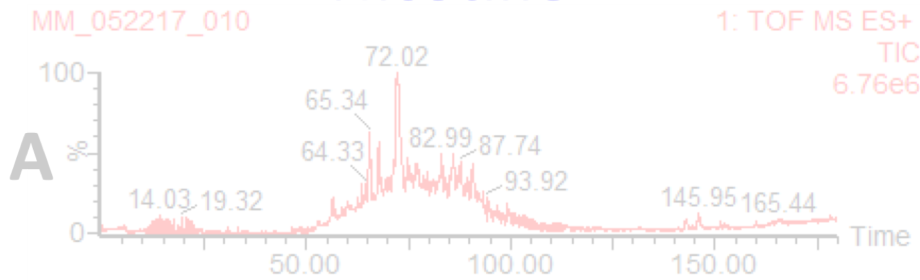
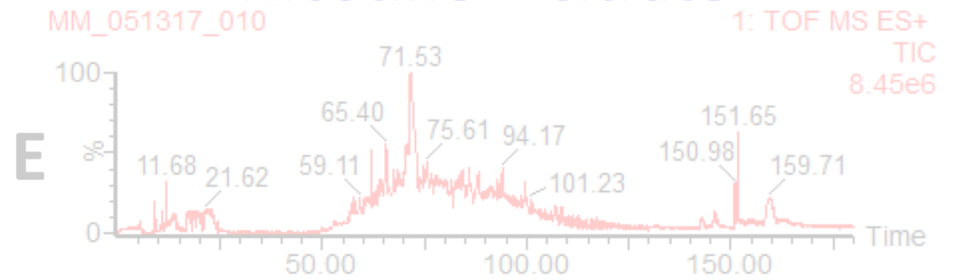
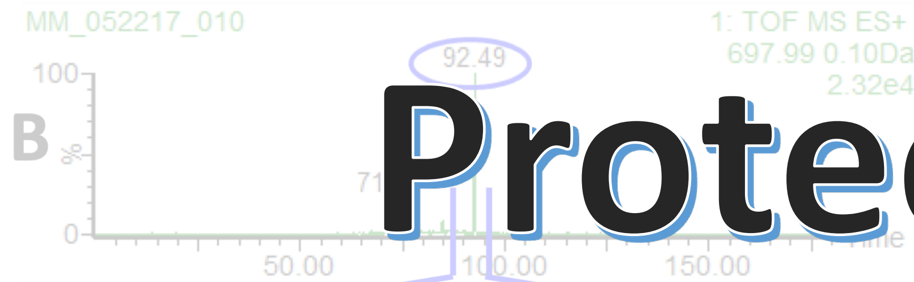
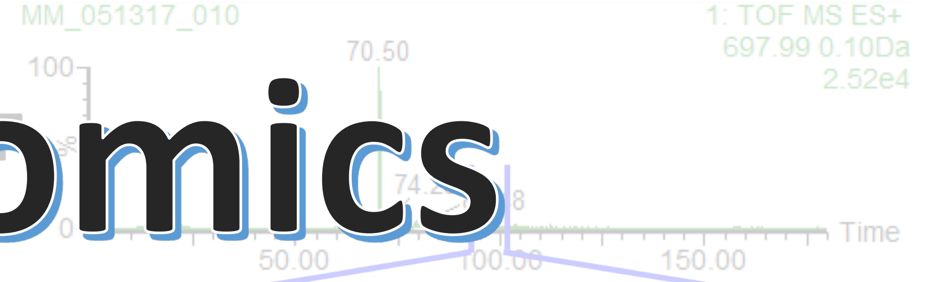
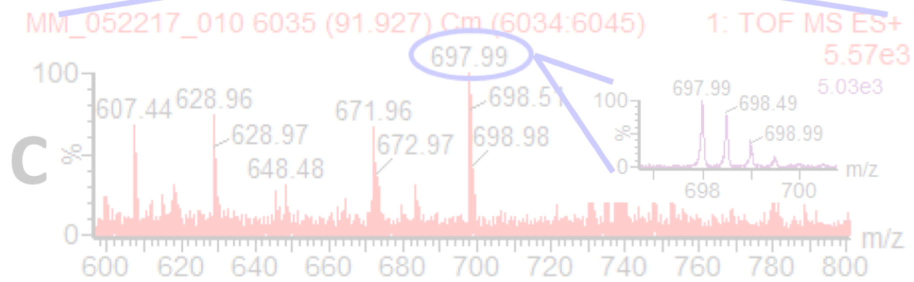
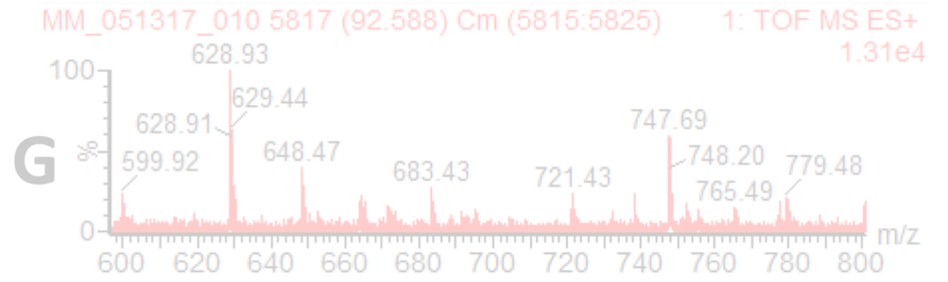


