UNIVERSITA' DEGLI STUDI DI VERONA

Facoltà di Medicina e Chirurgia

Dipartimento di Medina Clinica e Sperimentale

DOTTORATO DI RICERCA IN PROTEOMICA CLINICA

CICLO XX

TITOLO DELLA TESI DI DOTTORATO

Analisi comparativa dei metaboliti presenti nelle urine di

soggetti sani ed affetti da carcinoma della vescica mediante LC-MS/MS

S.S.D. MED/09

Coordinatore: prof. Oliviero Olivieri

Tutore:Prof. Franco Turrini.

Dottorando: dott. Salvatore Crobu

INTRODUCTION

For several years proteomics research has been expected to lead to the finding of new markers that will translate into clinical tests applicable to samples such as serum, plasma and urine: so-called in vitro diagnostics (IVDs). Attempts to implement technologies applied in proteomics as 2DE, Immuno Blotting, Mass Spectrometry have initiated constructive discussions on opportunities and challenges inherent in such a translation process also with respect to the use of multi-marker profiling approaches and pattern signatures in IVD. It is mandatory to fulfil requirements in routine IVD, including disease prevention, diagnosis, prognosis, and treatment monitoring or follow up among others. To fulfill IVD requirements, it is essential to provide diagnostic tests that allow for definite and reliable diagnosis tied to a decision on interventions (prevention, treatment, or non treatment), meet stringent performance characteristics for each analyte (in particular test accuracy, including both precision of the measurement and trueness of the measurement), and provide adequate diagnostic accuracy, i.e., diagnostic sensitivity and diagnostic specificity, determined by the desired positive and negative predictive values which depend on disease frequency. The fulfilment of essential IVD requirements is mandatory in the regulated environment of modern diagnostics. Addressing IVD needs at an early stage can support a timely and effective transition of findings and developments into routine diagnosis. IVD needs reflect features that are useful in clinical practice. This helps to generate acceptance and assists the implementation process[1].

The medical need for relevant biomarkers is enormous. This is particularly true for the many types of cancer, but other diseases such as Type 1 diabetes (DMT1) also lack useful and adequate diagnostic markers with high specificity and sensitivity. Despite advances in imaging technologies for early detection of diseases, proteomic and peptidomic multiplex techniques and metabolomics statistical analysis have evolved in recent years[2].

Diagnosis/Screening of Bladder Cancer

The bladder is an organ located in the pelvic cavity that stores and discharges urine. Urine is produced by the kidneys, carried to the bladder by the ureters, and discharged from the bladder through the urethra[3]. Bladder cancer accounts for approximately 90% of cancers of

the urinary tract (renal pelvis, ureters, bladder, urethra). Bladder cancer usually originates in the bladder lining, which consists of a mucous layer of surface cells that expand and deflate (transitional epithelial cells), smooth muscle, and a fibrous layer. Tumors are categorized as low-stage (superficial) or high-stage (muscle invasive). In industrialized countries (e.g., United States, Canada, France), more than 90% of cases originate in the transitional epithelial cells (called transitional cell carcinoma; TCC). In developing countries, 75% of cases are squamous cell carcinomas caused by Schistosoma haematobium (parasitic organism) infection. Rare types of bladder cancer include small cell carcinoma, carcinosarcoma, primary lymphoma, and sarcoma.According to the National Cancer Institute, the highest incidence of bladder cancer occurs in industrialized countries such as the United States, Canada, and France. Bladder cancer is the fourth most common type of cancer in men and the eighth most common type in women. The disease is more prevalent in Caucasians than in African Americans and Hispanics.

Causes and Risk Factors

Cancer-causing agents (carcinogens) in the urine may lead to the development of bladder cancer. Cigarette smoking contributes to more than 50% of cases, and smoking cigars or pipes also increases the risk. Other risk factors include the following:

- Age
- Chronic bladder inflammation (recurrent urinary tract infections, urinary stones)
- Consumption of Aristolochia fangchi (herb used in some weight-loss formulas)
- Diet high in saturated fat
- Exposure to second-hand smoke
- External beam radiation
- Family history of bladder cancer (several genetic risk factors identified)
- Gender (male)
- Infection with Schistosoma haematobium (parasite found in many developing countries)
- Personal history of bladder cancer
- Race (Caucasian)
- Treatment with certain drugs (e.g., cyclophosfamide - used to treat cancer)

Exposure to carcinogens in the workplace also increases the risk for bladder cancer. Medical workers exposed during the preparation, storage, administration, or disposal of antineoplastic drugs (used in chemotherapy) are at increased risk. The primary symptom of bladder cancer is blood in the urine (hematuria). Hematuria may be visible to the naked eye (gross) or visible only under a microscope (microscopic) and is usually painless. Other symptoms include frequent urination and pain upon urination (dysuria)[4].

Diagnosis of bladder cancer includes urological tests and imaging tests. A complete medical history is used to identify potential risk factors (e.g., smoking, exposure to dyes).
Laboratory tests may include the following:

- NMP22®BladderChek® (to detect elevated levels of tumor markers in the urine)
- Urinalysis (to detect microscopic hematuria)
- Urine cytology (to detect cancer cells by examining cells flushed from the bladder during urination)
- Urine culture (to rule out urinary tract infection)
- NMP22®BladderChek® is a urine test used to detect elevated levels of a nuclear matrix protein (called NMP22®). Bladder cancer increases levels of this protein in the urine, even during early stages of the disease.
- Results of this test, which is non invasive and is performed in a physician's office, are available during the patient's office visit. Studies have shown that when used with cystoscopy, NMP22®BladderChek® may be more effective than other diagnostic tests (e.g., urine tests or cystoscopy alone).

Various imaging tests may also be performed. Intravenous pyelogram (IVP) is the standard imaging test for bladder cancer. In this procedure, a contrast agent (radiopaque dye) is administered through a vein (intravenously) and x-rays are taken as the dye moves through the urinary tract. IVP provides information about the structure and function of the kidneys, ureters, and bladder. Other imaging tests include CT scan, MRI scan, bone scan, and ultrasound. If bladder cancer is suspected, cystoscopy and biopsy are performed. Local anesthesia is administered and a cystoscope (thin, telescope-like tube with a tiny camera attached) is inserted into the bladder through the urethra to allow the physician to detect abnormalities. In biopsy, tissue samples are taken from the lesion(s) and examined for cancer cells. If the sample is positive, the cancer is staged using the tumor, node, metastases (TNM) system[3].

Biomarker discovery

Biomarkers are important tools that can supply some of the needed information, especially when used in conjunction with traditional clinical and laboratory data. A biomarker is a biologic characteristic that is measured and evaluated objectively as an indicator of normal biologic processes, pathogenic processes, or pharmacologic response to therapeutic intervention. Biomarkers may be any parameter of a patient that can be measured, for example, mRNA expression profiles, proteins metabolites, peptides, proteomic patterns, lipids, imaging methods, or electrical signals. The best biomarkers are accurate, relatively noninvasive and easy-to-perform tests that can be done at the bedside or in the outpatient setting. These tests involve a blood or spot urine specimen, can be measured serially, and have a fast turnaround[5].

Recently, functional studies have emphasized analyses at the level of gene expression (transcriptomics), protein expression (proteomics) and now the metabolic network (metabolomics) with a view of the "systems biology" approach of defining the phenotype and bridging the phenotype- to - genotype divide. Metabolomics is the one of the latest in the string of "omic" technologies, involving the comprehensive analysis in which all the metabolites of an organism are identified and quantified[6]. The biochemical response of an organism to a perturbation can be characterised by its effect on the differential accumulation of individual metabolites[7]. Metabolomics aims at the comprehensive and quantitative analysis of wide arrays of metabolites in biological samples. These numerous analytes have very diverse physico-chemical properties and occur at different abundance levels. Consequently, comprehensive metabolomics investigations are primarily a challenge for analytical chemistry and specifically mass spectrometry has vast potential as a tool for this type of investigation.

