Post print (i.e. final draft post-refereeing) version of an article published on Journal of Chemistry and Chemical Engineering. Beyond the journal formatting, please note that there could be minor changes from this document to the final published version. The final published version is accessible from here:

http://davidpublishing.org/journals_show_abstract.html?5535-0

This document has made accessible through PORTO, the Open Access Repository of Politecnico di Torino (http://porto.polito.it), in compliance with the Publisher's copyright policy as reported in the SHERPA-ROMEO website:

http://www.sherpa.ac.uk/romeo/issn/0010-4825/

New software for the identification and characterization of peptides generated during Fontina cheese ripening using mass spectrometry data

Sabina Valentini[1], Massimo Natale[2,3], Elisa Ficarra[3] and Andrea Barmaz[1]

1. Institut Agricole Régional, Rég. La Rochère 1/A, 11100 Aosta, Italy

2. BioDigitalValley S.r.l., via Carlo Viola 78, 11026 Pont Saint Martin (AO) , Italy

3. Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy

Abstract:

The microbiological profile in raw milk cheese is typically characterized by a multitude of microbial groups; during cheese ripening a wide range of enzymes interact, hydrolysing caseins into peptides and free amino acids. Although a number of microbial enzymes are common to many cheese varieties, the final peptide composition of a cheese reflects its characteristic ripening process. The peptide profile composition differs according to different stages of ageing, type of manufacturing, and territory, resulting in the flavour and texture characteristic of the particular variety.

The peptide profile may cover thousands of peptides derived from the four original casein molecules of different genetic and chemical variants and other, recently discovered, of non-proteolytic origin but synthesised de novo in cheese by enzymatic activities.

An effective way of acquiring more information on the proteolytic process in cheese is to identify the peptides produced throughout ripening. Some proteolytic peptides have been identified in exploratory studies mainly using amino acid sequencing and mass spectrometry, but little information is available on the peptide profile of Fontina cheese, an Italian semi-hard and semi-cooked cheese.

Nowadays various software, available online, allow users to identify proteins but all of them are focused on human or human model proteins data sets. Furthermore the peptides data-base used for molecular weight matching are generated by in-silico digestion with only few proteolytic enzymes.

The aim of this work was to design and implement a new bioinformatics software which is able to identify the protein peptides from the peaks arising from in-source or MS/MS fragmentation.

Key words:

Fontina PDO, cheese ripening, LC-MS/MS, new bioinformatics software.

1. Introduction

Cheese maturation involves complex biochemical reactions, of which proteolysis is regarded as being the most important for many varieties. Proteolysis contributes to textural changes (breaking down the αs1, αs2, β and para-κ casein network), it decreases the

water activity through water binding by setting free carboxyl and amino groups, it increases the pH by producing NH3 from deamination of free amino acids. Moreover, proteolysis directly contribute to flavour (release of peptides and amino acids) and off-flavours (bitter hydrophobic peptides), also liberating substrates for others reactions. Thus for the development of an acceptable cheese flavour a well-balanced breakdown of the protein (i.e., casein) into small peptides and amino acids is necessary [1].

The progress of proteolysis in many ripened cheese is first catalysed by residual coagulant (chymosin, which has an hydrophilic character, has more affinity for the whey fraction, but it can also be partially trapped in the caseins network) and, to a lesser extent, by plasmin and milk proteinases, resulting in the formation of large and intermediate-sized peptides which are subsequently degraded by the coagulant and enzymes from starter and non starter microflora of the cheese, resulting in the production of small peptides and free amino acids. Although a number of microbial enzymes are common to many cheese varieties, the final peptide composition of a cheese reflects its characteristic ripening process. The peptide profile composition differs according to different stages of ageing, type of manufacturing, and territory, resulting in the flavour and texture characteristic of the particular variety [3].

An effective way of acquiring more information on the proteolytic process in cheese is to identify the peptides produced throughout ripening [4]. Some proteolytic peptides have been identified in exploratory studies mainly using amino acid sequencing and mass spectrometry, but little information is available on the peptide profile of Fontina cheese, an Italian semi-hard and semi-cooked cheese, marked with the quality P.D.O. label, typically produced in Aosta Valley, made from raw cow milk and ripened for at least 3 months in natural caves.

Nowadays various software, available online, allow users to identify proteins but all of them are focused on human or human model proteins data sets. Furthermore the peptides data-base used for molecular weight matching are generated by in-silico digestion with only few proteolytic enzymes.

The aim of this work was to design and implement a new bioinformatics software which is able to identify the protein peptides from the peaks arising from in-source or MS/MS fragmentation.