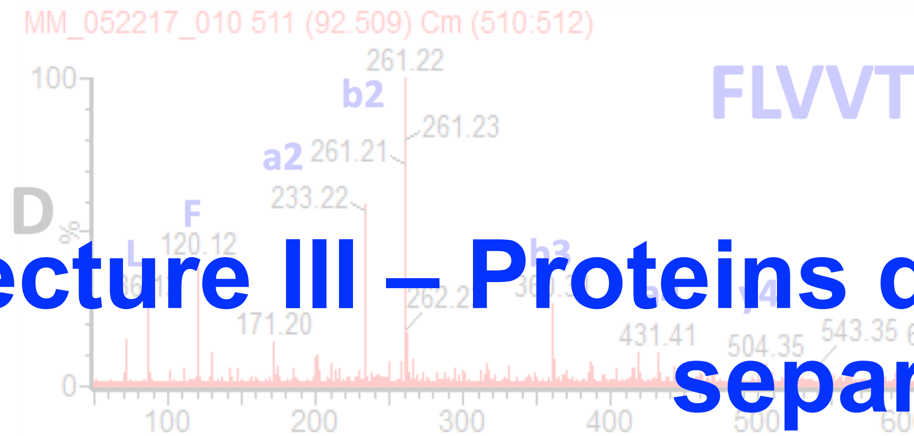


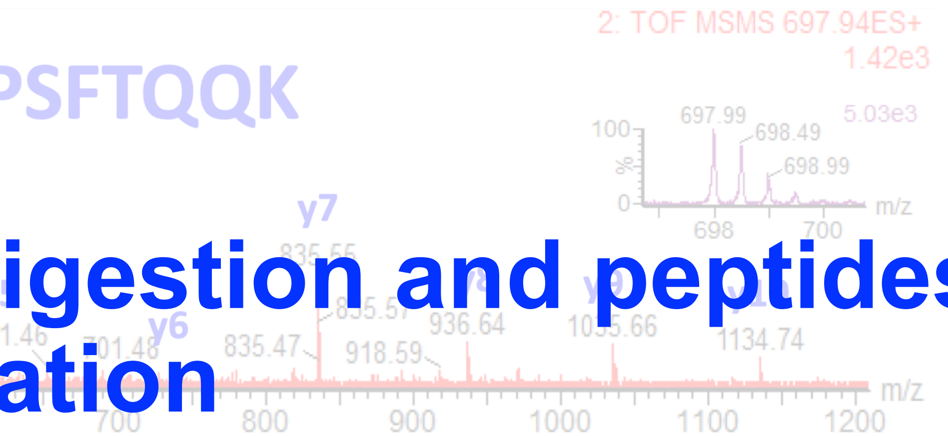








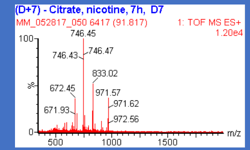




Proteomics

Lecture III – Proteins digestion and peptides separation

I. 3. Protein Digestion for Peptide Generation



Depending on the presence or absence of the protein fractionation step, proteomic methods can be grouped into two major types of approaches:

A. Methods that include a protein fractionation step

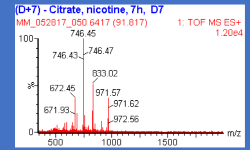
1. Fractionation is performed in gels - gel-based proteomics.
2. Fractionation is achieved through other methods (most commonly chromatography) - gel-free proteomics.