Mass spectrometry

Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample. For large samples such as biomolecules, molecular masses can be measured to within an accuracy of 0.01% of the total molecular mass of the sample i.e. within a 4 Daltons (Da) or atomic mass units (amu) error for a sample of 40,000 Da. This is sufficient to allow minor mass changes to be detected, e.g. the substitution of one amino acid for another, or a post-translational modification. For small organic molecules the molecular mass can be measured to within an accuracy of 3 ppm or less, which is often sufficient to confirm the molecular formula of a compound, and is also a standard[8].

Structural information can be generated using certain types of mass spectrometers, usually those with multiple analyzers which are known as tandem mass spectrometers. This is achieved by fragmenting the sample inside the instrument and analysing the products generated. This procedure is useful for the structural elucidation of organic compounds and for peptide or oligonucleotide sequencing.

Mass spectrometers are used in industry and academia for both routine and research purposes. The following list is just a brief summary of the major mass spectrometric applications:

Biotechnology: the analysis of proteins, peptides, oligonucleotides

Pharmaceutical: drug discovery, combinatorial chemistry, pharmacokinetics, drug metabolism

Clinical: neonatal screening, haemoglobin analysis, drug testing

Environmental: PAHs, PCBs, water quality, food contamination

Geological: oil composition

Especially mass spectrometry help biochemists in accurate:

Accurate molecular weight measurements: sample confirmation, to determine the purity of a sample, to verify amino acid substitutions, to detect post-translational modifications, to calculate the number of disulphide bridges

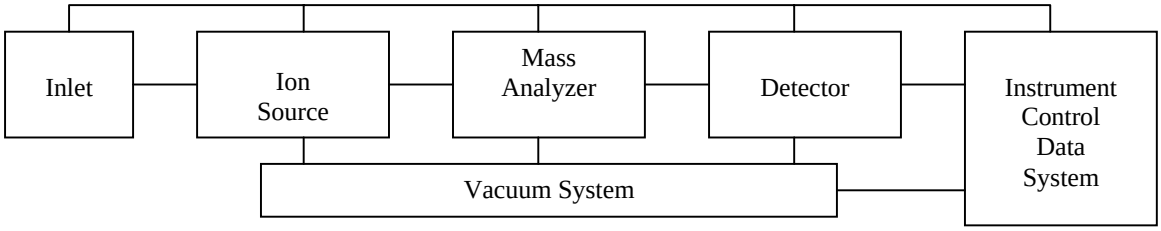
Reaction monitoring: to monitor enzyme reactions, chemical modification, protein digestion

Amino acid sequencing: sequence confirmation, de novo characterisation of peptides, identification of proteins by database searching with a sequence "tag" from a proteolytic fragment

Oligonucleotide sequencing: the characterisation or quality control of oligonucleotides
Protein structure: protein folding monitored by H/D exchange, protein-ligand complex formation under physiological conditions, macromolecular structure determination

Instrumentation

Mass spectrometers can be divided into five fundamental parts, namely the inlet system, the ionization source , the analyzer, the detector and the data system. The sample has to be introduced into the ionization source of the instrument. Once inside the ionization source, the sample molecules are ionized, because ions are easier to manipulate than neutral molecules. These ions are extracted into the analyzer region of the mass spectrometer where they are separated according to their mass (m) -to-charge (z) ratios (m/z) . The separated ions are detected and this signal sent to a data system where the m/z ratios are stored together with their relative abundance for presentation in the format of a m/z spectrum. The analyzer and detector of the mass spectrometer, and often the ionization source too, are maintained under high vacuum to give the ions a reasonable chance of travelling from one end of the instrument to the other without any hindrance from air molecules. The entire operation of the mass spectrometer, and often the sample introduction process also, is under complete data system control on modern mass spectrometers[8] (Fig 1)



(Fig 1) Mass spectrometer logical shìcheme

Sample introduction

The method of sample introduction to the ionizations source often depends on the ionizations method being used, as well as the type and complexity of the sample. The sample can be inserted directly into the ionization source like Target Plate in Matrix Assisted Laser Desorption Ionization source, or can undergo some type of chromatography en route to the

ionization source. This latter method of sample introduction usually involves the mass spectrometer being coupled directly to a high performance liquid chromatography (HPLC), gas chromatography (GC) or capillary electrophoresis (CE) separation column, and hence the sample is separated into a series of components which then enter the mass spectrometer sequentially for individual analysis[8]

Ionization methods

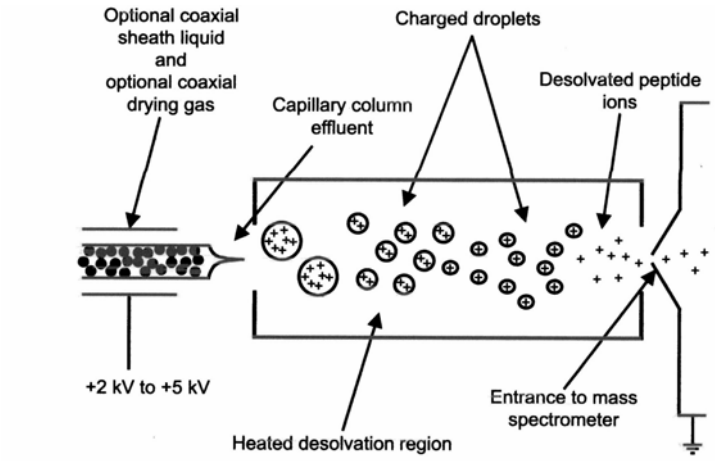
A fundamental challenge to the application of mass spectrometry to any class of analytes is the production of gas-phase ions of those species, and difficulties in producing gas-phase ions can prevent mass spectrometric analysis of certain class of molecules. Many ionization methods are available and each has its own advantages and disadvantages[9]. The ionization methods used for the majority of biochemical analyses are Electro Spray Ionization (ESI) and Matrix Assisted Laser Desorption Ionization (MALDI) , and these are described in more detail in Sections 5 and 6 respectively. With most ionization methods there is the possibility of creating both positively and negatively charged sample ions, depending on the proton affinity of the sample[8].

Electrospray ionization

Electrospray Ionisation (ESI) (Fig.2) is one of the Atmospheric Pressure Ionisation (API) techniques and is well-suited to the analysis of polar molecules ranging from less than 100 Da to more than 1,000,000 Da in molecular mass; ESI is known as a "soft" ionization method as the sample is ionised by the addition or removal of a proton. During standard electrospray ionisation[10], the sample is dissolved in a polar, volatile solvent and pumped through a narrow, stainless steel capillary (75 - 150 micrometers i.d.) at a flow rate of between 1 µL/min and 1 mL/min. A high voltage of 3 or 4 kV is applied to the tip of the capillary, which is situated within the ionization source of the mass spectrometer, and as a consequence of this strong electric field, the sample emerging from the tip is dispersed into an aerosol of highly charged droplets, a process that is aided by a co-axially introduced nebulizing gas flowing around the outside of the capillary. This gas, usually nitrogen, helps to direct the spray emerging from the capillary tip towards the mass spectrometer. The charged droplets diminish in size by solvent evaporation, assisted by a warm flow of nitrogen known as the drying gas which passes across the front of the ionization source. Eventually charged sample ions, free

from solvent, are released from the droplets, some of which pass through a sampling cone or orifice into an intermediate vacuum region, and from there through a small aperture into the analyzer of the mass spectrometer, which is held under high vacuum. To enhanced sensitivity more instrument were equipped with a low flow rate version of electrospray ionizations, called nanoESI[11].

In positive ionization mode, a trace of formic acid is often added to aid protonation of the sample molecules; in negative ionization mode a trace of ammonia solution or a volatile amine is added to aid deprotonation of the sample molecules. Proteins and peptides are usually analyzed under positive ionization conditions and saccharides and oligonucleotides under negative ionization conditions. In all cases, the m/z scale must be calibrated by analyzing a standard sample of a similar type to the sample being analyzed (e.g. a protein calibrant for a protein sample), and then applying a mass correction[8].



