Università di Genova

DEPARTMENT OF EXPERIMENTAL MEDICINE

PhD COURSE IN EXPERIMENTAL MEDICINE

Curriculum of Biochemistry

Development of a method for biomarkers characterization by mass spectrometry techniques

---

Candidate

Alberto Borassi

Tutor

Prof. Gianluca Damonte

PhD Program Coordinator

Prof. Ernesto Fedele

Academic Year 2021-2022

XXXV Cycle

1. INTRODUCTION

One of the most important aspects of the proteomics in post genomics era is Protein/peptide profiling. The proteomic approach to the study of human diseases through the analysis of tissues and body fluids and the translation of this method to the clinic created a new field of research defined as "clinical proteomics" [1].

One of the fundamental tasks of this field is to identify, within the complex protein components (proteoma), new molecules that can be used as biomarkers for a specific pathology useful in diagnosis, in prognosis and in the definition of a therapeutic treatment [2, 3, 4]

Over the past decade, the proteomic and peptidomic approach has been increasingly applied to clinical research thanks to the use of established and/or rapidly evolving technologies such as electrophoresis, chromatography and mass spectrometry of samples both in crystalline and liquid phase.

In particular, these methods are used on a daily basis for the identification and characterisation of possible biomarkers at molecular level. In fact the low molecular weight peptides often play key roles in biological and pathological processes, and the assessment of any qualitative or quantitative difference of these molecules in the plasma or serum of patients versus controls may be important in the study of the progression of diseases such as cancer or degenerative diseases.

1.1 The aim and pipeline of the project

The purpose of this study is to define an extractive approach for the detection of the low-molecule peptide fraction from human plasma or serum and the subsequent analysis and interpretation of the obtained data, with the ultimate aim of developing a standardised protocol for the identification of potential biomarkers.

During the initial phase a system of different extraction of the low molecular protein fraction was developed thanks to a series of standard peptides solutions and using silica magnetic beads techniques differently functionalised [5,6].

Initially the study focused on the development of the extraction method and used a small number of peptides coupled with the technique of extraction using one single type of magnetic silica beads. Then, in order to go deeper in the potential uses of this technique some specific peptides have been synthesized to obtain a standard solution with a larger number of molecules.

At the same time different types of magnetic beads, differently functionalized, have been used with the purpose to bind target molecules with a different type of intermolecular force.

During the implementation of the method, when applied on real biological samples, it became necessary to find a way to eliminate the high molecular weight fractions present in the body fluid.

The treatment of the samples, plasma or serum, took place without the use of proteases, as trypsin, in order to generate digested lysates, or electrophoresis and gel separation techniques, in such a way to avoid creating additional complexity in subsequent steps of data interpretation and to use the lower quantity of sample as possible.

Both the peptides contained in the standard solution and those in the low molecular weight fraction of the pre-treated biological sample were separated and characterized through high performance liquid chromatography (HPLC) coupled to full scan and tandem mass spectrometry at low and medium resolution equipped with an electrospray ion source (ESI-MS/MS). Samples from biological sources were subsequently analysed using the mass spectrometry MALDI-TOF technique [7].

The fundamental point for finding possible candidate biomarkers in proteomic studies is to assess their presence, absence, or different abundances between cases and controls: in this respect semi-quantitative proteomics is an essential function.

In this project the development of the extraction method was followed by its application to real samples. The presence of low-molecular-weight peptides in plasma samples, from dialysis nephrotic patients at various stages of Sars-COV2 infection, and in plasma from healthy donors, was evaluated.

The purpose of this study was to validate the extraction method and the analytical systems which have been developed to this aim, with the hope of finding significant differences between groups, especially in terms of qualitative/quantitative differences in the m/z ratios present in MS spectra [8, 9]; it has been essential to process, from a bioinformatics point of view, the mass spectrometry data extracted with the different types of magnetic beads from pools of samples from different patients.

1.2 Proteoma

The availability of the complete genome of several organisms, including humans, has made it possible to construct detailed and complete protein maps.

The term "proteoma" was coined by Mark Wilkins to describe the totality of proteins encoded by the Genome [10]. Proteomics is the science that aims to establish the identity, quantity, structure,

biochemical functions and interactions of the whole range of proteins expressed in an organ or cell at any given time.

Proteome, unlike genome, is not a fixed feature of an organ. A single genome can give rise to an essential infinitive number of qualitatively and quantitatively different proteomes, and whose high dynamism occurs depending on a precise external stimulus and/or during development.

The proteoma represents therefore a molecular image of the phenotype.

The study of proteoma, or the "protein complement encoded by the genome", is confronted with a series of difficulties generated by its vast complexity, as well as by the intrinsic characteristics of protein macromolecules [11, 12]. Proteins, in fact, have complex structures, heterogeneous chemical-physical properties and different possibilities of post-translational modifications, so no single method of protein extraction allows to capture the entire proteome.

The study of proteoma is linked to the development of targeted technologies, capable of isolating, separating and identifying the highest possible number of components. From the experimental point of view the proteomics is based on three aspects including techniques to fractionate protein or peptide mixtures (such as electrophoresis on gel), the mass spectrometry (MS) technologies to acquire the data necessary to identify individual proteins, and bioinformatics to analyse and assemble the MS data.

Recent research in the field of functional and structural proteomics, expanding the overview, has led to a better understanding of cellular processes and functions.

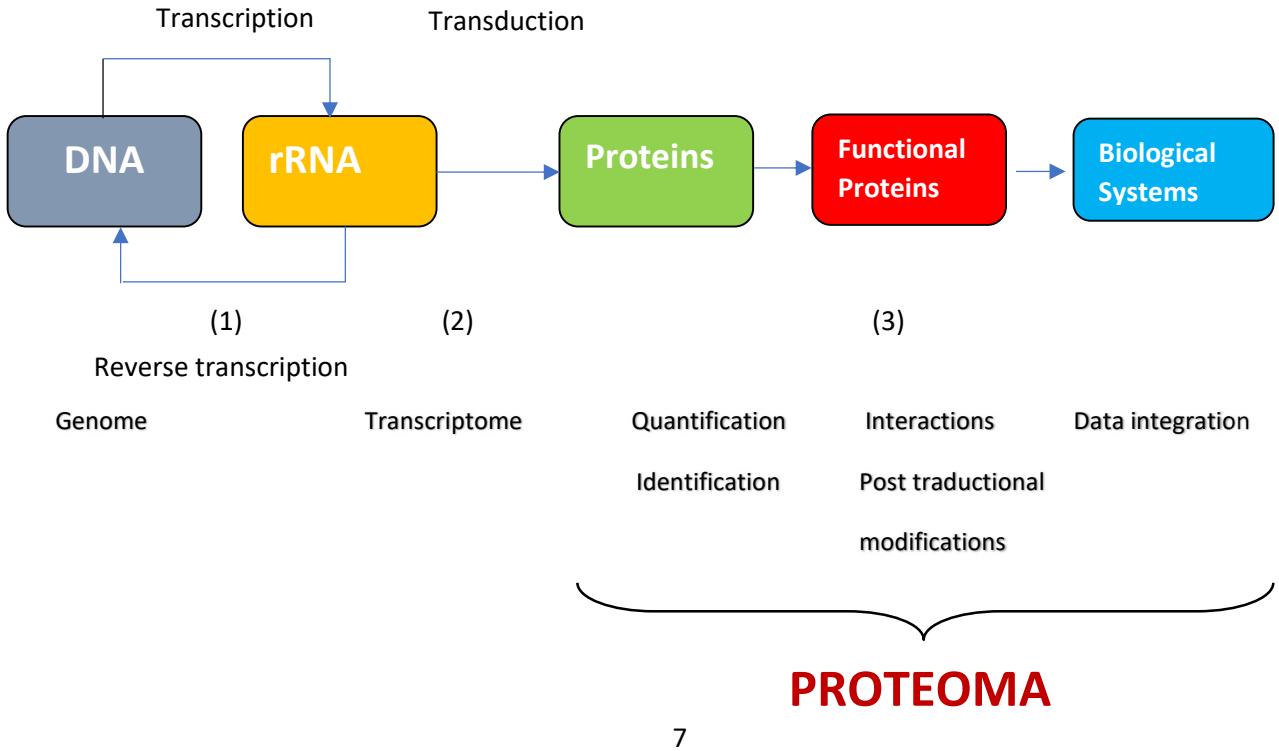


Figure 1. From Genomics to Proteomics. Integration of Genomics and Proteomics in a biological system: a protein is the result of a series of steps that originate as a product of transcription (1), translation processes (2) and post-translational modifications (3).

Proteomic investigation can be divided into:

1) systematic, when aim to identify all proteins expressed in the sample of interest;

2) differential, when it analyse the differential expression of proteins in samples of interest;

3) functional, when include the study of the interactions between proteins (interactomics), the study of the interactions between a protein and its substrates (metabolomics) and the study of the specific functions of proteins (enzymatic genomics and biochemistry).

The components of a proteomic profile are useful as a reference for human disease (diagnosis), as a therapeutic target (therapy) and important for controlling therapy (drug response).

1.3 Peptidomics

Endogenous peptides, including growth factors, cytokines, mediators of cellular response, hormones, and neurotransmitters, represent an important class of biomolecules and have crucial roles in human physiology [13,14]. The peptidome (or low-molecular-weight fraction, LMF) of the human plasma proteome is an invaluable source of biological information, especially in the context of identifying plasma-based markers of disease. In particular, the focus is on the analysis of biological fluids, with an approach based on the hypothesis that the human serum or plasma contains peptides generated by active synthesis and proteolytic processing, often producing proteolytic fragments that mediate or reflect a variety of physiological and pathological functions that occurred in any district of the body [15]. Because of their small size, peptides penetrate cell membranes and epithelial barriers more freely than proteins.

Degradomic studies, investigating cleavage products via peptidomics and top-down proteomics in particular, have warranted significant research interest. In the last years the study of the low molecular protein fraction has been very successful, especially in identifying bioactive peptides candidates to be potential disease markers (for example, oncomarkers) and even foreign peptides related to pathogenic organisms and infection agents [16-20]

Generally, the peptidomics can be defined as the systematic qualitative and quantitative analysis of endogenous peptides in a given biological sample at a well-defined location and time. It can be said that peptidomics is an integrative approach of proteomics in the classical sense and is intended to bridge the gap between this and metabolomics.

1.5 Bioinformatics approach and Python

As previously reported, no proteolytic enzymes were used in biological samples, as in a typical "bottom-up" proteomic approaches, and as a result, there are no peptides from protein lysates and there is no need to use algorithms for the reconstruction of sequences to be connected, see software Sequest [30]. Even considering that, in the case of *tryptic* peptides form protein, these would be mono charges sequences and it would also be easier to process data.

The bioinformatics approach has been implemented, both using statistical tools such as Venn diagram [21] or Significance Analysis of Microarrays (SAM) [22] and developing a series of appropriate codes written in Python language, for processing spectral data combined with algorithms with in silico fragmentation rules. The results of the analyses have been compared with the information coming from peptide databases in order to obtain a good agreement between the theoretical and the experimental spectrum.

Given the growing interest in proteomics has become indispensable a parallel bioinformatics approach that has produced an increase in the development of new software libraries, including freely available and open-source software and at the same time an increase in the number of sequences in protein and peptide databases. Different computational methods can now be used to identify peptides and proteins. The most popular ones are based on the use of search engines [23] and protein sequence databases, but there are other approaches such as de novo sequencing (especially used when the genome of the studied organism is not well known) [24, 25] and the spectral library searches [26, 27]. As a result, there are several well established software applications like Mascot [28], X!Tandem [29], Sequest [30], MaxQuant [31] and Andromeda [32]. These have shown partial peptide recognition capabilities and additional limitations when used for peptidomics bioinformatics approach research, probably as a result of their protein-focus centred; in fact their algorithms generate candidate peptides from an in silico digestion of sequence entries present in a protein database, simulating the proteolytic enzymes activities [33]. These peptides fragments are

then used for the subsequent experimental to the theoretical parent (initial) and fragment (next) mass matching.

Python is a versatile programming scripting language that is widely used in industry and academia. It was created by Guido van Rossum and released in 1991. It is used for several tasks in bioinformatics including academic research, data manipulation, protein sequencing, data analysis, data visualization, accessing databases, and statistical learning. . Python was designed for readability and has some similarities to the English language with influence from mathematics.

Python's use of data analysis in mass and proteomic spectrometry began to be widely used around 2012, when several stable Python libraries were released, including pymzML, Pyteomics, and pyOpenMS. This has led to a growing interest in the Python programming language in the proteomics and mass spectrometry community. These Python libraries contain both functions for the reading of the intensities and the values m/z and to the metadata of a file in mzML format coming from mass spectrometry analysis and functions for the in silico generation of possible CID fragments of a peptide.

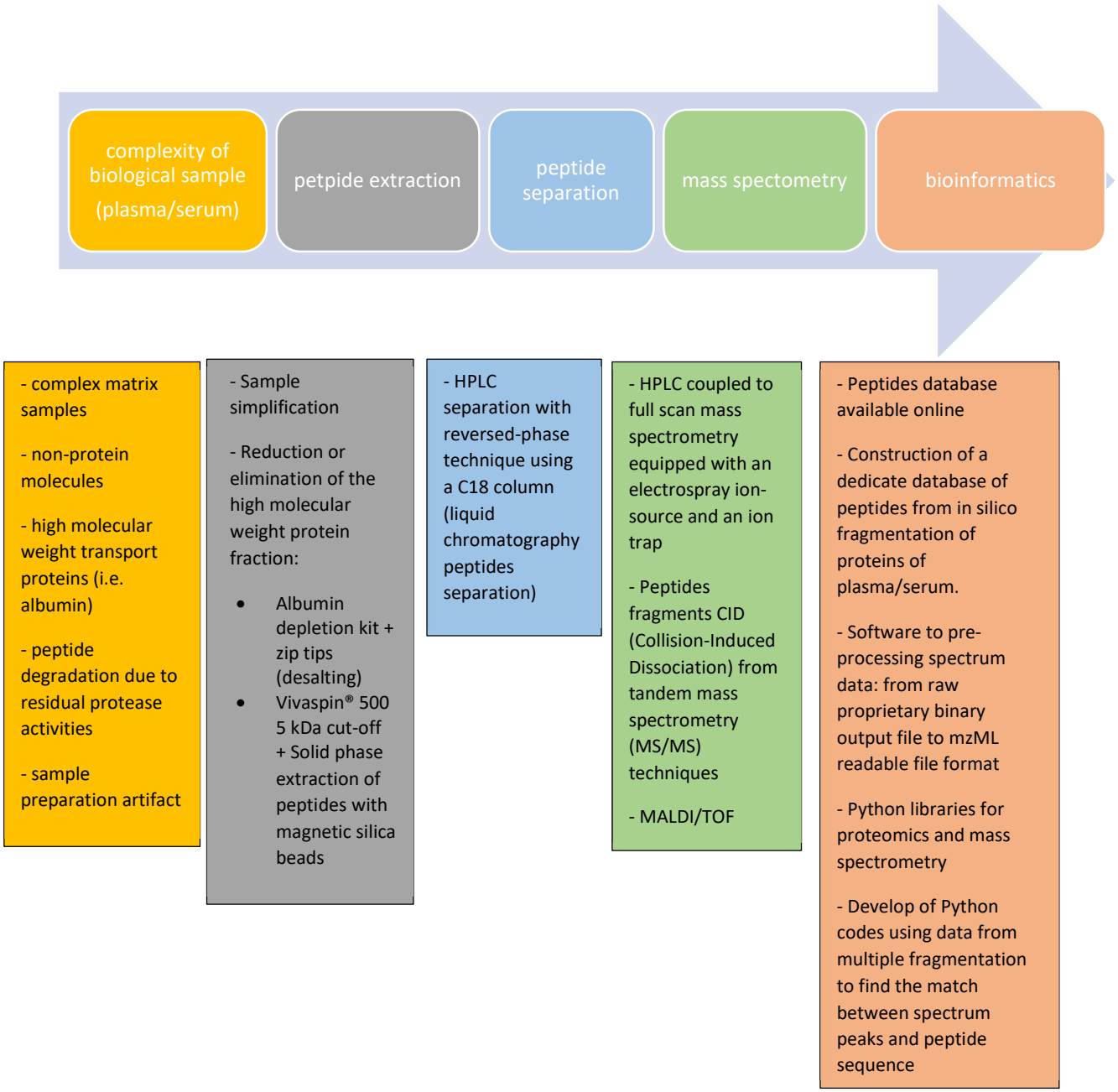


Figure 2. Pipeline scheme used in the current study to investigate the low molecular weight protein fraction presents in biological samples.

2. MATERIALS AND METHODS

2.1 Peptide synthesis

The peptides were manually synthetized using the standard method of solid phase peptide synthesis according to the 9-fluorenylmethoxycarbonyl (Fmoc) strategy with some modifications [ref]. Briefly, Wang- deprotected resin was treated with a coupling reaction mixture of 5 equivalents (eq.) of the appropriate Fmoc-aminoacid, 4.5 eq. of O-benzotriazol-N,N,N',N'-tetramethyluroniumhexafluorophosphate (HBTU), 5 eq. of N,N-diisopropilethylamine (DIPEA), at 0.2 M aminoacid final concentration in anhydrous N-methylpyrrolidone (NMP). A solution of 20% (v/v) piperidine in DMF was used to remove the Fmoc group. The resin was incubated with a cocktail of acetic anhydride, sym-collidine and DMF (0.5:0.5:9 ratio (v/v)) to block remaining amino groups on the resin [34].

All the resin was rinsed with dichloromethane and dried. The final cleavage of the peptide from the solid support and removal of all protecting groups were carried out with a trifluoroacetic acid (TFA): tri-isopropylsilane: H_2O: DODT (90:2.5:2.5:5) mixture. The peptide was precipitated in ice cold diethyl ether and then purified by preparative reverse phase high performance liquid chromatography (RP-HPLC) on an Agilent 1260 Infinity preparative HPLC equipped with a Kinetex C18 column (21.20 x 150 mm). The separation was obtained with a gradient starting with 5% solvent B for 5 min, linearly increasing to 70% solvent B in 30 min and up to 100% B in 10 min. The solvents used were 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The peaks of interest were then collected and evaporated under vacuum, then lyophilized and stored at 4°C. The identity and purity of the peptides were verified by RP-HPLC and the collected peak was analyzed in off-line mass spectrometry analysis using an Agilent 1100 series LC/MSD ion trap instrument.

With this method for this study the following peptides were synthesized, analyzed and purified:

peptide_accession	Peptide sequence H_2N-…-COOH	label	MW (g/mol)
pept_2_11	NITHFAIVASL	pept_AB0	1184.66
pept_15_9	FPSHANAAG	pept_AB1	870.40
pept_14_7	ATKEQLK	pept_AB2	816.47
pept_17_10	NGIGVTQNVL	pept_AB3	1013.55
pept_18_16	AQSTIEEQADTFLDYD	pept_AB4	1844.80
pept_19_11	NHEAEDLFYQS	pept_AB5	1352.53
pept_20_7	DATDQLS	pept_AB6	748.83

Table 1. List of peptides synthesized for this study.

2.2 Peptide Mixture solution

All peptides synthesized in the laboratory and used in mixture solutions have been individually stored in stock solutions at a concentration of 1mg/ml in water or in 10% ACN in the case of peptides not very soluble in water and stored at -18°C. All peptide mixtures used directly for mass spectrometry analysis (HPLC-ESI-MS) or used first in contact with magnetic beads have been used at a concentration of 1 pmol/µl in 90% chromatography eluent A (water + 0.1% FOA) and 10% chromatography eluent B (ACN + 0.05% FOA). For mixed peptides solution for MALDI-TOF spectrometry analysis has been used same concentration using 30 % ACN and 1% TFA in water as eluent.

2.3 Albumin depletion kit

The first analyses carried out on real samples were performed on the serum of healthy voluntary donors. The results showed spectra with the interference of high molecular weight proteins (i.e. albumin) despite the use of C18 silica magnetic beads, which should not bind such molecules.

Thermo Scientific Albumin depletion kit has been used to improve the analysis of serum components by quickly removing abundant albumin protein from serum samples thus simplifying the matrix.

The kit contents were as follows:

Pierce Albumin Depletion Resin, a liquid suspension, supplied as 50% resin slurry

Binding Capacity: 2.0mg of human serum albumin (HSA) for 200µL of settled resin

Binding/Wash Buffer (25mM Tris, 75mM NaCl; pH 7.5) has been optimized for binding HAS and reduced binding of non-albumin proteins

Pierce Spin Columns

The original protocol of the manufacturer has been adapted and optimized according to the type of sample with the aim of having a minimum consumption of materials with a significant spectrum signal with the least albumin interference.

The procedure we used is the following:

1. Shake the resin bottle to resuspend the resin. Using a wide-bore micropipette tip, transfer 260µL of the slurry (corresponding to 130µL settled resin volume) into a spin column and loosely cap the column.
2. Twist off bottom closure of the spin column and place spin column into a 1.5 mL collection tube. Centrifuge at 12,000 × g for 1 minute to remove excess liquid. Discard flow-through and place spin column back into the same collection tube.
3. Add 150µL of Binding/Wash Buffer to the spin column.
4. Centrifuge at 12,000 × g for 1 minute. Discard flow-through and place spin column into a new collection tube.
5. Apply 40 µL of serum sample to resin and incubate for 1-2 minutes at room temperature.
6. Centrifuge at 12,000 × g for 1 minute. Re-apply flow-through to spin column and incubate for 1-2 minutes at room temperature to ensure maximal albumin binding.
7. Centrifuge at 12,000 × g for 1 minute. Retain flow-through. Place spin column in a new collection tube.
8. Wash resin to release unbound proteins by adding 32µL of Binding/Wash Buffer for each 130µL of resin used.
9. Centrifuge at 12,000 × g for 1 minute. Retain flow-through. Place spin column in a new collection tube.
10. Repeat Steps 8 and 9 three to four additional times.

2.4 C18 tips

This tool became necessary after the use of the albumin depletion kit because many buffers and salts interfere with both MALDI-MS and ESI-MS.

The Thermo Scientific Pierce C18 Pipette Tips enable fast and efficient capture, concentration, desalting and elution of peptides . Each tip contains a monolithic C18 reversed-phase sorbent that provides binding and recovery of a wide range of peptides.

Material Preparation:

• Sample treatment solution: 2.5% TFA

• Wetting solution: 50:50 ACN:water; 200µL per sample

• Equilibration solution: 0.1% TFA in water; 200µL per sample

• Rinse solution: 0.1% TFA in 5% ACN:water; 200µL per sample

• Elution solution: 0.1% TFA in 50-95% ACN:water for MALDI-MS or 0.1% FA in 50-75% ACN:water for ESI-MS, up to 100µL per sample

Procedure

1. Set the pipettor to 100µL and secure the pipette tip tightly to the end of the pipettor for optimum tip-to-pipettor seal and sample aspiration.
2. Adjust sample to 0.1-1.0% TFA using 2.5% TFA.
3. Wet tip by aspirating 100µL of 50% ACN in water and then discarding solvent. Repeat once.
4. Equilibrate tip by aspirating 100µL of 0.1% TFA and discarding solvent. Repeat once.
5. Aspirate up to 100µL of sample (prepared in Step 2) into the C18 tip. For maximum efficiency, dispense and aspirate sample for 3-10 cycles.
6. Rinse the tip by aspirating 100µL of 0.1% TFA/5% ACN and discarding solvent. Repeat once.
7. Elute the sample as follows:

MALDI-TOF analysis: Slowly aspirate 5-100µL of 0.1% TFA in 50-95% ACN elution solution with or without matrix and dispense directly onto a MALDI plate.

LC/MS or LC/MS/MS analysis: Slowly aspirate 5-100µL of 0.1% formic acid or 0.1% acetic acid in a 50-95% ACN or methanol and dispense into an autosampler vial or well plate.

2.5 HPLC (High Performance Liquid Chromatography)

The acronym HPLC comes from High Performance Liquid Chromatography - chromatography high performance liquid. It is a chromatographic analytical technique that allows to separate, identify or quantify two or more compounds present in a mixture by exploiting the balance of affinity between a "stationary phase" (column filling) placed inside the chromatographic column and a "mobile

phase" (eluent) that flows through it. Depending on the chemical structure of the analyte, the molecules are delayed during the transition to the stationary phase, in particular a substance closer to the stationary phase than to the mobile phase takes longer to travel to the chromatographic column (retention time), compared to a substance with low affinity for the stationary phase and high for the mobile phase.

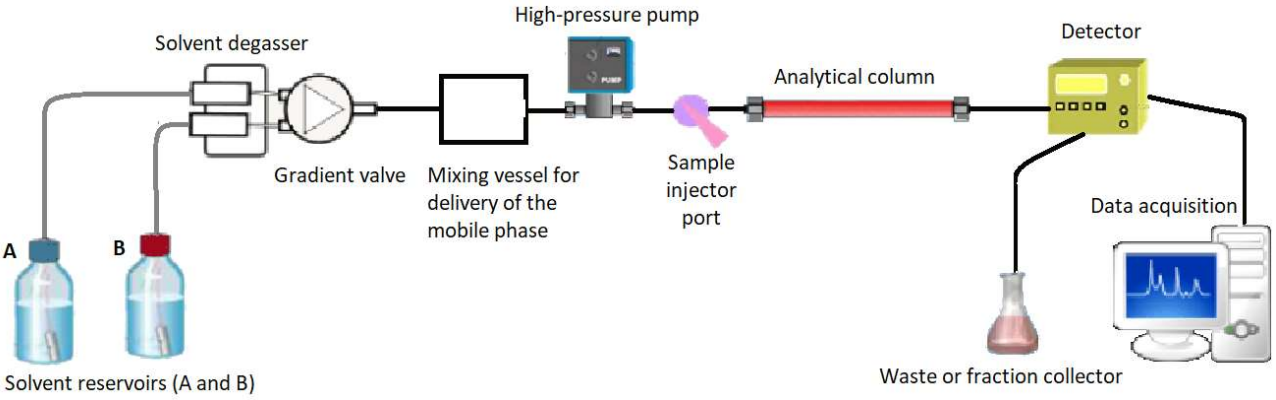


Figure 3. HPLC scheme

An HPLC instrument consists of the following parts:

- solvent bottle: one or more solvents (mobile phases) that can be used individually or in a mixture.
- There are usually devices for degassing solvents and eluent solutions and eliminate any undissolved particles. The preparation of the solution and replacement of a container of solvent, produce bubbles and slag. The latter are particularly dangerous since, if allowed to enter in the column, they could widen the bands; moreover, the efficiency of the pumping system would also be compromised.
- Pumping system with pressure up to 430 atm. Stable flow free of pulsations is between 0.1 and 10 ml/min, corrosion resistant components, high reproducibility flow control.
- Sample introduction system, consisting of a multi-way valve and a fixed volume circuit, or loop mode.
- Separative column and pre-column, if necessary.
- A detector that continuously monitor a certain measurable property of the column eluate and transforms it into an electrical signal. The detector requires high sensitivity, low noise (that is, insensitive to external changes such as temperature and flow), wide linear range,

good repeatability, and wide application range. The most common detectors consist of diode matrices that register absorption in UV-VIS (200-800 nm).

- PC to manage system and the data.

HPLC has two elution methods, which are called isocratic and gradient. Isocratic elution means that the composition of the mobile phase remains constant during the same analysis cycle, which is suitable for samples with a small number of components and little difference in properties. Gradient elution is a program to control the composition of the mobile phase over time within an analysis, such as solvent polarity, ionic force, and pH value. It is used to analyse complex samples with a large number of components and large differences in properties. The use of gradient elution can reduce analysis time, increase resolution, improve peak shape and increase sensibility.

The term high performance refers to the high speed and resolution in the separation of the different components of the solute and is obtained by using very small diameter particles of the order of tens of µm or less, packed in steel columns, glass or plastic, as stationary phase. Different chemical and physical parameters characterize the column type and thus the separation efficiency. Chemical properties, such as the type of surface of the stationary phase and pore size, affect sensitivity and retention. Physical properties of the column dimension, including particle size, column length, and inner diameter, affect efficiency and speed. Pore size is the average size of a pore in a porous packing, and it determines whether a molecule can diffuse into and out of the packing. Therefore, the pore size of the packing material in your HPLC column is important, since the molecules must fit into the porous structure in order to interact with the stationary phase. Smaller pore size packings (from 80 to 120Å) are best for small molecules with molecular weights up to 2000. For larger molecules with MW over 2000, wider pore packings are required; for example, a popular pore size for proteins is 300Å. For polypeptides and many proteins, it is best to choose 200-450 Å. With particle size one indicates the average size of the particles in the packing of the HPLC column. The analytical columns for HPLC have packages smaller than 2-3 µm, including 1,8 µm. Shorter columns with these particles can produce faster separations at high resolution. The smaller the diameter of the packed particles, the smaller the width of the solute bands and thus the amplitude of the resulting peaks. If the particle size of a column is reduced by half, the plate number doubles (assuming that the column length remains the same). However, if the particle size halves, the column backpressure increases four times. If the column length doubles, the plate number and analysis time also double. As column length increases, backpressure increases linearly.

The main chromatographic methods used for the separation of proteins and peptides are reported in table 2. In this study was used reversed phase method.

	Mechanism of separation
Ion exchange	Charge
Gel filtration	Molecular size
Reversed phase	Hydrophobic interaction

Table 2.

Reversed-phase liquid chromatography (RPLC) has become an essential tool in the separation and analysis of proteins and peptides because of its ability to separate proteins of nearly identical structure. Protein separation minimizes ion suppression, increases the dynamic range of detection, and reduces precursor spectral complexity, simplifying the data interpretation [35]. RPLC separates molecules based on surface hydrophobicity and is the most used and widely applicable LC technique. RPLC is most commonly applied as the final dimension of separation. This is due to the ability of RPLC to exchange the original solvent for the MS-compatible solvent. RPLC uses a hydrophobic or nonpolar stationary phase and a hydrophilic or polar mobile phase. The stationary phase is commonly composed of porous silica particles linked to alkyl chains between 4 and 18 carbon atoms or other inert nonpolar substances such as divinylbenzene (DVB). Retention is largely based on van der Waals interactions. Shorter alkyl chains (C4 and C8) are typically preferred for intact protein separation because they are less retentive than longer alkyl chains. The typical ranges of the main parameters of an inverse phase HPLC analytical column used for peptide samples are reported in table 3.

Column parameter	Dimension
length	150 mm
Internal diameter	1.8 – 4.6 mm
pore size	130 – 300 Å
particles size	2.0-3.5 µm

Table 3. Typical HPLC column parameters used for peptide samples

Larger proteins are usually much more hydrophobic and therefore strongly interact with the C18 matrix. This results in broad chromatographic peaks during the elution and in many cases to incomplete elution. Nonporous silica (NPS) particles with fused alkyl chains may also be used in the stationary phase. Use of NPS in the stationary phase gives the advantage of increased protein recovery, but is limited by loading capacity.

The mobile phases which were used are polar: water/organic solvent mixtures as acetonitrile, methanol, ethanol or isopropanol. When the sample is introduced to the stationary phase, the

hydrophobic molecules in the polar mobile phase adsorb to the hydrophobic stationary phase, permitting the more hydrophilic molecules to be eluted first. Decreasing the polarity of the mobile phase by increasing the percentage of organic solvent, usually acetonitrile, reduces hydrophobic interaction between the stationary phase and solutes, allowing for solute elution. More hydrophobic solutes will bind more strongly to the stationary phase, and thus a higher concentration of organic solvent is required in the mobile phase for their elution. In reference to the analysis of samples of amino acid sequences will be eluted first the small polar peptides followed by increasingly large and apolar peptides and finally proteins. [36-38].

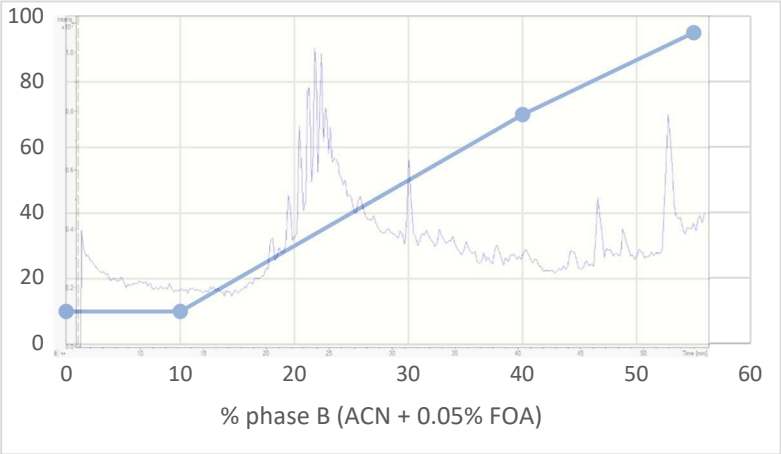
It is common to introduce in the eluents a small percentage (0.05 – 0.2% v/v) of organic acids such as formic acid or trifluoracetic, these ionic couplers significantly improve chromatographic efficiency (height equivalent to a theoretical plate). In HPLC the force that allows the eluent to flow in the column, is represented by the pressure that is applied by a head pump to the column and forces the mobile phase to flow inside the stationary phase. This allows not only to make the process faster but also to get more theoretical plates which means better resolution. Sample separation can be directly coupled to the mass spectrometry instrument with the advantage the strengths of both techniques can be utilised with reduced sample handling. However, time constraints due to continuously eluting molecules limit data collection and based on mass spectrometer electronics there is a limit to the number of ions that can be selected and further fragmented (tandem MS/MS, multiple fragmentation techniques).

2.6 HPLC-ESI-MS analysis

The HPLC-ESI-MS analysis has been performed using an Agilent 1100 chromatograph directly coupled to mass spectrometry at low and medium resolution, equipped with an electrospray ion-source and an ion trap analyser (ESI-MSD trap XCT). Chromatographic separation was conducted using a XBridge® Peptide, reversed phase C18 column (Waters) with dimension 1.0mm x 150mm and particle size of 3.5µm. The choice of the column has been made considering for the samples type and separation specificity of a wide variety of peptides and the reproducibility of the method. The injection volume of the sample was 8 µL.

The elution was carried out in linear gradient using mobile phases consisting in water with 0,01% formic acid (phase A) and ACN with 0,05% formic acid (phase B) with the following protocol: 0-10 min 10% B, going linearly to 70% B in 30 min, rising to 95% B in 15min, for a total of assay run time

of 55 min with a flow rate of 30 µL /min; the run was followed by 15 minutes of post-run to reach the initial conditions (Figure 4).



Time (min)	% phase A (H_2O + 0.1% FOA)	% phase B (ACN + 0.05% FOA)
0	90	10
10	90	10
40	30	70
55	5	95

Figure 4. Chromatographic conditions: Mobile phase gradient program

The eluent flow goes directly to the MS interface in the ESI source. All spectra were acquired in a mass range between 250- 2000 m/z. The main operating parameters of the ion trap are the following: drying gas flow, 8 L/min at 325 C; capillary voltage 3400 V; skimmer, 1:30; capillary exit, 130V; trap drive, 130V. The maximum accumulation time within the trap is set to 306 ms, while the maximum ionic target is set to 70,000. The acquisition is controlled by a data-dependent scanning procedure, 8 ions, among the most abundant. All measurements were performed in the positive ion mode. The full scan analysis has been setting in a range of m/z from 250 to 2000.

2.7 Solid phase extraction of peptides with magnetic silica beads (from serum or mixture of standard peptides)

- Part of protocol common to all types of beads:
  after a phase of homogenization by vortex, an aliquot of 20 µL of slurry containing the beads is transferred into a vial, placed in a Magnetic Separation Rack (MSR) and removed the supernatant. After washing with water we proceeded to the washing with the pre-load equilibration solution characteristic for each type of bead (with this step the subsequent

binding with peptides is facilitated). In the use of beads with biological samples the same volume of slurry of beads and biological sample was used.

Different protocols were used according to the type of chemical substituent present on the beads.

- Magnetic bead-based hydrophobic interaction MagSi-proteomics C18 beads (Bead Ligand: C18 alkyl group; Bead Matrix: Magnetic Silica particles; Bead Size: 1.2 µm)
  Beads were washed three times with 120 µL of 200 mM NaCl, 0,1 % TFA. The beads were re-suspended in 20 µL of peptides standard solution (all peptides were at a concentration of 1 pmol/µL) or in 15 µL of water, mixed with 20 µL of serum and an left at room temperature for 5 minutes. After incubation time, the tube was placed in the MSR to allow the separation of the beads from the supernatant, which is then discarded. The beads were washed three times with 300 µL of a 0.1% TFA solution in water and the bound fraction was eluted by incubation, at room temperature for two minutes, with 15 µL of a solution of 80% acetonitrile in water.

  2.7.1 Magnetic beads with weak anionic interactions. These magnetic silica beads contain in their surface carboxyl groups as the types B-COOH, B-GSH and B_PAB).
- B-COOH, Bead Ligand: Carboxyl group; Bead Matrix: Magnetic Silica particles; Bead Size: 1 µm; Alpha Nanotech inc.)
- B-GSH, Bead Ligand: glutathione ; Bead Matrix: Magnetic Silica particles; Bead Size: 1 µm)
- B_PAB, Bead Ligands: peptide P_AB4: AQSTIEEQADTFLDYD, peptide P_AB5: NHEAEDLFYQS; Bead Matrix: Magnetic Silica particles; Bead Size: 1 µm)
-

  The beads were washed three times with 200 µL a pre-loaded solution containing 0,02 M of buffer MES (2-morpholin-4-ylethanesulfonic acid) pH6 plus 1 M NaCl. Treat the beads with three washes of 200 µL of equilibration to adsorption buffer consisting of 0,02 M MES, pH 6 solution. The beads were re-suspended in 36 µL of adsorption solution and 4 µL of peptides standard solution (thus to have a final solution at the concentration of 1 pmol/ µL for all peptides) or with 20 µL of plasma with 1 µL of peptide standard (pept_7 at 1pmol/ µL concentration) and left a room temperature for 5 minutes. After incubation time, the tube was placed in the MSR to allow the separation of the beads from the supernatant that is

discarded. The beads were washed two times with 150 µL of adsorption solution, followed by a final wash with 100 µL of water. The bound fraction was eluted by incubation, at room temperature for two minutes, with 15 µL of desorption solution consisting in a 1% TFA in water.

2.7.2 Magnetic beads with weak cationic interactions. These magnetic silica beads contain in their surface amino group (B-NH_2, Bead Ligand: Amino group; Bead Matrix: Magnetic Silica particles; Bead Size: 1 µm; Alpha Nanotech inc.) . The beads were washed three times with 200 µL a pre-loaded solution containing 0,02 M of buffer Ammonium Acetate (AmAc) pH 4 plus 1 M NaCl. Treat the beads with three washes of 200 µL of equilibration to adsorption buffer consisting of 0,02 M AmAc, pH 4 solution. The beads were re-suspended in 36 µL of adsorption solution and 4 µL of peptides standard solution (thus to have a final solution at the concentration of 1 pmol/ µL for all peptides) or with 20 µL of plasma with 1 µL of peptide standard (pept_7 at 1pmol/ µL concentration) and an left a room temperature for 5 minutes. After incubation time, the tube was placed in the MSR to allow the separation of the beads from the supernatant that is discarded. The beads were washed two times with 150 µL of adsorption solution, followed by a final wash with 100 µL of water. The bound fraction was eluted by incubation, at room temperature for two minutes, with 15 µL of desorption solution consisting in a 1% TFA in water.

2.8 Preparation of GSH beads

This type of magnetic beads was obtained by preparing 150 µL of amino beads, previously dried by removing the supernatant of its slurry, in a conjugation buffer consisting of Sodium Phosphate 0,1 M solution plus Sodium Chloride 0,15 M, pH 7.2. Immediately before use, a volume of 150 µL of Sulfo-SMMC solution was prepared, obtained by dissolving 1,5 mg of substance in 150 µL in milliQ water. The crosslinker solution thus obtained was added to the NH_2 beads previously prepared and the reaction was allowed to proceed for 50 minutes at room temperature. After this time the product residuals which were not linked were removed with MSR removing the supernatant. A solution of glutathione was prepared by dissolving 3 mg of substance in 200 µL in milliQ water. The latter was added to the solution of beads with the crosslinker previously prepared after which the reaction was allowed to proceed for 30-40 minutes at room temperature (Figure 5). The GHS beads thus obtained was re-suspended in 150 µL of Phosphate buffer at pH 7.2 and conserved at 4°C.

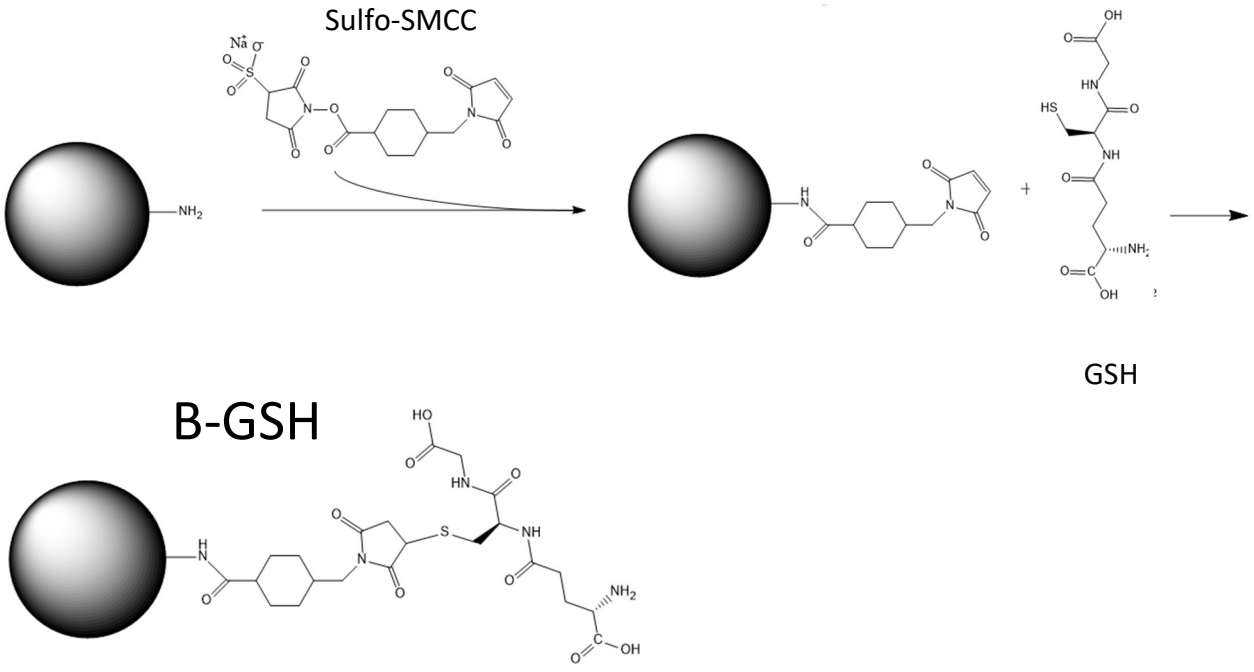


Figure 5. Preparation scheme of GSH functionalized beads

2.9 Preparation of beads functionalised with peptides (B-PAB)

To derivatize the magnetic silica beads with the peptides it was decided to use a carbodiimide crosslinker reagent. The water-soluble carbodiimide, 1-ethyl-3-(3-(dimethylaminopropyl)-carbodiimide (EDC) is widely used for labelling and crosslinking carboxylic acids to primary amines. In a volume of 250 µL of carboxyl beads, previously dried by removing the supernatant of its slurry, we added 1 mg of EDC in a conjugation buffer consisting of MES buffer (4-morpholinoethanesulfonic acid) at pH 6.0 plus 500mM NaCl solution. Reaction was allowed to proceed for 40 minutes at room temperature. The second phase of reaction is the conjugation with the aminic groups of peptides with the previous reaction product. Were weighed and added1,4 mg of P_AB4 (MW = 1844,80 g/mol) and 1,5 mg of P_AB5 (MW = 1352,53 g/mol) and allowed to react for 2 hours. After this time product residuals not linked to the beads was removed with MSR. The peptide beads thus obtained were re-suspended in 200 µL of MilliQ water and conserved at 4°C.

The preparation scheme of peptides functionalized beads using EDC crosslink reactive is reported in Figure 6. The Carboxyl-terminus of peptides can be coupled to NH_2-terminaed beads using EDC (1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride).

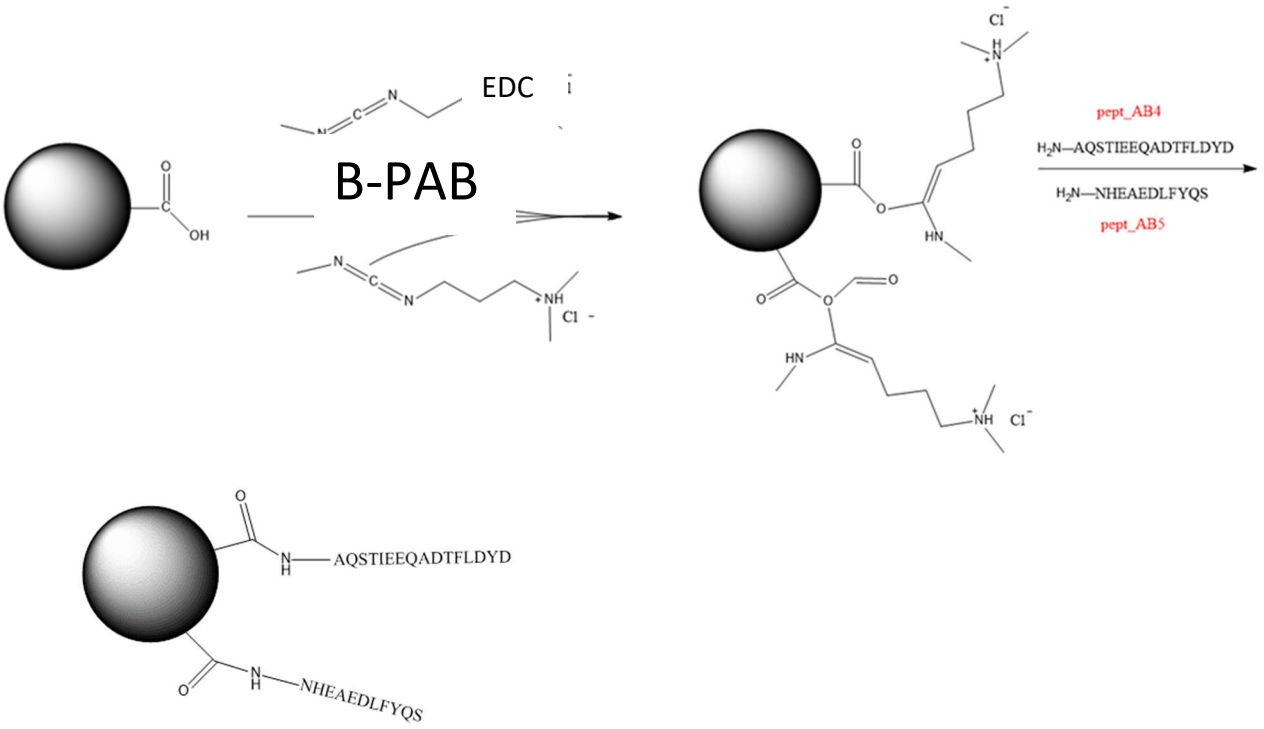


Figure 6. Preparation scheme of magnetic beads functionalized with peptides (B-PAB)

2.10 Preparation of serum from voluntary donors

These sample, at the time of collection, were treated following a standard protocol. After coagulation, which lasted for one hour at room temperature, the samples were centrifuged for 10 minutes at 1500 g. The serum samples were then aliquoted and frozen at -80°C until use.