B. Shot-gun methods

Tripsin
Proteins
Peptide
Tripsin
Proteins
Peptides
LC
MS/MS
MS
MS/MS
Protein identification using computers
Combine results

Regardless of the major approach used, one of the mandatory steps in a proteomic study is the fragmentation of proteins into peptides.

I. 3. Protein Digestion for Peptide Generation



Why protein digestion is required in the first place?

Mass spectrometry instruments are currently capable of measuring the molecular masses of protein molecules. Although this **parameter is unique for each protein and should therefore be sufficient to identify it from a mixture**, in reality, this is not possible because:

1. Despite the very high precision of mass spectrometry instruments, they still have a certain degree of error. In the molecular mass range of proteins, the accuracy of the instruments still does not allow the precise identification of proteins with very similar masses;
2. not all proteins can be measured by mass spectrometry - membrane proteins or those of very large sizes do not ionize properly;
3. the sensitivity of measurements is reduced when using proteins of large sizes.

Peptides, on the other hand, are significantly easier to analyze because:

1. Mass spectrometers have maximum precision in the peptide mass range;
2. **They can be relatively easily sequenced.**

In most proteomic methods, the extracted proteins are completely cleaved into peptides, and then the peptides are sequenced. Based on the amino acid sequence of the peptides, the original protein from which they originate is identified. Because identification starts from peptides to proteins, these methods are collectively referred to as **bottom-up methods**.

Review

A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field

Emmalyn J. Dupree [1,†], Madhuri Jayathirtha [1,†], Hannah Yorkey [1], Marius Mihasan [2,*], Brindusa Alina Petre [3,4,*] and Costel C. Darie [1,*]

1 Biochemistry & Proteomics Group, Department of Chemistry and Biomolecular Science, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699-5810, USA; dupreeej@clarkson.edu (E.J.D.); jayathm@clarkson.edu (M.J.); yorkeyh@clarkson.edu (H.Y.)
2 Laboratory of Biochemistry & Molecular Biology, Faculty of Biology, "Alexandru Ioan Cuza" University of Iasi, 700506 Iasi, Romania
3 Laboratory of Biochemistry, Department of Chemistry, "Alexandru Ioan Cuza" University of Iasi, 700506 Iasi, Romania
4 Center for Fundamental Research and Experimental Development in Translation Medicine–TRANSCEND, Regional Institute of Oncology, 700506 Iasi, Romania
* Correspondence: marius.mihasan@uaic.ro (M.M.); brindusa.petre@uaic.ro (B.A.P.); cdarie@clarkson.edu (C.C.D.)
† These authors contributed equally to this work.

Received: 1 June 2020; Accepted: 1 July 2020; Published: 6 July 2020

Abstract: Proteomics is the field of study that includes the analysis of proteins, from either a basic science prospective or a clinical one. Proteins can be investigated for their abundance, variety of proteoforms due to post-translational modifications (PTMs), and their stable or transient protein–protein interactions. This can be especially beneficial in the clinical setting when studying proteins involved in different diseases and conditions. Here, we aim to describe a bottom-up proteomics workflow from sample preparation to data analysis, including all of its benefits and pitfalls. We also describe potential improvements in this type of proteomics workflow for the future.

Keywords: mass spectrometry; proteomics; protein identification; protein characterization

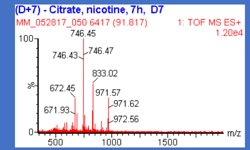
1. Proteomics

The term proteome refers to all proteins that are produced or modified by an organism (e.g., human [1], animal [2], plant [3], bacteria [4]) or living system (e.g., organ, cell culture, complex community from an environmental sample)). The term "proteome" and the first dedicated proteomics laboratory were introduced in 1994 by Wilkins et al. to describe proteins as a complement to genomic data [5]. However, the "whole" proteome of a particular cell, tissue, organ, or organism is yet to be identified. This is particularly difficult due to the vast variety of proteins and their isoproteins/proteoforms/protein species, which are expressed at different levels—from very abundant proteins, such as actin, to less abundant ones, such as transcription factors—in different cells, tissues, or organs. The variety of post-translational modifications (PTMs) in proteins, which may be stable or transient, is responsible for the vast number of proteoforms, which is an obstacle in most proteomics experiments. This, corroborated with the multiple limitations of proteomics methods, makes the quest to identify the proteome of any given organism a difficult one [6].