(Fig 2) ESI source

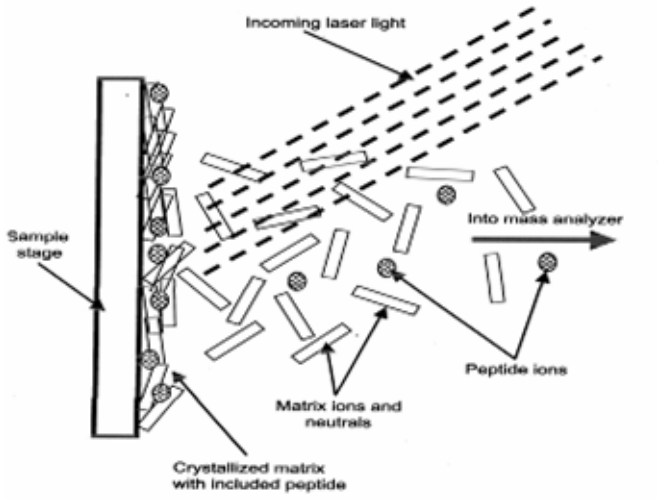
In ESI, samples (M) with molecular masses up to ca. 1200 Da give rise to singly charged molecular-related ions, usually protonated molecular ions of the formula $(M+H)^+$ in positive ionization mode, and deprotonated molecular ions of the formula $(M-H)^-$ in negative ionization mode. Samples (M) with molecular weights greater than ca. 1200 Da give rise to multiply charged molecular-related ions such as $(M+nH)^{n+}$ in positive ionization mode and $(M-nH)^{n-}$ in negative ionization mode. Proteins have many suitable sites for protonation as all of the backbone amide nitrogen atoms could be protonated theoretically, as well as certain amino acid side chains such as lysine and arginine which contain primary amine

functionalities. If the number of charges on an ion is known, then it is simply a matter of reading the m/z value from the spectrum and solving the above equation to determine the molecular weight of the sample. Usually the number of charges is not known, but can be calculated if the assumption is made that any two adjacent members in the series of multiply charged ions differ by one charge. This may seem long-winded but fortunately the molecular mass of the sample can be calculated automatically, or at least semi-automatically, by the processing software associated with the mass spectrometer called deconvolution[8].

MALDI

Matrix assisted laser desorption ionization[12] deals well with thermolabile, non-volatile organic compounds especially those of high molecular mass and is used successfully in biochemical areas for the analysis of proteins, peptides, glycoproteins, oligosaccharides, and oligonucleotides. It is relatively straightforward to use and reasonably tolerant to buffers and other additives. The mass accuracy depends on the type and performance of the analyser of the mass spectrometer, but most modern instruments should be capable The sample to be analysed is dissolved in an appropriate volatile solvent, usually with a trace of trifluoroacetic acid if positive ionisation is being used, at a concentration of ca. 10 pmol/µL and an aliquot (1-2 µL) of this removed and mixed with an equal volume of a solution containing a vast excess of a matrix. A range of compounds is suitable for use as matrices: sinapinic acid is a common one for protein analysis while alpha-cyano-4-hydroxycinnamic acid is often used for peptide analysis. An aliquot (1-2 µL) of the final solution is applied to the sample target which is allowed to dry prior to insertion into the high vacuum of the mass spectrometer. The laser (Fig.3) is fired, the energy arriving at the sample/matrix surface optimised, and data accumulated until a m/z spectrum of reasonable intensity has been amassed. The time-of-flight analyser separates ions according to their mass(m)-to-charge(z) (m/z) ratios by measuring the time it takes for ions to travel through a field free region known as the flight, or drift, tube. The heavier ions are slower than the lighter ones. The m/z scale of the mass spectrometer is calibrated with a known sample that can either be analysed independently (external calibration) or pre-mixed with the sample and matrix (internal calibration). of measuring masses to within 0.01% of the molecular mass of the sample, at least up to ca. 40,000 Da. MALDI is based on the bombardment of sample molecules with a laser light to bring about sample ionization. The sample is pre-mixed with a highly absorbing matrix compound for the

most consistent and reliable results, and a low concentration of sample to matrix works best. The matrix transforms the laser energy into excitation energy for the sample, which leads to sputtering of analyte and matrix ions from the surface of the mixture. In this way energy transfer is efficient and also the analyte molecules are spared excessive direct energy that may otherwise cause decomposition. Most commercially available MALDI mass spectrometers now have a pulsed nitrogen laser of wavelength 337 nm. MALDI is also a "soft" ionisation method and so results predominantly in the generation of singly charged molecular-related ions regardless of the molecular mass, hence the spectra are relatively easy to interpret. Fragmentation of the sample ions does not usually occur[8].



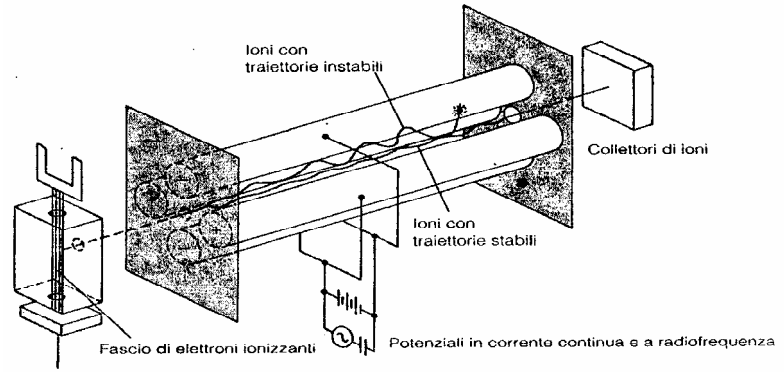
(Fig 3) MALDI Source

Analyzers

The analyzer uses dispersion or filtering to sort ions according to their mass-to-charge ratios or a related property. The most widely used analyzers are magnetic sectors, quadrupole mass filters, quadrupole ion traps, Fourier transform ion cyclotron resonance spectrometers, and time-of-flight mass analyzers.

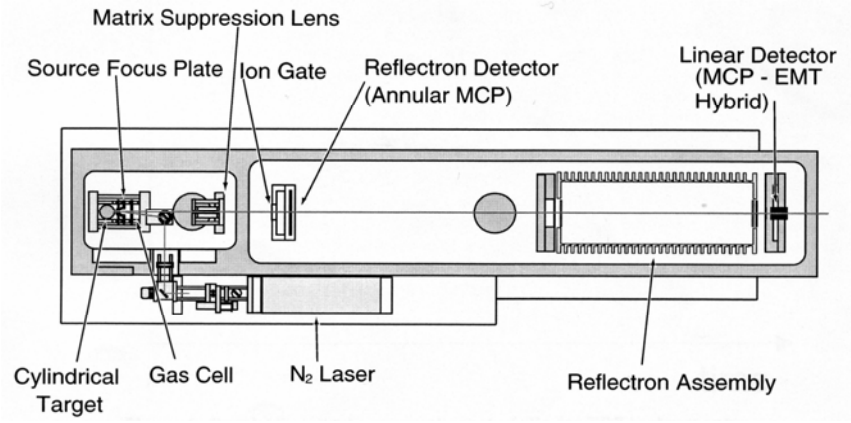
The first type of mass analyzer to consider is quadrupole mass filter, consists of four parallel poles or rods. In this device (Fig.4), mass sorting depends on ion motion resulting from simultaneously applied constant (dc) and radio frequency electric (rf) electric fields. Scanning is accomplished by systematically changing the field strengths, thereby changing the m/z value that is transmitted through the analyzer. Quadrupole mass spectrometers provide

lower resolution than double focusing instruments but tend to be more easily interfaced to various inlet systems and to be less costly.



