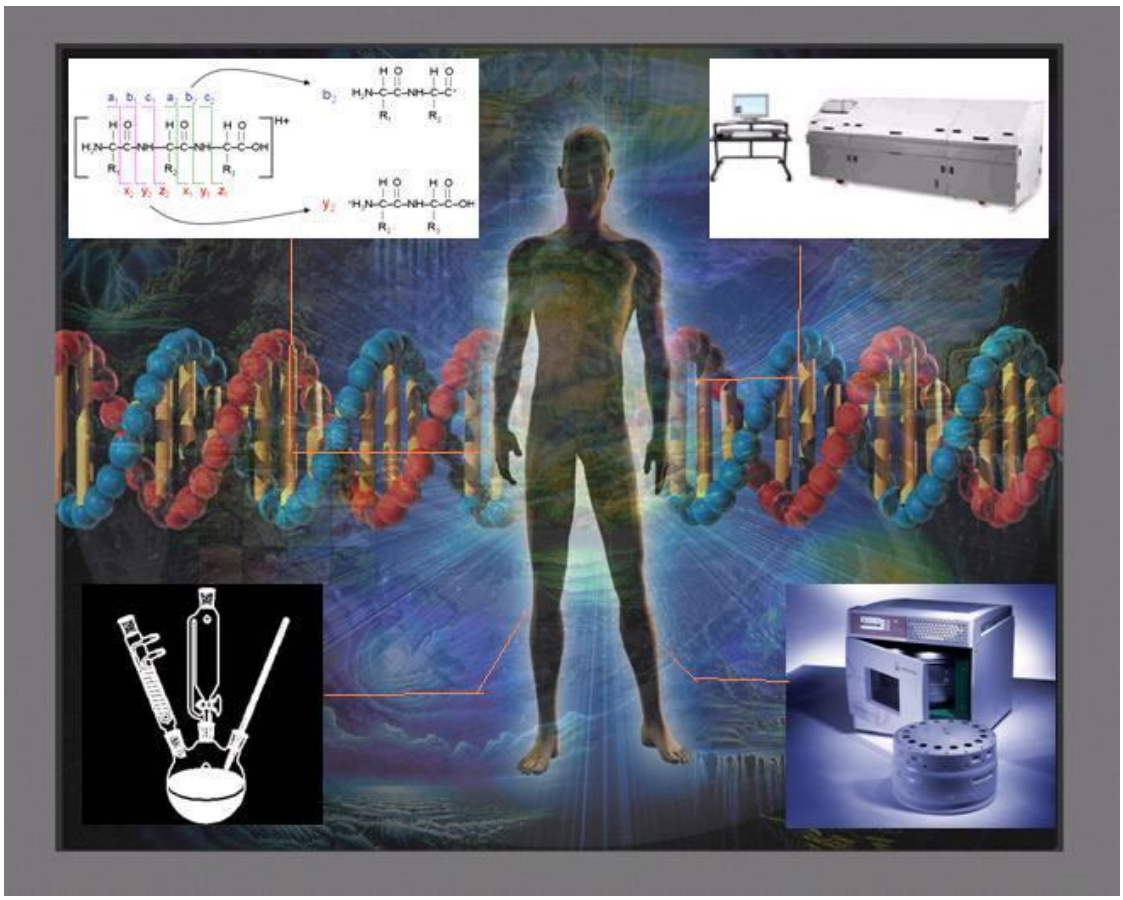
Synthesis and application of new solid phase techniques in quantitative proteomics using MALDI and ESI mass spectrometry



By

Petros Moschidis

Supervisor: Prof. Giovanni Sindona

Co-Supervisor: Dr. Constantinos Athanassopoulos

UNIVERSITÀ DELLA CALABRIA

Dipartimento di CHIMICA

Academic Year 2007-2008

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

UNIVERSITY OF PATRAS

UNIVERSITÀ DELLA CALABRIA

Dipartimento di CHIMICA

Submitted in the partial fulfilment of the requirements of the

Degree of Doctor of Philosophy

(PhD) in

Methodologies for the development of molecules of pharmacological interest

XXI cycle (CHIM06)