The proteomics field consists of a wide range of methodology, which has been largely driven by the modern development of involved technology. The concept of global protein analysis as a complete atlas of human proteins was proposed over 50 years ago [7]; however, proteomics research did not start until the mid-1990s. The beginning of proteomics research was sparked due to parallel

Proteomes **2020**, *8*, 14; doi:10.3390/proteomes8030014 www.mdpi.com/journal/proteomes

I. 3. Protein Digestion for Peptide Generation



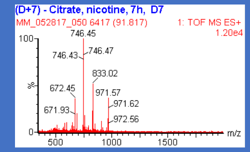
Proteins can be converted into peptides through a process of hydrolysis of a limited number of peptide bonds. Peptide bond hydrolysis can be achieved :

A. **Chemically** - by exposure to strong acids (e.g., HCl). Most commonly, hydrolysis occurs completely, releasing amino acids, which is why it has limited application in proteomics.
B. **Enzymatically** - by using proteases – the go to method for proteomics.

Proteases, also called peptidases or proteinases, **are enzymes that catalyze the cleavage of peptide bonds with the involvement of a water molecule.** Depending on the nature of the molecule involved in the reaction mechanism, proteases are classified into:

1. **Serine proteases**, or **serine endopeptidases**, have three amino acids involved in catalysis in active site forming a catalytic triad - a basic amino acid (His), a dicarboxylic acid (Asp or Glu), and a strongly nucleophilic amino acid - serine. The latter initiates the nucleophilic attack on the peptide bond. Examples include chymotrypsin, trypsin, and subtilisin;
2. **Cysteine proteases**, or **thiol proteases**, have a catalytic triad, but serine is replaced by cysteine. Ex: papain;
3. **Threonine proteases** have a catalytic triad, but serine is replaced by threonine. Ex: proteases domains from the inside the proteasome;
4. **Aspartate proteases** - catalysis is performed by an aspartic acid residue, no catalytic triad. For example, pepsin;
5. **Glutamate proteases** - catalysis is performed by a glutamic acid residue, no catalytic triad;
6. **Metalloproteases** - a metal ion is found in the catalytic site and is involved in catalysis;
7. **Asparagine-peptide lyases** - are a group of enzymes that contain an active site with an Asp residue involved in catalysis, but the actual reaction mechanism does not involve any water molecules.

I. 3. Digestia proteinelor pentru generarea de peptide



Proteases are **promiscuous enzymes** - **can hydrolyse a wide range of different protein molecules, regardless of the amino acid sequence**. However, **the peptide bond cleavage itself is highly specific - proteases recognize a specific amino acid within the amino acid sequence and hydrolyze the peptide bond either before or after it.** These properties are extremely useful in proteomic studies where all proteins in the sample need to be hydrolysed to generate peptides, in a sequence-independent manner..

Trypsin	/K-, /R-, \P
Chymotrypsin	/W-, /Y-, /F-, \P
Glu C (V8 protease)	/E-, /D[a]-, \P
Lys C	/K-, \P
Asp N	/D-

The most frequently used proteases in protemics studies are:

A. Trypsin - is by far the most used enzyme in proteomics because it can be obtained in large quantities relatively inexpensively from the pancreas of pigs or cattle;

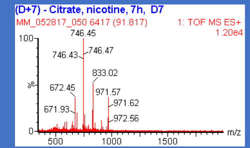
- hydrolyses the peptide bond before a Lys (K) or Arg (R) residue if they are not followed by a Pro (P) residue.

Due to this fact, trypsin action generates a large number of peptides from a protein molecule. The high frequency of the amino acids K and P generally results in a 50 kDa protein generating 30 peptides through trypsin action.

B. Glu-C (protease V8) - is an endoproteinase whose substrate specificity changes depending on the reaction conditions:

- In ammonium-based buffer solutions, the enzyme hydrolyses the peptide bond before a glutamate residue;
- In phosphate-based buffer solutions, the enzyme hydrolyses before Glu and Asp.

I. 3. Digestia proteinelor pentru generarea de peptide



Proteins fractions can be digested either in solution or, in the case where fractionation has been performed via electrophoresis, directly in gels. Most commonly, the gel is cut into pieces and treated with trypsin. By hydrolyzing the proteins from the gel fragment, peptides are generated, which additionally have the advantage that they can be very easily extracted or eluted from the gel (they are significantly smaller than the proteins and are no longer retained by the gel pores).