2. Experiment

2.1 Material

Fontina cheese samples (24) coming from 12 different cheese wheels, aged from 7 to 84 days, were produced in the same factory, on the same day, starting from the same batch of milk to reduce the variability usually present in Fontina cheese. Cheese wheels were then ripened in 3 different caves and sampled at regular intervals during the following 3 month. The cheese samples were grated and stored at –80°C until analysis. All solvents and reagents were HPLC or LC-MS grade. The oligopeptide fraction was extracted according to the method reported by Sforza et al. [5].

2.2 HPLC-Mass Spectrometry analysis

The HPLC-MS equipment consisted of a Thermofisher ACCELA AS separation module connected with a Thermofisher LTQ XL mass spectrometer. HPLC condition: Jupiter Proteo (Phenomenex) C18 column, 4 µm, 300 Å, 250 x 4.6 mm. Eluent A: H2O (0.2% CH3CN and 0.1% HCCOOH), eluent B: H2O : CH3CN 65 : 35 (0.1% HCOOH). Ms condition: ESI interface with 90% splitting of the column flow, positive ions, single quadrupole analyser.

2.3 Software development and implementation

Subsequently the LC-MS/MS analysis, the peptides were analysed by our software to determine their amino acid sequence. Due to in-source fragmentation oligopeptides gave rise to a pattern of multi-charged ions and/or fragments in the MS spectra, which

allowed the determination of the reconstructed peptide sequence. In fact, since all these peptides derive from the major milk proteins generated by proteolysis during the ageing period, the knowledge of their molecular weight offers a limited range of possibilities for the punctual identification of every single peptide.

Since it is impossible to forecast all the possible cutting sites of enzymes involved in cheese ripening, the software digested in-silico, in all possible peptides form 5 to 20 amino acid length, the major milk proteins (alpha-s1-casein, alpha-s2-casein, beta-casein, kappa-casein, alpha-lactalbdumin, and beta-lactoglobulin).

The software calculate the molecular mass of peptides by taking into account the average isotopic distribution of atomic weights.

Thanks to in-silico digestion, the software can match every pick in mass spectrum with a list of all the possible sequences contained in the caseins that were compatible with that particular molecular weight. When a match is established between the theoretical molecular weight of a peptide and a pick in MS spectra, the software tries to match all the others peaks in the MS spectra to the theoretical fragment list. The software calculated the molecular weight of the fragments generated by all the possible peptides for the loss of the first N- or C-terminal amino acids. In low-energy induced fragmentations, peptides are known to break up mainly at the peptidic junction, giving rise to two series of fragments indicated as b fragments (charge retained on the N-terminal) and y fragments (charged retained on the C-terminal). In figure 1 is shown an schema of peptide fragmentation occurring at the peptidic junction.

Depending on how many fragments are matched with peaks of MS spectra, the software assigns a score to the identification of peptide.

Analysis of the MS spectra showed that actually every molecular peak was associated to small fragmentation peaks, which usually corresponded to only one of the possible calculated b or y patterns, thus allowing in most cases a clear-cut identification of the peptide sequence from its original casein portion.

The program allows you to specify the mass tolerance, that is the error window for MS fragment ion and mass theoretical values. The unit is expressed in Dalton (Da).

The software was developed in Java as desktop applications using Java Swing technology for the graphic user interface.

3. Results and Discussion

The most usual approaches for peptides identification of MS/MS spectra are based on two different technologies: database dependent and de-novo sequencing, as shown in Table 1. The database-dependent identification strategy attempts to match experimental spectra obtained from the MS, with theoretical spectra representing hypothetical peptides generated from a protein or from genomic database that has been translated. The limitations of database-dependent method is the lack of databases for each organism of interest, and a list of enzymes for theoretical digestion of proteins. The peptides are not searchable using this methods if the enzymes of proteolytic processes are not known. De-novo sequencing approach has the ability to identify previously unknown peptides sequences and enzymes cutting sites. Unfortunately, for studying thousands of peptides, derived from cheese ripening, for instance, the in-source fragmentation cannot induce an extensive peptides fragmentations, while de-novo sequencing is dependent on the instrument mass accuracy and resolution as well as spectral quality.

Since instruments currently used are not able to identify accurately and quickly the large number of peptides generated by proteolytic processes during cheese ripening, a novel identification software tool has been developed. Our software, PeptideHunter, is a database-dependent method for the peptides identification generated by a set of known proteins. PeptideHunter identifies, for each analysis, all the

possible peptides masses produced by in-silico proteins' digestion, and when a peptide mass matches with a MS spectra peak, calculates also all the possible fragments of the peptide.