Synthesis and application of new solid phase techniques in quantitative proteomics using MALDI and ESI mass spectrometry

(PhD) Thesis

Department of Chemistry

University of Calabria

Italy

Supervisor
Prof Giovanni Sindona

Coordinator
Prof Bartolo Gabriele

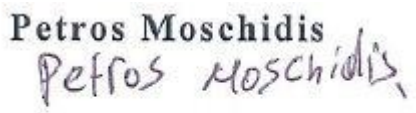
Candidate
Petros Moschidis

Academic Year 2007/2008

Declaration

I hereby declare that the thesis that follows is my own composition, that it is a record of the work done by myself, and that it has not been presented in any previous application for a higher degree.

Petros Moschidis



(26/11/2008)

ACKNOWLEDGMENT

I would therefore like to thank my supervisors Prof. Giovanni Sindona, Prof. Anna Napoli and Dr. Constantinos Athanassopoulos, whose unwavering support and enthusiasm were fundamental to my own enjoyment of the research, and whose near limitless knowledge of their own fields meant that the training I received on this project was unique.

I would like to thank the entire staff of the Mass Spectrometry lab of Unical which has contributed in no small measure both to the work presented in this thesis as well to the pleasant and vibrant atmosphere in the lab. In particular, Dr. Donatella Aiello for her valuable advices during all this 3 years that made me more precise while working in the “invisible” microscale of Proteomics (do not shake unknown vials!). The “swimming pool boys” Raf and Fabio for their lessons of Italian language, Art and Culture. I would also like to thank Prof. Leonardo, Domenico, Barbara, Naim and the very enthusiastic and capable graduate students Alessandra and Elvira.

An important part of my work has been made in the lab of Organic Synthesis of Unical. I would therefore like to thank my friends Amedeo and Bea for their illimitable help in Science and in the real life. I must not forget Marianna, Garofalo, Attilio, Monica, Francesca, Elvira and Prof Tagarelli. A special thank to Maurizio for the advanced Italian language course (Panti, Pasquale, etc).

For my stay as a visiting PhD student in Patras (where I have completed my undergraduate studies in 2005) under the supervising of Dr Costantinos Athanassopoulos, I have to thank Prof Dionissios Papaioanou for the hospitality in his laboratory. I would also like to thank the people of this lab: Giorgos, Stavros, Thomas, Katia, Mitsos and Antonis.

A PhD is not only Science but a life-experience. Thus I may acknowledge the friends in blocco 3 (dirty house): Attya, Ashif, Hisham, Cappo (Habib), Jose’, Butaflica and all the other persons that came and left. Sure I have to refer to my good Egyptian friend Mohammed Youssry Ali (…) and remember all the “rainy days” as well as the joyful moments we passed together in bankina this 3 year in Arcavacata di Rende.

I big thank to Mrs Sofia for her unique artistic touch on the cover layout of this thesis.

Finally, I owe all of my accomplishments to my parents. My father and mother, they sacrificed much to provide me with the quality of life from which I could choose what I want to be and to do.

My PhD scholarship was funded by the School of Hard Sciences “Bernardino Telesio” 1st cycle.