(Fig 4) Quadrupole analyzer

Time-of-flight mass analyzers (Figure.5), usually coupled with MALDI separate ions by virtue of their different flight times over a known distance. A brief burst of ions is emitted from a source. These ions are accelerated so that ions of like charge have equal kinetic energy and then are directed into a flight tube. Since kinetic energy is equal to 1/2 mv2, where m is the mass of the ion and v is the ion velocity, the lower the ion's mass, the greater the velocity and shorter its flight time. The travel time from the ion source through the flight tube to the detector, measured in microseconds, can be transformed to the m/z value through the relationships described above. Because all ion masses are measured for each ion burst, TOF mass spectrometers offer high sensitivity as well as rapid scanning. They can provide mass data for very high-mass biomolecules[8].



(Fig 5) Time-of-flight mass analyzers

Improved mass resolution in MALDI TOF-MS has been obtained by the utilization of a single-stage or a dual-stage reflectron (RETOF-MS). The reflectron, located at the end of the flight tube, is used to compensate for the difference in flight times of the same m/z ions of slightly different kinetic energies by means of an ion reflector. This results in focusing the ion packets in space and time at the detector. Enhancing the mass resolution can also increase the mass accuracy when determining the ion's mass. Another contribution to mass resolution loss in conventional (i.e., continuous ion extraction) linear MALDI TOF-MS is attributed to a range of flight times of identical m/z ions due to different initial velocities. A fraction of the final velocity that is attained for particular m/z ions as they are accelerated out of the ion source and into the field free region includes this initial velocity component. No compensation is made with continuous ion extraction linear TOF-MS for ions with the same m/z but different initial ion velocities. Improvements in mass resolution can be achieved by utilizing delayed pulsed ion extraction[1314] (DE) can compensate for the initial velocity distribution of the MALDI generated ion packet such that same m/z ions arrive simultaneously at a space focal plane located at the detector. Broadening of the ion velocity distribution due to collisional processes in the ion source can also be minimized by allowing the dense plume of MALDI generated ions/neutrals to dissipate prior to ion draw out from the ion source. This results in narrower ion arrival time distributions and provides better mass resolution when compared to continuous ion extraction.

Tandem mass spectrometry (MS/MS)

Tandem mass spectrometry (MS/MS) is used to produce structural information about a compound by fragmenting specific sample ions inside the mass spectrometer and identifying the resulting fragment ions. This information can then be pieced together to generate structural information regarding the intact molecule. Tandem mass spectrometry also enables specific compounds to be detected in complex mixtures on account of their specific and characteristic fragmentation patterns. A tandem mass spectrometer is a mass spectrometer that has more than one analyzer, in practice usually two. The two analyzers are separated by a collision cell into which an inert gas (e.g. argon, xenon) is admitted to collide with the selected sample ions and bring about their fragmentation (Fig.6).

The basic modes of data acquisition for tandem mass spectrometry experiments are as follows:

Product or daughter ion scanning:

the first analyser is used to select user-specified sample ions arising from a particular component; usually the molecular-related (i.e. (M+H)+ or (M-H)-) ions. These chosen ions pass into the collision cell, are bombarded by the gas molecules which cause fragment ions to be formed, and these fragment ions are analysed i.e. separated according to their mass to charge ratios, by the second analyser. All the fragment ions arise directly from the precursor ions specified in the experiment, and thus produce a fingerprint pattern specific to the compound under investigation. This type of experiment is particularly useful for providing structural information concerning small organic molecules and for generating peptide sequence information[15].

Precursor or parent ion scanning:

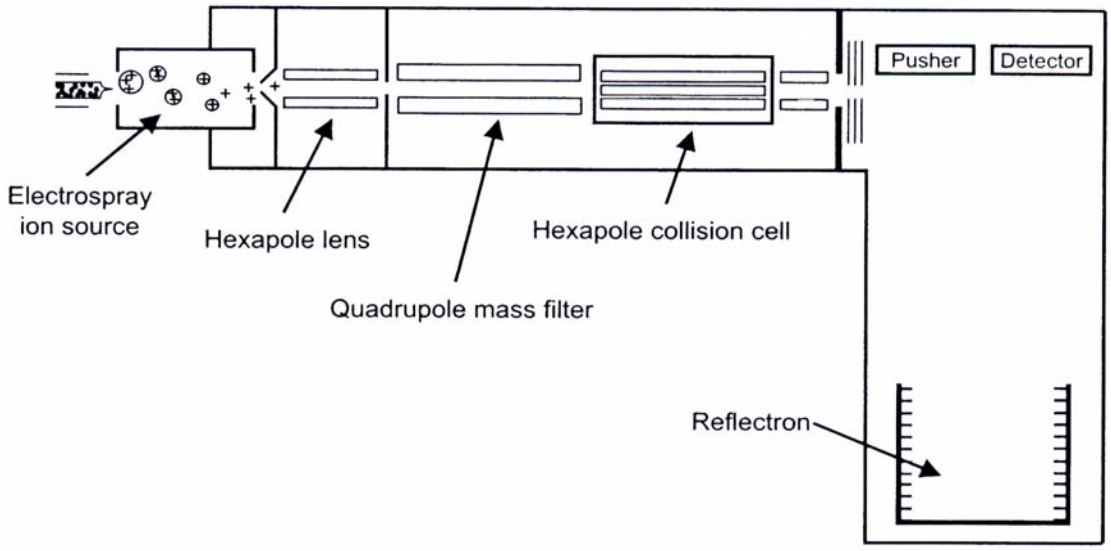
the first analyser allows the transmission of all sample ions, whilst the second analyser is set to monitor specific fragment ions, which are generated by bombardment of the sample ions with the collision gas in the collision cell. This type of experiment is particularly useful for monitoring groups of compounds contained within a mixture which fragment to produce common fragment ions, e.g. glycosylated peptides in a tryptic digest mixture, aliphatic hydrocarbons in an oil sample, or glucuronide conjugates in urine[15].

Constant neutral loss scanning:

this involves both analyzers scanning, or collecting data, across the whole m/z range, but the two are off-set so that the second analyser allows only those ions which differ by a certain number of mass units (equivalent to a neutral fragment) from the ions transmitted through the first analyser. e.g. This type of experiment could be used to monitor all of the carboxylic acids in a mixture. Carboxylic acids tend to fragment by losing a (neutral) molecule of carbon dioxide, CO_2, which is equivalent to a loss of 44 Da or atomic mass units. All ions pass through the first analyser into the collision cell. The ions detected from the collision cell are those from which 44 Da have been lost[15].

Selected/multiple reaction monitoring:

both of the analyzers are static in this case as user-selected specific ions are transmitted through the first analyser and user-selected specific fragments arising from these ions are measured by the second analyser. The compound under scrutiny must be known and have been well-characterised previously before this type of experiment is undertaken. This methodology is used to confirm unambiguously the presence of a compound in a matrix e.g. drug testing with blood or urine samples. It is not only a highly specific method but also has very high sensitivity[15].

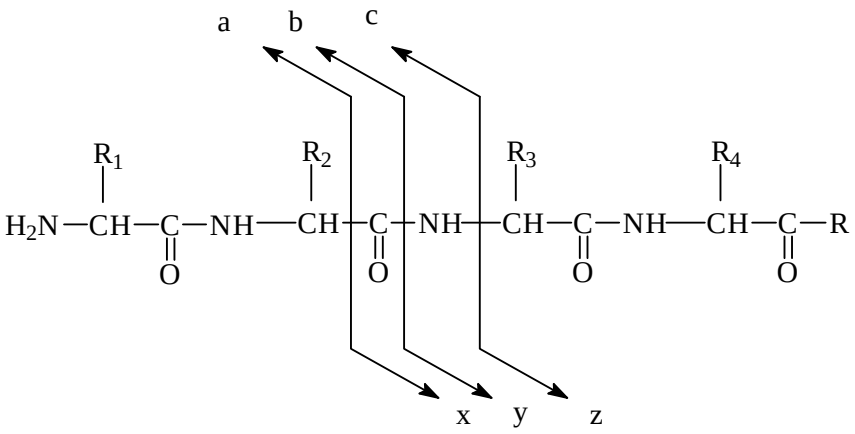


(Fig 6) Tandem mass spectrometry

Peptide Sequencing by Tandem Mass Spectrometry.

