Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
BIOCHIMICA, BIOLOGIA MOLECOLARE E BIOTECNOLOGIE

CICLO XXV

COORDINATORE Prof. Francesco Bernardi

"Basic and clinical applications of proteomic research. Our experience"

Settore Scientifico Disciplinare Bio/12

Dottorando
Dott. Ura Blendi

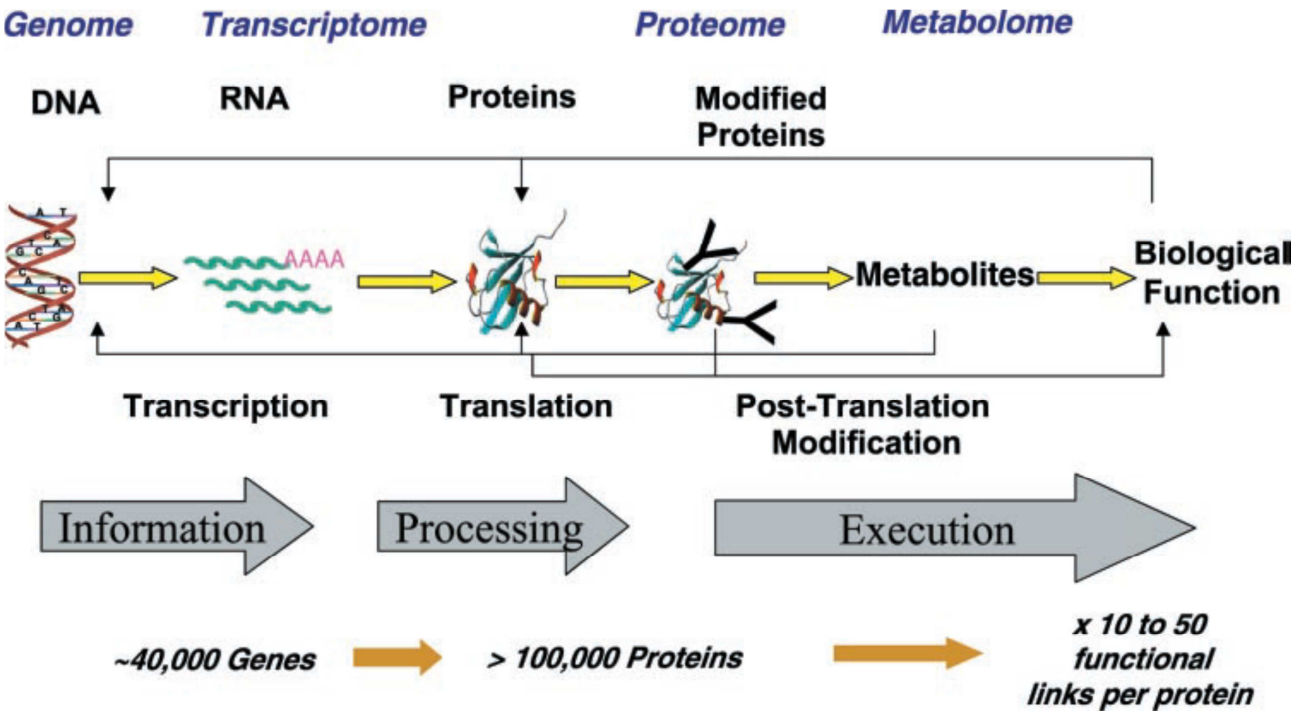
Tutore
Prof. Bergamini Carlo

Anni 2010/2012

Index

GENERAL ISSUES IN PROTEOMICS

In this first chapter we will deal with the basics for biologic and clinical applications of proteomics. The technique was developed in the early ‘90ies as a development of the high resolution bidimensional electrophoretic techniques of protein identification originally set-up by O’Farrell to study expression of proteins in E.coli cells and its coupling with mass spectrometric identification of proteins. At the early time the procedure was applied to basic studies of proteins chemistry, including sequencing and folding as described by Wilkins and associates (1996), who were among the first authors to apply the term proteomics for a general application of the concept to biologic investigations. The potentials of the technique for elucidation of complex issues was definitely apparent when Nair et al (2004) reviewed in Am. J. Physiol. their application in the field of physiology and therefore in pathology. In their review these authors state that this advancement is a consequence of the whole human genome sequencing so that it is possible to identify any protein from their sequence to couple its variation in tissue levels to physiologic pathologic events. Actually these authors state that *“ the term “proteome” refers to the collection of proteins within a cell , tissue, or entire organism and was first coined to describe large-scale protein identification and amino acid analysis. The study of the proteome, or proteomics, has moved beyond cataloguing of proteins in an organism and now includes the comprehensive and detailed analysis of proteins…..”* in an issue that is now appreciated as more complex than genomics because it includes also proteins that are post-translationally modified for regulatory purposes.



Thus, if the main concept is the utilization of proteomics to identify tissular proteins, this i s also a procedure to obtain information about the functions of cells and tissues they are contained in under physiological and pathological conditions as summarised in the following table

Proteomics: objectives for potential application in human research

***1*) Identify proteins or protein involved in diseases**

***2*) Identify modifications of proteins causing complications, e.g., nonenzymatic glycosylation of proteins in diabetes and structural modification (folding changes)**

***3*) Identify molecular pathways involved in functional changes, e.g., signal transduction**

***4*) Target drugs to specific proteins or pathway involved in disease**

At the same time, as it is evident in points 2 and 3 of the table, the potential of the approach is so efficient to include fine modification of the protein structure that are relevant to protein function (i.e. in the field of "functional proteomics"). All this stems from the concept that proteins are indeed functional molecules, and that their study is of major interest to physiologists and physician scientists. The new technologies now available offer new perspectives because one protein may cause cascade effects and change

the synthesis rate, modifications, and folding of several other proteins, so that large scale of proteomics may yield crucial information on regulation of body functions and the mechanism of diseases. However, the proteomic technologies are still not sufficiently precise and sensitive enough to detect subtitle changes in concentrations of proteins that occur during acute physiological interventions. At the same time one must be aware that proteins that are newly synthesized may get degraded immediately (fast turnover of proteins), and changes in detectable levels of concentration may not occur. In addition many protein functions are also altered by modifications (i.e., phosphorylation or glycation) without a change in concentration. Metabolic labeling in vivo with stable isotopes and identification of proteins whose synthesis rates change during an intervention may enable us to detect changes that could not be detected by profiling and quantification alone. Improvements in the techniques to accurately purify proteins from a mixture, quantification of samples with greater precision, and detection of the enrichment of stable isotope tracers in small sample sizes will help to advance our existing capabilities. Further advances in simultaneous detection of modification of multiple proteins in biological fluids and tissue proteins are also essential to fully understand the regulation of functions and the alterations that occur in diseases.

Since the main information that is obtained through proteomic investigations is related to changes in tissue concentration of single proteins, from the biological point of view an immediate interest of the proteomic research is related to the field of modulation of protein expression through variation in their biosynthetic rate. This is exemplified by the issues of Systems biology, that is the analysis of the relationships among the elements in a system in the response to genetic or environmental perturbations, with the goal of understanding the system or the emergent properties of the system. A system may be a few protein molecules carrying out a particular task (such as metabolism of galactose or any other specific metabolite); a complex set of proteins and other molecules physically associated or working together as a molecular machinery (such as the ribosome); a network of proteins operating together to carry out an important cellular function such as giving the cell shape (protein network), or a cell or group of cells carrying out particular phenotypic functions. Thus, a biological system may encompass molecules, cells, organs, individuals, or even ecosystems. System biology will continually improve our capacity to understand and to model biological systems on a more global and in-depth scale than ever before. The systems biology in medicine will be continual with the development of new technologies, which will enhance efficiency, scale and precision with which cellular measurements are

made. This later influence will facilitate all aspects of health care, including the detection and monitoring of diseases, drug discovery, treatment evaluation, and ultimately, predictive and preventative medicine.

As already pointed out a main usefulness of proteomic investigations is related to the analysis of qualitative/ quantitative changes in tissue protein content. These changes are usually ascribed to variations in the cell protein synthetic potential, but this is certainly an oversimplification, since several other effects might contribute to a final pattern documented by the experimental techniques, including chiefly protein breakdown, as in red blood cells. What is now clearly emerging is that - thanks to their high sensitivity and versatility - proteomic techniques can provide answers to complex sets of queries when rationally employed within an experimental design. In the perspective of biological and medical applications, results of proteomic experiments can be interpreted to solve experimental and clinical issues. This subdivision displays certainly borderline effects and is therefore questionable so that we prefer to merge it somewhat in this short discussion of the usefulness of proteomic investigations, taking into account a few selected examples.

This thesis is devoted to these issues and will proceed hrough a short description of the experimental approaches that are employed in proteomic research (Chapter 2), and further with a few examples of topics in the field of experimental biologic and of clinical research that might be faced through proteomic techniques (chapter 3) before entering the experimental investigations I have carried out both in the field of basic and clinical research. The aim is certainly not that to be exhaustive in the treatise but to summarize the few issues I took in account which include protein posttranslational modification (focusing on the action of human transglutaminase on troponin, tubulin and dyphtheria toxin) and issues in cancer research spanning from identification of cancer biomarkers - in relation to breast cancer - to the event of Epidermal Mesenchimal Transition to which now much interest is dedicated because this process is believed to be crucial in determining aggressive behaviour of cancers in mammals.

Some of these results have already been published. Other reports are now in progress to describe completely my efforts in the scientific literature

References

Nair KS, Jaleel A, Asmann YW, Short KR, Raghavakaimal S. Proteomic research: potential opportunities for clinical and physiological investigators. Am J Physiol Endocrinol Metab. 2004 Jun;286(6):E863-74

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, YanJX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, and Hochstrasser DF. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology 14: 61–65, 1996.

Chapter 2

METHODOLOGIES FOR PROTEOME ANALYSIS

The steps involved in basic proteomic investigations include: 1) a pre-analytical stage that involves solubilization of the samples in appropriate solvents and eventually steps of partial purification and enrichment in the particular sub-fractions that are submitted to analysis; 2) an analytical stage itself comprising two phases; in the first the proteins are either resolved from each other by 2D electrophoresis or directly submitted to proteolysis before separation of the released peptide by liquid chromatography; in the second individual spots in the stained 2D gels are digested with proteases and the released peptides or selected peptides from LC chromatograms are fed to matrixes from which the peptides are detached in ionized form and analyzed by MS/MS and TOF to deduce their *m/z* ratios; and 3) final post-analytical stage in which the data derived by the successive fragmentation procedures are compared with those present in PDB to deduce the identity of the protein itself.

These considerations apply to any protein sample including those that have been selected for any peculiarity such as phosphorylated proteins, glycosylated proteins or proteins that are secreted from cultured cells. In other words these kinds of research are extremely flexible and can be applied to any significant experimental system.

The methodologies that are employed in proteomic investigations will now be basically described according to this general scheme and the main variations that are employed when dealing with peculiar samples or with different aims are also briefly outlined.

THE PRE-ANALYTICAL STEPS

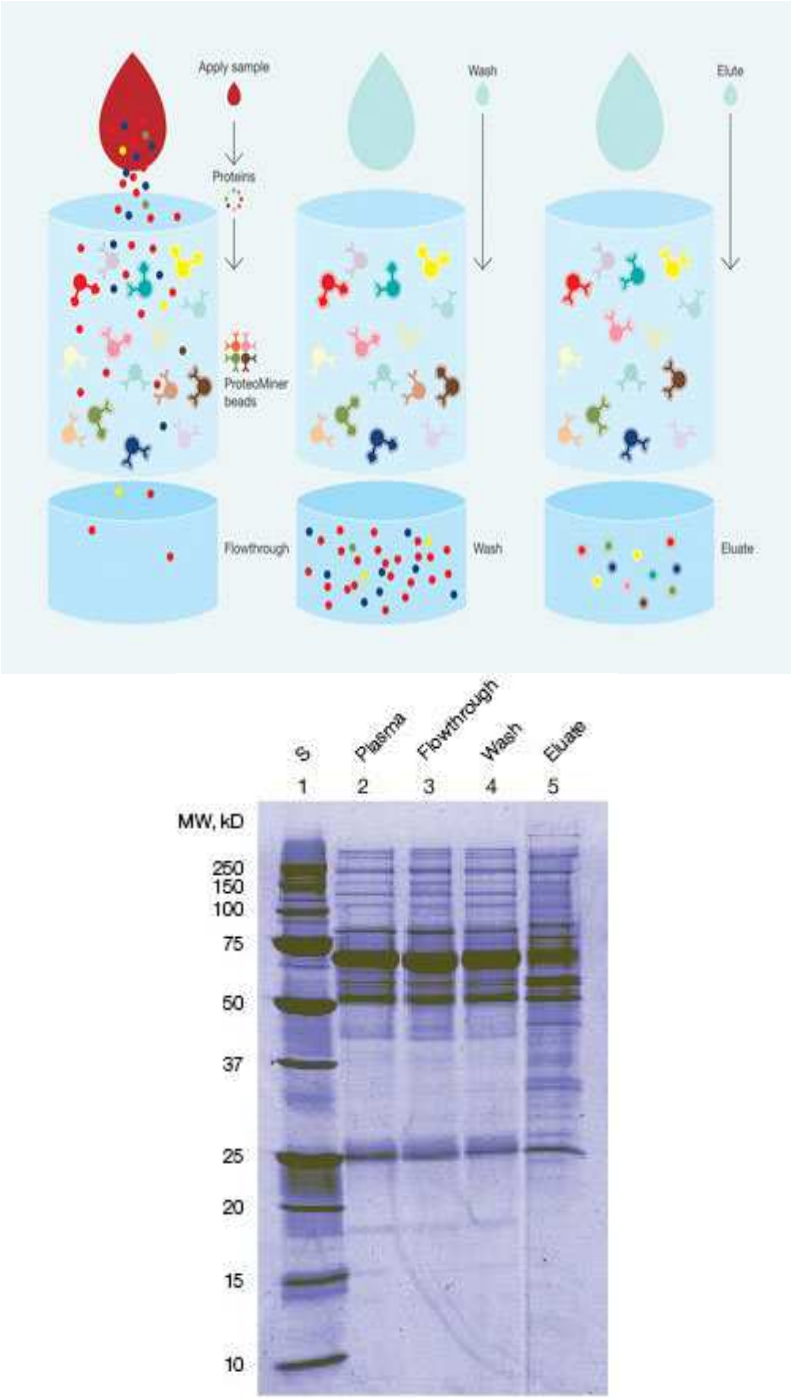
In general terms the sample might be represented by a solution of protein (e.g. plasma), a tissue or even a fixed tissue sample of the type that is employed in pathologic research and diagnosis.

In the case of ***solutions of proteins*** they may be mixed with solubilization buffer at a rather basic pH (around 9.0) containing chaotropic reagents (urea and thiourea) and non ionic detergents (usually Chaps or ASB 14) and mix of proteases inhibitors to avoid digestion of the sample proteins that would be facilitated by the presence of the denaturants that unfold proteins and make them more susceptible to proteolysis. The settlement is followed in the

case the sample has suitable concentration of proteins to allow optimal loading in the subsequent separation step and that the interst of the research is focusing on the major proteins in the sample.

In other conditions it may be important to concentrate the proteins or eventually to enrich the fraction of minor proteins if these represent the aim of the determination.

Procedures for concentration may be represented by lyophilization or high performance precipitation. This last step is usually carried out using solution of strong acids like TCA in acetone. This mix is capable of quantitative protein precipitation even from very dilute protein solutions such as 50 micrograms/ml. Procedures for enrichment of minor proteins in the sample are usually denoted as depletion of major proteins. Examples for these requirements are represented for instance by the analysis of the protein in red blood cell lysates (where the enormous amounts of hemoglobin – about 95% of total proteins - would prevent the possibility to analyze the cytosolic enzymes) or the analysis of plasma where albumin and immunoglobulins would represent about 80% of the total proteins. In these cases removal of the major proteins might be achieved by immunoprecipitation or by chromatographic depletion by cation exchange chromatography of cyano-Methemoglobin. Alternatively the sample might be treated with Proteominer, which is a library of peptides which are immobilized on a matrix in equivalent moderate amounts. This issue is represented schematically below



Schematic representation of Proteominer Treatment using Proteominer beads and SDS-PAGE analysis of Proteominer bead-treated plasma.

The Proteominer technology is a novel sample preparation tool used for the compression of the dynamic range of the protein concentration in complex biological samples. High-abundance proteins present in complex biological samples like sera or plasma, make the detection of medium- and low abundance proteins extremely challenging. This technology provides a method of overcoming this challenge, allowing for the exploration of the entire proteome.

This is accomplished through the use of a large, highly diverse bead-based library of combinatorial peptide ligands. When complex biological samples are applied to the beads, the high-abundance proteins saturate their high affinity ligands and excess protein is washed away. In contrast, the medium- and low abundance proteins are concentrated on their specific affinity ligands. This reduces the dynamic range of protein concentrations while maintaining representatives of all proteins within the original sample.

In the case the sample is represented by ***a tissue or isolated cells*** (these are models that have been analysed in the present thesis) the proteins must be extracted from the cells by homogenization with the same lysis buffer as above employed mechanical protein dispersion with a Potter-Elverhjem apparatus in the case of solid tissues or vortexing and long-lasting stirring in a rotatory wheel in the case of isolated cells (e.g. blood cells or the cholangio-carcinoma cells we have submitted to proteomic analysis, see below). After these steps of protein extraction, cellular debris, collagne or DNA are removed by centrifugation or syringe needle squeezing, depending on the individual requirements.

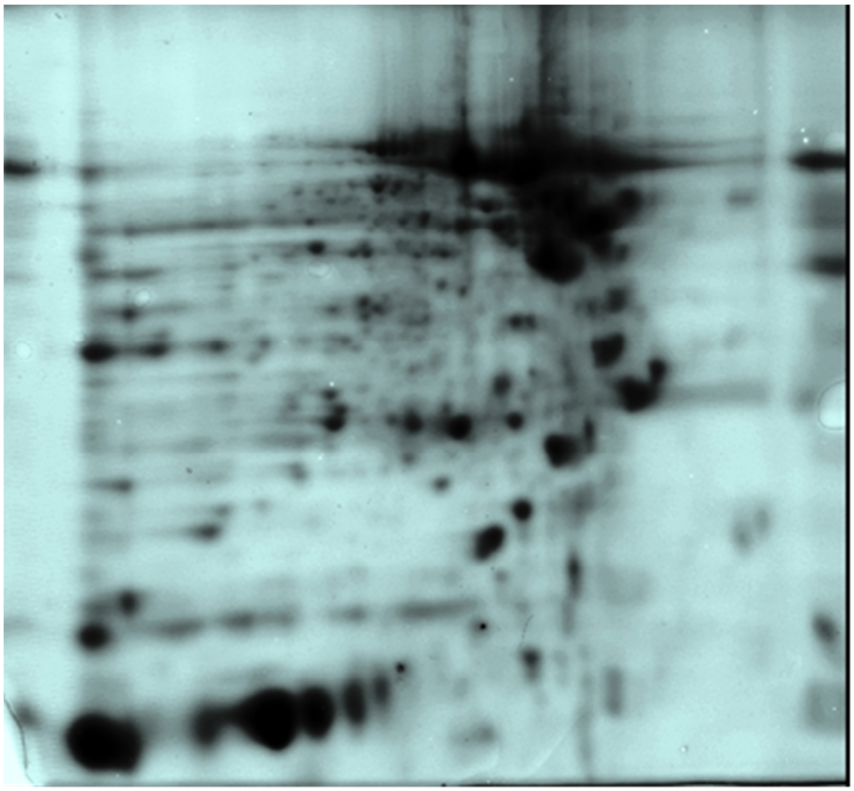
If the proteins are contained in a fixed sample as the Formalin Fixed Paraffin Embedded (FFPE) tissues that are employed in pathologic research tissue slices are extracted with solutions containing SDS and reducing agents at 100 to 80 °C and the detergent is removed by Acid Acetone protein precipitation. Released proteins are thereafter separated by IEF-SDSPAGE. Alternatively for shotgun approaches slices of FFPE tissues are treated with SDS-DTT-ammonium carbonate buffer at lower temperature after hexane extraction to remove paraffin (this treatment effectively breaks down protein crosslinks by formalin) and treated with trypsin in the presence of high concentrations of $CaCl_2$. In this way soluble peptides are released that can be directly analysed by LC-MS.

THE ANALYTICAL STEPS

These obviously vary according to the 2D electrophoretic or to the LC approach. In the first instance the protein mixture is resolved in a first dimension by IsoElectricFocusing (IEF) and in a second dimension by SDS-PAGE in a slab gel, which is stained by a procedure compatible with digestion by proteinases to identify and select isolated proteins to be analyzed, directly within gel fragments (in situ digestion). The released peptides are then submitted to MS analysis. In the LC approach instead the digestion is performed directly on the sample before resolving the peptides by hplc. The main difference is therefore that the first approach (the one we have employed in the present research),

provides directly some basic information on the differential tissue protein expression, which is not available for the LC-MS approach unless procedure of difference sample labeling are employed.

High-resolution two-dimensional electrophoresis (2-D PAGE) for the separation of complex protein mixtures was introduced by (O'Farrell et.al 1975) by combining isoelectric focusing (IEF) in the first dimension in presence of urea, detergents and DTT, with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension.



Representation of 2D-electrophoresis of breast cancer tissue

Proteins are separated according to isoelectric point (pI) and molecular mass (Mr), and quantified according to relative abundance. Depending on the gel size and pH gradient used, 2-D PAGE can resolve more than 5,000 proteins simultaneously (~2,000 proteins routinely), and can detect <1 ng of protein per spot.

Furthermore, it delivers a map of intact proteins, which reflects changes in protein expression level, isoforms or post-translational modifications. An additional strength of 2-D PAGE is its capability to study proteins that have undergone some form of post-translational modification (such as phosphorylation, glycosylation or limited proteolysis) and which can – in many instances - be readily located in 2-D gels as they appear as distinct 'spot trains' in the horizontal and /or vertical axis of the 2-D gel. In addition, 2-D

PAGE not only provides information on protein modifications and/or changes in their expression levels, but also permits the isolation of proteins for further structural analyses by MALDI-ToFMS or ESI-MS .

Despite these benefits, a major problem of 2-D PAGE has been the exchange of 2-D gel data between laboratories due to spatial irreproducibility between 2-D gels generated by the conventional method of 2-D PAGE using carrier ampholyte (CA) IEF. Equilibrium CAIEF cannot be achieved because of pH gradient instability with prolonged focusing time, as the pH gradient moves towards the cathode ('cathodic drift') and flattens in the centre ('plateau phenomenon'). Consequently, time-dependent protein patterns are obtained. In addition, reproducibility of pH gradient profiles is limited by the batch-to-batch variability of CA preparations.

The problems of pH gradient instability and irreproducibility were overcome by the introduction of immobilized pH gradients (IPG) for IEF (Bjellqvist et al. 1982), (Görg et al. 1988). IPGs are based on the principle that the pH gradient is generated by a limited number (<10) of well-defined chemicals (the 'Immobilines') which are co-polymerized with the acrylamide matrix. Thus cathodic drift is eliminated, reproducibility enhanced and pattern matching and inter-laboratory comparison simplified. IPGs allow the generation of pH gradients of any desired range (broad, narrow or ultra-narrow) between pH 2.5 and 12. Since sample loading capacity of IPG-IEF is also higher than with CA-IEF, especially in combination with narrow (1 pH unit) or ultra-narrow (0.1 pH unit) IPGs, 2-D PAGE with IPGs is the method of choice for micropreparative separation and spot identification.

The protein spots that are considered to vary in the sample according to the experimental parameters taken into account are identified in stained gels (either colloidal Coomassie stain or Silver nitrate staining are acceptable) are then digested with trypsin or other suitable proteinases (e.g. staphylococcal V8 proteinase) with high sequence and amino acid specificity.

As already mentioned in the ***LC-MS approach*** the peptides that will be later employed for protein identification are generated by digestion of the soluble proteins in the sample extract at an early step, before any attempt to fractionation procedures. LC-MS is used in proteomics where again components of a complex mixture must be detected and identified in some manner. The LC-MS approach to proteomics generally involves protease digestion and denaturation (usually trypsin as a protease, urea to denature tertiary structure and iodoacetamide to cap cysteine residues) followed by LC-MS with peptide mass fingerprinting or LC-MS/MS (tandem MS) to derive sequence of individual peptides. LC-

MS/MS is most commonly used for proteomic analysis of complex samples where peptide masses may overlap even with a high-resolution mass spectrometer. Samples of complex biological fluids like human serum may be run in a modern LC-MS/MS system and result in over 1000 proteins being identified.

Once the peptides have been obtained they are submitted to HPLC chromatographic which is technique used to separate a mixture of compounds in analytical chemistry and biochemistry with the purpose of identifying, quantifying and purifying the individual components of the mixture. HPLC typically utilizes different types of stationary phases (i.e. sorbents) contained in columns, a pump that moves the mobile phase and sample components through the column, and a detector capable of providing characteristic retention times for the sample components and area counts reflecting the amount of each analyte passing through the detector. The detector may also provide additional information related to the analyte, (i.e. UV/Vis spectroscopic data, if so equipped). Analyte retention time varies depending on the strength of its interactions with the stationary phase, the composition and flow rate of mobile phase used, and on the column dimensions. HPLC is a form of liquid chromatography that utilizes small size columns (typically 250 mm or shorter and 4.6 mm i.d. or smaller; packed with smaller particles), and higher mobile phase pressures compared to ordinary liquid chromatography. With HPLC, a pump (rather than gravity) provides the higher pressure required to move the mobile phase and sample components through the densely packed column. The increased density arises from the use of smaller sorbent particles. Such particles are capable of providing better separation on columns of shorter length when compared to ordinary column chromatography.

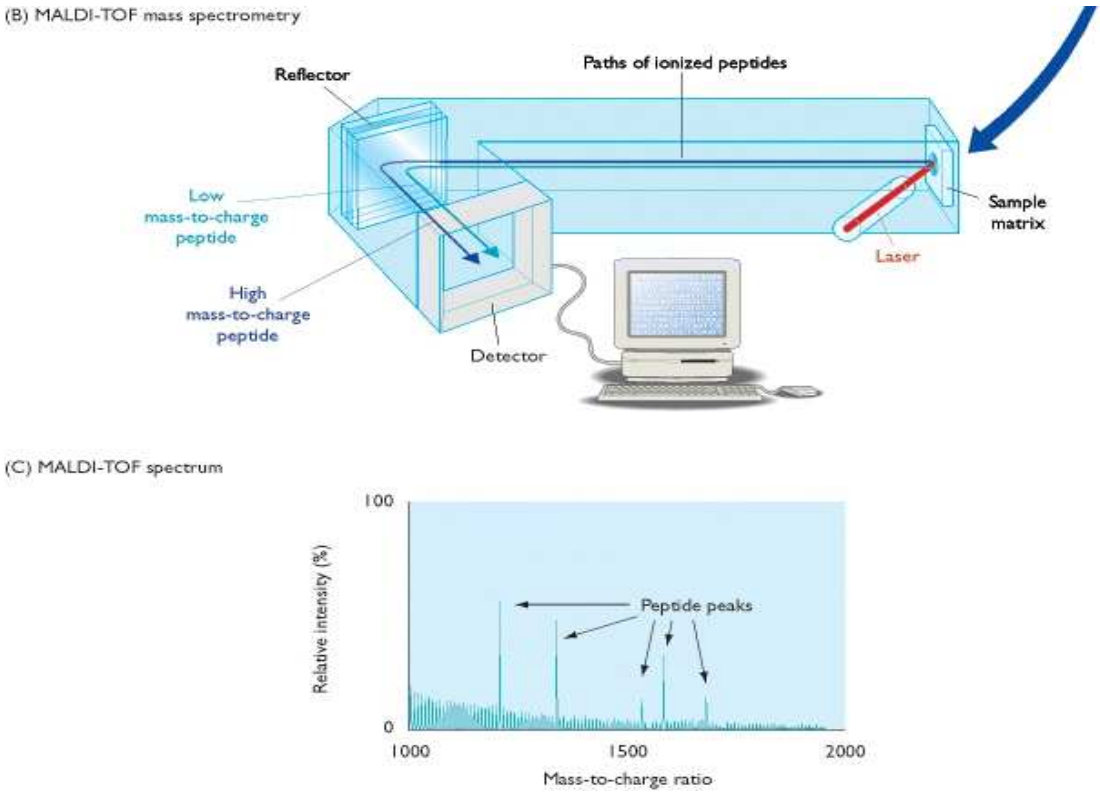
In this instance the peptides that are resolved during the chromatographic procedure are directly to the mass spectrometer.

The identification of the parent protein is then achieved by ***Mass Spectrometry*** of which several types have been described based basically on the procedure employed to generate ionized peptides on which *m/z* ratio are determined. They are mainly the MALDI and the ESI procedures.

In the case of ***Matrix-assisted laser desorption/ionization (MALDI)*** **t**he technique used in mass spectrometry is a soft ionization procedure allowing the analysis of biomolecules (biopolymers such as DNA, proteins, peptides and sugars) and large organic molecules (such as polymers and other macromolecules), which tend to be fragile and fragment when ionized by more conventional ionization methods. It is similar in character to electrospray ionization both in relative softness and the ions produced (although it causes many fewer

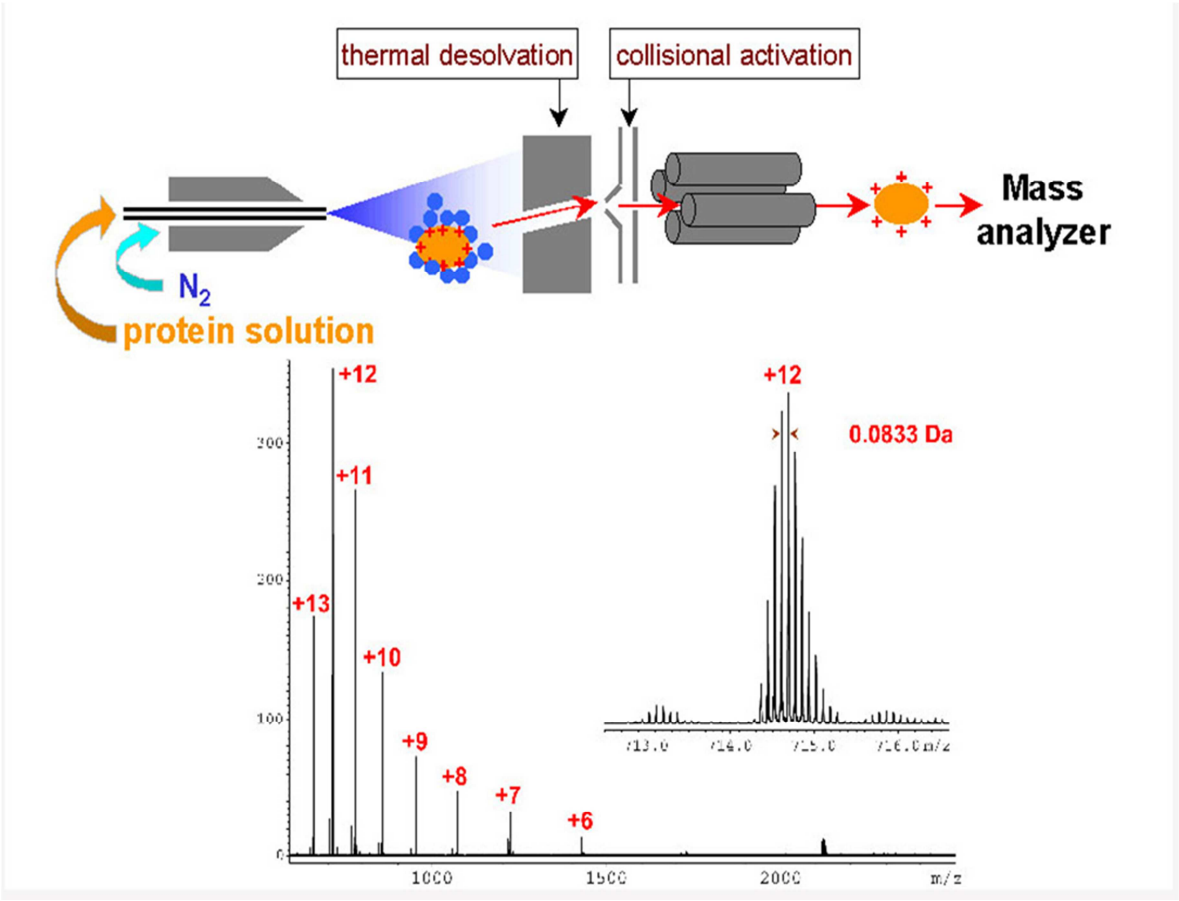
multiply charged ions. MALDI is a two step process. In the first, the sample is absorbed on a matrix, from which desorption (the second step) is triggered by a UV laser. Matrix material heavily absorbs UV laser light, leading to the ablation (removal of material from the surface of an object by vaporization, chipping, or other erosive processes). of upper layer (~micron) of the matrix material. A hot plume produced during the ablation contains many species: neutral and ionized matrix molecules, protonated and deprotonated matrix molecules. The second step is ionization (more accurately protonation or deprotonation). Protonation (deprotonation) of analyte molecules takes place in the hot plume. Some of the ablated species participate in protonation (deprotonation) of analyte molecules.

The ***Matrix*** is a crucial component whose function is adsorption of energy from laser pulse, and then transfer to sample thereby causing desorption of the analyte molecules in an expanding plume, to ionize the desorbed analyte molecules and to prevent aggregation of the analyte molecules. The matrix molecules for MALDI are chosen on the basis of fulfillment of requirement that matrix molecules must be able to absorb ultra violet wavelength of usually 237 nm, low volatility and ability to transfer protons to the sample molecules. For proteins samples typical MALDI matrix consist of cinnamic acid and hydroxylated benzoic acid derivatives. The whole process is summarized below. The signal that is generated and analysed in this kind of mass spectrometry is known as ***Time-of-flight mass spectrometry (TOFMS)*** which is a method of mass spectrometry in which an ion's mass-to-charge ratio is determined via a time measurement. Ions are accelerated by an electric field of known strength. This acceleration results in an ion having the same kinetic energy as any other ion that has the same charge. The velocity of the ion depends on the mass-to-charge ratio. The time that it subsequently takes for the particle to reach a detector at a known distance is measured. This time will depend on the mass-to-charge ratio of the particle (heavier particles reach lower speeds).



Representation of MALDI-TOF

An alternative procedure for ionization of peptides and feeding to the mass spectrometer is ***Electrospray ionization (ESI)*** which is a technique used to produce ions in mass spectrometry. It is especially useful in producing ions from macromolecules because it overcomes the propensity of these molecules to fragment when ionized. The liquid containing the analyte(s) of interest is dispersed by electrospray into a fine aerosol. Because the ion formation involves extensive solvent evaporation, the typical solvents for electrospray ionization are prepared by mixing water with volatile organic compounds (e.g. methanol, acetonitrile). To decrease the initial droplet size, compounds that increase the conductivity (e.g. acetic acid) are customarily added to the solution. Large-flow electrosprays can benefit from additional nebulization by an inert gas such as nitrogen. The aerosol is sampled into the first vacuum stage of a mass spectrometer through a capillary, which can be heated to aid further solvent evaporation from the charged droplets. The solvent evaporates from a charged droplet until it becomes unstable upon reaching its Rayleigh limit. At this point, the droplet deforms and emits charged jets in a process known as Coulomb fission. During the fission, the droplet loses a small percentage of its mass (1.0-2.3%) along with a relatively large percentage of its charge (10-18%) (Kebarle P.et.al 2009).



