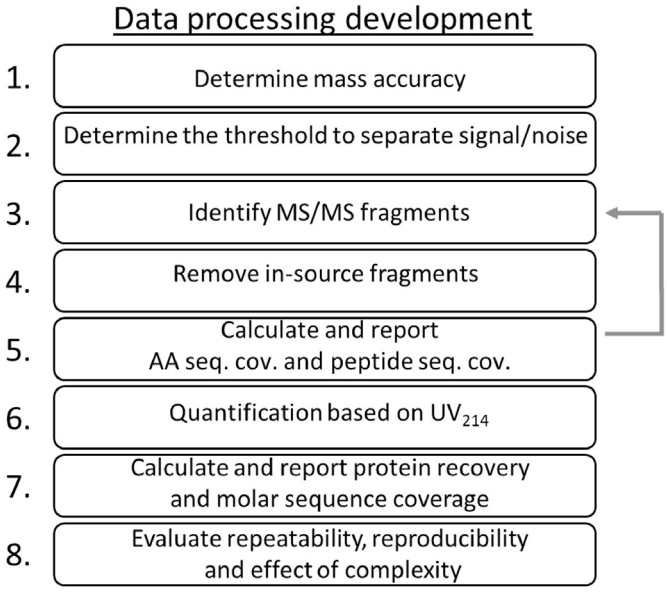



**Fig. 1.** The proposed steps for the development of an automated UPLC-PDA-MS data processing method.

most identified MS/MS fragments. This list still contains many entries of which some are not considered sufficiently reliable (e.g. if only one fragment was identified in a 6 AA peptide). These were removed by applying a filter selection constructed by the user. All annotations that were not confirmed with (at least) 2 b/y fragments were excluded. Peptides were included when 2 b/y fragments were identified (relevant for peptides with 2–6 amino acids) and when more than 15% of the possible b/y fragments were identified (relevant for peptide with 7–16 amino acids) or when at least 5 b/y fragments were identified (relevant for peptides ≥17 amino acids). The peptide-length dependence of the fragmentation criteria were set based on the peptides that could theoretically be formed using a-specific hydrolysis. Listing these peptides showed that with these settings there was a negligible chance to have isobaric peptides that still meet the requirements. In addition, an additional mass error restriction of 5 ppm was set for peptides eluting within 15.00 min after the injection. The filter also removed in-source fragments recognised by UNIFI and annotations with a $H_2O$ or $NH_3$ adduct. Annotations with a methionine oxidation were only included when originating from α-LA, while annotations with serine phosphorylation were only included when originating from β-cas. In-source fragments that were not recognised by the UNIFI software as such, were removed using PeptQuant, an in-house developed script in Matlab v2018b. Annotations were considered as in-source fragments if the parent peptide and potential in-source fragment eluted at a similar retention time and the in-source fragment included the same sequence as the peptide and the in-source fragment had a lower parent ion MS intensity than the peptide. In case a unique peptide was annotated twice, the peptide with the lowest MS intensity was removed. The presence of intact protein was manually evaluated and if present, added to the UNIFI output.

### 2.9. Peptide quantification

Peptides were quantified based on UV absorbance at 214 nm. The UV peaks between 1 and 20 [min] were integrated using the peak integration option in Masslynx. The peak integration was performed using a peak to peak baseline noise ratio of 500, a peak width of 0.28 min and a baseline increase of 1% (all values determined manually). The UV peaks corresponding to tris, DTT and aprotinin were excluded. The list of UV peak areas and retention times was coupled to the filtered UNIFI output using PeptQuant. The coupling was based on the start and end retention time [min] of the integrated UV peak and the retention time of the annotated peptide [min], taking into account the retention time offset between UV and MS (0.08 min). If multiple peptides were linked to the same UV peak, the UV peak area was divided over the co-eluting peptides based on their total ion count and molar extinction coefficient ϵ (Equation (2)).

$$A_{214,i}[\mu\text{AU}\cdot\text{min}] = \left( \frac{\varepsilon_{214,i} \cdot MS_{tic,i}}{\sum \varepsilon_{214} \cdot MS_{tic}} \right) \times A_{214,\,tot} \qquad (2)$$

where $A_{214,i}$ [μAU·min] is the UV peak area at 214 nm assigned to co-eluting peptide i, $A_{214,tot}$ [μAU·min] is the total UV peak area at 214 nm, $\varepsilon_{214,i}$ [L $Mol^{-1}$ $cm^{-1}$] is the molar extinction coefficient at 214 nm and $MS_{tic,i}$ [counts] is the total ion count for co-eluting peptide i.

The concentration of each peptide, $C_{peptide}$ [μM], was calculated with Equation (3).

$$C_{peptide\ [\mu M]} = \frac{A_{214} \cdot Q}{\varepsilon_{214} \cdot l \cdot V_{inj} \cdot k_{cell}} \qquad (3)$$

where $A_{214}$ [μAU min] is the UV peak area at 214 nm, $V_{inj}$ [μL] is the volume of sample injected, Q [μL $min^{-1}$] is the flow rate and l [cm] is the path length of the UV cell, which is 1 cm according to the manufacturer. The molar extinction coefficient $\varepsilon_{214}$ [L $Mol^{-1}$ $cm^{-1}$] for each peptide was calculated according to Kuipers et al. [21]. The cell constant, $k_{cell}$ for the UV detector was 0.78. The $k_{cell}$ was determined with a concentration series of α-LA and angiotensin II, with known concentrations. Corrected for protein content, purity and dilution during hydrolysis, the expected protein concentrations were 86 μM for α-LA, 78 μM for β-LG and 50 μM for β-cas. Equation (3) was also used to calculate the expected total UV based on the starting protein concentrations. The molar extinction coefficients $\varepsilon_{214}$ [L $Mol^{-1}$ $cm^{-1}$] of the hydrolysates were corrected for the degree of hydrolysis, resulting in a coefficient of 294,089 L $Mol^{-1}$ $cm^{-1}$ for α-LA, 281,944 L $Mol^{-1}$ $cm^{-1}$ for β-LG and 412,089 L $Mol^{-1}$ $cm^{-1}$ for β-cas.