Contents

Abbreviations i

Abstract iii

Chapter 1. ***Proteomics*** 1

1.1 Proteomic Analysis Techniques 2

1.1.1 Protein Separation 3

1.1.2 Protein identification 4

1.2 Mass spectrometry 5

1.2.1 Sample introduction 6

1.2.2 Methods of sample ionisation 7

1.2.3 Analysis and separation of sample ions 7

1.2.4 Detection and recording of sample ions 7

1.3 Electrospray 8

1.3.1 Electrospray ionisation 8

1.4 Nanospray ionisation 9

1.5 Matrix assisted laser desorption ionisation (MALDI) 10

1.6 Positive or negative ionisation? 13

1.7 Tandem mass spectrometry (MS-MS) 14

1.7.1 Tandem mass spectrometry 14

1.7.2 Product or daughter ion scanning 14

1.7.3 Precursor or parent ion scanning 15

1.7.4 Constant neutral loss scanning. 15

1.7.5 Selected/multiple reaction monitoring 15

1.8 Peptide Sequencing by Tandem Mass Spectrometry 16

1.8.1 Peptide sequencing: H_2N-CH(R')-CO-NH-CH(R")-CO_2H 16

1.8.2 Immonium ions 17

1.8.3 Protein identification common procedure 17

1.8.4 Peptide sequencing in summary 19

1.8.5 Peptide Mass Fingerprinting (PMF) 20

Chapter 2. ***Quantitative Proteomics*** 25

2.1 Quantitation via 2DE and MS 26

2.2 Quantitation without stable isotope tags 26

2.3 Stable Isotope Labelling 27

2.3.1 O^{18} Labelling 29

2.3.2 Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC) 30

2.4 Solid-Phase Chemical tools for proteomics 32

2.4.1 Solid Phase Synthesis 32

2.4.2 Solid phase support (resin) 33

2.4.3 TentaGel resins 35

2.5 Quantitative proteome analysis by solid-phase isotope tagging and MS 36

Chapter 3. ***Results and discussion*** 41

3.1 Mass spectrometry equipment 43

3.2 Dansyl Solid Phase Tag 44

3.2.1 Synthesis of d6-Dansyl Chloride 44

3.2.2 Development of Dansyl containing resins 49

3.2.3 Reactivity test of Dansyl –OSu Tags 50

3.2.4 Reactivity test of Dansyl –Iodo Tags 53

3.3 Cysteine-Cysteine Isobaric Solid Phase Tag 57

3.3.1 Development of Cys-Cys Isobaric resins 57

3.3.2 Reactivity test of Cys-Cys Isobaric Tags 58

3.4 Glycine-Alanine Isobaric Isotopic Label-Free Solid Phase Tag 63

3.4.1 Development of Glycine-Alanine Isobaric Isotopic Label-Free resins 63

3.4.2 Reactivity test of Glycine-Alanine Isobaric Isotopic Label-Free resins 64

Chapter 4. ***Experimental part*** 75

Chapter 5. ***References*** 89

Chapter 6. ***Appendix*** 97

Abbreviations

2D	Two-dimensional
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
AA	Amino acid
ACN	Acetonitrile
BSA	Bovine Serum Albumin
CID	Collision-induced dissocation
CHCA	Alpha-cyano-4-hydroxycinnamic acid
C-terminal	Carboxy terminal
DCC	N,N'-dicyclohexylcarbodiimide
DCM	Dichloromethane
DIC	N,N'-diisopropylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethyl Formamide
DNA	Deoxy-Ribonucleic Acid
DNS	5-dimethylamino-1- naphthalenesulfonic acid
Dansyl	5-dimethylamino-1- naphthalenesulfonic acid
ESI	Electrospray ionisation
FA	Formic Acid
FAB	Fast Atom Bombardment
Fmoc	9H-fluoren-9-ylmethoxycarbonyl
HBTU	O-Benzotriazole-N,N,N’,N’-tetramethyl-uronium-hexafluoro-phosphate
HPLC	High Performance Liquid Chromatography
ICAT	Isotope coded affinity tag
iTRAQ	Isotope Tagging for Relative and Absolute Quantitation
IEF	Isoelectric focusing
i-PrOH	Isopropanol
LC	Liquid chromatography
m/z	Mass-to-charge ratio
MALDI	Matrix-Assisted Laser Desorption and Ionisation
MS	Mass Spectrometry
MS/MS	Tandem MS
Mtt	Methyltrityl
N-terminal	Amino terminal
NHS	*N*-Hydroxysuccinimide
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
RAM	Ring Amide Linker
RNA	Ribonucleic Acid
RP	Reverse-phase
SILAC	Stable Isotope Labelling by Amino Acids in Cell Culture
SDS	Sodium dodecyl sulphate
SPSS	Solid Phase Peptide Synthesis
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Tri-fluoro acetate
TFMSA	Trifluoromethanesulfonic acid
TIC	Total Ion Chromatogram
TOF	Time-Of-Flight
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

