Published in final edited form as:

Methods Mol Biol. 2018 ; 1719: 223–240. doi:10.1007/978-1-4939-7537-2_15.

Milk Peptidomics to Identify Functional Peptides and for Quality Control of Dairy Products

David Dallas and Søren Drud Nielsen

Abstract

Human milk and dairy products are important parts of human nutrition. In addition to supplying - nutrients, milk proteins contain fragments—peptides—with important biological functions that are released during processing or digestion. Besides their potential functional relevance, peptides released during processing can be used as markers of ripening stage or product deterioration. Hence, identification and quantification of peptides in milk can be used to assay potential health benefits or product quality. This chapter describes how to extract, identify, and analyze peptides within breast milk, dairy products, and dairy digestive samples. We describe how to analyze extracted peptides with liquid chromatography-mass spectrometry, to use software to identify peptides based on database searching, and to extract peak areas for relative quantification of each peptide. We describe methods for data analysis, including predicting which enzymes are responsible for protein cleavage, identifying the site specificity of protein breakdown, mapping - identified peptides to known bioactive peptides, and applying models to predict novel functional peptides.

Keywords

Peptidomics; Milk; Mass spectrometry; Peptide; Bioactive; Identification; Quantification; X! Tandem; Skyline; Proteome Discoverer

1 Introduction

Human milk has evolved to match the newborn infant's nutritional needs. Beyond providing basic nutrients, milk also serves as a source of biologically active molecules with major impacts on the infant's health and development. Fragments of milk proteins produced by enzymatic cleavage—peptides—have an array of functions, including immunomodulation [1, 2], opioid-like activity [3, 4], and antimicrobial [5–7], antihypertensive [8], and - antioxidant actions [9]. These functions are often different from the functions of the parent protein. In effect, these peptides are functional units encrypted within the parent sequence. Most previous studies of bioactive milk peptides are based on peptides released from in vitro digestion of isolated milk proteins; thus they neither represent in vivo digestion nor identify which peptides are biologically relevant. Recently, we employed peptidomics to examine the in vivo gastric digestion of milk in infants. For this study, we used milk and gastric samples collected from tube-fed infants and demonstrated that milk protein digestion begins within the mother's mammary gland, where milk proteases release hundreds of peptides [10,11]. Within the infant gut, milk proteins are further digested into smaller fragments, producing hundreds of additional peptides [12]. Our bioinformatic analyses demonstrated that these

peptides derived from specific milk proteins at specific sites within the proteins, and that many matched known functional peptides with antimicrobial and immunomodulatory actions.

Peptidomics is used not only to study digestion of proteins—it has many uses in dairy science. Dairy scientists apply peptidomics to identify peptides in dairy products, including cheeses [13–16], yogurts [17], and kefir [18]. These peptidomics studies revealed that differences in starting materials, production techniques, and storage time resulted in major differences in the peptide profile. For example, peptidomic analysis of lactose-hydrolyzed ultra-high temperature heat-treated milk indicated that residual proteases in the lactase preparation caused protein degradation that resulted in bitter flavor development with storage time [19–21]. Peptidomics can also be used to monitor inter-individual variation among cows [22] and as a marker for their health status—we demonstrated that peptide profiles differed in milk from healthy and mastitic quarters (teats) within the same cow [23].

This chapter describes how to extract peptides from dairy products and their digesta, to analyze them by liquid chromatography (LC) coupled with mass spectrometry (MS), and to identify and quantify them by label-free relative quantification (overview in Fig. 1). Several different MS instruments are appropriate for peptidomics, including Orbitrap, Q-TOF, and others [24]. Label-free relative quantification can be accomplished using peptide signal intensities. The ion-signal intensity approach uses extracted chromatographic areas to compare peptide abundances across samples. The software used for identification and quantification depends on the instrument used as the output file type differs among instruments from different vendors. This chapter describes two methods for identification and label-free relative quantification of peptides. The first method uses X! Tandem and Skyline, which are two popular noncommercial programs (Subheadings 3.7 and 3.8). Within Skyline, we create a library of X! Tandem-identified peptides across all samples. This library is used to match peaks in other samples based on retention time and mass. This approach identifies peptides that may be present in the MS data as parent molecule masses but not identified in that sample by X! Tandem. The second method uses the Proteome Discoverer software from Thermo Fisher (Subheading 3.9). This method determines only peptide abundance from tandem-identified peptides within the sample as no library-based peak searching is available. Selecting which program to use depends on the specific research needs, sample type, and instrument combination. This chapter includes protocols for further data analysis using bioinformatics techniques. The first tool visualizes from where identified peptides derive across the parent protein sequence. A homology search tool identifies the peptides that match previously identified functional peptides. Several tools use structure-function relationships to predict the likelihood that each identified peptide is functional. Two tools predict which proteases in a sample were responsible for protein cleavages based on the peptidomic data. Finally, we suggest that the peptidomics results of any study should be uploaded to ProteomeXchange and shared with the public.