### 2.10. Limits of detection, annotation and quantification

The limit of detection (LOD) and limit of annotation (LOA) of peptides were determined using a dilution series of the α-LA hydrolysate with a hydrolysate concentration from 0.00005 to 5 g $L^{-1}$. The MS intensity was reported for the highest dilution in which a peptide was respectively detected or annotated. The limit of detection (LOD) was defined as the lowest MS intensity of a peptide at which the precursor ion was recognised as signal in UNIFI. To be detected as a signal, the datapoints in the spectra had to form a recognisable (Gaussian) peak shape and the MS peak height had to be above the minimum detector count threshold in the MS (>250 counts). The limit of annotation (LOA) was defined as the lowest MS intensity for a peptide to be annotated and meet the criteria on MS/MS fragmentation as stated in the UNIFI filters. The individual LOD and LOA of the peptides were averaged to determine the general LOD and LOA for this method. The average LOD was used in this study to differentiate signals from the noise and the average LOA was used to differentiate abundant from non-abundant annotations. The limit of quantification (LOQ) was defined as 10 x the standard deviation of the noise in the UV chromatogram and was determined to be $3 \times 10^1$ μAU · min. Since the peptides had large differences in molar extinction coefficient (from $2 \cdot 10^3$ to $3 \cdot 10^5$ L $Mol^{-1}$ $cm^{-1}$), the LOQ was not expressed in μM for individual peptides.

### 2.11. Reproducibility

The individual hydrolysates and mixed hydrolysates were injected in four replicates. The repeatability of automatically annotated peptides was expressed as the percentage of unique peptides that were annotated similarly in all 4 replicates. The repeatability was calculated for peptides above the average LOA and for peptides between the LOD and LOA. The standard deviations over the total UV area, annotated UV area and absolute peptide concentrations were calculated based on the individual α-LA hydrolysate. To calculate the error on the concentration, annotations were used that were annotated similarly in all four replicates.

### 2.12. Tools to assess the completeness of peptide annotation and quantification

The completeness of the peptide analyses was evaluated by calculating the amino acid sequence coverages, peptide sequence coverages, protein recoveries and molar sequence coverages, as previously introduced by Butré et al. [20].

The amino acid sequence coverage, also used in proteomics [28], was calculated by dividing the number of unique amino acids

annotated in the peptides by the total number of amino acids in the protein sequence (Equation (4)).

$$\textit{Amino acid sequence coverage}\ [\%] = \frac{\#\ \textit{unique annotated amino acids}}{\#\ \textit{amino acids in protein sequence}} \cdot 100\,\% \quad (4)$$

When a peptide is annotated, other peptides should be present that cover the amino acids directly before and after this peptide. When this is not the case, the amino acids that should be covered form a 'missing' sequence. Moreover, a certain unique amino acid could be covered by multiple peptides. A 100% amino acid sequence coverage does therefore not necessarily imply that all peptides in the hydrolysate are identified. To include both aspects in the sequence coverage, the peptide sequence coverage was calculated. This was calculated by dividing the number of unique annotated peptides by the number of expected peptides (Equation (5)).

$$\textit{peptide sequence coverage}\ [\%] = \frac{\#\ AA\ (\textit{annotated peptides})}{\#\ AA\ (\textit{annotated peptides}) + \#\ AA\ (\textit{missing peptides})} \cdot 100\,\% \quad (5)$$

To assess the completeness of quantification, the concentration of the peptides has to be considered. Based on the law of mass conservation, all the amino acids [μM] in the initial substrate should end up after hydrolysis as free amino acids, peptides or remaining intact protein. The protein recovery was calculated to assess to what extent the measured average AA concentrations matched the injected protein concentration (Equation (6)).

$$\textit{Protein recovery}\ [\%] = \left( \frac{\left( \frac{\sum C_n}{\#AA_{protein}} \right)}{C_0} \right) \cdot 100\,\% \quad (6)$$

where $C_n$ [μM] is the concentration of each individual AA (n) in the protein sequence, and $\#AA_{protein}$ is the number of amino acids in the initial protein and $C_0$ [μM] is the initially injected protein concentration. At last, the molar sequence coverage was calculated, which considers that certain parts of the protein sequence might be over-quantified whereas other regions are quantified with a lower concentration compared to the expected concentration.

The molar sequence coverage represents to what extent the peptides that cover an amino acid in the protein sequence are quantified relative to the injected molar concentration [μM] (Equation (7)).

$$\textit{Molar sequence coverage}\ [\%] = \left( 1 - \frac{\sqrt{\frac{\sum (C_n - C_0)^2}{(\#AA_{protein} - 1)}}}{C_0} \right) \cdot 100\,\% \quad (7)$$

where $C_n$ [μM] is the concentration of each individual AA (n) in the protein sequence, $C_0$ [μM] is the initially injected protein concentration and $\#AA_{protein}$ is the number of amino acids in the initial protein.

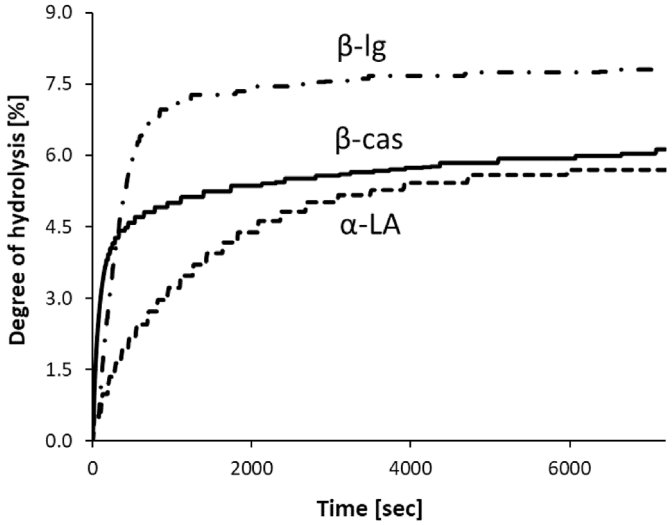



**Fig. 2.** Degree of hydrolysis ($DH_{stat}$) versus time of 1% α-LA (- - -), β-LG (— ·) and β-cas (—) hydrolysed with bovine trypsin.

## 3. Results & discussion

### 3.1. Characterisation of the starting protein isolates and hydrolysates

The protein isolates of α-LA, β-LG and β-cas had a protein content [w protein/w DM] of 93%, 96% and 90% and a protein purity of 90%, 100% and 90% respectively (Table 1). The remaining 10% of protein in the α-LA isolate was identified as β-LG with UPLC-MS. The remaining proteins in the β-cas had masses between 25 and 35 kDa. Analysis of the intact proteins showed that β-LG was equally present as genetic variant A or B. Literature indicates that the methionine residue [M90] in α-LA is prone to oxidation. Uniprot indicated that the serine residues [S15,S17,S18,S19,S35] in β-cas were phosphorylated, which was confirmed by the intact protein mass in the MS. The protein isolates of α-LA, β-LG and β-cas were hydrolysed with bovine trypsin and reached a $DH_{stat,max}$ of respectively 5.6% (±0.1%), 7.7 (±0.2%), and 6.2% (±0.1%) (Fig. 2). The $DH_{stat,max}$ values were in line with previous results under the same conditions [30].