2.11 Treatment of plasma samples (patients and voluntary healthy donors)

Plasma samples were thawed and 300 µl ultrafiltered using a Vivaspin® 5 kDa cut-off centrifugal filter device (Sartorius) to remove most high molecular weight proteins. The use of this tool involves centrifugation for 15 min at a speed of 12,000 (xg), as reported in the manufacturer's instructions. The obtained filtrate was divided into different aliquots from 20 µL and used later with distinct types of magnetic beads (functionalized differently). The types of magnetics beads functionalization used for this study were those with GSH, PAB and NH_2.

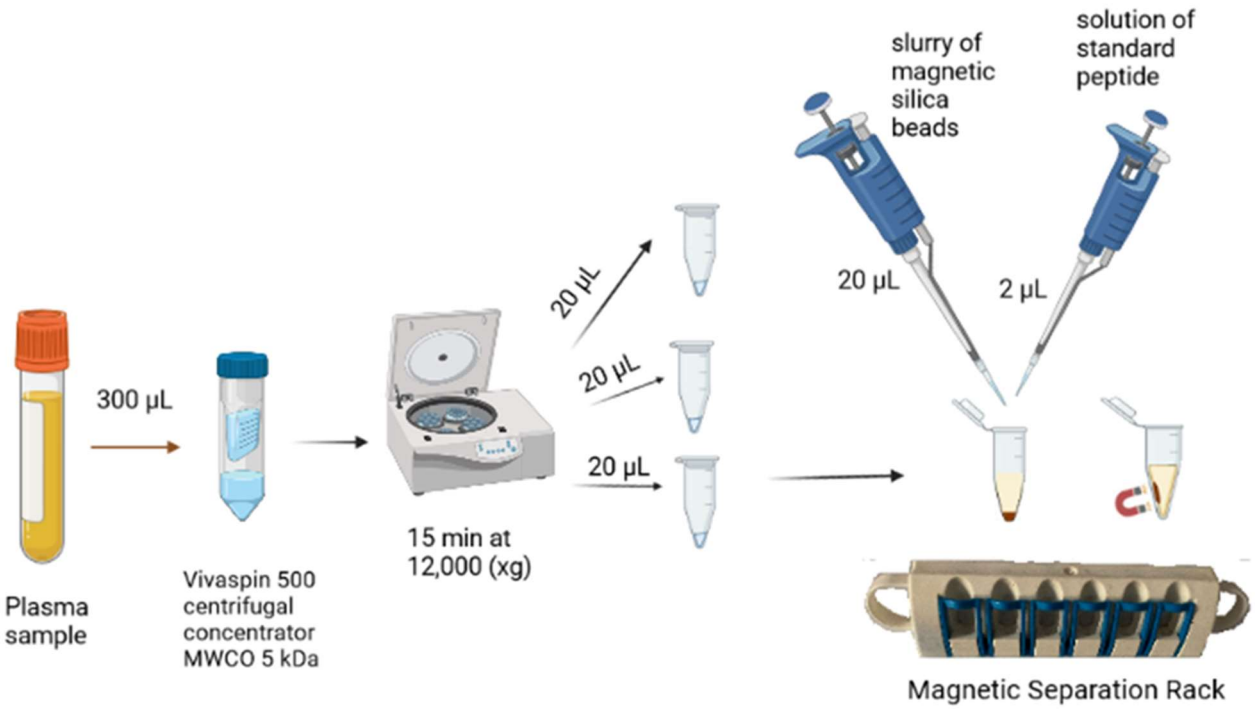


Figure 7. Scheme of sample treatment (created in BioRender.com)

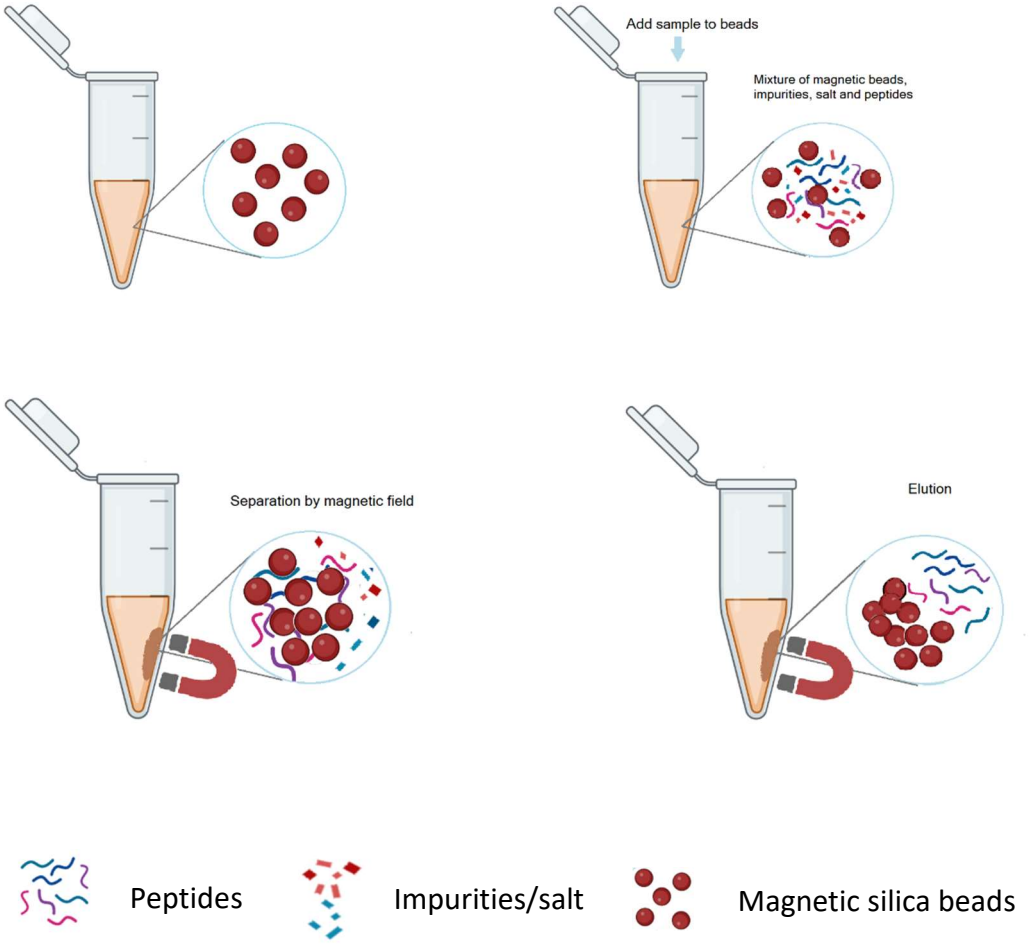


Figure 8. Scheme of separation mechanism with magnetic silica beads.

In case of MALDI-TOF mass spectrometry the eluate obtained by the extraction process with the magnetic beads was mixed in a ratio of 1:2 with the matrix solution (consisting of 6.2 mg of alpha-cyano-4-hydroxycinnamic acid in 1 ml of 36% methanol, 56% acetonitrile and 8% water). 1 µl of this mixture was spotted onto the MALDI target plate and left to dry under vacuum at room temperature for about 20 minutes.

2.12 MASS SPECTROMETRY

Mass spectroscopy is an essential analytical technique widely used in life sciences to identify and quantify proteins or peptides present in biological samples to determine their structure and properties and to compare their profiles in samples different [39]. Mass spectroscopy has existed for about 100 years, it is a common tool in chemistry, but for proteomics it has come into greater use in the last few decades. A mass spectrometer is an instrument used to measure the mass of a molecule after it has been ionized.

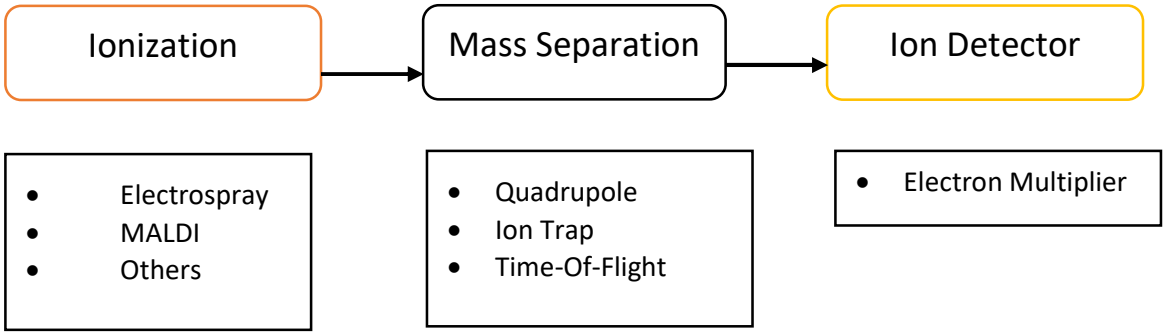


Figure 9. Block diagram of the components of a typical mass spectrometer

Regardless of the type of technique, the pattern of successive phases identifies the following 3 basic steps, as illustrated in the block diagram in figure 9. They are ionization (molecules receive a charge and eventually are fragmented), mass separation of ions which are accelerated and deflected in a vacuum under the influence of electric and magnetic fields, and ion detection. There are different types of ionization, mass separation, and ion detection used in various combinations with various strengths and weaknesses and different uses. A mass spectrometer measures mass-to-charge ratio or m-over-z (m/z) which can be used to determine the mass. Ion deflection is based on charge, mass and velocity, ions separation is based on mass to charge (m/z) ratio, and detection is proportional to the abundance of ions.

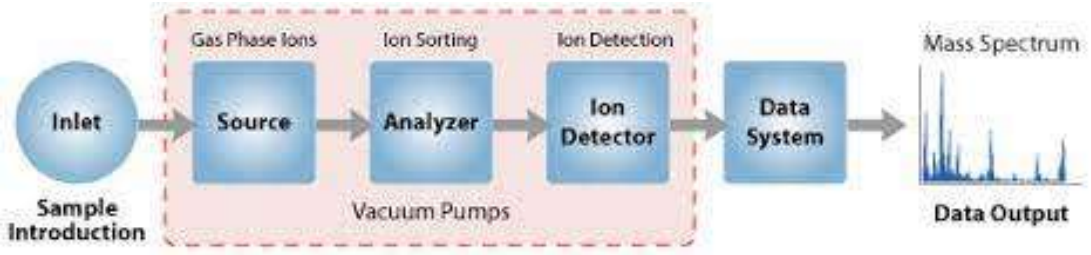


Figure 10. Instrumentation of Mass Spectrometry and pipeline from sample induction to data output

A more detailed scheme of a mass spectrometric analysis: from the sample to the spectrum display

- Introduction of the sample: the sample can be introduced in two ways.
  - on-line, if the spectrometer is directly interfaced with a separation system (for instance a liquid chromatography system) where everything that is eluted by the chromatographic column enters directly into the instrument [40]
  - off-line, if the sample is introduced directly into the instrument without the use of a previous separation system.

- Ion Source: molecules present in the solid or liquid sample are vaporized and ionized using different techniques, and may undergo any fragmentation. These ions can then be manipulated using electric and/or magnetic fields. Generating ions can be done in many ways, for instance using *laser desorption (LDI)* or *electrospray (ESI)*. For a given ion source geometry, the ionization efficiency is primarily a function of the number of electrons in the atom or molecule.

- Mass Analyser: the ions are separated based on their mass over charge ratio (m/z). A particular approach uses the so-called flight time, which consists in measuring the time needed for the ions to travel through a linear path immersed in a constant electric field. When two ions have the same charge, the ion with the highest mass will take longer to travel the same distance.

- Detector: the mass separated ions are converted into measurable signal proportional to the number of ions present that is transmitted and amplified to the acquisition system

- Data analysis.

The data processing system records these electrical signals as a function of the m/z ratio. The diagram produced by the m/z ratios, and their respective relative abundances, represents a mass spectrum that is displayed by a computer as a diagram of peaks, reporting in the x-coordinate the ratio m/z and in the y-coordinate the abundance of ions.

The accuracy and overall performance of a mass spectrometer are a function of the sections it is composed of.

In the instrument, the parts going from the ion source to the ion detection are placed under vacuum in order to avoid signal losses due to collision with residual gas molecules, and hence generating a long mean free path, but also to keep the ability to maintain a uniform electric potential over the path

.2.12.1 Ion formation

The ionization phase occurs in the source:

$$\text{MOLECULE} \rightarrow \text{ION}$$

$$[M] \rightarrow [M+H]^+$$

Depending on the energy deposited in the ion a hard ionization may occur, where the excess of energy causes fragmentation of the molecule, or a soft ionization, where the fragmentation process is reduced to a minimum or non-existent.

The principles of modern analysers have been known since the 1940s [41], the ionization of proteins and peptides has long been a limiting step for their analysis. Towards the end of the 1960s ionization was by electronic impact (EI) [42, 43] or chemical ionization (CI) [44].

EI is a hard technique that uses a high-energy electron beam (70eV) that impacts the analyte in the gas phase. In this technique the large excess of energy on the molecular ion $[M]^{+\cdot}$ causes an extensive fragmentation, useful for obtaining repeatable structural information, for the same molecule, at a qualitative and quantitative level. It must be remembered that the fragmentation, for the same compound, is not a random process. Operating in the same experimental conditions brings to the same fragments at the same intensity and therefore the same spectrum.

The CI technique, on the other hand, produces ions by transferring protons, creates proton molecules $[M+H]^+$ and deprotonates $[M-H]^-$ with minimal excess energy. CI is achieved by a reaction between the analyte molecules and the ions of a reagent gas in high concentration, generated by pressure EI. The number of ion-molecule collisions is sufficiently high during the residence time of the reagents within the source.

Both methods require a volatile sample, and create an extensive fragmentation of the same, therefore they are suitable for relatively small molecules and in the gas phase; conditions which are very far from the main biomolecules. It was necessary to wait for the mid-1980, with the advent of two ionization techniques, to allow mass spectrometry to predominately enter the biological sphere. It is the advent of MALDI (Matrix assisted laser desorption ionisation) [45, 46] and ESI (electrospray ionisation) [47]. These two techniques are the most important ionization sources used in proteomics/peptidomics. in both techniques a soft ionization is provided, and this is useful to prevent the formation of molecular fragments.

2.12.2 ESI

After the first studies on ionization by electronebulization by Dole and colleagues in the late 1960s [48], it was Fenn (Nobel Prize winner) who realized what is considered the modern electronebulization at the base of the ESI technique [47, 49].

The Process of Electrospray Ionization can be divided into three stages: droplet formation, desolvation, and gas phase ion formation.

Droplet formation: Under the action of a high voltage electric field (about 5000 V), the analyte solution is infused into a thin metal capillary to form a positive charged droplet where the molecules are induced by an electric field to pick up extra protons from the solvent. The solution is attracted to the electrode with opposite charge, while undergoing an opposite force dictated by the surface tension (figure 11).

This spraying process is usually assisted by a coaxial gas flow. The initial ESI droplets have radii in the micrometer range.

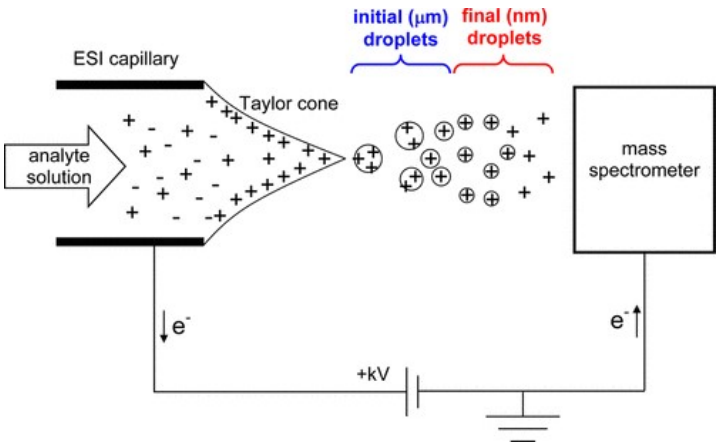


Figure 11. Scheme of an Electrospray Source

When the electrostatic force is higher than the surface tension, a cone of the solution of interest (called Taylor cone) is generated and from these small drops are created, as shown in figure 12.

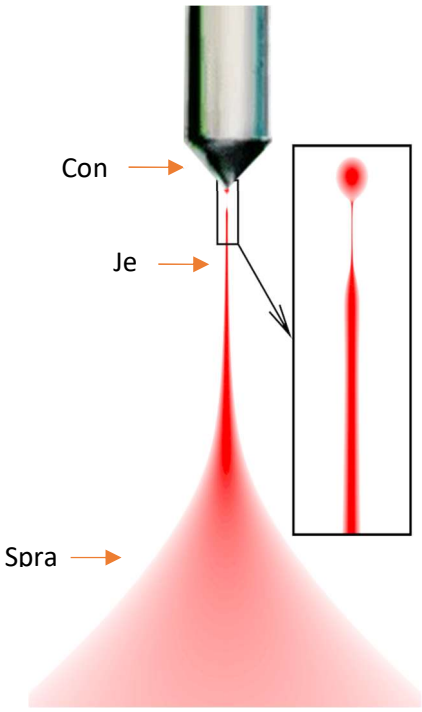
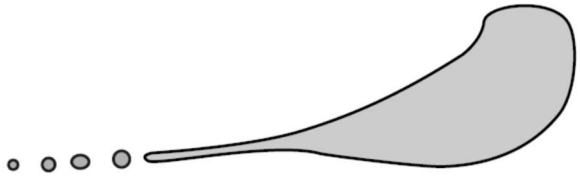


Figure 12 Representation of the Taylor cone

Desolvation. The droplets emitted from the Taylor cone undergo rapid solvent evaporation by the counter current of the heated dry gas (like nitrogen gas). In the case of aqueous/organic mixtures, the organic component usually evaporates more readily, causing a gradual increase in water percentage. The evaporation phenomenon continues to a point where the electrostatic repulsion is

greater than the surface tension of a single drop, exploding it, and generating smaller and smaller drops. This limit is given by the Rayleigh equation. [50].



$$q^2 = 8\pi^2 \varepsilon_0 \gamma D^3$$

Figure 13 Decomposition of the Taylor cone in drops and its coulombian explosion in accordance with the Rayleigh equation. q is the charge; ε_0, the permittivity of the environment; γ, the surface tension, and D, the diameter of the supposedly spherical drops

It should be noted that the phenomenon continues as long as the ion is removed from the drop itself (ion evaporation model), [51, 52], or the solvent completely evaporates (residual charge model) [53].

Gas phase ion formation. The size of the charged droplets reaches the nm level, and the ions in the droplets turn into gas phase ions.

The electronebulization depends on the concentration of the sample, therefore it was sought during its evolution to reduce the volume of the flow that reaches the source, increasing the Instrumental sensitivity.

Recently Tempst and collaborators have realized the "injection adaptable fine ionization source", pushing up to a flow of a few nL/ min [54]. Ions (generally positive) are accelerated towards the cathode, here they pass through the skimmer, where they access to the analyser, in which a vacuum is applied more pushed (10^{-7}, 10^{-8} torr). This ionization, unlike the MALDI one, can generate multi-charge species, therefore allowing to analyse molecules normally not measurable in the range of an analyser mass (generally limited to m/z 3000).

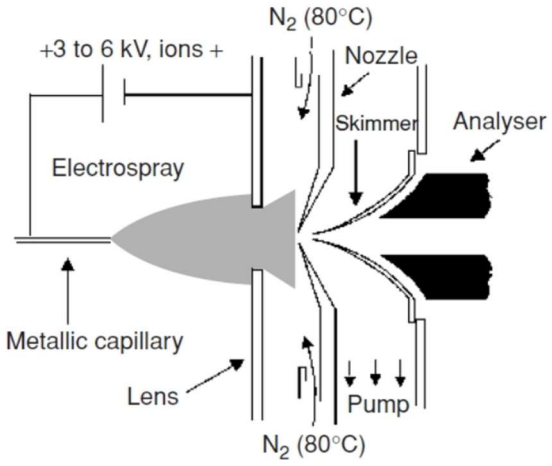


Figure 14. scheme of an ESI source

2.12.3 MS/MS tandem and type of fragmentation

Collision-induced dissociation (CID) is one of the most common techniques in tandem mass spectrometry (MS/MS) [55, 56, 57], and consists in a procedure for examining individual ions in a mixture of ions. Some ions selected during full scan are isolated from their characteristic values m/z, activated by the collision and allowed to fragment. The resulting product ions are examined in a second mass measurement step. The resulting product ions are examined in a second mass measurement step. In addition to the ion scanning function, the ion trap can select a precursor ion in a step that ejects all other ions from the trap. Two important discrete states can be distinguished in CID: the first represents the energy gained in collision events that increases vibrational and stretching modes of the peptide backbone due to the inelastic impact against an inert gas (helium); the second is given by this increased motion and causes molecular breaks of the ion that occur typically at the peptide bond.

The product ion fragments are recorded by scanning the RF voltage to perform a second mass-analysis scan (level 2 spectrum). Technically, two types of CID, high and low energy, can be distinguished. Since the high energy ones are basically very specialized analysis. Low energy CID techniques (< 200 eV) are the most common in the biological field and they are used, for example in the analysis of peptide sequences.

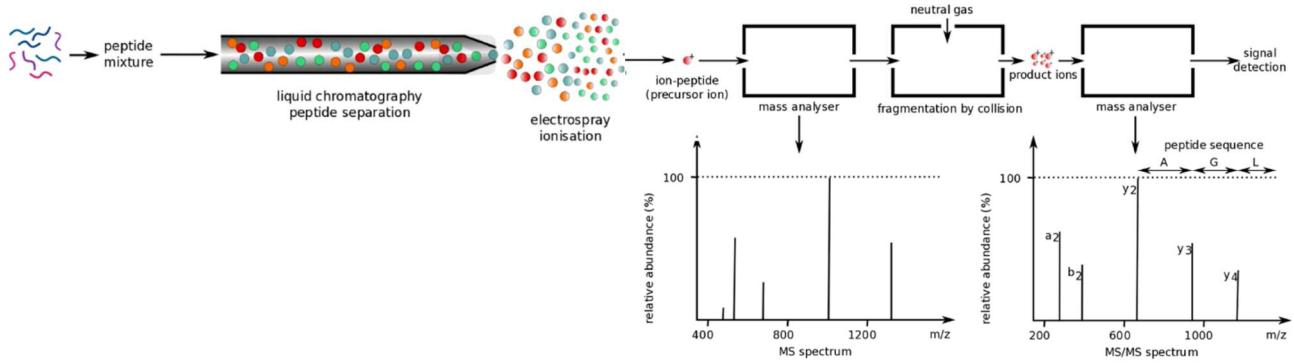


Figure 15. Diagram of a tandem mass spectrometer, showing the entrapment of an ion of interest (the signal of the first mass spectrum from first mass analyser); its fracture and subsequent analysis (second spectrum from second mass analyser)

The mixture of extracted peptides is fractionated by size or other properties by liquid chromatography and then enter the mass spectrometer. First, the fractions off the column are ionized—-they're given a charge—-and then they enter into the mass analyser, where the initial precursor ion spectra of the entire peptide is produced.

Then, the peptide is fragmented further by collision with a neutral unreactive gas like argon. Its fragment ions are analysed to produce a second spectra that can be used to determine or estimate the peptide sequence.

The available instrument 1100 LS/MSD TRAP XCT selects three ion precursors from each full scan spectrum level 1, and individually proceeds in the ion-trap to a fragmentation of them giving rise to a spectrum MS/MS of level 2 as illustrated in figure 16.

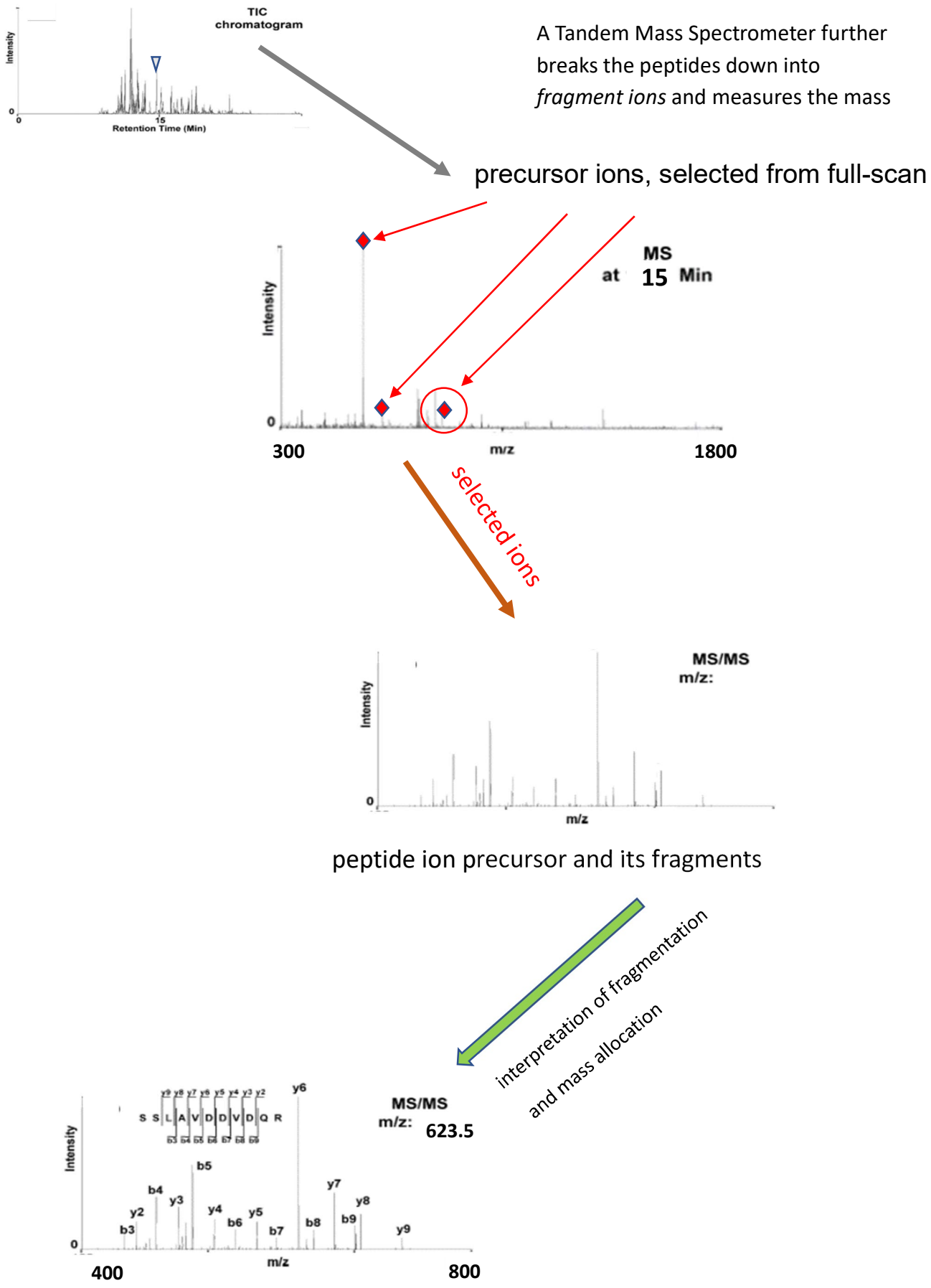


Figure 16. Scheme of obtaining a MS/MS spectrum instrument 1100 LS/MSD TRAP XCT

2.12.3.1 Types of fragmentation

These are the fundamental processes that support the peptide fragmentation:

– Selected ions are accelerated into a field of inert gas (He N Ar Xe) at moderate pressure.

– The energy gained in collision events increases vibrational and stretching modes of the peptide backbone

– The increased motion of the energized peptide causes its breaks

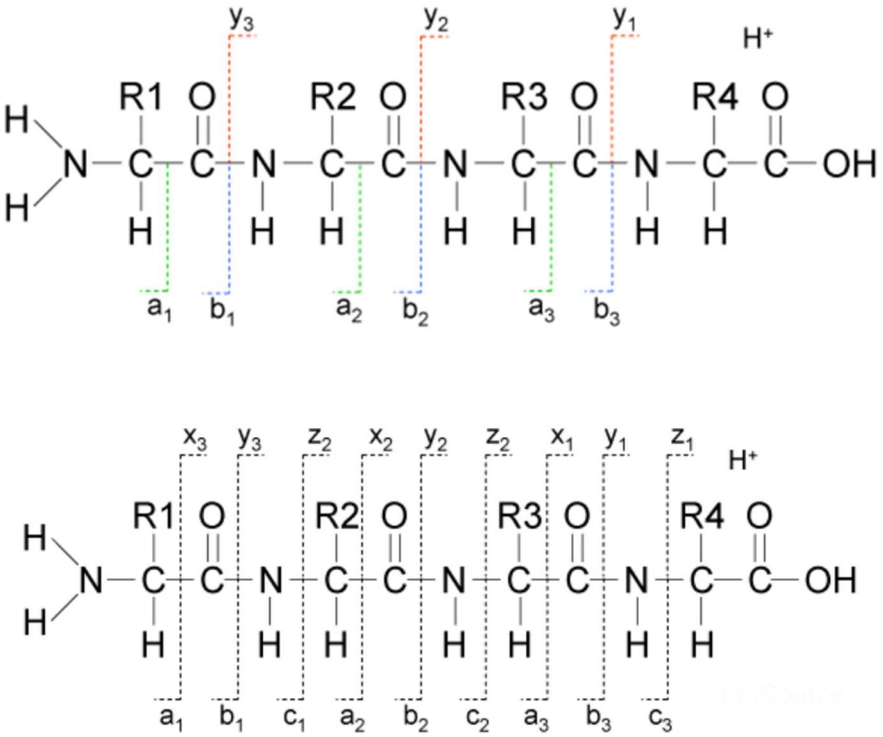


Figure 17. Ions of peptide fragmentation according to the nomenclature proposed by Biemann [58,59]

The most common bond breakage fragmentation observed in low energy collisions concerns the direct cleavage of peptide bonds which results in the formation of **b** ions (containing the N-peptide terminal) and **y** ions (containing the C-peptide terminal). The other splits of the spine bond result in ions **a**, **x**, c and **z**. The rupture of two peptide bonds simultaneously gives rise to internal fragments along with b and y ions [60]. While readily observed and diagnostic for b ions, an ion occurs at a lower frequency and abundance in relation to b ions. Another typical non-ionic loss and therefore not visible as ion is the loss of a water molecule and the side chain groups (Rn) can also be broken, sometimes more easily than the peptide chain.

2.12.4 Detectors

After separation, the ions must be recorded by the instrument to be measured. Detectors are based on electron cascade multiplication (EMT). They can be distinguished on this basis: flat or linear detectors (Farady cup), "channeltron" that have a cornucopia shape, and microchannel plate detectors (MCP).

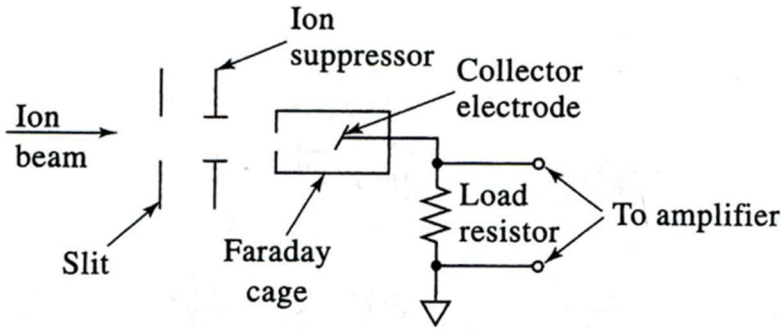


Figure 18. Faraday cup detector

The "Faraday cup"(figure 18) is not an EMT, Its mechanism consists in relating the ions entering the detector with the current generated by them outside the detector. In practice the ion enters the cup through a hole and its charge is neutralized, generating a small current. By measuring this current one can calculate how many ions have entered the detector. The "Faraday cup" is not as sensitive as an electromultiplier, but is accurate because of the direct relationship between ion and generated current. The response is independent of the energy, mass and chemical nature of the ion. The "channeltron", in fact, is nothing more than a coated dynode of a semiconductor. For obvious reasons, the entrance is kept on the ground (it must be able to accommodate both positive and negative ions). The ions incident on the inner walls cause a cascade amplification, emitting other electrons until the final signal is realized. This amplification is remarkably high, up to 108 electrons per single ion.

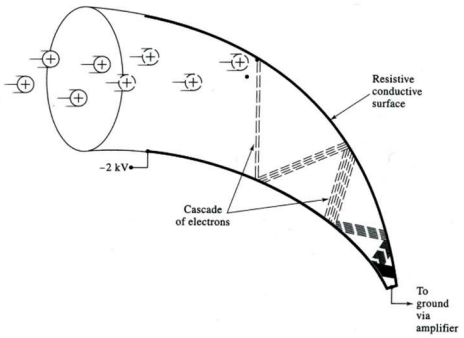


Figure 19. Illustration of a channeltron ion detector.

MCP is a planar element closely related to an EMT. In practice it is as if there were many EMTs placed spatially next to each other.

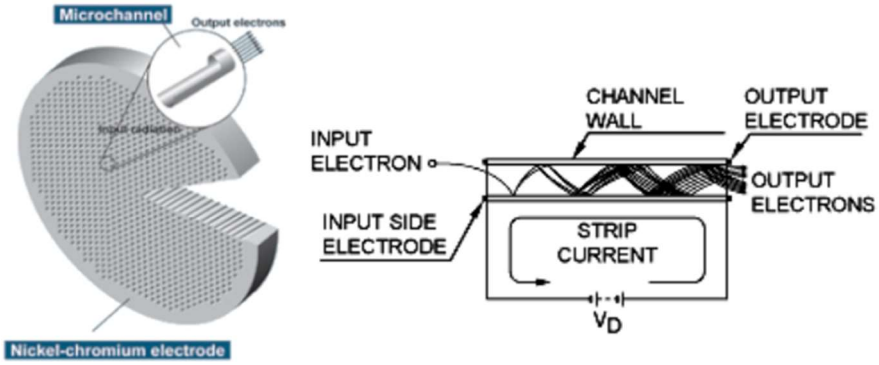


Figure 20. Image and scheme of a microchannel plate detectors (MCP)

It consists of a plate in which are placed thousands of channels with a diameter ranging from 10 µm to 100 µm and a length of the channel calculated so that the ratio of this to the diameter is between 40 and 100. It has been shown that the characteristics of the single channel do not depend on its length but only on the ratio between it and the diameter. Each channel is covered with a semiconductor material (dynode), which is used for the production of additional cascading electrons. Compared to traditional EMT, MCPs they have two advantages: they are small in size, and have lower time resolutions. Several MCP are often aligned in an opposite direction to amplify the signal.

2.12.5 Mass spectrum

From incoming ion flow from the coupled HPLC instruments, based on the set accumulation time value, the signals of mass-to-charge of ions and their abundance are registered. Various spectra are scanned in tandem MS/MS spectrometry and for each of them some precursor ions are selected, typically those associated with peaks m/z having a higher abundance.

The mass spectrum is a graph (in this case a histogram) in which on the x-coordinate is reported the ratio m/z and on the y-coordinate the abundance. The highest peak is defined as the base peak, which is the one with the highest relative abundance. The number of separate peaks observed is greater than the number of species. This can be understood by the presence of what are called isotopic peaks (M+2, M+3,. . .) which may accompany the main one.

2.12.5.1 From raw spectrum data to extracted signals.

The study of the mass spectrum aims at determining the presence, the position and the relative height of peaks, providing in this way a measurement of the mass and the indication of the relative abundance of a specific analyte. The large volume of data produced in a typical mass spectrometry experiment requires computers to be used for data storage and processing.

The mass spectra can be recorded in profile (also called continuum) mode but are often 'centroided.' Centroiding is, in effect, a process of peak detection for a profile mode mass spectrum (hence in the *m/z* dimension, not in a chromatographic dimension).

Data processing includes all steps that transform raw data into meaningful information. Its inputs are raw or converted data files, and the output is a list of characteristics in each sample (with associated chromatography, MS, and MS/MS data for each feature) for further analysis. There are several steps involved in converting the profile spectrum into a mass list, and many metadata information about the mass spectrum that can and should be collected during the data analysis process, which can prove useful to improve the precision of the resulting mass list.

Data format conversion refers to conversion of raw data files to usable formats for any subsequent data processing that does not intentionally interpret the raw data. Most instruments output their acquired and stored data in a very specific and often proprietary binary file format. For this reason it is needed to develop an open XML-based format to encode mass spectrometer output files data and then write software to use this format for archiving, sharing and processing. For example, conversion of data files from proprietary vendor format to open data formats such as mzML [61] or mzXML [62,63]. Data reduction such as peak centroiding and deisotoping is often performed during this transformation from proprietary formats to peak lists. Information about MS precursor signals, details of a given scan, raw data m/z data and associated metadata (e.g., instrument settings and description, capture modes, etc.) with respect to the files from which they were derived is added to the peak list. The lists of peaks indexed by spectrum number (index) are then used as input for the next data analysis.

2.12.6 Analysers

A mass analyser is a component that allows to separate the ions according to their mass to charge ratio and outputs them to the detector where they are detected and later converted to a digital output. There are two big categories: the analysers that separate in space, and those that separate in time. The most common space-separating analysers are quadruples, and TOF (time of flight). The analysers over time are tools that can trap ions (ion traps) and allow repeated analysis on the same ions, or on their fragments: MS/ MS or in general MS^n. Ion selection with an m/z ratio in the spectrum will generate different peaks because they reach the detector at different times.

Five concepts concern the performance and analysis of a mass spectrometer and are generally related to mass analysers.

• the mass range;

• the rate of analysis;

• the transmission;

• the accuracy;

• the resolving power;

The mass range defines the limit m/z in which the instrument can operate. It is expressed in Thomson (m/z).

The rate of analysis, or "scan speed", is the ability of an analyser to record measurements in a given mass range. It is expressed in units per second (us^{-1}). The transmission expresses the ratio of ions reaching the detector and ions entering the analyzer. Accuracy indicates how close the measured value [(m/z) measured] by the instrument is to the theoretical value [(m/z) theoretical]. It is expressed in ppm.

Resolution is the ability of the instrument to provide two distinct signals for two different m/z. Two distinct signals for two different m/z, meaning that the valley created between the two peaks must not exceed 10% of the peak height with lower intensity. It essentially defines how well two peaks are separated from each other. So if ΔM is the smallest mass difference between two peaks M and M + ΔM solved, the Mass Resolving Power is defined as:

$$R = \frac{M}{\Delta M} \qquad (2.1)$$

A specific m/z value and also the method like 10% valley or 50% valley or full width at half maximum (FWHM) must be given. Resolving power is usually a large number (up to 2,000,000 for FT-MS) and is used as “performance” parameter.

ΔM is usually derived from a mass peak´s full width at half maximum (FWHM), as shown below for the exemplary analysis of a mixture of two compounds X and Y:

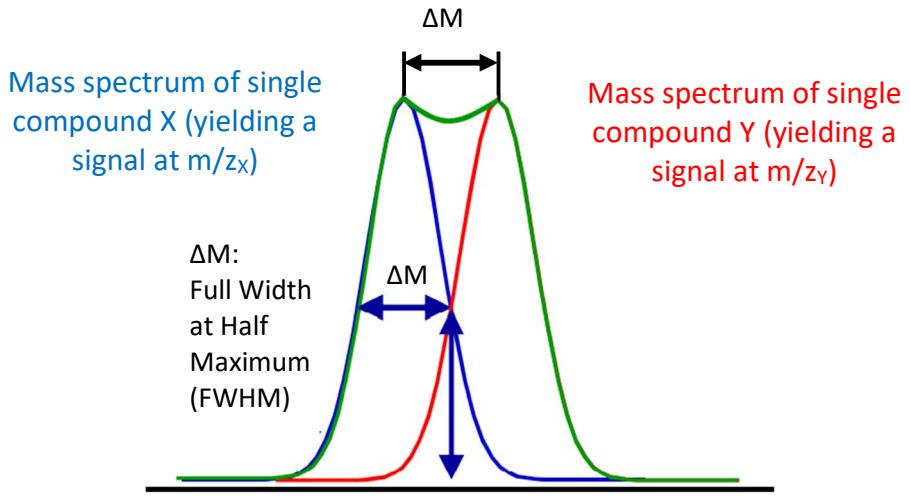


Figure 21. Mass spectrum obtained from mixture [X+Y]: The resolution (i.e. FWHM) provided by the mass spectrometer used here does not allow baseline separation of X and Y. To achieve this, narrower peaks (i.e. smaller FWHM = higher resolution) would be required.

2.12.6.1 Ion trap analyser

The quadrupolar ion trap consists of three electrodes, in which appropriate electromagnetic fields create a potential hole that allows the incorporation of ions. They can be confined in such hole for a long period of time (in the order of hundreds of *ms*) as shown in the following figures.

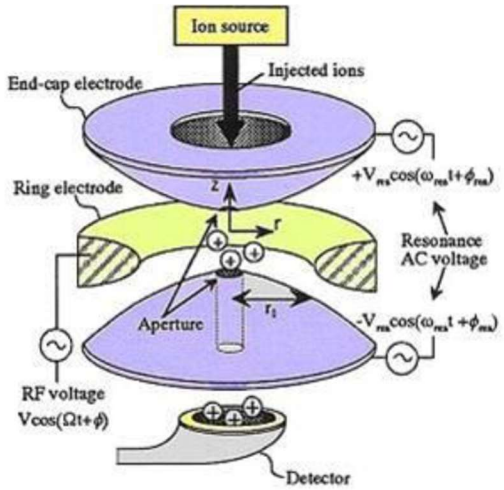


Figure 22. An ion trap scheme

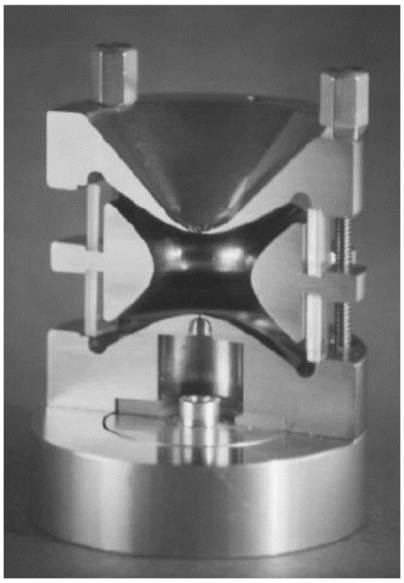


Figure 23. An ion trap cut along the vertical axis, note the hyperboloid geometry of its electrodes.

Two electrodes are defined as "end-cap", (placed along the vertical axis, one at the bottom and one at the top) and resemble two saucers, while the central circular electrode consists of a toroid.

The QITMS (quadrupole ion trap mass spectrometer) can function as a mass spectrometer and as an ion "storage". It is formed by three electrodes with hyperbolic inner surface where the two "end-caps" are perfectly identical.

An "end-cap" has a single small hole in which the desired ions can be selected, while the other electrode has several small holes that allow the ions product to reach the analyser.

The geometric choice of electrodes was designed to produce an electric field closer to the ideal quadrupole, when applying an RF (radio frequency) to the central electrode while a continuous electric field (DC) is applied to the two lateral electrodes. Such a potential constantly generates a

focusing (towards the center) and a defocusing (outward) that is a hyperbolic field able to trap ions. Ions that weigh less will be closer to the center of the trap, since they will "swing" less; vice versa for heavier ions. In the separation process the lowest ratio m/z ions will be extracted first (the process takes place by increasing the RF of the central electrode, generating a ramp of potential) while heavier ratio ions later. Ions with higher kinetic energy generating an outflow of ions that are ejected sequentially according to their increasing ratio m/z and directed towards the detector.

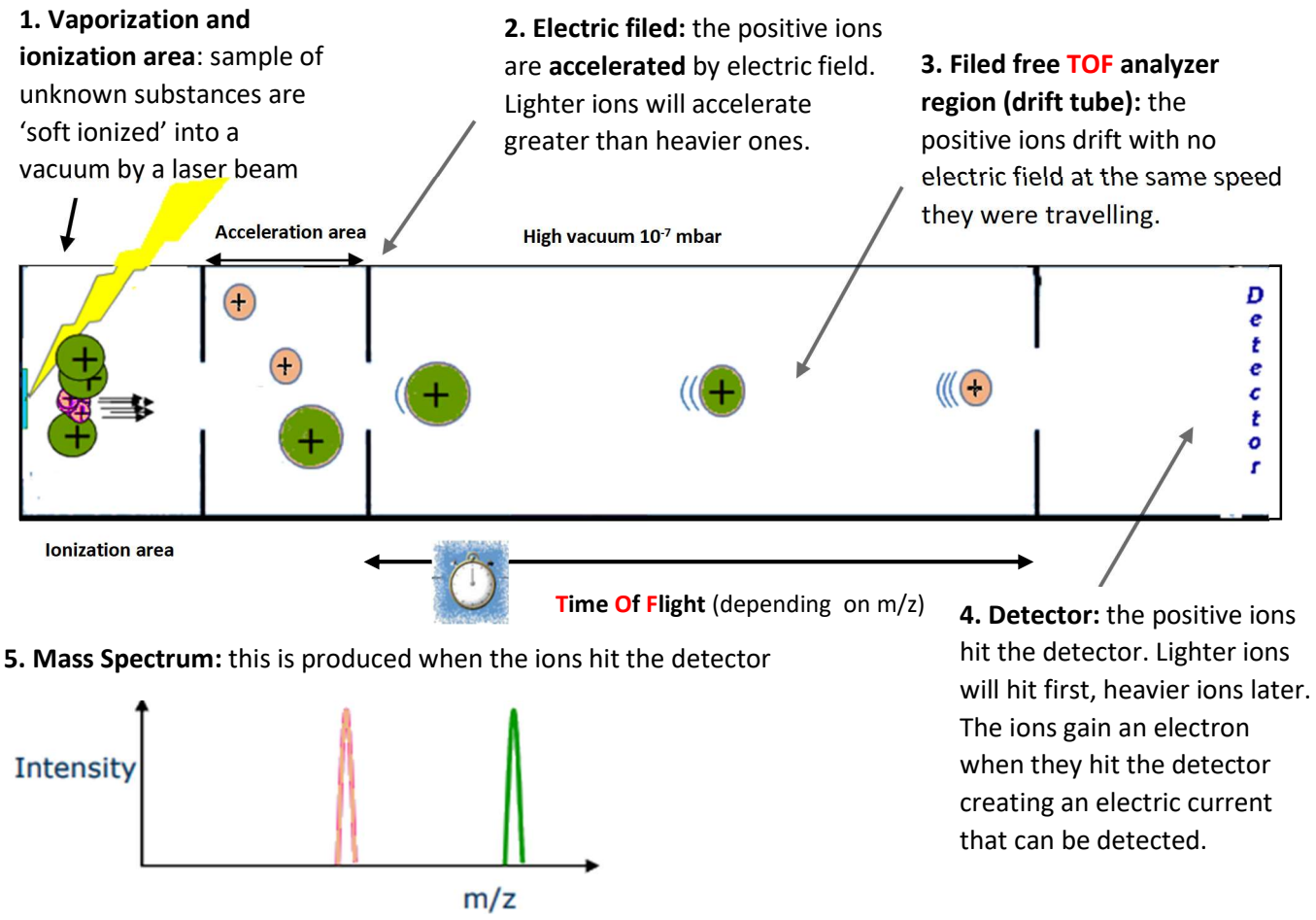


Figure 24. Scheme of basic principle of MALDI Time-of-Fight based mass analysis.