Schematic representation of ESI-MS

Ion Trap

An **ion trap** is a combination of electric or magnetic fields that captures ions in a region of a vacuum system or tube. Ion traps have a number of scientific uses such as mass spectrometry and trapping ions while the ion's quantum state is manipulated. The two most common types of ion traps are the Penning trap and the Paul trap (quadrupole ion trap). When using ion traps for scientific studies of quantum state manipulation, the Paul trap is most often used. An ion trap mass spectrometer may incorporate a Penning trap (Fourier transform ion cyclotron resonance),(Blaum,Klaus. 2006) Paul trap or the Kingdon trap. Other types of mass spectrometers may also use a linear quadrupole ion trap as a selective mass filter. These procedures are complicate and will not further dealt with in this study which is amied to demonstrate applicability of proteomic procedures to solve biologically or clinically relevant issues.

The complexity of the signals that are obtained is particularly evident during attempts of ***Gel-independent quantitative profiling of the proteome.*** The combination of LC-MA/MS and sequence database searching has been widely adopted for the analysis of complex peptide mixtures generated from the proteolysis of samples containing several proteins. This approach is often referred to as 'shotgun' proteomics and has the ability to catalog hundred, or even thousands, of components contained in sample isolated from very different sources. Specific examples include the identification of proteins in the periplasmic space of bacteria, yeast ribosomal complexes, murine nuclear interchromatin granule cluster, human urinary proteins, proteasomal proteins etc.

Such studies have also highlighted the limitations of shotgun proteomics, including, the difficulty of detecting by collision-induced dissociation (CID) mass spectrometry all of the peptides in a sample, the qualitative nature of data-dependent experiments, and the challenge of processing the tens of thousands of CID spectra generated in a typical experiment. On average, a protein digested with trypsin will generate 30-50 different peptides. A tryptic digest of the of a typical human cell will therefore generate a peptide mixture containing at least hundreds of thousands of peptide. Even the most advanced LC-MS/MS system cannot resolve and analyze such complexity in a reasonable amount of time.

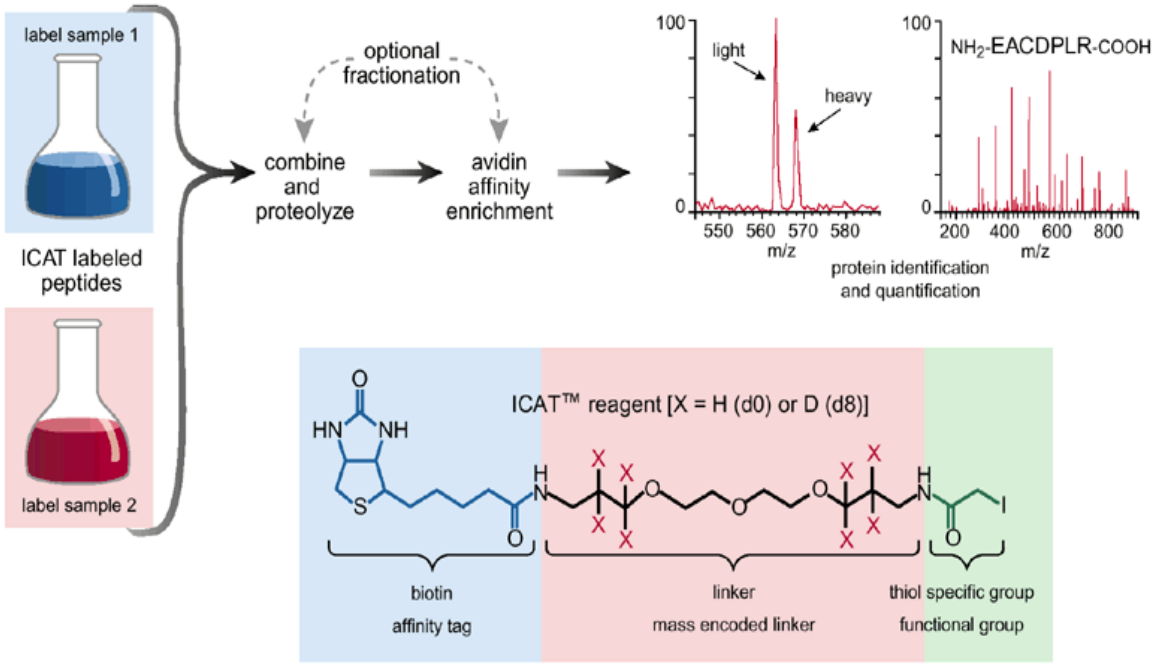
To use LC-MS/MS for the analysis of most proteomes, therefore, a form of complexity reduction (fractionation) is required. Two approaches have been developed to tackle this problem. The first approach is the selective enrichment of a subset of peptides from a complex mixture. This has been mostly achieved by specifically targeting peptides that contain a distinguishing feature, such as chemically reactive sulfhydryl groups (cysteine residues) or residues that has been modified post-translationally with phosphate or carbohydrate. Ideally, such strategies would select precisely one idiotypic peptide per protein. Although this has not been realized, substantial reductions in same complexity have been achieved. The trade-off is the loss of proteins that do not contain the selected feature.

The second approach relies on extended upstream fractionation of the complex peptide mixtures, with the aim of increasing the potential of the mass spectrometer to detect and to sequence all of the components of the sample. This has been implemented by using two or three orthogonal peptide separation methods in sequence, the most powerful of which have involved cation exchange and capillary reverse-phase chromatography. Although both approaches either by themselves or in combination have produced evidence that proteome

cataloguing using such proteomic strategies is feasible, a complete map of the proteome of any species has yet to be produced by any method.

To turn shotgun proteomics into a quantitative protein profiling method, therefore, stable isotope dilution has been combined with the complexity reduction technique.

The first such approaches was based on a class of reagent termed ‘isotope-coded affinity tags or ICAT’, LC-MS/MS and sequence database searching. The reagents consist of an alkylating group (iodoacetic acid) that covalently attaches the reagent to reduced cysteine residues, a polyether mass-encoded linker containing either eight hydrogen (d0) or eight deuterium (d8) that represents the isotope dilution and a biotin affinity tag through which tagged peptides are selectively isolated.



Quantitative proteomics using ICAT reagents.

After optional protein enrichment and enzymatic digestion of the combined samples, the biotinylated ICAT-labeled peptides are enriched by means of avidin affinity chromatography and analyzed by LC-MS/MS. Each cysteinyl peptide appears as a pair of signals differing by the mass differential encoded in the mass tag. The ratio of these signal intensities precisely indicates the ratio of abundance of the protein from which the peptide originates, and the MS/MS spectrum of either isotopic form of the peptide allows the protein to be identified. Thus, in a single, automated operation this method identifies the proteins present in two related samples and determines the ratio of relative abundance.

Although isotope-tagging methods based on chemical labeling after isolation are compatible with essentially any protein sample, including organelles, body fluids or subcellular and biochemical fractions, the application of metabolic labeling is limited to those situations in which cells can be in isotopically defined media.

THE POST-ANALYTICAL STEPS

These are represented by the computational approaches that are employed for the identification of any protein structure from the fragments (peptides) that have been evidenced during the fragmentation and mass spectrometry steps. Although additional procedures have been developed, that we have relied on is the ***Mascott software*** which is a powerful search engine which uses mass spectrometry data to identify proteins from primary sequence databases. By using an appropriate scoring algorithm, the closest match or matches can be identified. If the "unknown" protein is present in the sequence database, then the aim is to pull out that precise entry. If the sequence database does not contain the unknown protein, then the aim is to pull out those entries which exhibit the closest homology, often equivalent proteins from related species. The algorithm MOWSE is more selective and sensitive than other algorithms calculating only number of matching peptides. Calculating estimates it takes into consideration both relative excess of peptides of the same molecular mass in the database and size of protein. Program MASCOT is based on the MOWSE algorithm; this program also evaluates a possibility of random matching of experimental and theoretical peptide masses. Using sequential queries MASCOT can optimize search parameters. Using this program it is possible to identify proteins in relatively simple mixtures.

Mascot is a search engine based on mass spectrometry data to identify proteins using primary sequence databases. There are primarily three types of searches namely peptide mass fingerprint method (based on a list of peptide mass values generated from an enzymatic digest of a protein) sequence query method (based on one or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses) and MS/MS ion search method (based on identification based on raw MS/MS data from one or more peptides). Mascot looks for the highest scoring set of peptide matches which are within a contiguous stretch of sequence less than or equal to the specified protein molecular weight. The optimum data set for a peptide mass fingerprint is, of course, all of the correct peptides and

none of the wrong ones. The sequence query mode of Mascot supports both standard and error tolerant sequence tags. It also allows arbitrary combinations of fragment ion mass values, amino acid sequence data and amino acid composition data to be searched.

Given an absolute probability that a match is random, and knowing the size of the sequence database being searched, it becomes possible to provide an objective measure of the significance of a result. A commonly accepted threshold is that an event is significant if it would be expected to occur at random with a frequency of less than 5%. This is the value which is reported on the master results page. The master results page for typical peptide mass fingerprint search reports that "Scores greater than 67 are significant ($p<0.05$)". The protein with the high score of 108 is a 26 kDa heat shock protein from yeast. This is a nice result because the highest score is highly significant, leaving little room for doubt.

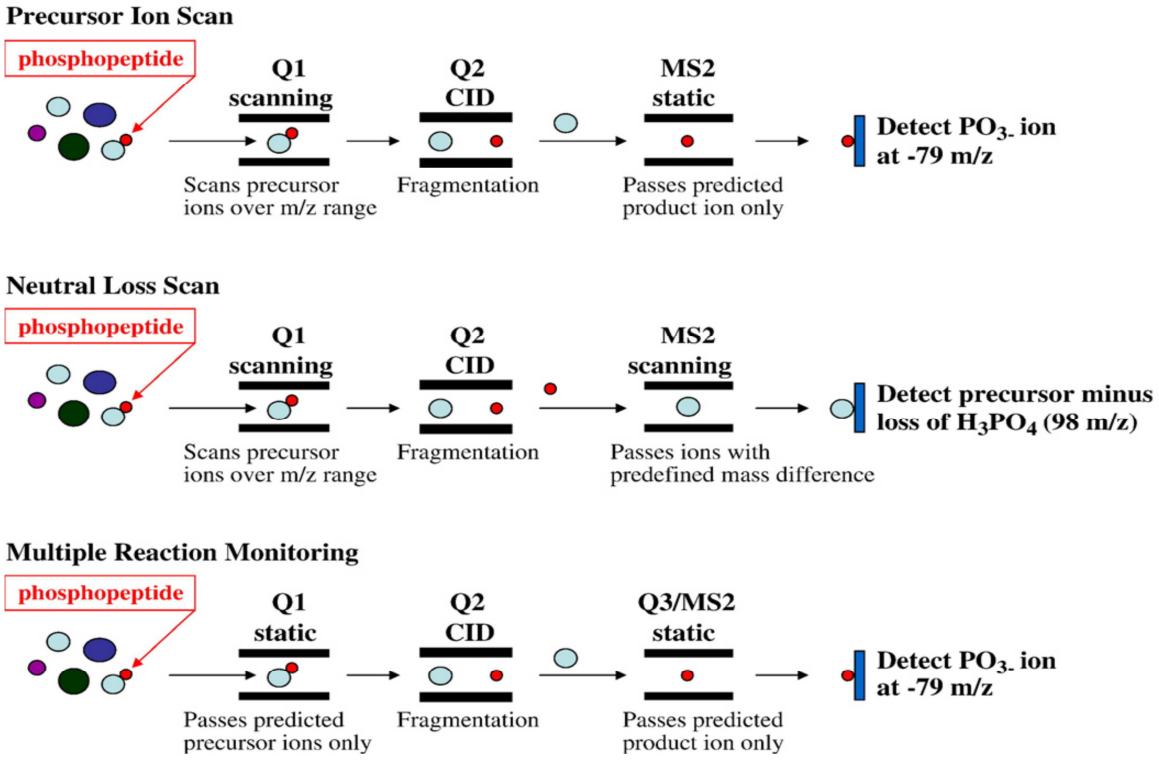
SOME SPECIALIZED APPLICATIONS OF PROTEOMICS

One of these is represented by the study of ***protein phosphorylation*** which has prominent importance in the biological events of signal transduction. For these purposes identification of phosphopeptides and of site(s) of phosphorylation has been achieved using either MALDITOF or electrospray mass spectrometry. For the analysis of phosphopeptides by MALDI-TOF, the peptide must ionize at high level after co-crystallization with a UV absorbing matrix as 4-hydroxycyanno-cinnamic acid or 2,5-dihydroxybenzoic acid. Phosphopeptide ions undergo a characteristic loss of phosphoric acid after ionization by a process referred to as post source decay (PSD) and when analyzed by reflectron MALDI TOF MS, the poorly resolved metastable fragments can be observed (R.S. Annan et.al 1996). By using the new generation MALDI-TOFTOF mass spectrometers the phosphopeptide ions can be fragmented by either PSD or collision induced dissociation (CID) to generate MS/MS spectra and the position of phosphorylation can be determined from the fragmentation series.An alternative to this is to use a MALDI-QqTOF mass spectrometer with 2,5-dihydroxybenzoic acid as the matrix, which has been reported to yield higher quality MS/MS spectra than that obtained by a MALDI-TOF-TOF mass spectrometer and hence makes the assignment of the site of phosphorylation easier.

The main limitation of using MALDI-TOF mass spectrometry for the detection and sequencing of phosphopeptides is they are often difficult to detect in unfractionated protein digests.

Detection of phosphopeptides by specific mass spectral scanning methods based on either the neutral loss of phosphoric acid from phosphoserine or phosphothreonine (neutral loss

scan) during collision induced dissociation (CID) in positive mode, or the generation of a phosphate specific reporter ion after collision induced dissociation in negative mode (precursor ion scan).



Scan modes used for phosphorylation site mapping.

Additional interesting applications are represented by the ***interaction proteomics*** and the ***subcellular proteomics***. The former is investigated taking advantage of immunoprecipitation studies are another useful way to exploit the resolving power of proteomics, dealing with the protein-protein interactions in cells and in lysates that might play important roles in metabolic regulation. In this instance, very large quantities of protein can be subjected to incubation with antibodies against chosen signaling molecules. This allows high-affinity capture of these proteins, which can subsequently be eluted and electrophoresed on a 2D gel to provide a high-resolution proteome of a specific subset of proteins. Detection by blot analysis allows the identification of extremely small amounts of defined signaling molecules. Again, the different isoforms of even very low abundance proteins can be seen, and, very importantly, the technique allows the investigator to identify multiprotein complexes or other proteins that co-precipitate with the target protein. These coassociating proteins frequently represent signaling partners of the target proteins, and their identifications by mass spectrometry can lead to invaluable information on the signaling processes involved.

The depth of signal transduction analysis offered by proteomics, and the utility for target validation studies can be extended even further by applying cell fractionation studies (Corthals, G,L et.al 1997). By purifying subcellular fractions, such as membrane, nuclear, organelle and cytosolic, it is possible to assign a localization to proteins of interest and to follow their trafficking in a cell. Enrichment of these fractions will also allow much higher representation of low abundance proteins on the proteome. These signal transduction analyses can be of additional value in experiments where inhibitors derived from a screening program against the target are being evaluated for their potency and selectivity. The inhibitors can encompass small molecules, antisense nucleic acid constructs, dominant-negative proteins, or neutralizing antibodies microinjected into cells. In each case, proteome analysis can provide unique data in support of validation studies for a chosen candidate drug target.

Another approach which has been of special interest in our laboratory is that of ***Secretomics***, that might be useful for the identification of several clinically relevant biomarkers as for instance in the case of cancer proteomics. "Secretome" is referred to as the rich, complex set of molecules secreted from living cells, the term also includes molecules shed from the surface of living cells. Proteins of secretome play a key role in cell signaling, communication and migration. The need for developing more effective cancer biomarkers and therapeutic modalities has led to the study of cancer cell secretome as a means to identify and characterize diagnostic and prognostic markers and potential drug and therapeutic targets.

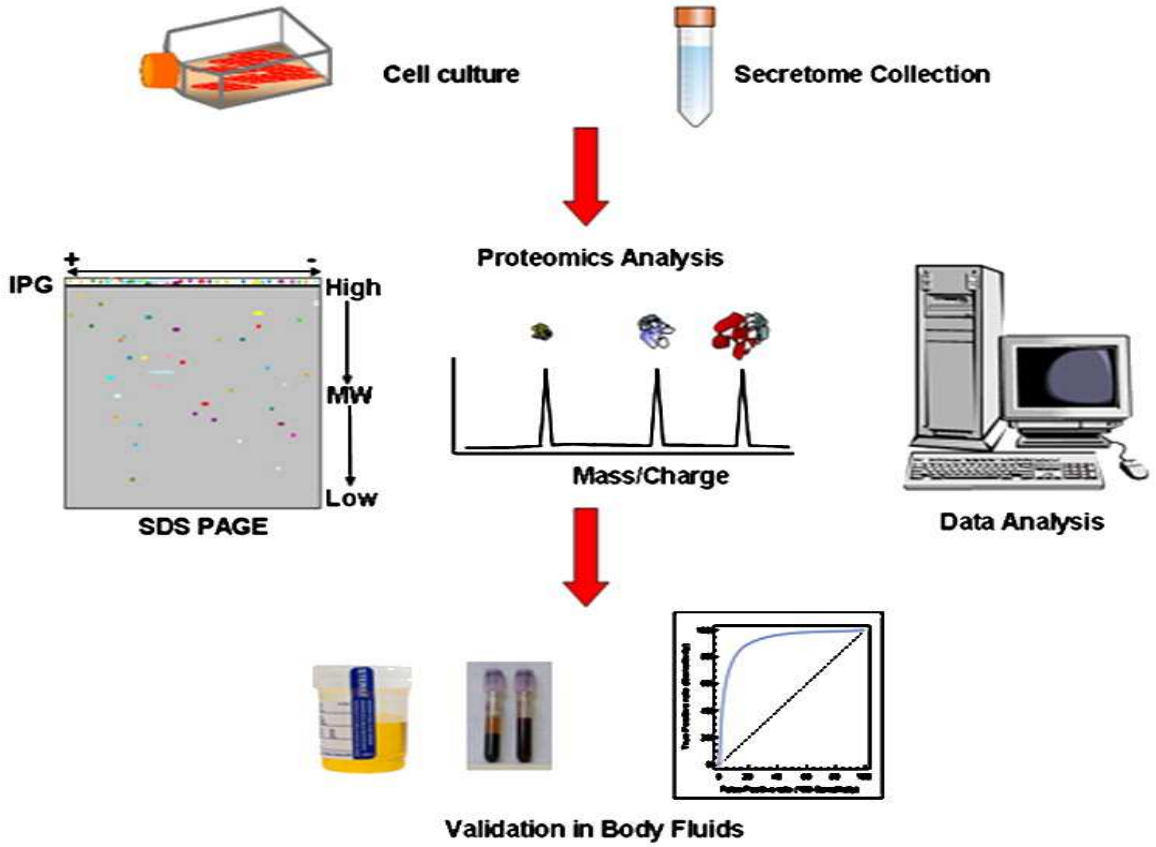
Secreted proteins are responsible for the cross talking among cells and understanding this language could largely increase our knowledge on the molecular mechanisms of neoplasia. In addition, extracellular matrix components and other molecules secreted by tumor cells are a rich source of potential markers and drug targets for cancer treatment. Secreted proteins are also most likely excellent candidate serological tumor markers as they are released by the cells thus having highest possibility to enter the circulation.

Analysis of secreted proteins is quite challenging; this may be attributed to technical difficulties such as a) presence of the proteins at frequently low concentrations due to their high dilution (in body fluids or cell culture media), b) their masking and contamination by cytoplasmic or other normally non secreted proteins released following cell lysis and death, and c) masking by serum proteins (i.e. fetal bovine serum) normally present in the culture media (Chevallet M,et.al 2007). These issues bunderscore the importance of procedures that have to be followed for cell culture and conditioned medium (CM)

collection so as to avoid contamination from dead cells and cytosolic proteins as well as for effective protein concentration, regardless the protein separation technology in use.

PROTEOMICS SEPARATION TECHNIQUES FOR SECRETOME ANALYSIS

A variety of proteomics methodologies have been utilized for secretome analysis including 1D and 2D electrophoresis and liquid chromatography, all of them coupled with different types of MS analysis. Each of these methods has its own advantages and disadvantages with regards to reproducibility, quantification, identification of differentially expressed proteins, sensitivity and robustness.



Schematic diagram of the experimental secretomic strategy.

Recently, was performed a systematic comparison of 3 separation strategies prior to nano LC-MS/MS analysis: 1D gel electrophoresis (1DGE), peptide SCX (strong cation exchange) chromatography and tC2 protein reversed phase chromatography. 1DGE outperformed the SCX and tC2 methods with respect to the number of identified proteins, reproducibility in protein identification, and throughput. A variation of 1DGE for the study of secreted proteins were run on precast or home made SDS-PAGE gels and the

electrophoresis was stopped after the sample had barely passed from stacking into the resolving gel. All sample proteins were concentrated in a tight band at the top of the resolving gel, stained with Coomassie blue, excised, digested with trypsin and processed for MS analysis. This methodology increased the throughput of secretome analysis by 1 order of magnitude compared to previous methodologies.

All currently known biomarkers are secreted or shed proteins (Kulasingam V.et.al.2005), which makes cancer secretome analysis quite promising in biomarker research. The table below review the main findings from the analysis of cancer cell secretome with respect to identification of potential disease biomarkers.

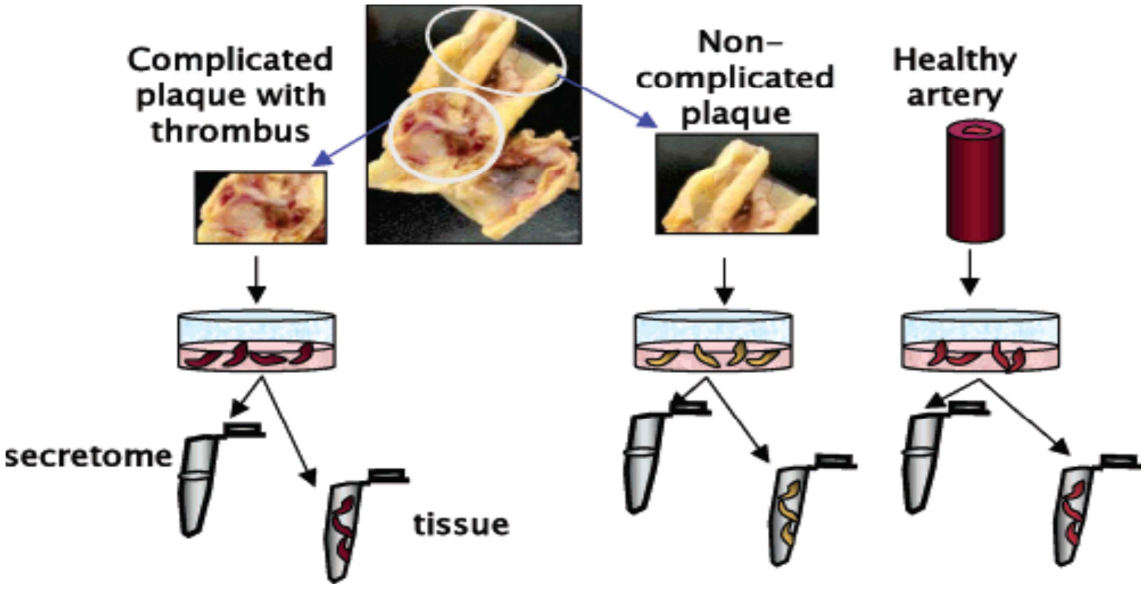
Cell line	Malignancy	Screening methods	Validation methods	Biological fluid	Candidate biomarkers
RT112, T24	Bladder	LC-MS/MS	IHC, Western blot	Tissue, urine	CXCL1
U1, U4	Bladder	SDS-PAGE/ MALDI-TOF MS	Western blot	Urine	Pro-u-PA
MCF-10A, BT474, and MDA-MB-468	Breast	LC-MS/MS	ELISA	Serum, tissue	Elafin, KLK6, ALCAM
TIF, NIF	Breast	2DE/ MALDI-TOF MS	Tissue Microarrays	Tissue	Calreticulin, cellular retinoic acid-binding protein II, CLIC-1, EF-1-beta, galectin 1, peroxiredoxin-2, platelet-derived endothelial cell growth factor, protein disulfide isomerase and ubiquitin carboxyl-terminal hydrolase 5
SW480 SW620,	Colorectal	LC-MS/MS	IHC, ELISA	Tissue, serum	TFF3, GDF1L5
Colo205, SW480 SW620, LS174T	Colorectal	SDS-PAGE/ MALDI-TOF MS	IHC, ELISA	Tissue, plasma	CRMP-2, Mac-2BP
SW948, SW620, SW480, HT29, CaCo2	Colorectal	2DE/ MALDI-TOF MS, Western blot	Western Blot	Serum	Glod 4, C-terminal agrin
HuCCA-1, HCC-S102, HepG2, SK-Hep-1, Alexander	Cholangiocarcinoma	SDS-PAGE/ LC-MS/MS	Western blot	Tissue	Lipocalin 2
SK-Hep-1, Hep G2, Hep-3B	Hepatocellular	SDS-PAGE/ LC-MS/MS	ELISA	Plasma	CD14
L1236, KMH2, L428, DEV	Hodgkin Lymphoma	SDS-PAGE/ LC-MS/MS	IHC, ELISA	Tissue, plasma	ALCAM, cathepsin S, CD26, CD44, IL1R2, MIF, TARC
H23, H520, H460, H1688	Lung	LC-MS/MS	ELISA	Serum	ADAM 17, osteoprotegerin, pentraxin 3, follistatin, tumor necrosis factor receptor superfamily member 1A
Lung cancer primary cells	Lung	SDS-PAGE/ ESI MS/MS	ELISA	Plasma	MMP-1, 14-3-3 sigma, beta, eta, fascin, LAMC2
CL_{1-0}, CL_{1-5}	Lung	SDS-PAGE/ LC-MS/MS	IHC, ELISA	Tissue, serum	RbAp46
M-BE	Lung	2DE/ MALDI-TOF MS	Tissue Microarrays, ELISA	Tissue, plasma	Cathepsin D
NCI-H226, H226Br	Lung	SDS-PAGE/ MALDI-TOF MS	ELISA	Serum	LDHB
1198, 1170-I, BEAS-2B, 1799	Lung	2DE/ MALDI-TOF MS	Western blot, ELISA	Tissue, plasma	PGP9.5, TCTP, TIMP-2, TPI
CL_{1-0}, CL_{1-5}	Lung	SDS-PAGE/ LC-MS/MS	ELISA	Plasma	SDF-1/CXCL12
NPC-TW02, NPC-TW04	Nasopharyngeal	SDS-PAGE/ MALDI-TOF MS	ELISA	Serum	Fibronectin, Mac-2BP, PAI-1
NPC-TW04	Nasopharyngeal	SDS-PAGE/ LC-MS/MS	IHC, ELISA	Tissue, plasma	CLIC-1
NPC-TW02, NPC-TW04 NPC-BM1	Nasopharyngeal	SDS-PAGE/ LC-MS/MS	ELISA	Plasma	Cathepsin L1, ISG15
OEC-M1, SCC4	Oral	SDS-PAGE/ MALDI-TOF MS	IHC, ELISA	Tissue, serum	Mac-2BP
HTB75, TOV-112D, TOV-21G, RMUG-S	Ovarian	LC-MS/MS	ELISA	Serum	Clusterin, IGFBP6

HPDE, Panc1	Pancreatic	NuPAGE/ LC-MS/MS/ SILAC	Tissue Microarrays	Tissue	CD9, perlecan, SDF4, apoE, fibronectin receptor
PC3, LNCaP, 22Rv1	Prostate	LC-MS/MS	ELISA	Serum	Follistatin, chemokine ligand 16, pentraxin 3, spondin 2
PC3(AR)6	Prostate	LC-MS/MS	ELISA	Serum	Mac-2BP, KLK5, KLK6, KLK11
CAL54	Renal	2DE/ MALDI-TOF MS, Western blot	Western blot, Homogeneous fluorescent immunoassay	Serum	Pro-MMP-7

Putative human cancer biomarkers identified by secretome analysis

TISSUE SECRETOME.

The patterns of protein secretion could be different between atherosclerotic plaques and normal endarteries. Thus, normal arteries and carotid endarterectomy samples (either complicated or fibrous) were incubated in protein-free medium and the supernatants subsequently analyzed by 2-DE. This approach allows the concentration of proteins released into the supernatant, corresponding to pathological or normal arterial wall. By focusing only on the secreted proteins found in the tissue culture media, there is an intended bias toward those molecules that would have a higher probability of later being found in the plasma.



Schematic diagram of the experimental strategy used to harvest the proteins released by different artery segments.

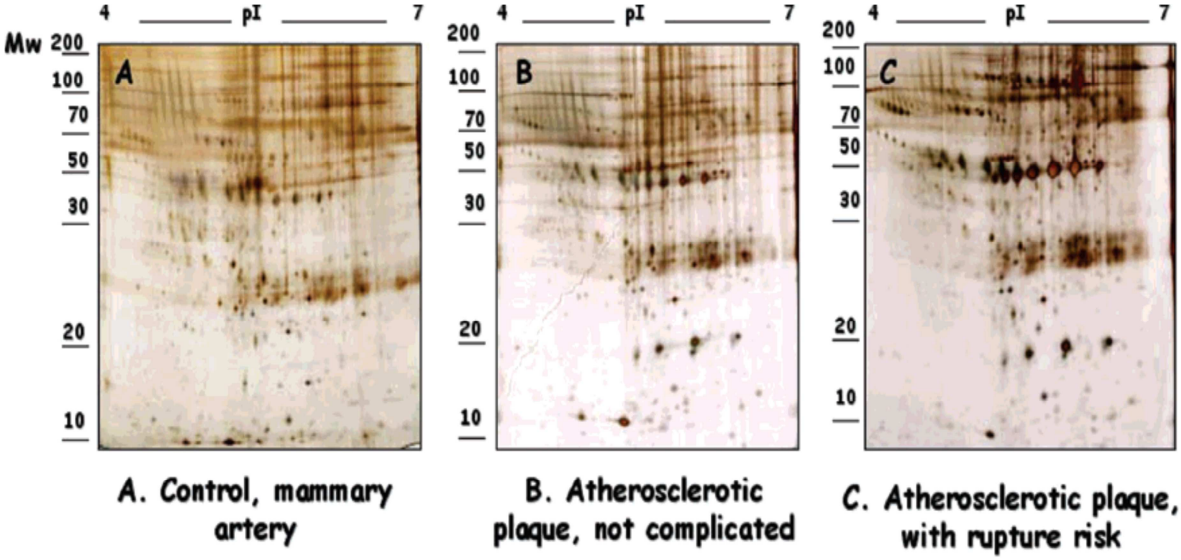
Two-dimensional electrophoresis followed by MS have enabled us to identify, in the secretomes, more than 80 proteins differentially expressed in complicated atherosclerotic

plaques versus adjacent fibrous plaques, of which 34 proteins exhibited increased release by complicated atherosclerotic plaques, while another 31 proteins were decreased (Duran, M. C.et.al 2005).

Proteins released by the atherosclerotic plaque, as well as their classification into five different groups according to their function, is presented: protein transporters, structural proteins, metabolic enzymes, transcription factors, antioxidant, and stress defense proteins. Interestingly, variations in protein levels were correlated with function and an increased secretion of protein transporters, enzymes and signal transductors during atherosclerosis was observed.

In contrast, the majority of antioxidant, stress defense and structural proteins, showed lower release by atherosclerotic areas than by the fibrous region (control). One of the most interesting proteins which showed an increased secretion in atherosclerotic plaques versus the control region is protein disulfide isomerase (PDI), which is considered as one of the major proinflammatory mediators involved in atherosclerosis (Higuchi, T..et.al 2004). Cathepsin D is another of the proteins which was secreted in higher levels by atherosclerotic plaques than in control regions. However, several isoforms were detected which were expressed at different levels, with increased or decreased secretion, depending on the considered isoform.

Among the differentially secreted proteins, decreased heat shock protein-27 (HSP27) was identified as a potential marker of atherosclerosis (Martin-Ventura, J. L.et.al 2004).



2-DE analysis of the proteins released by pathological (panels B and C) or normal (panel A) artery segments obtained by endarterectomy.

To confirm our hypothesis that plasma protein content can reflect arterial wall secretion, we measured soluble HSP27 level in the plasma of patients with carotid stenosis and healthy controls. Circulating HSP27 levels were decreased 20-fold in patients with carotid atherosclerosis relative to healthy subjects (Martin-Ventura, J. L.et.al 2005).

Similar findings have been recently made in diabetic patients. Two isoforms of HSP-27 were identified by MALDI-MS peptide mapping, corresponding to the nonphosphorylated (isoform 2) and the mono-phosphorylated state of HSP-27, with Ser82 as the modified residue (isoform 1). Recently, both isoforms were further studied showing a significant decrease in atheroma plaque secretion versus the nondamaged arterial segment. The more acidic isoform (pI 5.65) was released at lower levels than the nonphosphorylated analogue in the control region. Furthermore, the intensity of the decrease in phospho- HSP-27 was much greater than that of the nonphosphorylated isoform in atherosclerotic plaque conditioned medium. Although the exact role of HSP-27 in atherosclerosis is unknown, previous data suggest that HSP-27 could play a protective role. This function would be performed not only by regulating the mobility of SMCs and coordinating actin dynamics within the cells, but mainly by interacting with IKK protein and inhibiting the activation of NF-κB, widely involved in plaque instability and rupture. Interestingly, all these potential atheroprotective functions are exerted by the phosphorylated isoform of HSP-27. In this sense, the decrease in HSP-27 levels, and especially the lower levels of phospho-HSP-27 release observed in atheromatous plaques compared to controls, supports the hypothesis of a protective role for HSP27 against atherosclerosis. Appearing analyze of the protein profiles of mammary (control) endarteries and carotid plaques by SELDI-TOF MS (Surface-Enhanced Laser Desorption- Ionization Time-of-Flight Mass Spectrometry) and again a marked reduction in the amount of HSP27 was observed in the carotid plaque samples.

Identification of new biomarkers was restricted to the ability of finding and isolating new candidate proteins. However, in recent years, with the application of genomic and proteomic techniques, biomarkers can be identified without previous knowledge of their involvement in the pathogenesis of the disease. The list of biomarkers for atherosclerosis and ACS is expanding rapidly. However, the field is in the early stages of evolution and large-scale clinical studies are required to validate the usefulness of newly identified biomarkers in diagnosis, risk stratification, treatment and follow-up of cardiovascular diseases.

References

Annan RS, Carr SA: Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. Anal Chem. 1996 ,68(19),3413-21.

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Görg A, Westermeier R, Postel, W: Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. J Biochem Biophys Methods 1982, 6, 317-339

Blaum, Klaus (2006). "High-accuracy mass spectrometry with stored ions". Physics Reports 425 (1): 1–78.