Abstract

Proteomics, the analysis of the protein complement of a cell or an organism has grown rapidly as a category of the life sciences. Mass spectrometry (MS) is one of the central detection techniques in proteome analysis, yet it has to rely on prior sample preparation steps that reduce the enormous complexity of the protein mixtures obtained from biological systems. For that reason, a number of tagging (or labeling) strategies have been developed that target specific amino acid residues or post-translational modifications, enabling the enrichment of subfractions via affinity clean up, resulting in the identification of an ever-increasing number of proteins. In addition, the attachment of stable isotope-labeled tags now allow the relative quantitation of protein levels of two samples, e.g. those representing different cell states, which is of great significance for drug discovery and molecular biology.

This work presents the research for new stable isotope-labeled (as well as stable-isotope free) solid phase strategies for the identification and relative quantitation of complex protein mixtures using MALDI and ESI-MS, MS/MS technology.

1

Proteomics

1. Proteomics

Proteomics is the large-scale study of all expressed proteins. Proteins are vital parts of living organisms, as they are the main components of the physiological metabolic pathways of cells. The term "proteomics" was coined to make an analogy with genomics, the study of the genes. The word *proteome* is a combination of *prot*ein (from the Greek word *πρώτα* ("prota"), meaning "of primary importance") and gen*ome*. The proteome is the entire complement of proteins, including all the proteins in the cell, all protein isoforms and their post-translational modifications. Proteomics is often considered the next step in the study of biological systems, after genomics. It is much more complicated than genomics, mostly because while an organism's genome is constant - with exceptions such as the addition of genetic material caused by a virus, or the rapid mutations, transpositions, and expansions that can occur in a tumor - the proteome differs from cell to cell. This is because proteome is changing through its biochemical interactions with the genome and the environment. One organism has different protein expression in different parts of its body, in different stages of its life cycle and in different environmental conditions.

The study of the proteome presents us with several technical challenges: the complexity and heterogeneity of the proteome, the instability of individual proteins, and the difficulty of performing quantitative analyses. The template for the proteome, the genome, contains a vast number of protein coding sequences. The human genome, for example contains approximately 30,000 genes, most of which will be translated to protein. Complexity is increased at the transcript level by splice variants. Additionally, once translated from mRNA, proteins are commonly *post-translationally modified* by factors such as acetylation[1], methylation[2], glycosylation[3], phosphorylation[4], truncation events[5] and sulphation[6], vastly increasing the complexity of the array of proteins expressed even in a single cellular state.

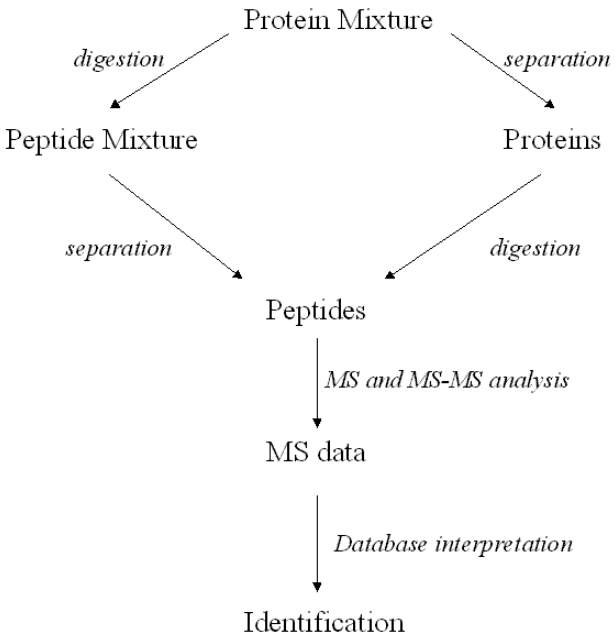
Proteins perform thousands of different tasks, such as structural maintenance, molecular trafficking and transcriptional regulation, required to maintain a functioning cell, and are required to be heterogeneous in structure and physiochemical properties. This provides many difficulties for researchers. Not all proteins will remain soluble