2.12.6.2 TOF

The MALDI source is typically associated with a Time Of Flight (TOF) analyzer. The concept of separation of ions according to their mass/charge ratio versus their flight time dates back to 1946 [64] but the acquisition and detection electronics were not yet sufficiently developed to make this type of analyzer usable. After two years the first TOF instrument was built by Eggers and Cameron [65]. However, the first instruments did not arouse the expected interest in scientific community. The main reasons were their low resolution and some problems of electronics.

Following the development of the MALDI source [45], the TOF analyzer began to be effectively used: after irradiation with laser light, essentially two processes take place, or the ablation or desorption of the crystal and the ionization of the analyte. The fundamental improvements that were introduced: were the delayed focusing [66], delayed ion extraction [67,68], and ion mirror [69].

The ions products are ejected in packets that are accelerated by a decreasing potential difference (from 20 to 15kV) that gives the same kinetic energy to the ions formed. These ions are then transferred to the flight tube where a high vacuum is generated (10^{-7} mbar). At this point they separate according to their mass and the time taken to reach the detector is proportional to the square root of their mass to charge ratio (m/z). The greater the mass of the protein, the more slowly it will reach the detector. The flight time is therefore inversely proportional to the mass. The qualities of MALDI-TOF lie in its sensitivity (sub picomolar), tolerance to salts and buffers, while a weak point is the resolution (R). The resolution for a TOF is of the order of 1000 parts per million (ppm) [70] and is referred to as the amplitude of the peak at half height. The main causes of this limit are the distribution of the initial uneven kinetic energy, small differences in flight time (ions are not exactly generated at the same time or place and therefore do not necessarily acquire the same initial kinetic energy) and verifiable collisions in the process of desorption and ionization that can increase dispersion in time and energy. In light of these imperfections of the system some devices allow you to greatly increase the resolution and accuracy of the measurement like using a longer flight tube or introducing an electrostatic reflector. The electrostatic reflector designed by Mamyrin Karataev in the 1971 [71] is composed of a series of rings or grids that act as an electrostatic mirror making more or less penetrate the ions in depth depending on of their kinetic energy to give it a null energy, and then reflect them in the opposite direction.

Ions with the same mass but with more energy result to be reflected and re-accelerate in subsequent times with the result of getting to the detector ions with equal m/z in the same moment. With this device the resolution can reach up to 20000 ppm.

All these peculiar innovations allowed the TOF to be more and more used in biological and proteomic research [72,73].

Principles of TOF

The treatment will concern the classical TOF in so-called linear mode (Figure 24).
The ions are produced and then accelerated by a potential V, and enter the flight tube, ideally with the same kinetic energy (we will then see how to overcome differences of this) Ek, coming separated according to the speed. So remember the definition of electric potential energy we can write:

$$E_{el=}zeV \qquad (2.2)$$

Where z is the number of charges (normally 1), and the charge of the electron and V the potential applied during acceleration. Keep in mind that at this juncture all electrical potential energy is transformed into kinetic energy Ek, which in classical mechanics is defined as:

$$E_k = \frac{mv^2}{2} \qquad (2.3)$$

Where m is the mass and v is the speed. Equalling 2.2 with 2.3 and rearranging obtains:

$$v = \sqrt{\frac{2zeV}{m}} \qquad (2.4)$$

After initial acceleration, the ions pass through the flight tube in time t that will have a distance L. Therefore, using the basic laws of motion you can write:

$$t = \frac{l}{v} \qquad (2.5)$$

Replacing the speed of the 2.4 with the 2.5 and rearranging you get:

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2eV}\right) \qquad (2.6)$$

The 2.6 shows how, considering the ratio between the brackets constant, m/z is calculated by measuring the square of time. Not only, but, considering invariable other factors, lighter masses will reach the detector faster.

The reason for the initial difficulty in establishing this instrument, was mainly due to low instrumental resolution. This phenomenon was linked to factors that generated a flight time distribution of ions having the same m/z.

Three factors generated the problem posed earlier:

• The time distribution;

• the spatial distribution;

• The distribution of kinetic energies;

The last of these points was solved by the realization of the mirrors ionics, building the reflectron.

Reflectron

The reflectron was first proposed by Mamyrin. He devised a retardant field, acting as an ionic mirror, reflecting the ions back, sending them back into the flight tube. The simplest reflectron consists of circular, equi-spaced electrodes.

The reflectron is placed in front of the ion source and, as a result, the detector is placed on the side where the ions exit from the ion mirrors. The reflectron corrects the dispersion of different Ek ions with the same m/z, such as shown in Figure 25.

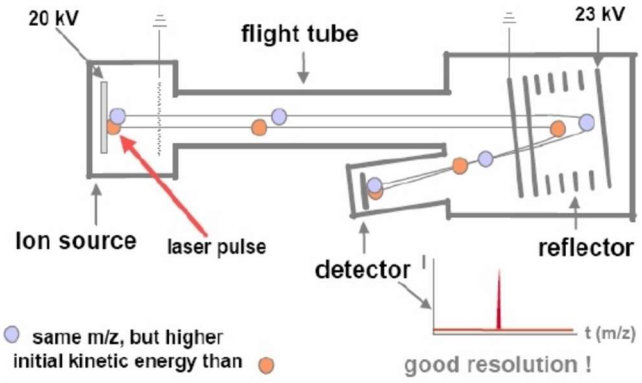


Figure 25. Diagram of a TOF equipped with a reflectron (RTOF), note the off-axis position of the detector and the improvement of the resolution.

Ions with higher kinetic energy will have higher speed and will penetrate deep into the reflectron. Slower ions, of course, will spend less time inside; but all ions with the same m/z will arrive at the same time to the detector. Another observation, which will not escape the careful eye is dictated by the fact that the reflectron allows to increase the length of the flight tube without increasing the size of the instrument. The total time (2.7) that an ion with data m/z takes a reach the detector is given by the sum of the time spent in the flight tube t plus the time in reflectron tr:

$$t_{TOT} = t + t_r$$

If we have a difference in kinetic energy we can express it with the ratio of the kinetic energy of the two ions with identical m/z, and therefore it can be demonstrated that

$$t_{TOT} = \frac{t}{a} + at_r \qquad (2.8)$$

As you can easily notice regardless of the value of a the journey time will be the same for identical values of m/z. [74]

2.12.7 Ionization for Matrix Assisted Laser Desorption (MALDI)

MALDI-TOF mass spectrometry is an analytical technique that allows the analysis of polar macromolecules such as peptides, proteins and DNA, in general of biopolymers in a theoretically unlimited mass range but that in practice ranges from 1 to 500 kDa. It is an extremely sensitive method, allowing very small sample quantities to be used (10^{-3} to 10^{-6} pmol); with an accuracy of 0.1 - 0.01 %. MALDI-TOF mass is considered one of the most powerful techniques for proteomic studies.

Laser desorption, the principle on which MALDI technology is currently based, was first used by Posthumus in 1978 [75] to produce intact peptides in the gas phase from solid samples. This technique of volatilization required close knowledge of the physical parameters of the peptides analysed (such as absorption wavelength, volatility, etc.) and did not allow the observation of intact peptides with PM > 1000 Da. These limitations have been overcome by the development of FAB (Fast Atom Bombardment) [76] and the PD techniques (Plasma Desorption) [77, 78]. The FAB method, introduced by Michael Barber, uses a liquid matrix to incorporate the analyte before its

ionization by irradiation with argon or xenon atoms [76]. The matrix concept was also present in the PD when nitrocellulose [79] and glutathione [80] were used to trap proteins below 20 kDa.

Matrix-assisted laser desorption/ionization (MALDI) was introduced as a 'soft ionization' strategy for the analysis of large biomolecules [45, 81] and subsequently adapted for a range of smaller analytes, such as peptides [82].

They combined laser technology with the use of a matrix to assist the desorption/ionization of analytes allowing the visualization of molecules with masses even greater than 100,000 Da. This technology underwent rapid development with the change in laser wavelength [82, 83], the discovery of new matrices [83–86] and the evolution of sample preparation methods [87-88]. The scheme of MALDI desorption is shown in figure 26.

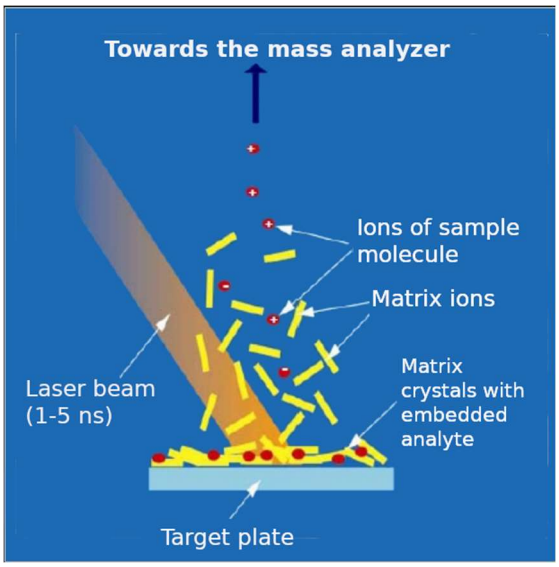


Figure 26. scheme of MALDI desorption

The matrices used for the MALDI technology are low molecular weight acids that allow the absorption of a UV/IR radiation and the transmission of the photons emitted by the laser to the analyte, allowing the passage to the gaseous state without damaging it.

The matrix co-crystallizes the sample avoiding the aggregation of analyte molecules and prevents the destruction of the sample by absorbing the laser power and transferring the energy needed for desorption to the sample. Desorption thus becomes independent of the intrinsic properties of the analyte, while the choice of the matrix becomes a crucial point both in function of the type of sample to be analysed and in function of the type of analysis conducted (linear mode or reflector). The classification of the matrices depends on their sublimation temperature: the increase in this value is directly proportional to the energy released. A large number of materials were tested as matrices

and only a few were found to be satisfactory. The main matrices now used routinely in proteomic analysis are organic molecules with low molecular weight that generally contain acid functional groups, able to easily yield a proton to the analyte. Furthermore, the matrix must not exhibit chemical reactivity to the sample, it must have high vacuum stability and high laser wavelength absorbance [89]. The matrix used for this study α-Cyano-4-hydroxycinnamic acid and a wavelength of 337 nm.

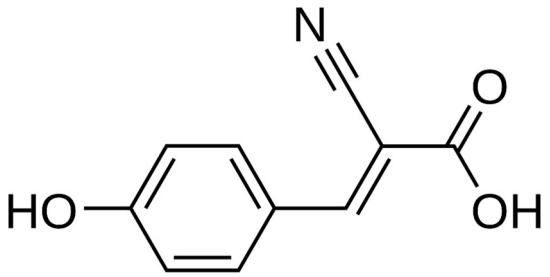


Figure 27 Structure of α-Cyano-4-hydroxycinnamic acid

To prepare the sample, usually, in the "dried-droplet" method, a dilute concentration of the sample is used against a more concentrated matrix. Everything is then laid (for a volume in the order of µl) on a metal plate (target), and left to crystallize.

Desorption/ionization is described on the basis of a rapid transition from solid to gaseous state, the energy required is provided by the laser, as observed in Figure 28 [90-96]. In the process of transition into the gaseous state, the molecules are ionized, but the principle by which this happens is still being studied [95]. The time in which energy is transferred is extremely small (on the order of ns), so it does not undergo thermal decomposition. The matrix performs basically three tasks:

• absorbs the energy of the laser protecting the sample from photo-dissociation;

• avoids the formation of aggregates;

• ensures the ionization of the analyte;

One of the focal points of MALDI spectrometry is the preparation of the sample, an ideal method should allow a high sensitivity, a wide range of mass, high accuracy, as well as optimal reproducibility. Another unique feature is the ability to automate the process. One thing to be kept in mind is that this ionization generates monocharged species $[M+H]^+$. There are many strengths of this process, in addition to the "soft" ionization , such as versatile adaptability to different analysable molecules, remarkable sensitivity, ability to cope with complex biological systems [79], and high

yields [97]. This gives a way of understanding why this methodology is considered among the most powerful in proteomic investigation.

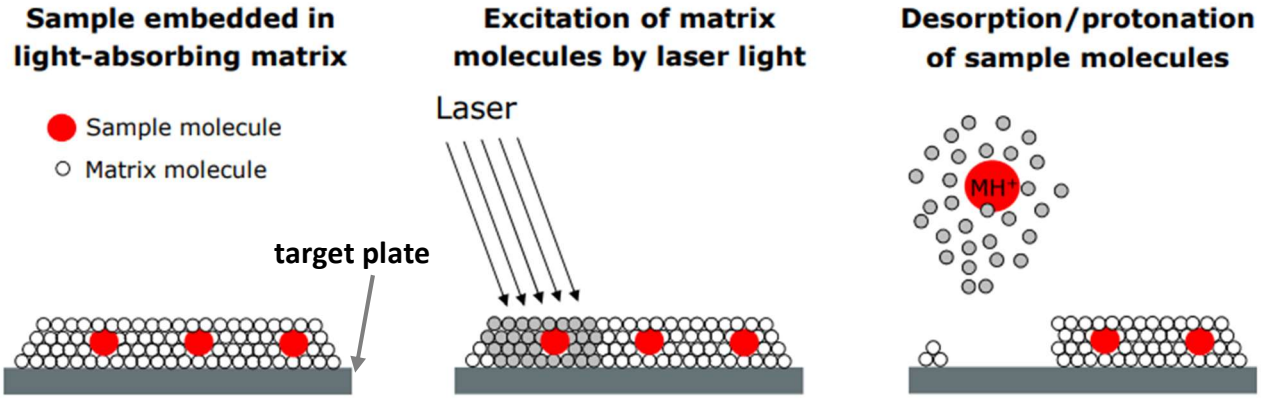


Figure 28. The matrix transfers the energy needed for ionization from the laser light to the sample molecules.

$$M \xrightarrow[\text{Matrix}]{\text{Laser energy}} [M+H]^+$$

2.12.7.1 MALDI/TOF analysis

The procedure of an analysis follows the following steps:

1. Mix sample and matrix solutions in suitable ratio.

2. Put a sub µl aliquot of this mixture on the target plate.

3. Let the mixture co-crystallize.

4. Insert the target plate in the MALDI mass spec (running under high vacuum).

5. Shoot with the laser on the MALDI preparation to generate ions (singly charged predominantly).

6. Collect, analyze and detect the resulting ions.

A simple illustration of the flow is given in figure 29.

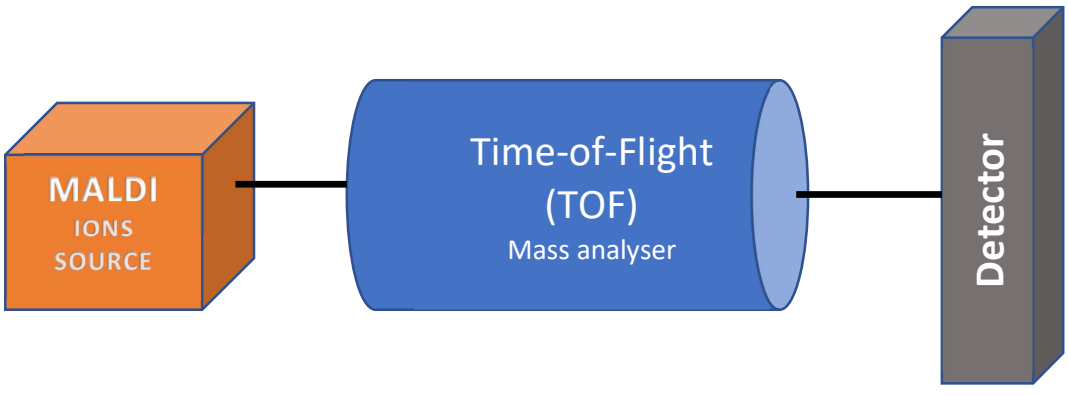


Figure 29. Generic diagram of a MALDI TOF instrument

The elute obtained by the extraction process with the magnetic beads is mixed in a ratio of 1:2 with the matrix solution (consisting of 6.2 mg of alpha-cyano-4-hydroxycinnamic acid in 1 ml of 36% methanol, 56% acetonitrile and 8% water). At this point 1 µl of this mixture was spotted in triplicate onto the MALDI target plate and left to dry under vacuum at room temperature for about 20 minutes. The samples eluted by the magnetic beads, after being mixed with an appropriate amount of matrix, were deposited, in triplicate, on the target plate of the MALDI source and allowed to dry at room temperature in a vacuum desiccator for about 20 minutes. For each spot, a volume of mixture (elute+matrix) equal to 1 µl was deposited on target plate (figure 30)..

Figure 30. Target frame

In order to optimize the analytical conditions, many parameters have been tested such as the type of matrix, the quantitative ratio between sample volume and matrix volume and the laser power. MALDI/TOF analysis was performed in positive mode on a mass spectrometer ultrafleXtreme™ (Bruker Daltonics Inc.).

Figure 31. UltrafleXtreme™ (Bruker Daltonics Inc.)

The instrument uses the high-vacuum (2.1×10^{-6} mBar) and it is equipped with a solid-state, so-called smartbeam-II™, laser allowing a very high data acquisition speed in both MS and MS/MS with a frequency range of 1-2000 Hz. It features a mass range of up to 50kDa in linear mode up to 80kDa in reflector mode. For peptide analysis the system can exceed 40,000 Full Width at Half Maximum (FWHM) resolution with corresponding measurement accuracy of <1.5 ppm.

Combining true 2 kHz speed in TOF mode and 1 kHz in TOF/TOF mode the instrument is able to significantly increase the dynamic range thanks to the 10bit 5GHz digitizer. Is equipped with a self-cleaning ion source based on IR laser (second laser) as shown in the figure 32.

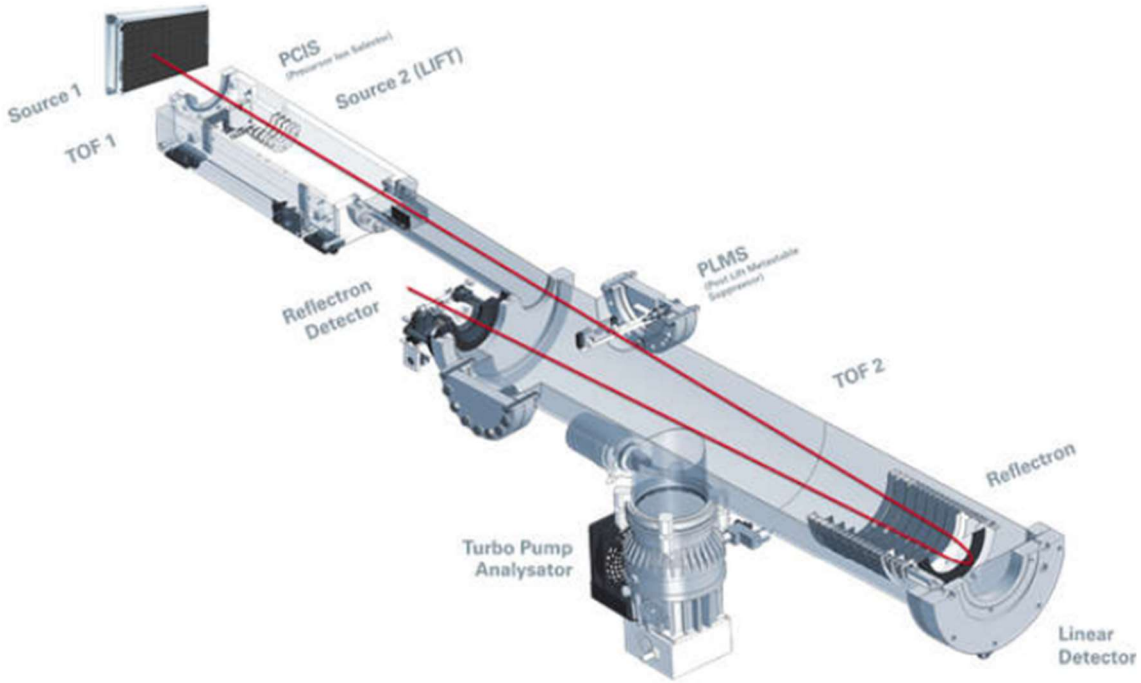


Figure 32. Diagram of TOF/TOF MS system

Data were acquired in the *m/z* range 600 to 3500 using these parameters: ion source 1 voltage 20.09 kV, ion source 2 voltage 17.81 kV, reflector 1 voltage 20.73 kV, reflector 2 voltage 11.18 kV and lens voltage 9.94 kV.

Five laser shots per raster spot were fired yielding 1000 laser shots per spectrum with a sample rate of 2000 Hz and from the sum of these data is generated a raw spectrum. All measurements were performed in positive ion and the acquisition laser power was set at 30% of maximum.

2.13 Data Analysis with a bioinformatic approach using Python and libraries ad hoc.

In addition to expanding the number of standard peptides to be included in the mixture of peptides to be analysed has been developed a bioinformatic approach writing a series of Python codes to improve the understanding and recognition of know sequences to be associated with m/z signals present in the TIC.

The development of the Pyhton codes was possible thanks to the conversion of the proprietary binary file format containing the raw data generated from instrument (in this workAgilent) to mzML standard format [61] by using MSconvert, a tool of the software suite ProteoWizard [98] and use of two open-source Python libraries, pyOpenMS [99] and pymzML [100] specific for processing mass spectrometry data, specifically for the proteomics analysis. From a computational point of view, a peptide sequence can be considered as a character string with decoding to a letter, this corresponds to one of the twenty amino acids. All Python codes were written on Spyder, an IDE (Integrated Development Environment) open-source distribution for scientific computing.

The idea of developing codes for the assignment of the m/z values of the spectrum peaks to m/z values of ions of synthetic peptides from standard mixtures. Using both the in house Data Analysis software equipping the ion-trap instrument and the python codes that I developed, I have performed computational research of the masses of standard peptides in the two levels of mass spectrum, this for an intrinsic property of a peptide that cannot undergo fragmentation or be among the three selected by the instrument to give rise to the spectrum of second level (MS/ MS in tandem). The developed codes process the signals coming from the different points of the analyte path inside the instrument, from the full scan to the MS/MS spectrum on the selected precursors, gradually improving the possibility of match between m/z value and peptide sequence.

The first code (Code I) developed takes into account the values of experimental m/z coming from the full scan chromatogram and associates them to the m/z values present in a peptide list of known

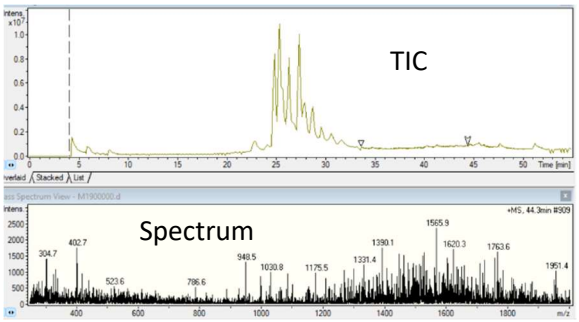
sequences obtained from MIX19 considering them also at different degrees of protonation (from mono- to tetra-charged).

The analysis of the spectra from standard peptides showed the presence of abundant values of m/z not attributable to whole masses more or less protonated. Hence the need to also research the masses of their possible fragments (CID), simulating in silico what probably happens in the solution or ionization chamber in case of loss of one or more amino acids at the end or beginning of the sequence. As already seen, not all ions contributing to TIC are selected by the instrument and subsequently fragmented in the ion trap in tandem (MS/MS, level 2 spectrum). In our case, it is even more important for the type of electronics of the instrument used, where only three ions are selected and fragmented in the quadrupole ion trap, giving rise to the spectrum of second level belonging to a single molecule. In fact, the identification of a peptide sequence by tandem mass spectrometry relies on the knowledge of products produced by collision-induced dissociation (CID) of peptide ions. Therefore, for very large mixtures of also known peptides or in a case of biological samples with a complex matrix most sequences can be present only in the full spectrum level 1 scan without undergoing fragmentation at level 2 as spectrum of MS/MS level 2.

It was also necessary to write auxiliary codes for the fragmentation of sequences with the aim of:

- creating fragments of entire sequences by removing one or two amino acids from the N-terminal site and C-terminal site (occurred, for example, in sequences of synthetic peptides present in standard solution samples).
- fragmenting amino acid sequences of proteins to obtain peptide sequences to create an extended database of string-fragments to be used as decoy for the recognition of unknown sequences in biological samples.

"Find masses"
For each peak its values of m/z and abundance in the level 1 spectrum are extracted



list of peaks

# index	m/z	abundance
1	250.4	119
2	251.3	189
3	254.4	145
4	255.4	112
……	…..	…..
600	1652.0	116
601	1673.1	102

match_find_masses_frag.py (Code I)

Calculate m/z values with different degree of protonation

$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$ $[M+4H]^{4+}$

Peptide accession	Peptide sequence
pept_2_11	NITHFAIVASL
pept_3_9	SLELGDSAL
pept_4_9	MSAPRKVRL
pept_5_9	AMAPIKVRL
pept_6_25	MIPDIPTDISDQIKKEKSLLVDFFL
……	……
pept_19	NHEAEDLFYQS
pept_20	DATDQLS

generate _frag.py

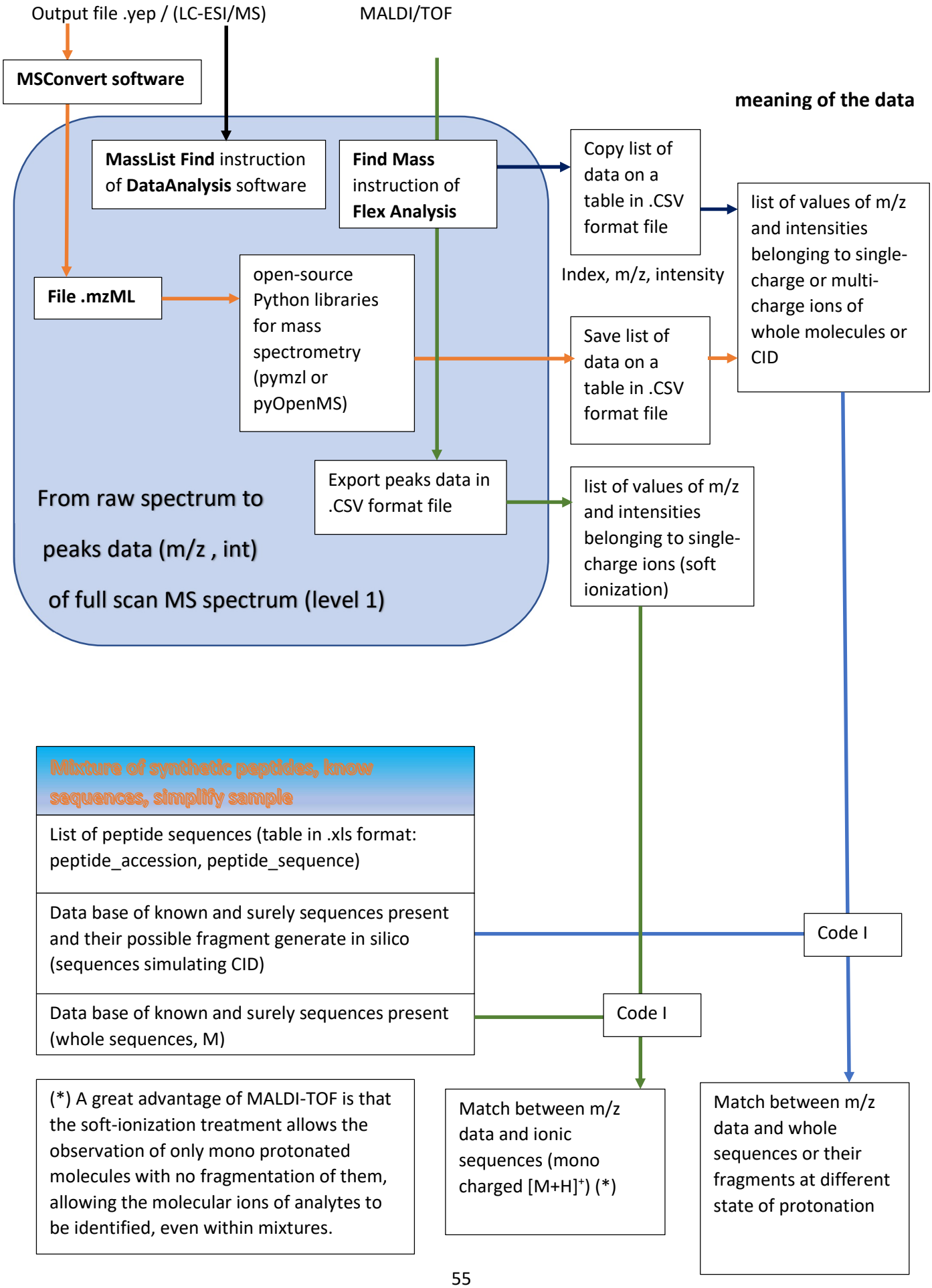
Peptide accession	Peptide Sequence [M]
pept_2_11	NITHFAIVASL
pept_2_11_i_1	ITHFAIVASL
pept_2_11_f_1	NITHFAIVAS
……..	….
pept_3_9	SLELGDSAL
pept_3_9_i_1	LELGDSAL
……..	……
pept_6_25	MIPDIPTDISDQIKKEKSLLVDFFL
pept_6_25_i_1	IPDIPTDISDQIKKEKSLLVDFFL
pept_6_25_f_1	MIPDIPTDISDQIKKEKSLLVDFF
…….	…….

Table of full sequences of synthetic peptides as present in mixture solution.

List of whole peptide sequences and their possible fragments, created by eliminating the first and/or the second amino acid N-terminal or C-terminal, simulating possible CID structures.

Label of standard peptide, number of amino acids eliminated from initial or final residue (referred to the whole sequence).

The diagram presented in this page and the one presented below show the flow of operations of the first algorithm applied to the full scan spectral data from HPLC-MS ESI or MALDI/TOF and using as information in input the list of peptide sequences presented in the analysed samples.


Output file .yep / (LC-ESI/MS)
MALDI/TOF
MSConvert software
meaning of the data
MassList Find instruction of DataAnalysis software
Find Mass instruction of Flex Analysis
Copy list of data on a table in .CSV format file
list of values of m/z and intensities belonging to single-charge or multi-charge ions of whole molecules or CID
Index, m/z, intensity
File .mzML
open-source Python libraries for mass spectrometry (pymzl or pyOpenMS)
Save list of data on a table in .CSV format file
Export peaks data in .CSV format file
list of values of m/z and intensities belonging to single-charge ions (soft ionization)
From raw spectrum to peaks data (m/z , int) of full scan MS spectrum (level 1)
Mixture of synthetic peptides, know sequences, simplify sample
List of peptide sequences (table in .xls format: peptide_accession, peptide_sequence)
Data base of known and surely sequences present and their possible fragment generate in silico (sequences simulating CID)
Data base of known and surely sequences present (whole sequences, M)
Code I
Code I
(*) A great advantage of MALDI-TOF is that the soft-ionization treatment allows the observation of only mono protonated molecules with no fragmentation of them, allowing the molecular ions of analytes to be identified, even within mixtures.
Match between m/z data and ionic sequences (mono charged [M+H]+) (*)
Match between m/z data and whole sequences or their fragments at different state of protonation

In what follows I present an example of code output applied to the HPLC-ESI MS spectrum data of the mixture of synthetic peptides with known sequences (table 4). In the generated output it is interesting to point out that several values of m/z have been assigned to the CID fragments with one or more charge.

peptide_sequence [M]	peptide_accession	Peak (m/z)	n_charge
SLELGDSAL	pept_3_9	302.3	$[M+3H]^{3+}$
SLELGDSAL	pept_3_9	904.5	$[M+H]^{+}$
MSAPRKVRL	pept_4_9	529.5	$[M+2H]^{2+}$
AMAPIKVRL	pept_5_9	333.6	$[M+3H]^{3+}$
MIPDIPTDISDQIKKEKSLLVDFFL	pept_6_25	1453.2	$[M+2H]^{2+}$
VMAPRTLIL	pept_7_9	338.6	$[M+3H]^{3+}$
……	…….	……..	..
VIPDIPKDISQQIHKEKVLMVELFM	pept_10_25	984.4	$[M+3H]^{3+}$
VIPDIPKDISQQIHKEKVLMVELFM	pept_10_25	1476	$[M+2H]^{2+}$
VMAPRILIL	pept_11_9	513.5	$[M+2H]^{2+}$
VMAPRILIL	pept_11_9	1025.8	$[M+H]^{+}$
…….	……	……	..
IPDIPTDISDQIKKEKSLLVDFFL	pept_6_25_i_1	694.3	$[M+4H]^{4+}$
IPDIPTDISDQIKKEKSLLVDFFL	pept_6_25_i_1	1387.7	$[M+2H]^{2+}$
MIPDIPTDISDQIKKEKSLLVDF	pept_6_25_f_2	882.6	$[M+3H]^{3+}$
MIPDIPTDISDQIKKEKSLLVDF	pept_6_25_f_2	1323.2	$[M+2H]^{2+}$
MAPRTLIL	pept_7_9_i_1	305.5	$[M+3H]^{3+}$
MAPRTLIL	pept_7_9_i_1	914.5	$[M+H]^{+}$
…….	……	……	…
IPDIPKDISQQIHKEKVLMVELFM	pept_10_25_i_1	713.5	$[M+4H]^{4+}$
VIPDIPKDISQQIHKEKVLMVELF	pept_10_25_f_1	705.6	$[M+4H]^{4+}$
PDIPKDISQQIHKEKVLMVELFM	pept_10_25_i_2	685.5	$[M+4H]^{4+}$
MAPRILIL	pept_11_9_i_1	309.5	$[M+3H]^{3+}$
……..	……	…….	..

Table 4. Sorting the table by values of m/z it is possible to see that there are unique values. Some of these values are not found in the precursor ion list because they are not fragmented in the ion trap.

The multicharge assignment makes sense, from the chemical point of view, if in the sequence there are several protonable residues such as lysine (table 5).

peptide_sequence [M]	peptide_accession	Peak (m/z)	n_charge
PDIPKDISQQIHKEKVLMVELFM	pept_10_25_i_2	685.5	$[M+4H]^{4+}$
IPDIPTDISDQIKKEKSLLVDFFL	pept_6_25_i_1	694.3	$[M+4H]^{4+}$
VIPDIPKDISQQIHKEKVLMVELF	pept_10_25_f_1	705.6	$[M+4H]^{4+}$
IPDIPKDISQQIHKEKVLMVELFM	pept_10_25_i_1	713.5	$[M+4H]^{4+}$
VIPDIPKDISQQIHKEKVLMVELFM	pept_10_25	984.4	$[M+3H]^{3+}$
MIPDIPTDISDQIKKEKSLLVDF	pept_6_25_f_2	1323.2	$[M+2H]^{2+}$
IPDIPTDISDQIKKEKSLLVDFFL	pept_6_25_i_1	1387.7	$[M+2H]^{2+}$
MIPDIPTDISDQIKKEKSLLVDFFL	pept_6_25	1453.2	$[M+2H]^{2+}$
VIPDIPKDISQQIHKEKVLMVELFM	pept_10_25	1476	$[M+2H]^{2+}$

Table 5. m/z values of peaks belonging to pept_6 and pept_10 used in the MIX19.

If the mass spectrometer has enough resolution in the spectrum, one can observe the isotopic peaks of the main peak and confirm a multi-charge ion.

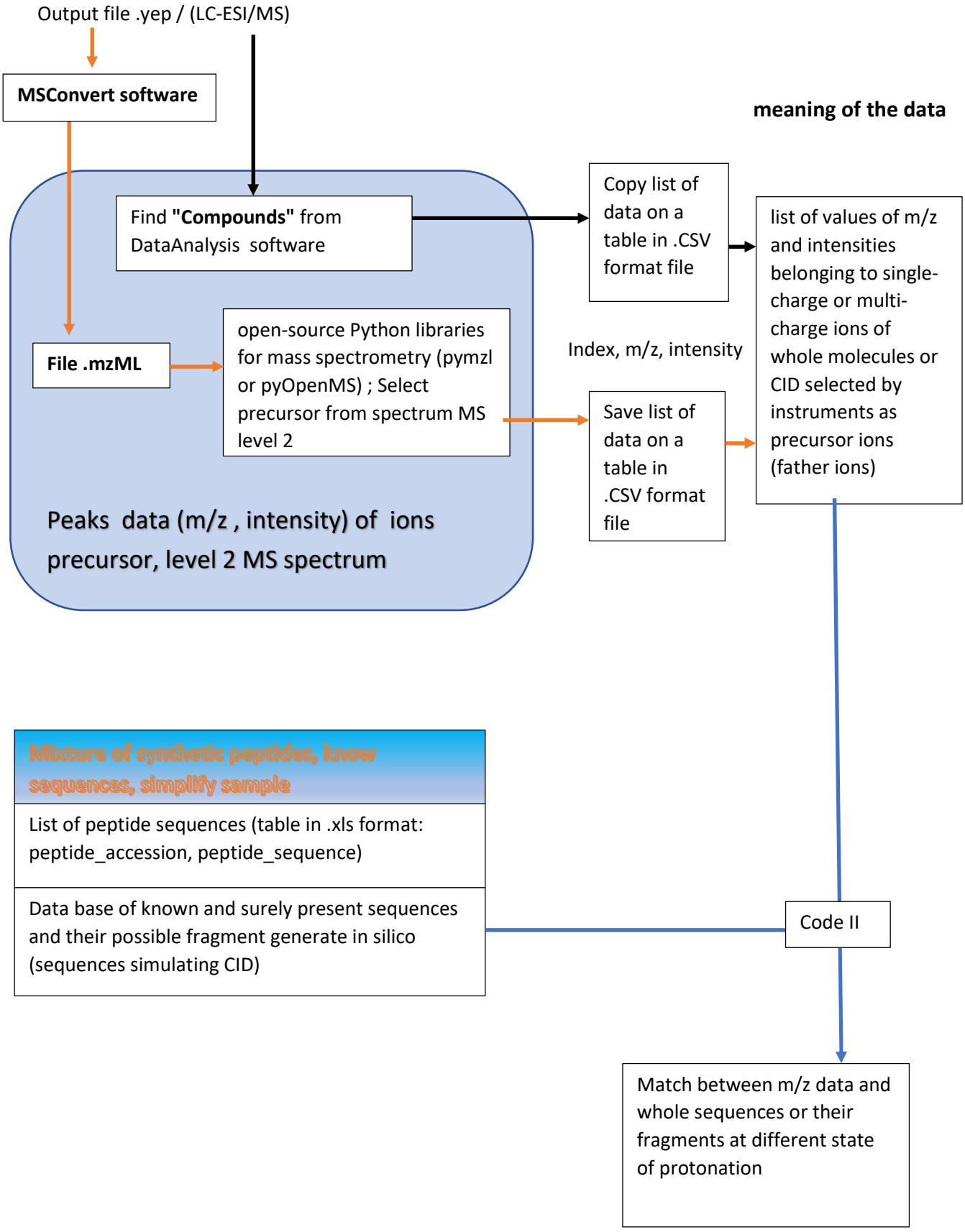


Diagram of second code (Code II), similar to the previous one, considers only the ions that have been selected as precursors and consequently fragmented in a second step giving rise to a level 2 spectrum.

Obviously, the computational approach of the algorithm in code I for the assignment of the ionized sequences to the m/z values in the MS level 1 spectra to search for sequences in the biological samples becomes possible only using an appropriate database constructed taking account of peptide sequences known for tissue, from time to time examined, as was done in this work for human plasma (see below). Nevertheless, to make the search for peptide correspondences more robust and reliable, it was decided to develop an additional code that also takes into account the MS/MS tandem spectra for each selected precursor ion (Code III).

Code III:

For the next algorithm, designed to search for unknown sequences in biological samples, it was necessary to take into consideration the spectral data obtained by MS/MS tandem techniques, in particular m/z data and abundance of the precursor ion and its fragments as showed in level 2 spectra. In fact, the identification of a peptide sequence by tandem mass spectrometry is based on the information obtained by collision-induced dissociation (CID) of peptide ions.

For writing this code two open-source Python libraries for mass spectrometry were used, specifically for the analysis of proteomics and metabolomics data on mzML data file format. In particular the PymzML library has been implemented to use certain functions that make it possible to manipulate the information in the mzML file concerning data and metadata of the spectrum acquired by the spectrometer such as values of the spectral indices and, for a peak, the value of m/z and its abundance. The functions of the pyOpenMS library was used on the sequences present in the data base of peptide sequence to produce in silico the ions fragments a, b and y or losses of neutral mass (e.g. the water molecule) and with their values m/z it was possible to simulate a theoretical spectrum.

Each spectrum has an index code which has a consecutive numerical value comprising both level 1 and level 2. The m/z value of each precursor present in level MS 1, considered as single charge and double charge, is compared with the in silico calculated mass for a series of possible peptides and associated with a match if the experimental m/z value is within a range of ±0.3 Da. This is the first filter on sequences list stored in the local database. If a match exists, the scan procedure on peptide list is suspended and a portion of higher peaks of the selected MS/MS spectrum are compared with a spectrum generated in silico for the selected peptide as decoy following the most common fragmentation CID rules that give rise above all to b ions (N-terminal part from the peptide bond break) and y-side ions (C-terminal part from the peptide bond break). In this phase, each

experimental m/z value of the level 2 spectrum, above a certain threshold of intensity, representing fragments of a single precursor ion is compared with the masses of fragments generated in silico for the peptide selected by the list during the first phase of data analysis. The m/z ratios are considered to be corresponding, and therefore the experimental data is accepted, if compared values are within a range of ±0.18 Da.

Finally, a sum of match scores above a certain percentage, calculated as scores=selected experimental peaks/ predicted theoretical peaksx100 is assigned. The output file is a table in CSV format that contains several data including index of experimental peak, its m/z value, string of the peptide selected, some its physical chemical properties and other properties according to the type of amino acids that constitute it such as the isoelectric point, the aliphatic index of the molecule, the percentage of hydrophilic, hydrophobic amino acids, the portion of aromatic amino acids or with acidic and basic features (see section “How the data were analysed”).

The output results of the code III on mixtures of standard peptides for the recognition of sequences and also for a comparison of the data obtained with the use of different magnetic beads will be examined later in the text. An example of output obtained from serum of an healthy donor is reported below.

index	prec	peptide_sequence	peptide_accession	count	peaks_%
561	733.3313	DSGEGDFLAEGGGVR	fibr_alpha1_a_15_21	25	62.5
548	733.3313	DSGEGDFLAEGGGVR	fibr_alpha1_a_15_21	24	60
556	733.3313	DSGEGDFLAEGGGVR	fibr_alpha1_a_15_21	24	60
520	539.2698	GDFLAEGGGVR	fibr_alpha_11_25	18	45
525	510.7591	DFLAEGGGVR	fibr_alpha_10_26	17	42.5
541	510.7591	DFLAEGGGVR	fibr_alpha_10_26	17	42.5
525	510.7591	FIDAAQEAR	PAp02405433	13	32.5
541	510.7591	FIDAAQEAR	PAp02405433	13	32.5
660	521.7744	VSGPDLDLNL	PAp04409257	13	32.5
529	510.7591	FIDAAQEAR	PAp02405433	12	30
561	733.3985	LESLSYQLSGLQK	PAp00430084	12	30
562	500.3135	TATQPLLKK	PAp03103048	12	30
616	507.3124	VMAPRTLIL	pept_std_7	12	30
520	539.7655	ETMAAAWTVV	interf_10_785	11	27.5
520	539.2603	CIDLEVEEK	PAp02392383	11	27.5
528	603.7911	EGDFLAEGGGVR	PAp00156094	11	27.5
532	603.7911	EGDFLAEGGGVR	PAp00156094	11	27.5
536	510.7591	FIDAAQEAR	PAp02405433	11	27.5
544	733.3641	AVTDEEPFLIFAN	PAp04474030	11	27.5

Table 6. The index of MS level 2 (first column), the m/z of precursor (second column), the peptide sequence (third column), the code of accession of sequence in the table (fourth column), the m/z values with match

(fifth column) and peaks % (sixth column) are reported. (”PAp…” is a sequence within data base from PeptideAtlas) other that was created from fragmentation of proteins, pept_std_7 is the sequence of a peptide used as internal standard.

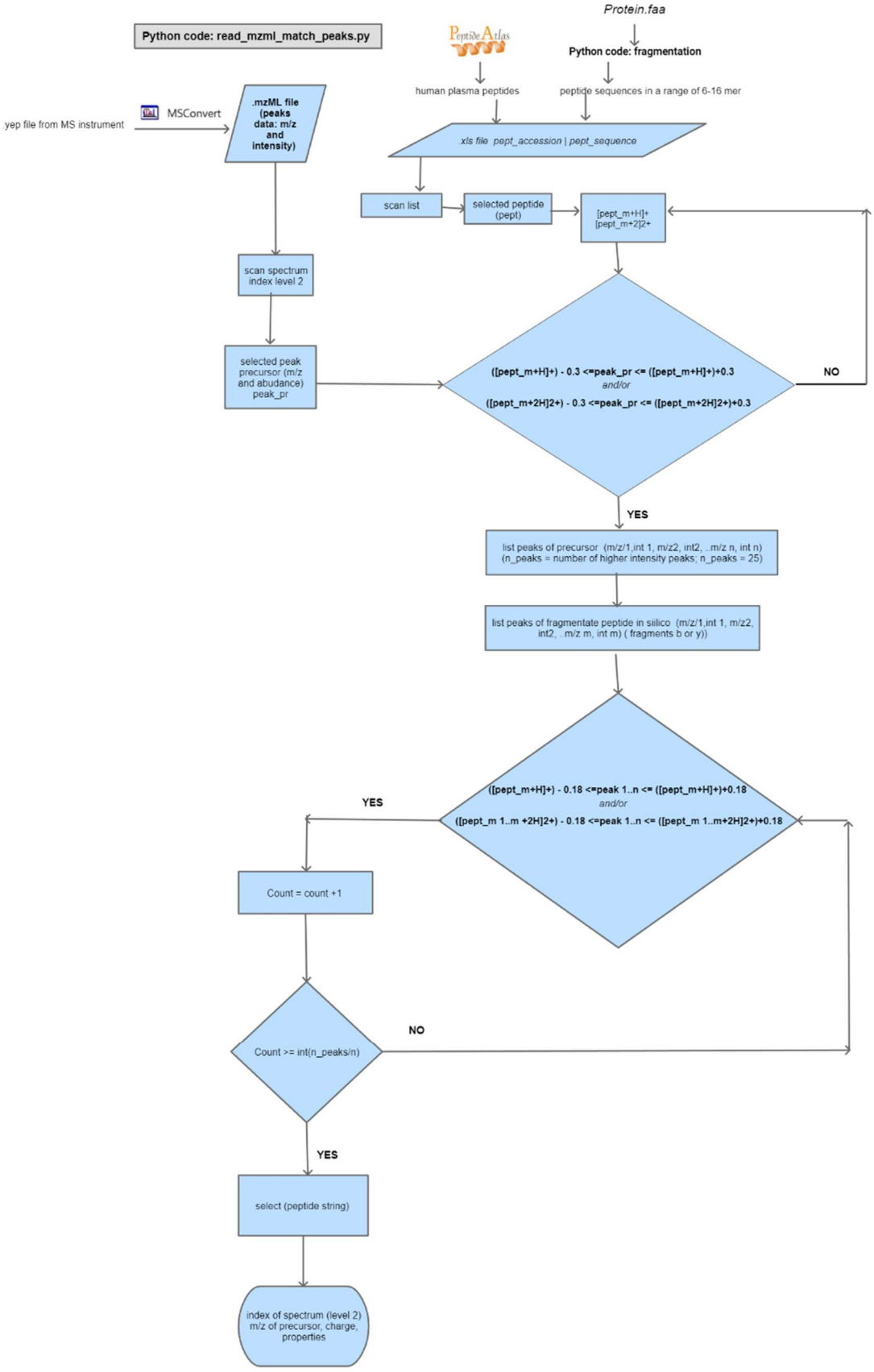


Figure 33. Flow chart of Code III used to determine the detail of the precursor sequence.