It is worth noting that, as far as we know, this is the first time that a software is implemented to overcome peptides detection limitations. The database-dependent identification strategy could now rely on a useful tool for the peptides and enzymes involved in complex biological processes identification).

4. Conclusions

In this work we presented a novel software useful for identification of peptide produce by the broad range of enzymes active during cheese ripening. With this tool we obtained useful insight into the proteolytic processes which occur during Fontina cheese aging, arriving at a better understanding of the functional features of the proteolysis end product.

However, this method can be effectively used for other research whose goal is the study of complex proteolytic processes.

References

[1] Mc Sweeney, P.L.H.; Sousa, M.J. Biochemical pathways for the production of flavour compounds in cheese during ripening: A review. *Lait*. **2000**, 80: 293-324

[2] Fox, P.F. Proteolysis during cheese manufacture and ripening. *Journal Of Dairy science*. **1989**, 72: 1379-1400.

[3] Ardö, Y. Cheese ripening. General mechanisms and specific cheese varieties. *Bulletin of the International Dairy Federation*. **2001**, 369: 7-12.

[4] Ardö, Y.; Lilbaek, H.; Kristiansen, K.R.; Zakora, M.; Otte, J. Identification of large phosphopeptides from β-casein that characteristically accumulate during ripening of the semihard cheese Herrgård. *International Dairy Journal*. **2007**,17: 513-524.

[5] Sforza, S.; Galaverna, G.; Neviani, E.; Pinelli, C.; Dossena, A. Study of oligopeptide fraction in Grana Padano and Parmigiano Reggiano cheeses by liquid chromatography-electrospray ionization mass spectrometry. European Journal of Mass Spectrometry **2004**, 10: 421-427

New software for the identification and characterization of peptides generated during Fontina cheese ripening using mass spectrometry data

Table 1 . Tandem mass spectrometry data analysis software

Software	Company	Type	Web link
PEAKS	Bioinformatics Solutions Inc.	de-novo / database	http://www.bioinfor.com/
Spectrum Mill / Sherenga	Agilent Technologies	de-novo / database	http://spectrummill.mit.edu/
SEQUEST	Thermo Finnigan	database	http://fields.scripps.edu/sequest/
Phenyx	GeneBio	database	http://www.genebio.com
Mascot	Matrix Science Inc.	database	http://www.matrixscience.com

Figure 1 . Software workflow.

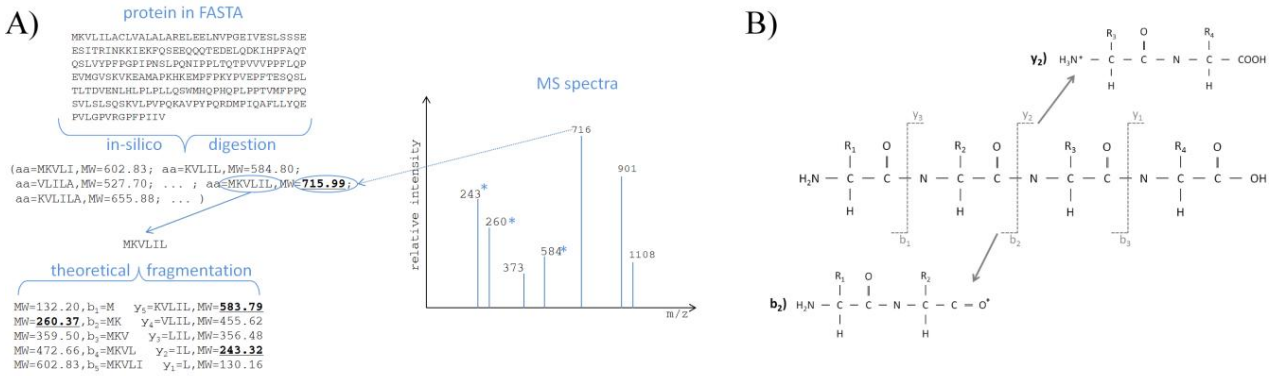

Figure 01) A. The types of fragment ions observed in an MS spectrum are usually produced by break up at the peptidic junction. Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either b, if the charge is retained on the C terminal, the ion type is y. A subscript indicates the number of residues in the fragment. The accepted nomenclature for fragment ions was first proposed by Roepstorff and Fohlman [Roepstorff, 1984], and subsequently modified by Johnson et. al. [Johnson, 1987]. B) The protein sequence is submitted to the software in the FASTA format. The software digests the protein in-silico and only when it is established a matching between a MS spectra pick and a calculated peptide mass, the software fragments the peptide and tries to match theoretical fragments with the others MS spectra picks. The MS spectra data are submitted to the software as a series of numbers.