under the same conditions, remain stable and intact under the same conditions, or be effectively analyzed by the same instrumentation. Data derived from proteomic analyses can be enhanced by direct comparison of multiple phenotypes. For this reason, comparative quantitation[7] is an important aspect of proteomics. Obtaining abundance data from a heterogeneous, extremely complex mixture is not trivial, and several methods have been employed with variable success.

Despite this, proteomics has developed into a highly successful area of research, which provides the unique ability to identify and quantify hundreds or thousands of proteins simultaneously. Indeed, proteomic studies have provided invaluable insight into many important areas of molecular biology, for example the functionality of the centrosome[8], flagellar motility in bloodstream trypanosomes[9] and protein degradation using the ubiquitin ligase assembly[10].

1.1 Proteomic Analysis Techniques

Attempts to overcome the challenges of heterogeneity, instability and dynamic range of proteins lead to the two main requirements of any technique that approaches a ‘proteomic’analysis (Scheme 1) technique: the ability to reduce sample complexity to manageable levels, and the ability to identify proteins once they are separated. The most popular methods for separating proteins are gels and chromatography columns, and in both cases, the principle of their action is to split a complex mixture of proteins into multiple lower complexity fractions. The most effective tool for protein identification has become the mass spectrometer. In simple terms, a mass spectrometer is a device for detecting the number of ions present with a particular mass. In essence, mass spectrometers separate or filter ions by mass.



Scheme 1. Proteomics workflow

1.1.1 Protein Separation

Various approaches are possible for protein separation including 1-D SDS-PAGE. However, for quantitative and comparative proteomic analysis, there is still no widely available technology that surpasses 2-D gel electrophoresis (2-DE), which has been the method of choice for separation of complex protein mixtures for several decades[11]. In the 2-D-PAGE the protein mixtures resolved first in one dimension isoelectrofocusing (IEF),that separates proteins on the basis of isoelectric point (*pI*). These proteins are then resolved on a second dimension SDS- PAGE, which separate according to molecular weight. The proteins, that run on the gel, that have been separated during electrophoresis, can be visualized using different staining techniques, most common are Coomassie Brilliant Blue staining is relatively easy, cost-effective and compatible with subsequent protein identification by mass spectrometry (MS), but it is only moderately sensitive, with a limit of approximately 10 ng protein. The other alternative has been silver-staining, which is more sensitive, detecting as little as 0.5 ng protein[12] but not particularly quantitative and less suitable for MS[13].

Moreover the availability of improved stationary-phase materials and hardware has greatly improved the performance of LC systems for protein purification. Although HPLC of intact proteins has not become a widely used technique for analytical proteomics, it is nevertheless highly applicable as an initial step to fractionate protein mixtures. Diverse chromatographic separations are available, including RP, anion and cation exchange, size exclusion, and affinity chromatography.The latter is particularly attractive as a means of pulling a subset of proteins from a complex mixture.

1.1.2 Protein identification

The mass spectrometry is widely used and is the method of choice for both protein identification and characterization. The first step toward protein identification is typically excision of 1-D or 2-D gel containing the protein bands or spots of interest, in-gel digestion with a site-specific protease (commonly trypsin), and finally MS analysis of the resultant eluted peptides (Figure 1). Two MS platforms in particular represent powerful tools for proteomic studies. The first, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS is typically used to measure the masses of the peptides derived from the trypsinized parent protein spot, generating a 'peptide mass fingerprint' (PMF). A common alternative to the PMF approach is de novo sequencing by electrospray ionisation tandem mass spectrometry (ESI) MS/MS, which yields amino acid sequences of selected tryptic peptides (Figure 1). The first step of tandem MS involves ionisation of a sample and separation based upon the mass-to-charge ratio (*m/z*) of the primary ions. An ion with a specific *m/z* value is then selected, fragmented, and the fragment ions detected after passing through the second mass spectrometer. This process produces a series of fragment ions that can differ by single amino acids allowing a portion of the peptide sequence, termed an 'amino acid sequence tag', to be determined and used for database searching[11].