2.13.1 Local data base of peptide sequences

In addition to the mzML file, the peptide list is the second input file that constitutes the local data base of possible peptide sequences. Initially this local database was obtained from the PeptideAtlas site (http://www.peptideatlas.org/) [101], by downloading data from over 200,000 distinct peptides observed and validated [102] in human plasma. The PeptideAtlas project comprises a publicly accessible database set of peptides identified in many tandem mass spectrometry proteomics studies. Also, with another code I wrote, countless peptide sequences have been added to this data. These sequences were obtained by fragmenting in silico sequences downloaded in FASTA format [103] from the Uniprot site [104]. The FASTA file of proteins considered and downloaded were the most abundant such as albumin/globulins, and those most likely to be traced in serum or plasma such as coagulation factors, complement proteins, interleukin, interferons, peptide hormones, etc. The overall file you get is a table in CVS format that contains for each row a first column with a peptide code and a second with a peptide sequence (Table 7).

It is important to underline that during the data processing amino acid sequences over 20-mer were not considered and that analysed peptides are not fragments obtained by enzymatic digestion. Once the optimal computational parameters for the recognized of the known peptides of the standard mixture were found, the assignment algorithm was improved and applied to real samples previously depleted of the high molecular weight fraction, using, in the meantime, to find possible correspondences, a local database of peptides already found in the plasma from previous studies.

Below, the calculation of the fragments generated for a protein consisting of n AA is reported

L = length of protein downloaded from Uniprot site (number of amino acids that it is composed of)

*l*1 = length of the smallest fragment

*l*2 = length of the largest fragment

$Nseq\_{}_n$ = numbers of sequences with length *n*

NT = Numbers of generate fragment for protein

$Nseq\_{}_{l2}$ = (L-*l*2)+1

$Nseq\_{}_{l2-1}$ = (L-*l*2-1)+1

$Nseq\_{}_{l2-2}$ = (L-*l*2-2)+1

- - - - - - - -

$Nseq\_{}_{l2-n}$ = (L-*l*2-n)+1 (*l*1 =*l*2- n)

$$NT = [L \cdot (l2 - l1)] - \sum_{i=0}^{i=l2-l1} l\,2 - i + (l2 - l1) + 1$$

e.g. albumin (P02768) 609 AA

l1=6

l2=20

NT = [609 * (20-6)] –(21+20+19+….6) + (20-6)+1 = 8914 fragments of a length between 6 and 20

Extract of peptides data filtered from PeptideAtlas

peptide_accession	peptide_sequence
PAp00000001	AAHEEICTTNEGVMYR
PAp00000004	ALPGEQQPLHALTR
PAp00000005	AQETSGEEISK
PAp00000006	CDPHEATCYDDGK
PAp00000007	CFQTENPLECQDKGEEELQK
PAp00000008	CHAANPNGR
PAp00000009	CHAGHLNGVYYQGGTYSK
PAp00000010	CHEGGQSYK
PAp00000011	CLPDRETAASLLQAGYK
PAp00000012	CLQSGTLFR
PAp00000013	CNDQDTR
PAp00000014	DFFLANASR
……	……..

Extract of peptides data from fragmentation of selected proteins from FASTA format

peptide_accession	peptide_sequence
A5PL27HUMA_7_1	MKILILG
A5PL27HUMA_7_2	KILILGI
A5PL27HUMA_7_3	ILILGIF
A5PL27HUMA_7_4	LILGIFL
A5PL27HUMA_7_5	ILGIFLF
A5PL27HUMA_7_6	LGIFLFL
A5PL27HUMA_7_7	GIFLFLC
A5PL27HUMA_7_8	IFLFLCS
A5PL27HUMA_7_9	FLFLCST
A5PL27HUMA_7_10	LFLCSTP
A5PL27HUMA_7_11	FLCSTPA
A5PL27HUMA_7_12	LCSTPAW
……	……

Table 7. The list of peptides that make up the local database consists of a sample table in .CVS file format that contains for each row the first column with an accession peptide code and the second one with a peptide sequence. The peptide accession code in the table can be traced back to the sequence identification code, as reported in the header of the FASTA file of the fragmented protein. In the right pane, the peptide accession code numbers show the length of the fragment and the number from which the cut occurs, relative to the initial group of the amino terminal.

In summary, the advantages of exploring a stand-alone data set in your home are that you don't need a link to an external protein database, limiting your search to a set of proteins of interest in a range of residue number.

3. Results and discussions

3.1 First experimental evaluations on serum sample with C18 magnetic silica beads

After a brief evaluation of the standard peptide solution, which was subsequently resumed, the first experimental phase of the study on true biological samples consisted of active extraction of peptides from serum from healthy voluntary donors. This was done with the aim to develop and optimize a good reproducible extraction method.

The peptide fraction of serum was first extracted and purified using the ability of magnetic beads to bind peptides through hydrophobic interaction with chromatographic C18 (RPC 18) solid phase[2]. This first trial gave inferior quality results due to sample complexity. In particular the results were partially invalidated by the interference with high molecular weight proteins (mainly albumin) that linked at the beads in a non-specific way. Therefore, using the dynabeads tool to extract the peptide fraction, a problem of contamination of the chromatographic column presumably due to the presence of proteins with high molecular weight occurred. The following LC-MS analysis showed a total ion chromatogram (TIC) with a wide peak centred a 32 min of retention time, subsequently attributed to albumin (Mr 66000), that covers, in part, the signals of peptides of interest, as illustrated in the following figure.

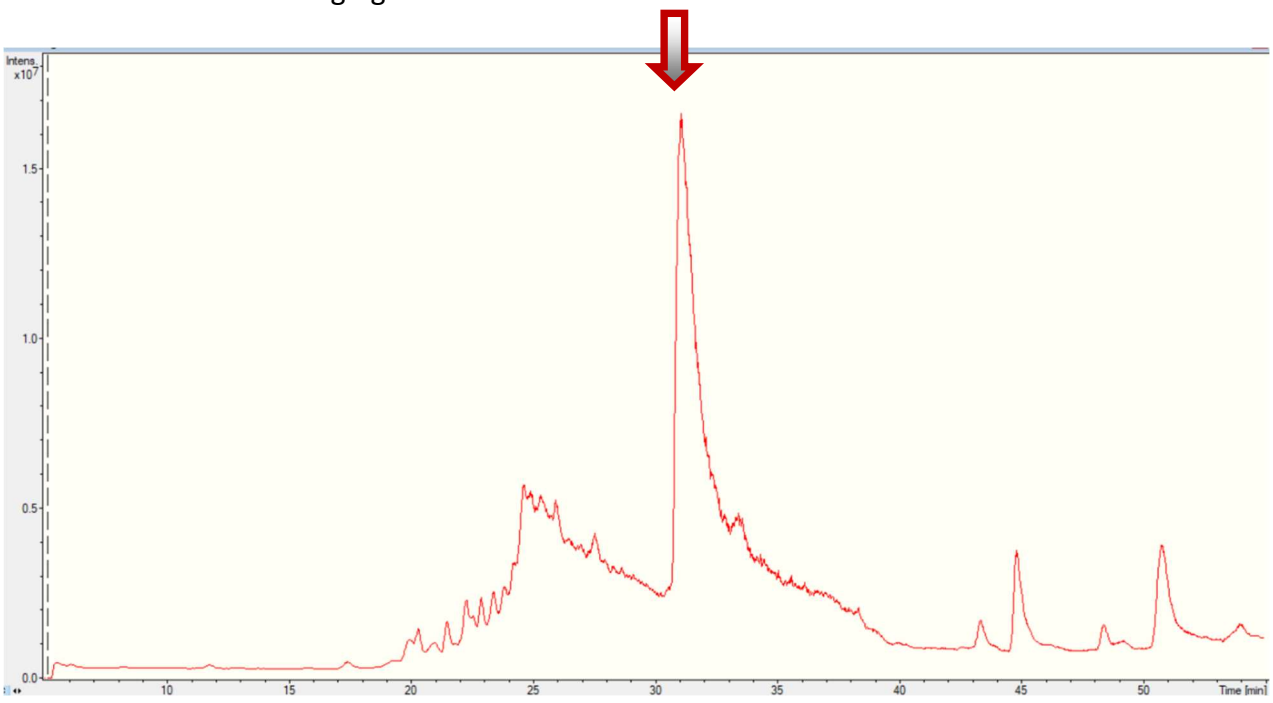


Figure 34. Total ion chromatogram (TIC) of a LC-MS analysis of serum with the presence of wide peak centred a 32 min.

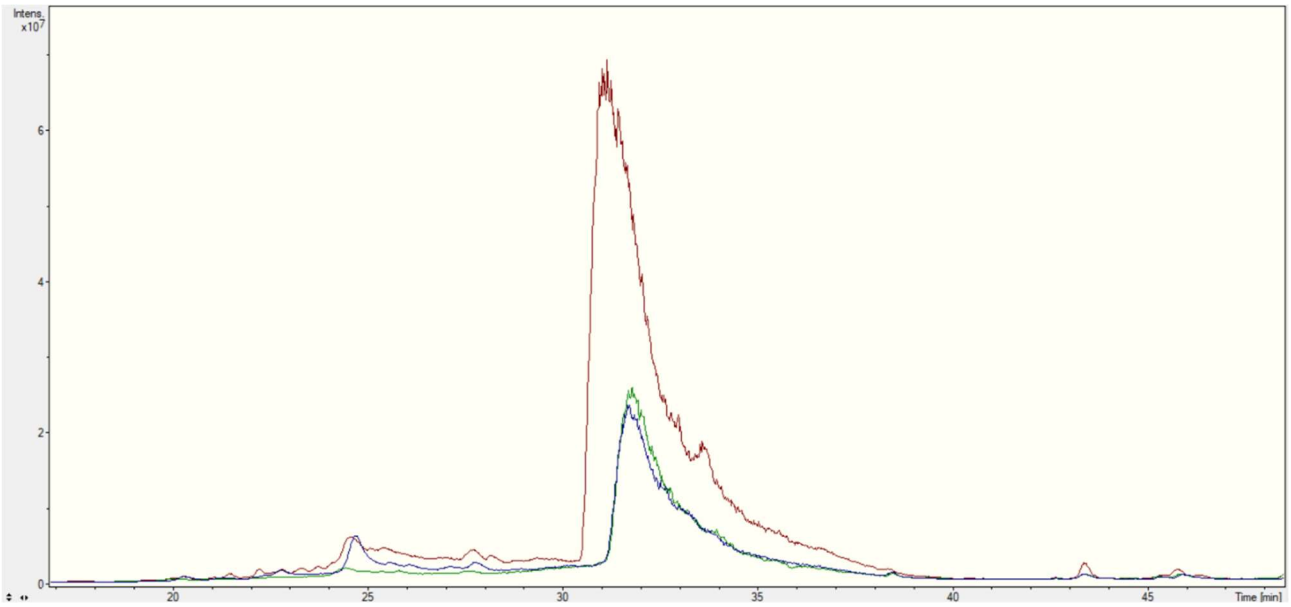


Figure 35. Total ion chromatogram (TIC) of a LC-MS analysis of different sample of serum

This shouldn't happen because this kind of beads should bind only peptides and proteins with a molecular weight lower than 12000 Da. A possible explanation for the presence of this contaminant could be the formation of beads aggregates that bind albumin on their surface; In addition, the albumin is retained inside the HPLC C18 column which also compromises the next analysis (figure 35).

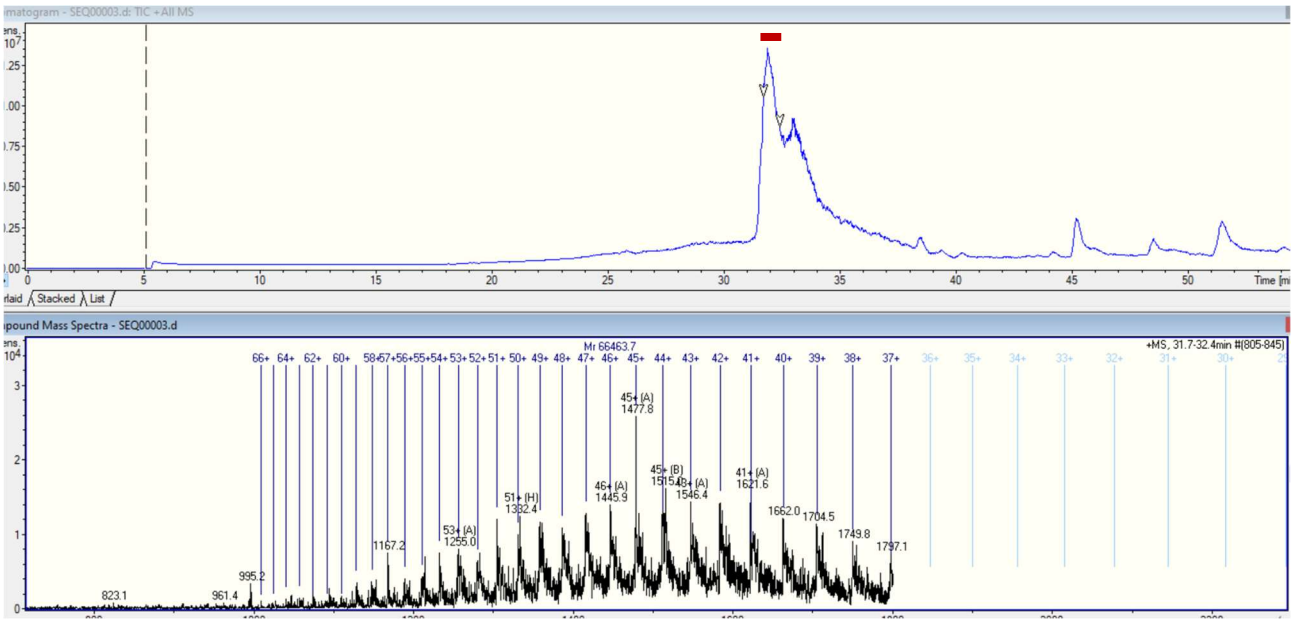


Figure 36. Identification of albumine (MW~66000) through deconvolution (- represents the range of peaks extracted for convolution ΔRT 31.7 – 32.4)

3.2 Research of other pre-treatment tool and experiments with standard peptides mixtures

To reduce the presence of albumin we needed to define a good reproducible method useful to minimize the interferences. Different tools were tried, in particular an albumin depletion kit followed by a desalting tool (zip tips C18). The method has been developed by analysing serum from healthy donors and/or using a solution of standard peptides synthesised in our laboratory. We tried to use different amounts of beads and resin suspension (slurry) for depletion albumin in ratio with different volumes of serum. The sample/beads ratio was also changed, with the aim of having a minimal consumption of materials with a significant presence of signals. In order to obtain useful information on the efficacy of the peptides extraction tools employed, a very simplified sample was used, consisting exclusively of a standard solution of 13 synthetic peptides (Mix 13), also in order to measure the recovery percentage for each peptide (figure 37).

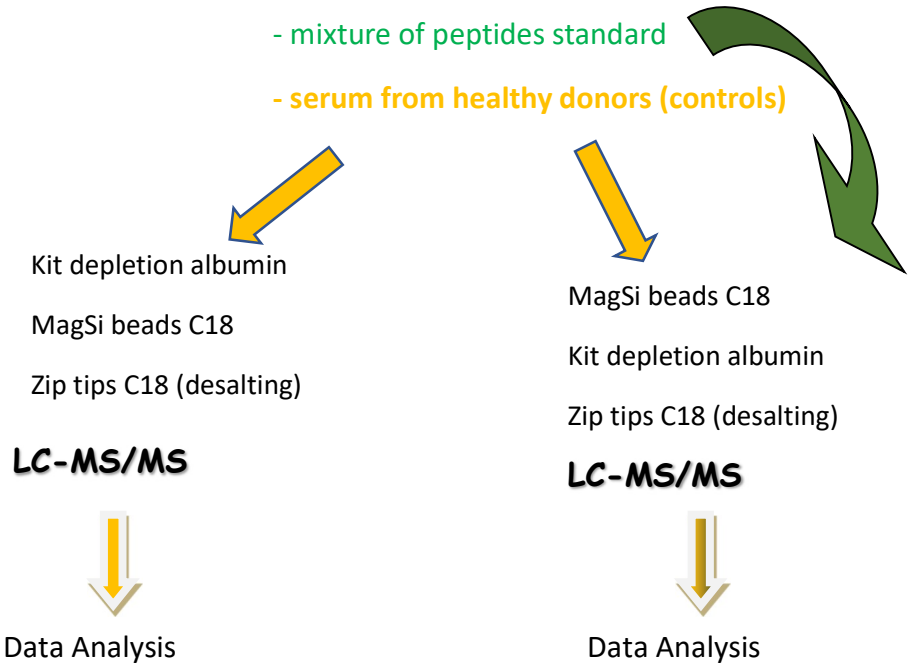


Figure 37. Comparison between samples of serum/mix of standard peptides in different extraction conditions

The Mix 13 was analysed under different experimental conditions, and the chromatographic/spectrometric data obtained were compared.

Three different approaches were used:

- The peptide mixture was analysed by HPLC-MS as such, without the use of any extraction tool (figure 38).
- The peptide mixture was analysed by HPLC-MS after having undergone only extraction with the dynabeads C18 extraction tool (figure 39).
- The peptide mixture was analysed by HPLC-MS after extraction with the magnetic C18 beads followed by albumin depletion kit tool (rPkab) and zip tips C18, as desalting treatment (figure 40).

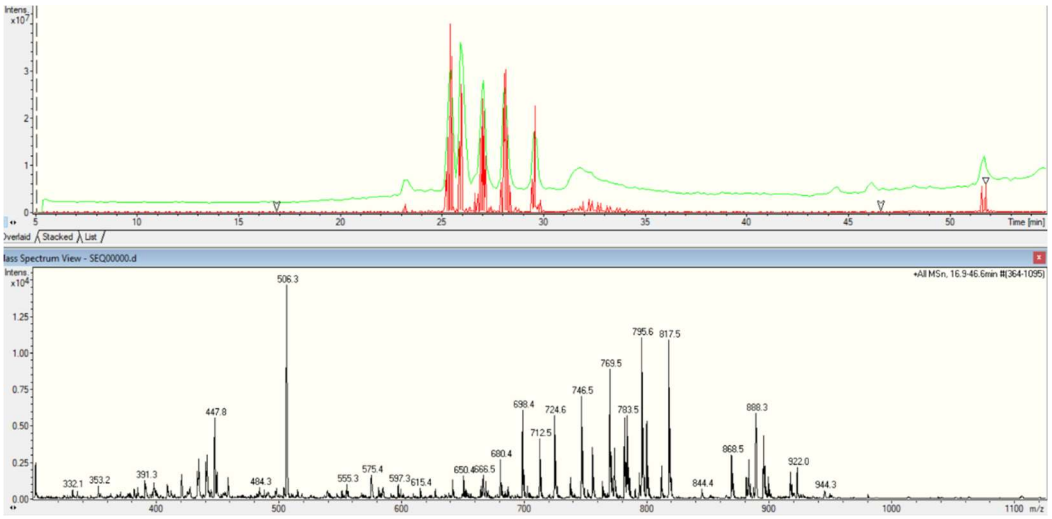


Figure 38. Chromatogram and Mass Spectrum (average abundance of peaks between r.t. 16.9 – 46.9 min) of Mix 13 as such, without the use of any extraction

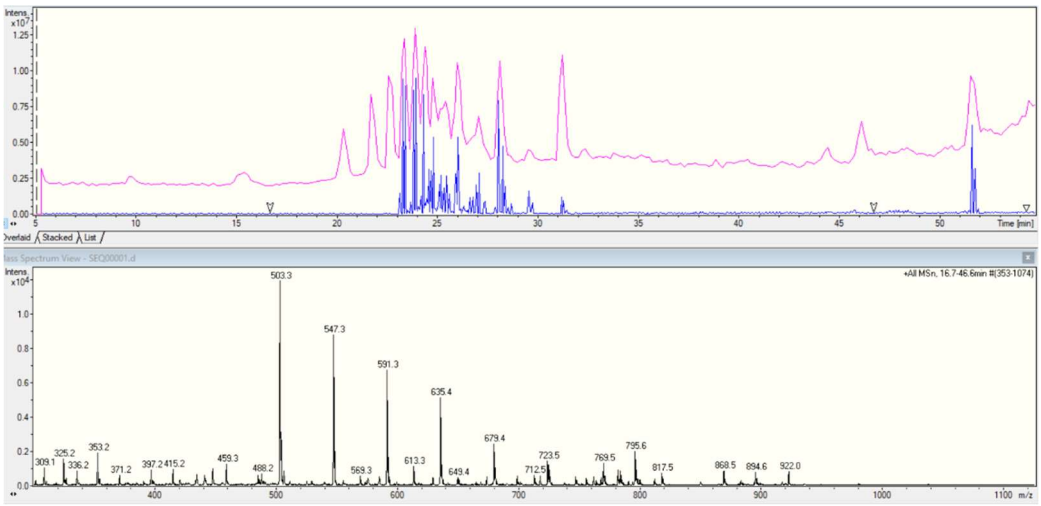


Figure 39. Chromatogram and Mass Spectrum (average abundance of peaks between r.t. 16.9 – 46.9 min) of Mix 13 after extraction with the dynabeads

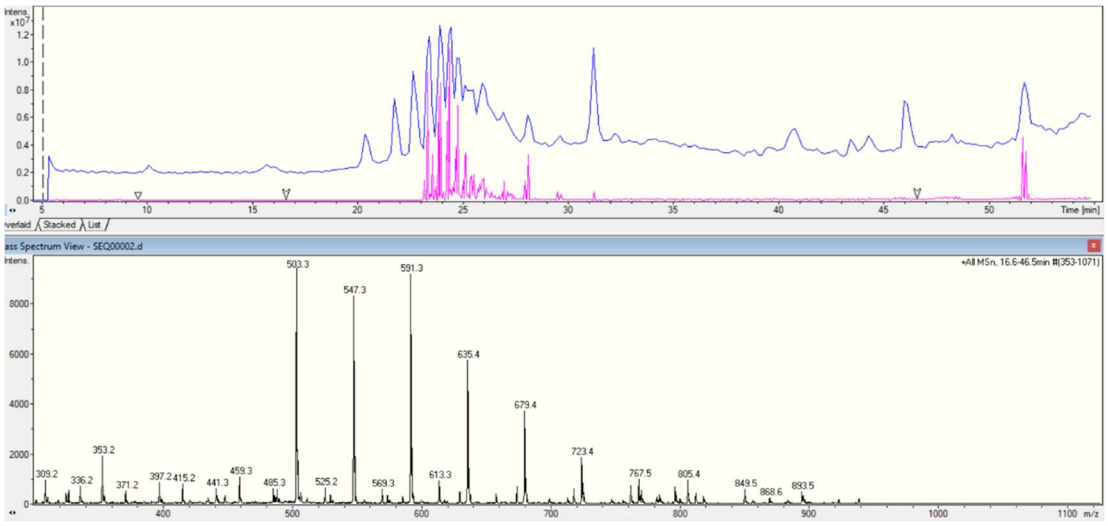


Figure 40. Chromatogram and Mass Spectrum (average abundance of peaks, between r.t. 16.9 – 46.9 min) of Mix 13 after after extraction with dynabeads followed by albumin depletion kit tool and zip tips C18, as desalting treatment.

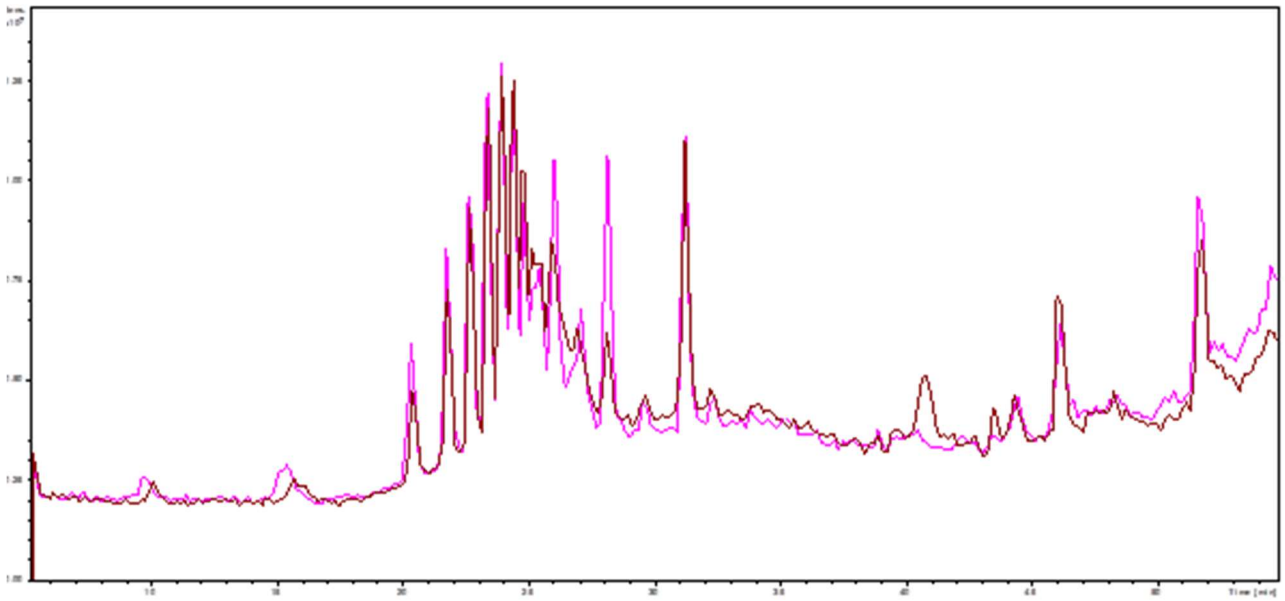


Figure 41.

Figure 41, shows the overlapping of the TIC obtained analysing Mix 13 after treatment with only dynabeads extraction tool with the TIC obtained analysing Mix 13 after treatment with dynabeads and albumin depletion kit tool and zip tips C18.

There is a low qualitative variation between the two TIC obtained using the two different approaches; this means that the use of the albumin depletion kit does not affect the subsequent extraction phase with beads.

There is a strong variation between the TIC coming from the analysis of Mix 13 as such if compared with the other two conditions, as can be seen in the figure 42 where the three chromatograms have been overlapped.

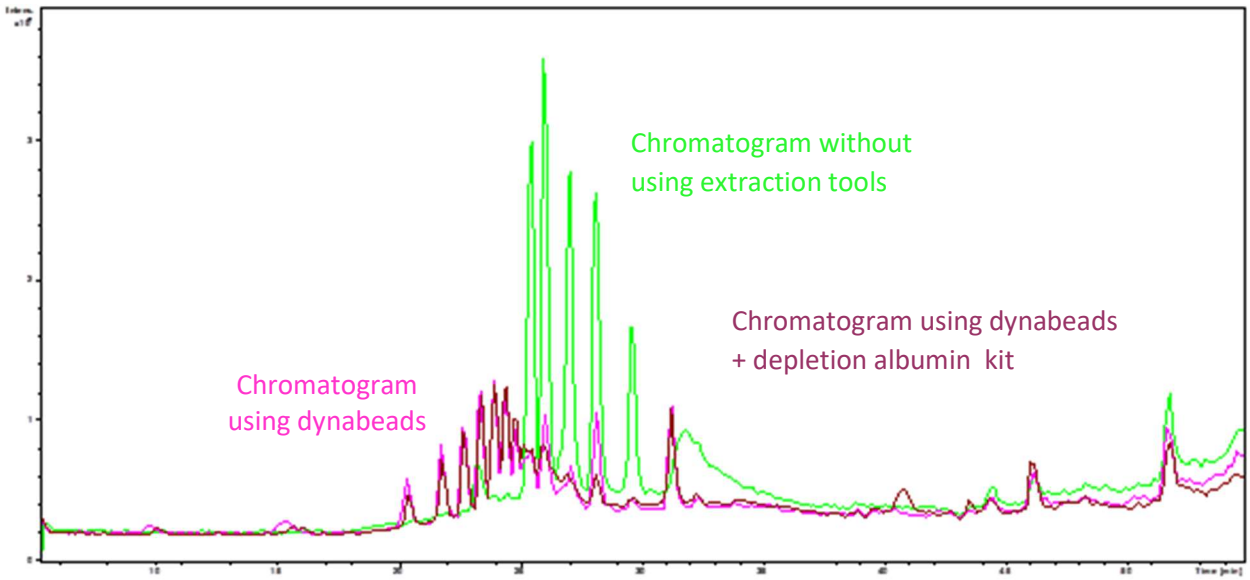


Figure 43. overlapping of the three chromatograms

Analysing the MS spectra from a quantitative point of view (relative measure of peak areas) resulted in a problem of low recovery of peptides.

As an example, Figure 44 shows the peak areas of the extracted ion chromatogram (EIC) of the standard peptide VMTPRTLVL (RT= 25.4 m/z 515.3 $[M+2H]^{2+}$) under the three different experimental conditions mentioned above.

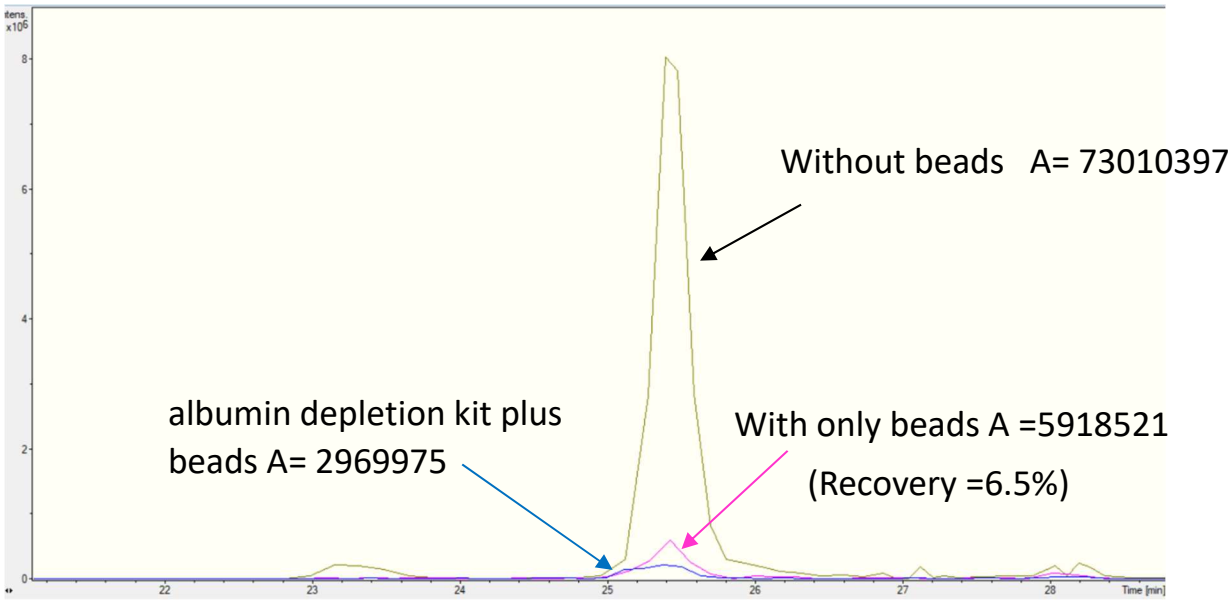


Figure 44.

A table of five representative synthetic peptides with the corresponding recovery and some of their properties is reported below

H2N-…-COOH	N° AA	PI	$[M+H]^{+}$ / $[M+2H]^{2+}$	Retention Time (minutes)	rPwab %	rPkab%
AMAPIKVRL	9	11.45	999.63 / 500.15	23.2	10.8	4.5
SLELGDSAL	9	3	905.47 / 453	25.3	5.9	2.2
VMTPRTLVL	9	10.8	1030.62 / 515.31	25.4	6.5	3.3
VMAPRTLIL	9	10.8	1014.62 / 508.31	26	22.9	3.3
VMAPRTLFL	9	10.8	1048.61 / 525.3	27	8.1	3.7

Table 8: First column: peptide sequence from amino terminal to carboxy terminal; second column: number of amino acids in the chain; third column: isoelectric point of peptide; fourth column: m/z value of mono protonated molecule and double protonated molecule; fifth column: retention time of extracted peak; rPwab% = percentage of recovery without treatment to remove the albumin; rPkab% = percentage of recovery after treatment with tools to remove the albumin

Looking at the last two columns of the table it is clear that the treatment with albumin depletion kit further lowers the recovery of the analyte.

Since these initial figures not encouraging in terms of peptide recovery and in order to avoid an extraction protocol that involves the use of three tools that lead to a further low recovery (decreasing), the decision was made to use different types of magnetic beads, expand the number of synthetic peptides with different characteristics in terms of polarity and chain lengths and to minimize the number of treatment tools to be used in sample preparation. For this it was decided to use a non-specific tool for the reduction of albumin avoiding the use of the tool tips C18. This was put in place in order to a reliable and reproducible extraction procedure that could also ensure acceptable recoveries. Also different types of beads were used, as such or derivatized.

3.3 Development of a new extraction method

To obtain useful information on the extraction tools efficacy a new peptides mixture has been designed and synthesized in our laboratory with the Fmoc method (see materials and methods). The potential outcomes could be also used to get a subsequent assignment of more circumscribed m/z values through the bioinformatic flow, as will also be discussed in the "materials and methods" section and described in detail later in this chapter. The new peptides mixtures contained 19 synthetic peptides ranging from 800 to 3000 Da (MIX19), taking into account that on real biological samples will be used a kit for the reduction of the protein component above 5000 Da. Seven of these peptides have been synthesized in our laboratory.

To set up the extraction method, MIX19 was analysed in two different conditions:

- The peptide mixture was analysed by HPLC-MS as such, without the use of any extraction tool.
- The peptide mixture was analysed by HPLC-MS after having undergone only extraction with different types of beads.

Moreover, of the nineteen peptides present in this new mixture, two were also used as a binding sequence during beads functionalization (Pept_AB4 and Pept_AB5).

Many are the advantages of using a mixture of known analytes

- Use a simplified sample (avoid the complexity of matrix sample and interference from high molecular weight proteins)
- Know the molecular mass limits related to different beads used
- find the best standard peptide to be added as internal standard to the sample of serum/plasma before extraction
- Check the effectiveness of the extraction of the beads in quantitative terms. Analysing the mixture as such or after the treatment with beads, allows to measure the recovery of the extraction tool.
- An immediate assignment of experimental peaks after tandem MS/MS analysis.

- Thanks to the different length, amino acids composition, polarity, hydrophobicity and isoelectric point of the peptides present in the MIX, understanding the correlations between the type of magnetic beads and the feature of extracted peptides.

-

peptide_accession	peptide_sequence [M]	label	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$ /$[M+4H]^{4+}$
pept_2_11	NITHFAIVASL	pept_AB0	1185.66	593.33	
pept_3_9	SLELGDSAL	Vβ1	904.46	452.73	
pept_4_9	MSAPRKVRL	En168B	1057.62	529.32	
pept_5_9	AMAPIKVRL	M12	998.61	499.81	
pept_6_25	MIPDIPTDISDQIKKEKSLLVDFFL	16B	2905.56	1453.29	969.18/727.13
pept_7_9	VMAPRTLIL	CW3	1013.61	507.31	
pept_8_9	VMAPVTVLL	B7	942.56	471.79	
pept_9_9	VMAPRTLFL	HLA-G	1047.59	524.30	
pept_10_25	VIPDIPKDISQQIHKEKVLMVELFM	16A	2950.60	1475.81	984.2/738.4
pept_11_9	VMAPRILIL	U2	1025.65	513.33	
pept_12_9	VMGPRTLIL	U3	999.59	500.30	
pept_13_9	VMTPRTLVL	U4	1029.60	515.31	
pept_14_7	ATKEQLK	pept_AB2	817.47	409.24	
pept_15_9	FPSHANAAG	pept_AB1	871.40	436.20	
pept_16_9	ILGKVFTLT	FLU	991.62	496.31	
pept_17_10	NGIGVTQNVL	pept_AB3	1014.55	507.78	
pept_18_16	AQSTIEEQADTFLDYD	pept_AB4	1845.80	923.4	
pept_19_11	NHEAEDLFYQS	pept_AB5	1353.53	677.27	
pept_20_7	DATDQLS	pept_AB6	749.83	375.415	

Table 9: List of peptides used as standard mixture (MIX19). The mass-to-charge values in red correspond to the protonated state of the peptide used to detect the analyte in the chromatogram.

All peptides were analysed through high performance liquid chromatography coupled to full scan mass spectrometry equipped with an electrospray ion source, all at the same concentration. The total count of the detected peptide ions constitutes the total ion current or total ion chromatogram (TIC). Each compound can be easily identified extracting the specific m/z current from the TIC (in red in the table 5).

3.4 Different types of Magnetic silica beads

With the aim to avoid the drawback of peptides low recovery and, overall, to differentiate the qualitative aspect of recovery in terms of type of the same, the successive extractive step has taken in consideration the use of different magnetic beads as such or functionalized in different ways with protocols set up according to their type.

The idea of using different types of magnetic beads boils down to considering different types of binding mechanism, for example affinity, and trying to extract specific fractions of peptides in terms of chain length, molecular weight, hydrophobicity, isoelectric point, type of amino acids that make up their chain.

label used to identify the magnetic silica beads	Sizes (µm)	type of interaction
RPC18	1.0	hydrophobic
B-NH2	1.0	cation exchange
B-COOH	1.0	anion exchange
B-GSH	1.0	anion exchange
B-PAB (with two different types of di peptides bonded on its surface)	1.0	affinity

Table 10. Types of magnetic silica beads used for this study.

The greatest advantages in using beads are the small volume of the sample used and the short incubation times for peptide adsorption, desorption and magnetic collection. This significantly reduce the time compared to the use of conventional protocols based on column solid phase extraction.

As previously reported in materials and methods, in addition to using them as such, some beads have been functionalized in our laboratory by adding amino or carboxyl groups.

Before proceeding to the use of magnetic beads on biological samples, several standard mixture solutions of synthetic peptides were tested, to evaluate their effectiveness both in terms of quality and selectivity and in terms of recovery. Magnetic beads commercially functionalized with C18 are the only ones to which the manufacturer's protocol has been applied.

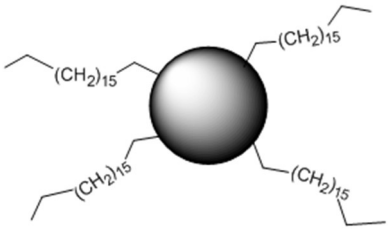


Figure 45. C18 bead

In the case of beads functionalized with aminic or carboxyl groups, not being commercialized to be used for peptides extraction, the suppliers did not provide a specific protocol; therefore it was necessary to develop one.

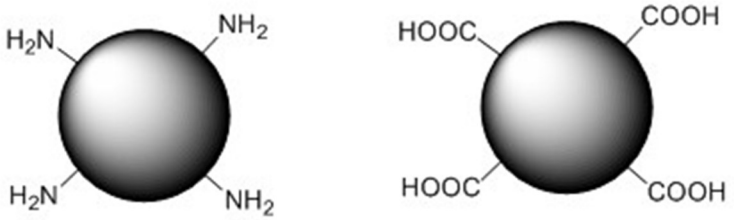


Fig.46. BNH2 and BCOOH beads

The other two types of beads (B-GSH and B-PAB) are derived from these two types for peptide insertion using crosslinking reagents (see materials and methods section for preparation scheme). For similar beads, being functionalized with a weak cation exchange (WCX) or a weak anion exchange (WAX), their extraction protocols have been found in the literature (https://www.magtivio.com/wp-content/uploads/2023/01/PS0014-0021_MagSi-WCX.pdf; https://www.magtivio.com/wp-content/uploads/2023/01/PS0013-0021_MagSi-WAX.pdf) slightly acidic pH buffer were considered the silica magnetic beads with NH2 functional groups operating with the principle of WCX, while for the COOH, BGSH and BPAB ones with WAX using a slightly basic pH buffer.

As previously reported, BGSH beads (functionalized with glutathione) and BPAB beads (functionalized with two different peptides) were obtained in our laboratory.

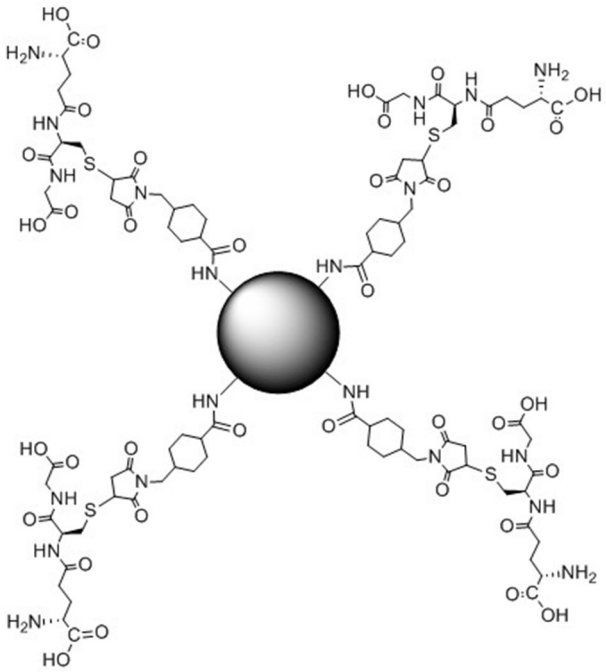


Figure 47. BGSH bead

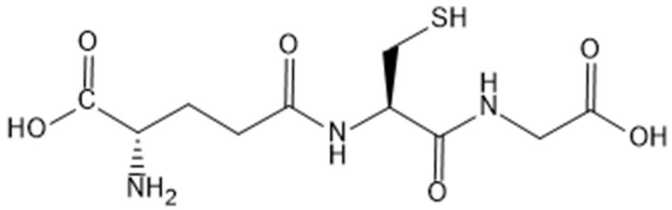


Figure 48. glutathione structure

For the type of chemical groups exposed on the surface, the functionalized GSH beads can be considered as chelating agent bidentate.

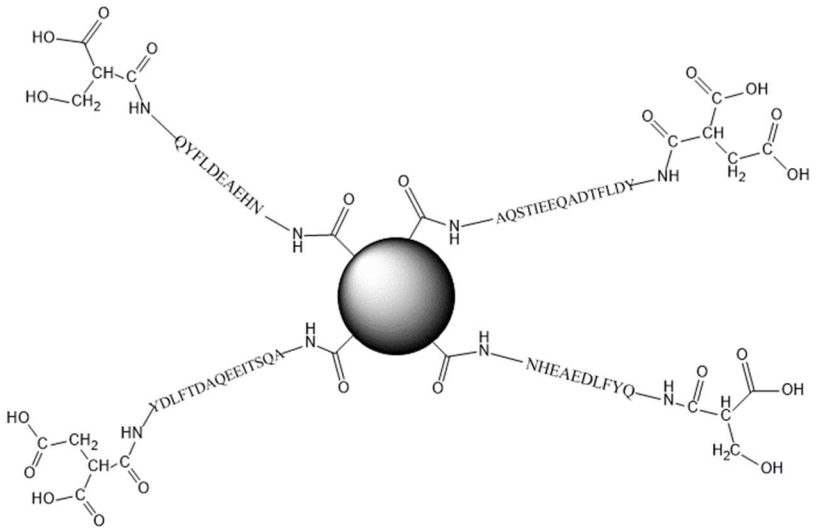


Fig.49. BPAB bead

A particular discussion concerns the magnetic beads labelled as B-PAB, designed and functionalized with specific peptides with the aim of creating an affinity interaction with the Receptor Binding Domain (RBD) portion of the spike glycoprotein of the Sars-Cov2 virus (figure 50).

Figure 50. General topology of the SARS-cov-2 spike monomer

In particular, the two different peptides exposed on the surface of the beads should mimic the receptor portions of ACE2 that is recognized and subsequently binds the two Receptor Binding Motif (RBM) of the of SARS-Cov-2. These beads were, obviously, also used with the mix standard solution containing the 19 synthetic peptides (MIX 19).

During the design and synthesis of these two peptides the portions of structure of ACE2 receptor, binding to RBD of SARS-Cov-2, was considered (figure51).