2 Materials

All solutions should be prepared in glassware using nanopure water (18 M Ω) and analytical-grade reagents. Be careful to avoid any residual detergents and other contaminants as these will deteriorate the MS signal. Solutions can be stored at room temperature.

2.1 Sample Preparation

1. Many types of samples from milk or dairy products can be analyzed, including fresh milk, milk digesta, in vitro enzymatic digests, cheese, yogurt, and kefir.
2. Protease inhibitor cocktail, complete mini EDTA-free 50 $\times$ stock solution (we use product 4693124001 from Roche).
3. Folch solution: 2:1 chloroform:methanol (MeOH).
4. Freeze drier or vacuum centrifuge.
5. 200 g/L trichloroacetic acid (TCA) in nanopure water.
6. C18 solid-phase extraction 96-well plate (we use Glygen, Columbia, MD).
7. Activation solution: 99% acetonitrile (ACN), 0.1% trifluoro-acetic acid (TFA), 0.9% water (v/v).
8. TFA solution: 0.1% TFA in nanopure water (v/v).
9. Equilibration/wash solution: 1% ACN, 0.1% TFA, 98.9% water (v/v).
10. Peptide elution solution: 80% ACN, 0.1% TFA, 19.9% water (v/v).
11. Peptide rehydration solution: 2% ACN, 0.1% TFA, 97.9% water (v/v).

2.2 Mass Spectrometry

1. Solvent A: 0.1% formic acid (FA).
2. Solvent B: 100% ACN.
3. Liquid chromatography coupled to nanoelectrospray (we use a Waters Nano Acquity UHPLC (Waters, Milford, MA) with a Proxeon nanospray source).
4. A reversed-phase trap column (we use a 100 μ m $\times$ 25 mm Magic C18 100 $\text{\AA}$ 5U column) for online desalting and a reversed-phase analytical column for peptide separation (we use a 75 μ m $\times$ 150 mm Magic C18 200 $\text{\AA}$ 3U column, Waters, Milford, MA).
5. Mass spectrometry instrument (we use a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer from Thermo Fisher Scientific, Waltham, MA, but many other instruments can be used).

3 Methods

3.1 Liquid Milk Sample Collection and Handling

1. For human milk collection, nipples should be cleaned prior to collection. For bovine milk collection, teats should be washed in water and then dipped in an antiseptic solution. Collection volumes can vary, but we typically collect at least 1 mL of sample.
2. Milk samples must be frozen as soon as possible after collection and kept at -80°C to limit milk protease activity, which could alter the peptide profile (*see Note 1*).

3.2 Peptide Extraction and Sample Preparation

1. Thaw samples and bring to 4°C (*see Note 2*).
2. Pipette 100 μL of each liquid milk sample into a 1.5-mL tube. Add 5 μL of protease inhibitor stock solution and mix with a vortex on low speed for 10 s.
3. For solid samples, like cheese, weigh out 100 mg of sample and cut into fine pieces. Combine with 1 mL of nanopure water. Shake at 40°C for 1 h. Agitate in an ultrasonic bath for 15 min at 40°C . Shake at 40°C for another 1 h. To eliminate large particles, centrifuge solution at $3080 \times g$ at 4°C for 30 min. Collect liquid supernatant (approx. 0.9 mL), skip **steps 4–6** and proceed to **step 7**.
4. Skim milk samples by centrifugation at $16,000 \times g$ for 15 min at 4°C (*see Note 3*).
5. Carefully insert a thin pipette tip below the upper lipid layer (cream) and collect the infranate (skim milk). If necessary, repeat the centrifugation procedure to remove any residual cream.
6. Take 25 μL of the skim milk sample. Add 100 μL of nanopure water (*see Note 4*).
7. For some sample types, like gastric digesta and cheese, removing any remaining lipids by applying a Folch liquid-liquid extraction avoids LC column clogging and signal deterioration (*see Note 5*). Add four times the sample volume of Folch solution to the sample. Mix with a vortex for 10 s. Centrifuge at $16,000 \times g$ for 15 min at 4°C . If the upper phase is not clear, centrifuge again until clear. Collect the top layer (MeOH and water). Be careful not to collect any of the middle protein layer. Discard the middle protein layer and bottom chloroform/lipid layer. Dry collected upper layer by centrifugal evaporation at 44°C . Rehydrate samples in 100 μL nanopure water and mix using a vortex until completely solubilized.
8. Precipitate intact proteins from the samples by acid precipitation. Add 1:1 v/v (sample to solution) 200 g/L trichloroacetic acid, mix using a vortex for 5 s, then

centrifuge at $4000 \times g$ for 10 min at 4 °C. Collect the supernatant and discard the protein pellet.