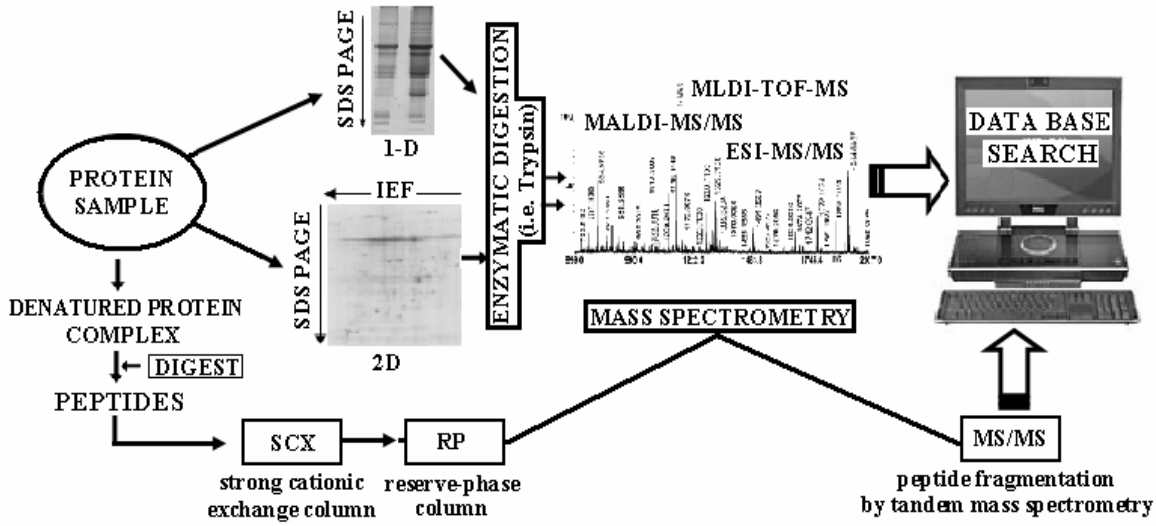
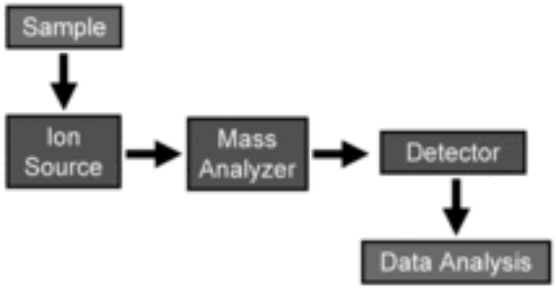


Figure 1. Peptide identification

1.2 Mass spectrometry

Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample. For large samples such as biomolecules, molecular masses can be measured to within an accuracy of 0.01% of the total molecular mass of the sample *i.e.* within a 4 Daltons (Da) or atomic mass units (amu) error for a sample of 40,000 Da. This is sufficient to allow minor mass changes to be detected, *e.g.* the substitution of one amino acid for another, or a post-translational modification. For small organic molecules the molecular mass can be measured to within an accuracy of 5 ppm or less, which is often sufficient to confirm the molecular formula of a compound, and is also a standard requirement for publication in a chemical journal. Structural information can be generated using certain types of mass spectrometers, usually those with multiple analysers which are known as tandem mass spectrometers. This is achieved by fragmenting the sample inside the instrument and analysing the products generated. This procedure is useful for the structural elucidation of organic compounds and for peptide or oligonucleotide sequencing.