### 3.2. Manual peptide identification

The manual annotation of peptides in individual protein hydrolysates of α-LA, β-LG and β-cas yielded 27, 39, and 24 unique annotated peptides respectively (Supplementary Tables S1–3). Of these 90 peptides, 67 peptides resulted from specific hydrolysis for trypsin and 23 peptides from semi-specific hydrolysis, i.e. either the peptide bond on the C- or on the N-terminal side that was hydrolysed did not match trypsin specificity. The methionine residue in α-LA was present in both the oxidised and in the non-oxidised form. The serine residues [S15, S17, S18, S19 and S35] in β-cas were always phosphorylated. The traditional amino acid sequence coverage was 100% for all three substrates. The peptide sequence coverages, which take into account peptides that should be present based on the other formed peptides present, were respectively 91%,

97%, and, 97% for the α-LA, β-LG and β-cas, respectively. These sequence coverage values were comparable to previous sequence coverages of studies of Butré and Deng, who used the same manual approach [20,30].

### 3.3. Automated peptide annotation with UNIFI - default run

In the **default** analysis of the α-LA hydrolysate 2034 unique *m/z* signals were identified. Of these *m/z* values, 843 were not matched to a peptide sequences, 279 were recognised by UNIFI as in-source fragments and 912 were tentatively matched to peptide sequences. Among these 912 annotations were 56 peptide annotations with adducts ($H_2O$, $NH_4$) and 157 non-unique annotations. The remaining 699 unique peptide annotations had an absolute average mass error of 47 ± 31 ppm and an average MS/MS fragment recovery of 4 ± 14%. Of the 27 manually annotated α-LA peptides, 26 peptides were also identified in the default analysis. The number of annotations was clearly higher in the default automated analysis than in the manual analysis. The question arises or all the (new) annotations should be considered valid and how to create confidence in the identified peptides.

### 3.4. Evaluation of different lock mass components

The 26 manually confirmed α-LA peptides in the **default** analysis had an average absolute mass error of 12 ± 6 ppm and showed a negative dependency with mass with a slope of −0.005 ppm $Da^{-1}$ (Fig. 3). Therefore, to ensure that large peptides were included in the analysis a high mass error threshold (100 ppm) was used. At the same time, the high mass error threshold would result in multiple tentative peptides that could be matched with a parent ion mass within the mass error, and potentially result in wrong annotations. To reduce the increase in mass error with increasing peptide mass, different lock mass combinations were evaluated using a β-LG hydrolysate. Without a lock mass, the peptides in the β-LG hydrolysate yielded an average absolute mass error of 5.1 ppm, with a maximum of 12.1 ppm and a slope in the mass residuals of −0.0052 ppm $Da^{-1}$ (Table 2). Analysis of the same hydrolysate with lock mass yielded an average absolute mass error of 5.1 ppm for Insulin [3+], 2.0 ppm for Angiotensin [1+] and 2.0 ppm for LeuEnk [1+]. The insulin [3+] was not effective as a lock mass, probably because the *m/z* was higher than the majority of the peptides. The average mass error was efficiently decreased when Angiotensin [1+] or LeuEnk [1+] was used, but the mass error still showed a dependency with increasing mass, respectively −0.0017 ppm $Da^{-1}$ for Angiotensin and −0.0015 ppm $Da^{-1}$ for LeuEnk. Therefore the lock mass processing was performed with two components. Processing the same data with LeuEnk [1+] and Insulin [3+] decreased the average absolute mass error to 1.4 ppm and reduced the slope to −0.0007 ppm $Da^{-1}$. It was observed that the insulin was mainly present in the [M+4] [M+5] and [M+6] state in the lockspray spectrum, whereas the charge state of interest [M+3] comprised only ~ 0.2% [MS intensity] of the mass spectrum's signal intensity. Therefore, the charge state was altered by changing the solvent conditions and the addition of diethylamine. The relative abundance of [M+3] increased from 0.2% to 90% of the mass spectrum's signal intensity (Supplementary Figs. S1–2). This change in solvent composition yielded a final average mass error of 1.2 ± 1.1 ppm for the β-LG hydrolysate. Processing of the other samples showed a comparable average mass error of respectively 1.0 ± 0.9 ppm for α-LA and 1.5 ± 1.9 ppm for β-cas. For the manually confirmed α-LA peptides, the slope was reduced from −0.005 to −0.0007 ppm $Da^{-1}$ using the optimised lock mass combination of Leu-Enk with Insulin (Fig. 3). Based on this, the mass error threshold was set at 10 ppm for the analyses with double lock mass in further sections.

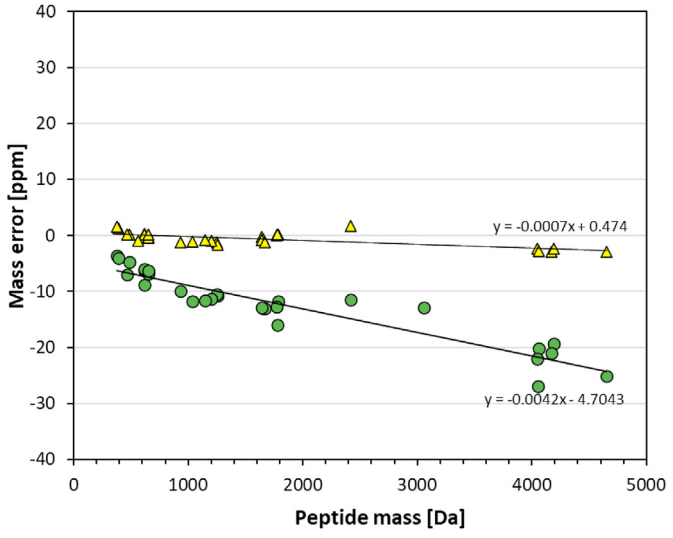



**Fig. 3.** The mass error [ppm] plotted for manually confirmed α-LA peptides as function of the peptide mass [Da]. The green dots (●)represent α-LA peptides in the default run without lock mass, the yellow triangles (▲) represent α-LA peptides analysed with the lock mass combination of LeuEnk [1+] and Insulin [3+] with diethylamine. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

### 3.5. Separate signals from the noise based on the LOD/LOA

The limit of detection (LOD) was on average 1.6 $\cdot 10^5$ ±1.7 $\cdot 10^5$ counts, and was used to filter the signals from the noise. The limit of annotation (LOA) was on average 2.4 $\cdot 10^6$ ± 3.5 $\cdot 10^6$ counts which is~15x higher than the average LOD (Fig. 4). Without applying cut-offs for the LOD or LOA, analysis of the α-LA hydrolysate with double lock mass yielded 288 unique annotated α-LA peptides, 389 β-LG peptides and 343 β-cas peptides. The MS intensities of 599 of these annotations were below the LOD and were therefore excluded, leaving 421 peptides with MS intensities above LOD. It was observed that for these remaining 421 peptides, part of the annotations were not confirmed with a sufficient number of b/y fragments. Therefore, a filter was introduced to include only annotations with sufficient identified b/y fragments. Of the total MS intensity above the LOA, 90.0% was attributed to peptides that passed the applied filter, which resulted in the identification of 73 peptides, of which 43 from α-LA, 21 from β-LG peptides and 9 from β-cas.