9. Clean up extracted peptides via C18 solid-phase extraction. This step eliminates sugars, salts, and TCA. To prepare the 96-well plates, first wash them by adding 200 μ L of the activation solution to each well and centrifuging at $2800 \times g$ for 30 s. Repeat this step for a total of three times. Next, equilibrate the columns by washing three times with equilibration solution using 200 μ L each time at the same centrifuge speeds and times. Next, add all of each sample to a well (200 μ L) and centrifuge. To wash off salts, sugars, and TCA, wash three times with 200 μ L wash solution. Finally, elute three times using 200 μ L peptide elution solution and centrifuge each time. Collect this fraction.
10. Transfer eluted samples to 1.5-mL tubes and dry with a vacuum concentrator or freeze drier. Rehydrate peptides by adding 40 μ L of 2% ACN, 0.1% TFA in nanopure water, and mix with a vortex for 1 min. Store at -80 °C until use for MS analysis.
11. *Optional step:* At this point, you may want to measure the extracted peptide concentration to determine how much sample to inject for LC-MS (*see* Subheading 3.3). We commonly used the Bradford assay, the bicinchoninic assay, or absorbance at 280 nm.

3.3 Mass Spectrometry

1. Add samples to sampling vials.
2. Set up analysis parameters for LC-MS. Set up the elution gradient using solvent (A) 0.1% FA and solvent (B) 100% ACN. Design a 60-min gradient as follows: 5–35% B over 50 min, 35–80% B over 3 min, 80% B for 1 min, 80–5% B over 1 min, and then hold at 5% B for 5 min. Flow rate: 300 μ L/min. Set to collect mass spectra in data-dependent mode with one MS precursor scan followed by 15 MS/MS scans, using dynamic exclusion of 20 s. Set MS spectral acquisition resolution to 70,000 and a target of 1×10^6 ions or a maximum injection time of 30 ms. Set MS/MS spectral resolution to 17,500 with a target of 5×10^4 ions or a maximum injection time of 50 ms. Apply higher-energy collision dissociation with a normalized collision energy value of 27% for peptide fragmentation. Exclude unassigned charge states as well as ions $>+7$ from MS/MS fragmentation.
3. Load approximately two micrograms of each sample onto the enrichment column for online desalting and then onto the analytical column for analytical separation.
4. After each sample, perform a column wash using a blank sample containing 0.1% FA. Wash gradients will depend on sample type (*see* **Note 6**). We use a saw-tooth gradient as follows: 5–80% B over 6 min, 80% B for 7 min, 80–5% B over 2 min, 5% B for 3 min, 5–80% B over 6 min, 80% B for 7 min, 80–5% B over 2 min, 5% B for 7 min.

3.4 Build a Custom Protein Database

1. To perform a database search for peptide identification, several premade databases are available in the various software programs. However, searching for peptides with no cleavage specificity (as is required in peptidomics) greatly increases the number of possible peptides, increasing the search space and slowing the search from minutes or hours to days when searched against a full organism-level protein library. We recommend creating a custom milk-specific and species-specific protein library for your searches to decrease search time. This library can be made by gathering the fasta format sequences from www.uniprot.org based on available proteomics literature. At www.dallaslab.org, we have human, cow, and pig milk protein databases available for use (select the tab “Resources”).

3.5 Convert Data to .mgf Format with MSconvert

1. If proceeding with X! Tandem analysis for peptide identification, raw data needs to be converted to mgf files. In order to do this, install ProteoWizard (<http://proteowizard.sourceforge.net/tools.shtml>). Open MSConvert, select your raw files and the location for saving the new mgf files. Select output format as mgf. Use all other settings as the defaults, making sure not to include any filters.

3.6 Peptide Identification with X! Tandem Database Search

1. For peptide identification via database searching (see **Note 7**), we use the downloadable GUI version of X! Tandem, called GPM Manager (<http://www.thegpm.org/TANDEM/instructions.html>). Within this program, we work inside the “advanced” tab in the “directory” section.
2. Select the files for analysis. Select “yes” to Skyline data file annotation to allow the results to import correctly into Skyline.
3. Select the protein library you wish to search against. If the library you want to search against is unavailable, this file must be added to the folder GPM Fury/the gpm fasta and added to the files taxonomy.xml and species.js within the folder GPM (/Fury/the gpm/tandem).
4. Apply a peptide $\log(e)$ value of less than -2 . The “e” stands for “expectation value.” $A - 2 \log(e)$ corresponds to an e-value threshold of 0.01 (a 99% confidence level threshold).
5. For measurement errors, in case of Orbitrap data, allow 10–20 ppm for the fragment mass and 10–20 ppm for the parent mass. Allow isotope error, which allows for identification when the instrument fragments an isotope rather than the parent ion. Select “fragment type” as monoisotopic.
6. For signal processing, select a maximum parent charge of 7, a minimum parent $M+H$ of 275, a minimum fragment m/z of 50, total peaks for evaluation as 50, the minimum number of peaks for evaluation as 15 and fragment types b and y (see **Note 8**).