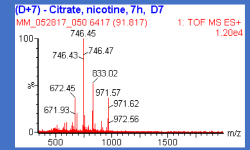
cut out band

1. wash
2. trypsin
3. elute peptides

MS analysis

Fig. 1. Schematic representation of in-gel digestion.

I. 4. Peptides fractionation

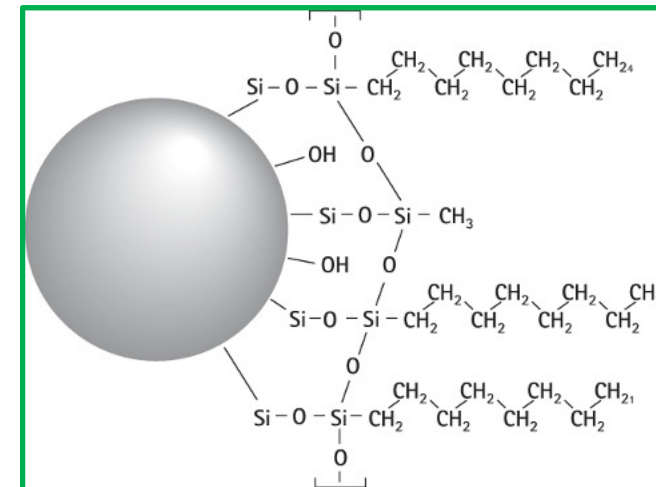


Regardless whether the proteins in the sample are fractionated or not, the number of peptides resulting from enzymatic hydrolysis is still too large to be exhaustively analyzed and sequenced by today's mass spectrometers. Therefore, an additional step of peptide fractionation is required, which is carried out in a chromatographic system similar to the one described earlier.

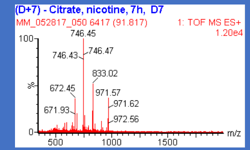
Traditionally, **peptides are separated by reverse-phase chromatography** using chromatographic columns where the stationary phase is attached to inert silica beads with dimensions of 1.8–5 µm. **The mobile phase is pumped through the chromatographic column at a flow rate of 0.1-10 ml/min, and the liquid is forced to pass through the narrow spaces between the beads, creating pressure inside the chromatographic system. For this reason, the separation technique is called HPLC** - **high-pressure liquid chromatography**. Although the term has become almost synonymous with peptide separation, HPLC refers to all chromatographic separations using a liquid mobile phase (LC-liquid chromatography, as opposed to GC - gas chromatography) where the operating pressure is high.

Basic principles for peptide separation using RP-HPLC:

- **The stationary phase** consists of **hydrophobic molecules** - **alkyl chains** (-CH2-, alkanes, or hydrocarbons with various degrees of unsaturation) ranging in length from 2 to 18 carbon atoms. The most commonly used are denoted as C2-C18.
- **The mobile phase** consists of a **variable mixture of a polar solvent such as water (aqueous solvent) and a nonpolar, organic one (acetonitrile, ACN, or methanol, MeOH - organic solvent).** At the beginning of the separation, the ratio is favorable to water, with the mobile phase predominantly polar; at the end of the separation, the ratio is favorable to the nonpolar solvent. The change in the ratio between the two solvents is automated by the chromatographic system in two different ways:

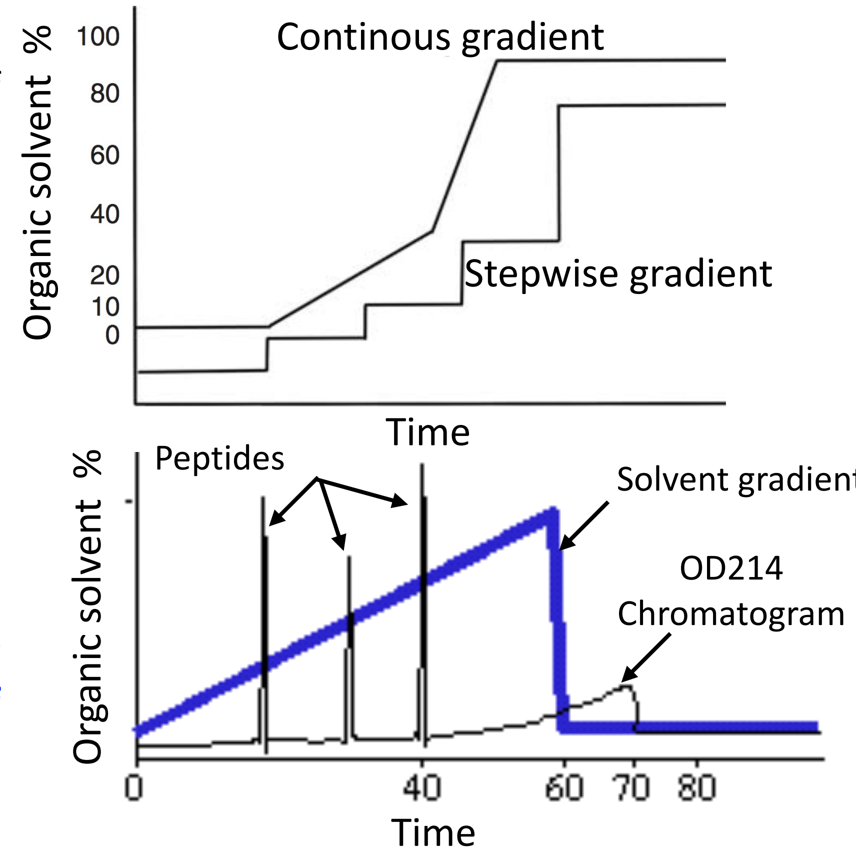


I. 4. Peptides fractionation



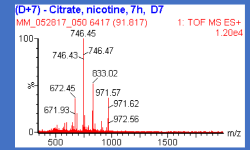
A. As a continuous concentration gradient;
B. As stepwise changes in concentration;

- The peptide mixture is injected into the column in the presence of the polar, aqueous mobile phase, causing the hydrophobic amino acids in the peptide structure to interact with the hydrophobic C18 chains and be retained by the stationary phase;
- As the concentration of organic solvent increases, the hydrophobic interactions between peptides and the hydrophobic chains of the stationary phase decrease, and the peptides are fractionated from the chromatographic column based on sequence: peptides with few hydrophobic amino acids at low concentrations of polar solvent, those with a large number of hydrophobic amino acids towards the end, at high concentrations of polar solvent. **The process by which peptides are removed from the chromatographic column by modifying a physicochemical parameter (in this case, the polarity of the solvent) is called elution.**
- The peptides eluted from the chromatographic column are recorded by the detector as signals or peaks.



- Most commonly, in RP-HPLC, the detector is a UV-VIS spectrometer, and peptides are detected at 214nm;
The short time from sample injection into the chromatographic column to the elution of a peptide is recorded and is called **retention time**. The intensity of the signal from the detector, represented by the peak height, is strictly proportional to the concentration of the eluted peptide.

I. 4. Peptides fractionation

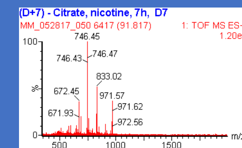


Traditional HPLC systems operate at flow rates of 0.1-10 ml/minute, and because of this, they cannot be directly connected to mass spectrometers - the amount of solvent is too high and cannot be evaporated quickly enough for the peptides to ionize. Additionally, columns in traditional HPLC have a large diameter (4.6 mm most commonly), which results in sample dilution following separation, making it undetectable by MS instruments.

Fractionation of peptides in proteomic experiments is carried out using nano-HPLC systems, which operate on the same principles as traditional HPLC systems but:

- operate at much lower flow rates, typically in the range of **nanoliters per minute** (most frequently 200-350 nL/min), compared to milliliters per minute (mL/min) for traditional HPLC. This reduces solvent consumption and allows for more efficient coupling with mass spectrometers;
- use columns with smaller internal diameters, often in the range of **75-300 µm**, compared to the larger diameters used in traditional HPLC.
- the elution of peptide peaks occurs in a volume of approximately **100 nL**, a volume that can be easily evaporated to ionize the contained peptides. This reduces sample dilution and improves sensitivity, making the eluted peptides more suitable for detection by mass spectrometers.
- are often coupled directly to mass spectrometers, allowing for online analysis of the eluted peptides. This eliminates the need for sample transfer and minimizes sample losses, resulting in improved analytical sensitivity and throughput.

I. 4. Fracționarea peptidelor



Separare LC bidimensională

peptide ions

ion exchange column

FIRST SALT PULSE
ELUTE +

REVERSE PHASE LC

SECOND SALT PULSE
ELUTE 2+

REVERSE PHASE LC

SECOND SALT PULSE
ELUTE 3+

REVERSE PHASE LC