### 3.6. Removal of in-source fragments

For some peptides, formed in-source fragments were recognised as such by UNIFI or were excluded since the MS intensity was below the LOD. In other cases, these fragments were incorrectly annotated as unique peptide. From the 73 annotated peptides in the α-LA hydrolysate, 6 in-source fragments (~10% of total) were incorrectly identified as peptide by UNIFI and therefore removed in PeptQuant.

### 3.7. Peptide identification in the individual hydrolysates

The analysis of the individual α-LA, β-LG and β-cas hydrolysates using the proposed data-processing method yielded in total 77 peptides above the LOA, (26 of α-LA, 29 of β-LG and 22 of β-cas) and 56 peptides between the LOD and LOA (14 α-LA, 32 β-LG and 10 β-

**Table 2**
The mass error of β-LG peptides analysed with different lock mass components.

| Lock mass [charge state] | Maximum observed mass error (absolute, ppm) | Average mass error (absolute, ppm) | Trendline slope (ppm/Da) |
|---|---|---|---|
| *No lock mass* | 12.1 | 5.1 | −0.0052 |
| Insulin [3+] | 7.8 | 5.1 | −0.0013 |
| Angiotensin [1+] & Insulin [3+] | 5.5 | 2.1 | −0.0001 |
| LeuEnk [1+] | 7.4 | 2.0 | −0.0015 |
| Angiotensin [1+] | 7.3 | 2.0 | −0.0017 |
| LeuEnk [1+] & Angiotensin [1+] | 7.3 | 1.8 | −0.0011 |
| LeuEnk [1+] & Insulin [3+] | 4.8 | 1.4 | −0.0007 |
| LeuEnk [1+] & Insulin [3+] (with diethylamine) | 4.8 | 1.2 | −0.0008 |

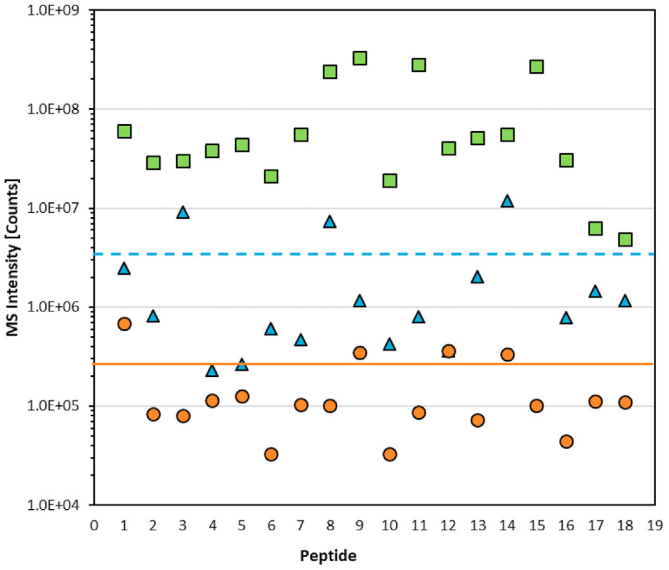



**Fig. 4.** MS intensities of α-LA peptides in the highest dilution at which the MS signals were detected (●) and the peptides were annotated with MS/MS fragments (▲). The MS intensities as in the individual α-LA hydrolysate were also indicated (■). The average LOD (—) and average LOA (– –) were calculated based on the average of the individual peptides.

cas) (Supplementary Table S1). The total MS intensity of these peptides was described for 99 ± 0.6% by peptides above the LOA and 1 ± 0.6% by the peptides between the LOD and LOA. 97% of the manually identified peptides were also identified with the automated annotation. The automated analysis identified in total 4 additional peptides above the LOA and 40 additional peptides below the LOA that had not been found in the manual analysis.

### 3.8. Repeatability of peptide identification in the individual hydrolysates

The 77 peptides identified above the LOA were consistently annotated in the four replicates with a repeatability of 100% (Fig. 6). For the 56 peptides between the LOD and LOA, 50% of the peptides was annotated consistently in all four replicates. The LOA could therefore be used as an MS intensity threshold to describe the confident and repeatable part of the annotations. The repeatability of peptides below the LOA implies that a peptide could be annotated below the LOA, but that the inclusion or exclusion is not as consistent as for peptides above the LOA. The repeatability of peptide identification in this work (100% > LOA, 50% < LOA) is higher than the repeatability in peptides for proteomics purposes. In work of Tabb and co-workers, a typical repeatability of 35–60% was described between two technical replicates [33].

### 3.9. Completeness of peptide identification in the individual hydrolysates

Using the automated annotation, the amino acid sequence coverages were 100 ± 0% for α-LA and β-LG; and 99 ± 0% for β-cas for the individual hydrolysates (Table 3). The peptide sequence coverages were 91.4 ± 1% for α-LA, 95.4 ± 2% for β-LG and 88.8 ± 1% for β-cas. The peptide sequence coverages of α-LA and β-LG were in line with the manual peptide sequence coverage (91.0 for α-LA & 97.0% for β-LG). The peptide sequence coverage of the automated β-cas analysis (88.8 ± 1) was lower than that with the manual analysis (97%). This is probably due to 4 'missing' peptides, that were expected based on the 10 peptides that were annotated additionally to the manually annotated peptides. The variation in annotation of peptides between replicates did not result in substantial variation in the amino acid and peptide sequence coverages between the replicates. In the analysis of the individual α-LA hydrolysate, also β-LG was identified, with an amino acid sequence coverage of 98 ± 0% and a peptide sequence coverage of 88 ± 2%. This gives a first insight that the identification of peptides in a mixture of two proteins works. The amino acid sequence coverages in the analysis of the individual hydrolysates were higher than the typical amino acid sequence coverages reported in proteomics, which were typically below 50% [34,35].