7. For protein modifications, disallow all complete (required) modifications, as the procedure did not make any chemical changes to the peptides. Allow for oxidation of methionine as a potential modification. Select “no” for using sequence annotations and “yes” to allowing protein N-terminal acetylation.
8. For refinement, select “yes” to “refine model,” “no” to “use sequence annotations,” “no” to “use point mutations,” “no” to “use single amino acid polymorphisms” and allow round 1 modifications to include serine and threonine phosphorylation and possibly asparagine and glutamine deamidation. Select “no” for “semi-cleavage,” “no” for “use mods throughout.” Use a valid expectation of <-2 .
9. For protein cleavage specification, select “No Enzyme [X][X],” which must be added into the source code (see **Note 9**). Leave the other settings as defaults.
10. Search spectra.

3.7 Label-Free Peak Quantification with Skyline

1. Install Skyline [25] at <https://skyline.gs.washington.edu/labkey/project/home/begin.view> and open. Create a blank document and save as a .sky skyline file.
2. Go to Settings > Peptide Settings and click the “Modifications tab.” Select the structural modification you would like to include: typically, we use oxidation (M), deamidation (NQ), and phosphorylation (S,T) (see **Note 10**).
3. Click the “Library tab” and then click “Build” Insert a name for your library and specify a save location. Click “next” and then click “Add files” to select the files you wish to analyze. For our protocol, we use the X! Tandem output files (.xml), but other file types are possible. These files must be manually renamed to have the file extension .xtan.xml instead of .xml. To find your .xml files, go to GPM Fury/the gpm/ gpm/archive. After importing your .xtan.xml files, place a checkmark in the field “include ambiguous results,” click “finish,” and wait for the library to build. Select your new library from the list and select “pick peptides matching library” and leave “rank peptides by” blank. On the filter tab, set min length as 4 and max length as 50. Click “OK.”
4. Click View → Spectral libraries. Click “Add all...” and click “add all/include all” in the following boxes.
5. In Settings → Transition settings, go to the Prediction tab and select precursor mass: monoisotopic; product ion mass: monoisotopic; collision energy: Thermo TSQ Vantage; declustering potential: none; optimization library: none.
6. In the filter tab, fill in precursor charges 1–7; ion charge 1; ion type p (for precursor) (see **Note 11**).
7. For the library tab, select ion match tolerance 0.5 m/z , check the box “if a library spectrum is available, pick its most intense ions”; pick 6 product ions from filtered ions charges and types.

8. In the instrument tab, select min m/z : 50; max m/z : 1600; method match tolerance m/z : 0.055.
9. For the Full-Scan tab, select isotopic peaks included: count; precursor mass analyzer: Orbitrap; peaks: 3; resolving power: 60,000 at 400 m/z ; MS/MS filtering acquisition method: none; retention time filtering: use only scans within 1 min of MS/MS IDs.
10. Go to File > Import > Results to add your RAW data file to peak volume extraction. Select “add single-injection replicates in files” and select your .raw files. After confirming, Skyline will extract the peak areas for each identified peptide in all samples.
11. If possible (depending on the number of samples and number of peptides identified), manually inspect for appropriate peak picking of the MS1-filtered peptides. Peaks that do not match criteria or are too close to the noise level to be visually discernable can be deleted from the data set. The criteria we typically use for a match are a 10 ppm mass error and an idotp (iso-tope dot product) score 80. This filter step also can be done within Excel in the exported results from the following step.
12. To export the results, go to File > Export > Report. Select “edit list...” and in the new window click “import” and load the file “Peptidomics_standard_output.skyr.” Click “OK.” This standard file can be downloaded from our website dallaslab.org under Resources. Select export report as “Peptidomics_standard_output” and export as a .tsv file (tab-separated).

3.8 Alternate Peptide Identification and Label-Free Quantification Approach with Proteome Discoverer (v2.1.0.81)

1. Open Thermo Proteome Discoverer.
2. Go to file > New study/analysis. Fill in Study Name and Study Root Directory.
3. Select a consensus workflow. You can create or modify a new workflow by dragging new nodes into your workflow tree. Our consensus workflow can be viewed in Fig. 2.
4. In the consensus workflow, keep settings not described as default.
5. In the peptide and protein filter node, select “Peptide Confidence at Least”: High. Set “Minimum Peptide Length” to 4.
6. In the Peptide and Protein Annotation node, set the “Annotate Flanking Residues of the Peptide” to “True.” Set “Protein Modifications Reported” to “For All Proteins.” Set “Protein Position for Peptides” to “For All Proteins.”
7. Select a processing workflow. Our processing workflow is shown in Fig. 3.
8. In the processing workflow, keep settings not described as default.
9. In the Event Detector node, set “Mass Precision” to 2 ppm.