The most common usage of MS/MS in biochemical areas is the product or daughter ion scanning experiment which is particularly successful for peptide and nucleotide sequencing. Peptides fragment in a reasonably well-documented manner[1617].The protonated molecules fragment along the peptide backbone and also show some side-chain fragmentation with certain instruments[18]. There are three different types of bonds that can fragment along the amino acid backbone: the NH-CH, CH-CO, and CO-NH bonds. Each bond breakage gives rise to two species, one neutral and the other one charged, and only the charged species is

monitored by the mass spectrometer(Fig.7). The charge can stay on either of the two fragments depending on the chemistry and relative proton affinity of the two species. Hence there are six possible fragment ions for each amino acid residue and these are labelled as in the diagram, with the a, b, and c" ions having the charge retained on the N-terminal fragment, and the x, y", and z ions having the charge retained on the C-terminal fragment. The most common cleavage sites are at the CO-NH bonds which give rise to the b and/or the y" ions. The mass difference between two adjacent b ions, or y"; ions, is indicative of a particular amino acid residue (see Table of amino acid residues at the end of this document). The extent of side-chain fragmentation detected depends on the type of analysers used in the mass spectrometer. A magnetic sector - magnetic sector instrument will give rise to high energy collisions resulting in many different types of side-chain cleavages. Triple quadrupole and quadrupole-time-of-flight mass spectrometers generate low energy fragmentations with fewer types of side-chain fragmentations.



(Fig 7) Peptide fragmentation

Statistical methods

Over the last decade or so a new body of practice has become standard in computational analysis. It is known variously as corpus-based, empirical, or statistical methods of data analysis, most common being the simple rubric statistical methods. Present-day computational analysis differs from “traditional” computational methods in the pervasiveness of probabilities in its theoretical models, the centrality of large data collections, including multidisciplinary data and integrated databases, and the emphasis on rigorous empirical evaluation. The change in computational analysis is part of a larger shift to statistical methods in computer science, particularly in artificial intelligence, pattern recognition, speech recognition, and machine learning. In this study were employed PCA and LDA like statistical approach to analyze our data.

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components*. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible PCA is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. But this is not necessarily the case, depending on the application. This technique allows to discover or to reduce the dimensionality of the data set and to identify new meaningful underlying variables[19].

AIMS

This study focuses on the development and application of metabolomics technologies to peptides in urine of Bladder Cancer. Numerous attempts made in the understanding of the molecular mechanisms of many pathologies, among which tumours; much still remains unknown with regards to pathogenesis, with the end results being the development of an efficient strategy for early diagnosis and early cure. The analysis of small molecule, peptide and metabolites expressions of tissues from patients affected with tumours and various diseases are surely an ideal starting point. With this aim, proteomics coupled statistical methods is therefore a good method for this study:

- knowledge of metabolites variations rising from pathological states;
- the development of new biological markers for early diagnosis detection of the disease;
- identification of new therapeutic targets;
- the evaluation of therapeutic and toxic effects of new drugs.

For the research of tumour markers, the most common approach used so far has been the comparison of healthy and unhealthy tissues in other to identify particular protein patterns that are characteristic of tumour cells with respect to healthy cells.

Peptidomics technologies provide new opportunities for the detection of low-molecular-weight proteome biomarkers (peptides) by mass spectrometry. Improvements in peptidomics research are based on separation of peptides and/or proteins by their physicochemical properties in combination with mass spectrometric detection, identification and sophisticated bioinformatics tools for data analysis.

MATERIALS AND METHODS

Sample Preparation

In order to isolate selectively small organic molecules fraction of a complex substrate, like urine, was necessary to develop a purification method, to produce the highest content of metabolites and minor amounts of interference, we had testing a wide variety of protocols and techniques.

The first step of each tested method had required to eliminate intact cells (endogen or hexogen) or epithelium piece, so spontaneous urine samples were collected, immediately frozen, and stored at -20 °C. After thawing, the urine samples (10ml) were centrifuged for 20 min at 1000 rpm.

The second step involves the removal of the protein fraction. The supernatant was slowly denaturated with a linear temperature ramp: 35 ° C to 95 ° C per 120 minutes in a gas-chromatography oven to obtain most reproducibility (DT).:It is important that the denaturating temperature is reached slowly in order to favour the digestion of the colloidal precipitates and to avoid the inclusion of peptides inside the proteic clusters. Or as a method to remove protein fraction tested was chemical denaturation: we preformed these by three time volume of CH_3CN in agitation for 30 min. Each sample was concentrated to 3 ml, centrifugated at 13000 rpm for 15 min and the supernatant was applied onto a Size Exclusion Chromatography (SEC) column packed “in-house” with Sephadex G10 (Sigma-Aldrich) to remove salts, urea, electrolytes, and other interfering matrix components. Alternatively we performed salt elimination by Solid Phase Extraction (SPE) in two kind of Reverse Phase preparative microcolumn: Strata C_{18} by Phenomenex and Amprep^{TM} C_2 by Amersham Bioscences[20]. Collected fractions were lyophilized and recovered with 500 μL HPLC-grade water, to enhance the amount of peptides injected

MS/MS analysis

We have tested more instrument setup as so HPLC column, calibration and ionization technology; the goal is follow described

Metabolite mixture in mQ water are the better substrate to LC-MS/MS analysis. 5 μL of this solution were analyzed using a nanoflow capillary liquid chromatography coupled with nanoelectrospray quadrupole time of flight tandem mass spectrometry (nanoESI Q-TOF MS/MS). The Q-Tof Ultima (Waters, UK) was equipped with CapLC with autosampler (Waters, UK) using a ten port zero dead volume valve (Valco-Vici) enabling fast sample loading on a precolumn (Opti-Pack, Symmetry 300 C_{18}, 5 μm, Waters) with a isocratically flow rate of 20 μl/min for 5 min with solvent A delivered by auxiliary pump C to wash sample from residuals salt. After back-flushing washing of the pre-column, the ten-port valve was switched allowing delivery of a linear solvent B gradient 10-95% in 55 min[21] (solvent A: H_2O with 0.2% formic acid and 2% CH_3CN; solvent B: CH_3CN with 0.2% formic acid) with a flow rate of 200 nl/min after split onto the analytical column (NS-AC-12 BioSphere C18, 150 mm x 75 μm, 5 μm, by NanoSeparation). NanoESI MS/MS tandem spectra were recorded in the automated MS to MS/MS switching mode, with a m/z-dependent set of collision energy offset values. Singly to quadruply charged ions were selected and fragmented, with argon in collision gas cell. External calibration was performed in MS/MS mode by a solution of Glu1-Fibrinopeptide B (Sigma-Aldrich) 300 fM in H_20/CH_3CN 50/50 and 0.2% HCOOH; the same solution was employed to "Lock Mass Calibration" by *PressurePump* (Waters^{TM}). NanoESI source was equipped by *Universal Spray* (Waters ^{TM}) to increase stability of spray through PicoTip ^{TM} Emitter (metallized silica tip i.d. 10 to 5 μm by New Objective)

MALDI-MS analysis

We have trialed if was possible acquire the same information by MALDI-TOF analysis since this technologies is faster and cheaper than LC-MS/MS. 1 μL of purified sample was added to 1 μL of MALDI matrix solution (a-Cyano-4-Hydroxycinnamic Acid, 10 μg/μl in H_2O/CH_3CN/TFA:60/40/0.1) and spotted onto target plate. Spectra were acquired by MALDI Micro MX (Waters, UK) in positive ionization and reflectron mode. Eternal calibration was performed with PEG-NaI (Waters mix calib. Solution) from 600 Da to 5000 Da; we used the same range to acquired sample spectra.