### 3.10. Reproducibility of peptide quantification

The identified peptides were quantified with the corresponding UV areas at 214 nm and the predicted molar extinction coefficients. Over the four replicates of the individual α-LA hydrolysate, the average relative standard deviation of an integrated UV peak was 5%. The relative standard deviation of the total UV area in a chromatogram was 6.4%. The relative standard deviation of the calculated peptide concentrations was 3.7% for the absolute concentration of peptides above the LOA and 10.2% for that of peptides between the LOA and LOD. The relative standard deviations are comparable to those obtained with quantification techniques that require metabolic or chemical labelling (<10% RSD), and lower than those obtained with (relative) label-free quantification approaches in proteomics (10–30% RSD) [36].

### 3.11. Quantification of the peptides in the individual hydrolysates

The total UV area in the chromatograms of α-LA, β-LG and β-cas were respectively 105 ± 7%, 104 ± 4% and 112 ± 1% of the expected UV. The UV area attributed to the automated analysed peptides was 100 ± 5% of the expected amount of UV for the individual α-LA hydrolysate, 99 ± 4% for β-LG and 99 ± 2% for β-cas. These values indicate that the amount of UV included in the analysis was in line with the expected amount based on protein concentrations. The protein recoveries yielded comparable values for the manual and automated analysis (Table 3). Peptide losses due to insolubility

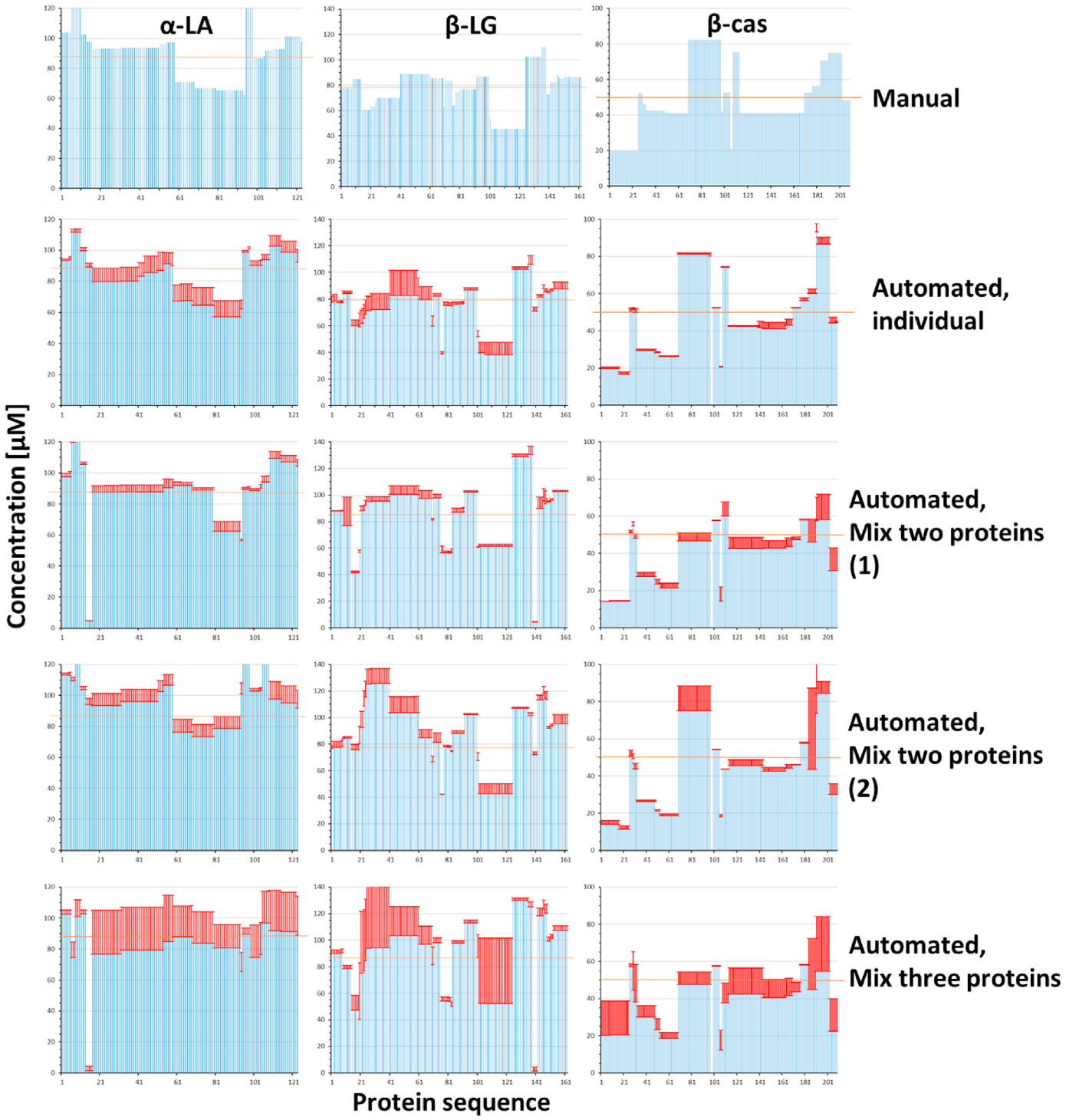



**Fig. 5.** Concentration of all amino acids $C_n$ for α-LA (left), β-LG (middle) and β-cas (right) of the manual analysis (row 1), individual protein hydrolysates analysed automatically (row 2), the 1:1 mixtures: α-LA in α-LA + β-LG (row 3, left), β-LG in α-LA + β-LG (row 3, middle), β-cas in α-LA + β-cas (row 3, right), α-LA in α-LA + β-cas (row 4, left), β-LG in β-LG + β-cas (row 4, middle), and β-cas in β-LG + β-cas (row 4, right), and the mixture of α-LA, β-LG and β-cas (row 5). The standard deviation shown was calculated over the four injections analysed with the automated annotation method. The orange line indicates the initial protein concentration in µM. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

during sample preparation or peptide instability during the analysis seem to be neglectable for these hydrolysates from relatively 'clean' protein isolates. The protein recovery values do not indicate whether certain regions of the parental protein were over-quantified or under-estimated by the peptide composition. Therefore, the concentration of the amino acids in the peptides were plotted against the protein sequence (Fig. 5). To describe the completeness of this plot with a quantitative parameter, the molar sequence coverages were calculated, which were 80%, 64% and 62% for α-LA, β-LG and β-cas using the manual input and 81 ± 2%,