10. In the Spectrum Selector node, set “Min. Precursor Mass” to 300 Da. Set the “Max Precursor Mass” to 5000 Da.
11. In the Sequest HT node, set the protein database to your custom database (examples for human, cow, and pig milk can be found on www.dallaslab.org under “Resources”). Custom databases can be added through the “maintain fasta file” icon in the menu bar then press add and select your .fasta file. Back in the Sequest HT node, set Enzyme Name to No-Enzyme (Unspecific). Set “Min. Peptide Length” to 4. Set “Max Peptide Length” to 144 (the maximum allowed by Proteome Discoverer). Set “Precursor Mass Tolerance” to 10 ppm. Set “Fragment Mass Tolerance” to 0.8 Da. Allow oxidation of Met and phosphorylation of Ser and Thr as “Dynamic Modifications” 1 and 2, respectively
12. In the Percolator node, set “Validation based on” to q-Value. Quantification method is “(no quantification)” for label-free quantification. Click “OK.”
13. In the “Input Files” tab > add files > and select your files (see **Note 12**). Select all your files and drag them to the “Input Files” field on the right. Click “Run.”
14. When the search is finished, go to your study tab > then analysis results tab and double click on the result from your run to open it. In the new window you can examine your results.
15. To export your results, in the results window, go to file > export > to Microsoft Excel... > Choose a path for your file and save your results. Export a file with protein groups and one with peptide groups.

3.9 Peptide Mapping

1. To better visualize peptides identified, map the peptides to the protein sequences. Mapping where the fragments from a protein derived in relation to the overall sequence can support biological insight into the enzymatic processes occurring in a system. Previously we used PepEx to map the endogenous peptides in human milk and revealed that the release of peptides was highly specific to regions of the parent protein [11].
2. To access the tool to map the peptides, go to <http://mbpdb.nws.oregonstate.edu> and choose PepEx.
3. Click “PepEx Add Fasta files.” Insert the .fasta file that you used for your X! Tandem search.
4. A .tsv file should be made with the following columns and information on your peptides (Fig. 4). Column A, Protein; column B, Peptide; column C, Precursor; column D, Precursor Charge; column E, Precursor Mz; column F, Peptide Modified Sequence; column G, Peptide start; column H, Peptide end; column I, Modifications; column J, Library Name; column K, Sample1 Total Area; column L, Sample2 Total Area (Fig. 4). Only columns A, G, H, and K need to be filled. You can add more samples adjacent to column L.

5. This program will output a report in Excel where, for each protein, the total abundances of peptides at each amino acid in the sequence are shown. Remember the result file comes as a .txt file, which needs to be opened with comma separation. Use Excel to make line graphs plotting the abundances (y -axis) against the peptide sequence (x -axis) to allow visualization (Fig. 5).

3.10 Enzyme Predictor

1. Enzyme predictor (<http://bioware.ucd.ie/~enzpred/Enzpred.php>) is an online bioinformatic tool [26] that predicts which enzymes most likely cleaved the sample proteins into the identified peptides. The predictions are based on a comparison of the protease cleavage specificity with the cleavage sites at both ends of each identified peptide.
2. Upload a tab-delimited text file with two columns ("prot_acc" and "pep_seq") containing the protein accession number from www.uniprot.org and the peptide amino acid one letter sequence in each row, respectively (see **Note 13**).
3. The output file shows how many identified cleavages mapped to each protease.

3.11 Peptidomics Enzyme Estimator

1. Peptidomics Enzyme Estimator [11] is another program for estimating protease activity based on peptidomics data. This program allows more analysis options than Enzyme predictor. For example, you can use either peptide count or abundance, and you can add new proteases (and their specificity patterns) as needed.
2. Download Peptidomics Enzyme Estimator from the eparker05 Github repository (<https://github.com/eparker05/Peptidomics-enzyme-estimator>) and save it to your python directory. Peptidomics Enzyme Estimator requires that Python 2.7 and Biopython are installed (see **Note 14**).
3. Create a .csv file with your identified peptide information. Row 1 in Column A should be "Sequence," in column B "Intensity," in Column C "Protein_id," and in column D "Sample_id." In the Sequence column, add the single amino acid sequence of your peptide. In the Intensity column, add the quantification. In the Protein_id, add the protein information (should be in fasta format, i.e., sp.|P05814|CASB_HUMAN). In the Sample_id column, add the name of your sample. The program output distinguishes between the results from each sample_id so that many samples can be analyzed simultaneously.
4. In Python IDLE, use the shell window and import the Peptidomics Enzyme Estimator module using the code: `import PeptidomicsEnzymeEstimator as pee`
5. Open the relevant library files:


```
fastaFile = open("YourProteinLibrary.fasta", "rU") #(see Note 15)
csvFile = open("YourPeptidomics.csv", "rU")
inputEnzymeList = ["Enzyme1", "Enzyme2"] #(see Note 16)
```