Chevallet M, Diemer H, Van Dorssealer A, Villiers C, Rabilloud T. Toward a better analysis of secreted proteins: the example of the myeloid cells secretome. Proteomics. 2007;7(11):1757-70

Corthals GL, Molloy MP, Herbert BR, Williams KL, Gooley AA. Prefractionation of protein samples prior to two-dimensional electrophoresis. Electrophoresis. 1997;18(3-4):317-23

Duran, M. C.; Mas, S.; Martı´n-Ventura, J. L.; Meilhac, O.; Michel,J. B.; Gallego-Delgado, J.; Lazaro, A.; Tun˜on, J.; Egido, J.; Vivanco, F. Proteomic analysis of human vessels: application to atherosclerotic plaques. Proteomics 2003, 3, 973-978.

Görg A, Postel W, Günther S: The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 1988, 9, 531-546.

Higuchi, T.; Watanabe, Y.; Waga, I. Protein disulfice isomerase suppresses the transcriptional activity of NF-κB. Biochem. Biophys. Res. Commun. 2004, 318, 46-52.

Kebarle P, Verkerk UH (2009). "Electrospray: from ions in solution to ions in the gas phase, what we know now". Mass Spectrom Rev 28 (6): 898–917.

Kulasingam, V.; Diamandis, E. P. Mol.Cell. Proteomics 2007, 6, 1997.

Martin-Ventura, J. L.; Duran M. C.; Blanco-Colio, L.; Meilhac, O.; Leclercq, A.; Michel, J. B.; Jensen, O. N.; Hernandez-Mèrida, S.; Tunon, J.; Vivanco, F.; Egido, J. Identification by a differential proteomic approach of heat shock protein 27 as a potential marker of atherosclerosis. Circulation 2004, 110, 2216-2219.

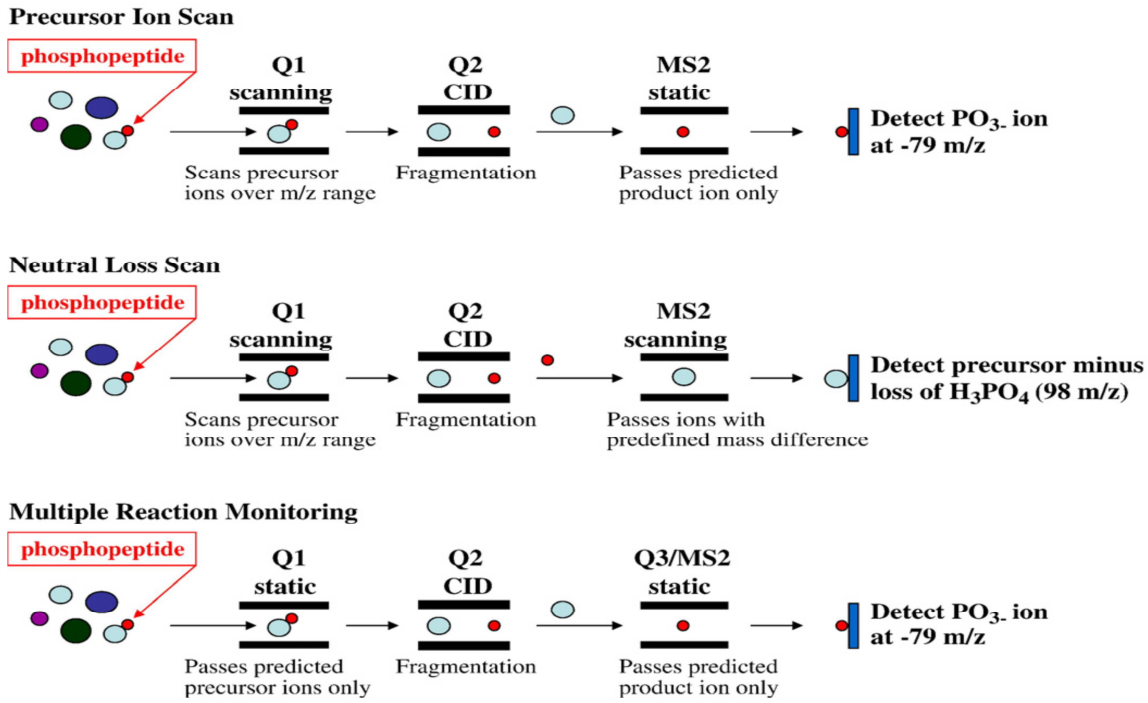
O' Farrell PH: High resolution two-dimensional electrophoresis of proteins. J Biol Chem 1975, 250, 4007-4021.

Chapter 3

PROTEOMICS APPLIED TO STUDY PROTEIN POST TRANSLATIONAL MODIFICATION

As already mentioned proteomics is quite useful in the study of protein posttranslational modification (PTM). This topics is particularly relevant because about 15 % of the human genes whose function is known (roughly 60% of the human genome) are engaged in signal transduction and communication, which takes place prevalently by PTM, particularly in the acute effects are taken into account. Reversible protein phosphorylation is one of the most abundant post translational modifications (PTM), with as much as one third of all proteins in a mammalian cell thought to be phosphorylated (P. Cohen 1999). The transfer of a phosphate group from adenosine triphosphate (ATP) to serine, threonine or tyrosine residues occurs through enzymes called protein kinases of which there are 518 encoded in the human genome. This process is reversible and dephosphorylation occurs through a slightly smaller family of enzymes called protein phosphatases. This dynamic phosphorylation/dephosphorylation often results in the sub-stoichiometric phosphorylation of proteins in vivo, as only a small subset of the cellular pool of a particular protein may be involved in any signal transduction cascade. The identification of sites of protein phosphorylation in vitro by protein kinases or in vivo, in response to various cellular stimuli, has been of central interest, yet presented a formidable challenge to MS based proteomics. One of the first routinely successful approaches combined HPLC separation of radiolabelled tryptic protein digests with Edman degradation and MS analysis. Two main kinds of phosphorylation reactions are known, those involving serine/threonine residues and hose related to phosphorylation of tyrosine residues by tyrosine kinases that are frequently associated with membrane receptors that undergo autophosphorylation triggered by ligand binding to generate phosphotyrosine docking sites that recruit SH2 and PTB domain containing proteins. The result is the assembly of large signaling complexes that feed receptor activity into different pathways (W.X. Schulze, et.al 2005). Specific techniques have been developed to investigate occurrence of phosphopeptides, by using specific scans on certain types of mass spectrometers. These scans are based on either the neutral loss of phosphoric acid from phosphoserine or phosphothreonine (neutral loss scan) during collision induced dissociation (CID) in positive mode, or the generation of a

phosphate specific reporter ion after collision induced dissociation in negative mode (precursor ion scan).



These processes of loss of phosphate moieties do not usually occur when phosphorylation of peptides at tyrosine residues is considered.

In addition to PTM by phosphorylation, proteomics techniques display potential for investigation of other reaction of PTM. In my guest laboratory interest in PTM is focusing mainly into signaling by protein oxidation and by transamidation. The former is usually known as Redox Proteomics. I will describe briefly this topics (although I was not personally involved in these experiments) because of its biologic relevance.

Protein oxidation: role in signaling and detection by mass spectrometry.

Proteins can undergo a wide variety of oxidative post-translational modifications (oxPTM); while reversible modifications are thought to be relevant in physiological processes, non-reversible oxPTM may contribute to pathological situations and disease. The oxidant is also important in determining the type of oxPTM, such as oxidation, chlorination or nitration. The best characterized oxPTMs involved in signalling modulation are partial oxidations of cysteine to disulfide, glutathionylated or sulfenic acid forms that can be reversed by thiol reductants. Some proteins regulated by cysteine oxidation, and the

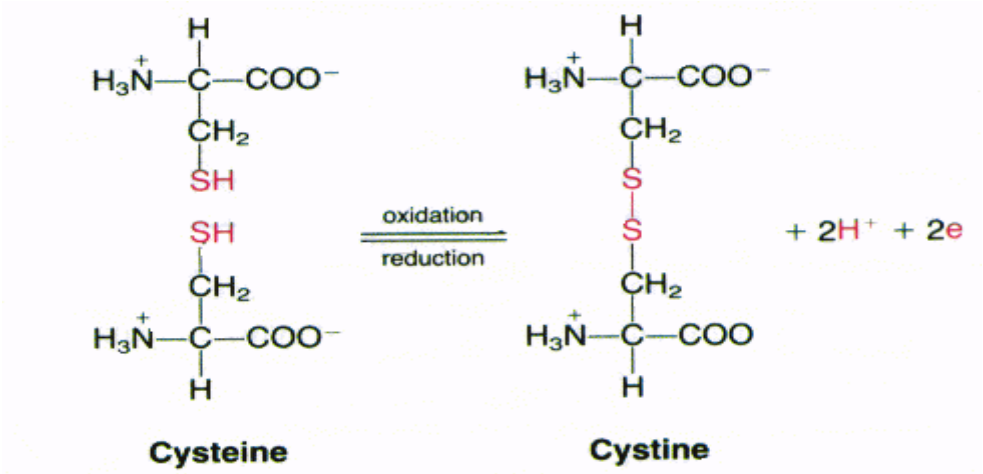
residues and molecular mechanism involved, have been extensively studied and are well understood, e.g. protein tyrosine phosphatase PTP1B, MAP3 kinase ASK1 and the transcription factor complex Keap1–Nrf2.

Understanding of the role oxPTMs in signalling has been facilitated by advances in analytical technology, in particular tandem mass spectrometry. Combinations of peptide sequencing by collision induced dissociation and precursor ion scanning or neutral loss to select for specific oxPTMs have proved very useful fodentifying oxidatively modified proteins and mapping the sites of oxidation.

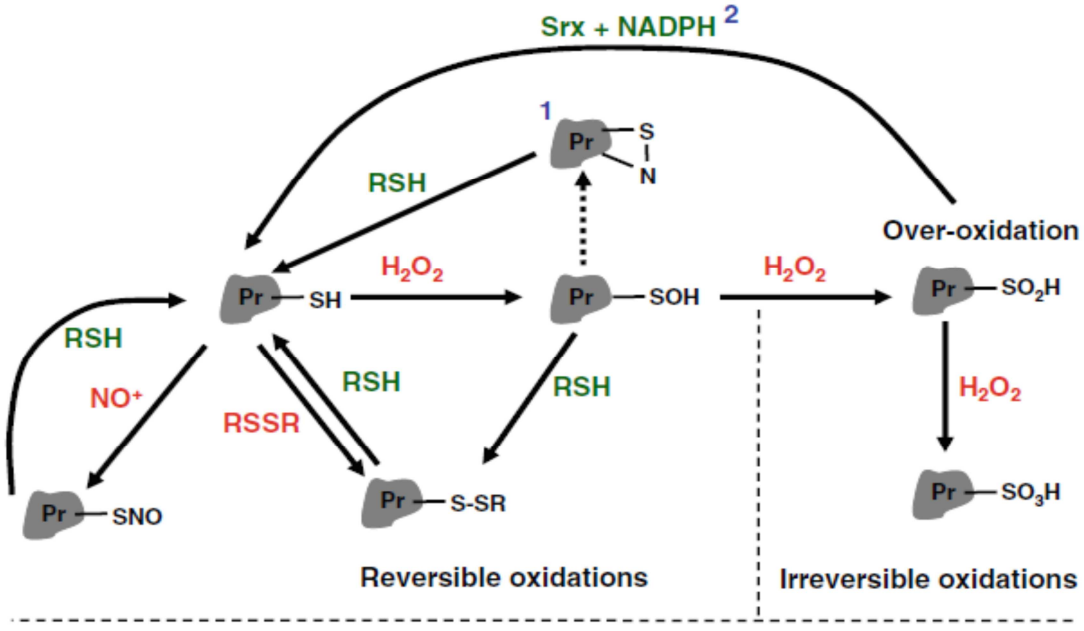
The oxidative PTM are not simply damaging to proteins and deleterious to cells, but under appropriate conditions can constitute important signalling mechanisms involved both in response to stress situations, such as infection and inflammation, and in normal physiological control of cell behavior. A variety of partially reduced oxygen species [superoxide ($O2^{-}$), hydrogen peroxide (H2O2) and hydroxyl radical (OH.)], reactive nitrogen species [nitric oxide (NO), nitrogen dioxide (N2O) and peroxynitrite (ONOO-)] and reactive chlorine species (e.g. hypochlorous acid, HOCl) can be produced by enzymatic and chemical reactions during phagocyte activation in inflammation, most of which are capable of reacting with the side chains in proteins. Moreover, superoxide, hydrogen peroxide and nitric oxide are produced during normal metabolism, by electron leakage during mitochondrial respiration, redox cycling enzymes such as cytochrome P450, xanthine oxidase and metabolite autoxidation can all result in superoxide formation . Also nitric oxide is produced by nitric oxide synthases (NOS), which are either constitutive or inducible enzymes. No thereafter reacts with protein amino acids, leading prevalently to their S-nitrosylation.

Protein oxidations at sulfur containing residues are very common oxPTMs, at cysteine and methionine residues. Oxidation of these residue may influence catalytic activity of enzymes, either directly at the active site or through conformational changes, or affect protein–protein interaction, protein degradation, and other post-translational modifications.

Redox-dependent behaviour has been reported for proteins involved in most aspects of cellular processes: metabolism, inflammatory responses, cell proliferation and differentiation, survival or apoptosis, and most signaling pathways can be modulated by cellular redox status. These effects depend on the activation, or more commonly, inactivation of specific proteins, usually by reversible cysteine oxidation.



Cysteine oxidation and reduction are chiefly reversible events but the overall rearrangement in cellular metabolism are more complex events, including both reversible and irreversible steps, probably in relation to the final accumulation of oxidant and reductant levels, with glutathione playing significant roles in these processes.



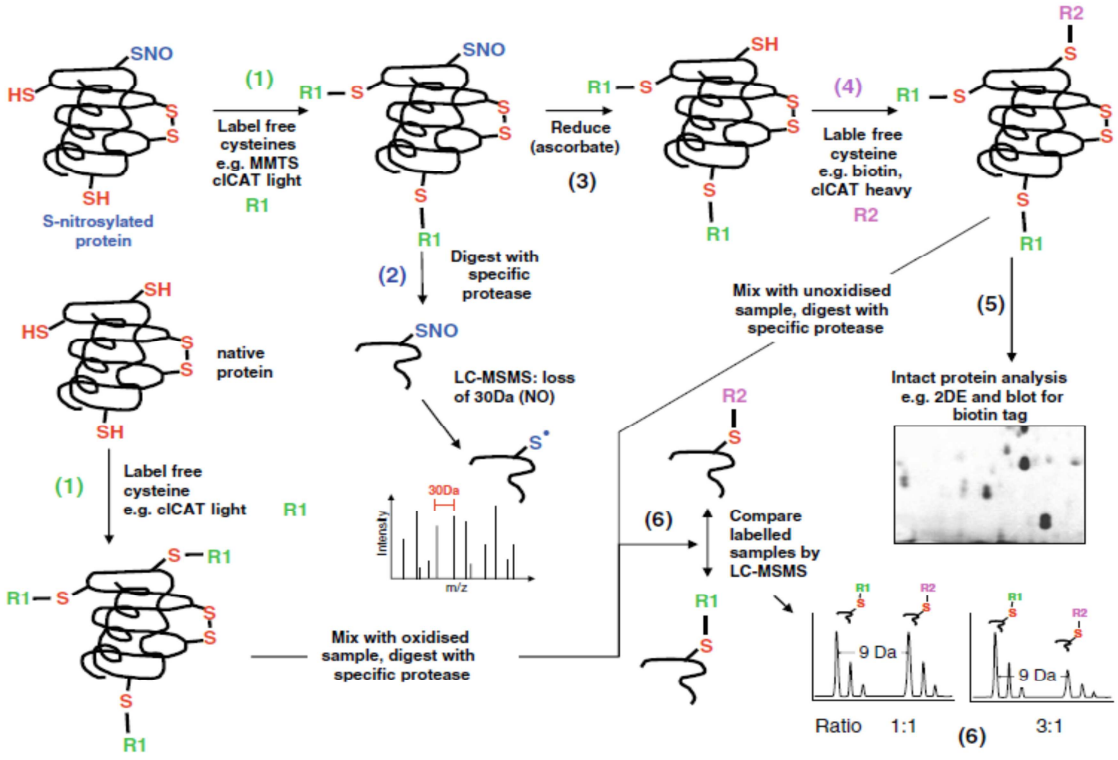
Scheme of protein thiol oxidation and regeneration reactions.

Oxidation of **Methionine residues** can also occur and this damage can be **c**an be repaired by methionine sulfoxide reductases; consequently methionine oxidation has previously been considered to exist as a protective mechanism for proteins, and one that in some proteins has comparatively little adverse on their function (Stadtman et al. 2003). Evidence for an additional role of certain methionines as oxidation sensors in the redox regulation of enzyme activity is accumulating, with particular emphasis on calcium transport and signaling. A notable example in human pathology is related to the susceptibility of patients

to cigarette smoke with development of lung emphysema in patients bearing the ZZ genotype of the serpin α1-Antitrypsin. Calmodulin is also very sensitive to oxidation at methionine residues.

Procedures to detect oxPTM of proteins may proceed through 2D Gel based analysis for thiol oxidation. In these approaches proteins are usually labeled with residue or modification-specific reactive fluorescent chromophores (e.g. maleimide- or iodoacetamide-based reagent for cysteine thiols) with different colors, and then the two samples are mixed and run on the same gel, thus removing problems caused by variation between gels. Differential colour scanning is then used to detect the proteins in a technique known as DIGE (differential in gel electrophoresis).

Spots of interest can be excised from the gel and the proteins identified by mass spectrometry, and (with luck) the site and nature of the modification might even be determined if the modified peptide is seen in the mass spectrometric analysis. Specific techniques (whose description would be outside the scopes of this Thesis) have been developed for Mass spectrometric detection of disulfides and higher oxidation state sulfur, S-Nitrosothiols (SNOs), based either on "D electrophoresis or on LC MS approaches, as in the following scheme.



Proteomics and protein crosslinking. I will detail some more on the procedures to study PTM by protein crosslinking that can be investigated by combining with mass spectrometry for structural proteomics. Protein crosslinking can be achieved either by employing bifunctional reagents or alternatively enzymes that catalyze formation of crosslinks between distant regions in a single protein or better covalent polymers between different proteins. In this perspective I took advantage of Transglutaminase, that is a PTM-ing enzyme that was the focus of years long investigations in the lab. In these cases the proteomic approaches can be employed not only to answer the traditional question of which proteins are getting crosslinked, but also to determine at which sites the protein crosslinking takes place.

The location of the created cross-links imposes a distance constraint on the location of the respective side chains and allows drawing conclusions on the distance geometries of a protein or a protein complex structure (Sinz, 2003). Analysis of cross-linked peptides by mass spectrometry makes use of several advantages associated with MS analysis: (1) The mass of the protein or the protein complex under investigation is theoretically unlimited because it is the proteolytic peptides that are analyzed (in case a bottom-up strategy is employed). (2) Analysis is generally fast and, in favorable circumstances, requires only femtomole amounts of total protein. (3) It is possible to gain insights into three dimensional structures of proteins in solution and flexible regions are readily identified. (4) Membrane proteins and proteins that exist as mixtures of different species (post-translational modifications, splice variants) are amenable to analysis. (5) The broad range of specificities available for crosslinking reagents towards certain functional groups, such as primary amines, sulfhydryls, or carboxylic acids, and the wide range of distances that different cross-linking reagents can bridge, offer the possibility to perform a wide variety of experiments . In these studies a first protein fragmentation step is usually performed by means of a protease of high specificity of cleavage (usually trypsin or V8 proteinase) and this might represent the main limitation of the approach since usually crosslinked proteins are characterized by high structural rigidity that hampers degradation by proteases.

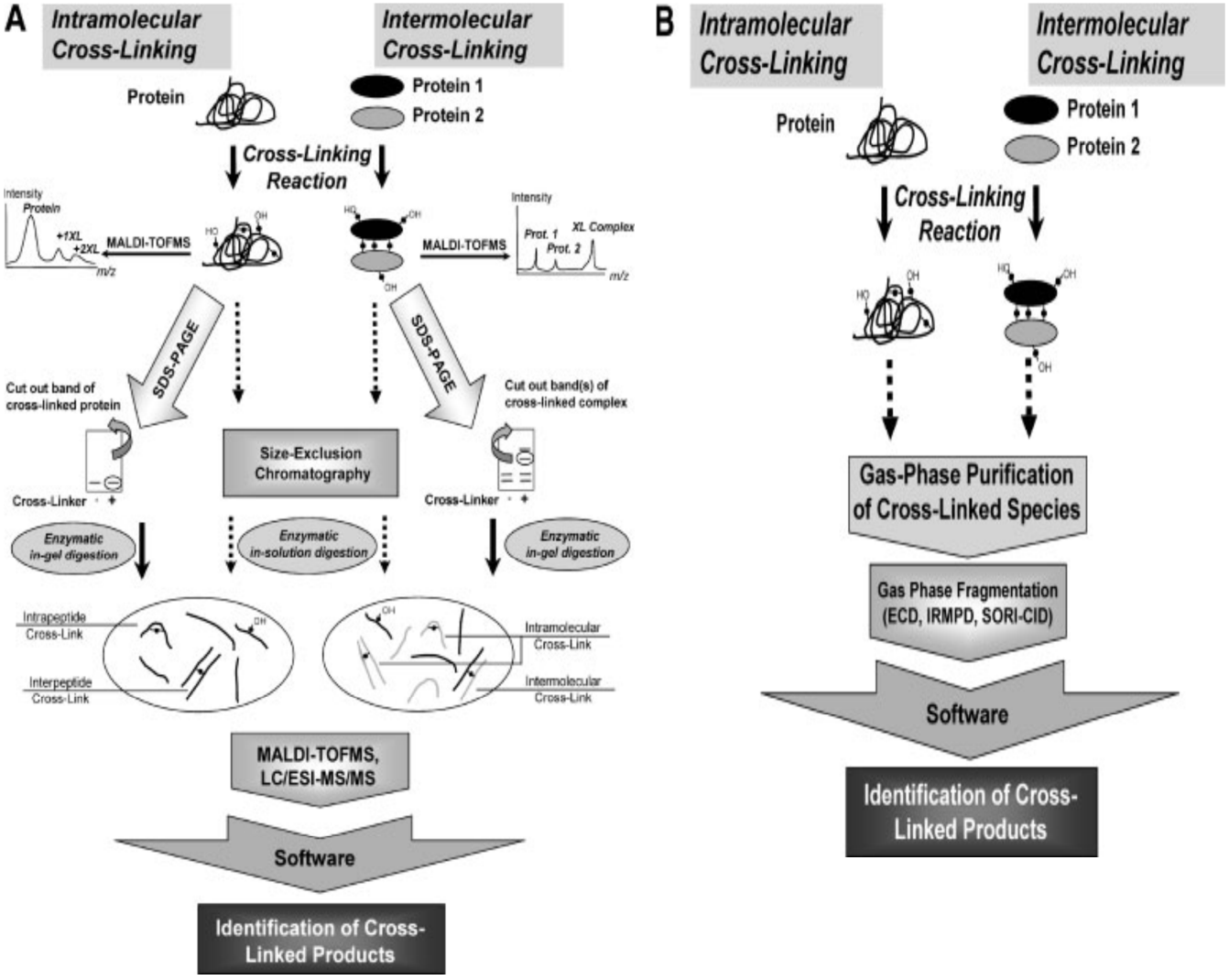
Several strategies have been employed to enrich crosslinker-containing species by affinity chromatography or to facilitate the identification of the cross-linked products; for example by using isotope-labeled cross-linkers or proteins, fluorogenic cross-linkers, or cleavable cross-linkers.

In recently developed innovative strategies, cross-linking reactions are conducted in living cells by directly incorporating reactive groups into the protein, using the cell own biosynthetic machinery.

Bottom-Up Approach.

In the bottom-up approach, the protein reaction mixture is enzymatically digested after the cross-linking reaction, and mass spectrometric identification of the cross-linked products is performed, based on the resulting proteolytic peptides. The bottom-up approach has been applied to map protein interfaces, but it has also proven especially valuable to determine low-resolution three-dimensional structures of proteins (Young et al., 2000). The most important prerequisite to successfully conduct crosslinking experiments is a detailed description of the respective amino acid sequences of the proteins under investigation. Full sequence coverage should be envisioned to fully characterize the protein with respect to possible amino acid variants, post-translational modifications, or splice variants. When conducting cross-linking reactions, control samples must be included, to which no cross-linker is added, to exclude the formation of any non-specific aggregates. After the cross-linking reaction, one-dimensional gel electrophoresis (SDS–PAGE) and MALDI-TOF (matrix assisted laser desorption/ionization time-of-flight) MS analysis of the reaction mixture can be used to check for the extent of cross-linked product formation and to optimize the reaction conditions. After the cross-linking reaction, there are several ways to isolate the cross-linked proteins from the reaction mixture. If SDS–PAGE of the cross-linking reaction mixture is performed, the band of the cross-linked protein or the crosslinked protein complex is excised from the 1D-gel and subjected to enzymatic in-gel digestion. Alternatively, the crosslinked protein or protein complex is separated from the reaction mixture by size-exclusion chromatography, and the digestion is performed in the solution. In some occasions in-solution digestion of the cross-linked proteins was more efficient than in-gel digestion, where appreciable amounts of protein can be lost, likely is high molecular weight aggregates. Peptide mixtures that originated from proteolytic digestion of cross-linking reaction mixtures were analyzed by MALDI or ESI mass spectrometry. One of the inherent problems of the bottom-up strategy is that large peptides are commonly created from cross-linked proteins during enzymatic proteolysis due to a high frequency of missed cleavages. Missed cleavages occur because the most commonly employed cross-linking reagents react with primary amine groups at lysine residues and the

N-termini of proteins, and trypsin the most commonly used proteolytic enzyme will not cleave C-terminal to a modified lysine residue. Another limitation of the bottom-up approach is that cross-linked products with low charge states are frequently created during electrospray ionization due to a loss of positive charge after modification of the e-amino groups of lysine residues; that modification might cause large peptides not to be detected.



General analytical strategies for protein structure characterization by chemical cross-linking and mass spectrometry. (A) Bottom-up approach, and (B) top-down approach using FTICR-MS.

Moreover, the number of peptides with the same nominal mass but different amino acid sequence, increases with the rising number of amino acid residues in the peptide.

Top-Down Approach.

One of the most direct techniques to analyze cross-linked products is the top-down approach, in which the cross-linked proteins are analyzed intact rather than being digested before the mass spectrometric analysis. Electrospray ionization-Fourier-transform ion-cyclotron resonance (ESI-FTICR) mass spectrometry is the method of choice for this kind of analysis. So far, the top-down approach has been exclusively employed to determine low resolution three-dimensional structures of proteins from intramolecular cross-linking experiments. The cross-linking reaction mixture is presented to the FTICR mass spectrometer, and the cross-linked product is isolated in the ICR cell before it is interrogated with one of the various fragmentation techniques, such as sustained off-resonance irradiation collision-induced dissociation (SORI-CID), infrared multi-photon dissociation (IRMPD), or electron capture dissociation (ECD). Determination of the accurate mass of the intact cross-linked product provides hints on the number of incorporated cross-linker molecules as well as on the number of modifications caused by partially hydrolyzed cross-linkers. The top-down approach presents some advantages over the bottom-up approach in that it eliminates the need to separate the reacted protein from the cross-linking reaction mixture before the mass spectrometric analysis, because this separation is accomplished by a ‘gas-phase purification’ in the mass spectrometer. Top-down approaches have been successfully employed to assign intramolecular crosslinked products of bovine rhodopsin (Novak et al., 2005) as well as ubiquitin. ECD seems to be especially favorable in conjunction with FTICR-MS because it allows a comprehensive fragmentation of large peptides, while post-translational modifications are kept intact. One limitation of the top-down approach is that analyses of large protein assemblies are difficult to perform (Novak et al., 2005). That combination of bottom-up and mass spectrometry and top-down analysis will most likely become the strategy with the greatest potential for a rapid and efficient analysis of a wide variety of cross-linking reaction mixtures.

Chemical Cross-Linking in Living Systems.

Several highly appealing approaches have been developed, which permit cross-linking of interacting proteins in their natural environment, and thus, give insight into the way cellular processes are organized. In one strategy reported, live cells are treated with formaldehyde, which rapidly permeates the cell membrane and generates cross-links between interacting proteins in the cell. Proteins that are cross-linked to a myc-tagged protein of interest are co-purified by immunoaffinity chromatography, the cross-linked complexes are subsequently dissociated, the bound proteins are separated by one-dimensional gel electrophoresis and identified by tandem MS. Another intriguing strategy is based on the incorporation of a unique chemical into a protein of interest, using the cell's own biosynthetic machinery, followed by a chemical reaction with a small-molecule probe. Only a handful of chemical motifs are known to possess the requisite qualities of biocompatibility and selective reactivity to function as chemical reporters in living cells. Such chemical motifs include the tetra-Cys motif that reacts selectively with biarsenicals or azides that react with phosphines in a Staudinger ligation or in a copper-catalyzed 'click' chemistry type reaction with alkynes (Prescher & Bertozzi, 2005). In another interesting study, three new photoactivatable amino acids that contain a diazirine moiety were designed and termed ''photo-methionine'' and ''photo-leucine,'' and ''photo-isoleucine''. The preference for methionine, leucine, and isoleucine implies that transmembrane domains as well as hydrophobic contact areas between proteins are preferentially cross-linked. In this preference, the method is complementary to the most commonly employed amine-reactive chemical cross-linkers that target lysine residues and the N-termini of proteins.

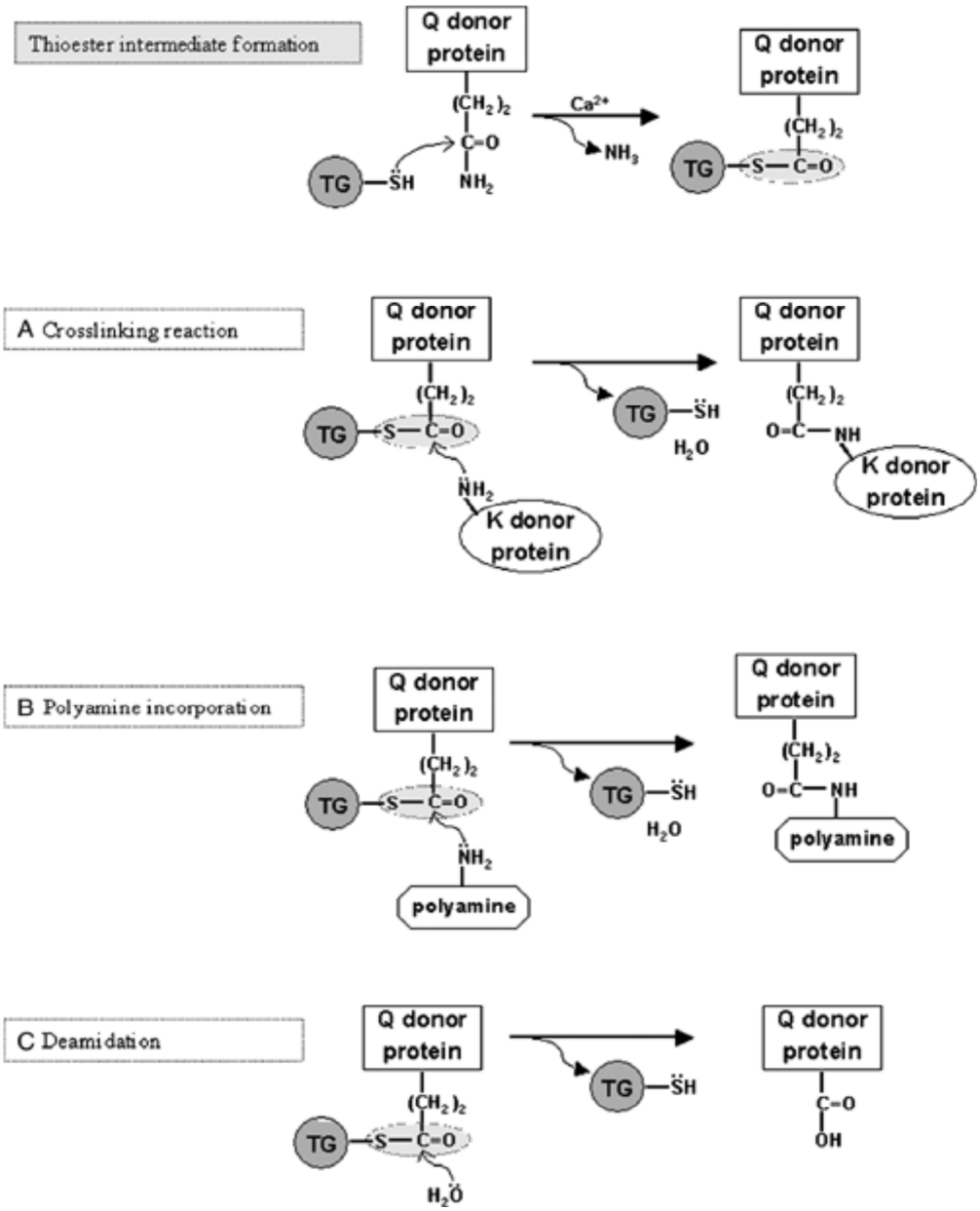
A major point of interest in this Thesis is the analysis of protein crosslinks formed by **Transglutaminases** (Tgases) which are widely distributed enzymes that catalyze PTM of proteins through isopeptide bonds which include protein cross-linking via ε-(γ-glutamyl)-lysine bonds or incorporation of primary amines at selected peptide bound glutamine residues. The cross-linked products, often of high molecular mass, are highly resistant to mechanical stress and to proteolytic degradation, leading to their accumulation is found in tissues during processes in which such properties are important, including skin, hair, blood clotting and wound healing. However, deregulation of enzyme activity, which is generally associated with major disruptions in cellular homoeostatic mechanisms,

explains the contribution of these enzymes to human diseases, including chronic neurodegeneration, neoplastic diseases, autoimmune diseases, diseases involving progressive tissue fibrosis and diseases related to the epidermis of the skin.

The term transglutaminase was first coined by (Clarke et al.,1957) to describe the transamidating activity observed in guinea-pig liver. Later studies on the stabilization of fibrin monomers during blood clotting (Pisano et al.1968), demonstrated that transamidation is brought about by enzymes which cross-link proteins through an acyl-transfer reaction between the γ-carboxamide group of peptide-bound glutamine and the ε-amino group of peptide-bound lysine, resulting in the formation of a ε-(γ-glutamyl)-lysine isopeptide bond. Proteins with Tgase activity have now been found in micro-organisms, plants, invertebrates, amphibians, fish and birds (Griffin et al, 1992).

In common with many other important cellular functions found in mammalian cells, Tgases require high concentration of Ca^{2+} for their activity. Moreover their Ca^{2+} activation is also modulated by further regulatory processes, which in essence means that they are virtually inactive under normal conditions and only activated following major disruptions in physiological homoeostatic mechanisms. Once activated, Tgases catalyze the formation of the isopeptide bonds as already described.