Spectra Interpretation

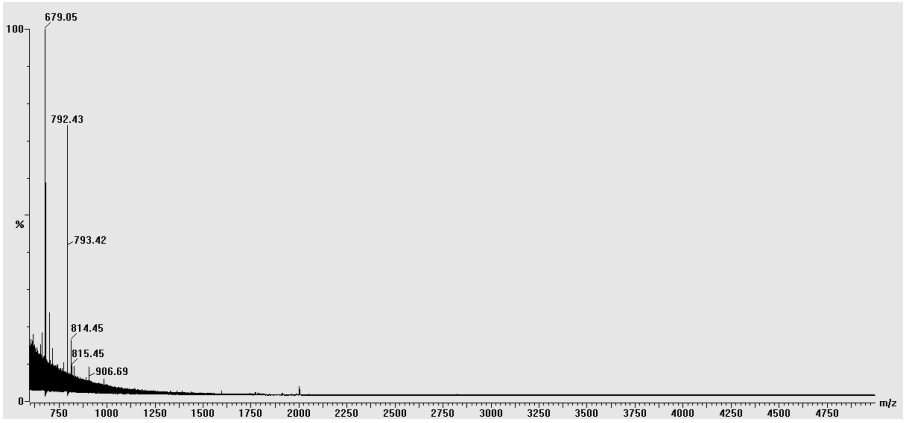
Mass data collected during an RP-LC-MS/MS analysis were processed by Protein Lynx Global Server (Waters, UK) and converted into a PKL file to be submitted:

- statistical elaboration with PCA approach,
- mathematical elaboration to feature selection: to identify "target" signal
- automated database searching "Mascot , MS/MS Ions Search" (available in house).

Data base search parameters were: parent tolerance 0.2 Da, fragment tolerance 0.1 Da, database SWISSPROT[22]. PCA was performed by SimcaP+ (Umetrics) to obtain variable space reduction in a plots of the first three components that allowed visualisation of the data and to establish whether there were any intrinsic healthy/unhealthy-related differences in the metabolic composition of the urine.

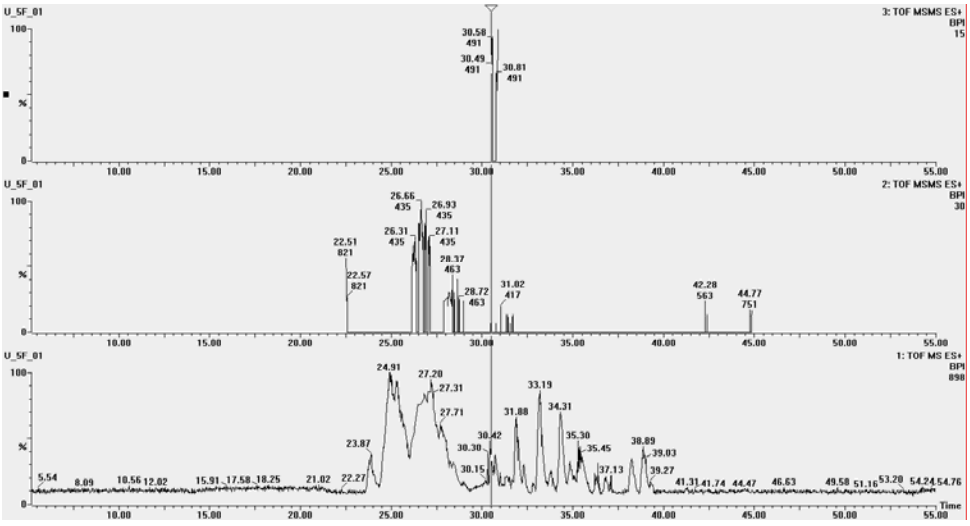
RESULTS & CONCLUSION

MALDI spectra showed in fig.8. is representative of all tested purification method. The number of signals is unsatisfactory, probably because presence salt residual decrease ionization whereas in LC-MS/MS analysis are removed during washing on precolumn

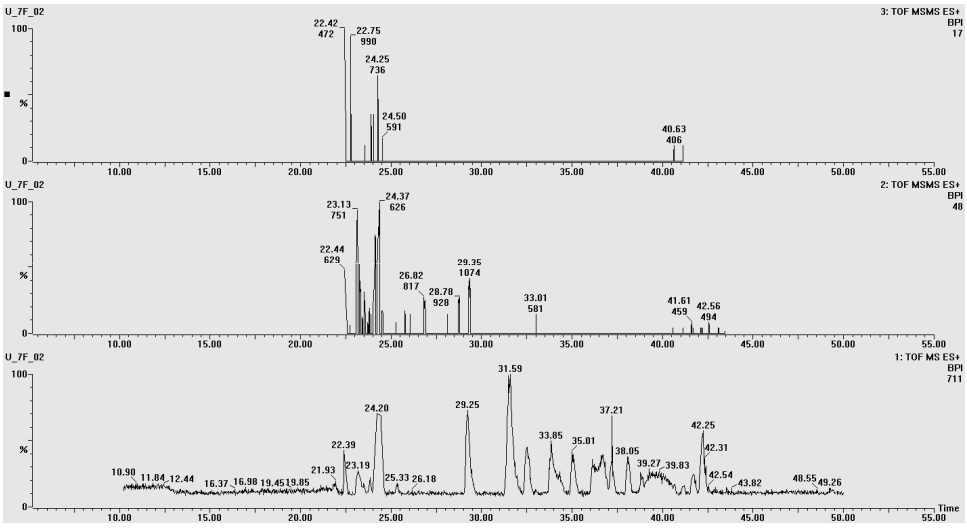


(Fig 8.) MALDI-TOF spectra

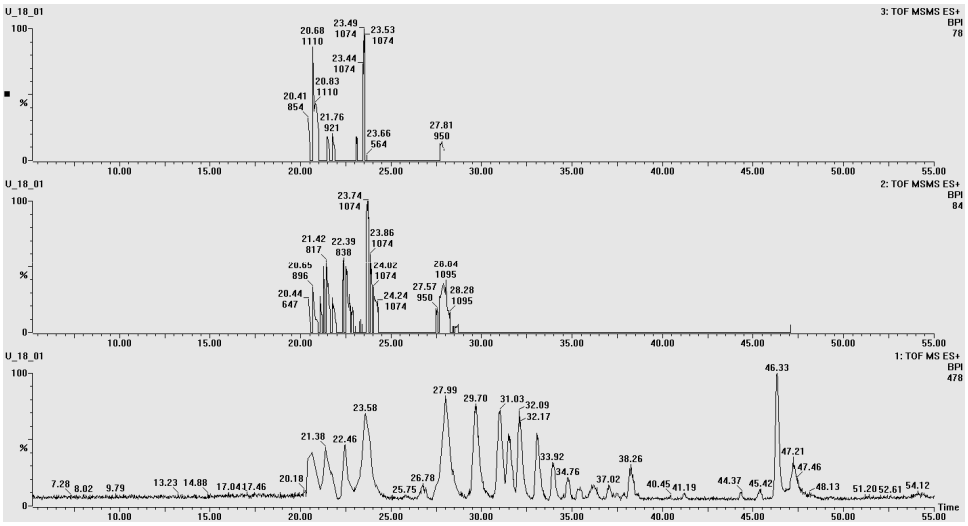
LC-MS/MS chromatograms showed below (Figg.9-14) display the trend signal according to the methods of purification tested