6. Load and preprocess the peptides and retrieve the results as a .csv file:

```
peptideList = pee.import_peptides_and_preprocess(csvFile, fastaFile,  
inputEnzymeList)
```

```
resultCSV=pee.extract_data_from_processed_peptides(peptideList,  
inputEnzymeList, result = "list")
```

```
import csv
```

```
with open('output.csv', 'wb') as csvfile:
```

```
resultWriter = csv.writer(csvfile)
```

```
resultWriter.writerows(resultCSV)
```

The results will show how much of the peptide profile (by count or abundance) was mapped to the proteases you searched against. These results will be output to a .csv file "output.csv" located in your Python directory.

3.12 Database Search

1. Milk contains a large number of functional peptides with a wide range of biological activities. One method for predicting peptide function is through a simple homology search against a database of known functional peptide sequences to determine whether the peptides identified in a sample have been previously associated with a biological action.
2. We have collected a complete database of human milk- and dairy-derived bioactive peptides. This database can be accessed at <http://mbpdb.nws.oregonstate.edu>. Select the Resources tab, which will direct you to a list of peptide tools. Choose the milk bioactive peptide database.
3. The search function in this database includes a search for the specific peptide sequence, specific protein, species, or function. It is possible to search a single peptide sequence or multiple peptide sequences. Furthermore, the search function for peptide sequence includes three search options. One option searches for bioactive peptides matching the input peptide sequence. The second option searches for bioactive peptides that contain the input peptide sequence. The third option searches for bioactive peptides contained within the input peptide sequence. These three search options can be combined with a similarity search option that allows identification of peptides with minor sequence modifications. This option is useful as homologous peptides may retain the functionality of the original peptide.

3.13 Antimicrobial Prediction

1. Antimicrobial activity of peptides is one of the main biological functions studied from milk-derived peptides. We apply CAMPR3 [27], a collection of antimicrobial peptide (AMPs) prediction tools (<http://www.camp.bicnirrh.res.in/>) to identify potential antimicrobials. Under the Tools menu, select AMP

prediction. Choose predict antimicrobial peptides in the opening window. Input your peptide list in a fasta file format or upload it. Click “submit.”

2. The program will produce a list with a score for the likelihood that each peptide is antimicrobial. A score above 0.5 is a positive score.
3. The database of antimicrobial activity and structure of peptides (<http://dbaasp.org/home.xhtml>) is another database of antimicrobial peptides that contains a tool for AMP prediction under the prediction tab.
4. For this search, paste in your fasta file and click “submit.” The program determines whether or not each peptide is potentially antimicrobial.

3.14 Data Sharing

1. After peptides are identified, we suggest that you upload your data to ProteomeXchange [28] or a similar program for public use. Public data sharing is very important for omics data in particular.

4 Notes

1. If possible, add the sample directly to a mixture of protease inhibitors to prevent any proteolytic action, mix, and then freeze.
2. Unless extensively heat-treated or previously treated with protease inhibitors, breast milk and dairy samples will contain active proteases, so avoid keeping the samples at room temperature as it may change the peptide profiles.
3. Cold centrifugation aids separation of the cream layer.
4. We added 100 μ L of nanopure water to assist in dissolving thick samples, such as digestive samples, kefir, yogurt, or cheese. If only milk is being examined, this addition is not necessary. However, in studies where we have compared milk to digestive samples, we have added 100 μ L of water to keep sample volumes constant for relative quantification.
5. The Folch step may not be necessary, depending on the sample type. For example, we did not employ this step for kefir peptidomics [18].
6. Samples vary in how difficult they are to fully elute from the column. Therefore, it is essential to test different blank method gradients and verify that all residual peptides are removed during the blank assessment to avoid contaminating the next sample.
7. Database searching matches tandem spectra by comparison with theoretical spectra derived from predicted peptides in a protein library. Several software tools are available for identification and quantification of peptides from the spectra obtained with mass spectrometry.
8. For collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) commonly used in peptide and protein mass spectrometry, “b” and “y” type ions are the most common.