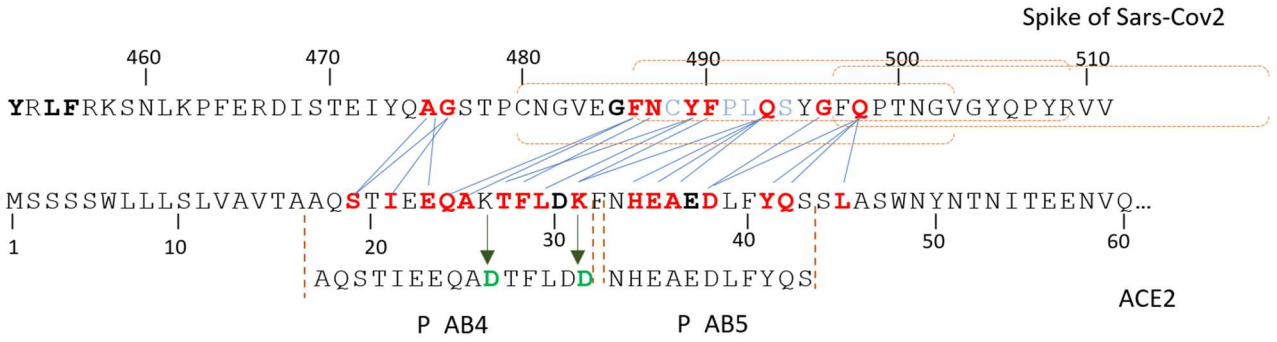


Figure 51. The sequence at the top represents one of the two Receptor Binding Motif portions within the RBD of glycoprotein spike of Sars-CoV-2. RBM interacts with ACE2 receptor at the binding interface. The blue lines of connection between the two opposing sequences highlights the specific interaction. Obviously, the spatiality and the folding of the secondary structure is missing. In bold red letters are indicated the amino acids that interact with each other giving rise to the binding among virus spike and ACE receptor. The bold black letters indicate the amino acids that also interact, but their interactions are in another portion of the respective sequences In brackets are reported the three sequences (from 15 residues each) that were captured by the PAB beads for affinity by using a mixture of 15-mer peptides from the spike protein of Sars-cov-2 virus. P_PAB4 and P_AB5 are the two peptides synthesized that follow two consecutive portions of RBM. [105]

However the original sequences of structure of portion of ACE2 receptor was modified to create optimal chemical conditions of binding to the carboxyl beads using a crosslinking reagent. It was decided to make magnetic silica beads functionalized with both peptides (P_AB4 and P_AB5)

starting from the beads having the carboxyl groups exposed on its surface (see materials and method section for details). In particular, in case of beads with carboxyl groups, it is important not to use amino acids with amino group in the side chain in order to avoid undesired ligation or by-products during the synthesis. These two aspects were carefully considered during the design phase of the peptide sequence; for instance, the substitution of lysine in the peptide labelled as PAB4 with amino acids that would seem to give a better binding was considered, as reported in the literature [106]. For this reason, the two lysine residues (K) in the peptide P_AB4 were replaced with aspartic acid (D).

3.5 How data were analysed

An in-depth qualitative and quantitative study of the chromatographic and spectral results of MIX 19 was carried out comparing the data obtained analysing the peptide mixture as such and those obtained analysing MIX19 after extraction with the use of different kinds of beads. An important part of this work consists in correctly assigning the m/z values present in the full scan chromatogram to the peptide sequences of the MIX19 whole or fragmented, considering the mass values having signal/noise ratios of 5. The assignment was made with a computational method based on the Python Code I. Through this approach it was possible to identify the m/z ratios of interest by comparing the analytical data with data obtained in silico. In addition to the peptides not fragmented, the use of our algorithms has allowed to characterize fragments generated in silico eliminating only one or two amino acids from the amino- or from the carboxyl-terminus and considering the mono-, double-, triple- and quadruple charged spectral signals within a range of ± 0.1 Da.

The calculations carried out on each peptide sequence have considered its main chemical properties, starting from the properties of the individual amino acids, regardless of their protonation state.

The calculation of the physical chemical properties was obtained using some open-source libraries and implemented in Code Python I. For the calculation of hydrophobic parameters (Cowan [107] and Wilson [108]), aliphatic index [109], Instability index (index of Boman [110]) the library "peptides.py" was used, this is a pure-Python package to compute common descriptors for protein sequences. The calculation of logP (the logarithm of the partitioning coefficient between n-octanol and water) [111, 112] was obtained with the library rdkit.Chem while Gravy (Grand Average of

Hydropathy) [113] and PI (isoelectric point) were obtained with library Bio.SeqUtils including in Biopyton package.

In addition to the calculation of the physical chemical parameters, the properties of the sequences in terms of the properties of the amino acids that compose it. The amino acids of the sequence were classified in aliphatic ('A', 'I', 'L', 'M', 'P', 'V','G'), aromatic ('F', 'Y', 'W'), hydrophobic ('A', 'F', 'G', 'I', 'L', 'M', 'P', 'V', 'W'), hydrophilic ('C', 'N', 'Q', 'S', 'T', 'Y'), basic ('H', 'K', 'R'), acid ('D', 'E') and protonable ('H','K'), of these were calculated their percentages in the sequence.

Below is an example of a typical output file from Code I; each column contains the sequence information that has been associated with a given value of m/z (peak column).

peptide_sequence	peptide_accession	peak	charge	lenght	log_P	aliph_idx	pept_Boman	IP	H_Cowan	H_Wilson	Gravy	%aa_aliph	%aa_arom	%aa_hydrophobic	%aa_hydrophilic	%aa_basic	%aa_acid	%aa_prot
TDQLS	pept_20_7_i_2	563.5	1	5	-4.50	78.00	3.06	3.75	-0.44	0.44	-0.94	20.00	0.00	20.00	60.00	0.00	20.00	0.00
ATKEQ	pept_14_7_f_2	576.5	1	5	-4.00	20.00	3.73	6.41	-0.87	-1.26	-1.96	20.00	0.00	20.00	40.00	20.00	20.00	20.00
SHANAAG	pept_15_9_i_2	627.5	1	7	-6.17	42.86	1.19	7.55	-0.17	-0.26	-0.36	57.14	0.00	57.14	28.57	14.29	0.00	14.29
KEQLK	pept_14_7_i_2	645.5	1	5	-2.23	78.00	3.71	9.54	-0.83	-0.16	-2.20	20.00	0.00	20.00	20.00	40.00	20.00	40.00
AQSTIEEQADTFLD	pept_18_16_f_2	1567.8	1	14	-8.96	70.00	2.45	3.30	-0.37	0.69	-0.61	28.57	7.14	35.71	35.71	0.00	28.57	0.00
STIEEQADTFLDYD	pept_18_16_i_2	1646.9	1	14	-7.82	62.86	2.81	3.20	-0.46	1.13	-0.83	21.43	14.29	28.57	35.71	0.00	35.71	0.00
SLELGD	fragm_3_6	317.4	2	6	-3.12	130.00	1.36	3.55	-0.19	2.07	-0.10	50.00	0.00	50.00	16.67	0.00	33.33	0.00
SLELGDS	pept_3_9_f_2	360.6	2	7	-4.65	111.43	1.65	3.55	-0.25	1.69	-0.20	42.86	0.00	42.86	28.57	0.00	28.57	0.00
TKEQLK	pept_14_7_i_1	373.5	2	6	-3.36	65.00	3.52	9.54	-0.74	-0.50	-1.95	16.67	0.00	16.67	33.33	33.33	16.67	33.33
DATDQLS	pept_20_7	375.4	2	7	-6.04	70.00	3.17	3.49	-0.57	0.07	-0.91	28.57	0.00	28.57	42.86	0.00	28.57	0.00
APRTVLL	pept_8_9_i_2	385.5	2	7	-1.77	167.14	0.26	10.55	0.62	2.53	0.97	71.43	0.00	71.43	14.29	14.29	0.00	0.00
GPRTLIL	pept_12_9_i_2	385.5	2	7	-1.77	167.14	0.26	10.55	0.64	2.51	0.70	71.43	0.00	71.43	14.29	14.29	0.00	0.00
APRTLIL	pept_7_9_i_2	392.4	2	7	-1.38	181.43	0.13	10.55	0.69	2.30	1.01	71.43	0.00	71.43	14.29	14.29	0.00	0.00
MSAPRKV	pept_4_9_f_2	394.5	2	7	-3.37	55.71	2.24	11.65	-0.01	0.71	-0.41	57.14	0.00	57.14	14.29	28.57	0.00	14.29
SLELGDSA	pept_3_9_f_1	396.4	2	8	-5.14	110.00	1.22	3.55	-0.18	1.44	0.05	50.00	0.00	50.00	25.00	0.00	25.00	0.00
APRILIL	pept_11_9_i_2	398.5	2	7	0.28	237.14	-0.94	10.55	0.99	3.23	1.76	85.71	0.00	85.71	0.00	14.29	0.00	0.00

(I) (II) (III)

Figure 52. A typical output file from Code I obtained considering the m/z signals from the full scan chromatogram. In the first block of columns (I) the peptide sequence, peptide code of accession in the list of sequences, m/z value of the peak, degree of protonation of peptide are reported. The second block of columns (II) contains other important parameters: length of sequence, log_P, aliphatic Index, index of Boman, Isoelectric point, Cowan hydrophobic parameter, Wilson hydrophobic parameter, Gravy index.

Finally, in the third block of columns (III) the percentages of amino acids classified as aliphatic, aromatic, hydrophobic, hydrophilic, basic, acidic and protonable amino acids are reported.

For a comparison of the peptides extracted from the various magnetic beads, from the qualitative point of view, the Venn diagram was also used.

3.6.1 HPLC-MS TIC and spectrum full scan of MIX 19 analysed as such

Initially, MIX19 was analyzed as such to obtain a total ion current chromatogram (TIC) containing all the useful information for the qualitative-quantitative analysis of samples subjected to extraction with several types of beads.

The TIC analysis allowed, through the use of Python Code I, the identification of different signals used as reference. Figure 53, reports the TIC obtained.

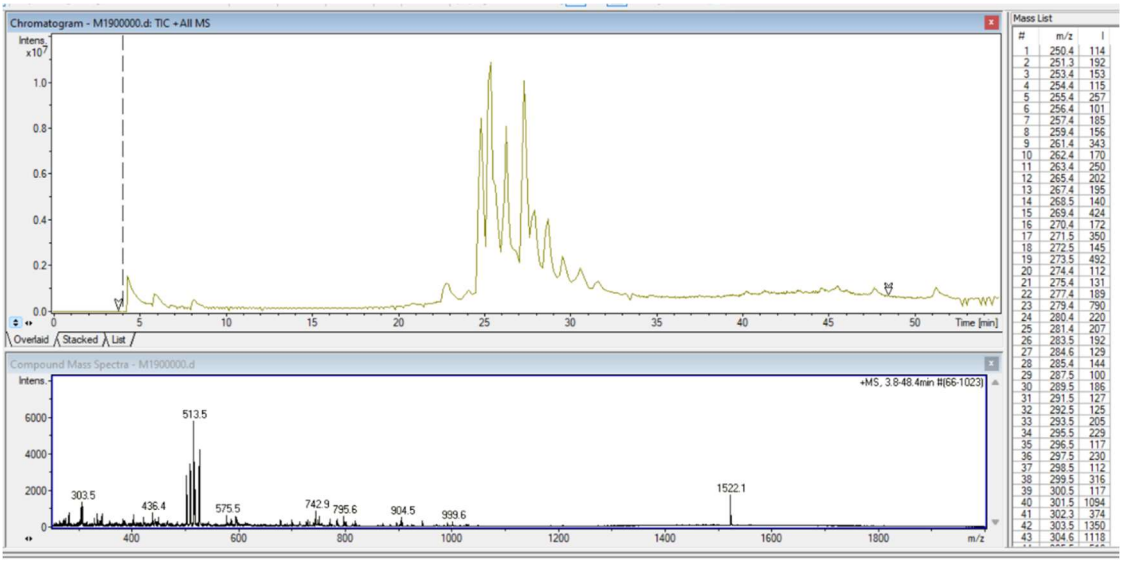


Figure 53. MIX 19 TIC and extracted peaks form spectrum level 1 in a range of RT 3.8 - 48.4 min.

Range RT	Number of peaks	Lower mass/higher mass	Lower intensity/higher intensity
3.8 – 48.4	630	257.2/949.5	3550/702865

Table 11. Parameters used to obtain data to be analysed with the Code I

In the following table the parameters used to obtain the data to be analysed with the Code I are reported

peptide_sequence [M]	peptide_accession	Peak (m/z)	State of charge
SLELGDSAL	pept_3_9	302.3	$[M+3H]^{3+}$
SLELGDSAL	pept_3_9	904.5	$[M+H]^{+}$
MSAPRKVRL	pept_4_9	529.5	$[M+2H]^{2+}$
AMAPIKVRL	pept_5_9	333.6	$[M+3H]^{3+}$
MIPDIPTDISDQIKKEKSLLVDFFL	pept_6_25	1453.2	$[M+2H]^{2+}$
VMAPRTLIL	pept_7_9	507.5	$[M+2H]^{2+}$
VMAPRTLIL	pept_7_9	1013.7	$[M+H]^{+}$
VMAPVTVLL	pept_8_9	471.7	$[M+2H]^{2+}$
VMAPRTLFL	pept_9_9	524.5	$[M+2H]^{2+}$
VMAPRTLFL	pept_9_9	1047.7	$[M+H]^{+}$
VIPDIPKDISQQIHKEKVLMVELFM	pept_10_25	984.4	$[M+3H]^{3+}$
VIPDIPKDISQQIHKEKVLMVELFM	pept_10_25	1476	$[M+2H]^{2+}$
IPDIPTDISDQIKKEKSLLVDFFL	pept_6_25_i_1	694.3	$[M+4H]^{4+}$
IPDIPTDISDQIKKEKSLLVDFFL	pept_6_25_i_1	1387.7	$[M+2H]^{2+}$
MIPDIPTDISDQIKKEKSLLVDF	pept_6_25_f_2	882.6	$[M+3H]^{3+}$
MIPDIPTDISDQIKKEKSLLVDF	pept_6_25_f_2	1323.2	$[M+2H]^{2+}$
SLELGD	frag_pept_3_6_1	633.5	$[M+H]^{+}$

Table 12. List of some m/z values of the peaks extracted from the instrument output and

corresponding calculated m/z of the sequences of the standard peptides considering also their different state of charge and fragmentation. Note that some same sequences are present and detected with different value of m/z corresponding to their different states of protonation.

Of the 630 spectral peaks, 105 were matched considering only the ionic sequences of the peptide and peaks with values of m/z > 250. The distribution is reported in Table 12.

m/z > 250 (TIC)	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
Whole sequence	8	12	6	5
Fragment (CID)	22	26	17	9
Tot.	30	38	23	14

Table 13. Statistics of the number of ions detected, classified by their charge and belonging to whole or fragmented sequences.

The 105 peaks recognized by the Code I, belong to 65 structures, 16 of which as whole sequences and 49 as their fragments.

At this level of study, only full scan data were analysed, omitting the characterization of the fragments obtained following tandem analysis (MS/MS).

In house Data Analysis software equipping the ion-trap instrument was used to determine the area of the most intense peaks. These areas were used as a reference for the analysis on samples submitted to extraction with beads.

The first two graphs (figure 54 and figure 55) show the properties of peptides detected derived from the block II columns of the output of the Python Code I. In the first graph we have on the x-axis the number of amino acids of the peptide chain while on the y-axis the properties as reported in the graph legend. The second graph (figure 55) represents a set of boxes and whiskers charts, one for each property, showing the distribution of data in quartiles, highlighting the average and outliers. These plots are efficient tools that helps to determine how property values are distributed in a dataset. In figure 54 and 55 the aliphatic index values is shown in logarithmic scale for the sake of clarity.

The last two graphs (figure 56 and figure 57) show the properties of the amino acids composition of the peptides, extracted from the columns of block III of the Python Code I output file. In particular the third chart (figure 56) shows the properties of the amino acid composition (y-axis) of the sequence distributed over the length of the peptide chain (x-axis) while the fourth graph (figure 57) represents a set of boxes and whiskers charts, one for each property.

Statistic diagram of 65 unique peptide structures detected.

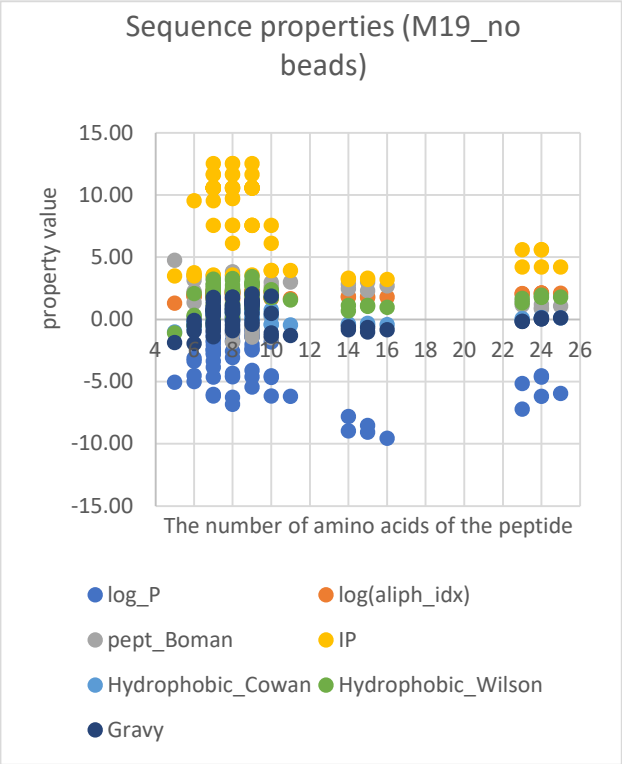


Figure 54. Scatter plot: properties/length of the peptide chain

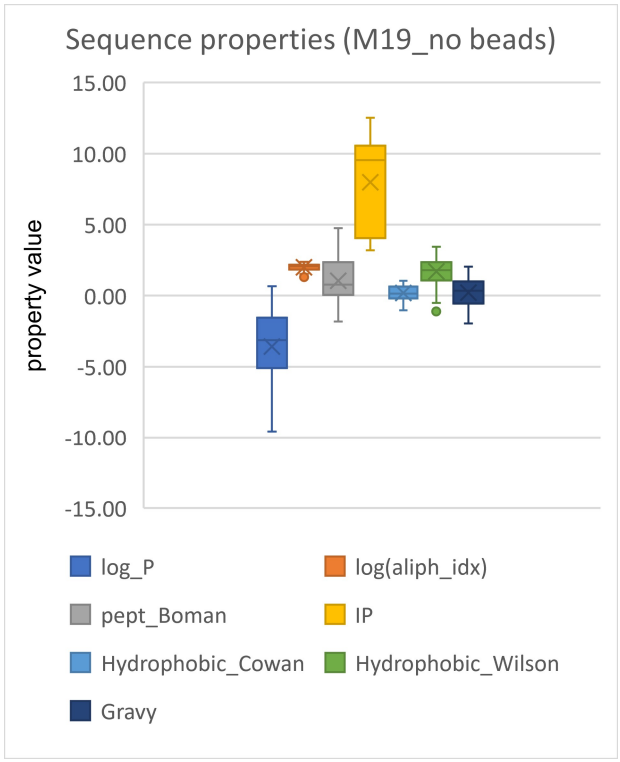


Figure 55. Boxes and whiskers charts of properties of the peptide with respect to its length

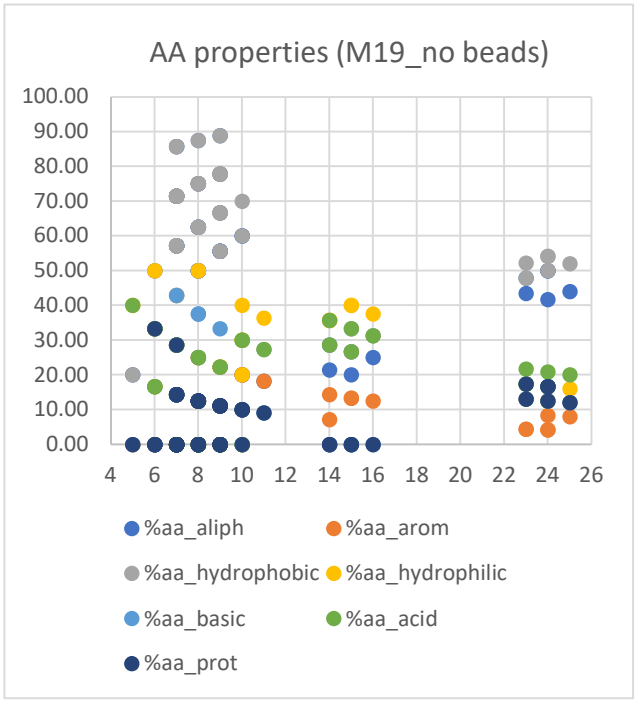


Figure 56. Scatter plot: percentage composition of amino acids classified by properties of the peptide with respect to its length.

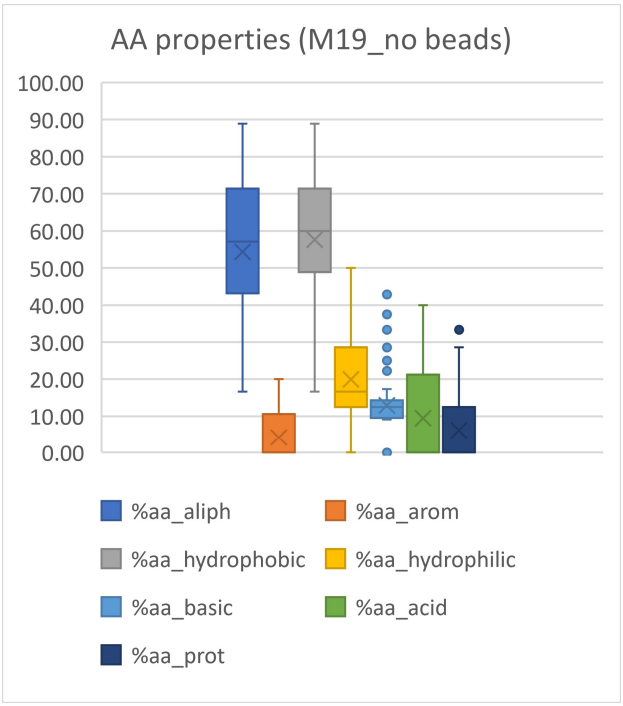


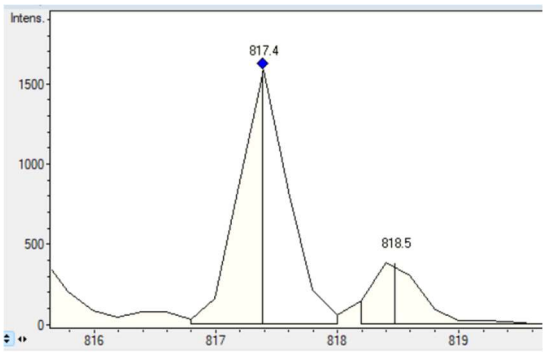
Figure 57. Boxes and whiskers charts of percentage composition of amino acids classified by properties of the peptide with respect to its length.

The peptide concentrations in the MIX19 used to obtain these data was the same for extraction experiments carried out with the different kinds of beads.

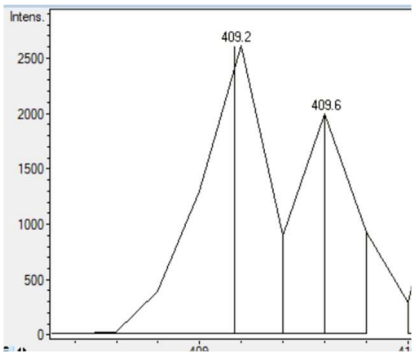
peptide_accession	peptide_sequence	m/z	RT [min]	Area
pept_2_11	NITHFAIVASL	593.33	25.8	7320835
pept_3_9	SLELGDSAL	904.45	24.9	7197523
pept_4_9	MSAPRKVRL	529.31	8.9	349923
pept_5_9	AMAPIKVRL	499.81	23	3650564
pept_6_25	MIPDIPTDISDQIKKEKSLLVDFFL	969.18/727.13(*)	28.5	1048902(*)
pept_7_9	VMAPRTLIL	507.31	25.3	49729854
pept_8_9	VMAPVTVLL	942.56	28.7	5087679
pept_9_9	VMAPRTLFL	524.30	26.4	53797340
pept_10_25	VIPDIPKDISQQIHKEKVLMVELFM	984.2/738.4(*)	29.5	2329772 (*)
pept_11_9	VMAPRILIL	1025.64	27.3	1144220
pept_12_9	VMGPRTLIL	500.30	26.4	53797340
pept_13_9	VMTPRTLVL	515.30	27.5	70300673
pept_14_7	ATKEQLK	817.47	26.5	266710
pept_15_9	FPSHANAAG	436.20	4.5	7213488
pept_16_9	ILGKVFTLT	990.60	4.5	7213488
pept_17_10	NGIGVTQNVL	507.78	24.4	37950044
pept_18_16	AQSTIEEQADTFLDYD	923.40	25.9	1095849
pept_19_11	NHEAEDLFYQS	677.27	23	9195734
pept_20_7	DATDQLS	749.83	29.7	3826214
frag_pept_3_6_1	SLELGD	633.30	21	542862
frag_pept_11_7_3	APRILIL	795.55	27.5	4543485
frag_pept_13_3	TPRTLVL	799.50	25	3602358
frag_pept_15_7_3	FPSHANA	743.35	27.8	2323830
frag_pept_14_6_1	ATKEQL	689.38	30.1	406182
frag_pept_14_6_2	TKEQLK	746.44	26.5	266710
frag_pept_14_5_4	ATKEQ	576.30	25.8	2983486
frag_pept_16_8_2	ILGKVFTL	890.57	45.4	533821

Table 14. List of standard peptides: for each peptide the m/z value as detected in the spectrum, the retention time and the peak area of extracted ion chromatography (EIC) is reported. The areas of the peaks will be compared with the corresponding area of same sequence also with different degree of protonation, if exists, extracted with the different types of beads. Note that thanks to Code I and the list of possible fragments of MIX19, CID ions from the fragmentation of the peptides synthetic present in the solution were also considered. For the m/z values that bear this symbol (*) see the explanation in the text below.

To note that, to compare the recovery of a standard peptide in the several types of beads it was necessary to take into account also the presence of multi-charge ions belonging to the same sequence, especially for sequences containing amino acids with protonable functional groups in the side chains. Thus, all quantitative data are the result of the sum of signals with different charge states related to the same peptide (Figure 58).



[ATKEQLK+H]$^{+}$ mono charge



[ATKEQLK+2H]$^{2+}$ double charge

Figure 58. Spectrum detail and magnification at instrument resolution level

Evidence of CID fragments in the full scan spectrum and assignment of fragments to portions of structures of whole peptide sequences of MIX19

In the TIC, in addition to values of m/z belonging to entire known sequences, several signals whose values of m/z belong to their fragments were found. Some standard peptides are not detected as whole molecule in the spectrum, due to their a priori degradation in the pure peptide solution or because they are detected as fragments (labelled later as frag_pept_n… or pept_n_i_ / pept_n_f…). The peptides ions fragments detected at MS spectrum level 1 are produced by collision-induced dissociation (CID) in the ionization source. It is possible to manually assign the m/z values of the CID signals to specific sequences through the Data Analysis software equipping the ion-trap instrument: however, this is in principle a work requiring a lot of resources and is not free from the danger of losing useful information. However, it was possible to proceed more quickly and reliably thanks to the generation of sequences of fragments in silico with the Python Code I so that the values of m/z could also be assigned to the sequence fragments and not only to the entire sequences. The following is an example of two extracted ion chromatograms (EIC), both related to mono-charged ions, belonging to the same peptide chain. Since they have different retention times they are not CID fragments; presumably the formation of the smallest ion has occurred in the peptide solution, before injection.

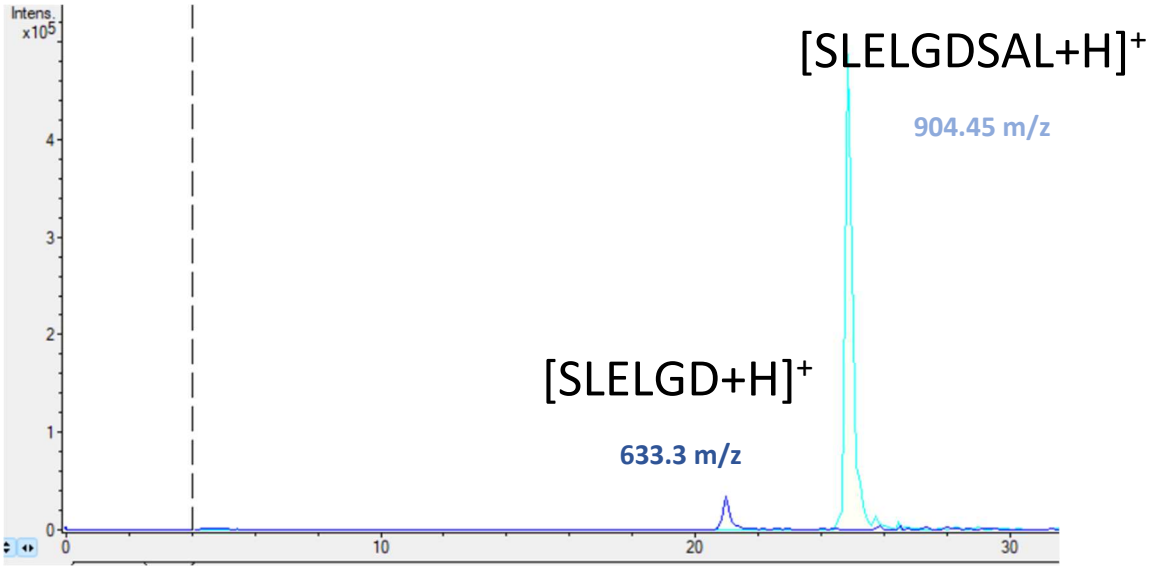


Figure 59. Detail of the current relative to two mono-charged ions belonging to the same peptide chain, extracted from the TIC

$[M+H]^+$	m/z		CID (coelutes) m/z value	
$[ATKEQLK+H]^+$	817,47		$[ATKEQLK+2H]^{2+}$ 409.24	
$[TKEQLK+H]^+$	746,5		$[ATKEQLK+H]^+$ 817,47 (*)	
$[ATKEL+H]^+$	689,4			
$[KEQLK+H]^+$	645,3			
$[ATKEQ+H]^+$	576,3		$[TKEQLK+H]^+$ 746,5	

Figure 60. Some extracted ion chromatograms that represent peptides fragments that origin from the same standard peptide (e.g. ATKEQLK), Some of these are CID, in fact they co-elute in the chromatogram.

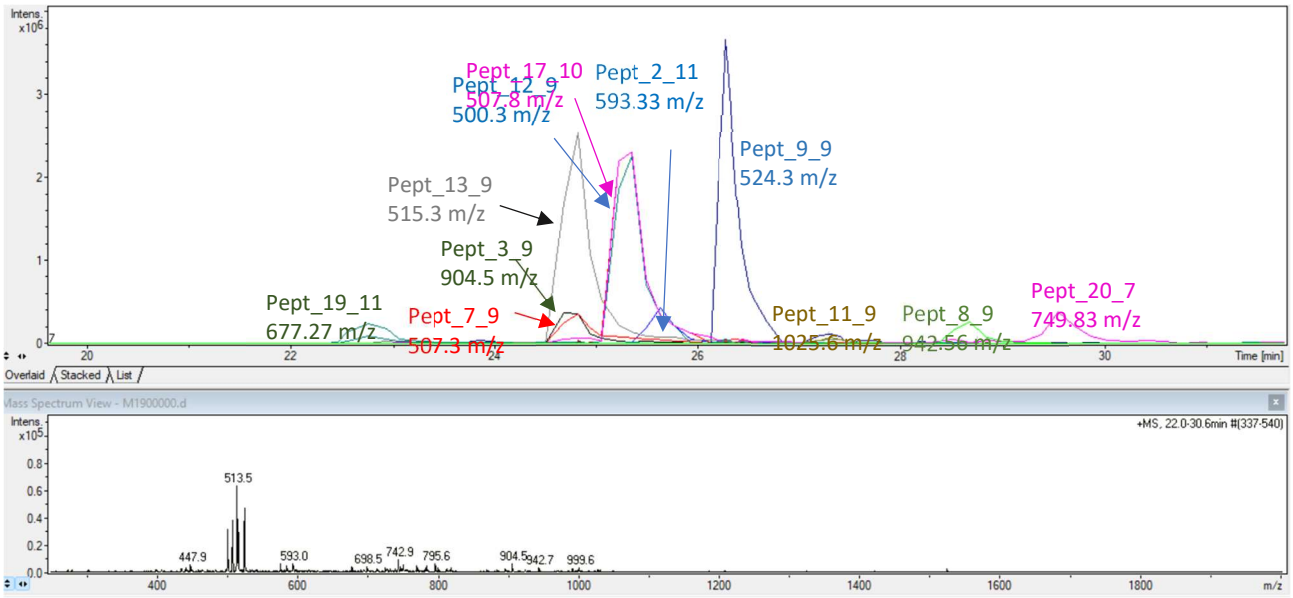


Figure 61. Some peptide ions extracted by TIC and a portion of full spectrum in the RT range between 22.00 – 30.6 min

A particular case regarding the sequences with 25 AA (*)(pept_6 and pept_10, table 14). They are present only as triple-charged and tetra-charged ions; therefore, for their quantification, the total area was obtained by adding the contribution of the areas of these two ions. Observing the amino acid composition of the two sequences this is predictable for the presence of three lysine residues (protonable amino acid)

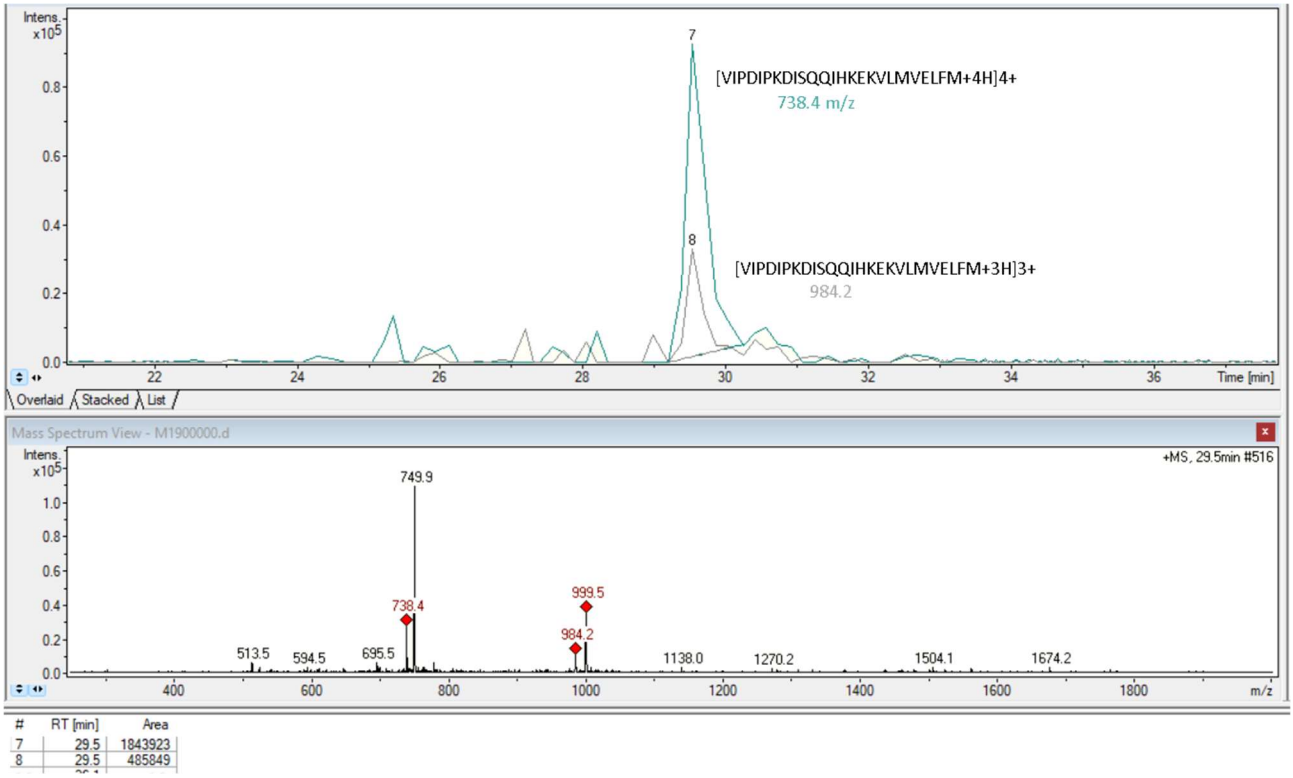


Figure 62. EIC of Pept_10_25 triple-charged and tetra-charged ions

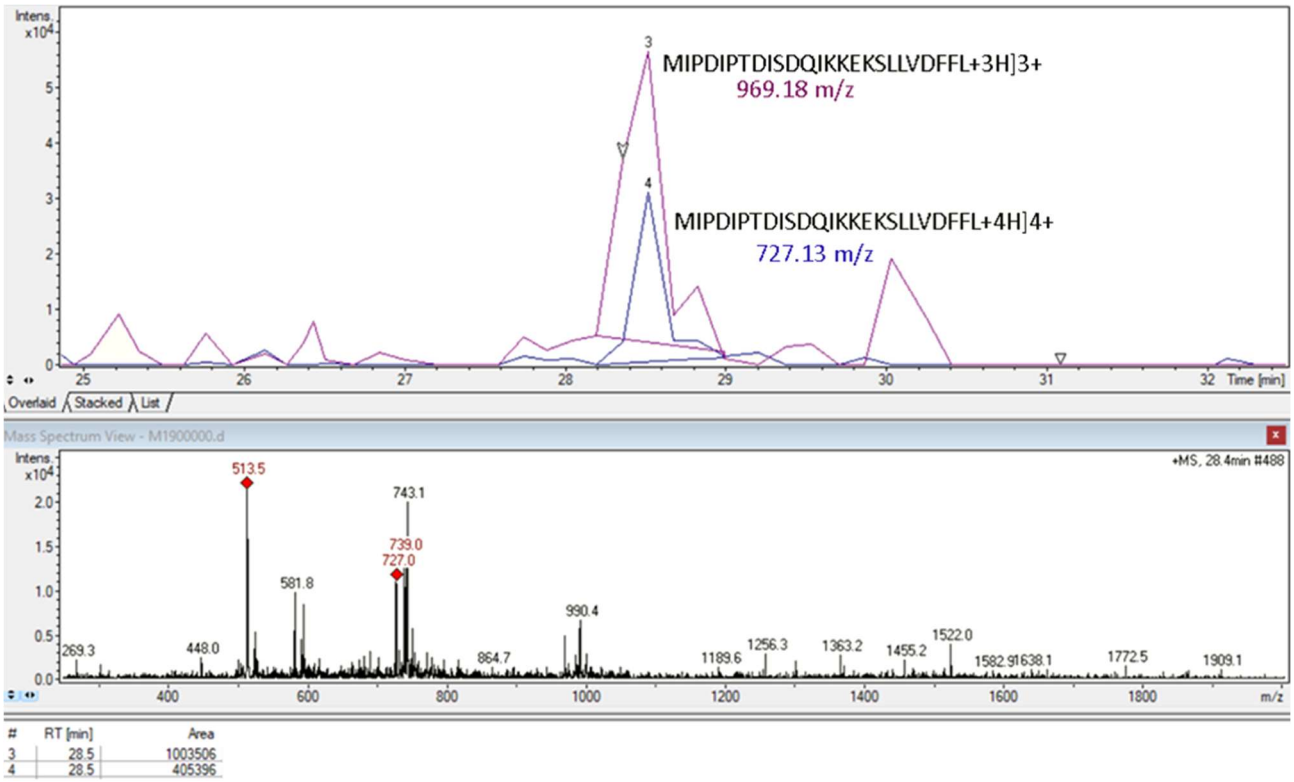


Figure 63. EIC of Pept_6_25 triple-charged and tetra-charged ions

3.6.2 Studies on recovery of standard peptides with different types of beads, qualitative and quantitative aspect

The percentage recovery value is derived from the ratio of the peak area extracted from a certain type of bead to the peak area of the corresponding value of m/z, obtained analysing MIX 19 without any extraction with beads. Remembering that in the samples of the standard solutions with or without the use of beads, the same concentration of 1 pmol/µL has always been used.

It is important to highlight that the recoveries values are constant and reproducible.

In the tables of the next sections, we can observe that some peptides, especially CID fragments, are extracted only from some types of magnetic beads selectively: in fact, their corresponding values of m/ z are not present in the spectrum of the standard mixture without the use of beads and therefore the data of recovery is missing. These are fragments of peptide sequences which presumably represent degradation products already present in small quantities in MIX 19. Probably, they are found only in the analyses related to samples extracted with beads thanks to their ability to concentrate the bound peptides.

3.6.2.1 HPLC-MS TIC and spectrum full scan of MIX 19 with BNH2 beads

This is the only beads type, among those used, that has been applied the Weak Cation eXchange (WCX) protocol with the use of a buffer at pH 4 (Ammonium Acetate).

Volume of slurry containing BNH2 magnetic beads: 20 µL

Volume of solution standard peptides (MIX19) all at a concentrantion of 1pmol/µL: 40 µL

Contact time between beads and peptides solution: 5 minutes

Desorption Solution: 1 % TFA (incubate for 2 minutes at room temperature before separation and remove the eluate).

Please refer to the materials and methods section about details on the different steps in the operating procedure, the washing solutions and their composition.

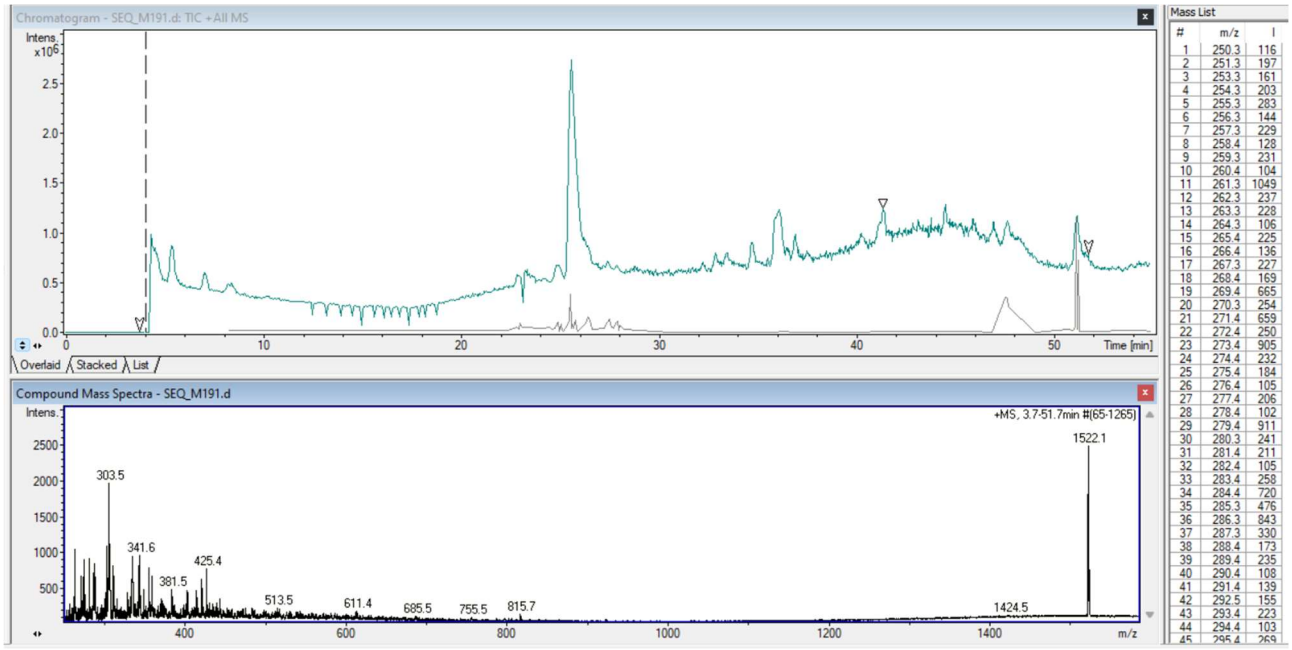


Figure 64. TIC from MIX 19 and extracted peaks from spectrum level 1 in a range of RT 3.7 – 51.7 min.

Range RT	Range index	Number of peaks	Lower mass/higher mass	Lower intensity/higher intensity
3.7 – 51.7	65-1265	620	250.0/1668	100/2503

Table 14. Parameters used to obtain data to be analysed with the Code I

Of the 620 spectral peaks, 89 were matched considering only the ionic sequences of the peptide and peaks with values of m/z > 250. The distribution is reported in Table 15.

m/z > 250 (TIC)	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
Whole sequence	0	10	9	7
Fragment (CID)	6	25	26	6
Tot.	6	35	35	13

Table 15. Statistics of the number of ions detected, classified by their charge and belonging to whole or fragmented sequences.

The 89 peaks recognized by the Code I, belong to 66 structures, 16 of which as whole sequences and 50 as their fragments.

Statistic diagram of 66 unique peptide structures detected

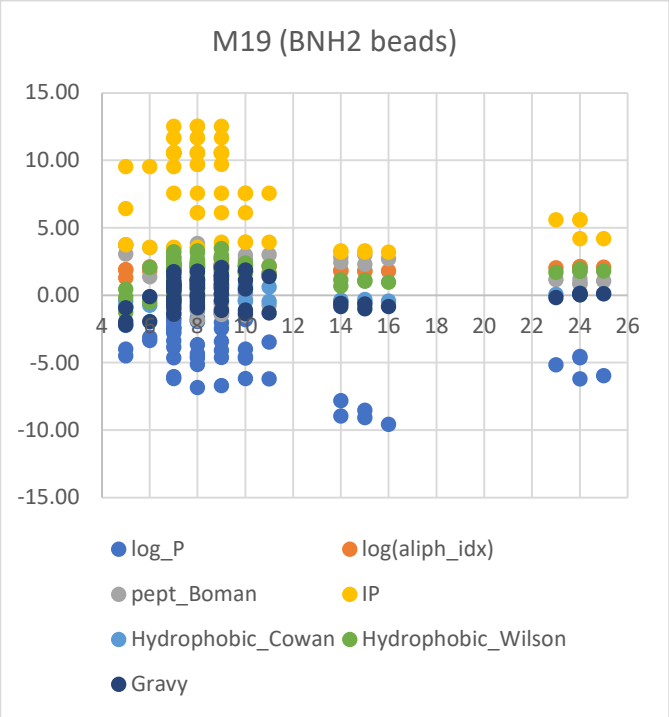


Figure 65. Scatter plot: properties of the peptide with respect to its length

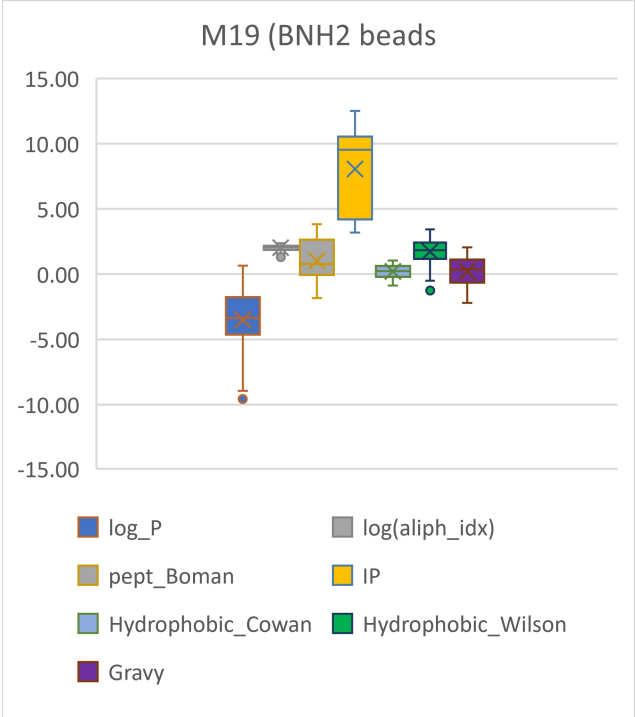


Figure 66. Boxes and whiskers charts of properties of the peptide with respect to its length

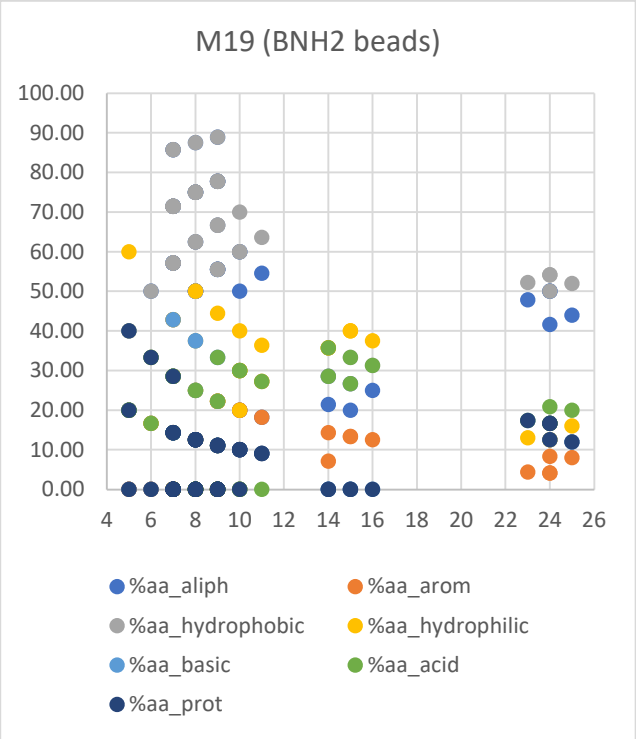


Figure 67. Scatter plot: percentage composition of amino acids classified by properties of the peptide with respect to its length.