Two points that deserve consideration in the case of the transamidation reaction (see the above schema) are the irreversibility of the reaction (some reversibility has been demonstrated only in the of PTM reaction leading to protein ammine adducts) and the strict selectivity in the recognition of protein glutamine substrates, and poor specificity for the acyl-acceptor amine group, which can either be the e-amino group of peptidyl lysine or a low-molecular mass primary amine (frequently a polyamine). Alternatively even water can act as nucleophile.



The transamidation reactions catalyzed by Tgases.

In the first instance, the products are often cross-linked high molecular-mass protein aggregates, while in the latter, protein-polyamine conjugates are generated, which can also be further polymerized. Biochemical and cell-biological studies indicate that both reactions involving protein cross-linking and polyamidation are relevant *in vivo*, and competition between these amine substrates may take place within cells in a number of important physiological functions where they act as a ` biological glue', including that of cell death, cell-matrix interactions, in the stabilization of the epidermis and of hair and in the general maintenance of tissue integrity.

Tgases represent a family of enzymes with eight distinct Tgase isoforms as summarized in table below, which display appreciable degree of homology. My discussion will be restricted to type 2 Tgase , which is wide distributed in mammals and is the only form I employed in my studies.

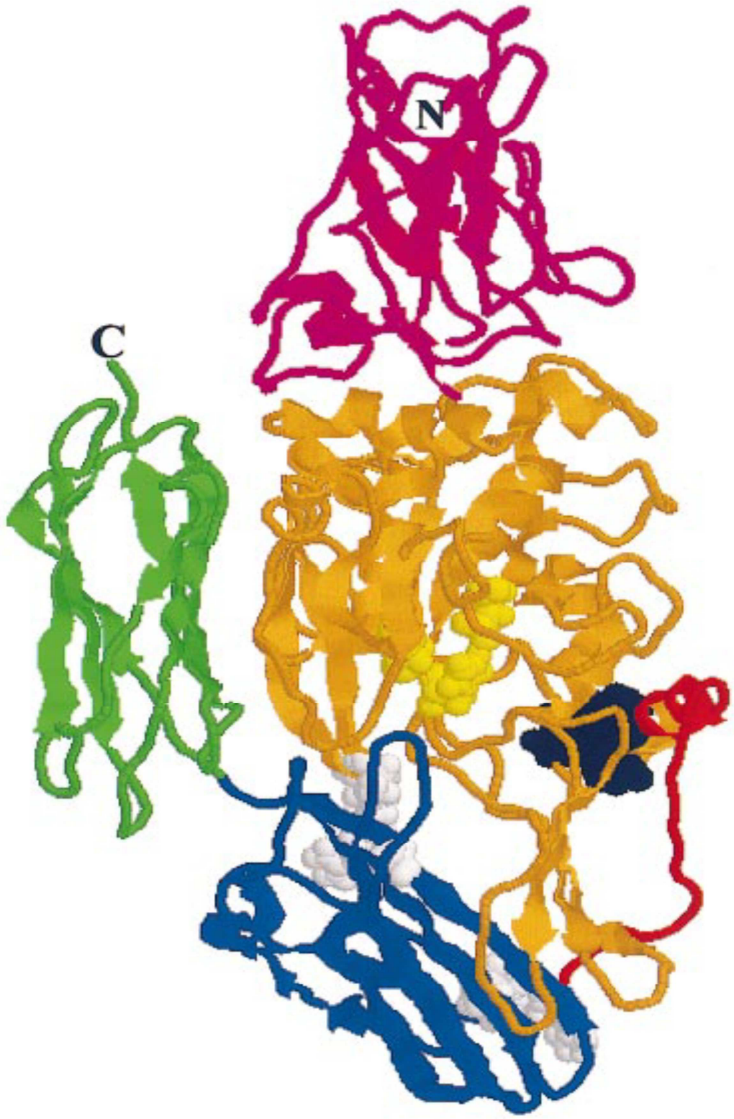
In addition to the eight different enzymes listed below, a further Tgase-like protein has been characterized from red blood cells. This protein, named erythrocyte-bound 4.2, has strong sequence identity with the Tgase family of proteins, but is inactive because of a substitution of alanine for the active-site cysteine: it forms a major component of the erythrocyte membrane skeleton [251].

Identified forms of Tgase	Synonyms	Residues (molecular mass in kDa)	Gene	Gene map locus	Prevalent function
Factor XIII A	Catalytic A subunit of Factor XIII found associated with B subunit in plasma as A2B2 heterotetramer. Fibrin stabilizing factor	732 (83)	F13A1	6p24-25	Blood clotting and wound healing
Type 1 Tgase	Keratinocyte Tgase	814 (90)	TGM1	14q11.2	Cell envelope formation in the differentiation of keratinocytes
Type 2 Tgase	Tissue Tgase	686 (80)	TGM2	20q11-12	Cell death and cell differentiation, matrix stabilization, adhesion protein
Type 3 Tgase	Epidermal Tgase	692 (77)	TGM3	20q11-12	Cell envelope formation during terminal differentiation of keratinocytes
Type 4 Tgase	Prostate Tgase	683 (77)	TGM4	3q21-22	Reproductive function involving semen coagulation particularly in rodents
Type 5 Tgase	Tgase X	719 (81)	TGM5	15q15.2	Epidermal differentiation
Type 6 Tgase	Tgase Y		TGM6	20q11 15	Not characterized
Type 7 Tgase	Tgase Z	710 (80)	TGM7	15q15.2	Not characterized

A point that is important to underline is the Tgase2 (which is actually a multifunctional enzyme, Bergamini et al, 2011) is silent within the cell for its transamidation activity because it requires high concentration of calcium (above 1 mM to display appreciable activity) and is also inhibited by GTP. Additional levels of control of activity are represented by the cell oxidative power and by eventual variation in its expression.

The structure of the enzyme is now known, as well as the general outline of its reaction mechanism.

Which involves formation of thioester intermediate between the peptidylglutamine substrate and the thiol group of Cys 277 that is engaged in a catalytic triad with His 335 and Asp 358. The catalytic center is present in the core domain of the protein.



Backbone structure of tTgase (the picture was produced with Rasmol). Domains I-IV are coloured respectively in magenta, orange, blue and green ; the regulatory loop between domain II and III is coloured red. Amino acids involved in the active site (Cys^{277}, His^{335} and Asp^{358}), in Ca^{2+} binding (Ser^{449}, Pro^{446}, Glu^{451} and Glu^{452}) and in interaction with GTP (Ser^{171}, Lys^{173}, Arg^{478}, Val^{479} and Arg^{580}) are coloured yellow, black and light grey respectively.

The mechanism of activation of tTgase by Ca^{2+}, mediated through the dislocation of the Tgase inhibitory domains, can be counteracted by the ` allosteric ' inhibitor GTP (see Bergamini et al, 2011 for further details).

Since Tgase 2 is by one side a latent enzyme, by the other can catalyze protein PTM, it has long been a tempting perspective to ascribe the physiologic roles played by the enzyme to its transient activation with changes in structure and function of target proteins that should mediate the roles played by the protein. A point in contrast with this view is represented by the irreversibility of the process since in most instances studied so far, events of “signal transduction” are usually related to reversible regulatory cascades. In any

case, this perspective stimulated interest in identifying substrate proteins (Esposito and Caputo, 2005) and an additional trigger for this search was represented by the possibility to utilize Tgase mediated formation of protein-amine adducts for biotechnological purposes (Porta et al, 2011).

Although I will discuss further with this issue in the experimental section (in chapter 5 I will analyze the modification by Tgase 2 of some purified proteins including skeletal muscle troponins, brain tubulin and diphtheria toxin) it is nevertheless important to note that the general opinion is that Tgase 2 is kept inactive at near physiological states both in the intracellular and in the extracellular compartment because of its complex regulatory network (calcium, GTP, reductants) as detailed elsewhere (Bergamini et al, 2011). Under these conditions it plays a role in cell survival through its complementary activity of signal transduction (Johnson swiss). In contrast activation of the enzymes takes place under conditions of severely deranged cell homeostasis switching-on the programs leading to cell death, usually through apoptosis. It is exactly under these conditions that proteins in the extra- or in the intracellular compartments are actively modified by Tgase 2. As detailed in some recent review the proteins that undergo PTM by Tgase 2 in these occasions include mainly enzymes involved in energy metabolism, cytoskeletal proteins and stress proteins (Esposito and Caputo, 2005).

Additional points of interest in Tgase research are represented by the involvement of the enzyme in several human important pathologic processes (Iismaa et al, 2009) which include: (i) autoimmune reactions as celiac disease, CD, a pathology characterized by hypersensitivity to dietary gliadin and development of autoantibodies directed against Tgase2 which also play a role in the malabsorption syndrome related to CD; (ii) involvement in several neurodegenerative which depend on the accumulation of crosslinked proteins in the brain; (iii) modulation of cell growth and programmed death and more generally in biology of cancer growth; and finally (iv) modulation of vascular responses, related to vascular remodeling in hypertension and control of permeability.

Reference.

Bergamini CM, Gambetti S, Dondi A, Cervellati C. (2004) Oxygen, reactive oxygen species and tissue damage. Curr Pharm Des. 2004;10(14):1611-26.

Bergamini CM, Collighan RJ, Wang Z, Griffin M. (2011) Structure and regulation of type 2 transglutaminase in relation to its physiological functions and pathological roles. Adv Enzymol Relat Areas Mol Biol. 78:1-46

Casadio R, Polverini E, Mariani P, Spinozzi F, Carsughi F, Fontana A, Polverino de Laureto P, Matteucci G, Bergamini CM. (1999) The structural basis for the regulation of tissue transglutaminase by calcium ions. Eur J Biochem. 262 :672-9.

Clarke DD, Neidle A. Sarkar NK, Waelsch H (1957) Metabolic activity of protein amide groups. Arch Biochem Biophys. 71 : 277-9

Cohen P. (1999) The Croonian Lecture 1998. Identification of a protein kinase cascade of major importance in insulin signal transduction. Philos Trans R Soc Lond B Biol Sci. 354: 485-495.

Esposito C, Caputo I. (2005) Mammalian transglutaminases. Identification of substrates as a key to physiological function and physiopathological relevance. FEBS J. 272: 615-31

Griffin M, Casadio R, Bergamini CM. (2002) Transglutaminases: nature's biological glues. Biochem J. 368: 377-96

Iismaa SE, Mearns BM, Lorand L, Graham RM. (2009) Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. Physiol Rev. 89: 991-1023.

Novak P, Haskins WE, Ayson MJ, Jacobsen RB, Schoeniger JS, Leavell MD, Young MM, Kruppa GH. 2005. Unambiguous assignment of intramolecular chemical cross-links in modified mammalian membrane proteins by Fourier transform-tandem mass spectrometry. Anal Chem 77:5101–5108.

Prescher JA, Bertozzi CR. 2005. Chemistry in living systems. Nat Chem Biol 1:13–21.

Pisano JJ, Finlayson JS, Peyton MP. (1968) Cross-link in fibrin polymerized by factor 13: epsilon-(gamma-glutamyl)lysine. Science. 160 : 892-3

Porta R, Mariniello L, Di Pierro P, Sorrentino A, Giosafatto CV. (2011) Transglutaminase crosslinked pectin- and chitosan-based edible films: a review. Crit Rev Food Sci Nutr. 51: 223-238

Prescher JA, Bertozzi CR. (2005) Chemistry in living systems. Nat Chem Biol. 1: 13-21

Schulze WX, Deng L, Mann M (2005). Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol.1:2005.0008.

Sinz A (2003). Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. J Mass Spectrom. (12):1225-37

Stadtman ER, Moskovitz J, Levine RL (2003). Oxidation of methionine residues of proteins: biological consequences. Antioxid Redox Signal 5:577-82.

Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger D. 2000. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. Proc Natl Acad Sci USA 97:5802–5806.

Chapter 4

PROTEOMICS IN BIOLOGICAL AND CLINICAL APPLICATIONS

From the biological point of view an immediate interest of the proteomic research is related to the field of modulation of protein expression through variation in their biosynthetic rate. This is exemplified by the issues of Systems biology, that is the analysis of the relationships among the elements in a system in the response to genetic or environmental perturbations, with the goal of understanding the system or the emergent properties of the system. A system may be a few protein molecules carrying out a particular task (such as metabolism of galactose or any other specific metabolite); a complex set of proteins and other molecules physically associated or working together as a molecular machinery (such as the ribosome); a network of proteins operating together to carry out an important cellular function such as giving the cell shape (protein network), or a cell or group of cells carrying out particular phenotypic functions. Thus, a biological system may encompass molecules, cells, organs, individuals, or even ecosystems. System biology will continually improve our capacity to understand and to model biological systems on a more global and in-depth scale than ever before. The systems biology in medicine will be continual with the development of new technologies, which will enhance efficiency, scale and precision with which cellular measurements are made. This later influence will facilitate all aspects of health care, including the detection and monitoring of diseases, drug discovery, treatment evaluation, and ultimately, predictive and preventative medicine.

As already pointed out a main usefulness of proteomic investigations is related to the analysis of qualitative/ quantitative changes in tissue protein content. These changes are usually ascribed to variations in the cell protein synthetic potential, but this is certainly an oversimplification, since several other effects might contribute to a final pattern documented by the experimental techniques, including chiefly protein breakdown, as in red blood cells. What is now clearly emerging is that - thanks to their high sensitivity and versatility - proteomic techniques can provide answers to complex sets of queries when rationally employed within an experimental design. In the perspective of biological and medical applications, results of proteomic experiments can be interpreted to solve

experimental and clinical issues. This subdivision displays certainly borderline effects and is therefore questionable so that we prefer to merge it somewhat in this short discussion of the usefulness of proteomic investigations, taking into account a few selected examples.

A prominent field of application is that of the

HUMAN PLASMA PROTEOME

which has become a main goal of research in the proteomics arena. The plasma proteome is undoubtedly the most complex proteome in the human body; consisting not only of the resident transport and haemostatic proteins, but also of immunoglobulins, cytokines, protein hormones, secreted proteins and foreign proteins, indicative of infection. In addition, blood circulates through almost all tissues of the body, and therefore contains tissue leakage proteins, including those released from damaged or dying cells. Plasma should, as noted above, therefore contain information on the physiological state of all tissues in the body. This, combined with its accessibility makes the blood proteome invaluable for medical purposes. It will become possible to relate individual serum proteome profiles to the genome, environmental stimuli, lifestyle and aging of each individual. What is important to note is that plasma proteome is characterized by a wide dynamic range, i.e. the great difference in concentration between the most abundant and the trace proteins, which are difficult to identify because their presence is masked by that of the most abundant plasma proteins. The dynamic range of plasma proteins is on the order of 10^9, with serum albumin being most abundant (30-50 mg/mL) while low-level proteins are present at much lower concentrations as it is the case for interleukin-6 which is present at a concentration of 0-5 pg/mL. (Anderson and Anderson, 2002). The most meaningful subset of proteins among the entire plasma proteome to generate diagnostics tools should probably focus on the less abundant fraction of proteins in the blood plasma, as it is this fraction that best reflects tissue physiology and pathology, as these usually belong to the nonresident proteins that are either actively or passively released from tissue into the blood stream. (Liotta et.al 2003). In addition, since diseases are potentially associated with altered structure of plasma proteins through changes in their sugar components (e.g. cancer notably associated with enhanced glycosylation of proteins, and alcoholism with decreased glycosylation of transferrin) methods that focus on specific subfractions of the plasma proteome may also be useful. As proteomics techniques improve, the disease profiles generated should be continually related to the respective gene expression changes, the genome sequence information.

Other emerging experimental applications deal for instance with the general processes of

TISSUE AND BODY AGEING

The process of senescence for instance is characterized by wide individual variability, possibly under genetic control. Population genetics has widely documented major differences in familial longevity in humans, encompassing brief or prolonged life expectancies, as originally postulated by Hayflick in his limit theory (reviewed in Hayflick, 1985), in relation to the number of mitotic divisions in different species. If this is really the case it is likely that the proteomes of individuals genetically programmed for longevity should differ from those of subjects with constitutional short life expectancies.

Other aspects might be related to the issues of aging in specific tissues. E.g. in muscle a process of sarcoplasmic atrophy associates with aging leading to an increased tendency of old people to fall down with risk of fractures (Ohlendieck, 2011) In these peculiar instances which are not under definitive genetic control despite the known genomic instability in the elderly (Nyström et al 2012), it is likely that changes in the proteomes, that certainly take place, depend on the progressive in vivo accumulation of damage without appropriate repair, e.g. by unedited protein synthesis or by oxidative challenges (Butterfield and Dalle-Donne, 2012) in agreement with the alternative "Error catastrophe" theory of aging. The negative nitrogen balance in the elderly might also play a role.

Occasionally in vivo changes may be mimicked also in vitro facilitating the investigative approaches. This is best observed in the case of mammalian red blood cells that have prolonged life spans (average half-life of 120 days in humans) and are probably cleared from circulation by immunologic mechanism through recognition of a "senescence antigen" related to posttranslational modification of membrane Band 3 protein, along with release of membrane vesicles and a decline in intracellular metabolites like ATP and 2,3-diphosphoglycerate.

Thus in these perspectives also

POSTTRANSLATIONAL PROTEIN MODIFICATIONS

at single sites or in a more general way through effects of protein polymerization-aggregation eventually coupled with decreased proteolysis (as they typically take place in the connective tissue and in organ fibrosis, or in the lens during senile cataract) might play roles in physio-pathological changes in animals and humans. It is remarkable that protein

PostTranslational Modification (PTM) might occur both in a reversible and in an irreversible way. Reversible PTM – that is usually linked to signal transduction pathways – is more difficult to investigate by proteomic techniques because of the intrinsic instability of the products, such as phosphoproteins, while irreversible protein PTM, such as those brought about by protein proteolysis, protein oxidation or by transglutaminase mediated modification at glutamine residues are usually more easily managed.
These issues have been developed more extensively in the preceding chapter.
A special technique borderline between experimental and clinical applications of proteomics is

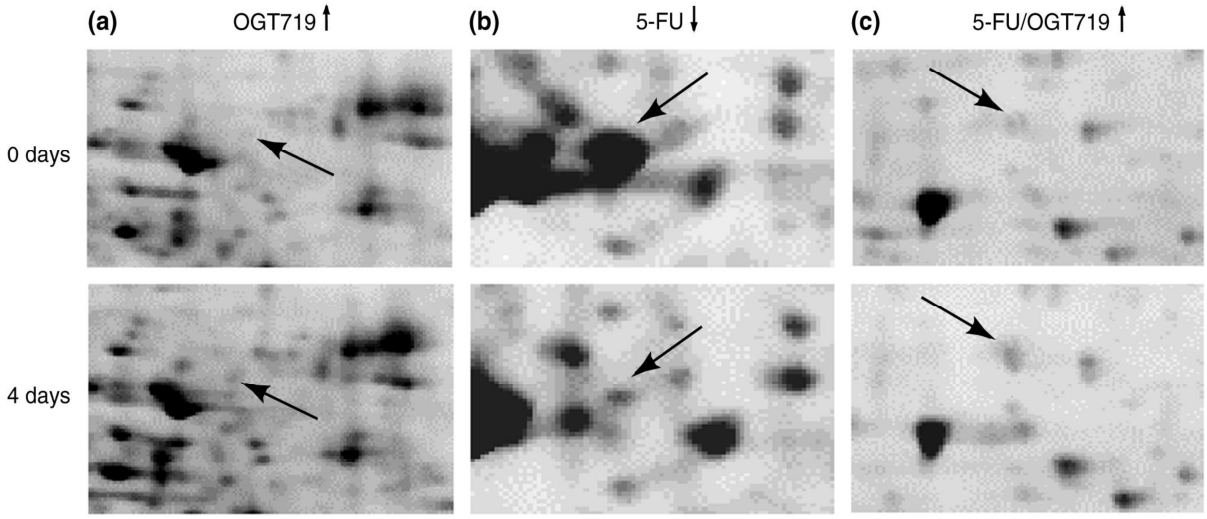
PHARMACOPROTEOMICS.

With this term we devise applications of proteomics in formal drug toxicology studies. As it is known most drugs are metabolized in the liver via the cytochrome P450 family of enzymes, which comprises a total number of ~200 different members, encompassing a wide array of overlapping specificities for different substrates. (Guengerich and Parikh 1997). They act during phase I of drug disposal promoting (beside drug clearance from plasma), also its metabolism that can eventually lead to the production and removal of toxic species. In some instances it was possible to correlate the ability or failure to remove such a toxin from circulatory fluids with a specific or a subgroup of P450 proteins. Each individual person will have a slightly different P450 profile, largely from polymorphisms and changes in expression levels, although other genetic and environmental factors aside from P450 also need to be taken into consideration (Vermes et.al, 1997). The marked variation of individuals in their ability to clear a compound can be one of the key factors in deciding the overall pharmacokinetic profile of a drug. Not only will this have a bearing on the likelihood of a patient responding to a treatment, but it will also be a factor in determining the possibility of experiencing an adverse effect and is also requirement to judge the drug half life and thereby its correct dosage.
In the case of the Pharmaproteomics approach for drug development, for instance animals can be dosed with increasing levels of an experimental drug over time, and serum samples can be drawn for consecutive proteome analyses. Using this procedure, it should be possible to identify individual markers, or clusters thereof, that are dose related and correlate with the emergence and severity of toxicity. Markers might appear in the serum at a defined drug dose and time that are predictive of early toxicity within certain organs and if allowed to continue will have damaging consequences. These serum markers could

subsequently be used to predict the response of each individual and allow tailoring of therapy whereby optimal efficacy is achieved without adverse side effects being apparent. This application can obviously extend to tracking toxicity of drugs in clinical trials where serum can be readily drawn and analyzed.

A relevant example in the field of proteomics applied to clinical pharmacology is represented by studies on OGT719. This is a novel galactosyl derivative of the cytotoxic agent 5-fluorouracil (5-FU), which is currently being developed for the treatment of hepatocellular carcinoma and colorectal metastases localized in the liver. The premise underpinning the design and rationale of OGT719 was to derive a 5-FU prodrug capable of targeting, and being retained in cells bearing the asialoglycoprotein receptor (ASGP-r), including hepatocytes, hepatoma Huh7 cells and some colorectal tumour cells. The growth of the human hepatoma cell line Huh7 is inhibited by either 5-FU or OGT719. If the inhibition by OGT719 were the result of uptake and conversion to 5-FU as the active component, then it would be expected that Huh7 cells would show similar proteome profiles following exposure of the cancer cell line to either drug, as depicted in the following figure

Additional relevant information might be provided by comparing the proteomes of cells in which the drug target has been eliminated by molecular knockout techniques, or by treatment with small molecule inhibitors believed to act specifically on the same target.



Proteins that are specifically up- or down-regulated in Huh7 cells by either 5-fluorouracil (5-FU) or OGT719: (a) elongation factor 1α2, (b) novel (three peptides by MS-MS) and (c) α-subunit of prolyl-4-hydroxylase. Arrows indicate up- or downregulated.

An even wider range of studies deals with proteomic research directed to management of

In most instances these studies deal with plasma and/or specific tissue proteome in relation to the classic goals of Clinical Biochemistry, to identify and measure the concentration of proteins useful in diagnosis and in assessment of prognosis, clinical course, response to therapy and risk of recurrence. This is particularly true in the case of chronic diseases, notably for cardiovascular and neoplastic diseases for which plasma proteome is the preferred research medium to investigate the presence of proteins whose "altered" levels correlate with the disease processes. Most importantly the results of threes investigations require to be clinically validated, i.e. the relevance of the identified proteins as disease biomarkers must correlate with clinical assessments. Improvements in the resolution, sensitivity, and mass accuracy of mass analyzers, an increase in the sample size used as the initial training set, and the combined use of multiple SELDI chip surfaces can all be expected to improve the specificity and predictive value of this approach. This technique can be used along with any number of other indicators such as genetic defects, or histopathologic findings, to make more accurate diagnoses. In any case, the initial findings using this method are encouraging, and to generate mass spectra requires only a small volume of serum and is relatively fast, taking as little as 30 min (Petricoin.E.F.et.al 2002). Equally encouraging is the opportunity to apply this approach to a broad spectrum of diseases. In this regard, a proteome "signature" has already been developed to discriminate individuals with prostate cancer from healthy individuals (Petricoin et al, 2002). The method of serum proteome pattern generation more comprehensively exploits the information contained within the serum proteome compared to single molecule identification, and the approach exploits a major characteristics of complex diseases, namely, that there is a whole cohort of molecular changes ranging from different protein levels to changes in protein cleavage and other modifications.

The concept behind this diagnostic approach is that the blood plasma proteome reflects tissue and organ pathology, causing patterns of protein changes that have diagnostic potential without even knowing the identities of the individual proteins. Since MS-based approaches provide a pattern of peaks, the idea is that these patterns can discriminate certain diseases and that this diagnostics involving the pattern or "signature" of the proteins rather than their identities might improve their clinical relevance (Petricoin and Liotta, 2004). To apply this approach, researchers have taken advantage of surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) analyzing for proteome diagnostic pattern samples of serum from affected and unaffected

patients individually applied to a matrix with analysis of retained proteins that are subsequently ionized and detected by TOF-MS. New computer-based artificial intelligence algorithms were used to identify patterns among a “training set” of mass spectral data, generated by SELDI-TOF-MS, for instance in the case of ovarian cancers. The algorithm generated a proteomic pattern that was then used to identify ovarian cancer in individuals from a second independent group individuals, some of whom were affected. In this instance the positive predictive value of the approach was 94%, the specificity was 95%, and the sensitivity was 100% .

In more general terms the rational of these studies is to compare the pattern of proteins release from the pathologic lesion into the blood stream to identify proteins which are landmarks of the disease either because of their unique features or thanks to variations in their levels in plasma.

Once their identity is achieved and their relevance is established in clinical validation studies they can likely supplement clinical assessment by classic routine immunologic analyses. This issue is virtually the same that was applied for detection of classic **cancer biomarkers** and newly discovered proteins will likely suffer the same limits of classic biomarkers, since they will probably be indicative more or organ of the lesion rather than of the lesion itself. Actually no true cancer biomarker really exists, although a few (PSA, ferritin and CA 125) tend to approach this goal. In any case classic as well as newly developed biomarkers are not themselves useful as diagnosis but are better employed for follow-up patients after effective therapy (levels of circulating markers should drop-down after therapy with a clearance kinetics reflecting their half-life in vivo) or to monitor local or systemic recurrence of the disease. In addition it must be appreciated that the rise of serum biomarkers might involve not only molecules directly released from pathologic lesions but also moieties reactive against the lesion as a result of its presence or possibly with the teleological role to limit its spreading. This is the case of cancer autoantibodies (Murphy et al, 2012).

Although, as already indicated, identification of temptative biomarkers usually proceeds through analysis of the plasma proteome of affected/unaffected individuals, there are also additional approaches. Among them it is important to recall those dealing with patterns of direct comparative protein analysis or of the secretome. In the direct comparison portions of both the lesion and of surrounding normal tissue are homogenized and extracted and their protein abundance is estimated by direct inspection in 2D gels or by techniques of fluorescence merging after the proteins in the normal tissue and the lesion are labeled by

different fluorophores. In the secretome approach (that begins to find wide application also in the field of cardiovascular diseases, see below for a more detailed description) the lesions are incubated for short periods of time in primary culture, usually without supplementation of serum, and the secretome the proteins that are secreted by the lesion (including also molecules that are shed from the surface) are collected for analysis, since they should mimic the release of proteins in vivo from the lesion. Proteins of secretome play a key role in cell signaling, communication and migration. The need for developing more effective cancer biomarkers and therapeutic modalities has led to the study of cancer cell secretome as a means to identify and characterize diagnostic and prognostic markers and potential drug and therapeutic targets. In addition, extracellular matrix components and other molecules secreted by tumor cells are a rich source of potential markers and drug targets for cancer treatment (May, 2009). Secreted proteins are also most likely excellent candidate serological tumor markers as they are released by the cells thus having highest possibility to enter the circulation (Kulasingam V, Diamandis EP 2008).

Analysis of secreted proteins is quite challenging; this may be attributed to technical difficulties such as a) presence of the proteins at frequently low concentrations due to their high dilution (in body fluids or cell culture media), b) their masking and contamination by cytoplasmic or other normally non secreted proteins released following cell lysis and death, and c) masking by serum proteins (i.e. fetal bovine serum) normally present in the culture media. These issues underscore the importance of procedures that have to be followed for cell culture and conditioned medium (CM) collection so as to avoid contamination from dead cells and cytosolic proteins as well as for effective protein concentration, regardless the protein separation technology in use.)

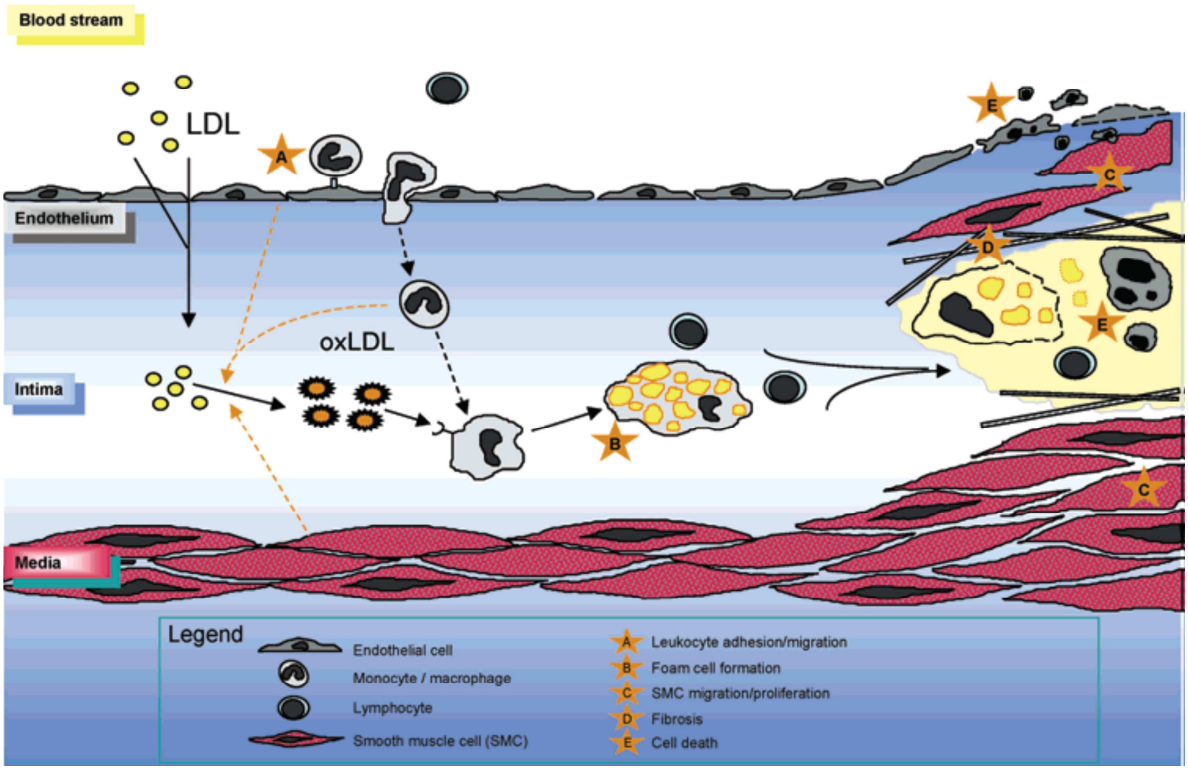
A drawback in the secretome approach is represented by possible tissue damage because of hypoxia during the time interval between surgery/biopsy and laboratory procedures. This time lapse could variably associate with cell necrosis or apoptosis so that the collected proteins might not belong to the secretome but to intracellular proteins released through damaged membranes.

Proteomic techniques have found wide applications also in the search of

NOVEL CARDIOVASCULAR BIOMARKERS

particularly in relation to Atherosclerosis the pathology resulting coronary heart disease and stroke, that represent the most common cause of death in developed countries. Atherosclerosis is an inflammatory process that results in the development of complex lesions or plaques, containing lipids that protrude into the arterial lumen. Therapeutic

programs have been developed (largely on the bases of the Framingham study) to check factors that promote development of atherosclerosis with the aim to limit its incidence and adequate biomarkers are available for this primary prevention. In contract major problems still in relation to assessment of the evolution of the atherosclerotic plaques, that may be two-sides leading to plaque remodeling (that means formation of an endothelial and fibrous cap that separating the lipid and inflammatory core of the lesion from the blood stream, protecting from additional damage) or the plaque instability and rupture that



proceeds through plaque ulceration and subsequently thrombosis resulting in the acute clinical complications of myocardial infarction (MI) and stroke, as represented schematically in the picture above, that shows the Arterial wall (composed of three layers: the intima, the innermost layer lined with endothelial cells (ECs) on the luminal side; the media, consisting of several layers of smooth muscle cells (SMCs) and the adventitia, which mainly consists of connective tissue) and the subendothelial wall invasion by oxidized LDL protein leading to inflammation and lesions of the upper layers.

Thus atherosclerosis is now looked at as a form of chronic inflammation resulting from interaction between lipoproteins (oxidized LDL; oxLDL), monocytes/macrophages, T lymphocytes, and the normal cellular elements of the arterial wall. This inflammatory process leads to the development of complex lesions or plaques that protrude into the

arterial lumen (Aikawa and Libby, 2004). It has been shown that endothelial cells (EC), upon activation, express adhesion molecules, that favor the attachment of monocytes which migrate into the subendothelial space and differentiate into macrophage. Uptake of oxLDL by these macrophages, via scavenger receptors (SR-A, CD36), leads to formation of foam cells and cytokines secreted by lymphocytes and macrophages, attract smooth muscle cells (SMC) which secrete extracellular matrix proteins forming the fibrous cap. Macrophages and SMC cell death by apoptosis or necrosis induced by a combination of injuries (oxLDL, proteolytic environment, etc.) leads to the formation of a necrotic core and accumulation of extracellular cholesterol. Secretion of matrix metalloproteinases (MMPs) and cathepsins by macrophages weaken the fibrous cap, favoring plaque rupture. This is dangerous because it exposes prothrombotic molecules as tissue factor, platelet adhesive matrix molecules to the blood, induce the formation of a thrombus, and elicit an acute coronary syndrome (Ruggeri, 2002). Therefore, plaque composition rather than the degree of arterial stenosis predicts the probability of rupture: large, fibrotic lesions can be relatively stable. In contrast, plaque rupture frequently occurs in relatively minor but lipid rich lesions with marked inflammatory infiltrates.

References

Aikawa M, Libby P. (2004) Atherosclerotic plaque inflammation: the final frontier? Can J Cardiol. 20: 631-634.

Anderson NL, Anderson NG 2002 The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 1: 845-867

Bergamini CM, Gambetti S, Dondi A, Cervellati C. (2004) Oxygen, reactive oxygen species and tissue damage. Curr Pharm Des. 10: 1611-1626

Berlett BS, Stadtman ER. (1997) Protein oxidation in aging, disease, and oxidative stress. J Biol Chem. 272: 20313-20316

Butterfield DA, Dalle-Donne I. (2012) Redox proteomics. Antioxid Redox Signal. 17: 1487-1489

Campbell DG, Morrice NA. (2002) Identification of protein phosphorylation sites by a combination of mass spectrometry and solid phase Edman sequencing. J Biomol Tech 13: 119-130.