9. Add the X|X enzyme in the file `g_pcs.js` in the folder GPM Fury/thegpm/tandem. Bottom-up proteomics employs proteolytic enzymes, such as trypsin, with high specificity, which allows searching against only peptides matching those - specificity patterns. In peptidomic analysis, peptides are cleaved by an array of often-unknown endogenous proteolytic enzymes, which requires searching against all possible peptide fragments.
10. Our saved settings file (“PeptideRelativeQuantSettings.skys”) for peptide relative quantification can be downloaded at www.dallaslab.org and imported via Settings > Import.
11. Selecting “p” means we are searching based only on precursor ions and isotopes and not on the products in the particular spectra. Since we are only searching the precursor ions, the product ions box can be ignored.
12. You may want to note the ID of your sample as this ID will appear in your exported result file.
13. An example input file can be viewed at our homepage.
14. Biopython is a package for Python, which is required for PeptidomicsEnzymeEstimator. To install Biopython using pip for Windows users, open the command prompt. Locate your python directory using the change directory command (`cd`). When there, type in the command to install Biopython “`python -m pip install biopython.`”
15. This file is the same protein library used for the original database search in X! Tandem or Proteome Discoverer.
16. The proteases already defined in Peptidomics Enzyme Estimator are: Arg-C proteinase, Asp-N endopeptidase, BNPS-Skatole, Chymotrypsin specific, Chymotrypsin low-spec, Pepsin, Plasmin, Cathepsin D, Cathepsin B, Thrombin, Elastase, Trypsin, and Thrombin-OSP. In the `inputEnzymeList = [“Enzyme1”, “Enzyme2”, “etc...”]` code, substitute Enzyme1 and Enzyme2 with the desired enzyme (e.g., Elastase or Plasmin). Continue adding as many enzymes as needed from those defined in the database. The source code can also be modified to add additional enzyme cleavage patterns.

Acknowledgments

The authors thank C. J. Dillard for editing this manuscript. The authors gratefully acknowledge funding from the National Institutes of Health, Eunice Kennedy Shriver Institute of Child Health and Development (4R00HD079561) R00 Pathway to Independence Career Award.

References

1. Migliore-Samour D, Floch F, Jollès P (1989) Biologically-active casein peptides implicated in immunomodulation. *J Dairy Res* 56:357–362 [PubMed: 2668358]
2. Jørgensen ALW, Juul-Madsen HR, Stagsted J (2010) Colostrum and bioactive, colostral peptides differentially modulate the innate immune response of intestinal epithelial cells. *J Pept Sci* 16:21–30 [PubMed: 19904766]