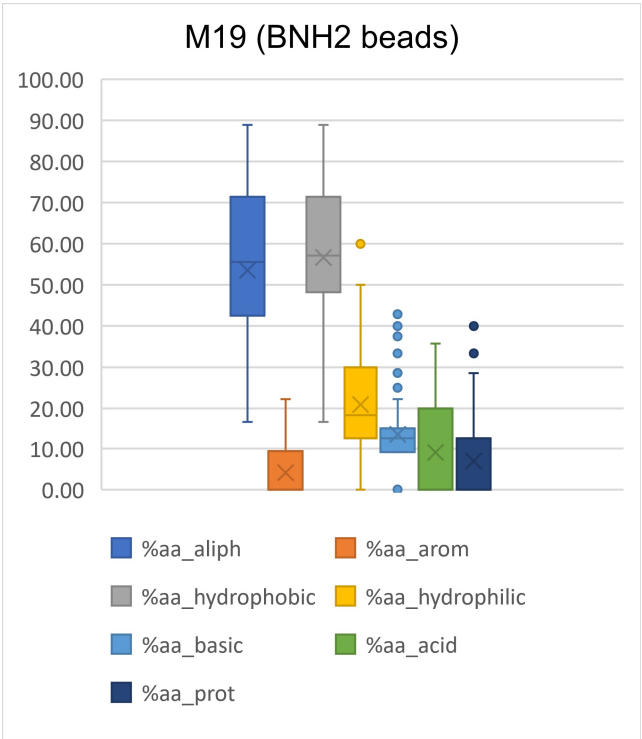


Figure 68. Boxes and whiskers charts of percentage composition of amino acids classified by properties of the peptide with respect to its length.

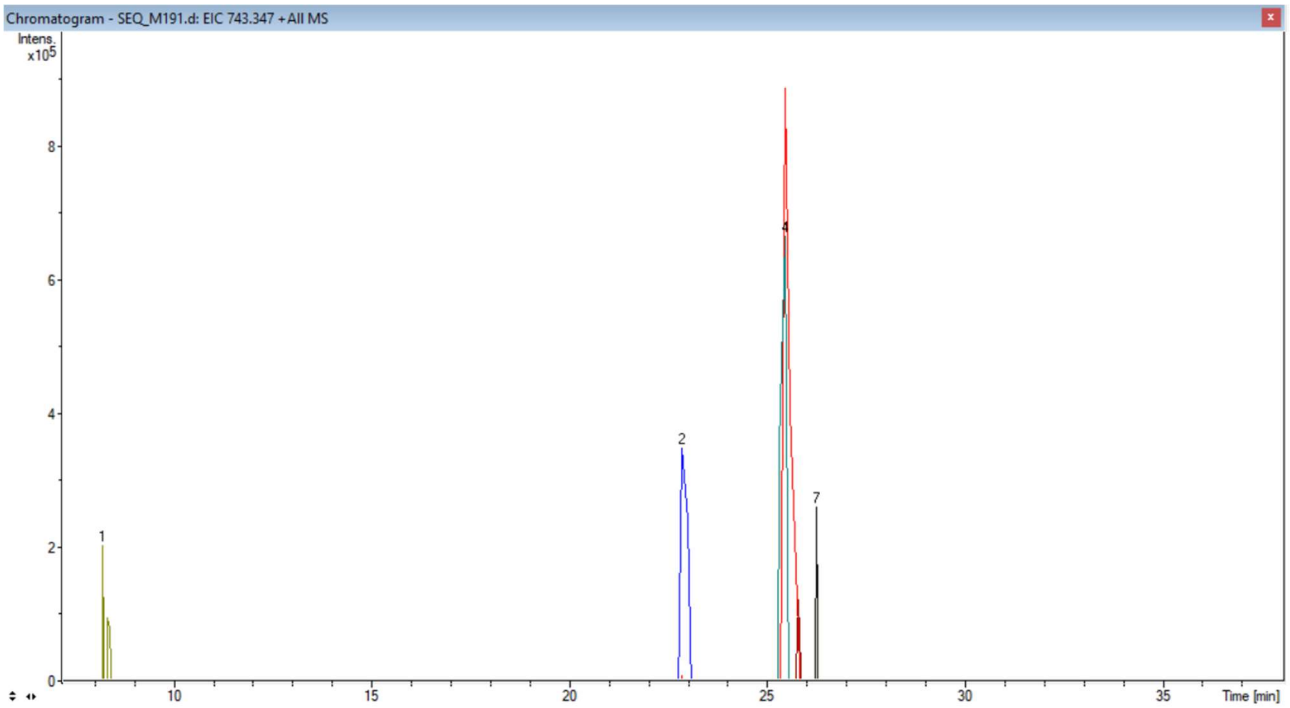


Figure 69. The chromatographic peaks of some ions as they appear in the extracted ion current chromatogram (EIC). EICs have been used to calculate areas and thus recoveries.

peptide_accession	peptide_sequence	m/z	RT [min]	recovery (%)
pept_4_9	MSAPRKVRL	529.3	8.6 (1)	6.7
pept_12_9	VMGPRTLIL	500.3	22.8 (2)	22.6
pept_20_f_2	DATDQ	549.2	24.8	
pept_17_10	NGIGVTQNVL	507.8	25.4	12.3
pept_2_11	NITHFAIVASL	593.3	25.8	42.7
frag_pept_14_5_4	ATKEQ	576.3	25.8(4)	40.8
pept_13_9	VMTPRTLVL	515.3	26.2(7)	8.3
frag_pept_15_7_3	FPSHANA	743.3	28	17.2
pept_9_9	VMAPRTLFL	524.3	27	14.4
pept_7_9	VMAPRTLIL	507.3	25.5	16.3

Table 16. percentage recoveries of the most intense peaks for BNH2 beads.

3.6.2.2 HPLC-MS TIC and spectrum full scan of MIX 19 with BGSH beads

Protocol used: Weak Anion eXchange

Buffer solution: MES (2-morpholin-4-ylethanesulfonic acid) at pH6

Volume of slurry containing BGSH magnetic beads:: 20 µL

Volume of solution standard peptides (MIX19) all at a concentrantion of 1pmol/µL: 40 µL

Contact time between beads and peptides solution: 5 minutes

Desorption Solution: 1 % TFA (incubate for 2 minutes at room temperature before separation and remove the eluate). Please refer to the materials and methods section about details on the different steps in the operating procedure, the washing solutions and their composition.

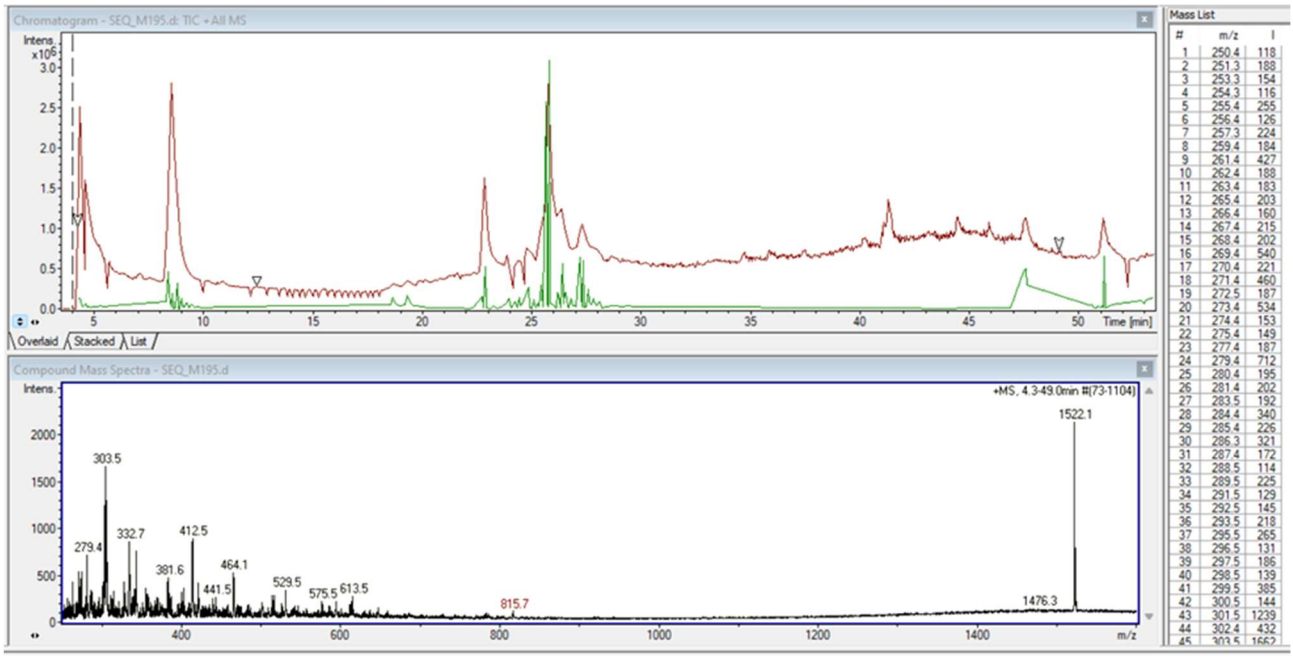


Figure 70 MIX 19 TIC and extracted peaks from spectrum level 1 in a range of RT 4.3 – 49.0 min.

Range RT	Range index	Number of peaks	Lower mass/higher mass	Lower intensity/higher intensity
4.3 – 49.0	74-1104	562	250.4/1702.0	100/2128

Table 17. Parameters used to obtain data to be analysed with the Code I

Of the 562 spectral peaks, 62 were matched considering only the ionic sequences of the peptide and peaks with values of m/z > 250. The distribution is reported in Table 18.

m/z > 250 (TIC)	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
Whole sequence	0	10	6	5
Fragment (CID)	2	18	19	3
Tot.	2	28	25	8

Table 18. Statistics of the number of ions detected, classified by their charge and belonging to whole or fragmented sequences

The 62 peaks recognized by the Code I, belong to the 49 unique structures of which 12 in whole sequences and 37 in its fragments.

Statistic diagram of 49 unique peptide structures detected.

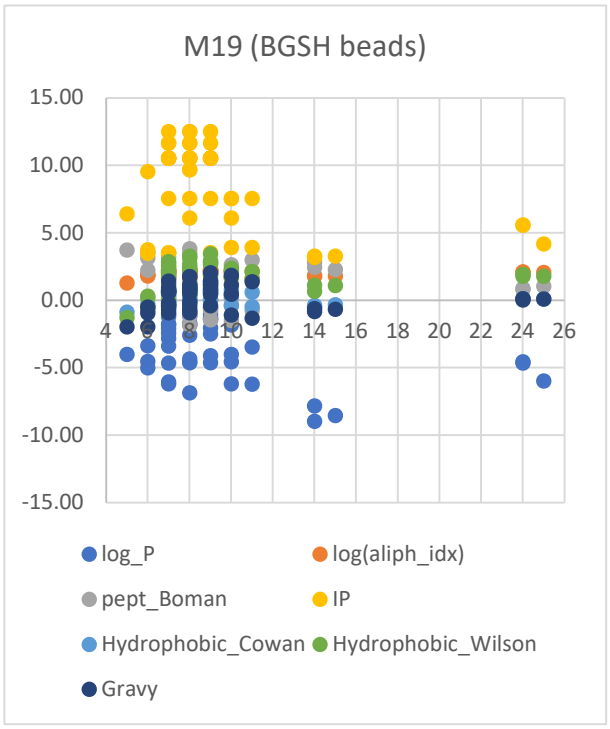


Figure 71. Scatter plot: properties of the peptide with respect to its length

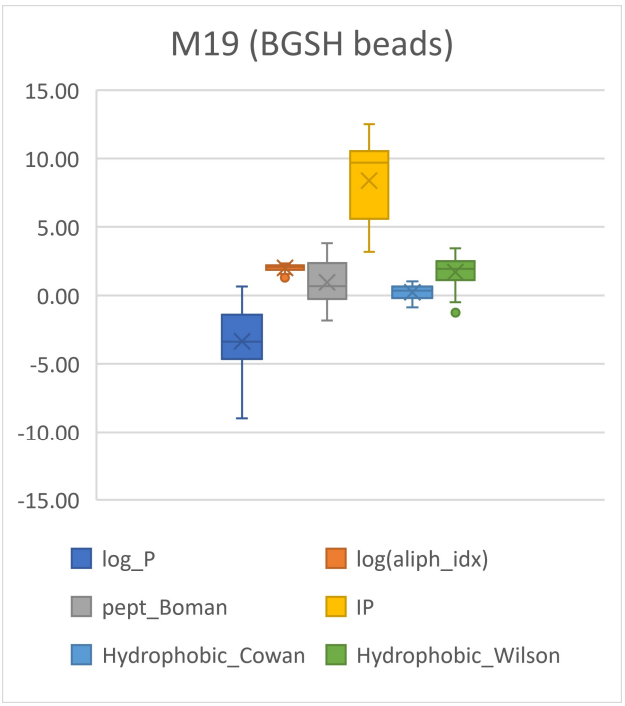


Figure 72. Boxes and whiskers charts of properties of the peptide with respect to its length

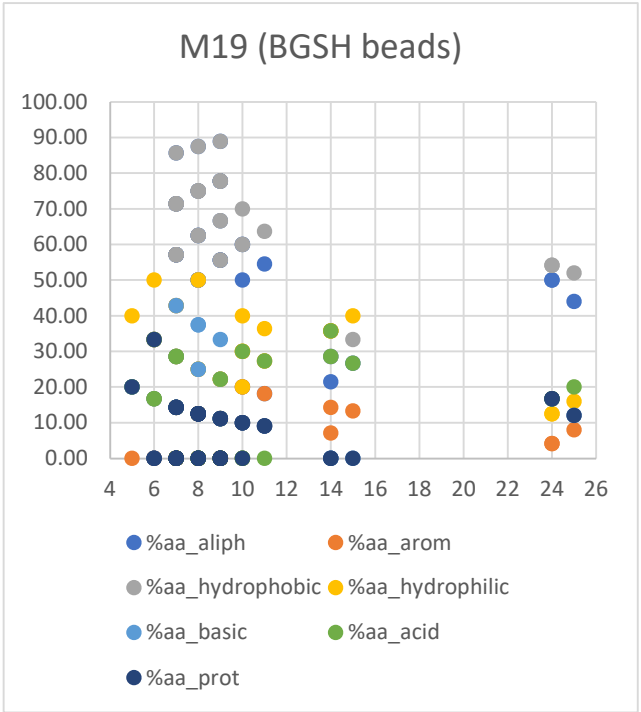


Figure 73. Scatter plot: percentage composition of amino acids classified by properties of the peptide with respect to its length.

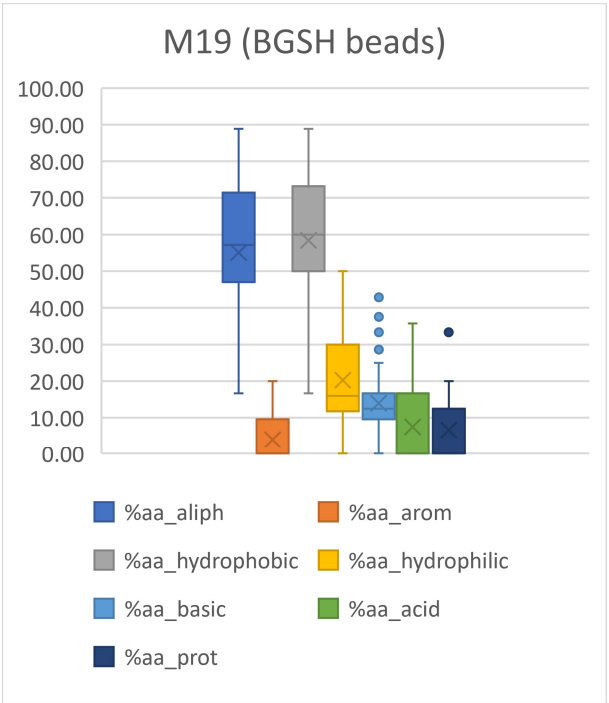


Figure 74. Boxes and whiskers charts of percentage composition of amino acids classified properties of the peptide with respect to its length.

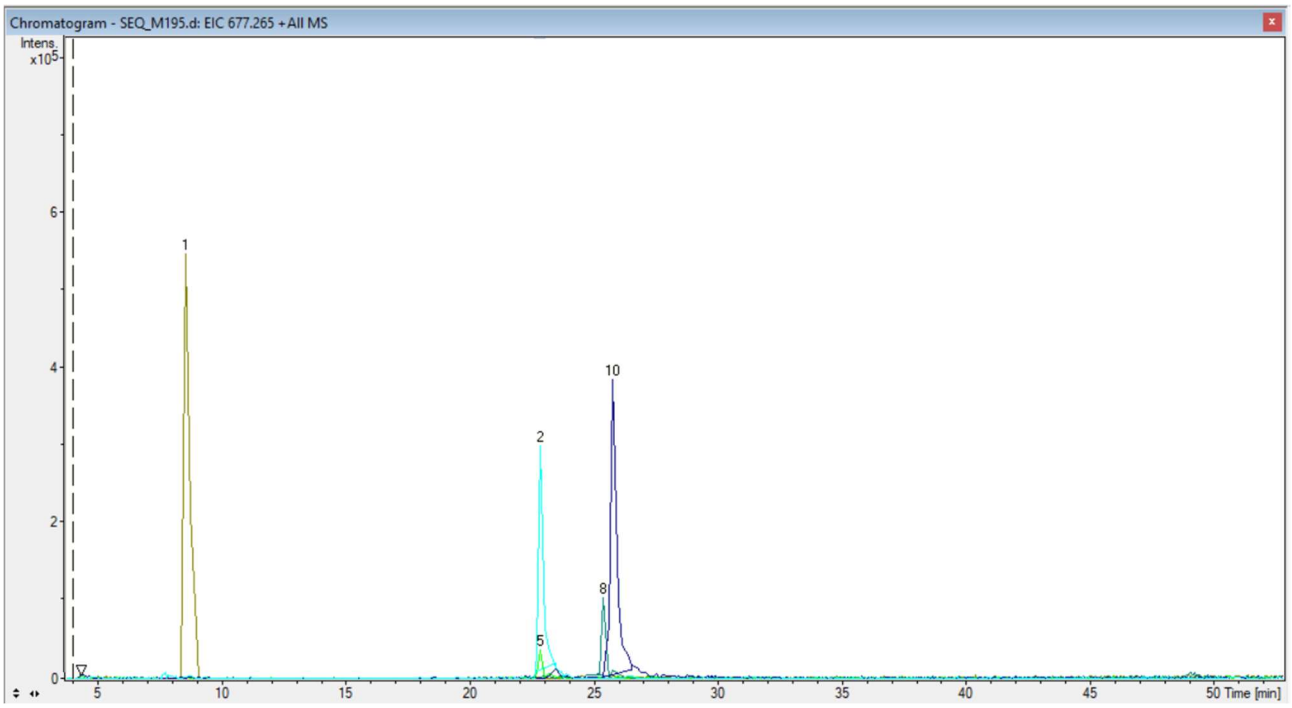


Figure 75 The chromatographic peaks of some ions as they appear in the extracted ion current chromatogram (EIC). EICs have been used to calculate areas and thus recoveries.

peptide_accession	peptide_sequence	m/z	RT [min]	recovery (%)
pept_4_9	MSAPRKVRL	529.3	8.6(1)	30.7
pept_12_9	VMGPRTLIL	500.3	22.8(2)	25.2
pept_19_11	NHEAEDLFYQS	677.3	22.8(5)	2.7
pept_17_10	NGIGVTQNVL	507.8	25.4(8)	17.6
pept_2_11	NITHFAIVASL	593.3	25.8(10)	42.7
pept_5_9	AMAPIKVRL	499.8	25.6	24.3
pept_9_9	VMAPRTLFL	524.3	25.4	29.6
pept_13_9	VMTPRTLVL	515.3	27.3	13.9

Table 19. percentage recoveries of the most intense peaks for BGSH beads.

3.6.2.3 HPLC-MS TIC and spectrum full scan of MIX 19 with BPAB beads

Remembering that this type of beads possesses two carboxyl groups exposed as terminal amino acids of the peptides AB4 and AB5 bound on the surface, also for this type of beads the WAXprotocol was used.

Buffer solution: MES (2-morpholin-4-ylethanesulfonic acid) at pH6

Volume of slurry containing BPAB magnetic beads: 20 µL

Volume of solution standard peptides (MIX19) all at a concentrantion of 1pmol/µL: 40 µL

Contact time between bads and peptides solution: 5 minutes

Desorption Solution: 1 % TFA (incubate for 2 minutes at room temperature before separation and remove the eluate).

Please refer to the materials and methods section about details on the different steps in the operating procedure, the washing solutions and their composition.

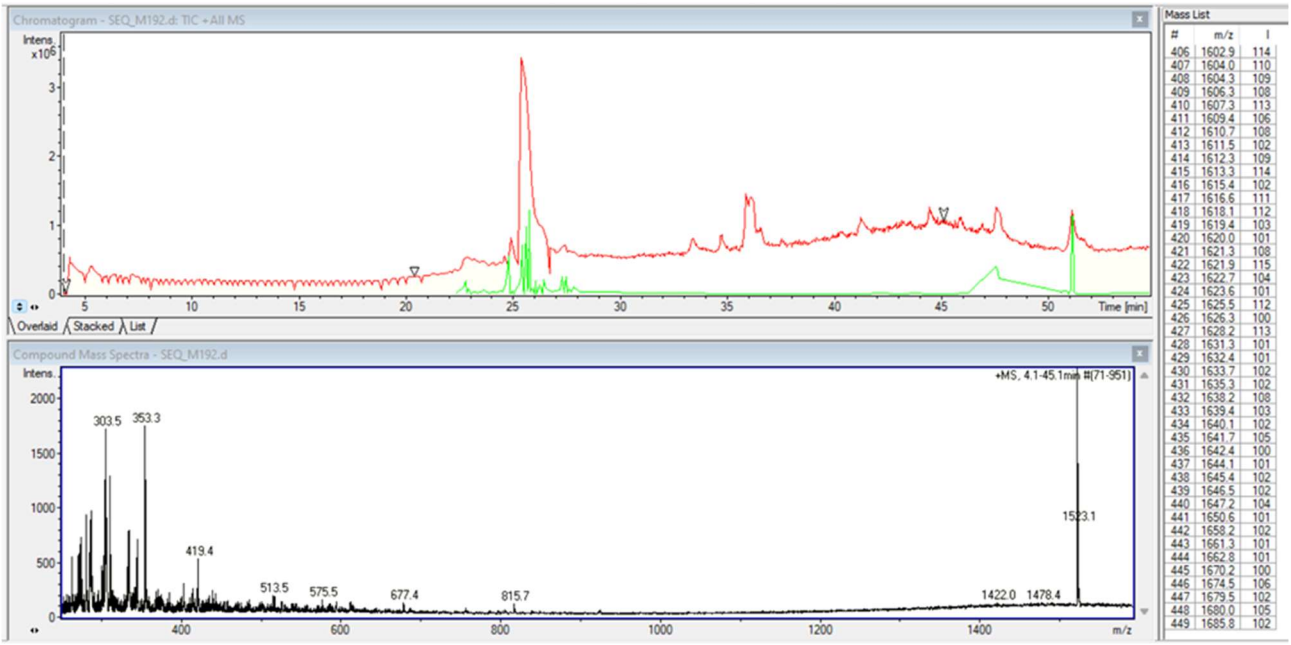


Figure 76. MIX19 TIC and extracted peaks from spectrum level 1 in a range of RT 4.1 – 45.1 min

Range RT	Range index	Number of peaks	Lower mass/higher mass	Lower intensity/higher intensity
4.1 – 45.1	71-951	444	250.4/1680	100/2342

Table 20. Parameters used to obtain data to be analysed with the Code I

Of the 444 spectral peaks, 60 were matched considering only the ionic sequences of the peptide and peaks with values of m/z > 250. The distribution is reported in Table 21.

m/z > 250 (TIC)	$[M+H]^{+}$	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
Whole sequence	0	9	9	6
Fragment (CID)	1	12	19	4
Tot.	1	21	28	10

Table 21. Statistics of the number of ions detected, classified by their charge and belonging to whole or fragmented sequences.

The 60 peaks recognized by the Code I, belong to the 45 unique structures of which 15 in whole sequences and 30 in its fragments.

Statistic diagram of 45 unique peptide structures detected from peaks of spectrum

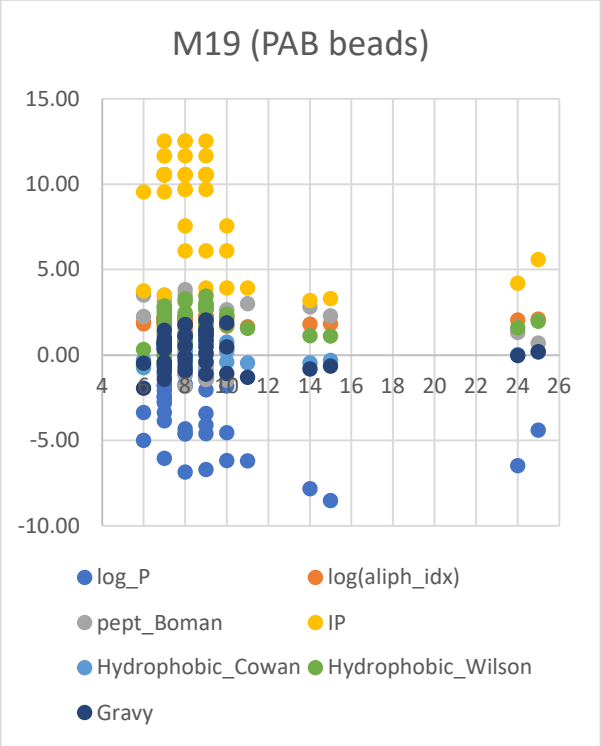


Figure 77. Scatter plot: properties of the peptide with respect to its length.

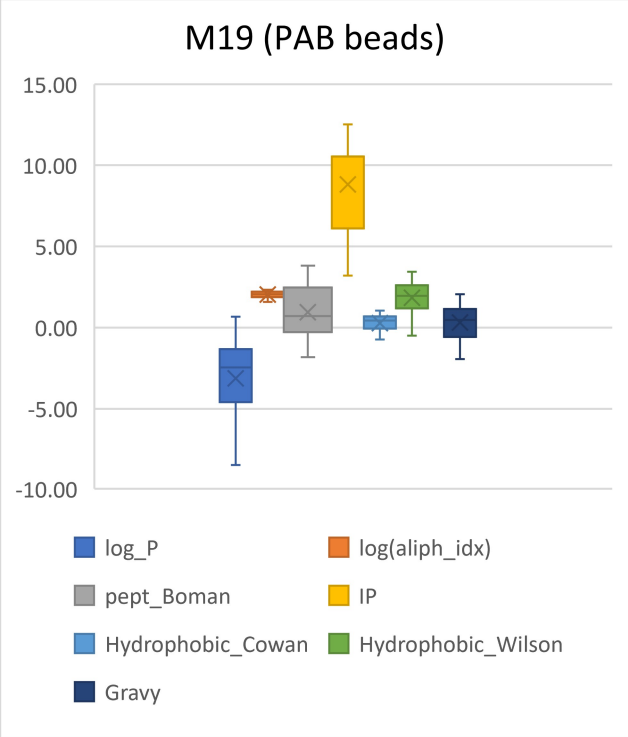


Figure 78. Boxes and whiskers charts of properties of the peptide with respect to its length.

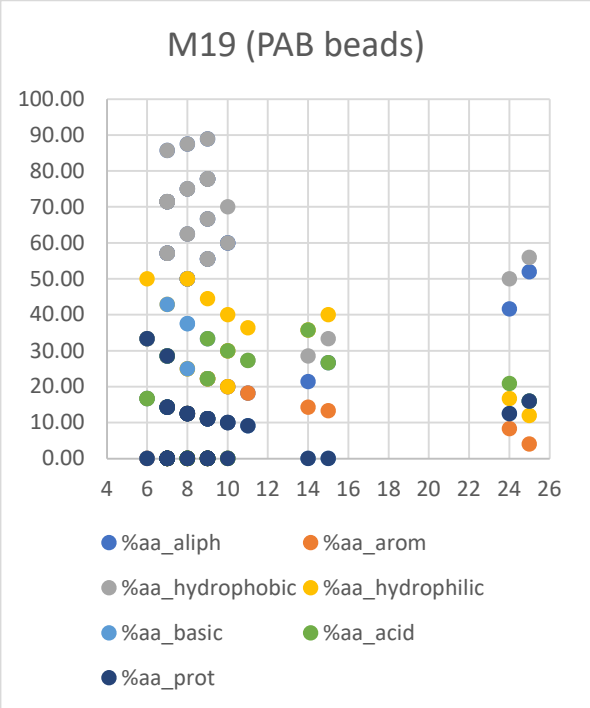


Figure 79. Scatter plot: percentage composition of amino acids classified by properties of the peptide with respect to its length.

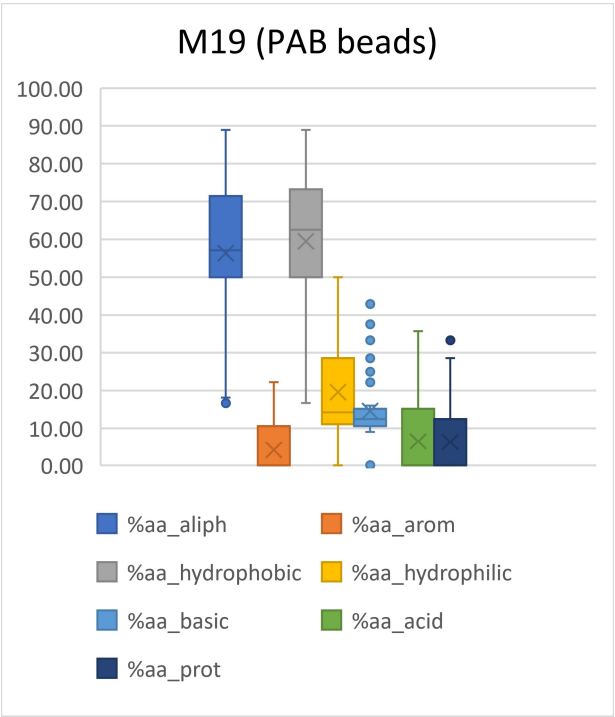


Figure 80. Boxes and whiskers charts of percentage composition of amino acids classified by properties of the peptide with respect to its length.

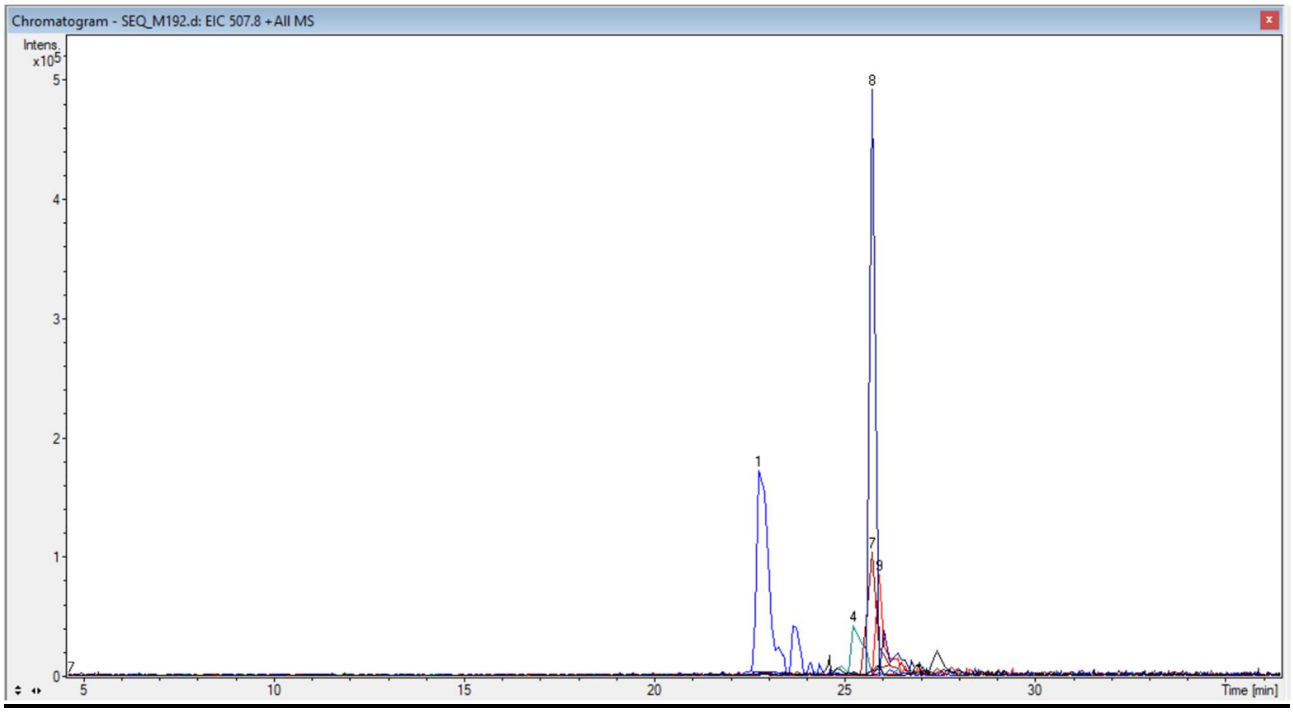


Figure 81. The chromatographic peaks of some ions as they appear in the extracted ion current chromatogram (EIC). EICs have been used to calculate areas and thus recoveries.

peptide_accession	peptide_sequence	m/z	RT [min]	recovery (%)
pept_19_11	NHEAEDLFYQS	677.3	22.7(1)	20.29
frag_pept_13_3	TPRTLVL	799.5	24.9	2.67
pept_17_10	NGIGVTQNVL	507.8	25.2(4)	8.49
pept_5_9	AMAPIKVRL	499.8	25.4	9.23
pept_12_9	VMGPRTLIL	500.3	25.4	1.33
frag_pept_14_5_4	ATKEQ	576.3	25.7(7)	
pept_2_11	NITHFAIVASL	593.3	25.7(8)	19.02
pept_18_16	AQSTIEEQADTFLDYD	923.4	25.9(9)	11.04
pept_9_9	VMAPRTLFL	524.3	26.4	1.90

Table 22. percentage recoveries of the most intense peaks for BPAB beads

3.6.2.4 HPLC-MS TIC and spectrum full scan of MIX 19 with BCOOH beads

Protocol used: Weak Anion eXchange

Buffer solution: MES (2-morpholin-4-ylethanesulfonic acid) at pH6

Volume of slurry containing BCOOH magnetic beads: 20 µL

Volume of solution standard peptides (MIX19) all at a concentrantion of 1pmol/µL: 40 µL

Contact time between beads and peptides solution: 5 minutes

Desorption Solution: 1 % TFA (incubate for 2 minutes at room temperature before separation and remove the eluate).

Please refer to the materials and methods section about details on the different steps in the operating procedure, the washing solutions and their composition.

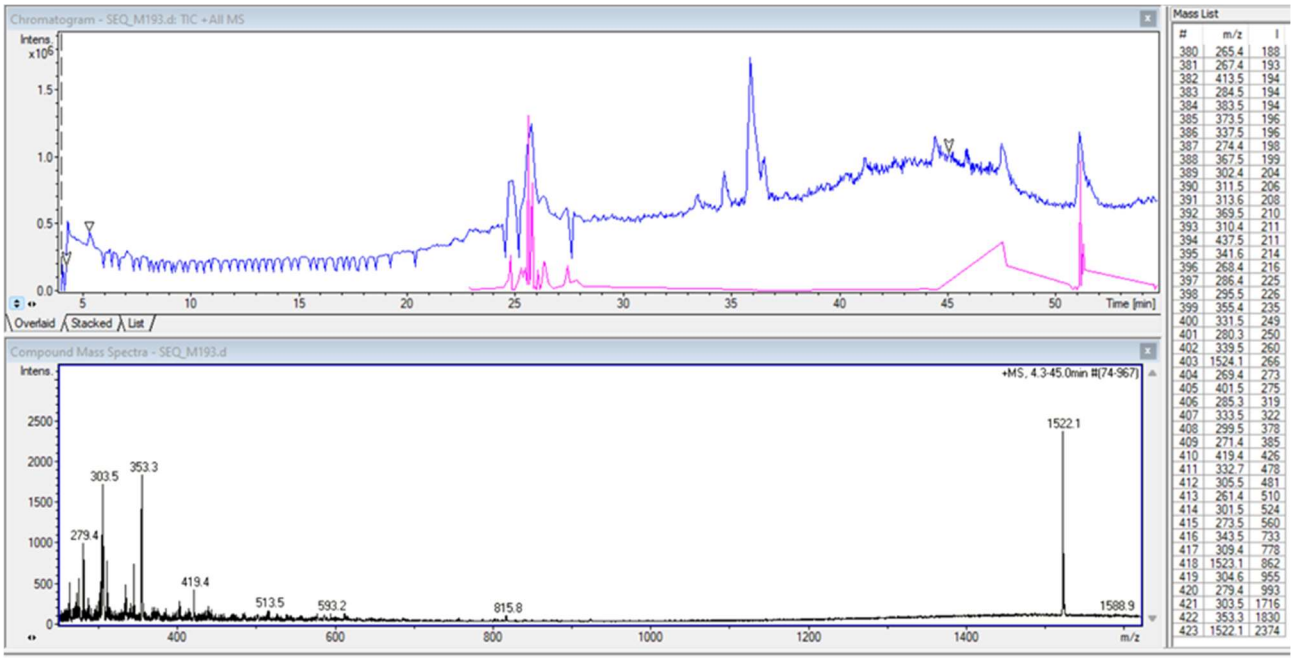


Figure 82. MIX 19 TIC and extracted peaks form spectrum level 1 in a range of RT 4.3 – 45.0 min.

Range RT	Range index	Number of peaks	Lower mass/higher mass	Lower intensity/higher intensity
4.3 – 45.0	74-967	449	250.4/1690	100/2349

Table 23. Parameters used to obtain data to be analysed with the Code I

Of the 449 spectral peaks, 57 were matched considering only the ionic sequences of the peptide and peaks with values of m/z > 250. The distribution is reported in Table 24.

m/z > 250 (TIC)	$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
Whole sequence	0	9	6	6
Fragment (CID)	0	15	18	3
Tot.	0	24	24	9

Table 24. Statistics of the number of ions detected, classified by their charge and belonging to whole or fragmented sequences.

The 57 peaks recognized by the Code I, belong to the 45 unique structures of which 13 in whole sequences and 32 in its fragments.

Statistic diagram of 45 unique peptide structures detected.

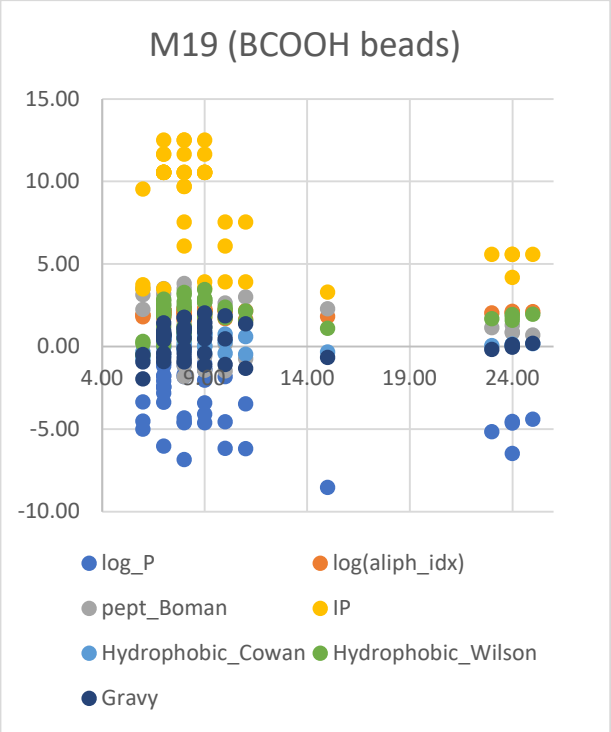


Figure 83. Scatter plot: properties of the peptide with respect to its length

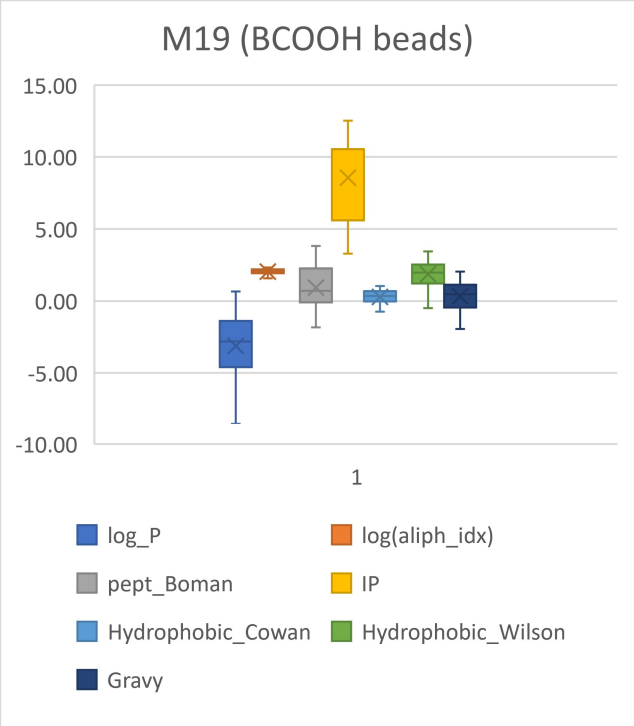


Figure 84. Boxes and whiskers charts of properties of the peptide with respect to its length

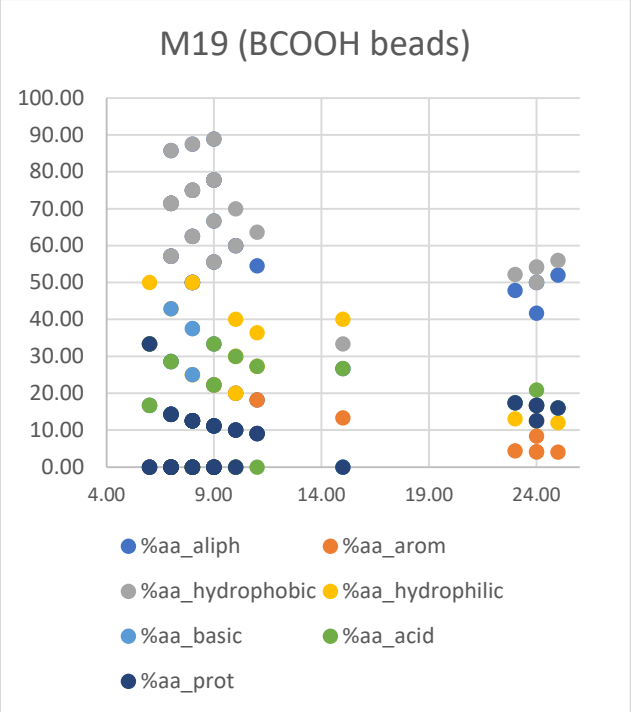


Figure 85. Scatter plot: percentage composition of amino acids classified by properties of the peptide with respect to its length.

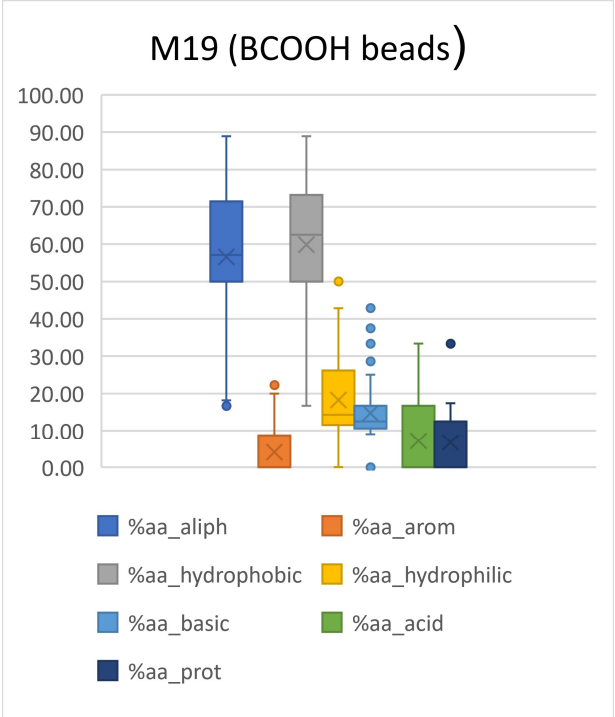


Figure 86. Boxes and whiskers charts of percentage composition of amino acids classified by properties of the peptide with respect to its length.