Held JM, Gibson BW. (2012) Regulatory control or oxidative damage? Proteomic approaches to interrogate the role of cysteine oxidation status in biological processes. Mol Cell Proteomics. 11: R111.013037. doi: 10.1074/mcp.R111.013037.

Hayflick L. (1985) Theories of biological aging. Exp Gerontol.20 : 145-159

Guengerich FP, Parikh A. (1997) Expression of drug-metabolizing enzymes. Curr Opin Biotechnol. 8: 623-628

Kulasingam V, Diamandis EP. (2008) Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol. 5:588-99. doi: 10.1038/ncponc1187

Larose E, Yeghiazarians Y, Libby P, Yucel EK, Aikawa M, Kacher DF, Aikawa E, Kinlay S, Schoen FJ, Selwyn AP, Ganz P. (2005) Characterization of human atherosclerotic plaques by intravascular magnetic resonance imaging. Circulation. 112: 2324-2331

Liotta LA, Ferrari M, Petricoin E (2003) Clinical proteomics: written in blood. Nature 425: 905

May M. (2010) Biomarkers still off the mark for detecting breast cancer. Nat Med. 16:3. doi: 10.1038/nm0110-3

Murphy MA, O'Leary JJ, Cahill DJ. (2012) Assessment of the humoral immune response to cancer. J Proteomics. 75: 4573-4579

Nyström T, Yang J, Molin M (2012), Peroxiredoxins, gerontogenes linking aging to genome instability and cancer. Genes Dev. 26: 2001-2008

Ohlendieck K.(2011) Proteomic Profiling of Fast-To-Slow Muscle Transitions during Aging. Front Physiol. 2:105

Orrù S, Caputo I, D'Amato A, Ruoppolo M, Esposito C. (2003) Proteomics identification of acyl-acceptor and acyl-donor substrates for transglutaminase in a human intestinal epithelial cell line. Implications for celiac disease. J. Biol Chem. 278: 31766-31773.

Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. (2002) Clinical proteomics: translating benchside promise into bedside reality. Nat Rev Drug Discov. 1: 683-695.

Petricoin EF, Liotta LA (2004) Clinical proteomics: application at the bedside. Contrib Nephrol. 141: 93-103

Ruggeri ZM. (2002) Platelets in atherothrombosis. Nat Med.8: 1227-1234

Vermes A, Guchelaar HJ, Koopmans RP. (1997) Individualization of cancer therapy based on cytochrome P450 polymorphism: a pharmacogenetic approach. Cancer Treat Rev. 23:321-339

Chapter 5

TRANSGLUTAMINASE AND SITE SPECIFIC PROTEIN MODIFICATION.

GENERAL CONSIDERATIONS

As summarized before Transglutaminases (Tgases) have been for several years a main focus of research in the laboratory (Griffin et al, 2002; Bergamini et al, 2011), with particular emphasis on a specific isoform, Tgase 2, which is considered a multifunctional protein involved in (i) protein PTM at glutamine residues through isopeptide bonds with low molecular weight amines (transamidation) or epsilon amino groups of lysine residues in other proteins (protein crosslinkage); (ii) intracellular transduction of external signals acting as a G-protein with GTP/GDP cycling; and (iii) as scaffold for the assembly of structural protein in the extracellular matrix (ECM) (Bergamini et al, 2011).

Details in the reciprocate regulation of the transamidase and signalling have already been provided. Just to summarize the Tgase crosslinking activity is considered silent in normal conditions inside the cells because calcium levels are very low (in the micromolar range) and GTP levels are elevated, thus allowing normal cellular life span. Indeed phases of massive activation of Tgase 2 in the intracellular compartment have been associated with waves of apoptosis, along with proein PTM through isopeptide bonds which are resistant to degradation unless Tgase 2 in involved acting in the back reaction, which has been proved to occur only in the case of substrate proteins couplped with low molecular weight amines, not in protein crosslinkage.

In the present section I will consider Tgase 2 as an enzyme active in PTM at specific sites and analyze its reaction looking particularly at the identification of the glutamine residues targeted for modification, in proteins whose structure is known. The model protein chosen for these investigations are the muscle Troponin complex (that was already known to be a substrate of Tgase 2, but its exact site of modification had not been previously determined) (Bergamini et al, 1995; Gorza et al, 1996); brain tubulin - which is a prominent protein of cell cytoskeleton – was taken into account because cytoskeletal proteins are frequently preferred substrates of Tgases (Esposito and Caputo, 2005) and the CRM 197 mutant of the Diphteria toxin (DT). The choice of this last protein derives from an observation we obtained several years ago in experiments that were carried out in collaboration with colleagues at the ISI, because of the interest in DT as a vaccine and a drug for the specific

killing of neoplastic cells, upon targeting with cell specific ligands. The present experiments were carried out for safety reasons with the non toxic mutant CRM197 bearing the amino acid substitution E52G that inactivates the ADPribosylating activity without disrupting the basic structural properties of the protein.
The experiments on troponin T have already been published (Squerzanti et al, 2013) while those on the other proteins Tubulin and DT are in advancement but yet complete.

OUTLINE OF THE GENERAL PROCEDURES EMPLOYED FOR THE EXPERIMENTS OF PROTEIN LABELLING IN VITRO BY HUMAN TGASE

Tgase was purified from outdated human erythrocyte by a method developed in our laboratory which consists of lysis of the erythrocytes, adsorption and elution from DEAE-cellulose, PEG fractionation, and subsequent chromatography on DEAE sepharose and on heparin sepharose (Gambetti et al., 2005). The enzyme is homogeneous and can be stored at – 20°C in 20% glycerol without appreciable loss of activity for at least one year, after re-activation of the enzyme by treatment with DTT. Activity is assayed by a filter paper procedure recording incorporation at timed intervals of radioactive putrescine (Gambetti et al., 2005) or continuously of fluorescent dansylcadaverine into a standard suitable protein acceptor as dimethylcasein (Curtis and Lorand 1976), or the test protein as reported in each experiment. For the fluorescent assays excitation and emission wavelengths were 360 nm and 546 nm.
Troponin was isolated from rabbit muscle ether powders by the procedure of Potter (1982) while Tubulin (devoid of MAPs that were separated by chromatography on cellulose-phosphate) was purified from fresh porcine brain according Williams and Lee (1982), respectively. Stabilization of tubulin in the assembled and disassembled states was achieved by thermal cycling (Lee and Timasheff, 1975). The mutant Diphtheria toxin CRM 197 was a kind gift of Dr. Rino Rappuoli, from Novartis, Siena.
Incubations for incorporation of probe amines into the selected substrate proteins were carried out as detailed in the description of the experiments dealing with the specific protein. At the required time intervals samples of the native and of the labelled proteins were submitted to standard SDS-PAGE and the gel stained with colloidal Coomassie blue, a staining solution that does not interfere with the MS proteomic analysis, to identify labelled proteins (eventually evidenced by fluorescence staining under UV light), excise the band and digesting them with either trypsin or V8 proteinase, as reported in the specific

method section of the Troponin T investigation. The peptides released from the digests of the native protein and of the protein labelled with Spermine or with Dansyl-cadaverine were submitted to MALDI TOF and the peptides with differing *m/z* ratios were further submitted to MS/MS after ionization-induced fragmentation, to precisely identify the site of labelling. The identification of the residues labelled in the involved peptides can be achieved by applying the Mascot software as described elsewhere in this thesis.

STUDIES WITH TROPONINS.

Aim of the study

Troponins play relevant roles in the physiological regulation of contractility of skeletal and cardiac muscle and in monitoring of pathologic changes in muscles. From the physiological point of view they control calcium sensitivity and regulate the onset of filament sliding. These effects are brought about by modulation during ontogenesis of expression of different isoforms generated by alternative splicing of individual troponin subunits and by modification in PTM of troponins usually by phosphorylation. Both Troponin I and troponin T contribute to these regulatory effects (Toyota et al 2008; Oliveira et al 2012). From the pathologic point of view, assay of circulatory levels of cTnT and cTnI are relevant in monitoring and in diagnosis of cardiac damage, usually but not constantly due to ischemic attacks. Investigations on perfused hearts submitted to hypoxic stress (Gorza et al, 1996) revealed previously that immunoreactivity of Troponins was modified and that PTM of these proteins involved not only phosphorylation but also calcium dependent proteolysis and polymerization that was due to activation of Tgase 2.
The occurrence of this PTM in an in situ system and its relevance in cardiac physiology and pathology and eventually in molecular diagnostic in CV diseases prompted us to investigate in more detail the effects of Tgase 2 on muscle troponins.
Results of these studies have been published and the text of the paper is reproduced below.

Amino Acids (2013) 44:227–234
DOI 10.1007/s00726-011-1144-3

ORIGINAL ARTICLE

The side chain of glutamine 13 is the acyl-donor amino acid modified by type 2 transglutaminase in subunit T of the native rabbit skeletal muscle troponin complex

Monica Squerzanti · Carlo Cervellati · Blendi Ura · Carlo Mischiati · Piero Pucci · Stefano Annunziata · Carla Iannone · Rita Casadio · Carlo M. Bergamini · Carla Esposito

Received: 28 July 2011 / Accepted: 28 October 2011 / Published online: 16 November 2011

Abstract Subunit T of the native muscle troponin complex is a recognised substrate of transglutaminase both in vitro and in situ with formation of isopeptide bonds. Using a proteomic approach, we have now determined the precise site of in vitro labelling of the protein. A preparation of troponin purified from ether powder from mixed rabbit skeletal muscles was employed as transglutaminase substrate. The only isoform TnT2F present in our preparation was recognised as acyl-substrate by human type 2 transglutaminase which specifically modified glutamine 13 in the N-terminal region. During the reaction, the troponin protein complex was polymerized. Results are discussed in relation to the structure of the troponin T subunit, in the light of the role of troponins in skeletal and cardiac muscle diseases, and to the rules governing glutamine side chain selection by tissue transglutaminase.

Keywords Transglutaminase · Skeletal troponin T · Protein post-translational modification

M. Squerzanti · C. Cervellati · B. Ura · C. Mischiati · C. M. Bergamini (✉)
Department of Biochemistry and Molecular Biology, University of Ferrara, Via Borsari 46, 44121 Ferrara, Italy
e-mail: bgc@unife.it

P. Pucci · S. Annunziata · C. Iannone
CEINGE Biotecnologie Avanzate and Department of Organic Chemistry and Biochemistry, University of Naples Federico II, Naples, Italy

R. Casadio
Department of Biology, Biocomputing Group, University of Bologna, Bologna, Italy

C. Esposito
Department of Chemistry, University of Salerno, Baronissi, SA, Italy

Introduction

Transglutaminases (Tgases) are enzymes which catalyse the modification of proteins at peptidyl-glutamine residues by a strictly calcium-dependent transfer of the acyl-moieties of the glutamine amido group to acceptor primary amines, with formation of isopeptide bonds. When the acyl-acceptors are constituted by soluble low-molecular weight amines, such as polyamines (PA), the reaction products are post-translationally modified proteins (protein transamidation). Alternatively, in the case the amine acceptor is the ε-aminogroup of protein-bound lysine residues, the Tgase reaction originates aggregates of polymerized proteins, which are frequently insoluble (protein crosslinkage) (Folk and Finlayson 1977).

$$\text{Protein} - \text{CONH}_2 + \text{PA} \rightarrow \text{protein} - \text{CONH} - \text{PA} + \text{NH}_3 \quad (1)$$

$$\text{Protein} - \text{CONH}_2 + \text{protein} - \text{NH}_2 \rightarrow \text{protein} - \text{CONH} - \text{protein} + \text{NH}_3. \quad (2)$$

The balance between the transamidation (1) and the crosslinkage (2) reaction catalysed by tissue transglutaminase (Tgase 2) is likely linked to the intracellular concentrations of PA (Lentini et al. 2007), which are the preferred amine substrates in transamidation. In this respect, we must remind that PA are considered "physiologic" substrates for transamidation (Folk et al. 1980) since the Km of Tgase 2 for polyamines falls in the cell physiologic range of polyamine concentration and PA-γ-glutamyl adducts are easily detectable in digested tissue extracts.

Tgase 2-mediated transamidation involves a few selected glutamine residues in substrate proteins (Esposito and Caputo 2005), because of strict rules of specific selection

not yet understood. Also the functional consequences of the type 2 Tgase-mediated protein transamidation are matter of discussion, since both enhancement and inhibition of the activity of modified proteins have been reported. Indeed, the original hypothesis that transglutaminases act as protein-modifying enzymes triggering physiologic responses by altering function in modified proteins (Esposito and Caputo 2005; Facchiano and Facchiano 2009) is now considered valid only in relation to the functions of Tgase 2 in the extracellular milieu. In contrast this "functional" effect is questionable in the intracellular space, as we discussed recently (Bergamini et al. 2011), since the transamidation reaction is irreversible. Tgase 2 is actually a cryptic enzyme under basic intracellular conditions, particularly because of effective inhibition by GTP coupled to the low concentrations of the essential activator calcium (Griffin et al. 2002). Under these conditions, the role of Tgase 2-dependent protein modifications is unlikely to affect protein/enzyme interconversion for metabolic regulation. These considerations seem to apply to all Tgase-mediated modifications at glutamine residues, including their hydrolysis to glutamic acid which is probably relevant in the pathogenesis of coeliac disease.

To further clarify the structural bases for the Tgase 2-mediated modification of peptidyl-glutamine residues, we explored the structural features around the site of labelling in proteins which are recognised Q-substrates for Tgases. In the present report we describe studies on skeletal muscle troponins, which we investigated in the past identifying the T subunit (TnT) as the preferred substrate of Tgase 2 in vitro and in situ (Bergamini et al. 1995; Gorza et al. 1996).

Materials and methods

Proteins

The troponin protein complex was purified from ethyl ether powders of rabbit skeletal muscle (Potter 1982) and Tgase 2 from human RBC (Gambetti et al. 2005) with the modification to incorporate a DTT-activation step before the final chromatographic separations. Sequence grade trypsin was obtained from Sigma-Aldrich.

Labelling of Tn subunits by Tgase 2

The reaction was carried out by incubating Tn complex (1–2 mg/ml) with purified Tgase in the presence of 5 mM calcium (and 1 mM spermine, when specified). Stoichiometry of glutamine labelling was checked in parallel incubations by measuring incorporation of radioactive putrescine into the troponin protein as previously described (Bergamini et al. 1995). To identify the glutamine residue whose side chain is modified by transglutaminase, incubations carried out in the presence of cold spermine were terminated by boiling in the presence of Laemmli denaturation buffer, and the troponin subunits were resolved by SDS-PAGE (Laemmli 1970). Relevant bands in gels stained according to (de Laurentiis et al. 2006), were excised from the gel, digested with trypsin and mass analysed by MALDI TOF and MALDI TOF–TOF techniques.

In-gel-digestion and MS identification of peptides

Coomassie blue-stained bands excised from the gel were washed in sequence with acetonitrile and with 50 mM ammonium bicarbonate pH 8.0. Protein samples in gel fragments were reduced by incubation in 10 mM dithiothreitol in ammonium bicarbonate for 45 min at 56°C and alkylated with 55 mM iodoacetamide in ammonium bicarbonate (30 min at room temperature in the dark) (de Laurentiis et al. 2006) and further washed with ammonium bicarbonate and acetonitrile. Enzymatic digestions were carried out with trypsin solution, 10 μg/ml, in 50 mM ammonium bicarbonate, pH 8.0, at 4°C for 1 h and further digested at 37°C for 18 h. A minimum reaction volume sufficient for complete rehydration of the gel was used. Released soluble peptides were acidified with 20% trifluoracetic acid (TFA), lyophilized, and resuspended in 10 μl of 2% TFA; 1 μl was loaded in the appropriate wells of the mass spectrometer plate. Peptides were mixed with 1 μl of matrix (α-cyano-4-hydroxyycinnamic acid 10 mg/ml) dissolved in a mixture of 70% acetonitrile and 30% TFA (0.2%).

MALDI-MS and MS/MS spectra were acquired using a 4800 plus MALDI TOF TOF mass spectrometer (ABI Sciex, USA). Protein identification was carried out by peptide mass fingerprinting using an in house version of the Mascot software. For troponins characterization, mass signals were mass mapped onto the anticipated sequences of rabbit skeletal muscle troponins as available in the Expasy site (entry P02641 for Troponin T).

Computational analysis

The sequence of Troponin subunits was numerated from the first amino acid of the coding sequence excluding the initial methionine present in the Swiss Prot data bank. Pairwise and multiple alignment of protein sequences was performed with CLUSTALW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) setting the BLOSUM protein weight matrix (Thompson et al. 1994). Disordered regions in protein sequences were predicted with DISOPRED2 (http://www.bioinf.cs.ucl.ac.uk/index.php?id=806) (Ward et al. 2004).

Results

Modification of troponins by Tgase

In these experiments, we proceeded with the identification of the reaction products of troponins following incubation with Tgase in the presence of spermine. The Tn complex was treated in vitro with purified Tgase in the presence of calcium ions and 1 mM spermine for 3 h at room temperature. The putative modified proteins were fractionated by SDS-PAGE. Figure 1 shows the corresponding gel stained with Coomassie blue, where the Tgase-modified troponins (lane M) were compared to a sample of unmodified troponins used as control (lane C).

Comparison of the two gel lanes showed a similar band patterns in the medium and low MW regions of the SDS-PAGE, whereas the troponins sample was specifically characterized by the occurrence of various bands in the high MW region that were absent in the control lane. Relevant bands in both the control and sample lanes were excised from the gel and submitted to the identification procedure as described in "Materials and methods".

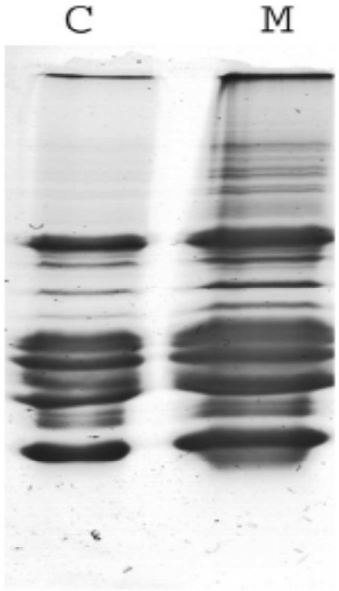


Fig. 1 Electrophoretic separation of troponin subunits by SDS-PAGE. The *lanes* labelled *C* and *M* refer, respectively, to a sample of control troponin and to a sample of troponin modified for 3 h by transglutaminase in the presence of calcium and spermine, under the conditions detailed in the "Materials and methods" section. The load of protein in the *lane M* was double than that in *lane C*

The peptide mass fingerprinting procedure led to the unambiguous identification of the proteins occurring in the three gel bands present in the medium–low MW regions of the gel. As reported in Fig. 1, the first band in this region corresponded to troponin T (Swiss Prot entry P02641) with the two other faster bands being troponin I (P02643) and troponin C (P02586), respectively. The pattern of protein identification obtained by MS analyses of the gel bands both in the control and sample lanes revealed that the troponin preparation contained the expected mixture of the three troponin subunits. Despite the theoretical possibility that multiple isoforms of the subunits are present as a consequence of the metabolic and functional heterogeneity of rabbit skeletal muscle (troponin T, the tropomyosin-binding subunit of troponin, in particular is reported to exist as nine isoforms produced by alternative splicing), the PMF procedure unambiguously identified the isoform present in the gel band as isoform 4 (also called isoform TnT2F).

Protein bands occurring in the high MW region were tentatively assigned to aggregates of troponin T generated by Tgase action. In addition to simple labelling of glutamine residues by incorporation of PA in the troponins monomer, it is well known that Tgase 2 also catalyses crosslinking of troponins into high MW polymers. These are likely stabilised by isopeptide bonds involving either directly lysine residues (Folk and Finnlayson 1977) or additional glutamine residues through a polyamine bridge to yield, under our conditions, aggregates stabilised by glutamyl-spermine-glutamine isopeptide bonds (Martinet et al. 1990). To confirm these assumptions, we tried to analyse peptides released by extensive trypsin proteolysis of the high MW bands. However, this approach was unsuccessful either because of the low amount of protein material or because of the resistance to proteolysis of the high MW aggregates. This is occasionally the case in protein aggregates generated by transglutaminase because the likely blockage of lysine residues and the tight assembly of the protein aggregates prevent efficient cleavage by trypsin (Ruoppolo et al. 2003). Conversely, it is unlikely that the crosslink involves a glutamine-PA-glutamine bridged isopeptide since sensitivity to proteolysis by trypsin should theoretically not be affected.

Identification of the specific Gln residues recognised by Tgase

Peptide mixtures from the tryptic digest of the protein bands corresponding to troponins in the Tgase-treated sample (lane M in Fig. 1) were carefully analysed to identify the Gln residues involved in transamidation in each individual troponin subunit.

Tryptic peptides from the gel band corresponding to troponin T were analysed by MALDI TOF and the corresponding mass signals mapped onto the anticipated sequence of the troponin T isoforms previously identified by PMF (entry P02641-4 in the Swiss Prot database). All mass signals recorded in the spectra could be assigned to predicted peptides from the troponin T sequence leading to the coverage of the entire protein primary structure. A single mass signal at m/z 5,247.6 could not be associated to any peptide within the troponin T sequence and was then tentatively interpreted as a Tgase-modified peptide.

The recorded mass value was 185.4 Da higher than the peak at 5,062.2 corresponding to the acetylated form of the N-terminal peptide 2-44, where Met 1 had been deleted. This mass difference could account for a spermine molecule covalently attached to the N-terminal peptide via an isopeptide bond catalysed by Tgase, with release of ammonia. Since this peptide contains a single Gln residue at position 13, this amino acid should be considered as the amino acceptor substrate of Tgase 2. However, this hypothesis had to be confirmed by tandem mass spectrometric analyses. The mass signal at m/z 5,247.6 was then isolated within the mass spectrometer and submitted to MS/MS analysis. Although the peptide mass value was particularly high, a very clear series of y ions could be detected in the daughter ion spectra, as shown in Fig. 2. Interpretation of the fragment ions led to confirm that this peptide indeed corresponded to the acetylated fragment 2-44. Moreover, the occurrence of the two key fragment ions at m/z 3,822.9 and 3,510.6 belonging to the y series clearly demonstrated that Gln13 was covalently modified by a spermine molecule via an isopeptide bond.

Similar analyses were carried out on the gel bands corresponding to troponin I and C from lane M of the gel depicted in Fig. 1. No altered peptides were detected whatsoever in the spectra of the tryptic digests corresponding to both isoforms where all the Gln residues were shown to be unmodified by Tgase. These data indicated that only troponin T is substrate of Tgase 2, with Gln13 at the N-terminus of the protein being the only labelled glutamine residue.

The structure of TnT around the site of labelling

The computational analysis of the region involved in labelling of the glutamine residue highlights some clues on the criteria for selection of glutamine residues by human Tgase 2. First of all, we compared the N-terminal sequences of the 3 subunits of skeletal troponin TnT, TnC and TnI. All these subunits include glutamine residues within the N-terminal sequence, as shown in Fig. 3a where the first 50 aminoacids of the three sequences are depicted. The most obvious peculiarity of the sequence of TnT around the labelled Gln 13 is the cluster of acidic residues (the sequence is rich in glutamic acid residues). This feature is not present in the other Tn subunits (TnI and TnC), while it is maintained also in other variants of TnT which are generated by alternative gene splicing (Fig. 3b). In these instances, the sequence preceeding Gln 13 (MSDEEVEHVEEEAQ) of the TnT2F variant is converted into the sequence MSDEEVEHVEEQYEEEEEAQ due to the expression of a glutamine containing peptide EEQYEE. On the basis that local information as derived from the primary structure is relevant for the recognition of TnT variants, the glutamine recognition should occur also in the other variants which conserve the same sequence motive around the conserved glutamine (Fig. 3b). We do not have this information yet because only the variant TnT2F was present in our preparation of troponins from rabbit fast skeletal muscle. Notably in human Annexin 1 which is a recognised transglutaminase substrate, Gln 19 is also surrounded by glutamic acid residues. It has been demonstrated in the pig counterpart, which shares a 88.7% sequence identity with human Annexin 1, that an identical Gln 19 flanked by glutamic acid residues is present in the N-terminal region. This region is disordered in the presence of calcium, but it undergoes folding to a flexible α helix in the absence of the cation (Rosengarth et al. 2001; Rosengarth and Luecke 2003).

To investigate additional features of TnT2F that might participate in glutamine recognition, we have considered secondary structure and local flexibility. Since these parameters are not directly available from studies in the crystalline state, we have taken advantage of predictors to gain this information. According to our data (Fig. 3a, where disordered residues are shown in italics), Gln 13 is predicted to be located in a long disordered region. This region should be then characterized by an appreciable flexibility, since the analysis of segmental "rumor" corroborates the notion that both substrate regions in TnT2F and human Annexin I are very disordered. In turn, the N-terminal regions of TnI and TnC are characterized by considerable rigidity. The analogous regions in the other skeletal muscle TnT variants which are presumed to be potential substrates are characterized by analogously high flexibility (data not shown).

Discussion

Gard and Lazarides (1979) described labelling of structural proteins of the Z-line in glycerinated chicken myofibrils by Tgase 2 through incorporation of the fluorescent probe Dansylcadaverine in the presence of calcium. The authors identified α-actinin, tropomyosin and actin as the labelled proteins. Labelling of actin was later employed to

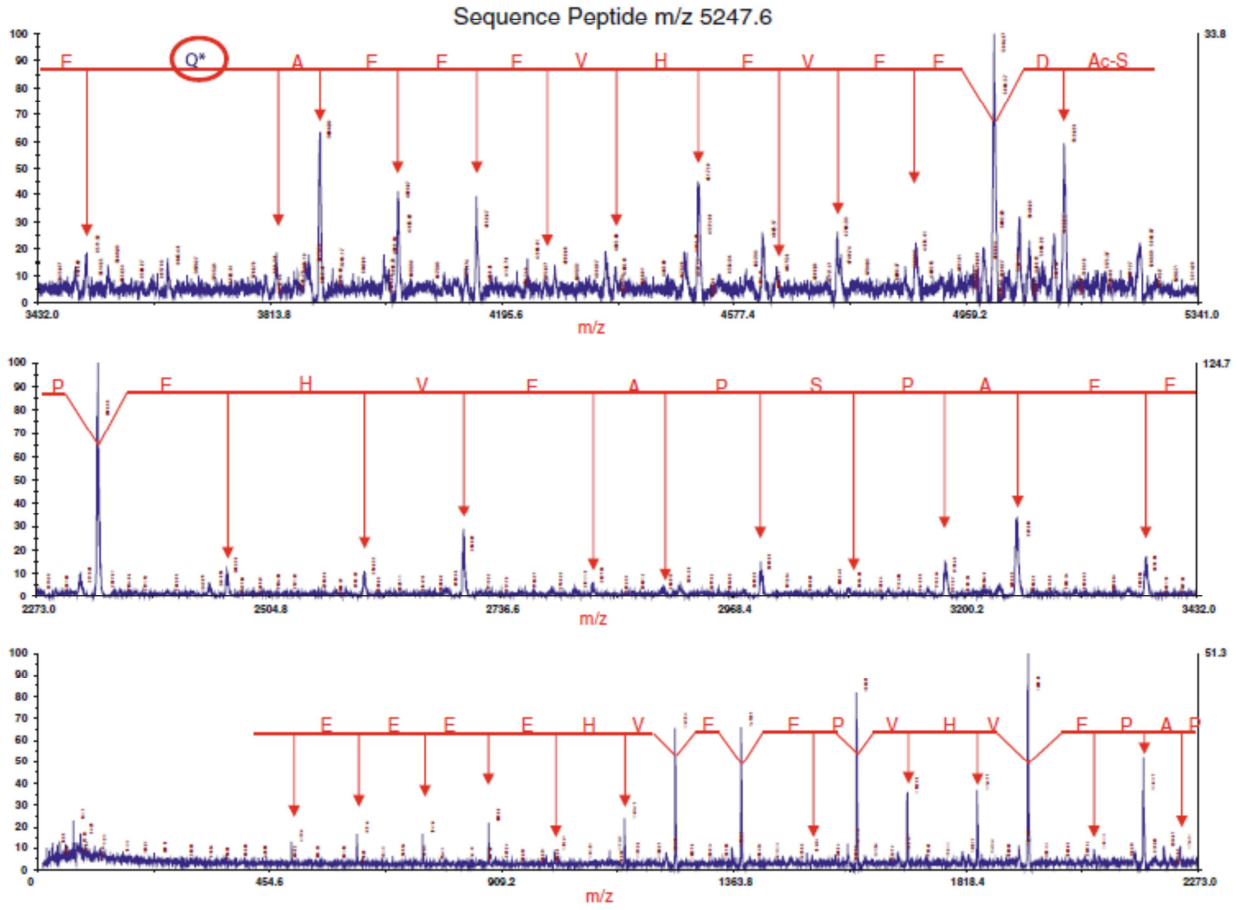


Fig. 2 Pattern of transamidation of glutamine residues in subunit T of skeletal troponin. The band corresponding to the monomeric subunit T of troponin was excised from the gel in Fig. 1 and processed for fragmentation MS as detailed in the "Materials and methods" section. The sequence identified in the only modified peptide with m/z 5,247.6 is reported in this picture starting from the top-right corner. The missing glutamine (residue Q13) is marked by *asterisk*

Fig. 3 **a** Sequence alignment of the amino-terminal region of the rabbit Tn subunits TnT, TnI and TnC. The sequence labelled in the TnT subunit is marked in *bold*, while regions of elevated disorder are presented in *italics*. **b** Comparison of the sequence of the known isoforms of the rabbit TnT subunit to document the conservation of the labelled region in TnT2F also in the other variants

A

```
P02641-4|TNNT3_RABIT   MSDEEVEHVEEEAQEEAPSPAEVHEPAPEVHVPEEVHEEEKPRPKLTAPK 50
P02643|TNNI2_RABIT     MGDEEKRNRAITARRQH LKSVMLQIAATELEKEEGRREAEKQ--NYLAEH 48
P02586|TNNC2_RABIT     MTDQQAEARSYLS-EE MIAEFKAAFDMFDADGGGDISVKELGTVMRMLGQ 49
                       * *:: .    : .:          : .      *        :
```

B

```
P02641      MSDEEVEHVEEQYEEEEAQEEAPSPAEVHEPAPEVHVPEEVHEDALEDMREEEEEEEKPR
P02641-2    MSDEEVEHVEEQYEEEEAQEEAPSPAEVHEPAPEVHVPEEVH------------EEEKPR
P02641-3    MSDEEVEHVEEQYEEEEAQEEAPSPAEVHEP-------EEVHEDALEDMREEEEEEEKPR
P02641-4    MSDEEVEHV------EEEAQEEAPSPAEVHEPAPEVHVPEEVH------------EEEKPR
P02641-5    MSDEEVEHVEEQYEEEEAQEEAPSPAEVHEP-------EEVH------------EEEKPR
P02641-6    MSDEEVEHV------EEEAQEEAPSPAEVHEP-------EEVH------------EEEKPR
P02641-7    MSDEEVEHVEEQYEEEEAQEE-----------------EEVHEDALEDMREEEEEEEKPR
P02641-8    MSDEEVEHVEEQYEEEEAQEE-----------------EEVH------------EEEKPR
P02641-9    MSDEEVEHV------EEEAQEE-----------------EEVH------------EEEKPR
            ********       *******                  ****            ******
```

investigate actin polymerization which proceeds unaltered, since the label does not grossly interfere with the functional properties of actin itself provided that an intramolecular crosslink between Q41 and K50 is not formed (Eli-Berchoer et al. 2000). At the same time, we investigated labelling of troponins by Tgase 2 in the perfused heart after calcium stress (Gorza et al. 1996), and in vitro employing skeletal muscle troponin as substrate (Bergamini et al. 1995). In our hands, troponin T was the main site of labelling. At variance with our results,

McDonough et al. (1999) reported that also TnI is a substrate for Tgase 2 and that it is crosslinked to the C-terminal (190-275) proteolytic fragment of TnT into aggregates which are released in heart perfusates following infarction/reperfusion injury. Such discrepancies prompted us to clarify what is the precise site of labelling and what factors can play a role about, since troponins are clinically relevant markers of muscular and cardiac cell dysfunction (Panteghini 2009).

The present findings confirm and extend our previous observations since modification of troponin occurs only at subunit T, labelling glutamine 13 at the N-terminus of the single isoform present in the protein preparation we have used. Identification of this glutamine Q13 as the site of labelling was achieved by analysing products of incubations carried out in the presence of exogenous polyamines. This glutamine is flanked by acidic residues (which might mitigate the effects of the positive charges brought about by labelling by Tgase in the case polyamines act as acyl-acceptors) at the end of a loose and flexible α-helix. As mentioned in the text, this same motive is present and substrate of Tgase 2 in human Annexin 1 (Pepinsky et al. 1989) can undergo reversible shifts between a random coil and a mobile flexible α-helix in porcine Annexin 1 (Rosengarth and Luecke 2003). These considerations are consistent with hypotheses that either sequence-charge or mobility effects are relevant for glutamine recognition (see also below). In contrast to the definitive identification of Glutamine 13 as the acyl-donor aminoacid, we could not reach any conclusion on the identity of acyl-acceptor in the crosslinked products because of the marked resistance of the polymer to proteolysis, which might be due to an intrinsic resistance of the polymer to degradation or of the inability of trypsin to digest aggregated proteins whose lysine residues are blocked in isopeptide bonds, as outlined in the "Results" section. We will investigate further this issue by means of additional proteinases of different specificity (e.g. the staphylococcal V8 protease) or of artificial glutamine acyl-donors such as CBZ-glutaminylglycine, which is an usual peptidyl-glutamyl substrate of Tgase 2. This last approach was useful in previous studies on the identification of substrate lysine residues using polyglutamine tails in other systems (Ruoppolo et al. 2003).

An issue deserving consideration is related to the specificity of glutamine labelling which might arise from (i) sequence and 3D structure around the labelled glutamine and from (ii) possible variation of substrate activity in relation to additional protein post-translational modifications. We must remind that in this respect controversial data have been published. In the case substrates are libraries of short random peptides containing Q residues (Sugimura et al. 2006), issues based on primary structure motives are probably relevant since preferential labelling of glutamines occurs at residues flanked by hydrophobic residues on the C-terminal side. In these studies, however, contributions arising from the limited folding of these short (decameric) peptides have been disregarded. In contrast, combined data obtained on intact proteins behaving as glutaminyl-substrates (Esposito and Caputo 2005), further summarised in the TRANSDAB database available at the University of Debrecen (Csosz et al. 2009), indicate that labelling takes place independently of primary structure, but are rather linked to the 3D arrangement, at variance with what happens with other protein modifying enzymes, e.g. the protein kinases (Pinna and Ruzzene 1996) which recognise sites of specific labelling from the primary structure of the protein substrate. The 3D structure is probably the predominant factor in dictating substrate specificity of mammalian transglutaminases since for instance labelling of fibronectin by Tgase 2 and by Factor XIII occurred at multiple different sites and was increased by limited proteolysis of the substrate, likely because of exposition of additional sites hindered in the native protein or of conformational changes (Fesus et al. 1986). These effects, which might associate with focal loss of compact structure, might explain the discrepancies between these and our previous in vitro findings (Bergamini et al. 1995) and those reported by McDonough et al. (1999) in the perfused heart under ischemic conditions, since in this last model activation of proteolytic cascades and nicking of substrate proteins likely occur. It is thus apparent that sequence issues are relevant determinants of glutamine-specific recognition only in short unstructured peptides but not in large proteins in which protein secondary and tertiary structures are important for selectivity.