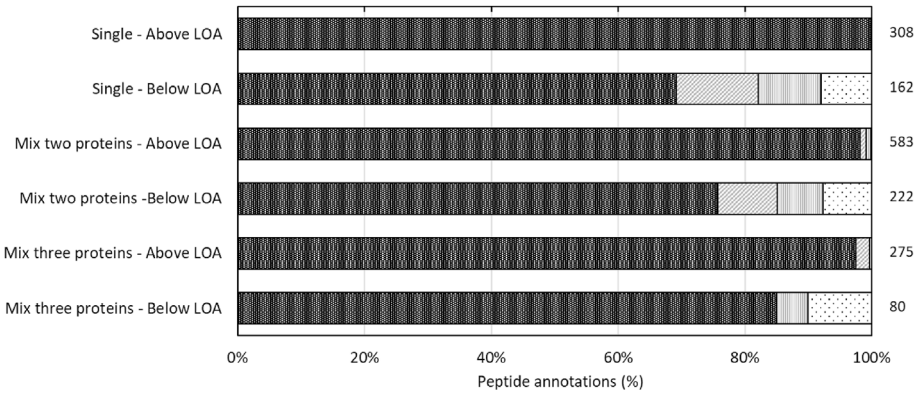



**Fig. 6.** Repeatability of peptide identifications in the individual hydrolysate and mixtures. The percentage (%/%) of α-LA, β-LG and β-cas annotations in 4/4 replicates (▩), 3/4 replicates (▨), 2/4 replicates (▥) or 1/4 replicates (□). The total number of annotations is given.

**Table 3**
Sequence coverages of α-LA, β-LG and β-cas in the individual hydrolysates and the mixtures. The standard deviation is calculated over the four replicate injections for the hydrolysates analysed with the automated peptide annotation method.

| Protein | Sample | AA sequence coverage [%] | Peptide sequence coverage [%] | Protein recovery [%] | Molar sequence coverage [%] |
|---|---|---|---|---|---|
| α-LA | Manual | 100% | 91% | 103% | 80% |
| | Automated | 100 ± 0% | 91 ± 1% | 101 ± 4% | 81 ± 2% |
| | + β-LG | 100 ± 0% | 93 ± 1% | 100 ± 1% | 75 ± 2% |
| | + β-cas | 100 ± 0% | 91 ± 1% | 110 ± 4% | 77 ± 2% |
| | + β-LG + β-cas | 100 ± 0% | 92 ± 3% | 106 ± 11% | 76 ± 8% |
| β-LG | Manual | 100% | 97% | 98% | 64% |
| | Automated | 100 ± 0% | 95 ± 2% | 101 ± 4% | 77 ± 1% |
| | + α-LA | 100 ± 0% | 95 ± 2% | 101 ± 1% | 67 ± 1% |
| | + β-cas | 100 ± 0% | 88 ± 1% | 114 ± 3% | 63 ± 2% |
| | + α-LA + β-cas | 100 ± 0% | 96 ± 1% | 122 ± 11% | 67 ± 5% |
| β-cas | Manual | 100% | 97% | 99% | 62% |
| | Automated | 99 ± 0% | 89 ± 1% | 94 ± 1% | 56 ± 0% |
| | + α-LA | 99 ± 0% | 88 ± 1% | 76 ± 4% | 63 ± 1% |
| | + β-LG | 99 ± 0% | 88 ± 1% | 85 ± 4% | 52 ± 3% |
| | + α-LA + β-LG | 99 ± 0% | 89 ± 1% | 82 ± 5% | 67 ± 6% |

77 ± 1% and 56 ± 0% for β-cas using the automated input (Table 3). The molar sequence coverages for α-LA and β-LG were in line with reported molar sequence coverages reported for BLP hydrolysates (70 ± 10% by Butré et al.), and for tryptic hydrolysates (79 ± 6% by Deng et al.), [20,30]. The molar sequence coverage for β-cas was lower than previously reported values, but was in the study of Deng et al. also mentioned to be below the average [30].

### 3.12. Peptide identification in the hydrolysate mixtures

In the mixtures there were no peptides annotated above the LOA that were not identified in the individual hydrolysates. However, the reverse is not true. Of the 308 peptide annotations above the LOA combined in the four individual hydrolysate replicates, 95% were similarly annotated in the mixtures of two proteins, and 89% were similarly annotated in the mixture of three proteins (Fig. 7). Peptide [14-16] of α-LA was one of the peptides above the LOA that disappeared upon mixing with β-LG and with β-LG + β-cas. In this particular case, the parent ion *m/z* was above the LOD in all samples, but the minimum requirement of 2 b/y fragments was not met. This could probably be a result of the co-elution with peptide [139−141] of β-LG. For the 162 peptide annotations below the LOA in the individual hydrolysates, 64% were similarly annotated in the mixtures of two proteins and 47% were similarly annotated in the mixture of three proteins. Mixing the hydrolysates resulted in a substantial loss of annotations between the LOD and LOA. The remaining peptides after mixing between the LOD and LOA, showed a relative improvement in the repeatability between replicates from 69% in the individual hydrolysates to 76% for the mixtures with two proteins to 85% for the mixtures with three proteins (Fig. 6). The repeatability of the annotated peptides above the LOA decreased from 100% in the individual hydrolysates to 98% for the mixture with two proteins and 97% for the mixture with three proteins (Fig. 6). For peptides above the LOA, the variation introduced by mixing three proteins (11%) was slightly higher than the variation between replicates (3%). For peptides between the LOD and LOA, the variation introduced by mixing three proteins (53%) was in line with the variation between replicates (28%). The repeatability of peptide annotations between the LOD and the LOA is in line with repeatability in proteomics studies [33]. In the end, the amino acid sequence coverages were identical for the mixed hydrolysates and the individual hydrolysates (Table 3). The peptide sequence coverages did not change significantly upon mixing, except for β-LG in the mixture with β-cas.