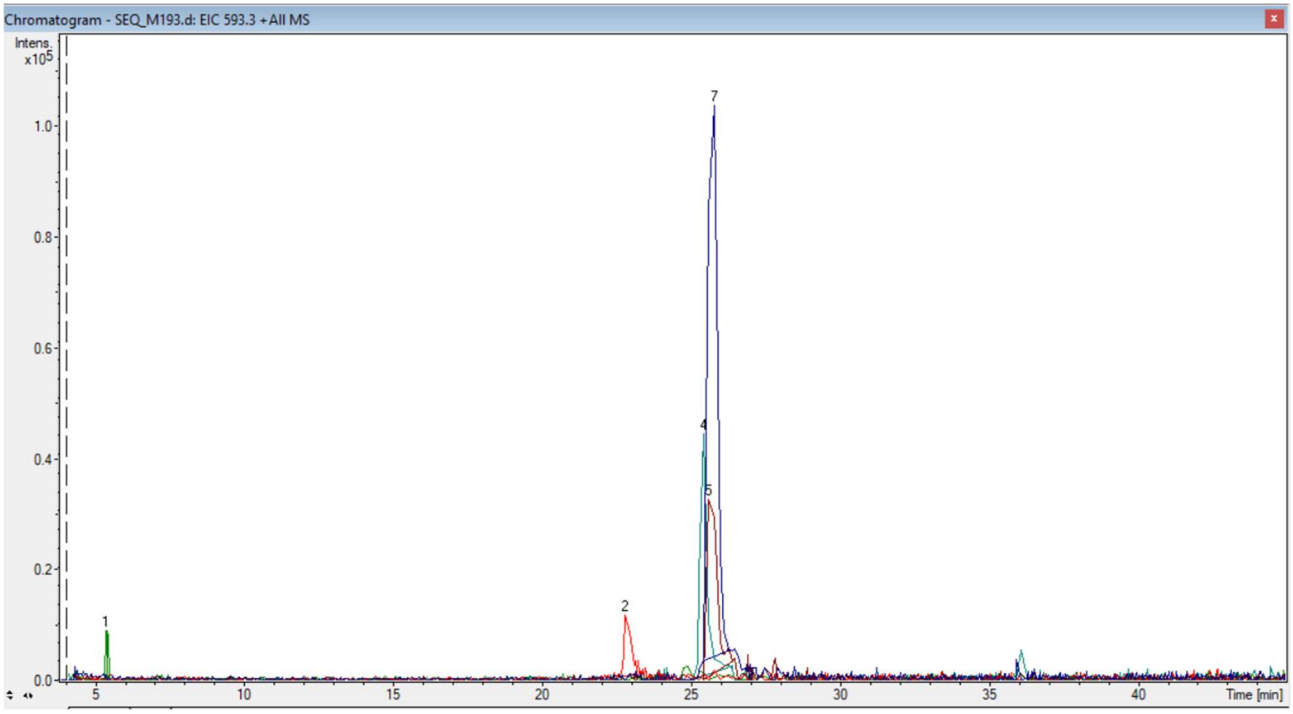


Figure 87. The chromatographic peaks of some ions as they appear in the extracted ion current chromatogram (EIC). EICs have been used to calculate areas and thus recoveries.

peptide_accession	peptide_sequence	m/z	RT [min]	recovery (%)
pept_15_9	FPSHANAAG	436.2	5.4(1)	1.9
pept_19_11	NHEAEDLFYQS	677.3	22.8(2)	4.3
pept_17_10	NGIGVTQNVL	507.8	25.4(4)	10.3
frag_pept_14_5_4	ATKEQ	576.3	25.6(5)	15.1
pept_2_11	NITHFAIVASL	593.3	25.7(7)	18.0

Table 25. percentage recoveries of the most intense peaks for BCOOH beads.

This type of beads, with anionic exchange mechanism, has not been considered efficient in the extraction both from the quantitative point of view and for the small number of extracted peptides compared to those present in the MIX 19 analyzed as such. Note how in the table of charges species detected in the full scan spectrum the mono-charged species are missing.

Analyzing and comparing the various graphs it can be observed that the NH2 beads release a mixture of peptides very similar to that coming from the analysis of the sample without the beads, both from the point of view of the composition and from the point of view of the lengths of the chains while the other beads preferentially release short chain peptides.

Although the properties of the amino acids extracted from the different types of beads are quite similar, for some peptides the recoveries, and therefore the quantitative aspects, are very different.

This will be clearer in the next paragraph where the Venn diagram tool was used for a qualitative analysis on the specificity of the beads to extract certain sequences.

3.6.3 Venn graph and specifity of beads

For a qualitative analysis on the specificity of beads to extract certain sequences, the Venn diagram was used considering, as input, the list of peptides extracted from MIX19 for each type of beads used. With this web tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) was possible to evaluate the eclusivity of the extraction of some peptides by only some types of beads or common sequences extracted from two or more types of beads.

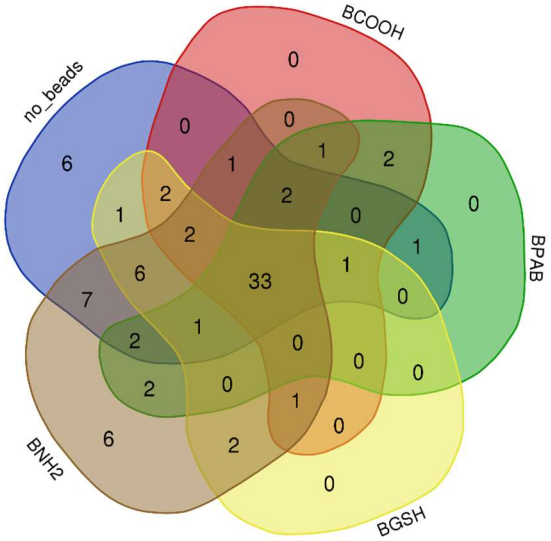


Figure 88. Venn diagram of all peptides extracted from MIX19 with all types of beads and peptides detected analysing MIX 19 without any extraction with beads

Note that 6 sequences were not extracted from any type of beads but it should be remembered that in the list of sequences used by Code I as input to perform the match with the m/z values there are 95 peptide sequences (19 whole sequence and 76 their possible fragments generated in silico). Analyzing the data of the intersections of the different data sets it is clear the three of the four types of pearls are more performing

BCOOH beads are definitely the least efficient. Actually, both BCOOH and BGSH beads cover a small portion of peptide extracts, but for further studies it was preferred to keep BGSH beads for their relatively high percentage of recovery.

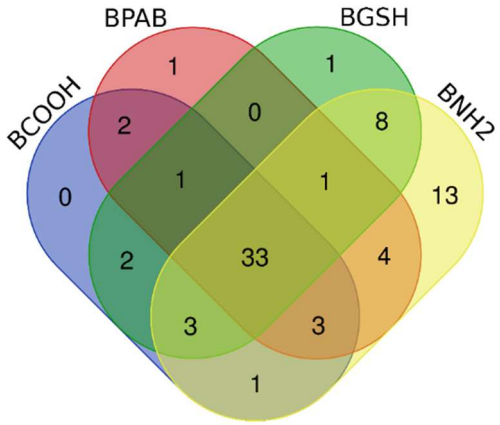


Figure 89. Venn diagram of all peptides extracted from MIX19 with all types of beads.

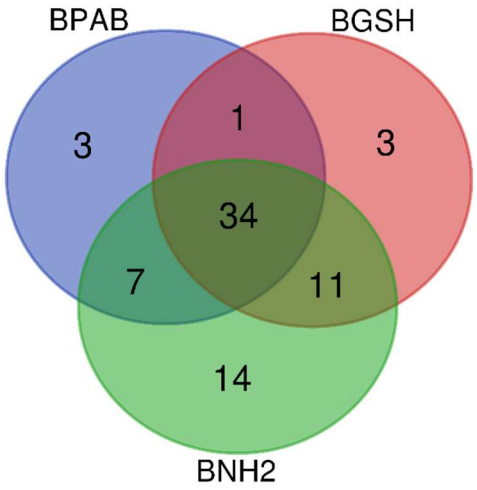


Figure 90. Venn diagram of all peptides extracted from MIX19 with all types of beads excluding BCOOH ones, not used on biological samples.

number of peptides of MIX19 (whole sequence)	19
number of fragments generated in silico from 19 peptides	76
Number of sequences recognised from m/z of spectrum MIX19 no beads	79
Number of sequences recognised from from m/z of spectrum MIX19 with all 4 beads	73
Number of sequences recognised from from m/z of spectrum MIX19 with 3 beads (excluding BCOOH)	73

Table 26.

As you can see from the table 26, even with the exclusion of BCOOH beads the number of recognized sequences does not change compared to the combined use of the other types of beads.

3.7 Experiments on a mixture of 15-mer peptides of protein spike S of SARS-CoV-2 virus

Another peptides mixture solution used to validate the previously developed method with the several types of beads was a sample containing peptide sequences from the spike S protein of Sars-Cov2 virus (MIX S1). This pool derives from the set of three commercial products developed for efficient in vitro stimulation of antigen-specific CD4+ and CD8+ T cells (PepTivator® SARS-CoV-2, Prot_S1, Prot_S+, Prot_S; Miltenyi Biotec). It is a mixture of lyophilized peptides, consisting mainly of 15-mer sequences with 11 amino acids overlap, covering the sequence domain aa 1–895 and the sequence domains aa 304-338, 421-475, 492- 519, 683-707, 741-770, 785-802, and 885 – 1273 (sequence end) of the surface glycoprotein ("S") of SARS-Coronavirus 2 (GenBank MN908947.3, Protein QHD43416.1).

For this mixture of peptides in solution no specification is given on the exact sequences present inside. For this reason, for computational purposes to determine the possible matches between experimental m/z values and m/z values calculated in silico, it was important to realize a small 15-mer sequence data base; the complete sequence being known, the problem has been solved by fragmenting in silico the entire sequence of the spike S protein of Sars-Cov2 virus. In order to create a data base containing the peptide list of 15-mer fragments I proceeded to treat the whole protein sequence available from the Uniprot site, in FASTA format. Using the Python code I wrote, the possible sequences were generated segmenting the protein sequence (SPIKE_SARS2) into sub sequences of 15 amino acid each starting from the amino terminal position with advances of a step-by-step residue thus generating 1260 in silico fragments stored in a table in CSV format. This table contained in the first column the access code with the information of the cutting and start position and in the second the peptide string.
The numbers of generated fragments was L-N+1 , with L the length of protein Spike (1273) and N the length of the fragments (15-mer).
The accession code saved in the table of output of these fragments was encoded in a string composed also by the number of fragment start:

S1_fragm_**number_of_start_fragment**

The numeration of the S1 spike is the same of the FASTA sequence as downloaded from UniProt.

3.7.1 Qualitative and quantitative comparative study of spectra coming from the mixture of peptides of spike protein with the use of beads

The work on the peptide mixture from the spike protein S1 of Sars-Cov2 was conducted following the same scheme as previously reported for MIX 19. The spectrum related to the TIC obtained analysing the mixture solution as such, was compared with those obtained analysing samples previously extracted with different types of beads. All peptides 15-mer of S1 spike had the same concentration in solution (1 pmol/µL). The experimental data obtained analysing MIX S1 as such and after extraction with magnetic beads is reported below. For each spectrum obtained a brief description on data analysis processing is given. For these experiments only the BNH2, BGSH and BPAB beads were used. Considering the high number of peptides obtained in silico for the spike S protein, in addition to the validation of the previously developed method, another goal was to demonstrate the effectiveness of BPAB beads in selectively extracting certain specific peptide sequences through an affinity mechanism, which was also why they were designed and synthesized.

3.7.2.1 HPLC-MS TIC and spectrum full scan of 15-mer peptides of spike S protein of Sars-Cov2 (S1) analysed as such

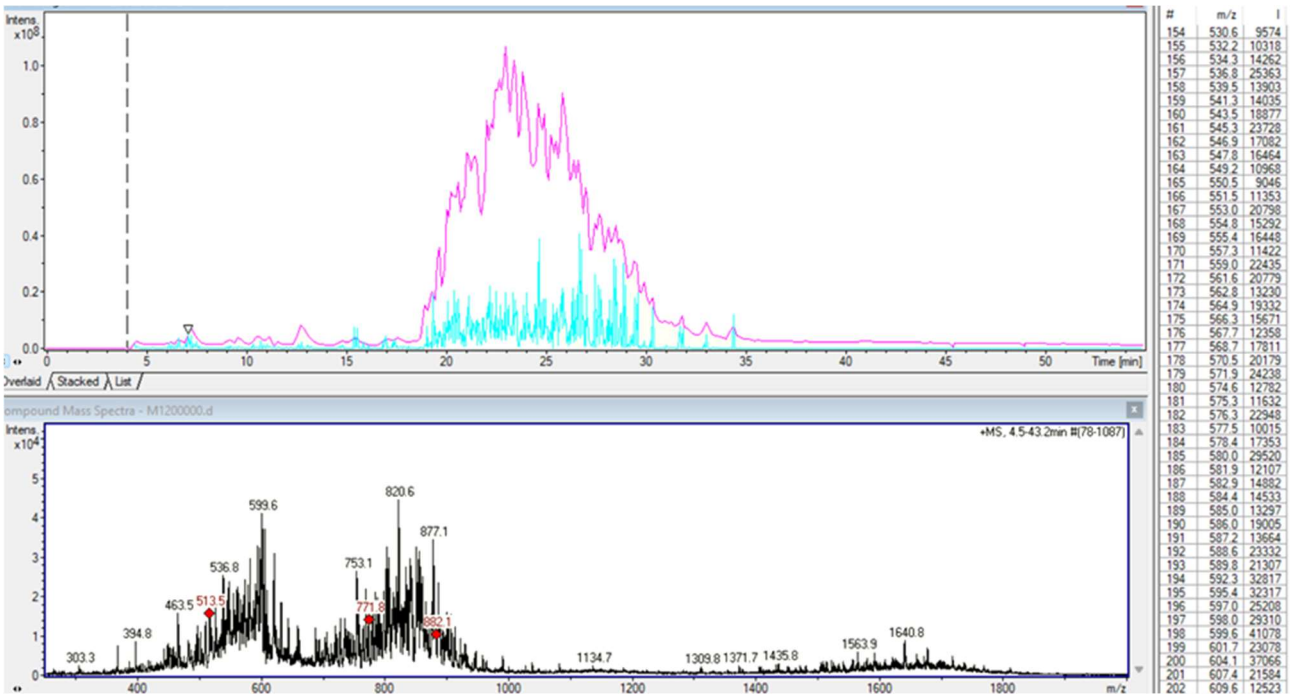


Figure 91 TIC and full scan spectrum of MIX S1 obtained without the use of beads. On the right is the list of extracted signals forms the spectrum MS level 1 in a range of RT 4.5 - 43.2 min.

Range RT (min)	Range index	Number of peaks	Lower mass/higher mass (m/z)	Lower intensity/higher intensity (absolute counts)
4.5 – 43.2	78 - 1078	2485	261.3/1822.9	874/44473

Table 27. Time intervals for which TIC has been analysed and spectra collected. The last three columns contain data on the number and type of m/z signals collected.

The m/z of peaks detected and associated with the sequences are those extracted as precursors ions and then subjected to further fragmentation to obtain the MS/MS spectra.

index	m/z	peptide_sequence	peptide_accession	count	peaks_%
309	771.38	TRAGCLIGAEHVNNS	S1_fragm_15_645	14	40.00
604	902.00	FRVQPTESIVRFPNI	S1_fragm_15_318	13	37.14
285	822.44	RALTGIAVEQDKNTQ	S1_fragm_15_765	12	34.29
465	777.87	KNHTSPDVDLGDISG	S1_fragm_15_1157	12	34.29
98	814.91	HVSGTNGTKRFDNPV	S1_fragm_15_69	11	31.43
166	761.92	LLHAPATVCGPKKST	S1_fragm_15_517	11	31.43
543	808.95	TLDSKTQSLLIVNNA	S1_fragm_15_109	10	28.57
596	718.35	PCSFGGVSVITPGTN	S1_fragm_15_589	10	28.57
881	957.01	YIKWPWYIWLGFIAG	S1_fragm_15_1209	10	28.57
169	787.39	KNFTTAPAICHDGKA	S1_fragm_15_1073	9	25.71
…..	……..	……..	……..	….	…….

Table 28. An extract from output file of the Python code III The index of MS level 2 (first column), the m/z of precursor (second column), the peptide sequence from amino terminus to carboxyl terminus (third column), the code of accession of peptide with length of sequence and starting number of the fragment with respect to the whole chain (fourth column), the m/z values with match (fifth column) and peaks % compared to a fixed number of more intense peaks (35) (sixth column) are reported.

3.7.2.2 HPLC-MS TIC and spectrum full scan of 15-mer peptides of S1 with BNH2 beads

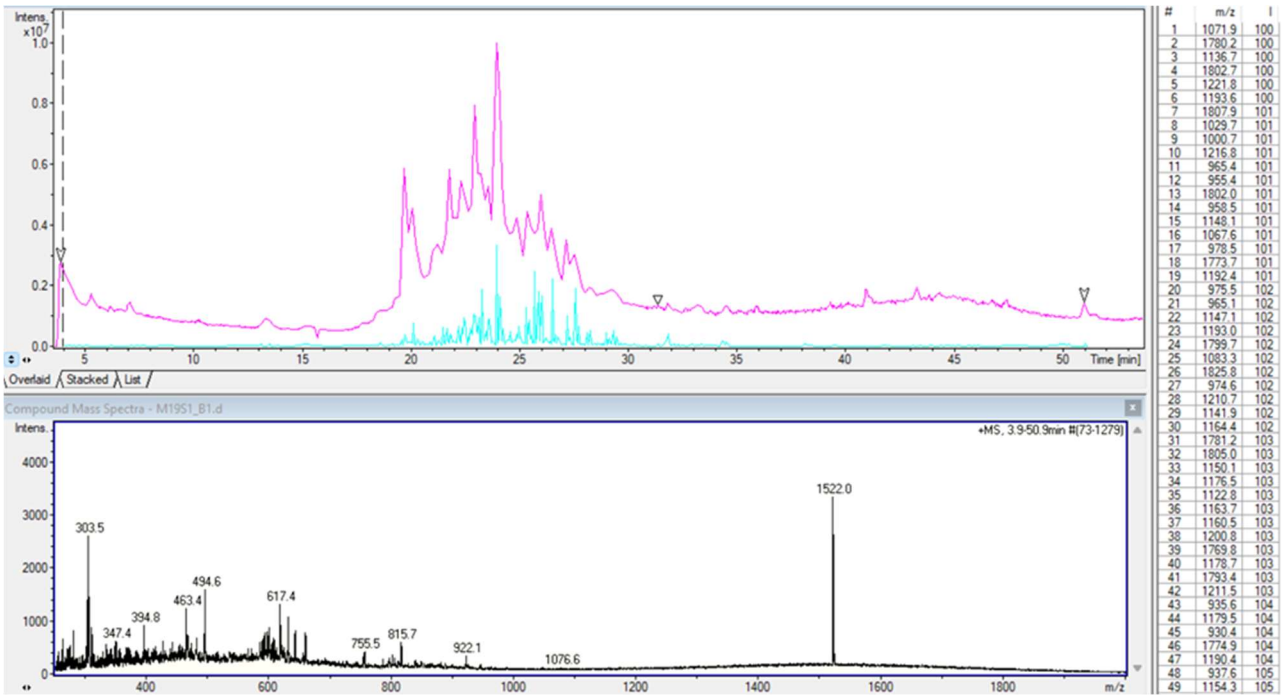


Figure 92. MIX S1 TIC and extracted peaks from spectrum level 1 in a range of 3.9 – 50.9 min.

Range RT (min)	Range index	Number of peaks	Lower mass/higher mass (m/z)	Lower intensity/higher intensity (absolute counts)
3.9 – 50.9	73 - 1279	1229	250.3/ 1825.8	100/3333

Table 29 Time intervals for which TIC has been analysed and spectra collected. The last three columns contain data on the number and type of m/z signals collected.

index	m/z	peptide_sequence	peptide_accession	count	peaks_%
668	815.938	DVVIGIVNNTVYDPL	S1_fragm_15_1127	14	40
637	795.883	GWTAGAAAYYVGYLQ	S1_fragm_15_257	13	37.14
756	920.9565	DLFLPFFSNVTWFHA	S1_fragm_15_53	11	31.43
582	795.3794	IAYTMSLGAENSVAY	S1_fragm_15_693	9	25.71
584	847.9063	NQVAVLYQDVNCTEV	S1_fragm_15_606	8	22.86
625	740.8752	TITSGWTFGAGAALQ	S1_fragm_15_881	8	22.86
692	810.4482	AIPTNFTISVTTEIL	S1_fragm_15_713	8	22.86
694	876.9332	HSTQDLFLPFFSNVT	S1_fragm_15_49	8	22.86
701	753.8906	GWTFGAGAALQIPFA	S1_fragm_15_885	8	22.86
542	801.4465	SKTQSLLIVNNATNV	S1_fragm_15_112	7	20
598	864.4412	LTPTWRVYSTGSNVF	S1_fragm_15_629	7	20
……	……	……	…….	….	…..

Table 30. An extract from output file of the Python code III. The index of MS level 2 (first column), the m/z of precursor (second column), the peptide sequence (third column), the code of accession of peptide (fourth column), the m/z values with match (fifth column) and peaks % compared to a fixed number of more intense peaks (sixth column) are reported.

3.7.2.3 HPLC-MS TIC and spectrum full scan of 15-mer peptides of S1 with BGSH beads

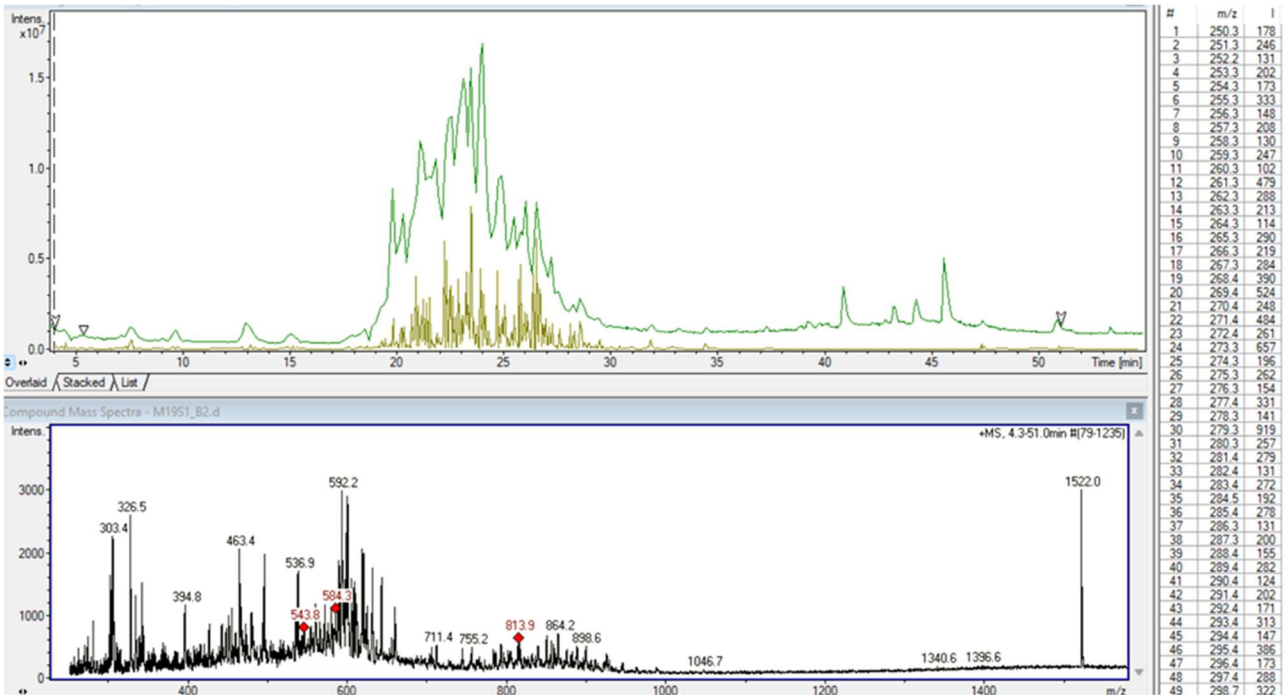


Figure 93. MIX S1 TIC and extracted peaks from spectrum level 1 in a range of RT 4.3 – 51.0 min.

Range RT (min)	Range index	Number of peaks	Lower mass/higher mass (m/z)	Lower intensity/higher intensity (absolute counts)
4.3 – 51.0	79 - 1235	1359	250.3/ 1836.1	100/3000

Table 31. Time intervals for which TIC has been analysed and spectra collected. The last three columns contain data on the number and type of m/z signals collected.

index	prec	peptide_sequence	peptide_accession	count	peaks_%
625	815.938	DVVIGIVNNTVYDPL	S1_fragm_15_1127	12	34.29
656	753.8906	GWTFGAGAALQIPFA	S1_fragm_15_885	11	31.43
667	782.4533	SALEPLVDLPIGINI	S1_fragm_15_221	11	31.43
673	830.4368	FNGLTVLPPLLTDEM	S1_fragm_15_855	11	31.43
585	894.9465	PFFSNVTWFHAIHVS	S1_fragm_15_57	10	28.57
721	920.9565	DLFLPFFSNVTWFHA	S1_fragm_15_53	10	28.57
629	932.0118	YQPYRVVVLSFELLH	S1_fragm_15_505	9	25.71
542	838.9252	GAAAYYVGYLQPRTF	S1_fragm_15_261	8	22.86
648	810.4482	AIPTNFTISVTTEIL	S1_fragm_15_713	8	22.86
558	839.9492	ASTEKSNIIRGWIFG	S1_fragm_15_93	7	20
648	810.4482	IAIPTNFTISVTTEI	S1_fragm_15_712	7	20

Table: 32. An extract from output file of the Python code III. The index of MS level 2 (first column), the m/z of precursor (second column), the peptide sequence (third column), the code of accession of peptide (fourth column), the m/z values with match (fifth column) and peaks % compared to a fixed number of more intense peaks (sixth column) are reported.

3.7.2.4 HPLC-MS TIC and spectrum full scan of 15-mer peptides of S1 with BPABbeads

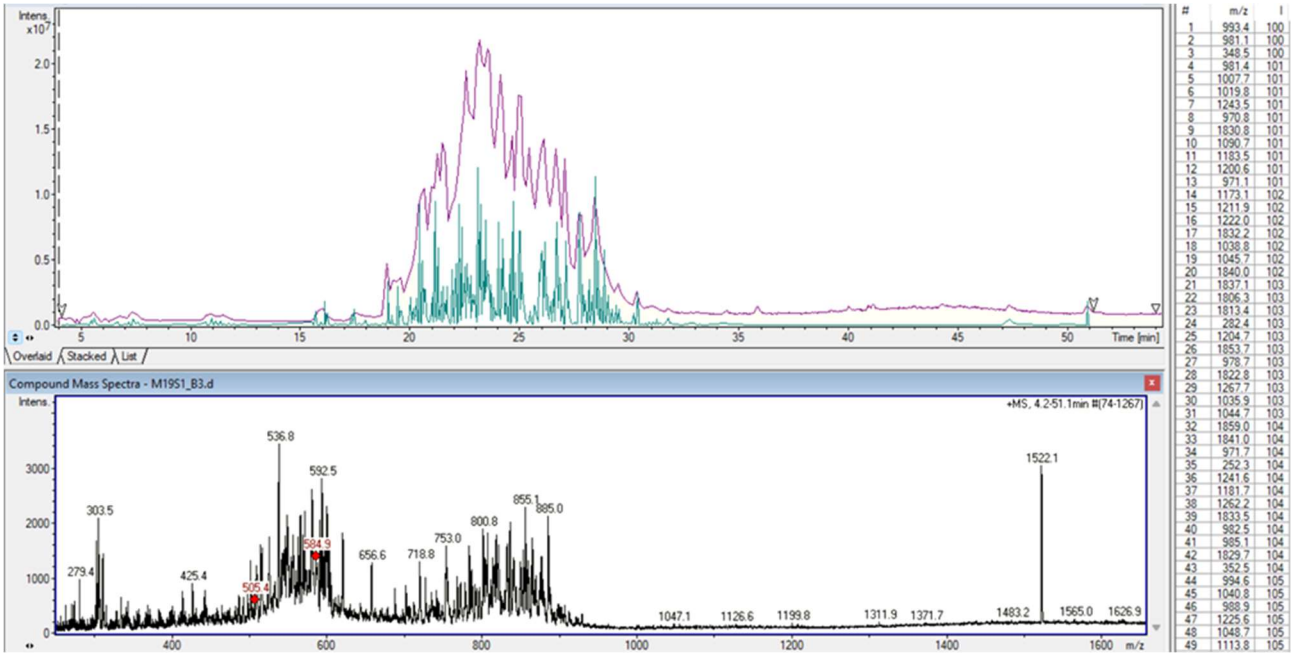


Figure 94 TIC and MS/MS spectrum of the MIX S1 after use of BPAB beads with list of extracted masses.

Range RT (min)	Range index	Number of peaks	Lower mass/higher mass (m/z)	Lower intensity/higher intensity (absolute counts)
4.2 – 51.1	74 - 1267	1499	250.3/ 1859	100/3451

Table 33. Time intervals for which TIC has been analysed and spectra collected. The last three columns contain data on the number and type of m/z signals collected.

index	prec	peptide_sequence	peptide_accession	count	peaks_%
331	771.607	TRAGCLIGAEHVNNS	S1_fragm_15_645	23	46
538	902.106	FRVQPTESIVRFPNI	S1_fragm_15_318	23	46
334	771.617	TRAGCLIGAEHVNNS	S1_fragm_15_645	22	44
207	762.031	LLHAPATVCGPKKST	S1_fragm_15_517	21	42
303	822.663	RALTGIAVEQDKNTQ	S1_fragm_15_765	21	42
672	800.271	LVLLPLVSSQCVNLT	S1_fragm_15_5	21	42
126	841.483	GICASYQTQTNSPRR	S1_fragm_15_669	20	40
347	771.677	TRAGCLIGAEHVNNS	S1_fragm_15_645	20	40
610	726.313	LLAGTITSGWTFGAG	S1_fragm_15_877	20	40
628	906.755	FNCYFPLQSYGFQPT	S1_fragm_15_486	20	40
215	787.693	KNFTTAPAICHDGKA	S1_fragm_15_1073	19	38
458	832.053	LVSSQCVNLTTRTQL	S1_fragm_15_10	19	38
460	863.145	FQPTNGVGYQPYRVV	S1_fragm_15_497	19	38
634	906.739	FNCYFPLQSYGFQPT	S1_fragm_15_486	19	38
658	810.667	AIPTNFTISVTTEIL	S1_fragm_15_713	19	38
204	762.159	LLHAPATVCGPKKST	S1_fragm_15_517	18	36
612	754.93	LDITPCSFGGVSVIT	S1_fragm_15_585	18	36
466	805.948	LGKLQDVVNQNAQAL	S1_fragm_15_945	17	34
533	887.223	YRFNGIGVTQNVLYE	S1_fragm_15_904	17	34
588	885.684	NLLLQYGSFCTQLNR	S1_fragm_15_751	17	34

Table 34: An extract from output file of the Python code III. The index of MS level 2 (first column), the m/z of precursor (second column), the peptide sequence (third column), the code of accession of peptide (fourth column), the m/z values with match (fifth column) and peaks % compared to a fixed number of more intense peaks (sixth column) are reported. The red rows report the detected sequences belonging to the protein binding zone for which the beads BPAB have been designed.

The quantitative aspect was also evaluated by analysing the recoveries of sequences selectively extracted from BPAB beads and demonstrating their effective binding in the expected sequence range. The recovery percentage is the ratio of the peak area extracted from a certain type of beads to the peak area of the corresponding value of m/z, obtained analysing MIX S1 as such.

In order to make a comparison between peaks areas the same values of m/z at the same retention times were considered.

S1 range sequence	Peptide sequence	m/z	RT [min]	Rec% S1_BNH2	Rec % S1_BGSH	Rec % S1_BPAB
486-501	FNCYFPLQSYGFQPT	906.4	28	0.26	0.85	19.39
497-512	FQPTNGVGYQPYRVV	862.94	22.3	0	6.37	52.05
480-495	CNGVEGFNCYFPLQS	839.35	23.3	1.9	9.65	30.73
473-488	YQAGSTPCNGVEGFN	772.32	24.1	2.07	3.5	44.98
485-500	GFNCYFPLQSYGFQP	884.39	20.5	0.59	2.07	19.38

Table 35. Comparison of recovery of specific sequences in the range 470-500 for protein S1 spike fragments, where PAB beads should act by affinity.

It is evident that beads designed to work with an affinity mechanism, allow much higher recoveries than other beads used. In the case of tissue studies, this result could be exploited to build beads useful for the selective extraction of any biomarkers once they had been identified using massive approaches such as those described above.

The affinity beads (in this case the PBAB beads) can drive the correct assignment of m/z values of the sequences for which they are specific; being able to obtain more significant values in the ratio of the EIC areas for the calculation of recoveries.

For example, the m/z ratio 772.32 corresponding to the 15_473 fragment may also derive from other sequences in the range of 0.1 Da, but as can be seen from the extracted ion current chromatograms (EIC) , the current relative to the m/z ratio that was expected to be extracted from the affinity beads (RT=24.1 min) has a much more intense area than those present at other retention times (figure 95).

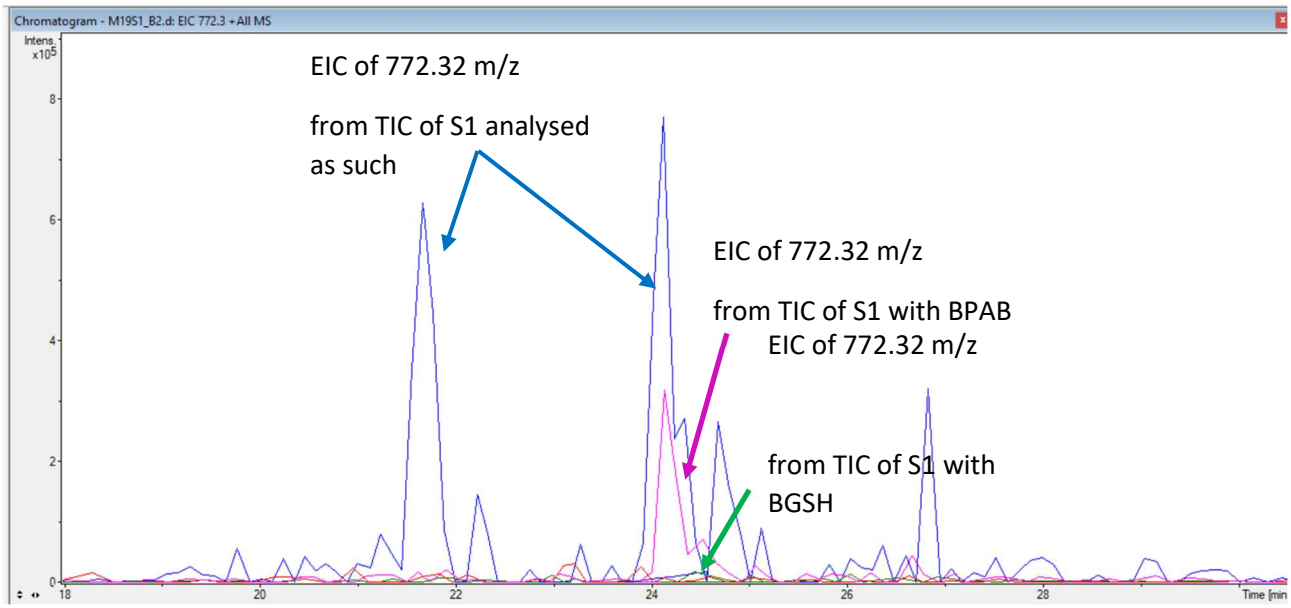


Figure 95 Overlap of the extracted ion current chromatograms of m/z 772.32 for different samples analysed (blue: MIX S1 as such; purple: MIX S1 extracted with BPAB beads; green: MIX S1 extracted with BGSH beads).

The specificity of action of BPAB beads is confirmed when we compare the recovery percentage, by different beads, of random peptides whose m/z values don't belong to the sequences for which the PAB beads show affinity.

The recovery percentages between the various beads are similar and those of the BPAB beads are recoveries lower respect to the recoveries of the corresponding peptides (m/z) for which they are specific for affinity (table 36).

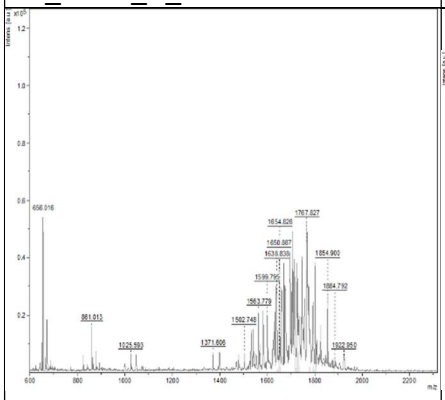
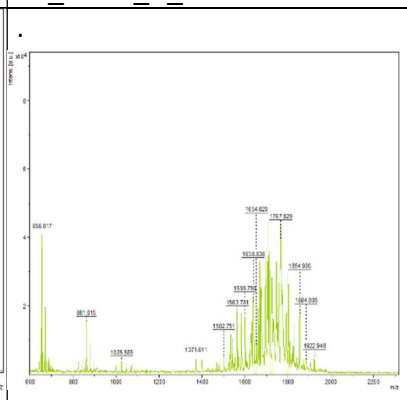
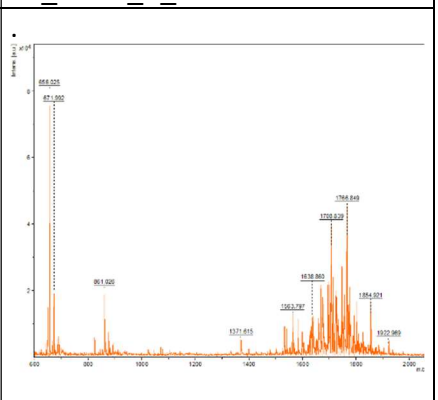
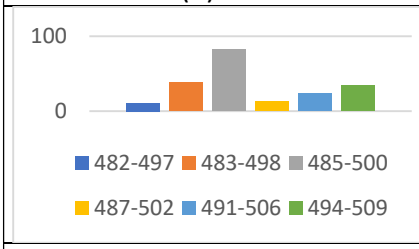
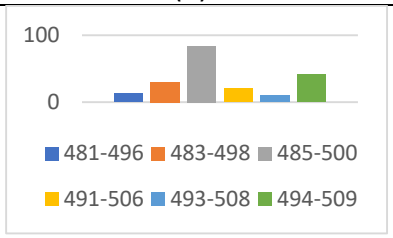
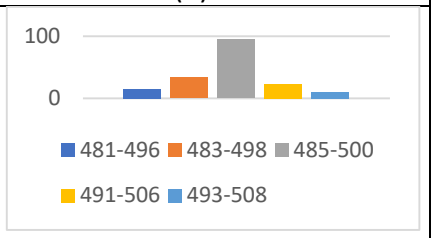
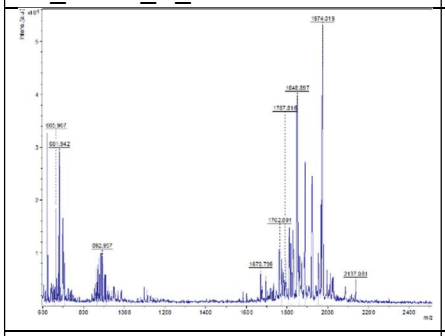
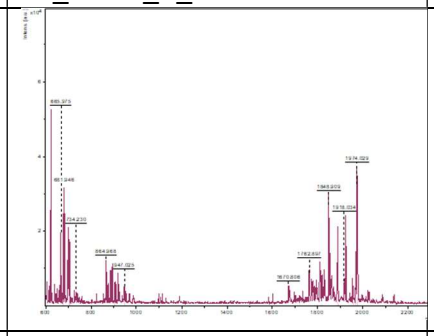
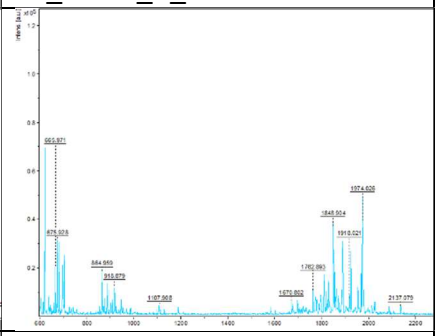
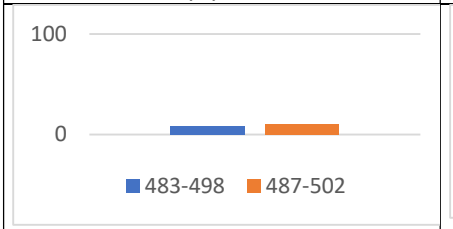
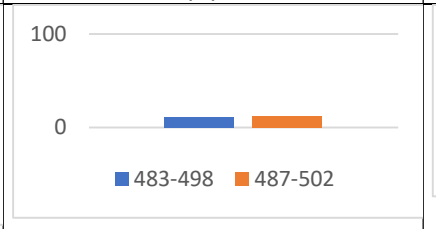
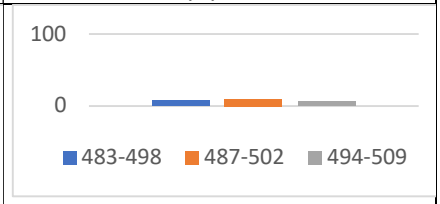
peak_m/z	RT (min)	Rec % S1_BNH2	Rec % S1_BGSH	Rec % S1_BPAB
822.44	15.5	5.4	15.5	22.4
902	24.6	12.2	3.1	9.5
957	33	6.4	17.9	8.15
787.39	22.2	16.8	10.9	13.15
761.92	20.8	9.6	4.7	5.0
592.6	23.4	14.94	26.3	33.5

Table 36. Comparison of recoveries data for signals with the same retention times and m/z, but belonging to random sequences, present outside the S1 PAB beads binding zone.

3.7.3 Further experimental evidence of the specificity of PAB beads by analysing data from MALDI-TOF technique.

MIX S1 extracted with various magnetic beads was also analysed with MALDI-TOF mass spectrometry using an UltrafleXtreme™ instrument (Bruker Daltonics Inc., Milan, Italy). The sample preparation, up to its analysis, was the same developed and described above for ESI-MS ion trap analysis. The elutes obtained from the magnetic beads' extraction process were analysed "off-line". Prior to the analysis, samples were mixed in a ratio of 1:2 with the matrix, deposited and crystallized on the target plate. This procedure, as well as the analysis itself, were described earlier in the chapter "Materials and methods".

Such a wide solution of peptides required that the processed data be analysed with the use of the Python code I developed, combined with a particular statistical web-based tool developed for bioinformatics (https://bioinformatics.psb.ugent.be/webtools/Venn/). Through the use of the Python code a match has been obtained between the values of m/z of the peaks of the experimental spectrum MALDI/TOF and the monocharged masses $[M+H]^+$ calculated for the peptides 15-mer present in the standard solution, considering a threshold range centred on 0.05 Da between experimental m/z value to calculated m/z value. The obtained results are summarized in Figure 96.

S1_BPAB_A_15	S1_BPAB_B_15	S1_BPAB_C_15
		
MALDI/TOF S1 spectrum with BPAB beads (1)	MALDI/TOF S1 spectrum with BPAB beads (2)	MALDI/TOF S1 spectrum with BPAB beads (3)
		
Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB
S1_BNH2_A_15	S1_BNH2_B_15	S1_BNH2_C_15
		
MALDI/TOF S1 spectrum with BNH2 beads (1)	MALDI/TOF S1 spectrum with BNH2 beads (2)	MALDI/TOF S1 spectrum with BNH2 beads (3)
		
Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB

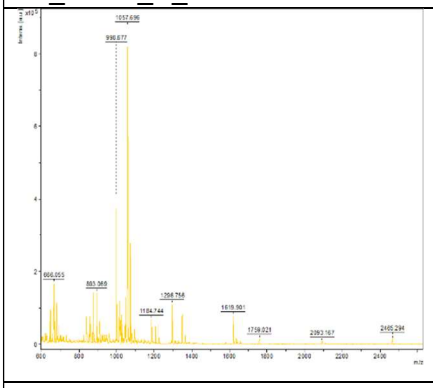
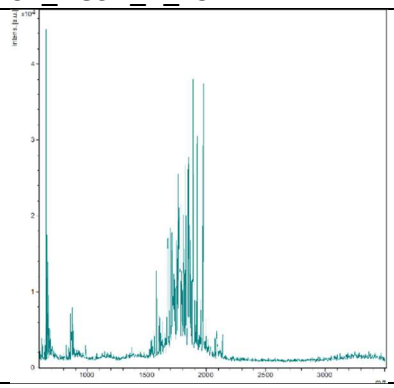
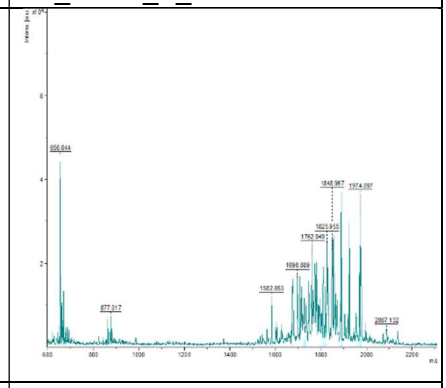
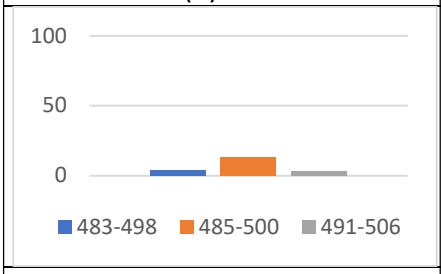
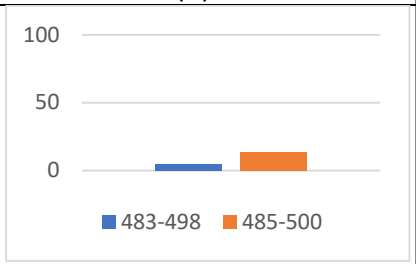
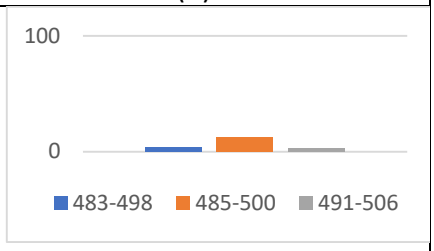
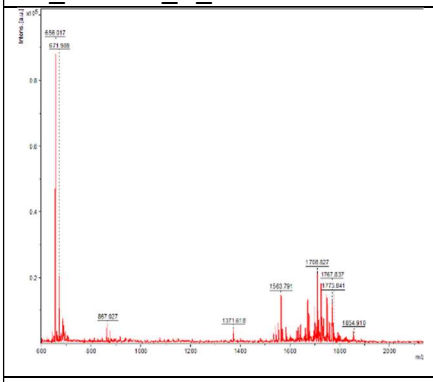
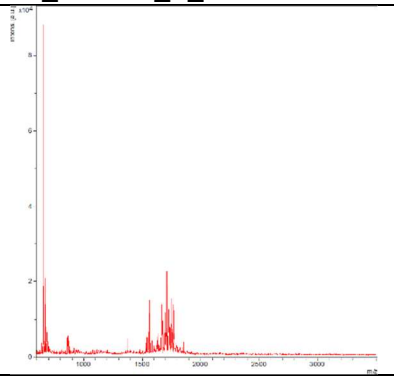
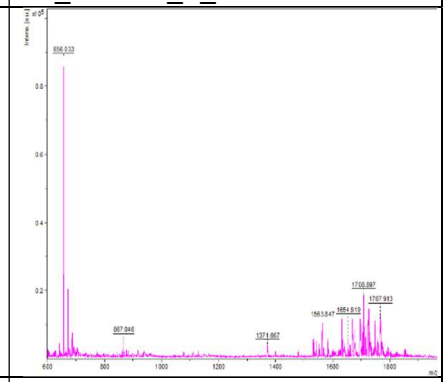
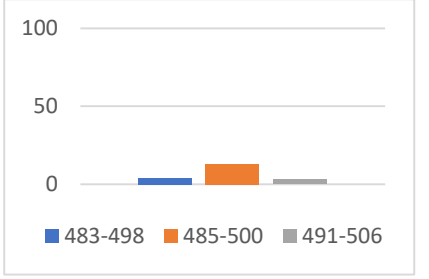
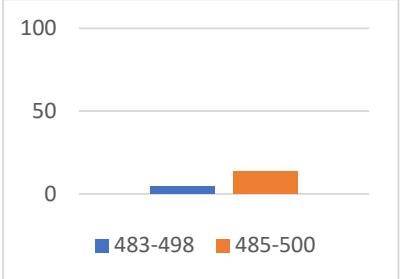
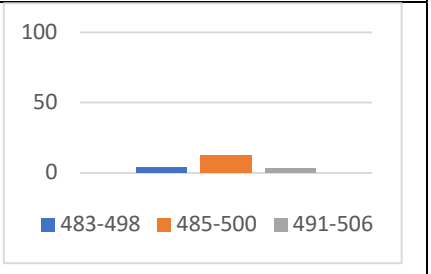
S1_BGSH_A_15	S1_BGSH_B_15	S1_BGSH_C_15
		
MALDI/TOF S1 spectrum with BGSH beads (1)	MALDI/TOF S1 spectrum with BGSH beads (2)	MALDI/TOF S1 spectrum with BGSH beads (3)
		
Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB
S1_BCOOH_A_15	S1_BCOOH_B_15	S1_BCOOH_C_15
		
MALDI/TOF S1 spectrum with BCOOH beads (1)	MALDI/TOF S1 spectrum with BCOOH beads (2)	MALDI/TOF S1 spectrum with BCOOH beads (3)
		
Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB	Relative abundance percentage of m/z peaks in the range of binding selective of BPAB

Figure 96. The raw spectra and data obtained from analysis with Python code of three replicates of the same sample for each type of beads tested are reported here. The range of the peptide sequence (15-mer) corresponding to the measured value of m/z is shown on the x-axes while on the y-axes the corresponding relative intensity of the peak is reported. The relative percentage intensity was calculated by reference to the base peak of the spectrum.