Data obtained with bacterial transglutaminase support this view. For instance, it has been proved (Mero et al. 2009) that bacterial Tgase is selective in modifying only one (Q134) out of the 17 glutamine residues within the 174 amino acids sequence of granulocyte-CSF. According to these authors, the most relevant parameter in Q-selectivity is represented by protein segmental flexibility around the substrate glutamine (estimated from a high value of crystallographic B-Factor), which dictates fitting of the substrate region in the enzyme active site, as it happens in the case of proteinases. In our instance, the glutamine residue acted upon by human transglutaminase in TnT4 is harboured in a region of high flexibility, as it happens for a mutant of yeast phosphoglycerate kinase (Coussons et al. 1991).

Other interesting implications might be related to the fact that only troponin T is acted upon by Tgase 2. This subunit is crucial in dictating the functional properties of the calcium-dependent ATPase activity of myosin. Indeed, the functional differentiation of muscle in fast- and slow-contacting forms is accompanied by shifts in TnT expression, which also occur during development in skeletal and

in cardiac muscles (Chaudhuri et al. 2005; Feng and Jin 2010). As mentioned, the TnT isoforms have different calcium sensitivity, and chiefly differ for their N-terminal regions (Wang and Jin 1998), actually that susceptible to the Tgase-mediated modification. If we also take into account the associations between over-expression of Tgase 2 and susceptibility to cardiac failure in transgenic animals (Small et al. 1999) as well as the presence at high title of anti-Tgase 2 antibodies in serum of patients with end-stage cardiac failure (Peracchi et al. 2002), it is possible that different tissular contents of expressed TnT isoforms with different reactivity towards Tgase-mediated transamidation, might be relevant in clinical pathology. Clearly, further experiments are required to verify these intriguing hypotheses, at the light of their possible clinical relevance and the three-dimensional assembly of the subunits of troponin complex, which is the object of intensive investigation in several laboratories (Kowlessur and Tobacman 2010).

Conflict of interest The authors declare that they have no conflict of interest.

References

Bergamini CM, Signorini M, Barbato R, Menabò R, Di Lisa F, Gorza L, Beninati S (1995) Transglutaminase-catalyzed polymerization of troponin in vitro. Biochem Biophys Res Commun 206:201–206

Bergamini CM, Collighan R, Wang Z, Griffin M (2011) Structure and regulation of type 2 transglutaminase in relation to its physiological functions and pathological roles. Advanced Enzymology, vol 78. John Wiley, New York, pp 1–46

Chaudhuri T, Mukherjea M, Sachdev S, Randall JD, Sarkar S (2005) Role of the fetal and alpha/beta exons in the function of fast skeletal troponin T isoforms: correlation with altered Ca2+ regulation associated with development. J Mol Biol 352:58–71

Coussons PJ, Kelly SM, Price NC, Johnson CM, Smith B, Sawyer L (1991) Selective modification by transglutaminase of a glutamine side chain in the hinge region of the histidine-388—glutamine mutant of yeast phosphoglycerate kinase. Biochem J 273:73–78

Csosz E, Meskó B, Fésüs L (2009) Transdab wiki: the interactive transglutaminase substrate database on web 2.0 surface. Amino Acids 36:615–617. http://www.genomics.dote.hu/wiki/

de Laurentiis A, Caterino M, Orrù S, Ruoppolo M, Tuccillo F, Masullo M, Quinto I, Scala G, Pucci P, Palmieri C, Tassone P, Salvatore F, Venuta S (2006) Partial purification and MALDI-TOF MS analysis of UN1, a tumor antigen membrane glycoprotein. Int J Biol Macromol 39:122–126

Eli-Berchoer L, Hegyi G, Patthy A, Reisler E, Muhlrad A (2000) Effect of intramolecular cross-linking between glutamine-41 and lysine-50 on actin structure and function. J Muscle Res Cell Motil 21:405–414

Esposito C, Caputo I (2005) Mammalian transglutaminases. Identification of substrates as a key to physiological function and physiopathological relevance. FEBS J 272:615–631

Facchiano A, Facchiano F (2009) Transglutaminases and their substrates in biology and human diseases: 50 years of growing. Amino Acids 36:599–614

Feng HZ, Jin JP (2010) Coexistence of cardiac troponin T variants reduces heart efficiency. Am J Physiol Heart Circ Physiol 299:H97–H105

Fesus L, Metsis ML, Muszbek L, Koteliansky VE (1986) Transglutaminase-sensitive glutamine residues of human plasma fibronectin revealed by studying its proteolytic fragments. Eur J Biochem 154:371–374

Folk JE, Finlayson JS (1977) The epsilon-(gamma-glutamyl) lysine crosslink and the catalytic role of transglutaminases. Adv Protein Chem 31:1–133

Folk JE, Park MH, Chung SI, Schrode J, Lester EP, Cooper HL (1980) Polyamines as physiological substrates for transglutaminases. J Biol Chem 255:3695–3700

Gambetti S, Dondi A, Cervellati C, Squerzanti M, Pansini FS, Bergamini CM (2005) Interaction with heparin protects tissue transglutaminase against inactivation by heating and by proteolysis. Biochimie 87:551–555

Gard DL, Lazarides E (1979) Specific fluorescent labeling of chicken myofibril Z-line proteins catalyzed by guinea pig liver transglutaminase. J Cell Biol 81:336–347

Gorza L, Menabò R, Vitadello M, Bergamini CM, Di Lisa F (1996) Cardiomyocyte troponin T immunoreactivity is modified by cross-linking resulting from intracellular calcium overload. Circulation 93:1896–1904

Griffin M, Casadio R, Bergamini CM (2002) Transglutaminases: nature's biological glues. Biochem J 368:377–396

Kowlessur D, Tobacman LS (2010) Troponin regulatory function and dynamics revealed by H/D exchange-mass spectrometry. J Biol Chem 285:2686–2694

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lentini A, Forni C, Provenzano B, Beninati S (2007) Enhancement of transglutaminase activity and polyamine depletion in B16–F10 melanoma cells by flavonoids naringenin and hesperitin correlate to reduction of the in vivo metastatic potential. Amino Acids 32:95–100

Martinet N, Beninati S, Nigra TP, Folk JE (1990) N1N8-bis(gamma-glutamyl) spermidine cross-linking in epidermal-cell envelopes. Comparison of cross-link levels in normal and psoriatic cell envelopes. Biochem J 271:305–308

McDonough JL, Arrell DK, Van Eyk JE (1999) Troponin I degradation and covalent complex formation accompanies myocardial ischemia/reperfusion injury. Circ Res 84:9–20

Mero A, Spolaore B, Veronese FM, Fontana A (2009) Transglutaminase-mediated PEGylation of proteins: direct identification of the sites of protein modification by mass spectrometry using a novel monodisperse PEG. Bioconjug Chem 20:384–389

Panteghini M (2009) Assay-related issues in the measurement of cardiac troponins. Clin Chim Acta 402:88–93

Pepinsky RB, Sinclair LK, Chow EP, O'Brine-Greco B (1989) A dimeric form of lipocortin-1 in human placenta. Biochem J 263:97–103

Peracchi M, Trovato C, Longhi M, Gasparin M, Conte D, Tarantino C, Prati D, Bardella MT (2002) Tissue transglutaminase antibodies in patients with end-stage heart failure. Am J Gastroenterol 97:2850–2854

Pinna LA, Ruzzene M (1996) How do protein kinases recognize their substrates? Biochim Biophys Acta 1314:191–225

Potter JD (1982) Preparation of troponin and its subunits. Methods Enzymol 85B:241–263

Rosengarth A, Luecke H (2003) A calcium-driven conformational switch of the N-terminal and core domains of annexin A1. J Mol Biol 326:1317–1325

Rosengarth A, Gerke V, Luecke H (2001) X-ray structure of full-length annexin 1 and implications for membrane aggregation. J Mol Biol 306:489–498

Ruoppolo M, Orrù S, Francese S, Caputo I, Esposito C (2003) Structural characterization of transglutaminase-catalyzed cross-linking between glyceraldehyde 3-phosphate dehydrogenase and polyglutamine repeats. Protein Sci 12:170–179

Small K, Feng JF, Lorenz J, Donnelly ET, Yu A, Im MJ, Dorn GW 2nd, Liggett SB (1999) Cardiac specific overexpression of transglutaminase II (G(h)) results in a unique hypertrophy phenotype independent of phospholipase C activation. J Biol Chem 274:21291–21296

Sugimura Y, Hosono M, Wada F, Yoshimura T, Maki M, Hitomi K (2006) Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGASE 2 and Factor XIIIA. J Biol Chem 281:17699–17706

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wang J, Jin JP (1998) Conformational modulation of troponin T by configuration of the NH2-terminal variable region and functional effects. Biochemistry 37:14519–14528

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337:635–645

Comments to the study on Troponin labelling

A significant aim of our research on protein labelling by human Tgase is the elucidation of the rules that dictate substrate recognition by the enzyme. As mentioned rules of recognition of glutamine residues are very selective while those for the acyl-acceptor are very flexible so that both primary amines (either amino groups in proteins or soluble low molecular weight amines, as well) can act on thio-ester catalytic intermediate. The data we have obtained in the present investigation indicate that the modified glutamine 13 is located in the N-terminal region and that the flanking sequence is particularly reach in acidic side chains because of clustering of glutamic acid residues.

In any case it is likely that this is not the only parameter relevant for the selection of substrate glutamine residues since Q13 in TnT is present in a flexible region, as it happens also for other proteins that are glutamine substrate of both microbial and eukaryotic Tgases. Actually, it is also because of our interest in these mechanisms of substrate recognition that we decided to expand this section of the research, as it is demonstrated below by the studies on labelling of tubulin ands of DT, that are described below.

The second point that deserves some comments is related to the possibility that Troponin undergoes

PTM by Tgase in living cells and to the possible functional consequences of this modification. Previous experiments demonstrated that this PTM may occur in the beating perfused heart since forms of TnT with high molecular weight consistent with the formation of dimers or even larger aggregates are detectable in Western blot of ischemic myocardial cytosol mostly if undergoing “calcium paradox” and stunning. What instead is not known is if polyamines, which are considered physiologic or para-physiologic Tgase substrates (Folk et al, 1980), may play any role in this perspective. The issue of cardiac polyamine metabolism has been already the object of investigation but not been so much explored in the perspective of ischemic/riperfusion injury damage, while data are available in relation to the pressure induced damage (e.g. Cetrullo et al, 2011), under conditions in which cardiac overload is known to affect expression of Tgase 2 (Iwai et al, 1995;Small et al, 1999; Nanda et al, 2001; Deasy et al, 2013). This is certainly an important topics and further studies are required to delineate completely its relevance beyond the present feeling of a general protective action of polyamines.

STUDIES WITH TUBULINS.

General aspects

Microtubules, largely composed of tubulin and associated proteins, exist in the cell cytosol as two main populations, that of the true microtubules which are the thickest cell filaments present at relatively stable levels in the inter-mitotic phase contributing to maintenance of cell shape and that of tubules of the mitotic spindles which are assembled in cells entering the mitosis employing also tubulin released from disassembled cytosolic microtubules (de Forges et al, 2012). In addition tubulin based cables contribute to cell trafficking of vesicles and organelles (Garnham and Roll-Mecak, 2012). This apparent stability of inter-mitotic microtubules is actually a dynamic effect (it is also called "dynamic instability") and arises from the contemporaneous occurrence of processes of assembly-disassembly at the extremities with different polarity of the microtubules. Posttranslational modifications (acetylation, phosphorylation, glutamylation, tyrosylation, etc) of tubulin in vivo by a number of enzymes and interaction of the microtubules with microtubule-associated-proteins (MAPs) are important processes in determining the relative size of cellular tubulin pools (Hammond et al. 2008).

As already remembered many proteins of the cytoskeleton are efficient Tgase substrates (Griffin et al, 2002; Esposito and Caputo, 2005). This observation and a formal similarity between the cycling mechanisms of assembly/disassembly of the microtubules and the actin microfilaments suggested us to investigate whether also brain tubulin might be an efficient substrate for Tgase 2 in vitro and to investigate in detail the conditions and the sites of labelling, employing chiefly proteomic approaches to obtain this important information. Indeed our results confirm these assumptions proving in particular that (i) Tgase 2 catalyzes incorporation of radioactive/fluorescent amines into tubulin and conjugation of tubulin itself with model glutamine peptide substrates; (ii) neither disassembled tubulin nor polymerized tubulin filaments are involved in formation of crosslinks most likely because of three dimensional hindrance between reactive side chains, despite a more efficient incorporation of the fluorescent label dansylcadaverine into polymerized than depolymerised tubulin; (iii) incorporation of primary amines occurs at multiple glutamine residues in subunit β, prevalently at several positions, at the C-terminal end of the peptide; (iv) aggregated ("polymerized") is more effective than disassembled tubulin in the incorporation of amine probes at glutamine residues.

EXPERIMENTAL PROCEDURES

To investigate modification of brain tubulin, the protein (1–2 mg/ ml) was incubated with purified Tgase 2, 5 mM calcium and 1 mM spermine or o.2 mM Dansylcadaverine as specified. Stoichiometry of glutamine labelling was checked measuring incorporation of radioactive putrescine into the tubulin protein by a filter paper assay (Bergamini et al. 1995). To identify the targeted glutamine residues incubations carried out in the presence of cold spermine or dansylcadaverine were terminated by boiling and the tubulin subunits were resolved by standard SDS-PAGE. Relevant bands in gels stained with colloidal Coomassie blue were excised, the proteins submitted to tryptic digestion in situ, and the peptide mass analysed by MALDI TOF and MALDI TOF–TOF techniques, as described for the labelling of muscle troponins. As we encountered difficulties in the analysis of labelled tubulin by this approach of in situ proteolytic digestion, also digestion in solution was carried out.

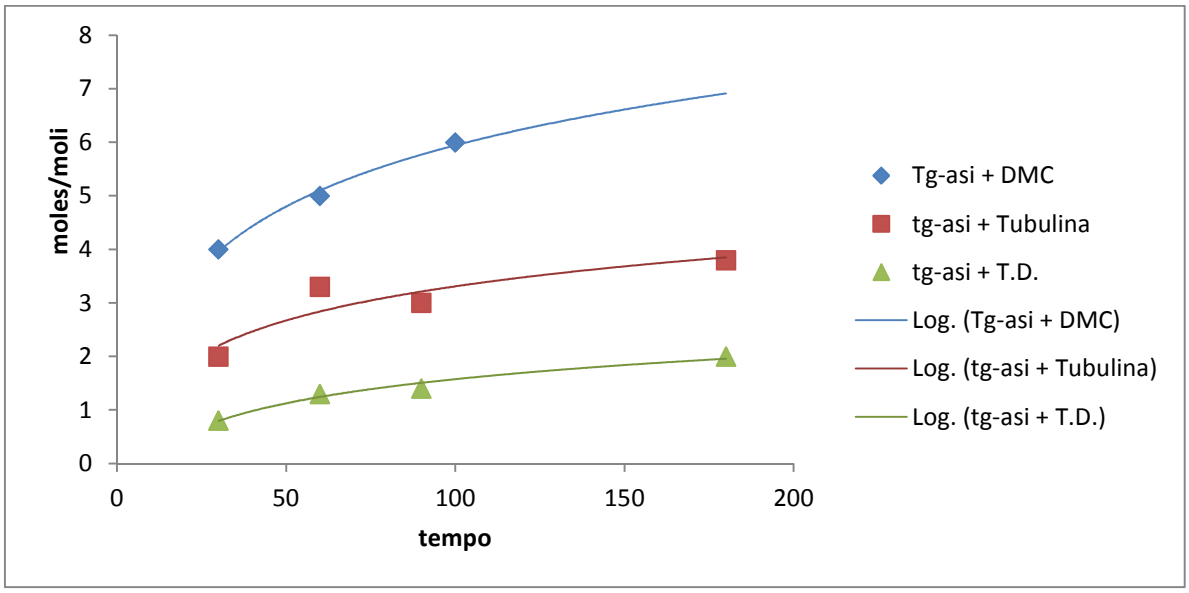
When searching for reactive lysine residues we have employed as probe the peptidyl-glutamine analogue Cbz-Gln-Gly, in experiments similar to those aimed to identify reactive glutamine residues. In this case bands in gel were processed by digestion with V8 proteinase and further processed as for tryptic peptides.

RESULTS

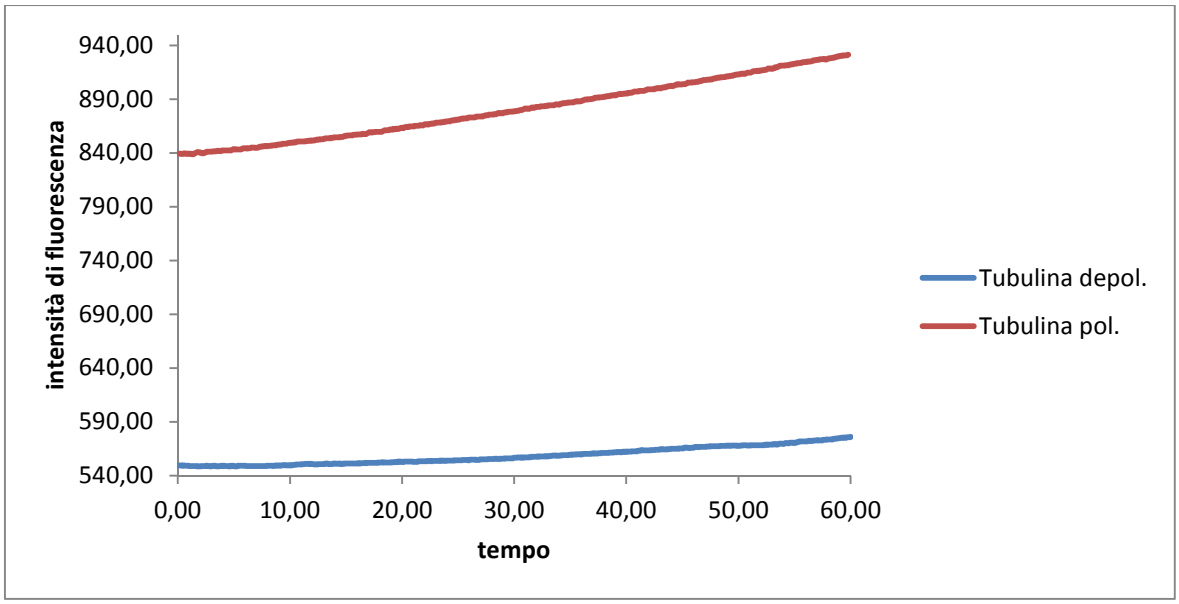
Labelling of glutamine and lysine residues in tubulin

I will employ the term monomeric tubulin to identify the complex of two α,β subunits in the unpolymerized form of tubulin. To incorporation of primary amines into brain tubulin was checked by incubating tubulin (1-2 mg/ml) with Tgase 2, in the presence of 5 mM calcium and either 0.5 mM radioactive putrescine or dansylcadaverine (which is fluorescent). Both radioactive and fluorescent assays were performed on the aggregated tubulin form (see Figure 5.1, in which data of the assay of incorporation of radioactive putrescine are presented as a function of time of incubation employing as substrates tubulin, Dyphtheria toxin and dimethylcasein as a model substrate). The data clearly demonstrate that both proteins tubulin and DT are glutamine substrates of Tgase 2, although incorporation levels are clearly lower than those obtained with dimethylated casein. Assays of tubulin labelling were also carried out by the continuous fluorimetric assay that was performed on both polymerized and depolymerised tubulin by changing the temperature in the instrument cell, verifying it by recording turbidity at 350 nm. The

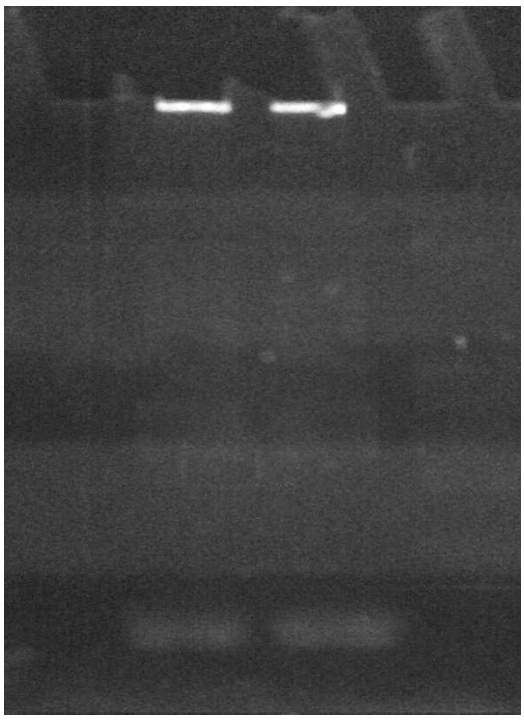
incorporation of radioactive putrescine (Fig. 5.1) yielded an apparent stoichiometry of 4 amine groups for monomer,



while the fluorescence assay confirmed effective incorporation of the probe and suggested in addition significantly higher labelling in tubulin assembled in filaments as compared to free disaggregated monomeric tubulin (Figure 5.2).



The analysis of the pattern of fluorescent labelling of tubulin in SDS denaturing electrophoresis demonstrated that the glutamine residues recognized as substrates by Tgase 2 are contained mostly if not exclusively in the β subunit, although very faint fluorescence is present also in higher aggregates and on the top of the stacking gel, because



of some form of major insolubility of the product (see Fig. 5.3).

MS spectroscopy analysis of the peptide released of gel fragments containing the tubulin subunits labelled with Dansylcadaverine that had been resolved by SDS-PAGE yielded a consistent coverage of the sequence of both subunits (over 80% of the sequence was covered) but we did not succeed in resolving the structure of the C-terminal region by in gel proteolysis with trypsin of the tubulin monomer labelled either with spermine or dansyl-cadaverine. This failure is probably dependent on poor ionizability of the C-terminal region, which likely is the labelled one because it is rich in glutamine residues. To overcome these difficulties we have already performed additional experiments in which proteomic search of the incorporated labels was performed both by MALDI-TOF analysis of peptides released by in gel proteolysis with V8 proteinase and by in solution proteolysis followed by ESI mass spectroscopy. Results are still in progress and will be the object of a future report.

In future studies it will be important to explore the nucleating ability of modified tubulin and the regular occurrence of thermal cycling for the conversion between the disaggregated and the aggregated states in polyamine modified tubulin as well.

Comments on the studies with tubulin

Our investigation with tubulin prove definitely that also this cytoskeletal protein is labelled by Tgase 2, but failed up to now to identify the exact site of labelling, in contrast with the relatively straight forward analysis of the labelling of TnT. This clearly demonstrates that the procedures of analysis must be must be carefully fitted to the experimental problem under investigation taking into account the properties of the protein that is employed in the labelling experiments.

Besides their interest at the light of the features of Tgase 2 as a protein modifying enzyme, these data wait for clarification for their functional interest (if any) at the light both of polyamines as physiologic substrate of Tgases and the general roles of polyamines as regulators of cell growth (Folk at al, 1980) and differentiation (Wallace, 2009).

Another point which clearly deserves investigation in this optics is represented by an analysis of transamidation of tubulin by polyamines in relation to the polymerization/depolymerization cycling and to the in vivo labelling of tubulin by polyamines as a novel PTM of this protein. This could be looked at more likely by analysis of the effects of increased calcium availability in the intracellular compartments of cell depleted or enriched with polyamines.

STUDIES WITH DIPHTHERIA TOXIN

General aspects

Diphteria toxin (DT) is a highly cytotoxic toxin produced by *Corynebacterium diphtheriae* as a monomeric single chain pro-toxin of 535 amino acids, that is cleaved by proteinases to the active toxin containing two fragments (A-B), further linked by a disulphide bridge. The N-terminal region (fragment A of 193 amino acids in the proteolyzed DT) is the catalytic C domain, i.e. the region toxic to the eukaryotic cells, while the remaining 342 amino acids in the C-terminal moiety contain two domains T and R that are involved in membrane translocation (T domain) of the C domain and in Receptor-mediated (R domain) anchorage to target cells, usually through heparin binding EGF receptors. DT is highly toxic bringing about death of host cells by blocking the protein biosynthesis machinery through inhibition of eukaryotic Elongation Factor 2 (eEF2). The inhibition is mediated by ADP ribosylation of histidine residues of eEF2 to the so-called diphtamide residues by the catalytically active domain C of the toxin after internalization. (Collier, 2001).

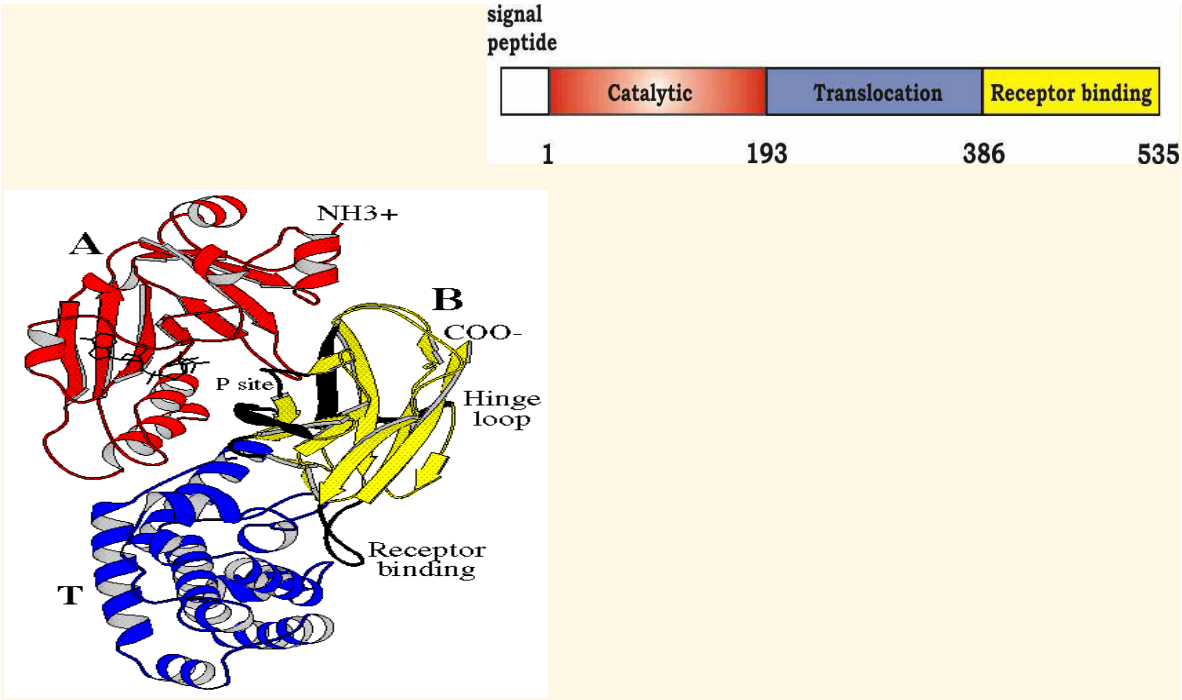


Fig. 5.4 The domain structure of the diphtheria toxin

Several features that have attracted interest of researchers in DT, include (i) its involvement in the pathogenesis of diphtheria; (ii) the mode of action as an effective blocker of protein synthesis; (iii) the availability of toxin mutants that retain immunogenic activity but are devoid of toxicity (including CRM 197, see below); (iv) the possibility to employ the T-R domains as a vehicle of peptides to the intracellular space (just like penetratin) and (v) the possibility to employ the catalytic C domain as a highly efficient weapon against cancerous cells, with various biotechnologic pharmaceutical applications. Thus the isolated domains of DT are suitable to be exploited by genetic manipulation to generate fusion proteins either between the translocating T-B moiety and appropriate therapeutic peptides (e.g. antibodies or cytokines fragments) or through fusion of the toxic A domain with ligands targeting cell-specific receptors, particularly for therapy of leukemias and lymphomas, sparing normal cells (Frankel et al, 2000; FitzGerald et al, 2004).

Since, as mentioned, we had proved previously in unpublished experiments that DT is a substrate of Tgase 2, we found of interest to evaluate the possibility to conjugate DT with carriers by the simple enzymatic procedure of treatment with Tgase with peptides, low molecular weight chemicals (e.g. drugs) or polymers (such as PEG) that might alter in vivo half-life (Fontana et al, 2008) or with peptides that might direct it to specific targets.

In this preliminary research I investigated the properties of DT as substrate of Tgase 2 with

aim to localize the sites of labelling as a prerequisite for specific coupling strategies. The studies have been carried out with the non toxic DT variant CRM 197 for the safe of the investigators since this mutants largely maintains the structure of the wild type toxin (Malito et al, 2012). At this purpose I wish to recall that the available mutants of DT have been obtained using the mutagenesis with N - methyl - N' - nitro - N – nitrosoguanidine, generating toxins whose functionality was altered but showed cross-reactivity with polyclonal antibodies anti – DT. These " Cross -reacting material" (CRM) can be divided into two groups depending on alterations in the fragment B with the function of receptor binding and translocation (CRM 30, 45, 228, 107, 102, 103, 9 and 1001) or in fragment A with impaired enzymatic activity (CRM 197, 228 and 176). CRM197 presents the substitution G52E, i.e. replacement of glycine 52 with glutamic acid, that prevents binding of the cofactor NAD^+ with almost complete loss of catalytic activity (Giannini et al., 1984).

EXPERIMENTAL METHODS

Modification of CRM 197 by Tgase 2 was carried essentially by the same general procedures outlined above at concentrations of DT between 0.2 and 0.5 mg/ml, in the presence of Tgase 2, Dansylcadaverine (or spermine, or Carboxybenzoyl-QG) and 5mM $CaCl_2$, monitoring the progress of the reaction by fluorescence assay (λ_{ex} 340 and λ_{em} 546), by radioactive incorporation of by electrophoretic analysis of protein polymerization. Procedure for proteomic analysis have been already described.

RESULTS

Initial attempts to screen CRM 197 as substrate of type 2 Tgase involved incorporation of radioactive or fluorescent amines into the protein, thus acting as Q substrate. The results of the radiochemical assay (using putrescine as the incorporated probe) demonstrated that CRM 197 is an appreciable glutamine substrate for tissue Tgase (see Fig 5.1, above) and were confirmed with the fluorescent assay, displaying time-dependent increase in the intensity of fluorescence emission at 546 nm increased in a time dependent way in incubations containing Tgase, CRM 197, and dansylcadaverine as the fluorescent probe. All these results therefore agree that some glutamine residues of CRM 197 are recognized as substrate by mammalian Tgase (results not shown).

A minimal stoichiometry of labelling was provided by incorporation of radioactive putrescine in the filter paper assay, with an apparent value slightly higher than one mole/mole despite the possible availability of free lysine residues that might compete with the exogenous amine as acyl-acceptors of the reactive glutamine residue(s). The contemporaneous occurrence of protein crosslinkage competing with the incorporation of external amines is likely as shown by progressive accumulation of high molecular mass protein aggregates which failed to enter the separation gel along the time of incubation as proved by SDS gel electrophoresis (Fig.5.5). This clearly suggests an aggregation of CRM 197 occurring through formation of γ-glutamyl-ε-lysine isopeptide bonds. As expected addition of amine (spermine) or glutamine (Cbz-QG) exogenous substrates inhibit protein crosslinkage (see also figure 3)

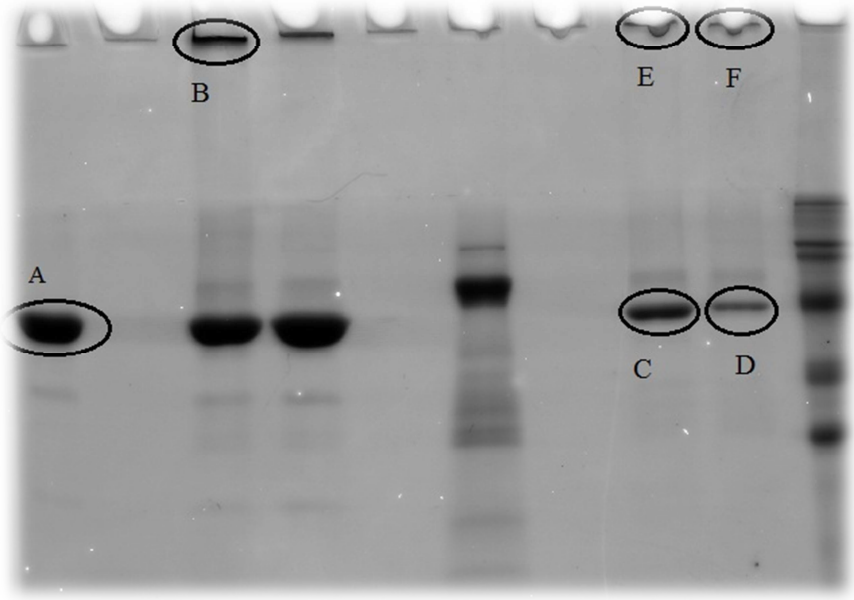


Fig. 5.5 Electrophoretic separation of DT by SDS-PAGE. **A**. Diphtheria Toxin**. B**. Crosslinking between DT and Tg-ase after 30 minutes of incubation. **C**. DT with Tg-ase and spermine after 30 minutes. **D**. DT with Tg-ase and spermine after 120 minutes. **E**. Crosslinking between DT, Tg-ase and spermine after 30 minutes of incubation**. F**. Crosslinking between DT, Tg-ase and spermine after 120 minutes of incubation.

We searched further for the specific sites of glutamine and lysine labelling within the sequence of CRM 197. For this issue we resolved electrophoretically incubation mixtures containing Tgase, calcium and either spermine or Cbz-QG to isolate the band of residual monomeric toxin, which presumably was at least partially labelled by the probes at selectively label glutamine and lysine residues of the toxin. In all instances bands were excised from the gel and submitted to proteomic analysis after peptide chain fragmentation in situ by treatment with trypsin.

The MS analysis of these data is still in progress.

References

Bergamini CM, Signorini M, Barbato R, Menabò R, Di Lisa F, Gorza L, Beninati S. (1995) Transglutaminase-catalyzed polymerization of troponin in vitro. Biochem Biophys Res Commun. 206: 201-206.

Bergamini CM, Collighan RJ, Wang Z, Griffin M. (2011) Structure and regulation of type 2 transglutaminase in relation to its physiological functions and pathological roles. Adv Enzymol Relat Areas Mol Biol. 78: 1-46.

Cetrullo S, Tantini B, Facchini A, Pignatti C, Stefanelli C, Caldarera CM, Flamigni F. (2011) A pro-survival effect of polyamine depletion on norepinephrine-mediated apoptosis in cardiac cells: role of signaling enzymes. Amino Acids. 40: 1127-1137

Collier RJ. (2001) Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. Toxicon. 39: 1793-1803.