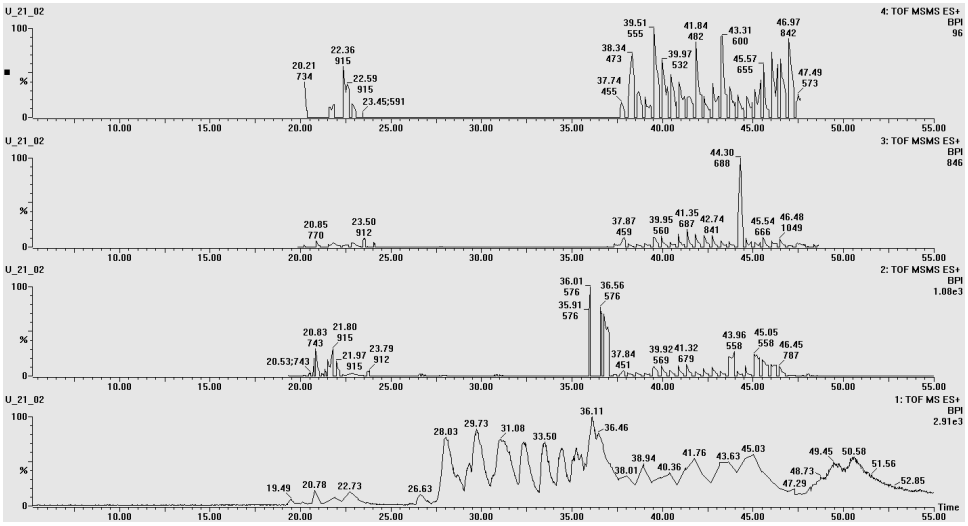
(Fig 9) MS/MS chromatogram generate after chemical denaturation and SPE C_{18}



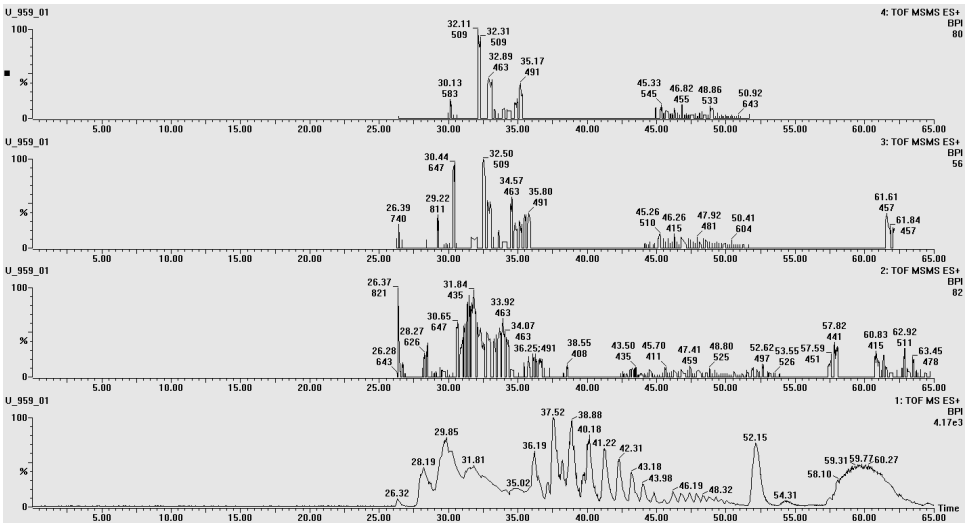
(Fig 10.) MS/MS chromatogram generate after chemical denaturation and SPE C_2



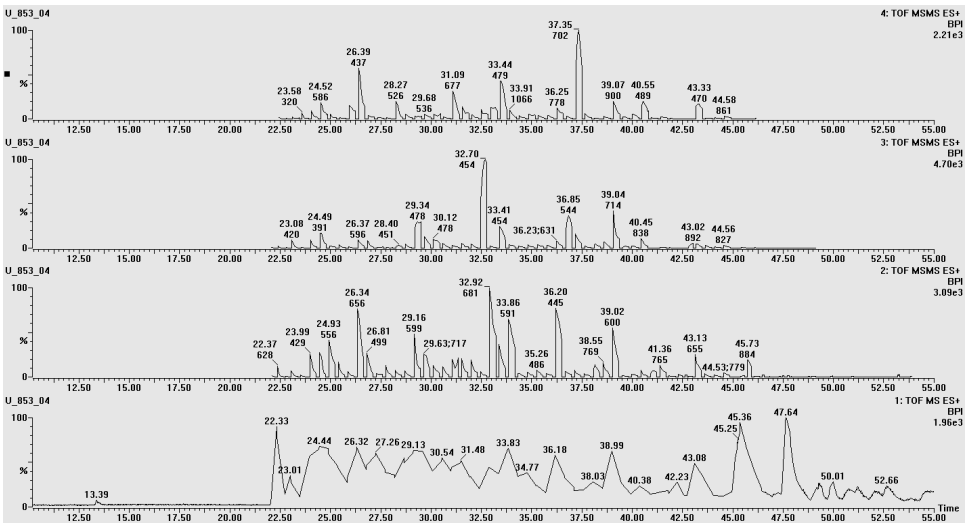
(Fig 11) MS/MS chromatogram generate after chemical denaturation and SEC



(Fig 12) MS/MS chromatogram generate after thermal denaturation and SPE C_{18}



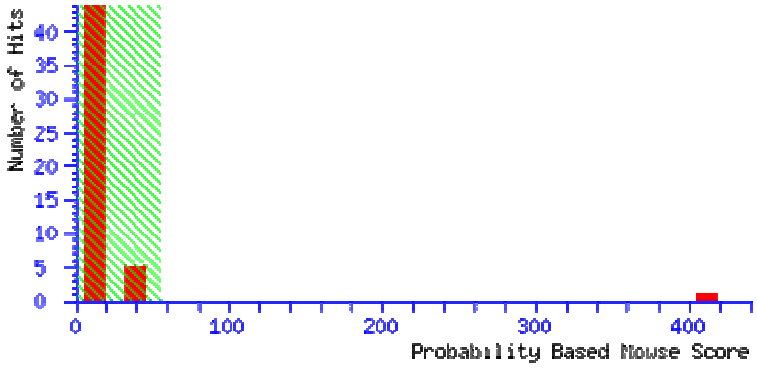
(Fig 13.) MS/MS chromatogram generate after thermal denaturation and SPE C_2



(Fig 14) MS/MS chromatogram generate after thermal denaturation and SEC

In each chromatogram can be seen the first channel due to the current generated by intact peptide onto detector, in the upper acquisition channels can see the signals generated by the amino acid chains fragmentation. Amount of signals presented by various chromatograms allow to say that the thermal denaturation followed by SEC is the most sensitive method. Mascott submission of PKL file generate by process of these chromatogram return “Uromoduline Precuror”; not significative hit because uromoduline is most aboundante protein in urine (Fig15).

Database:	SwissProt 46.2 (221901 sequences; 100660862 residues)
Taxonomy:	Homo sapiens (human) (48408 sequences)
Significant hits:	P07911-00-00-00 (UROM_HUMAN) Splice isoform Displayed; Variant Displayed; Conflict Displayed; from P07911 Uromodulin precursor (Tamm-Horsfall urinary glycoprotein) (THP)



Obser.	Mr(expt)	Mr(calc)	Delta	Score	Expect	Rank	Peptide
491.83	981.64	981.60	0.04	40	1.7	1	VLNLGPITR
561.01	1680.01	1679.97	0.04	111	1.2e-07	1	VIDQSRVLNLGPITR
884.54	1767.07	1767.00	0.07	81	0.00012	1	SVIDQSRVLNLGPITR
956.55	1911.08	1911.05	0.03	134	6.4e-10	1	SGSVIDQSRVLNLGPITR
680.74	2039.21	2039.15	0.06	57	0.035	1	SGSVIDQSRVLNLGPITRK

(Fig. 15) Uromodulin identification by match MS/MS spectra in *Mascott*

Mathematic elaboration of the molecular masses of the series of peptides obtained by pathologic and control urine to feature selection has given some particular or important abundant targets that are present in tumours and are relatively or scarsely present or absent in controls. The cluster formed by these series (Tab 1 A-B) of peptides might provide an early, rapid and non invasive diagnosis system at a relatively low cost.