3. Brantl V (1984) Novel opioid peptides derived from human beta-casein: human beta-casomorphins. *Eur J Pharmacol* 106:213–214 [PubMed: 6529969]
4. Kampa M, Loukas S, Hatzoglou A et al. (1996) Identification of a novel opioid peptide (Tyr-Val-Pro-Phe-Pro) derived from human alpha S1 casein (alpha S1-casomorphin, and alpha S1-casomorphin amide). *Biochem J* 319:903–908 [PubMed: 8920997]
5. Aniansson G, Andersson B, Lindstedt R, Svanborg C (1990) Antiadhesive activity of human casein against *Streptococcus pneumoniae* and *Haemophilus influenzae*. *Microb Pathog* 8:315–323 [PubMed: 2215182]
6. Stromqvist M, Falk P, Bergstrom S et al. (1995) Human-milk k-casein and inhibition of *Helicobacter pylori* adhesion to human gastric mucosa. *J Pediatr Gastroenterol Nutr* 21:288–296 [PubMed: 8523212]
7. Liepke C, Zucht H-D, Forssmann W-G, Ständker L (2001) Purification of novel peptide antibiotics from human milk. *J Chromatogr B Analyt Technol Biomed Life Sci* 752:369–377
8. Yamada A, Sakurai T, Ochi D et al. (2015) Antihypertensive effect of the bovine casein-derived peptide Met-Lys-Pro. *Food Chem* 172:441–446 [PubMed: 25442576]
9. Suetsuna K, Ukeda H, Ochi H (2000) Isolation and characterization of free radical scavenging activities peptides derived from casein. *J Nutr Biochem* 11:128–131 [PubMed: 10742656]
10. Dallas DC, Guerrero A, Khaldi N et al. (2013) Extensive in vivo human milk peptidomics reveals specific proteolysis yielding protective antimicrobial peptides. *J Proteome Res* 12:2295–2304 [PubMed: 23586814]
11. Guerrero A, Dallas DC, Contreras S et al. (2014) Mechanistic peptidomics: factors that dictate the specificity on the formation of endogenous peptides in human milk. *Mol Cell Proteomics* 13:3343–3351 [PubMed: 25172956]
12. Dallas DC, Guerrero A, Khaldi N et al. (2014) A peptidomic analysis of human milk digestion in the infant stomach reveals protein-specific degradation patterns. *J Nutr* 144:815–820 [PubMed: 24699806]
13. Combes C, Paterson E, Amadò R (2002) Isolation and identification of low-molecular-weight peptides from Emmentaler cheese. *J Food Sci* 67:553–559
14. Toelstede S, Hofmann T (2008) Sensomics mapping and identification of the key bitter metabolites in Gouda cheese. *J Agric Food Chem* 56:2795–2804 [PubMed: 18355023]
15. Gupta A, Mann B, Kumar R, Sangwan RB (2010) Identification of antioxidant peptides in cheddar cheese made with adjunct culture *Lactobacillus casei* ssp. *casei* 300. *Milchwissenschaft* 65:396–399
16. Sforza S, Cavatorta V, Lambertini F et al. (2012) Cheese peptidomics: a detailed study on the evolution of the oligopeptide fraction in Parmigiano-Reggiano cheese from curd to 24 months of aging. *J Dairy Sci* 95:3514–3526 [PubMed: 22720910]
17. Miclo L, Roux E, Genay M et al. (2012) Variability of hydrolysis of β -, α s1-, and α s2-caseins by 10 strains of *Streptococcus thermophilus* and resulting bioactive peptides. *J Agric Food Chem* 60:554–565 [PubMed: 22103626]
18. Dallas DC, Citerne F, Tian T et al. (2016) Peptidomic analysis reveals proteolytic activity of kefir microorganisms on bovine milk proteins. *Food Chem* 197:273–284 [PubMed: 26616950]
19. Rauh VM, Johansen LB, Ipsen R et al. (2014) Plasmin activity in UHT milk: relationship between proteolysis, age gelation, and bitterness. *J Agric Food Chem* 62:6852–6860 [PubMed: 24964203]
20. Jensen S, Sidsel J, Therese J et al. (2015) Storage-induced changes in the sensory characteristics and volatiles of conventional and lactose-hydrolyzed UHT processed milk. *Eur Food Res Technol* 240:1247–1257
21. Nielsen SD, Jansson T, Le TT et al. (2017) Correlation between sensory properties and peptides derived from lactose-hydrolyzed UHT milk during storage. *Int Dairy J* 68:1–108
22. Dallas DC, Guerrero A, Parker EA et al. (2013) Peptidomic profile of milk of Holstein cows at peak lactation. *J Agric Food Chem* 62:58–65 [PubMed: 24344900]
23. Guerrero A, Dallas DC, Contreras S et al. (2014) Peptidomic analysis of healthy and subclinically mastitic bovine milk. *Int Dairy J* 46:46–52

24. Dallas DC, Guerrero A, Parker EA et al. (2015) Current peptidomics: applications, purification, identification, quantification and functional analysis. *Proteomics* 15:1026–1038 [PubMed: 25429922]
25. Schilling B, Rardin MJ, MacLean BX et al. (2012) Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in Skyline application to protein acetylation and phosphorylation. *Mol Cell Proteomics* 11:202–214 [PubMed: 22454539]
26. Vijayakumar V, Guerrero AN, Davey N et al. (2012) EnzymePredictor: a tool for predicting and visualizing enzymatic cleavages of digested proteins. *J Proteome Res* 11:6056–6065 [PubMed: 23098558]
27. Waghui FH, Barai RS, Gurung P, Idicula-Thomas S (2016) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res* 44:D1094–D1097 [PubMed: 26467475]
28. Vizcaino JA, Côté RG, Csordas A et al. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* 41: D1063–D1069 [PubMed: 23203882]

Fig. 1. Overview of the steps described in this chapter for analysis and identification of milk peptides

Fig. 2.
Overview of nodes in Proteome Discoverer consensus workflow to conduct milk peptidomics

Fig. 3.
Overview of nodes in Proteome Discoverer processing workflow to conduct milk peptidomics

	A	B	C	D	E	F	G	H	I	J	K	L
1	Protein	Peptide	Precursor	Charge	Precursor Mz	Peptide Modified Sequence	Peptide start	Peptide end	Modifications	Library Name	Sample1 Total Area	Sample2 Total Area
2	sp P02666 CASB_BOVIN	DMPIQAF	821.4+	1	821.4	DMPIQAF	199	205		Milk	1.98E+10	2.15E+09
3	sp P02662 CASA1_BOVIN	DQAMEDIKQM	612.8++	2	612.8	DQAMEDIKQM[+16]	66	75	1O	Milk	3.51E+10	2.53E+08
4												

Fig. 4.
Example of a PepEx input file (.tsv) made inside Excel

Fig. 5. Example of how peptidomics data can be used in PepEx to map the relative abundance of peptides identified in an LC/MS run onto the sequence of human alpha s1-casein (not real data)