Curtis CG, Lorand L. (1976) Fibrin-stabilizing factor (factor XIII). Methods Enzymol. 45: 177-191

Deasey S, Shanmugasundaram S, Nurminskaya M. (2013) Tissue-specific responses to loss of transglutaminase 2. Amino Acids. 44: 179-187

de Forges H, Bouissou A, Perez F (2012) Interplay between microtubule dynamics and intracellular organization. Int J Biochem Cell Biol. 44: 266-274

Esposito C, Caputo I. (2005) Mammalian transglutaminases. Identification of substrates as a key to physiological function and physiopathological relevance. FEBS J. 272: 615-631

FitzGerald DJ, Kreitman R, Wilson W, Squires D, Pastan I. (2004) Recombinant immunotoxins for treating cancer. Int J Med Microbiol. 293: 577-582.

Folk JE, Park MH, Chung SI, Schrode J, Lester EP, Cooper HL. (1980) Polyamines as physiological substrates for transglutaminases. J Biol Chem. 255: 3695-3700

Fontana A, Spolaore B, Mero A, Veronese FM. (2008) Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. Adv Drug Deliv Rev. 60: 13-28.

Frankel AE, McCubrey JA, Miller MS, Delatte S, Ramage J, Kiser M, Kucera GL, Alexander RL, Beran M, Tagge EP, Kreitman RJ, Hogge DE. (2000) Diphtheria toxin fused to human interleukin-3 is toxic to blasts from patients with myeloid leukemias. Leukemia. 14: 576-585.

Gambetti S, Dondi A, Cervellati C, Squerzanti M, Pansini FS, Bergamini CM. (2005) Interaction with heparin protects tissue transglutaminase against inactivation by heating and by proteolysis. Biochimie. 87: 551-555.

Garnham CP, Roll-Mecak A. (2012) The chemical complexity of cellular microtubules: tubulin post-translational modification enzymes and their roles in tuning microtubule functions. Cytoskeleton 69: 442-463

Giannini G, Rappuoli R, Ratti G. (1984) The amino-acid sequence of two non-toxic mutants of diphtheria toxin: CRM45 and CRM197. Nucleic Acids Res. 12: 4063-4069.

Gorza L, Menabò R, Vitadello M, Bergamini CM, Di Lisa F. (1996) Cardiomyocyte troponin T immunoreactivity is modified by cross-linking resulting from intracellular calcium overload. Circulation 93: 1896-1904

Griffin M, Casadio R, Bergamini CM. (2002) Transglutaminases: nature's biological glues. Biochem J. 368: 377-396

Hammond JW, Cai D, Verhey KJ. (2008) Tubulin modifications and their cellular functions. Curr Opin Cell Biol. 20, 71-76

Iismaa SE, Mearns BM, Lorand L, Graham RM. (2009) Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. Physiol Rev. 89: 991-1023

Iwai N, Shimoike H, Kinoshita M. (1995) Genes up-regulated in hypertrophied ventricle. Biochem Biophys Res Commun. 209: 527-534.

Lee JC & Timasheff SN. (1975) The reconstitution of microtubules from purified calf brain tubulin.
Biochemistry. 14: 5183-5187

Malito E, Bursulaya B, Chen C, Lo Surdo P, Picchianti M, Balducci E, Biancucci M, Brock A, Berti F, Bottomley MJ, Nissum M, Costantino P, Rappuoli R, Spraggon G. (2012) Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197. Proc Natl Acad Sci U S A. 109: 5229-5234

Nanda N, Iismaa SE, Owens WA, Husain A, Mackay F, Graham RM. (2001) Targeted inactivation of Gh/tissue transglutaminase II. J Biol Chem. 276: 20673-20678

Oliveira SM, Zhang YH, Solis RS, Isackson H, Bellahcene M, Yavari A, Pinter K, Davies JK, Ge Y, Ashrafian H, Walker JW, Carling D, Watkins H, Casadei B, Redwood C. (2012) AMP-activated protein kinase phosphorylates cardiac troponin I and alters contractility of murine ventricular myocytes. Circ Res. 110: 1192-1201

Pappenheimer AM Jr. (1977) Diphtheria toxin. Annu Rev Biochem. 46: 69-94.

Potter JD (1982) Preparation of troponin and its subunits. Methods Enzymol 85B: 241–263

Small K, Feng JF, Lorenz J, Donnelly ET, Yu A, Im MJ, Dorn GW 2nd, Liggett SB. (1999) Cardiac specific overexpression of transglutaminase II (G(h)) results in a unique hypertrophy phenotype independent of phospholipase C activation. J Biol Chem. 274: 21291-21296

Squerzanti M, Cervellati C, Ura B, Mischiati C, Pucci P, Annunziata S, Iannone C, Casadio R, Bergamini CM, Esposito C. (2013) The side chain of glutamine 13 is the acyl-donor amino acid modified by type 2 transglutaminase in subunit T of the native rabbit skeletal muscle troponin complex. Amino Acids. 44: 227-234

Toyota N, Takano-Ohmuro H, Yoshida LS, Araki M, Yoshinobu K, Suzuki-Toyota F. (2008) Suppression of cardiac troponin T induces reduction of contractility and structural disorganization in chicken cardiomyocytes. Cell Struct Funct. 33: 193-201

Wallace HM. (2009) The polyamines: past, present and future. Essays Biochem. 46: 1-9

Williams RC Jr, & Lee JC. (1982) Preparation of tubulin from brain. Methods Enzymol. 85 Pt B, 376-385

Chapter 6

Experimental Investigations: Oncologic studies

Proteomics is emerging as an important field of investigation for almost every clinical issue. There are several relevant examples – as mentioned in the introductory chapters in this thesis – but the most relevant applications are those related to severe chronic diseases as cancer and cardiovascular diseases (CVD), which are at the same time frequent causes of morbidity and mortality. In the case of CVD, for instance proteomics is increasingly applied (White and Van Eyk, 2007) to assess difficult clinical entities as staging chronic heart failure (CHF) and risk stratification for atherosclerotic disease, in relation to the so-called "stability" of the atherosclerotic plaque i.e. its relative tendency to the formation of a fibrous cap and re-endothelization (the favourable evolution) or the formation of superficial ulcers that might progress towards thrombosis and formation of emboli, the usual cause of acute coronary syndrome and of stroke, depending on the affected arterial branch (Blanco-Colio et al 2009). The aim is that of the discovery of novel biomarkers of disease/risk for their clinical validation through analysis of the plasma proteome or eventually of the secretome of cultured vessels, since secretome is the pathway by which cells (in this case vascular cells) release specific proteins in the circulating blood stream (Kullo and Cooper, 2010). The close contact between vascular layering cells and plasma makes these approaches particularly attractive in the field of cardiovascular diseases.

Besides the field of CVD, major applications of clinical proteomics arise from the study of malignant tumours, again for selecting novel specific biomarkers for direct clinical application for tumour staging in terms of prognosis and of response to clinical therapy or of markers of basic biologic behaviour. In relation to direct clinical use, novel biomarkers are expected to perform better as guide-lines in patients stratification particularly in selection of adequate therapeutic protocols as for the sensitivity to radiotherapy by analysis of serum proteome (Cowan and Vera, 2008; Scaife et al, 2011) or chemotherapy through expression of MDR protein (Zheng et al, 2010) eventually by immunohistochemistry in the pathologic lesions or other markers in draining fluids (Kuang et al, 2012). Also tissues might be analyzed. In his perspective it is notable that data on experimental animals demonstrated tissue specific responses in gene expression (Lim et al, 2011). An interesting and important case in humans is represented by the triple negative phenotype in breast tumours, i.e. the negativity for E_2R, PgR and HER-2 (human epithelial growth factor

receptor-2) that accounts for more roughly 15% of total cases and poses great problems for its resistance to specific tailored therapies, as I will discuss further below. Notably present proteomics research is expanding utility of tumour markers also to cancer for which markers haven't yet been discovered and validated as for instance renal cell cancer (Craven et al, 2012).

In addition to these clinically oriented applications, appropriate careful applications of proteomics deal with the defining new emerging concepts in modern oncologic biology, those of cancer stem cells and of the conversion of epithelial-like into mesenchymal-like cells, the so-called EMT (Epithelial Mesenchimal Transition) and its reverse MET. These processes probably/frequently associate with vascular invasion and metastasis by the haematic rather than lymphatic routes, and with the onset of resistance to chemotherapy, as well as with the additional concept of the cancer stem cells. At the same time proteomic analysis of tissues from animals submitted the experimental radiation injury revealed possible involvement of EMT in the pathogenesis of radiation induced lung fibrosis (Almeida et al, 2013) a clinically relevant complication in patients undergoing thoracic radiation for malignancies.

DISCOVERY OF NOVEL PERFORMING CANCER BIOMARKERS BY PROTEOMICS. THE CASE OF THE INFILTRATING DUCT CARCINOMA OF THE BREAST

During the years, starting from the beginning of the '40ies last century, with the investigations of alkaline and acid phosphatase in patients affected by prostate and by biliary tract carcinomas, there has been raising interested in the release of tumour related molecules in the blood stream. In many instances, particularly in the earliest time of investigation, these were enzymes but later on also carbohydrates, mucins, onco-fetal antigens, hormones have been taken into account (Chan et al, 2006). What is now emerging is that in most cases these substances, that are improperly designed as tumours biomarkers, should not be considered as indicative of the type of pathology but rather – at best – as indicative of the affected tissue. The only *cancer biomarkers* to which a relatively confidence can be warranted particularly for population screening is PSA, the prostate specific antigen, a form of kallikrein that is commonly present in prostatic secretion and increases in the blood stream in the case of metastatic prostate adenocarcinoma, and possibly α-fetoprotein for screening of hepatocellular carcinoma in regions of high endemic frequency. In other words cancer biomarkers are not reliable for cancer diagnosis

and their use should best be related to screening and more properly in assessing effectiveness of initial antineoplastic therapy of whatever kind or in monitoring for recurrencies. An additional perspective for which determination of specific cancer biomarkers has found application is the estimation of sensitivity of tissues to therapeutic interventions particularly in the case of hormone dependent cancers, like prostate, breast and endometrial cancer. A notable example is the Triple Null phenotype in breast cancer which is an aggressive and therapy resistant form, characterized by the lack of expression of E_2R, PgR and HER-2.

These limitations further stimulate interest in genetic markers as indicative of the occurrence of specific mutations in the tumours, that are related to the pathogenesis of the lesions, such as B-Raf mutations (the amino acid substitution V600E) in papillary thyroid cancer (PTC) (Xing 2012) or the expression of specific proteins isoforms as cytokeratins and SCCA (Squamous Cell Carcinoma Antigen) in HNSCC, the common cancer of the oral cavity, pharynx and larynx (Chan et al, 2006). These markers are commonly determined on biopsy or surgical specimens by immuno-histochemical techniques.

Among the techniques that might be applied to the search of novel potential circulatory biomarkers (plasma is clearly the preferred biologic source to search because it can easily be obtained with minimal traumatism) interest is now increasing in the secretomic approaches (Stastna and Van Eyk, 2012) or in direct comparison between protein expression patterns in normal and pathologic tissues obtained from single patients by classic proteomic methodologies.

We have investigated in this topics employing the last approach of direct tissue comparison in a population of patients affected by infiltrating ductal adenocarcimas of the breast. The text that has been accepted for publication on Mol. Med. Reports is reproduced below.

Mol. Med. Reports (Accepted for publication)

Comparative proteomic analysis of ductal breast carcinomas evidenced altered expression of chaperonins and cytoskeletal proteins

Carcoforo P[1], Ura B[2], Squerzanti M[2], Lanzara V[2], Mischiati C[2]*, Cervellati C[2], Calza R[2], Polverino De Laureto P[3], Frare E[3], Portinari M[1], Feriotto G[1], Lanzara S[1], Agostinelli E [4], Bergamini CM[2]

1 Department of Morphology, Surgery and Experimental Medicine, University of Ferrara , Italy

2 Dept. Biomedical Sciences and Specialized Surgeries, University of Ferrara, Italy

3 CRIBI, University of Padova, Italy

4 Istituto Pasteur-Fondazione Cenci Bolognetti, Department of Biochemical Sciences, SAPIENZA University of Rome and CNR, Institute Biology and Molecular Pathology, Rome, Italy

**corresponding author address: Dept. Biomedical Sciences and Specialized Surgery, Section of Biochemistry, Clinical Biochemistry and Molecular Biology, via Luigi Borsari 46, University of Ferrara, 44100 Ferrara – Italy, e-mail:msc@unife.it , Tel: +39-532-454444*

keywords: proteomics, breast cancer, cancer biomarkers

Abstract

We have analyzed by comparative 2D proteomics the protein composition of ductal breast cancer and surrounding normal tissue in single patients to detect candidate disease biomarkers for diagnosis and prognosis. Chaperonins and cytoskeletal proteins predominate among the 11 proteins for which we detected major changes in abundance. In addition a few cytokeratins are also altered but they however cannot serve as specific circulatory biomarkers. The likely physiopathologic roles of these proteins are discussed.

Introduction

With incidence of about one million new cases/year breast cancer (BC) is a major cause of death in women diagnosed at the stage of infiltrating disease [1]. BCs are classically classified as lobular/ductal forms, with scirrhous, medullary and mucinous variants. Their biologic-clinical heterogeneity and variable response to therapy lead to refined classifications based on receptor status [2] as luminal A and B type (ER positive); HER2 type over-expressing EGF receptors; basal type (not expressing ER, PR and HER, also known as triple negative, TNBC) and normal-like cancer. Attempts to additional typing are based on markers for diagnosis and prognosis (cytokeratines and chaperonins), in relation to specific mutations detected by proteomic and cDNA microarrays approaches [3]. Interesting investigations are in progress.

Here we present proteomic analysis of normal and BC tissue from individual patients undergoing mastectomy at the Ferrara University Hospital, during the last 2 years. Changes in expression of selective proteins in most patients supports the search for their roles as tissue and serological markers for identification of aggressive tumors and as targets for therapy refractory cases.

Materials and Methods

Tissue Specimens

Samples of normal and cancer tissue were collected from 28 patients (indicated as P1 to P28) with ductal BC for proteomic investigation approved by the institutional ethical committee. Diagnosis was confirmed by histopathology, which proved that tumor specimens contained more than 50% tumor cells. Samples from 10 patients providing large amounts of tissues were snap-frozen in liquid N_2 and stored at -80°C until proteomic analysis for the present investigation.

Proteomics by 2D electrophoresis and mass spectrometry

Proteomics on BC and normal tissue from the same patient was performed by homogenization in 2.5 volumes of 7 M urea, 2 M thiourea, 4% CHAPS, 40 mM Tris, 1mM benzamidine, 1mM iodacetamide, 1mM EDTA, pH 8.8. After centrifugation, portions of supernatant corresponding to 260 μg protein, were submitted to IEF on precast pH 3-10 linear IPG strips at 40000 volts.h, according to manufacturer's instructions. After thiol reduction by 1% DTT and alkylation by 4% iodoacetamide, strips were submitted to 2nd dimension SDS-PAGE on 12,5%c polyacrylamide gels for 1 hour at 200V. Gels were stained by colloidal Coomassie, scanned with Molecular Imager Pharos FX Systems and analyzed by the ProteomeWeaver 4 program (Biorad). Spots excised from

the gel were processed by trypsin digestion for MS peptide identification. Briefly, gel fragments were rinsed in buffered acetonitrile, and dried. After additional thiol group reduction and alkylation, peptides were digested overnight with 12.5 ng/µl trypsin, resuspended in aqueous formic acid and analyzed with Ultimate 3000 Nano/micro HPLC apparatus Coupled with a LTQ-Orbitrap XL mass spectrometer [4,5]. By means of Excalibur 2.0.7 software spectra were submitted for peptides identification to the MASCOT program against the NCBI database.

Results

*In **Table 1** we summarize clinical characteristics of the patients along with the list of proteins with altered expression recognized by 2-DE proteomics. We achieved optimal resolution with total load of 260 µg of proteins on the IEF strip, running second dimension electrophoresis at constant voltage of 200 V for 1 hour, to avoid proteins smaller than 20 kDa to run out of the gel. Typically 2D electrophoresis of tumor and normal samples in single patients **(Figure 1)** resolved several hundreds proteins. By visual inspection 11 proteins had altered expression in BC tissue (circles, Figure 1), 2 proteins with lower and 9 higher levels of expression. These proteins were identified by MS spectroscopy. The nine proteins with increased expression (**Table 1)** are endoplasmin precursor (gp96), protein disulfide isomerase (PDI), TCP-1 subunit theta (TCP1-ϑ), F actin-capping protein subunit ϐ (FACP-ϐ), heat shock protein beta 1 (namely, HSP27), triosephosphate isomerase (TPI), RS/DJ-1, beta-tubulin (Tub-ϐ) and beta-actin (Act-ϐ) while carbonic anhydrase 1 (CA-1) and A-FABP (Adipocyte fatty acid binding-protein) are down-regulated. These proteins thus include molecular chaperones, cytoskeletal proteins and metabolic enzymes. We have identified also several cytokeratin peptides but we ignored them because of possible contaminations during sample processing. Actually cytokeratins are relevant in BC immunocytochemical approaches but not when searching for circulating biomarker discovery [6].*

Estimation of abundance of proteins with altered expression was achieved through normalization of intensity of spots from different gels against a virtual spot calculated as average on the same gel of 5-6 spots with identical expression in all analyzed samples, obtaining ratios of protein present in spots of BC and corresponding normal tissue in 2-DE gels (Table1). Ratios larger than 2 or smaller than 0.5 denote proteins significantly more/less expressed in BC versus normal tissue. Thus gp96 and TPI proteins were consistently modulated in at least 7 out of 10 patients, even if the score for TPI should be higher because its expression was not constantly detectable in normal tissue making impossible to calculate a precise ratio. PDI, FACP-ϐ, Tub-ϐ were highly expressed in 8 out of 10 tumors. HSP27, RS/DJ-1, CA-1 and A-FABP were consistently modulated in 9 out of 10 patients. Act-ϐ and TCP1-ϑ were always elevated in all tumors we analyzed.

Discussion

We resolved numerous proteins in normal and cancerous breast, 11 displaying consistently altered expression in tumors. The proteins with increased expression are

i) chaperonins, which are notably involved in increased tumorigenicity, metastatic potential and resistance to chemotherapy. They include HSP27 induced under unfavorable conditions to protect cells from death, preventing aggregation of denatured proteins, regulating caspase activity, intracellular redox state, polymerization of actin and cytoskeletal dynamics [7]; HSP27 is a proposed immunocytochemical discriminator to refine C3 and C4 categories in suspect BC aspirates [8]; and gp96 of the HSP 90 family [9] which represents a tool for active immunization against tumors by associating with cell surface peptides for presentation to cytolytic T lymphocytes and cell destruction. Notably Vitespen, a peptide vaccine based on gp96, prolongs survival in patients with early stage melanoma or renal cancer [10]. The presence of gp96 in infiltrating ductal BC is attractive for vaccine treatment in TNBC patients resistant to classic therapy while maintaining expression of gp 96 (see patient P20).

ii) the cytoskeletal proteins Act-β, FACP-β, Tub-β and TCP1-ϑ. The last protein assists ATP-dependent folding of actin and tubulin [11]. Interestingly some cytoskeletal proteins upregulated in BC are involved in estrogen receptor activation: Act-β binds to the ER α complex, contributing to ER nuclear functions [12]; chaperonin TCP1-ϑ (spot 3) is involved in folding Act-β [13] and estrogen-regulated HSP27 controls palmitoylation of ER, required for its interaction with membranes [14]. This indicates that cytoskeletal rearrangement as a key step in the motility mechanisms for metastatic spreading, but might also be involved in hormone receptivity [15].

iii) signaling proteins with altered expression are RS/DJ-1, TPI and A-FABP. RS/DJ-1 is an oncogene protein regulating RNA-protein interaction, present in sera from BC patients but not healthy subjects along with circulating antibodies against this protein [16]. Similarly, autoantibodies against TPI are present in BC patients [17] but also in patients with oral cavity and lung squamous cell carcinoma [18,19]. PDI catalyzes formation/ rearrangement of protein disulfide bonds, acting as a reductase at cell surface cleaving disulfide bonds with structural modifications of cell associated proteins and as chaperonin (inhibiting aggregation of misfolded proteins) or antichaperonin (facilitating aggregation) inside cells, depending on concentration. In addition PDI binds estrogen and thyroid hormones [20]. Interestingly, knockdown of PDI in BC MCF7 cells induces caspases-dependent apoptosis [21]. A-FABP is another protein whose expression is affected in a declining mode. It plays roles in intracellular lipid transport and metabolism, and signalling. Prognostic value for A-FABP was reported in bladder cancer since a decreased expression

correlated with poor prognosis [22]. It is induced by PPAR ligands (the promoter region of the A-FABP gene contains functional peroxisome-proliferator-responsive elements [23]) and its overexpression is likely beneficial in treatment of bladder cancer. Contrasting results are reported in BC since serum levels of A-FABP associate with tumor risk and aggressive behavior [24], but comparable values of expression were reported in ductal infiltrating carcinoma and normal tissue [25]. Our data document instead decreased levels of A-FABP in ductal BC than normal tissues in 9/10 patients, including the TNBC patient in our population. The association between BC and A-FABP and the possible selective induction by PPAR ligands might impact in strategies for therapy of TNBC if the data will be confirmed in larger populations.

In conclusion, despite the limited number of subjects investigated in the present study, the robustness of our findings suggests they might be also observed in future larger studies that include others kind of tumors and their multidrug resistant counterparts. This is an aspect of particular interest, since one of the problems of conventional anti-cancer therapy is the development of drug resistance. The proteins for which we documented an altered expression in the infiltrating ductal BC could be exploited as novel targets for therapeutic interventions or represent novel diagnostic/prognostic markers for early detection of aggressive tumors, particularly those with MDR phenotypes, in early stages of disease.

Acknowledgements

This research was supported by grants from Fondazione Cassa Risparmio di Ferrara to prof. Paolo Carcoforo

References

[1] Pisani P, Bray F, Parkin DM. Estimates of the world-wide prevalence of cancer for 25 sites in the adult population. Int J Cancer 2002; 97: 72–81.

[2] Sørlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001; 98: 10869–10874.

[3] Gast MC, Schellens JH, Beijnen JH. Clinical proteomics in breast cancer: a review. Breast Cancer Res Treat 2009; 116: 17-29

[4] Bozzo C, Spolaore B, Toniolo L, Stevens L, Bastide B. Nerve influence on myosin light chain phosphorylation in slow and fast skeletal muscles. FEBS J 2005; 272: 5771-5785.

[5] Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 2006; 1: 2856-2860.

[6] Galvao ER, Martins LM, Ibiapina JO, Andrade HM, Monte SJ. Breast cancer proteomics: a review for clinicians. J Cancer Res Clin Oncol. 2011; 137: 915-925.

[7] Acunzo J, Katsogiannou M, Rocchi P. Small heat shock proteins HSP27 (HspB1), αB-crystallin (HspB5) and HSP22 (HspB8) as regulators of cell death. Int J Biochem Cell Biol 2012, E-pub Apr 13.

[8] Keeling J, McKee GT. Heat shock protein (HSP)27: a further refinement in the diagnosis of suspicious fine needle aspirates of breast. Cytopathology 1999; 10, 40–49.

[9] Yang Y, Li Z. Roles of heat shock protein gp96 in the ER quality control: redundant or unique function? Mol Cells. 2005; 20: 173–182.

[10] Randazzo M, Terness P, Opelz G, Kleist C. Active-specific immunotherapy of human cancers with the heat shock protein Gp96-revisited. Int J Cancer 2012; 130: 2219-2231.

[11] Brackley KI, Grantham J. Activities of the chaperonin containing TCP-1 (CCT): implications for cell cycle progression and cytoskeletal organisation. Cell Stress Chaperones 2009; 14: 23-31.

[12] Ambrosino C, Tarallo R, Bamundo A et al. Identification of a hormone-regulated dynamic nuclear actin network associated with estrogen receptor alpha in human breast cancer cell nuclei. Mol Cell Proteomics 2010; 9: 1352-1367.

[13] Vainberg IE, Lewis SA, Rommelaere H et al. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell 1998; 93: 863-873.

[14] Razandi M, Pedram A, Levin ER. Heat shock protein 27 is required for sex steroid receptor trafficking to and functioning at the plasma membrane. Mol Cell Biol 2010; 30: 3249-3261.

[15] Jiang P, Enomoto A, Takahashi M. Cell biology of the movement of breast cancer cells: intracellular signalling and the actin cytoskeleton. Cancer Lett 2009; 284: 122-130.

[16] Le Naour F, Misek D, Krause M, Deneux L, Giordano T, Scholl S, Hanash S. Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. Clin Cancer Res 2001; 7: 3328-3335.

[17] Tamesa MS, Kuramitsu Y, Fujimoto M et al. Detection of autoantibodies against cyclophilin A and triosephosphate isomerase in sera from breast cancer patients by proteomic analysis. Electrophoresis 2009; 30: 2168-2181.

[18] Shukla S, Pranay A, D'Cruz AK, Chaturvedi P, Kane SV, Zingde SM. Immunoproteomics reveals that cancer of the tongue and the gingivobuccal complex exhibit differential autoantibody response. Cancer Biomark 2009; 5: 127-135.

[19] Zhang XZ, Xiao ZF, Li C et al. Triosephosphate isomerase and peroxiredoxin 6, two novel serum markers for human lung squamous cell carcinoma. Cancer Sci 2009; 100: 2396-401.

[20] Fu XM, Wang P, Zhu BT. Characterization of the estradiol-binding site structure of human protein disulfide isomerase (PDI). PLoS One 2011; 6 :e27185. doi:10.1371/journal.pone.0027185

[21] Hashida T, Kotake Y, Ohta S. Protein disulfide isomerase knockdown-induced cell death is cell-line-dependent and involves apoptosis in MCF7 cells. J Toxicol Sci 2011; 1: 1-7.

[22] Boiteux G, Lascombe I, Roche E, Plissonnier ML, Clairotte A, Bittard H, Fauconnet S. A-FABP, a candidate progression marker of human transitional cell carcinoma of the bladder, is differentially regulated by PPAR in urothelial cancer cells. Int J Cancer 2009; 124: 1820-1828.

[23] Schachtrup C, Emmler T, Bleck B, Sandqvist A, Spener F. Functional analysis of peroxisome-proliferator-responsive element motifs in genes of fatty acid-binding proteins. Biochem J 2004; 382: 239-245.

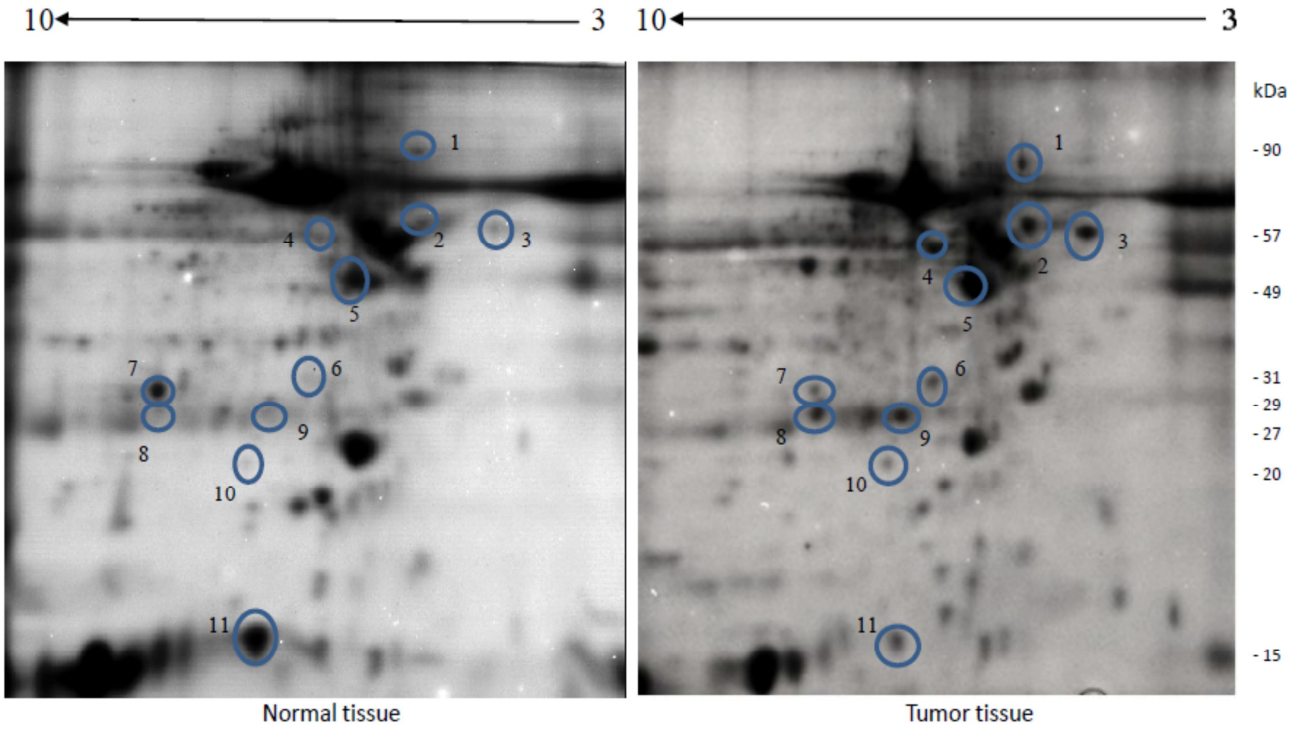
[24] Hancke K, Grubeck D, Hauser N, Kreienberg R, Weiss JM. Adipocyte fatty acid-binding protein as a novel prognostic factor in obese breast cancer patients. Breast Cancer Res Treat 2010; 119: 367-367.

[25] Li H, Lu Q, Dong LH, Xue H, Zhou HY, Yang HJ. Expression of fatty acid binding protein in human breast cancer tissues. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi 2007; 23: 312-316.

[26] Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991; 19: 403-410.

Figure and Table captions

Figure 1. Two-dimensional gel electrophoresis of mammary tissue proteins. Total proteins from homogenates of tumor or normal breast tissues from the same patient were loaded onto each gel and separated by two-dimensional gel electrophoresis. The spots circled in the image, representing the relevant up or down-regulated proteins, were excised and analyzed by LC-MS/MS. The identified proteins are listed in Table 1.



Case	P10	P13	P18	P19	P20	P21	P22	P23	P27	P28
Clinical Setting										
Citology	pT1c-pN0(i)(sn)	pT2-pN0(i-)(sn)	pT1c-pN0(i+)(sn)	pT1c-pN0(i-)(sn)	pT1mic-pN0(i-)(sn)	pT2-pN1a	pT1c-pN0(i+)(sn)	pT1c-pN0(i-)(sn)	pT1c-pN0(i-)(sn)	pT2-pN3a
NCI	n.d.	4.5	3.3	3.32	n.d.	5.56	3.28	3.3	3.24	6.44
Grade	G2	G3	G2	G2	micro	G3	G2	G2	G2	G3
ER	90	0	96	97	0	21	99	72	99	52
PR	50	0	36	94	0	8	98	64	56	68
HER2	1+	1+	0	0	0	0	0	0	0	3+
Identified proteins										
GP96	2,43	2,00	4,86	8,13	5,00	4,67	1,14	5,00	1,27	1,18
PDI	8,25	3.38	6.00	4.33	n.d.	9.33	2	4.76	n.d.	4.55
TCP1-θ	2.22	2.00	2.11	4.33	2.89	7.19	4.80	12.68	3.86	5.38
FACP-β	3.50	1.44	2.75	2.00	2.00	2.92	1.30	3.33	2.89	2.00
HSP27	2,43	1,29	5,33	6,77	2,57	2,00	2,29	4,67	3,60	2,73
TPI	3,50	n.d.	5,75	2,41	1,38	3,50	n.d.	3,33	6,00	3,93
RS/DJ 1	2,25	0,91	4,00	2,53	3,58	2,33	2,80	4,80	6,00	2,42
TUB-β	2,53	1,40	6,50	2,00	2,02	3,50	0,80	4,98	4,32	2,18
ACT-β	6,25	3,60	8,00	2,28	2,75	3,89	2,00	13,33	3,52	3,64
CA-1	0,50	0,28	1,11	0,50	0,50	0,18	0,50	0,19	0,50	0,45
A-FABP	0,23	0,18	0,50	0,11	0,25	0,13	0,36	0,89	0,15	0,18

Table1.

Table1. Population characteristics and identification of expression levels of proteins significantly altered in BC. Ratios of expression of individual proteins in normal and BC tissue is defined in the text. Underlined values indicate patients where expression deviates from the general trend observed in our cases

PROTEOMICS AND ADVANCES IN CANCER BIOLOGY. THE EMT CONCEPT

A main novel concept in relation to cancer biology is that of the occurrence of EMT. This is a relatively recent notion related to the capacity of cancer cells to reprogram their differentiation, in the sense of conversion of epithelial into mesenchymal cells. The former ones are characterized by close cell-cell adhesions that limit cell mobility through protein-protein interactions, while mesenchymal like cells are devoid of cell adhesive molecules and rather expressing at high levels proteinases which favour their mobility. During the transition there a overlapping progressive processes of loss of biomarkers of the epithelial cell and gradual acquisition of biomarkers of a mesenchymal phenotype. Among the stimuli that favour the EMT, cytokines play major roles, (notably TGFβ) with possible involvement of bone marrow derived cells, but also mechanical stimuli might contribute as it happens in the EMT in the cholangiocarcinomas cell lines that we have analyzed. I will come again to these concepts below along with the analysis of proteins differentially expressed in tumours undergoing EMT.

EMT has been recognized to play a role in several physiologically and pathologically relevant processes as outlined by Thiery et al, (2009). Kalluri and Weinberg (2009) have defined types of EMT. Type 1 plays a role in the prenatal development in which cells from the embryonic surface leaflets move deeply into the mesenchymal region to migrate and contribute to organ development. This process is not associated with inflammatory stimuli, which are instead present in type 2 EMT, which is associated with would healing with epithelial cells that converse into the mesenchymal ones for space filling during would healing by second intention repair. At the last extreme is type 3 EMT, that occurring in cancer tissues, in which dedifferentiated mesenchymal cells find easier access to blood vessels to spread the tumour by blood borne metastases, rather than by the lymphatic route. Besides its dedifferentiation and the acquisition of markers of cell mobility and of ECM degradation, EMT in cancer tissues is characterized by appearance of distinct phenomena in vitro in cultured cells, notably by formation of large globular cells aggregates commonly known as spheroids.