### 3.13. The effect of co-elution on peptide quantification in the mixed hydrolysates

The individual protein hydrolysates were mixed to evaluate the effect of co-elution on the quantification of individual peptides (Supplementary Fig. S3). The molar sequence coverage for mixed hydrolysates was on average 4 ± 8% lower than for the individual hydrolysates. The average standard deviation in molar sequence coverage over the four replicates increased upon mixing from ±2%

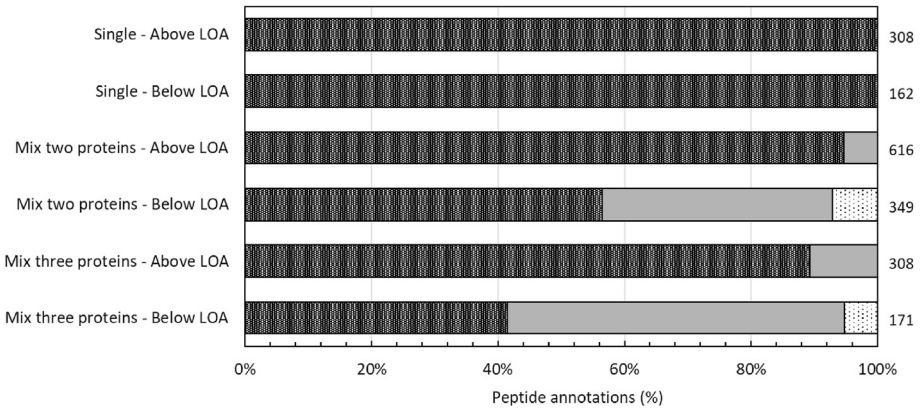



**Fig. 7.** The effect of mixing on peptide identifications. The percentage (%/%) of α-LA, β-LG and β-cas annotations in the mixed hydrolysates annotated similarly as in the individual hydrolysates (▩), missing in the mixed hydrolysates (■) or appearing in the mixed hydrolysates (□) The total number of annotations is given.

to ±8%. In mixtures of two hydrolysates, on average ~40% of the total annotated UV area was linked to co-eluting peptides. This value increased to 57% for the mixture with three hydrolysates. For 21 co-eluting peptides, the effect of co-elution on quantification was analysed by comparing the UV and MS signals as well as calculated concentrations in the individual and mixed hydrolysate. The total UV areas for each peptide in the individual differed on average 7% from that in the mixed hydrolysate. For some co-eluting peptides there was no ion suppression (0%), while for others there was (max 48%, average 21%). When ion suppression occurred, it affected all co-eluting peptides at that RT similarly. To calculate the concentration of co-eluting peptides, the UV area is divided over the peptides assuming that the peptides have a more or less similar ionisation efficiency (i.e. MS intensity per amount of peptide). At short RT the ionisation efficiency is typically lower ($1 \cdot 10^5$ Counts/μM) than at higher RT ($1 \cdot 10^7$ Counts/μM) (Supplementary Fig. S4). However, for each set of co-eluting peptides, the ionisation efficiencies differed maximally with a factor 2–3 between each set of co-eluting peptides. In the chromatogram maximum variation in ionisation efficiency of a factor 5 to 8 were observed at close retention times, although these were not co-eluting (Supplementary Fig. S4). The differences in ionisation efficiency of co-eluting peptides resulted in a difference in calculated concentration of on average 37 ± 21% compared to the concentrations in the individual hydrolysates. This error in calculated concentration of co-eluting peptides was substantially larger than the (relative) standard deviation over concentrations in replicate injections (6.3%) for the studied peptides.

## 4. Conclusion

A method was evaluated for reproducible automated annotation and absolute quantification of peptides. It was shown that using the LOA a distinction could be made between peptides with 100% repeatability in single hydrolysates (99% of the MS intensity assigned to peptides) and those with lower repeatability (50%, and ~1% of the MS intensity). For peptides above the LOA, mixing the hydrolysates resulted in an 11% loss of identified peptides and a 3% decrease in repeatability. The increased number of co-eluting peptides due to mixing had minor effects on amino acid, peptide or molar sequence coverage. However, calculated concentrations of individual co-eluting peptides in mixed systems varied on average 37%. The proposed approach enables automation of the hydrolysate compositional analysis while maintaining confidence in the repeatability of peptide annotations and completeness of the analysis. In addition, it opens up new possibilities for future research towards more complex protein hydrolysates.

## CRediT authorship contribution statement

**Gijs J.C. Vreeke:** Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing – original draft. **Wouter Lubbers:** Investigation. **Jean-Paul Vincken:** Supervision, Writing – review & editing. **Peter A. Wierenga:** Conceptualization, Software, Supervision, Writing – review & editing.

## 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.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2022.339616.

## References

[1] G. Mamone, G. Picariello, S. Caira, F. Addeo, P. Ferranti, Analysis of food proteins and peptides by mass spectrometry-based techniques, J. Chromatogr., A 1216 (43) (2009) 7130–7142, https://doi.org/10.1016/j.chroma.2009.07.052.
[2] J.-S. Kim, M.E. Monroe, D.G. Camp, R.D. Smith, W.-J. Qian, In-source fragmentation and the sources of partially tryptic peptides in shotgun proteomics, J. Proteome Res. 12 (2) (2013) 910–916, https://doi.org/10.1021/pr300955f.
[3] R.A. Zubarev, P. Håkansson, B. Sundqvist, Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements, Anal. Chem. 68 (22) (1996) 4060–4063, https://doi.org/10.1021/ac9604651.
[4] T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm, BMC Bioinf. 7 (1) (2006) 234, https://doi.org/10.1186/1471-2105-7-234.
[5] A.W.T. Bristow, K.S. Webb, Intercomparison study on accurate mass measurement of small molecules in mass spectrometry, J. Am. Soc. Mass Spectrom. 14 (10) (2003) 1086–1098, https://doi.org/10.1016/S1044-0305(03)00403-3.
[6] R. Zubarev, M. Mann, On the proper use of mass accuracy in proteomics, Mol. Cell. Proteomics 6 (3) (2007) 377–381, https://doi.org/10.1074/mcp.M600380-MCP200.
[7] A.M. Frank, M.M. Savitski, M.L. Nielsen, R.A. Zubarev, P.A. Pevzner, De novo peptide sequencing and identification with precision mass spectrometry, J. Proteome Res. 6 (1) (2007) 114–123, https://doi.org/10.1021/pr060271u.
[8] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis: Int. J. 20 (18) (1999) 3551–3567, https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551:AID-ELPS3551>3.0.CO;2-2.
[9] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, J. Am. Soc. Mass Spectrom. 5 (11) (1994) 976–989, https://doi.org/10.1016/1044-0305(94)80016-2.
[10] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann,