In the diagrams reported in Figure 97, each circle represents a set of a different category of data (strings that represent amino acid sequences or m/z values). The overlapping parts of the circles show where those categories of data have something in common. The non-overlapping parts represent the exclusive data of that category.

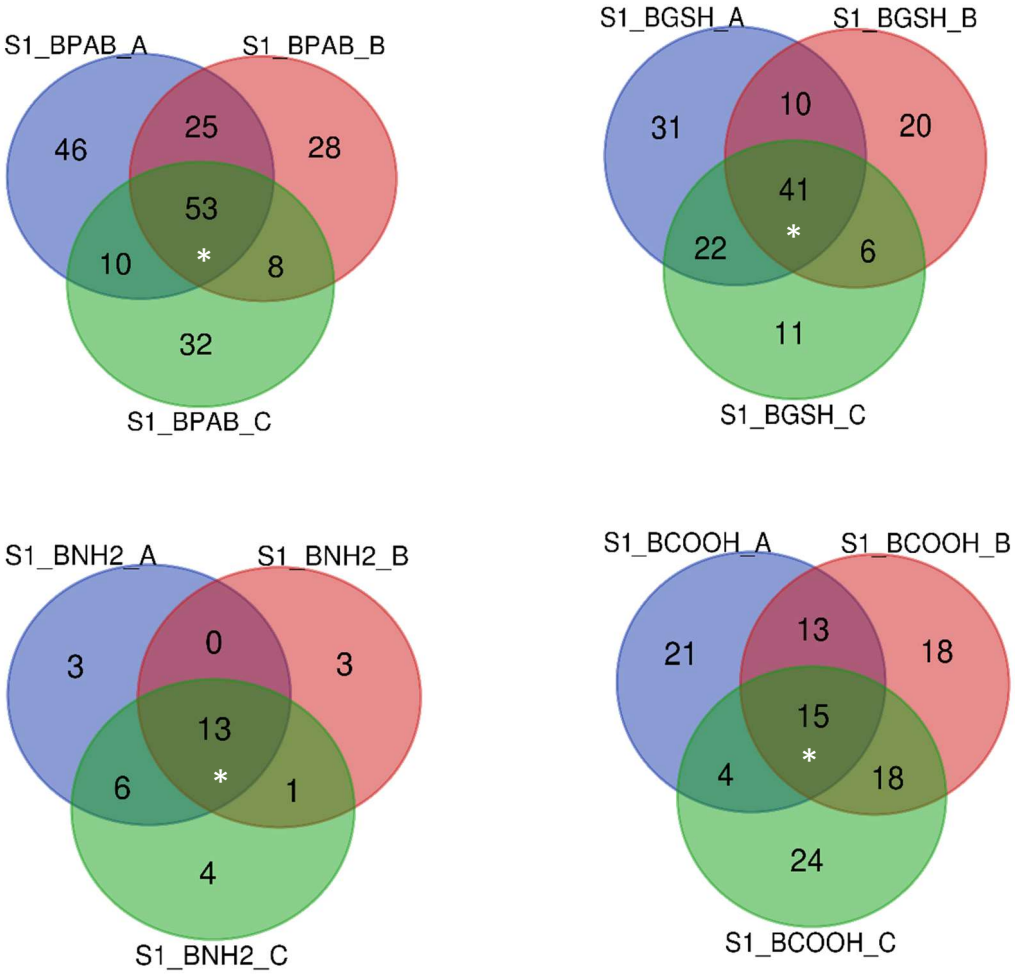


Figure 97. Venn diagram of peptides extracted from MIX S1. The match results between m/z data from MALDI/TOF peak spectrum and sequence library present in the standard peptide mixture are reported. The data in the circles refer to matches. All analyses were carried out in triplicate.

The central intersection (*) of the data represents the common sequences for the same type of beads (figure97). This data were used to perform another statistical test (Figure 98) that allowed to

show the main similarities/ differences between the type of peptides extracted from different types of magnetic beads. This has demonstrated, once again, the discriminating capacity of the various types of pearls and the need to use them all to obtain results that include as many analytes as possible.

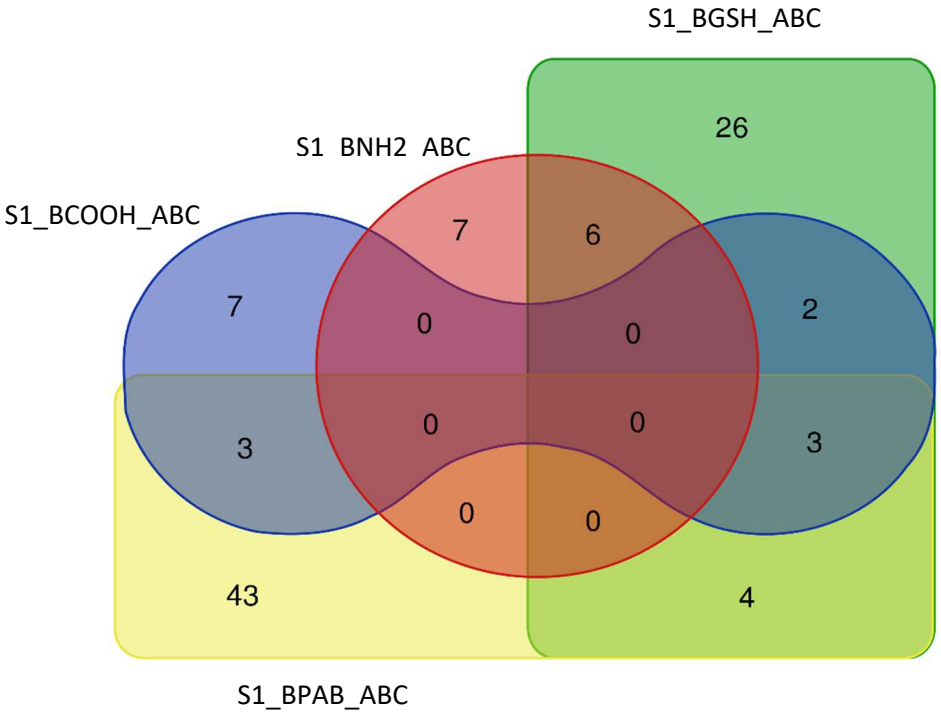


Figure 98. Venn diagram obtained analysing the central intersections of Figure 97.

As shown in Figure 98, in the central part of the intersection, no sequence is common to all types of beads, meaning that the several types of beads are selective towards peptide sequences. It is possible to deduce that the two types of beads with a high exclusivity towards certain sequences are BPAB (43 sequences) and BGSH (26 sequences). The relatively high number of peptides extracted with BGSH beads also shows that different spectrometric approaches must be used to obtain the most exhaustive results.

3.8 Conclusions I

From all the analyses carried out on the mixtures of standard peptides, and from the results obtained using different types of beads, it is clear that a systematic peptidome study of a certain tissue cannot disregard the use of different extraction protocols. In fact, it is necessary to consider the high complexity and variety of peptides which are potentially present in biological samples.

From the considerable amount of work carried out on standard mixtures of peptides, especially on MIX19, it emerges that the use of different magnetic beads, commercial or modified “in house”, is necessary in order to have a high probability of representative extraction of peptides present in biological samples, both qualitatively and quantitatively, this a fundamental condition for carrying out differential peptidomic studies aimed at the discovery of new biomarkers. The knowledge acquired thanks to the work done so far has been exploited, in the last part of this thesis, for a first validation on real samples.

The HPLC-MS-ESI approach, necessary for the development of extraction protocols, has been replaced by the use of a spectrometric system equipped with a matrix assisted laser desorption ionization (MALDI) source. This, in addition to the advantage of working only with monocharged ions, has allowed more measurements of the same sample, with the consequent advantage of obtaining statistically significant data.

3.9 An application study of magnetic beads on real biological samples

The last part of this thesis work has concerned a first application to real samples of the knowledge acquired during the development of the extractive and analytical method.
This preliminary study concerned the presence and difference of low-molecular-weight protein fraction in plasma on a reduced sample of twelve patients undergoing maintenance haemodialysis (HD) with a molecular diagnosis of COVID-19 during the second pandemic wave of 2020 (October-December). The obtained data were compared with those obtained by analyzing the plasma from a control group of six healthy volunteers not affected by Covid-19 at the time of sampling.

The twelve patients diagnosed with COVID-19 were further divided into two groups according to the severity of their symptoms (group t0 and group nt0).

On the filtrate obtained from plasma samples after a pre-treatment with an ultrafiltration device (Vivaspin 500, see materials and methods), in order to reduce the high molecular weight protein fraction (with a threshold of 5kDa), the extraction of the low molecular weight protein fraction was performed using different functionalized magnetic beads to find possible qualitative/quantitative differences in the peptide component. Although the ESI-MS spectrometer equipped with ion trap analyzer was extremely useful during the development of the extraction method, its relative low resolution as well as the generation of multi-charge ions (both factors that would have complicated the work of structural characterization) made it more optimal to analyse the eluates from the beads by means of MALDI-TOF mass spectrometry.

3.9.1 The differential approach in proteomics

The differential approach in proteomics consist in comparing data from different homogeneous groups of samples to find significant differences. As previously reported, in this study the groups considered were the following three: four patients in the acute phase of COVID infection (t0), eight covid patients in the absence of acute symptoms (nt0, blood draw a week or two weeks from acute symptoms) and six control samples (C, healthy people vaccinated and not affected by COVID at the time of blood sampling). On the basis of the results obtained previously by analysing the standard solution of known synthetic peptides (MIX19) and the mixture of peptides 15-mer from protein spike S of SARS-Cov-2 virus, in order to increase the probability of finding differences in the peptide component between these groups, the data from samples processed with three types of magnetic silica beads were considered: BGSH, BNH_2, BPAB. All measures were carried out in triplicate.
In this preliminary study, the processing of the spectral data and the subsequent comparison took place only on the acquired m/z signals and their relative abundances, without advancing any hypothesis on the possible peptide sequences to which they belong.

3.9.2 GEENA2 and SAM

From the output file of the MALDI instrument, a list of m/z values and their intensities. The raw data were generated and stored; the resulting peaks list was then used for subsequent processing. The first elaboration of the lists was carried out through GEENA2 [114] , a publicly available web tool that aims to solve the main problems encountered in the analysis of data from MALDI-TOF spectra. The main functions of GEENA2 are:

• pre-processing of the replicate spectra of each sample, based on the sum of the isotopic abundances attributable to a single peptide
• calculation, for each sample, of the average technical replicas of the spectra from the different experimental replicates, generating a single representative spectrum of the sample.
• the alignment along the m/z axis of the mediated spectra of each sample to produce a table of signals with corresponding abundances

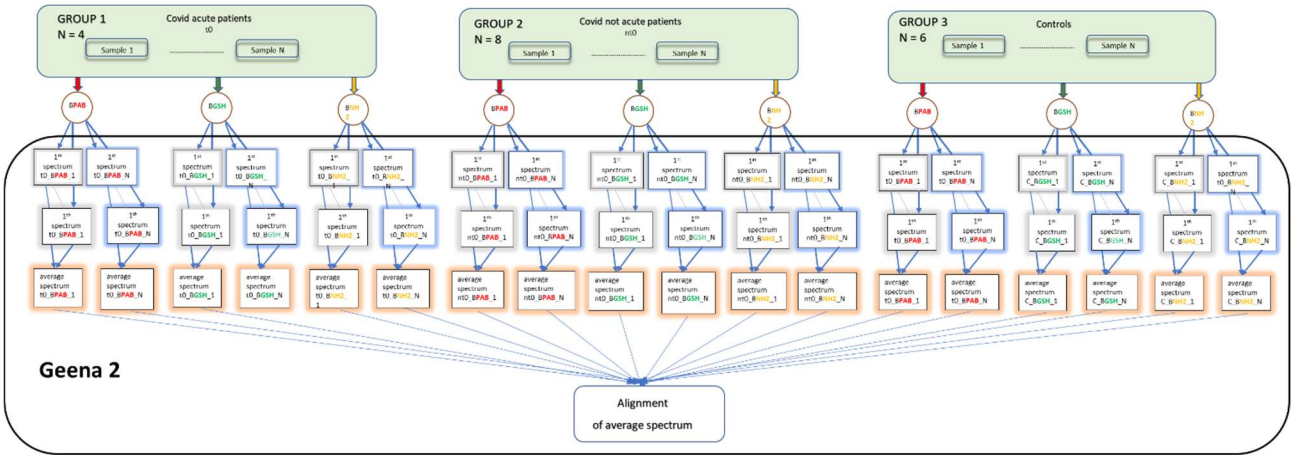


Figure 99 Data processing scheme with GEENA2 of the raw data coming from the analysis carried out in triplicate on plasma samples extracted with BGSH, BNH_2, BPAB magnetic beads.

The final alignment is achieved by averaging all m/z signals and their abundances from the same sample. Samples have been grouped as controls or according to different stages of the disease.

The reference to the original spectra is maintained in the final alignment, and this output data can therefore be passed to other tools for further analysis.

One of the results obtained by GEENA2 for each group of replicate includes the “average spectrum”, achieved by aligning and averaging all filtered spectra from the same sample, together with related alignment information. The average spectrum includes (m/z, abundance) pairs which are computed as average of respective m/z and abundance values for equivalent signals found in the filtered spectra.

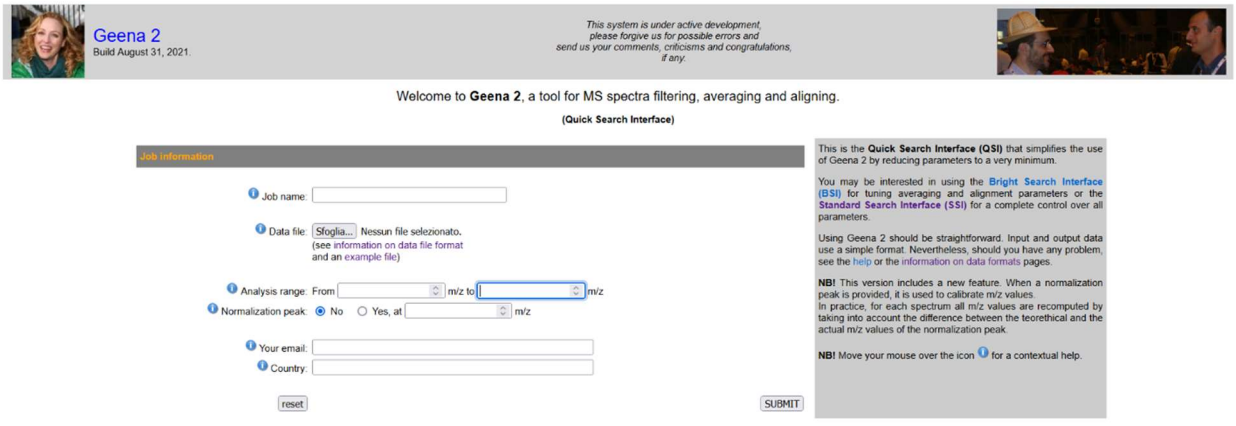


Figure.100 The main dialogue window for data input in GEENA2 (https://proteomics.hsanmartino.it/geena2/)

The next algorithm used to analyse data was SAM (significance analysis of microarrays) [115] (web MS-SAM, available at URL http://proteomics.hsanmartino.it/mssam/), which also proposes useful

algorithms for the "clustering" of data. The mode chosen to further analyze our data was the "two-class unpaired test" and a “multiclass comparison”. It is a methodology that allows to highlight any significant differences in terms of intensity in the expression of signals between different groups in comparison. The findings were also processed by a bootstrap resampling procedure, implemented in order to validate the output of the SAM analysis. The bootstrap method consists in the construction of new datasets by randomly resampling from the original one [116].

Starting with replacement from the original dataset, 100 new datasets of the same size as the original one were generated and SAM analysis performed for each of them. The original ratio between group sizes has also been preserved. Finally, the frequency of occurrence of the m/z selected as significant by SAM on the original dataset was evaluated in the results of the new 100 SAM analyses.

Beads	Number of signals in analysis range
BNH2	183
BPAB	109
BGSH	116

Table 37. Number of m/z signals extracted by each type of beads and processed with SAM algorithm.

Among the diverse types of beads used those that have allowed to obtain signals that show significant quantitative differences between the different types of beads used and where are manifested signals that are significantly higher or lower in the multiclass comparison are the BNH2 type.

Of the 183 peptides extracted with these beads, six were present in significantly different quantitates in the samples obtained from different groups of patients. This result further underlines how it is important to use different types of magnetic beads for the extraction of peptides from the samples of interest in order to have plurality of comparison in the research and identification of possible biomarkers. The results obtained by analysing the data from the three sample groups show significant quantities differences for six signals, four of which (m/z 628.347, m/z 636.76, m/z 614.799 and m/z 870.014) show a significant decrease in intensity in the group of patients with acute symptoms (t0) compared to the control group (C). A statistically significant decrease, but of lesser intensity, is observed for the same m/z ratios in the group of patients without acute symptoms (nt0) (Table 38). Two of the six significantly different m/z signals (m/z 614.084 and m/z 604.043) are present in smaller amounts in patients without acute symptoms (nt0) than in the control group (C),

but in the group of patients with acute symptoms (t0) the intensity of these signals is very similar to that found for the control group.

m/z	Contrast NCovid_t0_BNH2	Contrast NCovid_nt0_BNH2	Contrast Controls_BNH2
628.347	-0.811	-0.610	1.746
636.76	-0.914	-0.362	1.383
614.799	-0.744	-0.415	1.343
870.014	-0.627	-0.379	1.840
614.084	-0.122	-0.601	1.181
604.043	-0.164	-0.555	1.13

Table 38. The SAM processing allowed to get significant differences for six m/z signals within the three studied groups of patients.

The m/z values of the level 1 MS spectra reported in the table 38 were tested as input for Python Code I to try to identify the proteins that the peptides extracted with beads belonged to. This was done by searching for corresponding sequences on the local database of peptides obtained by fragmentation in silico of about 200 plasma proteins downloaded from the Uniprot site in FASTA format (see materials and methods). Considering the range of m/z values (600-900) of the peptides present in the Table 37, the fragmentation of whole proteins to obtain an appropriate list of peptide fragments was performed to get aminoacidic sequences between 5 and 11 residues.The mass interval to search for m/z values on the database sequences was set at ±0.2 Da. Analysing the output of the search performed with the Python Code I, only two m/z present in table 38 were associated with possible peptide sequences (table 39). Unfortunately, it was not possible to identify any sequence among the 200000 contained in the PeptideAtlas database, originating from the silico fragmentation of 4395 proteins. This is probably due to the fact that the peptide fragments of the PeptideAtlas database are composed of at least 7 amino acids.

peptide_sequence	UniProt FASTA code	m/z	
CNSSTC	P01130 · LDLR_HUMAN	614.191	**Low-density lipoprotein receptor**
SCSSCQ	P04196 · HRG_HUMAN	614.19	**Histidine-rich glycoprotein**
SYDDD	P35858 · ALS_HUMAN	614.194	**Insulin-like growth factor-binding protein complex acid labile subunit**
SYDDD	Q4ZHG4 · FNDC1_HUMAN	614.194	**Fibronectin type III domain-containing protein 1**
YECCP	Q15063 · POSTN_HUMAN	614.1948	**Periostin**
DYDSD	Q8IYJ1 · CPNE9_HUMAN	614.194	**Copine-9**
VGERGP	P39060 · COIA1_HUMAN	614.088	**Collagen alpha-1(XVIII) chain**
VNERP	Q6UVK1 · CSPG4_HUMAN	614.088	**Chondroitin sulfate proteoglycan 4**
VGGPRE	P11215 · ITAM_HUMAN	614.088	**Integrin alpha-M**
VGGRPE	P98160 · PGBM_HUMAN	614.088	**Basement membrane-specific heparan sulfate proteoglycan**
VVGDQP	O43157 · PLXB1_HUMAN	614.088	**Plexin-B1**
VGRPEG	Q6UVK1 · CSPG4_HUMAN	614.088	**Chondroitin sulfate proteoglycan 4**
VRENP	Q9Y4L1 · HYOU1_HUMAN	614.088	**Hypoxia up-regulated protein 1**
VVNPGE	Q8NFP4 · MDGA1_HUMAN	614.088	**MAM domain-containing glycosylphosphatidylinositol**
SQECGSFL	Q9HBE5 · IL21R_HUMAN	870.121	**Interleukin-21 receptor**
TYTCHMD	Q9HBE5 · IL21R_HUMAN	870.121	**Interleukin-21 receptor**
YAKCDSSP	P02765 · FETUA_HUMAN	870.121	**Alpha-2-HS-glycoprotein**
TDQFLCSG	P06681 · CO2_HUMAN	870.121	**Complement C2**
HCSAAAHSS	Q9BQ95 · ECSIT_HUMAN	870.121	**Evolutionarily conserved signaling intermediate in Toll pathway**
STYDRDN	P02675 · FIBB_HUMAN	870.121	**Fibrinogen beta chain**
MSFVGENS	Q9NZH6 · IL37_HUMAN	870.121	**Interleukin-37**

Table 39. List of possible amino acid sequences corresponding to m/z values present in Table 38, identified by the use of Python Code I. From left to right are reported in the columns: the matching amino acid sequence, the Uniprot code FASTA, the m/z value corresponding to the amino acid sequence, the protein containing this sequence.

Conclusions II

The main goal of this thesis project was to set up a method for comprehensive peptidomic analysis aimed to get a solid starting point on which to develop differential peptidomic methods on human plasma or serum. The detection and characterization of the low molecular weight fraction of a tissue is, in fact, increasingly used to identify possible biomarkers.

To pursue this goal the work proceeded through several stages:

- a high-performance liquid chromatography (HPLC) method has been optimized and two spectrometric methods have been developed: the first, at low-medium resolution, equipped with an electrospray ion source and an ion trap analyzer coupled with the HPLC (HPLC-ESI-iontrap MS), the second, off-line, equipped with a MALDI source and a time of flight analyzer (TOF) at medium-high resolution.

The HPLC-ESI-MS method was useful for the development of the extraction method while the MALDI-TOF method was used for the analysis of commercial peptide mixtures from biological samples and for the analysis of real samples.

- two different mixtures of standard peptides containing respectively thirteen and nineteen peptides were prepared, most of which were synthesized "ad hoc" for this project. The peptide mixtures were fundamental for the extraction phase set up. For the extraction it was decided to use a solid phase approach based on the use of magnetic silica beads, both for their ease and speed of use and for the high analytical throughput they can guarantee. The synthesis of two different functionalized magnetic beads was also done during this thesis work.

-the huge amount of data has made necessary the "in-house" development of informatic tools that allowed to obtain immediate and easy-to-read results, which is a necessary condition to carry on the development of the extraction method.

Three different Python codes have been developed. Code I for the analysis of level 1 MS spectra (full scan MS) code II for the analysis of level 2 MS spectra (tandem mass or MS/MS) and code III for the analysis of data from biological samples. Data analysis from biological samples also required the use of two powerful bio-informatic tools available on-line: GEENA2 and SAM.

A first phase of the work was carried out on serum of healthy patients. This initial phase was hampered by two main closely related factors: the presence of high molecular weight proteins in the eluate obtained following beads extraction, and the use of silica beads functionalized with reverse phase C18. The use of these beads on the serum led to the simultaneous extraction, by absorption, of high amounts of albumin. Moreover, the use of methods for depletion of the high molecular weight fraction of the serum sample led to very low recovery yields.

For these reasons it was decided to take a step back and start again by developing an extraction method based on the use of standard peptide mixtures and new magnetic beads. Commercial beads equipped with amino and carboxyl functional groups were further functionalized.

The use of different types of beads has proven successful to allow a wide recovery of peptides from standard mixtures. This data has been further confirmed by experiments performed on a commercial peptide mixture containing fifteen amino acids sequences of the spike S protein of Sars-Cov2 virus. The use of beads synthesized specifically to extract some specific peptides from this sample also demonstrated the importance of developing "affinity" magnetic beads for the extraction and quantification of biomarkers, once they were identified.

To validate the set procedures, the knowledge developed during this thesis project was finally applied to analyze real plasma samples from different groups of patients. Although the number of samples was limited, thanks also to a robust bio-informatic approach, it was possible to identify, among hundreds of extracted peptides, six signals present in significantly different amounts in different groups of patients; six potential biomarkers of disease progression.

BIBLIOGRAPHY

[1] Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature. 2008;452(7187):571-579.

[2] Curreem, S.O.T.; Watt, R.M.; Lau, S.K.P.; Woo, P.C.Y. Two-dimensional gel electrophoresis in bacterial proteomics. Protein Cell 2012, 3, 346–363

[3] Sirikaew, N.; Pruksakorn, D.; Chaiyawat, P.; Chutipongtanate, S. Mass Spectrometric-Based Proteomics for Biomarker Discovery in Osteosarcoma: Current Status and Future Direction. Int. J. Mol. Sci. 2022, 23, 9741

[4] Hu C, Dai Z, Xu J, Zhao L, Xu Y, Li M, Yu J, Zhang L, Deng H, Liu L, Zhang M, Huang J, Wu L and Chen G Proteome Profiling Identifies Serum Biomarkers in Rheumatoid Arthritis. Front. Immunol. 2022 13:865425.

[5] Bladergroen, Marco R., and Yuri EM van der Burgt. Solid-phase extraction strategies to surmount body fluid sample complexity in high-throughput mass spectrometry-based proteomics. Journal of Analytical Methods in Chemistry 2015.

[6] Rosa Terracciano, Luigi Pasqua, Francesca Casadonte, Stella Frascà, Mariaimmacolata Preianò, Daniela Falcone, and Rocco Savino. Derivatized Mesoporous Silica Beads for MALDI-TOF MS Profiling of Human Plasma and Urine, Bioconjugate Chemistry. 2009, *20* (5), 913-923

[7] Darie-Ion, L.; Whitham, D.; Jayathirtha, M.; Rai, Y.; Neagu, A.-N.; Darie, C.C.; Petre, B.A. Applications of MALDI-MS/MS-Based Proteomics in Biomedical Research. Molecules 2022, 27, 6196

[8] Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol. 2005; 1:252–262

[9] Bachi A, Bonaldi T. Quantitative proteomics as a new piece of the systems biology puzzle. J Proteomics. 2008; 71:357–367

[10] Wilkins MR, Pasquali C, Appel RD, et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)*. 1996;14(1):61-65.

[11] Kaltashov, I. A., Bobst, C. E., & Abzalimov, R. R. Mass spectrometry-based methods to study protein architecture and dynamics. Protein science: a publication of the Protein Society, 2013; 22(5), 530–544

[12] Luque-Garcia, J. L., & Neubert, T. A. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. Journal of chromatography. A, 2007; 1153(1-2), 259–276

[13] Savino R, Terracciano R. Mesopore-assisted profiling strategies in clinical proteomics for drug/target discovery. *Drug Discov Today*. 2012;17(3-4):143-152.

[14] Kastin AJ. Handbook of biologically active peptides. Amsterdam: Elsevier; 2013

[15] Geho DH, Liotta LA, Petricoin EF, Zhao W, Araujo RP. The amplified peptidome: the new treasure chest of candidate biomarkers. *Curr Opin Chem Biol*. 2006;10(1):50-55.

[16] Cunningham R, et al. Mass spectrometry-based proteomics and peptidomics for systems biology and biomarker discovery. Frontiers in biology, 2012, 7(4): 313-335.

[17] Gao Y, et al. Peptidome workflow of serum and urine samples for biomarker discovery. Analytical Methods, 2011, 3(4): 773-779.

[18] Lai X, Witzmann FA and Liangpunsakul S. Characterization of Peptides and Low Molecular Weight Proteins in Plasma from Subjects with Hepatocellular Carcinoma. A Proteomics. 2014;1(1): 6. ISSN:2471-0164

[19] Tammen H, Schulte I, Hess R, et al. Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. *Proteomics*. 2005;5(13):3414-3422.

[20] Apweiler R, Aslanidis C, Deufel T, et al. Approaching clinical proteomics: current state and future fields of application in cellular proteomics. *Cytometry A*. 2009;75(10):816-832.

[21] Heberle, H., Meirelles, G.V., da Silva, F.R. *et al.* InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16, 169 (2015)

[22] Roxas, B.A., Li, Q. Significance analysis of microarray for relative quantitation of LC/MS data in proteomics. *BMC Bioinformatics* 9, 187 (2008)

[23] Eng JK, Searle BC, Clauser KR, Tabb DL. A face in the crowd: recognizing peptides through database search. *Mol Cell Proteomics*. 2011;10(11):R111.009522.

[24] C. Hughes, B. Ma, G.A. Lajoie, De novo sequencing methods in proteomics, Methods Mol. Biol. 604 (2010) 105–121.

[25] J. Seidler, N. Zinn, M.E. Boehm, W.D. Lehmann, De novo sequencing of peptides by MS/MS, Proteomics 10 (2010) 634–649

[26] Lam H. Building and searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics*. 2011;10(12):R111.008565.

[27] H. Lam, R. Aebersold, Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics, Methods 54 (2011) 424–431.

[28] 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 20 (1999) 3551–3567.

[29] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, Bioinformatics 20 (2004) 1466–1467.

[30] Benjamin J. Diament and William Stafford Noble, Faster SEQUEST Searching for Peptide Identification from Tandem Mass Spectra, Journal of Proteome Research 2011 10 (9), 3871-3879

[31] Prianichnikov N, Koch H, Koch S, et al. MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics. *Mol Cell Proteomics*. 2020;19(6):1058-1069

[32] 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. Proteome Res. 10 (2011) 1794–1805.

[33] Menschaert G, Vandekerckhove TT, Baggerman G, Schoofs L, Luyten W, Van Criekinge W. Peptidomics coming of age: a review of contributions from a bioinformatics angle. J Proteome Res. 2010;9(5):2051-2061.

[34] Cardinali, B.; Lunardi, G.; Millo, E.; Armirotti, A.; Damonte, G.; Profumo, A.; Gori, S.; Iacono, G.; Levaggi, A.; Del Mastro, L. Trastuzumab Quantification in Serum: A New, Rapid, Robust ELISA Assay Based on a Mimetic Peptide That Specifically Recognizes Trastuzumab. Anal Bioanal Chem 2014, 406, 4557–4561.

[35] Krstulovic,Ante M, Brown, Phyllis R. Reversed-Phase High-Performance Liquid Chromatography: Theory, Practice, and Biomedical Applications. New York: John Wiley & Sons, Inc., 1982. Textbook.

[36] Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J. Cromatogr. 1990; 499; 177-196

[37] Cabrera, K., Lubda, D., Eggenweiler, H.-M., Minakuchi, H. and Nakanishi, K. (2000), A New Monolithic-Type HPLC Column for fast separations. J. High Resol. Chromatogr., 23: 93-99

[38] Hennessy TP, Boysen RI, Huber MI, Unger KK, Hearn MT. Peptide mapping by reversed-phase high-performance liquid chromatography employing silica rod monoliths. J Chromatogr A. 2003;1009(1-2):15-28.

[39] Han X, Aslanian A, Yates JR 3rd. Mass spectrometry for proteomics. Curr Opin Chem Biol. 2008 Oct;12(5):483-90

[40] Premstaller A, Oberacher H, Walcher W, et al. High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. Anal Chem. 2001;73(11):2390-2396.

[41] J.Stuart Grossert. A retrospective view of mass spectrometry and natural products—sixty years of progress, with a focus on contributions by R. Graham Cooks. International Journal of Mass Spectrometry 2001, 212 (1-3), 65-79.

[42] A. J. Dempster. Positive ray analysis of lithium and magnesium. Phys.Rev., Phys. Rev. , 1921, 18, 415.

[43] Nier, A.O. Mass Spectrometer for Isotope and Gas Analysis, Rev. Sci. Instrum. 1947, 18, 398–411.

[44] Munson M.S.B., Field F.H., Chemical Ionization Mass Spectrometry. I. General Introduction, J. Am. Chem. *Soc.* 88, 2621 (1966).

[45] Karas M., Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal. Chem., 1988, 60, 2299–2301

[46] Tanaka et al. Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry. Rapid Comm.Mass Spectrom., 1988, 2, 151–153.

[47] M. Yamashita, J. B. Fenn. Electrospray ion source. Another variation on the free-jet theme. J. Phys. Chem. 1984, 88: 4451–4459.

[48] Dole M., Mach L.L., Hines R.L. et al. Molecular Beams of Macroions. J. Chem. Phys., 1968, 49, 2240–2249.

[49] Banerjee S, Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. Int J Anal Chem. 2012, 2012:282574.

[50] Gaskell, S.J. Electrospray: Principles and Practice. J. Mass Spectrom., 1997, 32: 677-688.

[51] Thomson B. A. Iribane, J. V. On the evaporation of small ions from charged droplets. J.Chem.Phys., 1976, 64, 2287.

[52] Iribane J. V. Thomson, B. A. Field induced ion evaporation from liquid surfaces at atmospheric pressure. J.Chem.Phys., 1979, 71, 4451-4463.

[53] Schmelzeisen-Redecker et al. Desolvation of ions and molecules in thermospray mass spectrometry. Int.J.Mass Spectrom.Ion.Proc., 1989, 90, 139-150.

[54] Geromanos S, Freckleton G, Tempst P. Tuning of an electrospray ionization source for maximum peptide-ion transmission into a mass spectrometer. Anal Chem. 2000;72(4):777-790.

[55] K.L. Busch, G.L. Glish and S.A. McLuckey in “Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry” VCH Publishers, Inc., New York, 1988.

[56] Jennings, K.R. Collision-induced decompositions of aromatic molecular ions. *Int. J. Mass Spec. Ion Phys.* 1968,1, 227–235.

[57] Haddon, W.F. and McLafferty, F.W. Metastable Ion characteristics. VII. collision-Induced metastables. J. Am. Chem. Soc. 1968, 90, 4745–4746.

[58] Biemann K. Contributions of mass spectrometry to peptide and protein structure. Biomed Environ Mass Spectrom. 1988; 16(1-12):99-111.

[59] Biemann K. Appendix 5. Nomenclature for peptide fragment ions (positive ions) Methods Enzymol. 1990; 193:886–887.

[60] Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984; 11(11):601.

[61] Deutsch, Eric W. Mass spectrometer output file format mzML. Methods in molecular biology (Clifton, N.J.) 2010, 604: 319-31.

[62] Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R. A common open representation of mass spectrometry data and its application to proteomics research. Nat. Biotechnol. 2004, 22 (11)

[63] Lin SM, Zhu L, Winter AQ, Sasinowski M, Kibbe WA. What is mzXML good for? Expert review of proteomics 2005, 2 (6): 839–45.

[64] W. E. Stephens. A pulsed mass spectrometer with time dispersion. Bull. Am. Phys. Soc. 1946, 21, (2), 22.

[65] Eggers D. F. Cameron, A. E. An ion “Velocitron.” Rev.Sci.Instrum., 1948, 19, 605.

[66] McLaren I. H. Wiley, W. C. Time-of-flight mass spectrometer with improved resolution. Rev.Sci.Instrum., 1955, 26, 1150.

[67] Whittal RM, Li L. High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. Anal Chem. 1995;67(13):1950-1954.

[68] Cotter RJ. Time-of-flight mass spectrometry for the structural analysis of biological molecules. Anal Chem. 1992;64(21):1027A-1039A

[69] B.A. Mamyrin, V.I. Karataev, D.V. Shmikk, and V.A. Zagulin, Zh. Eksp. Teor. Fiz. 64, 82 (1973) SOV.Phys. JETP 1973, 37, 45.

[70] Zhou, J., et al. "Kinetic energy measurements of molecular ions ejected into an electric field by matrix-assisted laser desorption." *Rapid communications in mass spectrometry* 6.11 (1992): 671-678.

[71] Mamyrin, B. A., et al. "The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution." Zh. Eksp. Teor. Fiz 64.1 (1973): 82-89.

[72] Cotter RJ. Peer Reviewed: The New Time-of-Flight Mass Spectrometry. Anal Chem. 1999;71(13):445A-51

[73] Mamyrin B. A. Time-of-flight mass spectrometry (concepts, achievements, and prospects) International Journal of Mass Spectrometry. 2001;206(3):251–266.

[74] Safarik, I., & Safarikova, M. Magnetic techniques for the isolation and purification of proteins and peptides. Biomagnetic research and technology, 2004, 2(1), 7.

[75] M. A. Posthumus, P. G. Kistemaker, H. L. C. Meuzelaar, and M. C. Ten Noever de Brauw. Laser desorption-mass spectrometry of polar nonvolatile bio-organic molecules. Analytical Chemistry 1978, 50 .7, 985-991.

[76] Barber, M., Bordoli, R.S., Sedgwick, R.D., & Tyler, A.N. Fast atom bombardment mass spectrometry of cobalamines. Journal of Mass Spectrometry, 1981, 8, 492-495.

[77] Macfarlane R. D. Sundqvist, B. 252cf-plasma desorption mass spectrometry. Mass Spectrom., 1985.

[78] Beavis Vestal ML. Methods of ion generation. Chem Rev. 2001;101(2):361-375.

[79] Gunnar P. Jonsson, Allan B. Hedin, Per L. Hakansson, Bo U. R. Sundqvist, B. Goeran S. Saeve, Per F. Nielsen, Peter. Roepstorff, Karl Erik. Johansson, Ivan. Kamensky, and Maria S. L. Lindberg , Plasma desorption mass spectrometry of peptides and proteins absorbed on nitrocellulose. Anal.Chem., 1986 58 (6), 1084-1087

[80] Alai M, Demirev P, Fenselau C, Cotter RJ. Glutathione as a matrix for plasma desorption mass spectrometry of large peptides. Anal Chem. 1986;58(7):1303-1307.

[81] Beavis RC, Chait BT. Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation. *Rapid Commun Mass Spectrom*. 1989;3(12):436-439.

[82] Kaufmann R. Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. *J Biotechnol*. 1995;41(2-3):155-175.

[83] Karas, M., Bahr, U., Ingendoh, A., Nordhoff, E., Stahl, B., Strupat, K. and Hillenkamp, F., Principles and Applications of Matrix-Assisted UV-Laser Desorption/ Ionization Mass Spectrometry Analytica Chimica Acta 1990, 241, 175–185

[84] Beavis RC, Chait BT. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. Rapid Commun Mass Spectrom. 1989;3(12):432-435.

[85] Strupat et al. 2,5-dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. Int.J.Mass Spectrom., 1991, 111, 89-102.

[86] Beavis et al. Alpha-cyano-4- hydroxycinnamic acid as a matrix for matrix assisted laser desorption mass spectrometry. Org.Mass Spectrom., 1992.

[87] Xiang F, Beavis RC. A Method to Increase Contaminant Tolerance in Protein Matrix-Assisted Laser Desorption Ionization by the Fabrication of Thin Protein-Doped Polycrystalline Films. Rapid Communications in Mass Spectrometry. 1994;8:199–204

[88] Ole. Vorm, Peter. Roepstorff, and Matthias. Mann, Improved Resolution and Very High Sensitivity in MALDI TOF of Matrix Surfaces Made by Fast Evaporation. Anal. Chem. 1994, 66, 19, 3281–3287.

[89] Griffin, T. J., Goodlett, D. R., and Aebersold, R. (2001) Advances in proteome analysis by mass spectrometry. Curr. Opin. Biotechnol. 12, 607–612.

[90] Zenobi, R., Knochenmuss, R.: Ion formation in MALDI mass spectrometry. Mass Spectrom. Rev., 1998, 17, 337–366.

[91] Karas M. Gluckmann, M. The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization. J.Mass Spectrom., 1999, 34, 467-477.

[92] Rosinke, B.; Strupat, K.; Hillenkamp, F.; Rosenbusch, J.; Dencher, N., Krüger, U.; Galla, H. J. Matrix-Assisted Laser Desorption/Ionization, Mass Spectrometry (MALDI-MS) of Membrane Proteins and Non-Covalent Complexes. J. Mass Spectrom., 1995, 30, 1462–1468.

[93] Pfenninger A, Karas M, Finke B, Stahl B, Sawatzki G. Matrix optimization for matrix-assisted laser desorption/ionization mass spectrometry of oligosaccharides from human milk. J Mass Spectrom., 1999, 34(2):98-104.

[94] Preston LM, Murray KK, Russell DH. Reproducibility and quantitation of matrix-assisted laser desorption ionization mass spectrometry: effects of nitrocellulose on peptide ion yields. Biol Mass Spectrom., 1993, 22(9):544-550.

[95] Gusev A.I., Wilkinson,W.R., Proctor,A. and Hercules,D.M., Improvement of ion signal reproducibility and matrix/comatrix effects in MALDI analysis. Anal. Chem., 1995, 67, 1034–1041.

[96] Karas M, Glückmann M, Schäfer J. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. J Mass Spectrom., 2000, 35(1):1-12.

[97] Griffin TJ, Gygi SP, Rist B, et al. Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. Anal Chem., 2001, 73(5):978-986.

[98] Adusumilli R, Mallick P. Data Conversion with ProteoWizard msConvert. Methods Mol Biol., 2017; 1550:339-368.

[99] Röst, H.L., Schmitt, U., Aebersold, R. and Malmström, L., pyOpenMS: A Python-based interface to the OpenMS mass-spectrometry algorithm library. Proteomics, 2014, 14: 74-77.

[100] Bald T, Barth J, Niehues A, Specht M, Hippler M, Fufezan C. pymzML--Python module for high-throughput bioinformatics on mass spectrometry data. Bioinformatics., 2012, 28(7):1052-1053.

[101] Deutsch, E. W., Lam, H., & Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO reports*, *9*(5), 429–434.

[102] Farrah T, Deutsch EW, Aebersold R. Using the Human Plasma PeptideAtlas to study human plasma proteins. Methods Mol Biol., 2011, 728:349-374.

[103] Lipman, DJ; Pearson, WR , Rapid and sensitive protein similarity searches. Science., 1985, 227 (4693): 1435–41

[104] The UniProt Consortium, UniProt: the Universal Protein Knowledgebase in 2023, Nucleic Acids Research*,* 2023, 51; D1, D523–D531

[105] De Masi, L.; Argenio, M.A.; Giordano, D.; Facchiano, A. Molecular Aspects of Spike–ACE2 Interaction. Encyclopedia 2022, 2, 96–108

[106] Chan KK, Dorosky D, Sharma P, et al. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science*. 2020, 369(6508):1261-1265.

[107] Cowan, R., and R. G. Whittaker. *Hydrophobicity Indices for Amino Acid Residues as Determined by High-Performance Liquid Chromatography*. Peptide Research., 1990, 3(2):75–80

[108] Wilson, K. J., A. Honegger, R. P. Stötzel, and G. J. Hughes. *The Behaviour of Peptides on Reverse-Phase Supports during High-Pressure Liquid Chromatography*. The Biochemical Journal., 1981, 199(1):31-41.

[109] Ikai, A. *Thermostability and Aliphatic Index of Globular Proteins*. Journal of Biochemistry., 1980, 88(6):1895–98.

[110] Boman, H. G. Antibacterial peptides: basic facts and emerging concepts. Journal of Internal Medicine, 2003, 254(3), 197-215.

[111] Kujawski Jacek , Popielarska Hanna, Myka Anna, Drabińska Beata, Bernard Marek K, The logP Parameter as a Molecular Descriptor in the Computer-aided Drug Design – an Overview. Computational Methods in Science and Technology., 2012, 18 (2), 81-88.

[112] Sun H. A universal molecular descriptor system for prediction of logP, logS, logBB, and absorption. *J Chem Inf Comput Sci*., 2004, 44(2):748-757.

[113] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982, 157(1):105-132.

[114] Romano, P.; Profumo, A.; Rocco, M.; Mangerini, R.; Ferri, F.; Facchiano, A. Geena 2, improved automated analysis of MALDI/TOF mass spectra. BMC Bioinform., 2016, 17 (Suppl. 4)

[115] Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to transcriptional responses to ionizing radiation. Proc. Natl. Acad. Sci. 2001, USA., 98:5116–5121.

[116] Efron, B.; Tibshirani, R.J. An Introduction to the Bootstrap; Chapman & Hall/CRC: New York, NY, USA, 1993; ISBN 978-0-412-04231-7

ACKNOWLEDGEMENTS

My heartfelt thanks:

to Prof. Gianluca Damonte, Dr. Annalisa Salis and Prof. Enrico Millo for all teaching, patience and help.

A special thanks to

Dr. Aldo Profumo, Dr. Barbara Cardinali, Dr. Paolo Romano (Proteomics and Mass Spectrometry Unit, IRCCS Ospedale Policlinico San Martino)

for support and collaboration