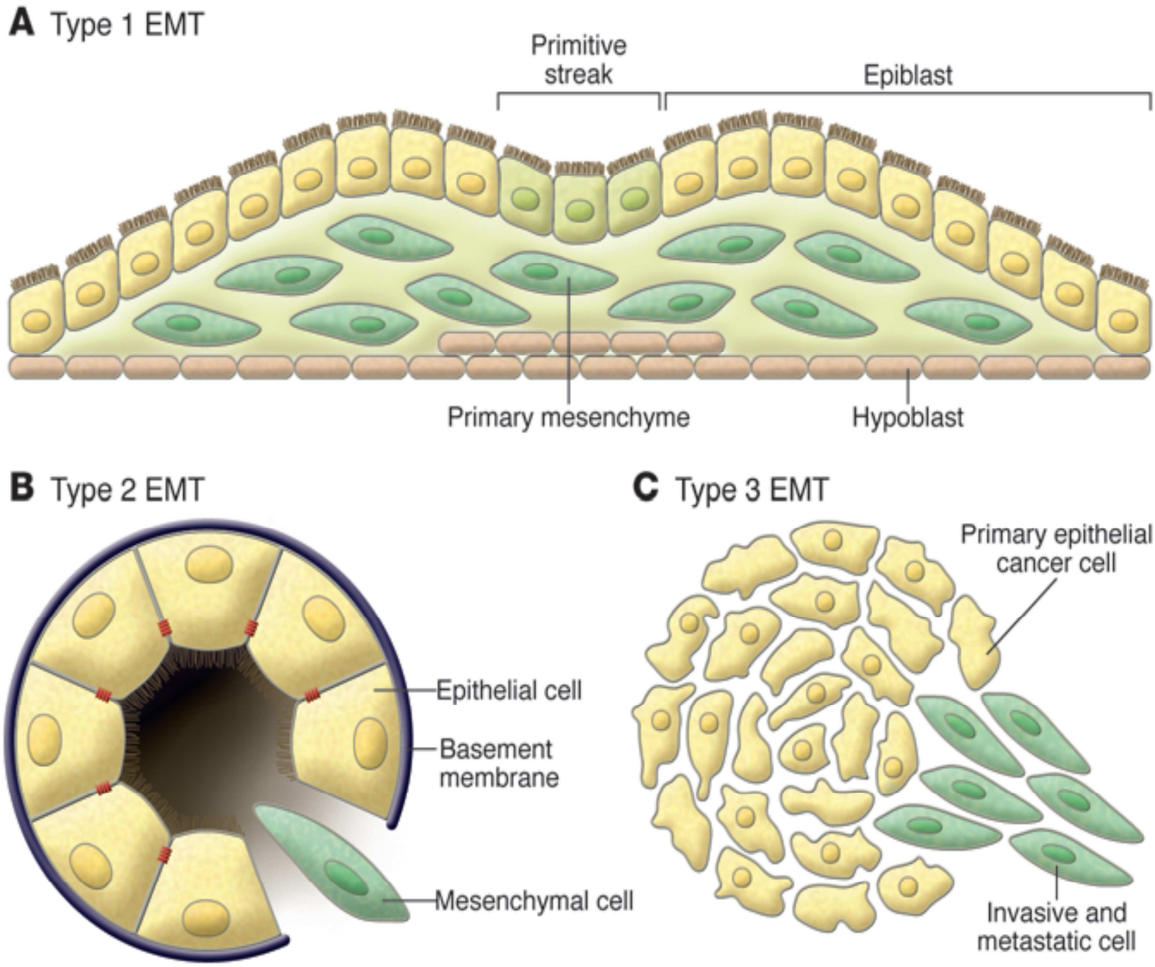


Fig. 6.1 Schematic view of the processes of type 1, type 2 and type 3 EMT.

Type 3 EMT is likely to contribute considerably to the heterogeneous appearance of poorly differentiated cancer, a feature which has been ascribed to either occurrence of EMT or also to an intrinsic property of tumours themselves considered as aggregates of differentiated and of poorly differentiated cancer cells that have a definitely higher ability to invade and to metastatize. These cells have therefore been described alternatively as products of EMT or as cancer stems cells to indicate their high intrinsic capacity to reproduce the tumour at distant sites. In this perspective differences in regulation and protein expression in lesions of primary tumours and in metastatic foci are easily understandable. At least in some tumours as for instance ovarian tumours spreading within the peritoneal cavity as ascites tumours (Shao et al, 2009) both the processes of acquisition of resistance to classic chemotherapy (or even to radiotherapy) and the secondary localization as coelomatic metastases are events associated with EMT (Zhang et al, 2008) as well as with the enhanced catalytic activity of type 2 Tgase.

For a few types of cancers, investigations of differential protein expression have already been published in lesions exhibiting epithelial phenotype or evidence of EMT, by either immuno-cytochemistry or proteomic profiling, employing several experimental systems, as exposition to TGFβ or comparison between primary tumours and metastatic lymph nodes which should be enriched in cancer cells that have undergone EMT, as noted above. This information is available for instance for HNSCC, prostate, pancreatic and lung cancer and expression of several proteins is frequently affected in a major way. Significant differences have been found in expression of cytokeratins, cadherins (particularly E-cadherin and fibronectin), MMPs, and focal adhesion kinases (Mathias and Simpson, 2009).

At our knowledge this kinds of investigations have never been carried out in the adenocarcinomas of the bile tract, tumours that are characterized by a high level of invasiveness. For this reason we decided to investigate this clinical entity, employing as models the cell lines SK-ChA-1 and MZ-ChA-1 to verify whether alterations in the protein expression pattern resemble those observed in other types of tumours, also because these cholangio-carcinomas cell lines are able to undergo EMT induction by simple mechanical stress as culturing under appropriate stirring conditions (Roncoroni et al, 2008), assuming a appearance of "spheroid" multicellular aggregates. These simple stimuli have not been reported as inducers of type 3 EMT in other cancer models.

EXPERIMENTAL STUDIES ON THE SPHEROID PHENOTYPE OF CHOLANGIOCARCINOMA CELLS

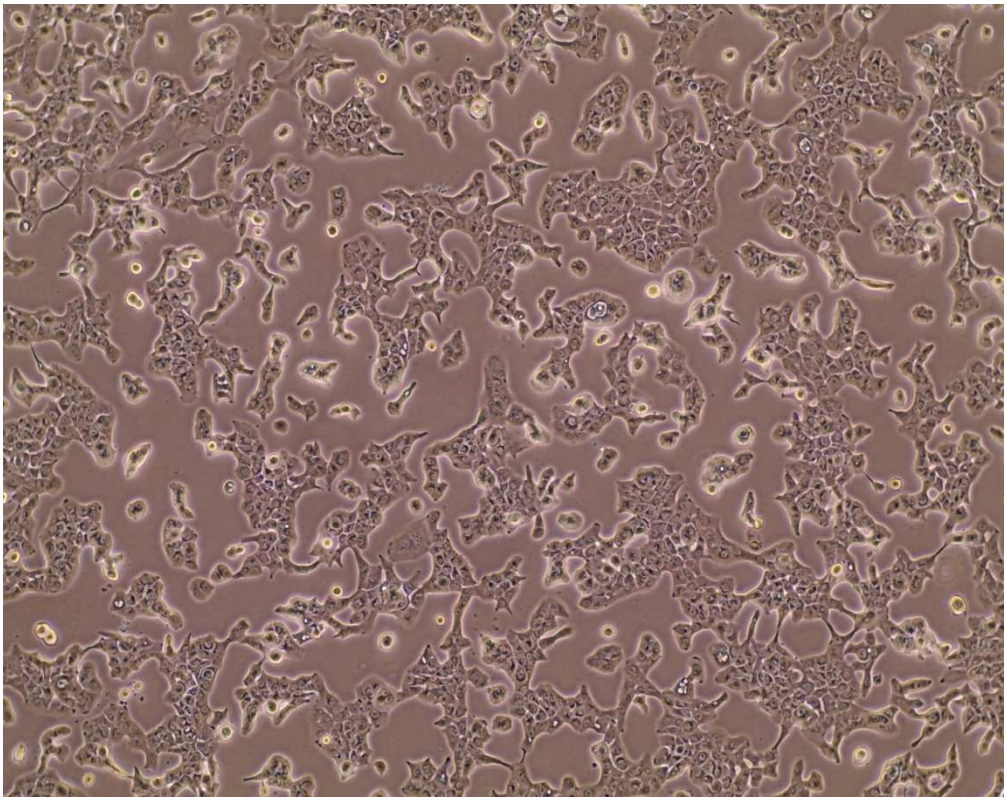
Introduction

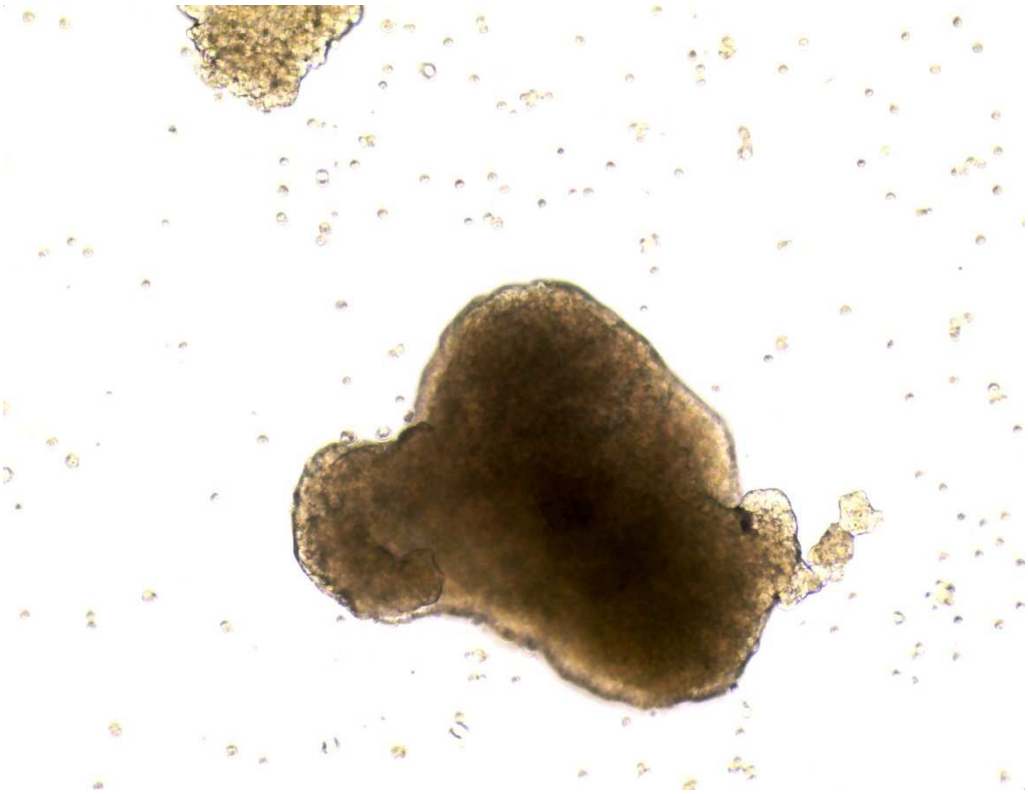
Cholangiocarcinomas are tumours originating from the epithelial cuboidal or columnar cells lining the surface of the bile ducts and account for 3% of all the gastrointestinal tumours. They can arise in the gallbladder, the hepatic and the common duct, the intraduodenal portion of the common duct or even the intrahepatic bile ducts roots. Cholangiocarcinoma is diagnosed with difficulty since the symptoms are rather aspecific (abdominal pain, jaundice, digestive disturbances, itching, laboratory signs of cholestasis) and arise late, making surgery suitable only in a limited number of patients, despite the critical location usually permits detection of the tumours when they are still low in size. Since both radiotherapy and traditional chemotherapy (5-fluorouracil and gemcitabine alone or in combination with cisplatin) showed disappointing results in terms of survival in unoperable patients, the development of new therapies is badly needed (Khan et al, 2005).

These considerations led us to explore the possibility of the occurrence of EMT in two models of cholangiocarcinoma, the cell lines SK- Cha-1 and MZ-Cha-1 that was originally supplied by prof. Alexander Knuth (Ludwig Institute, Frankfurt, Germany) and cultured for us by colleagues at the Gastroenterology and Celiac Disease Research Unit (Drs. Roncoroni, Elli and Bardella) of the Ospedale Maggiore Medical Polyclinic, University of Milan.

These cells were cultured under two different experimental settings to obtain either Two-Dimensional (2D) or Three Dimensional (3D) cultures (which are also called multicellular tumour spheroids). In the 2D setting cultures of SK-ChA-1 cells were grown in Dulbecco's modified Eagle's medium supplemented with 1% penicillin 100 U/ml, streptomycin 100 mg/ml, 2% L-glutamine 200mM and 10% foetal bovine serum in a humidified 5% CO_2 incubator at 37 °C, in six-well tissue culture plates. In the Three-Dimensional culture approach (that to obtain multicellular tumour spheroids or MCTSs) growth was started by seeding 2.10^5 cells/ml in 15 ml of complete Iscove modified Dulbecco's medium (supplemented with antibiotics and glutamine as above) in polycarbonate Erlenmeyer flasks and incubated in a gyratory rotation incubator (60 rev/min) at 37 °C in an air atmosphere. Homotypical aggregations became visible after 4 days of culture and MCTSs were usually complete within 7 days (mean diameter_standard deviation, 270 ± 53.5 μm).

Cells obtained by 2D and 3D were collected by centrifugation and lysed by treatment with TUC 2-D lysis buffer (7 M Urea, 2 M thiourea, 4% CHAPS, 40mM TRIS) and a cocktail of protease inhibitors. For 1 million cells we used 130μl of lysis buffer. After protein determination by the method of Bradford 300 μg of protein were loaded onto a 3-10 NL IPG strip 7 cm (Biorad), that was focused for 12 hours until reach the maximum voltage of 4000V, to obtain a total 60000 volt h. After thiol reduction by 1% DTT and alkylation by 4% iodoacetamide. The second dimension was run on a slab polyacrylamide gel (12,5%) at 200 volts, until the bromophenol reached the bottom of the gel to avoid loss of small MW peptides. In the case of the 3D spheroids lysis was performed with 150 μl of lysis buffer and the protein load was assayed only by the protein determination assay. After fixing the gels washed 2 times with bidistilated water and stained with colloidal Coomassie, scanned with Molecular Imager Pharos FX Systems and analyzed by the ProteomeWeaver 4 program (Biorad).

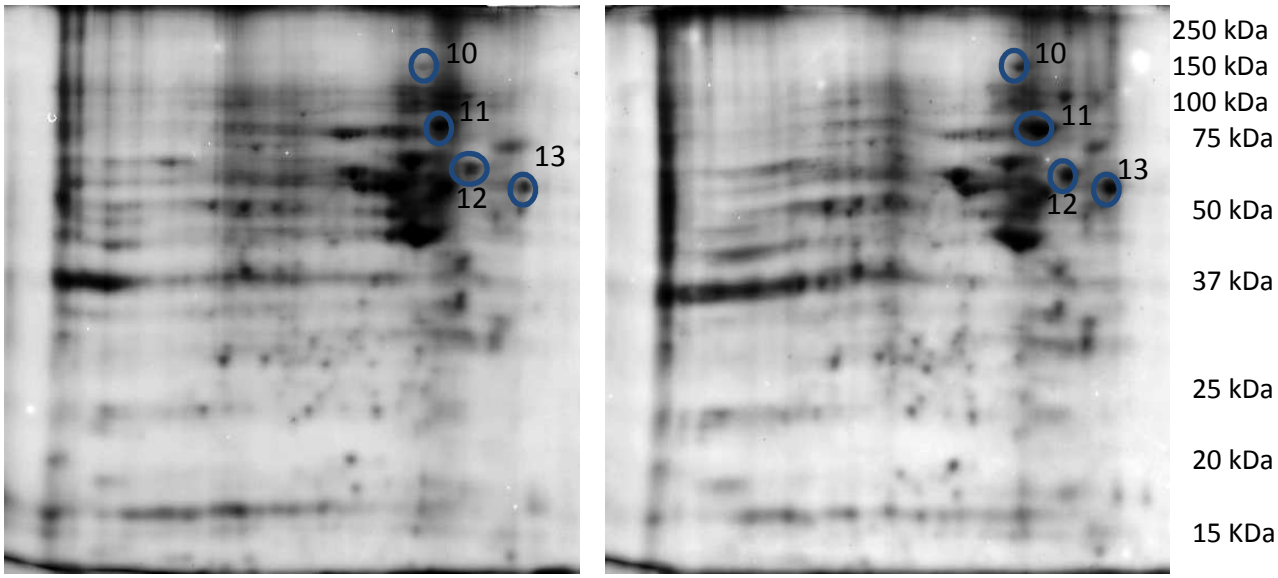




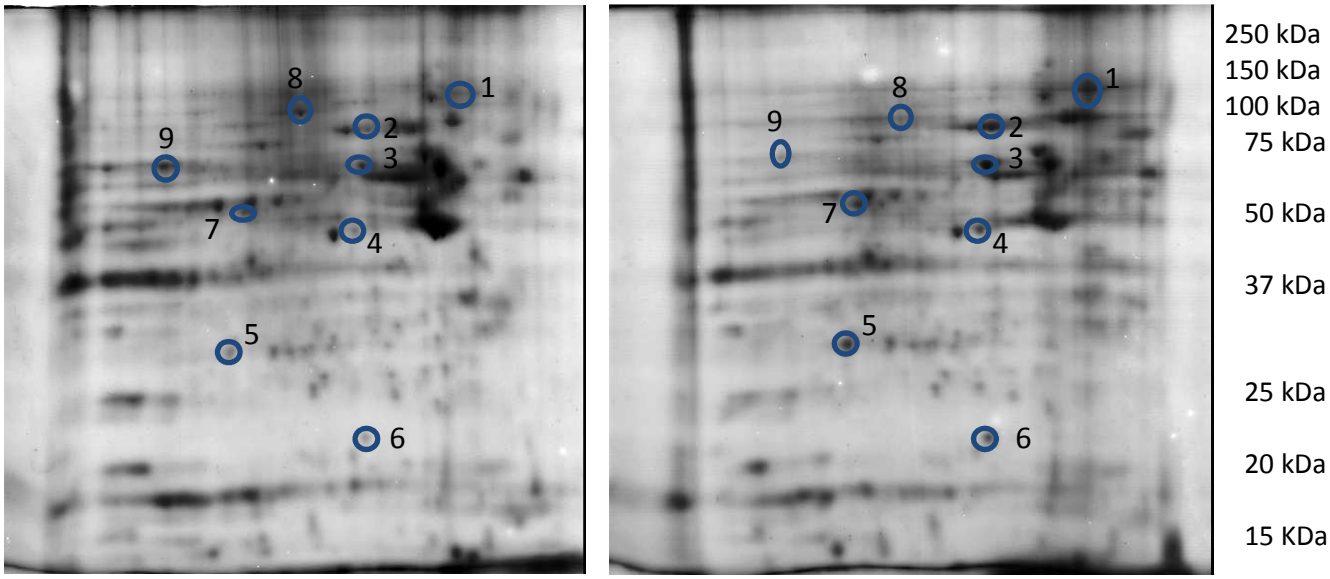
The above picture are those of the Cells of SK-ChA1 cell line growing as adherent pre-confluent cells in resting flasks (upper picture) and of cells of the same cell line growing as spheroids under continuous stirring (Courtesy of Dr. Leda Roncoroni, University of Milan, see the text for details).

Cells at these stages of growth have been employed in our experiments of bidimensional electrophoresis to evidence eventual alterations in the protein expression patterns.

Relevant gels reported below display protein expression patterns in the confluent cultures (left panel) and in the spheroid growing phenotype (right panel) of the MZ cell line (this page)



and the SK cell line



Although minor differences apparently exist in the protein expression levels between the two MZ and SK cell lines are evident (compare for instance the left panels), in both instances significant differences in the expression pattern appear as the cultures are grown under conditions of confluence or multicellular spheroid phenotype. The regions that appear to be most affected are those at pH 5 and at pH 7 to 8 (Note that the acid and the basic pH are on the left and the right side in each gel).

A few proteins that are differently expressed have been identified and submitted to proteome analysis to identify them on the basis of their sequence. Mass spectroscopy identification is in progress.

References

Almeida C, Nagarajan D, Tian J, Leal SW, Wheeler K, Munley M, Blackstock W, Zhao W. (2013)
The role of alveolar epithelium in radiation-induced lung injury. PLoS One. 8: e53628.

Blanco-Colio LM, López JA, Martínez-Pinna Albar R, Egido J, Martín-Ventura JL.(2009) Vascular proteomics, a translational approach: from traditional to novel proteomic techniques. Expert Rev Proteomics. 6: 461-464

Bradford MM. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 72: 248-254

Chan DW, Booth Ra, Diamandis EP (2006) Tumor markers, in Tietz textbook of Clinical Chemistry and Molecular Diagnostics (Burtis CA et al, eds), pp 745-795, Elsevier Saunders publisher, St. Louis Mo.

Cowan ML, Vera J (2008) Proteomics: advances in biomarker discovery. Expert Rev Proteomics. 5: 21-23

Craven RA, Vasudev NS, Banks RE. (2012) Proteomics and the search for biomarkers for renal cancer. Clin Biochem. doi:pii: S0009-9120(12)00665-0 (in press)

Gao D, Vahdat LT, Wong S, Chang JC, Mittal V. (2012) Microenvironmental regulation of epithelial-mesenchymal transitions in cancer. Cancer Res. 72 : 4883-4889

Khan SA, Thomas CH, Davidson BR, Taylor-Robinson SD. (2005) Cholangiocarcinoma. Lancet 366: 1303–1314.

Kalluri R, Weinberg RA. (2009) The basics of epithelial-mesenchymal transition. J Clin Invest. 119: 1420-1428.

Kuang P, Zhou C, Li X, Ren S, Li B, Wang Y, Li J, Tang L, Zhang J, Zhao Y. (2012) Proteomics-based identification of secreted protein dihydrodiol dehydrogenase 2 as a

potential biomarker for predicting cisplatin efficacy in advanced NSCLC patients. Lung Cancer. 77: 427-432

Kullo IJ, Cooper LT. (2010) Early identification of cardiovascular risk using genomics and proteomics. Nat Rev Cardiol. 7: 309-317

Lim YB, Pyun BJ, Lee HJ, Jeon SR, Jin YB, Lee YS. (2011) Proteomic identification of radiation response markers in mouse intestine and brain. Proteomics. 11: 1254-1263

Mathias RA, Simpson RJ. (2009) Towards understanding epithelial-mesenchymal transition: a proteomics perspective. Biochim Biophys Acta 1794: 1325-1331

Roncoroni L, Elli L, Dolfini E, Erba E, Dogliotti E, Terrani C, Doneda L, Grimoldi MG, Bardella MT. (2008) Resveratrol inhibits cell growth in a human cholangiocarcinoma cell line. Liver Int. 28: 1426-36.

Scaife L, Hodgkinson VC, Drew PJ, Lind MJ, Cawkwell L. (2011) Differential proteomics in the search for biomarkers of radiotherapy resistance. Expert Rev Proteomics. 8: 535-552

Shao M, Cao L, Shen C, Satpathy M, Chelladurai B, Bigsby RM, Nakshatri H, Matei D. (2009)
Epithelial-to-mesenchymal transition and ovarian tumor progression induced by tissue transglutaminase. Cancer Res. 69: 9192-201.

Stastna M, Van Eyk JE. (2012) Secreted proteins as a fundamental source for biomarker discovery.
Proteomics. 12: 722-735

Thiery JP, Acloque H, Huang RY, Nieto MA. (2009) Epithelial-mesenchymal transitions in development and disease. Cell. 139: 871-890

White MY, Van Eyk JE. (2007) Cardiovascular proteomics: past, present, and future. Mol Diagn Ther. 11: 83-95.

Xing M. (2012) BRAF V600E mutation and papillary thyroid cancer: chicken or egg? J Clin Endocrinol Metab. 97: 2295-2298.

Zeisberg M, Neilson EG. (2009) Biomarkers for epithelial-mesenchymal transitions. J Clin Invest. 119: 1429-1437.

Zhang S, Balch C, Chan MW, Lai HC, Matei D, Schilder JM, Yan PS, Huang TH, Nephew KP. (2008) Identification and characterization of ovarian cancer-initiating cells from primary human tumours. Cancer Res. 68: 4311-4320

Zheng G, Peng F, Ding R, Yu Y, Ouyang Y, Chen Z, Xiao Z, He Z. (2010) Identification of proteins responsible for the multiple drug resistance in 5-fluorouracil-induced breast cancer cell using proteomics analysis. J Cancer Res Clin Oncol. 136: 1477-1488

FINAL REMARKS

During the tenure of the PhD fellowship my main duty was to implement the use of proteomics in my guest laboratory, and to apply proteomic techniques to basic biological topics as well as to clinically relevant issues both in diagnostic and in investigative perspectives.

Thus I was not ascribed a single research time but rather the development of the technique and the optimization of its applications in relation to the scientific equipment available at the unit, which is suitable for proteomic investigations by the bi-dimensional electrophoresis approach to isolate specific proteins under investigation.

The final steps of the analysis consisted in the identification either of precise sites of labelling in irreversible protein PTM by Tgase 2 or of proteins involved in specific issues in oncology, i.e. in the altered protein expression in transformed cells that might be chosen as cancer biomarkers (in this case in breast ductal adenocarcinoma) and in cholangiocarcinoma cells undergoing EMT in vitro. This final analytic problems were performed for us by collaboration between our laboratory and the CRIBI Biotechnology Centre at the University of Padoa as well as the CEINGE Centre for Genetic Engineering at the University of Naples. I thank deeply these colleagues, who helped in my research. The outmost competence of these centres in mass spectroscopy is well recognized and was instrumental for our – still limited - success. However - as I explained - the analysis of the data is still in progress, mainly because of the busy activity in these host centres and therefore their time limitations. Data for additional publications will certainly be available as soon as these analyses will be completed.

Coming back to the experimental plans and to the applications of proteomics to the solution of the research queries that have been proposed for my activity, I would like to comments separately the issues concerning protein PTM by transglutaminases and the basic and clinical oncology topics. In relation to the **Protein Post-Translational Modification by Transglutaminases** (in this case Tgase 2), it is clear that these are largely irreversible reactions. At this purpose I remind that reversibility has been proved only in the case of proteins linked to primary amines, notably Dansylcadaverine, because of the back-reaction of Tgase itself, as proved by Lorand and associates (Parameswaran et al, 1997). PTM by Tgases are likely to affect greatly the properties of modified proteins because of changes in surface charge (conversion of the glutamine amide groups into the

carboxylic acid of glutamate or into a primary amine through linkage with polyamine) or in solubility because of massive aggregation stabilized by isopeptide bonds. The aggregated products are resistant to proteolysis because the isopeptide bond itself is not cleaved by proteases and because crosslinkage brings to an increased rigidity that is negative determinant of activity of proteases (Fontana). Historically the products of the transglutaminases reactions have been analyzed by extensive cleavage by a cocktail of proteases to degrade PTM proteins containing glutamyl-lysine or glutamyl-polyamines isopeptide bonds to free aminoacids and small peptides that are resolved by HPLC chromatography to isolate the crosslinking isopeptide (Beninati et al, 1988), which is later determined by reaction with fluorescamine or o-Phtalaldehyde. By this procedure it is possible to prove and to quantify formation of isopeptide crosslinks in PTM proteins but not to identify the sites in the protein sequence at which the crosslink occurs. For this purpose the proteomic approach seems ideal with the limitation that it is difficult to determine directly the sequences involved in glutamyl-lysine isopeptides owing to the resistance to proteolysis. This information is best deduced from separate determination of labelling of reactive glutamine and lysine residues through incorporation of suitable probes.

We followed this strategy in the published paper in the PTM of troponin T residues (Squerzanti et al, 2013) that is also reported in this Thesis, with the limitation that TnT is not a lysine substrate so that no protein crosslinkage occurs in this instance and determination of reactive lysine residues does not apply. In this study we have employed the MALDI-TOF MS/MS approach and we considered that this might be the general procedure for this kind of determination. This conclusion however was dismantled as we tried to apply the same procedure to the identification of the sites of PTM of tubulin. At this purpose it is worthy to note a clear discrepancy in the data we have obtained, for which we have not yet any definitive explanation. The matter is related to the eventual polymerization of the tubulin protein, which appears to be minimal by the analysis of the residual band of monomeric tubulin by SDS-PAGE after reaction with Tgase 2 both in the absence and in the presence of an amino substrate, which should obviously inhibit any reaction of protein crosslinkage. The point I want to discuss is related to the gel of fig. 5.5 which depicts the results of an experiment of PTM in the presence of dansylcadaverine. From the picture fluoresce is clearly present both in the separation and in the stacking gel. In the separation gel it is present with mild intensity in correspondence of the region of migration of the residual tubulin monomers while in the stacking gel it is present as a very

intense signal in the well, which however contains tiny amount of protein, which should be represented by aggregated insoluble protein which does not “enter” the gel. A priori this aggregated protein should arise from isopeptide crosslinked tubulin (i) aggregation should be very limited as already discussed and actually (ii) aggregation should also be inhibited by the presence of Dansylcadaverine which should compete with lysine for covalent binding to glutamine. At the same time it is important to realize that the very intense fluorescent signal produced by very tiny amount of protein actually displays a high “specific intensity” to employ an enzymologic concept, which is contrast with low labelling of monomeric tubulin subunits. In other words it is rather conceivable that an initial labelling of tubulin at a single or a few glutamine residues would affect protein solubility leading to large insolubility of the protein with conformational changes that allow a much more intense labelling by the fluorescence probe. Beside this effect, what is more important to note in relation to the concept of suitability of proteomics to detect sites of protein PTM by Tgase 2 is the failure we had to identify glutamine and lysine residues separately labelled in tubulin by Dansylcadaverine and by Cbz-Glngly, employing MALDI-TOF MS/MS as the analytical strategy. The more likely explanation of this failure, as explained before, is concerned with the primary structure of tubulin itself and with clustering of acidic residues in the C-terminal region producing tryptic peptides of low ionization capability. Attempts to solve this point by alternative procedures are still underway by employing other proteases of different sequence specificity with in-gel-digestion or alternatively to perform proteolysis in solution. By some authors (Sinz, 2006; Grebe and Singh, 2011) this last approach is considered superior to in-gel-digestion although it has drawbacks not to provide immediate imaging of the experimental results and to require a much more extensive access to mass spectrometry. On the bases of these results it is therefore important to underline the necessity of wide instrumental availability for extensive investigations of Tgase mediated protein PTM (Serpa et al, 2012).

The last experimental system of protein PTM on which I focused is related to CRM197. This protein is now employed also as a carrier for vaccine development for encapsulated Gram-positive and negative bacteria including *Haemophilus influenzae, Diplococcus pneumoniae* and *Neisseria meningiditis* (Malito et al, 2012), along with its use in the therapy of haematologic malignancies. In the case of the mutant DT CRM197, the PTM by Tgase 2 is characterized by extensive crosslinkage, which suggests the possibility to employ DT as a carrier for intracellular release of peptides for either vaccination of therapeutic purposes, after specific crosslinkage to CRM197. It is the aim of the laboratory

to explore further these effects in the perspective of the possible biotechnologic application of CRM197 in the vaccine and therapeutic field.

In contrast with these words of caution against the simple use of the MALDI-TOF MS/MS approach in the case of analysis of PTM by Tgase 2, this procedure proved reliable in our studies on **cancer**, for the analysis of the altered protein expression pattern both for the identification of candidate cancer cell markers and to investigate the basic feature of cancer cells. In this last aim we investigated the EMT process which determines their invasion characteristics of cancer cells, with relevance to the emergence of resistance to chemotherapy and radiation therapy as well as to the selective growth pressure of Cancer Stem Cells. An intriguing property of cancer growth is represented by the roles of proteases that are clearly involved in local spreading and in invasion of vessels. This role to control what could be called Cellular Adhesiveness involves chiefly extracellular proteases like the membrane Matrix Metallo Proteinases (MMP) (Knapinska and Fields, 2012). In addition also intracellular proteases serve important signalling functions (Turk et al, 2012) also in cancer cells, in particular in the regulation of cell growth and death, mainly through the apoptotic program (Fiandalo and Kyprianou, 2012).

I mentioned these effects to call attention on the cautions required to avoid artefacts when analysing protein composition in cancer specimens because the high content of proteases can eventually modify the pattern of expressed proteins, eventually degrading those that are more sensitive to proteolytic cleavage. This is particularly relevant when denaturation prone buffers are employed, as it is the case with the TUC (Thiourea, Urea, Chaps) which is mandatory for a nice solubilisation of proteins particularly from pathologic tissues that might have been exposed to inflammatory conditions in vivo and therefore display extensive fibrosis. The effective solubilisation brought about by TUC buffer promotes indeed protein unfolding thus exposing to attack by proteases of internal protein regions that are normally protected from nicking because they are not available to the protease active site (Fontana, 1991; Hubbard, 1998). These precautions obviously do not apply so stringently in *in vitro* reconstructed models, as in the case of protein PTM, where presence of proteases is certainly better controlled.

These difficulties can therefore be circumvented by means of addition of mixtures of inhibitors of proteases, the so-called inhibitor cocktails, that include inhibitors to most type of proteases (serine, cysteine, acidic and metal proteases). Under these conditions reliable and reproducible extraction of tissue proteins has been achieved with a solubilisation buffer that is suitable for IEF although containing a zwitterionic detergent.

Coming more closely to the experimental results in the two systems pertinent to clinical oncology that we have explored up to now we have confirmed altered protein expression both in comparison between healthy and affected tissue from the same patient or in cancer cell lines cultured under conditions that facilitate the assumption of a more aggressive phenotype. In the case of infiltrating duct breast carcinoma many proteins displayed altered expression - in most cases increased expression- when compared with normal breast, with proteins mainly related to the function of the cytoskeletal motility and of protein folding and transmembrane transport (including heat shock proteins). Also the protein expression in cholangiocarcinoma cell lines is altered during culture. At the time I am writing it was yet possible to give information of the proteins that are more clearly affected, but I hope this information will be available the day of my presentation.

References

Beninati S, Martinet N, Folk JE (1988) High-performance liquid chromatographic method for the determination of ε-(γ-glutamyl)lysine and mono- and bis-γ-glutamyl derivatives of putrescine and spermidine. J. Chromatogr. 443: 329–335.

Fiandalo MV, Kyprianou N. (2012) Caspase control: protagonists of cancer cell apoptosis. Exp Oncol. 34: 165-175.

Fontana A. (1991) Analysis and modulation of protein stability. Curr Opin Biotechnol. 2: 551-560.

Grebe SKG, Singh RJ (2011) LC-MS/MS in the Clinical Laboratory – Where to From Here? Clin Biochem Rev. 32: 5–31

Hubbard SJ. (1998) The structural aspects of limited proteolysis of native proteins. Biochim Biophys Acta. 1382: 191-206

Knapinska A, Fields GB. (2012) Chemical biology for understanding matrix metalloproteinase function. Chembiochem. 13: 2002-2020

Malito E, Bursulaya B, Chen C, Lo Surdo P, Picchianti M, Balducci E, Biancucci M, Brock A, Berti F, Bottomley MJ, Nissum M, Costantino P, Rappuoli R, Spraggon G. (2012) Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197.Proc Natl Acad Sci U S A. 109: 5229-5234.

Parameswaran KN, Cheng XF, Chen EC, Velasco PT, Wilson JH, Lorand L. (1997) Hydrolysis of gamma:epsilon isopeptides by cytosolic transglutaminases and by coagulation factor XIIIa. J Biol Chem. 272: 10311-10317

Serpa JJ, Parker CE, Petrotchenko EV, Han J, Pan J, Borchers CH. (2012)Mass spectrometry-based structural proteomics. Eur J Mass Spectrom 18: 251-267

Sinz A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrom Rev. 25: 663-682

Squerzanti M, Cervellati C, Ura B, Mischiati C, Pucci P, Annunziata S, Iannone C, Casadio R, Bergamini CM, Esposito C. (2013) The side chain of glutamine 13 is the acyl-donor amino acid modified by type 2 transglutaminase in subunit T of the native rabbit skeletal muscle troponin complex. Amino Acids. 44: 227-233

Turk B, Turk du SA, Turk V. (2012) Protease signalling: the cutting edge. EMBO J. 31: 1630-1643