Andromeda: a peptide search engine integrated into the MaxQuant environment, J. Proteomce Res. 10 (4) (2011) 1794–1805, https://doi.org/10.1021/pr101065j.
[11] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, Anal. Chem. 74 (20) (2002) 5383–5392, https://doi.org/10.1021/ac025747h.
[12] T. Koenig, B.H. Menze, M. Kirchner, F. Monigatti, K.C. Parker, T. Patterson, et al., Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics, J. Proteomce Res. 7 (9) (2008) 3708–3717, https://doi.org/10.1021/pr700859x.
[13] B. Cooper, The problem with peptide presumption and low Mascot scoring, J. Proteomce Res. 10 (3) (2011) 1432–1435, https://doi.org/10.1021/pr101003r.
[14] M. Berg, A. Parbel, H. Pettersen, D. Fenyö, L. Björkesten, Reproducibility of LC-MS-based protein identification, J. Exp. Bot. 57 (7) (2006) 1509–1514, https://doi.org/10.1093/jxb/erj139.
[15] N. Delmotte, M. Lasaosa, A. Tholey, E. Heinzle, A. van Dorsselaer, C.G. Huber, Repeatability of peptide identifications in shotgun proteome analysis employing off-line two-dimensional chromatographic separations and ion-trap MS, J. Separ. Sci. 32 (8) (2009) 1156–1164, https://doi.org/10.1002/jssc.200800615.
[16] T.M. Annesley, Ion suppression in mass spectrometry, Clin. Chem. 49 (7) (2003) 1041–1044, https://doi.org/10.1373/49.7.1041.
[17] B.C. Collins, C.L. Hunter, Y. Liu, B. Schilling, G. Rosenberger, S.L. Bader, et al., Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry, Nat. Commun. 8 (1) (2017) 291, https://doi.org/10.1038/s41467-017-00249-5.
[18] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster, Quantitative mass spectrometry in proteomics: a critical review, Anal. Bioanal. Chem. 389 (4) (2007) 1017–1031, https://doi.org/10.1007/s00216-007-1486-6.
[19] V. Brun, A. Dupuis, A. Adrait, M. Marcellin, D. Thomas, M. Court, et al., Isotope-labeled protein standards: toward absolute quantitative proteomics, Mol. Cell. Proteomics 6 (12) (2007) 2139–2149, https://doi.org/10.1074/mcp.M700163-MCP200.
[20] C.I. Butré, S. Sforza, H. Gruppen, P.A. Wierenga, Introducing enzyme selectivity: a quantitative parameter to describe enzymatic protein hydrolysis, Anal. Bioanal. Chem. 406 (24) (2014) 5827–5841, https://doi.org/10.1007/s00216-014-8006-2.
[21] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, J. Agric. Food Chem. 55 (14) (2007) 5445–5451, https://doi.org/10.1021/jf070337l.
[22] Y. Deng, H. Gruppen, P.A. Wierenga, Comparison of protein hydrolysis catalyzed by bovine, porcine, and human trypsins, J. Agric. Food Chem. 66 (16) (2018) 4219–4232, https://doi.org/10.1021/acs.jafc.8b00679.
[23] A. Bodin, X. Framboisier, D. Alonso, I. Marc, R. Kapel, Size-exclusion HPLC as a sensitive and calibrationless method for complex peptide mixtures quantification, J. Chromatogr., B: Anal. Technol. Biomed. Life Sci. 1006 (2015) 71–79, https://doi.org/10.1016/j.jchromb.2015.09.035.
[24] A.B. Nongonierma, C. Mazzocchi, S. Paolella, R.J. FitzGerald, Release of dipeptidyl peptidase IV (DPP-IV) inhibitory peptides from milk protein isolate (MPI) during enzymatic hydrolysis, Food Res. Int. 94 (2017) 79–89, https://doi.org/10.1016/j.foodres.2017.02.004.
[25] D. Dupont, G. Mandalari, D. Molle, J. Jardin, J. Léonil, R.M. Faulks, et al., Comparative resistance of food proteins to adult and infant *in vitro* digestion models, Mol. Nutr. Food Res. 54 (6) (2010) 767–780, https://doi.org/10.1002/mnfr.200900142.
[26] S.R.C.K. Rajendran, B. Mason, C.C. Udenigwe, Peptidomics of peptic digest of selected potato tuber proteins: post-translational modifications and limited cleavage specificity, J. Agric. Food Chem. 64 (11) (2016) 2432–2437, https://doi.org/10.1021/acs.jafc.6b00418.
[27] S. Gu, N. Chen, Y. Zhou, C. Zhao, L. Zhan, L. Qu, et al., A rapid solid-phase extraction combined with liquid chromatography-tandem mass spectrometry for simultaneous screening of multiple allergens in chocolates, Food Control 84 (2018) 89–96, https://doi.org/10.1016/j.foodcont.2017.07.033.
[28] B. Meyer, D.G. Papasotiriou, M. Karas, 100% protein sequence coverage: a modern form of surrealism in proteomics, Amino acids 41 (2) (2011) 291–310, https://doi.org/10.1007/s00726-010-0680-6.
[29] Y. Deng, C.I. Butré, P.A. Wierenga, Influence of substrate concentration on the extent of protein enzymatic hydrolysis, Int. Dairy J. 86 (2018) 39–48, https://doi.org/10.1016/j.idairyj.2018.06.018.
[30] Y. Deng, F. van der Veer, S. Sforza, H. Gruppen, P.A. Wierenga, Towards predicting protein hydrolysis by bovine trypsin, Process Biochem. 65 (2018) 81–92, https://doi.org/10.1016/j.procbio.2017.11.006.
[31] J. Adler-Nissen, Enzymic Hydrolysis of Food Proteins, Elsevier Applied Science Publishers, London, UK, 1986.
[32] C.I. Butré, P.A. Wierenga, H. Gruppen, Influence of water availability on the enzymatic hydrolysis of proteins, Process Biochem. 49 (2014) 1903–1912, https://doi.org/10.1016/j.procbio.2014.08.009.
[33] D.L. Tabb, L. Vega-Montoto, P.A. Rudnick, A.M. Variyath, A.-J.L. Ham, D.M. Bunk, et al., Repeatability and reproducibility in proteomic identifications by liquid chromatography– tandem mass spectrometry, J. Proteomce Res. 9 (2) (2010) 761–776, https://doi.org/10.1021/pr9006365.
[34] S. Garza, M. Moini, Analysis of complex protein mixtures with improved sequence coverage using (CE–MS/MS)n, Anal. Chem. 78 (20) (2006) 7309–7316, https://doi.org/10.1021/ac0612269.
[35] K. Krug, A. Carpy, G. Behrends, K. Matic, N.C. Soares, B. Macek, Deep coverage of the escherichia coli proteome enables the assessment of false discovery rates in simple proteogenomic experiments, Mol. Cell. Proteomics 12 (11) (2013) 3420–3430, https://doi.org/10.1074/mcp.M113.029165.
[36] W.X. Schulze, B. Usadel, Quantitation in mass-spectrometry-based proteomics, Annu. Rev. Plant Biol. 61 (2010) 491–516, https://doi.org/10.1146/annurev-arplant-042809-112132.