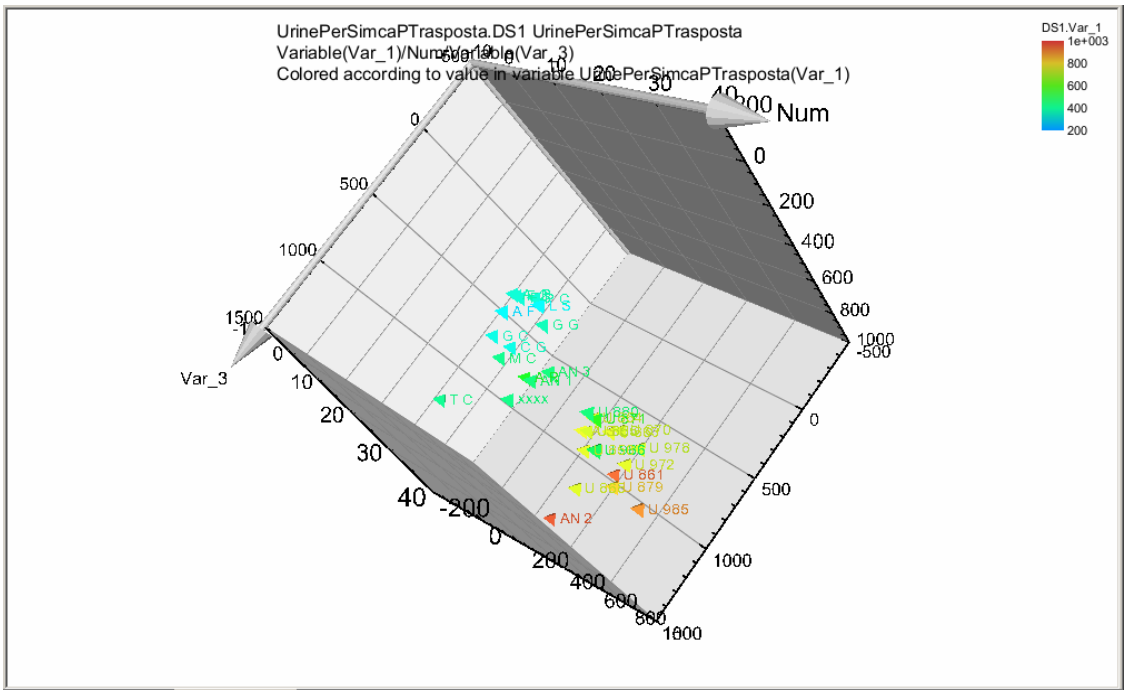
M.W. Peptide	Frequency Bladder Cancer	Frequency Control
1735.8	56.25	6.25
2190.8	56.25	50
2800.1	50	75
4949.9	50	25
2843.1	50	18.75
1022.4	43.75	25
6186.3	43.75	18.75
2305.9	37.5	18.75
2844.2	37.5	18.75
4984.1	37.5	0
1898.9	31.25	43.75
731.3	31.25	6.25
4844.0	31.25	0
2345.0	31.25	0
1489.0	25	6.25
1942.7	25	0
2017.1	25	0
2307.3	18.75	0

Tab. 1 A

M.W. Peptide	Frequency Bladder Cancer	Frequency Control
1735.8 731.3 2843	81.25	18.75
1735.8 4984.1	81.25	6.25
1735.8 731.3	68.75	6.25

Tab. 1 B

Moreover, the ms/ms data were elaborated to extract M.W. intact peptide information with “in-house” developed cig-win algorithm, to create matrix for PCA analysis. The elaboration carry out .a consistence reduction of data set, displayable in a tri-dimensional scatter plot (fig.16) showing an effective separation in two clusters (healthy and inhealthy). Thus PCA give us a predictive tool to discriminate sample.



(Fig 16) Tridimensional *Scatter plot.*

DISCUSSION

In conclusion, This study was focused on the development and application of metabolomics technologies to peptides in urine of Bladder Cancer.

In order to isolate selectively small organic molecules fraction of a complex substrate, like urine, was necessary to develop a purification method, to produce the highest content of metabolites and minor amounts of interference, we had testing a wide variety of protocols and techniques.

We have also tested more instrument setup to increase: number, repeatability and stability of information obtained by mass spectrometry. The was coupled SEC with DT to sample preparation and nanoLC-nanoESI-Q-TOF like analyzer.

Mathematic elaboration of the molecular masses of the series of peptides obtained by pathologic and healthy urine to feature selection has given some particular or important abundant targets that are present in tumours and are relatively or scarsely present or absent in controls. The cluster formed by these series of peptides might provide an early, rapid and non invasive diagnosis system at a relatively low cost.

The ms/ms data were elaborated to extract M.W. intact peptide information with “in-house” developed cig-win algorithm, to create matrix for PCA analysis. The elaboration carry out .a consistence reduction of data set, displayable in a tri-dimensional scatter plot showing an effective separation in two clusters (healthy and inhealthy). Thus PCA give us a predictive tool to discriminate sample.

References

1) Proteomics: from basic research to diagnostic application. A review of requirements & needs. Vitzthum F, Behrens F, Anderson NL, Shaw JH.J Proteome Res. 2005 Jul-Aug;4(4):1086-97

2) Peptides in body fluids and tissues as markers of disease. Schulte I, Tammen H, Selle H, Schulz-Knappe P. Expert Rev Mol Diagn. 2005 Mar;5(2):145-57.

3) http://www.urologychannel.com/bladdercancer/

4) http://www.lakecharlesurology.com/bladder_cancer.asp

5) Discovery of Protein Biomarkers for Renal Diseases. Stephen M. Hewitt, James Dear and Robert A. Star. J Am Soc Nephrol 15:1677-1689, 2004

6) Metabolomics-the link between genotypes and phenotypes.Fiehn O.Plant Mol Biol. 2002 Jan;48(1-2):155-71.

7)A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG.
Nat Biotechnol. 2001 Jan;19(1):45-50.

8) Protein Sequencing and Identeification Using Tandem Mass Spectrometry Kinter M, Sherman N.E. Wiley-Interscience Series on Mass Spectrometry USA 2000

9) Ionization Methods in Organic Mass Spectrometry", Alison E. Ashcroft, The Royal Society of Chemistry, UK, 1997

10) Electrospray Ion-Source - Another Variation on the Free-Jet Theme.
Yamashita, M. and J. B. Fenn. *J. Phys. Chem.* 1984, *88*(20), pp 4451-4459.

11) Analytical Properties of the Nanoelectrospray Ion Source. Wilm, M.; Mann, M.
Anal. Chem. 1996, 68(1), pp 1-8.

12) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymersHillenkamp F, Karas M, Beavis RC, Chait BT..Anal Chem. 1991 Dec 15;63(24)

13) Resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer Brown, R. S. and Lennon, J. J., Mass ". Anal. Chem. 1995, 67, 1998

14) Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry Vestal, M. L., Juhasz, P. and Martin, S. A. Rapid Comm. Mass Spectrom. 1995, 9, 1044

15) http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial.htm

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

17) Computer program (SEQPEP) to aid in the interpretation of high-energy collision tandem mass spectra of peptides. Johnson R.S.Biemann, K.Biomed Environ Mass Spectrom. 1989 Nov;18(11):945-57

18) Four-Sector Tandem Mass Spectrometry of Peptides, A. E. Ashcroft, P. J. Derrick in "Mass Spectrometry of Peptides" ed. D. M. Desiderio, CRC Press, Florida, 1990

19) The identification of novel biomarkers of renal toxicity usingautomatic data reduction techniques and PCA of proton NMR spectra of urine Elaine Holmes a, Jeremy K. Nicholson, Andrew W. Nicholls, John C. Lindon, Susan C. Connor, Stephen Polley, John Connelly - Chemometrics and Intelligent Laboratory Systems 44 1998 245–255

20)Discovery of biomarkers in human urine and cerebrospinal fluid by capillary lectrophoresis coupled to mass spectrometry: Towards new diagnostic and therapeutic approaches Stefan Wittke, Harald Mischak, Michael Walden, Walter Kolch, Thomas Rädler, Klaus Wiedemann. Electrophoresis 2005, 26, 1476–1487

21) Protein Structure Analysis R.M, Choli-Papadopoulou T, Wittmann-Liebold B. Sringer Lab Manual Berlin 1997 Kamp

22)Mass Spectrometry of Proteins and Peptides Chapman J.R. Ed Humana Press Totowa, New Jersey 2000,