Mass spectrometers can be divided into three fundamental parts, namely the ionisation **source,** the **analyser** , and the **detector**.



Ionisation source : MALDI , ESI , Chemical Ionisation , Electron Ionisation, FAB

Mass Analyzer : Quadrupole , TOF , Magnetic Sector , Ion Trap , MS/MS

Detector : Electron multiplier , Photomultiplier

Figure 2. Simplified scheme of a mass spectrometer

The sample has to be introduced into the ionisation source of the instrument. Once inside the ionisation source, the sample molecules are ionised, because ions are easier to manipulate than neutral molecules. These ions are extracted into the analyser region of the mass spectrometer where they are separated according to their mass (*m*) -to-charge (*z*) ratios (*m/z*) . The separated ions are detected and this signal sent to a data system where the *m/z* ratios are stored together with their relative abundance for presentation in the format of a *m/z* spectrum**.**

The analyser and detector of the mass spectrometer, and often the ionisation source too, are maintained under high vacuum to give the ions a reasonable chance of travelling from one end of the instrument to the other without any hindrance from air molecules. The entire operation of the mass spectrometer, and often the sample introduction process also, is under complete data system control on modern mass spectrometers.

1.2.1 Sample introduction

The method of sample introduction to the ionisation source often depends on the ionisation method being used, as well as the type and complexity of the sample.

The sample can be inserted directly into the ionisation source, or can undergo some type of chromatography *en route* to the ionisation source. This latter method of sample introduction usually involves the mass spectrometer being coupled directly to a high pressure liquid chromatography (HPLC), gas chromatography (GC) or capillary electrophoresis (CE) separation column, and hence the sample is separated into a series

of components which then enter the mass spectrometer sequentially for individual analysis.

1.2.2 Methods of sample ionisation

Many ionisation methods are available and each has its own advantages and disadvantages. The ionisation method to be used depends on the type of sample under investigation and the mass spectrometer available. The ionisation methods used for the majority of biochemical analyses are Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI*).*

1.2.3 Analysis and separation of sample ions

The main function of the mass analyser is to separate**,** or resolve, the ions formed in the ionisation source of the mass spectrometer according to their mass-to-charge (*m/z*) ratios. There are a number of mass analysers currently available, the better known of which include quadrupoles**,** time-of-flight (TOF) analysers, magnetic sectors , and both Fourier transform and quadrupole ion traps .

These mass analysers have different features, including the *m/z* range that can be covered, the mass accuracy, and the achievable resolution. The compatibility of different analysers with different ionisation methods varies. For example, all of the analysers listed above can be used in conjunction with electrospray ionisation, whereas MALDI is not usually coupled to a quadrupole analyser.

Tandem (MS-MS) mass spectrometers are instruments that have more than one analyser and so can be used for structural and sequencing studies. Two, three and four analysers have all been incorporated into commercially available tandem instruments, and the analysers do not necessarily have to be of the same type, in which case the instrument is a hybrid one. More popular tandem mass spectrometers include those of the quadrupole-quadrupole, magnetic sector-quadrupole, and more recently, the quadrupole-time-of-flight geometries.

1.2.4 Detection and recording of sample ions

The detector monitors the ion current, amplifies it and the signal is then transmitted to the data system where it is recorded in the form of mass spectra**.** The *m/z* values of the

ions are plotted against their intensities to show the number of components in the sample, the molecular mass of each component, and the relative abundance of the various components in the sample. The type of detector is supplied to suit the type of analyser; the more common ones are the photomultiplier**,** the electron multiplier and the micro-channel plate detectors.

1.3 Electrospray

Electrospray mass spectrometry (ESMS) has been developed for use in biological mass spectrometry by Fenn et al[14].

1.3.1 Electrospray ionisation

Electrospray Ionisation (ESI) is one of the Atmospheric Pressure Ionisation (API) techniques and is well-suited to the analysis of polar molecules ranging from less than 100 Da to more than 1,000,000 Da in molecular mass (Figure 3).

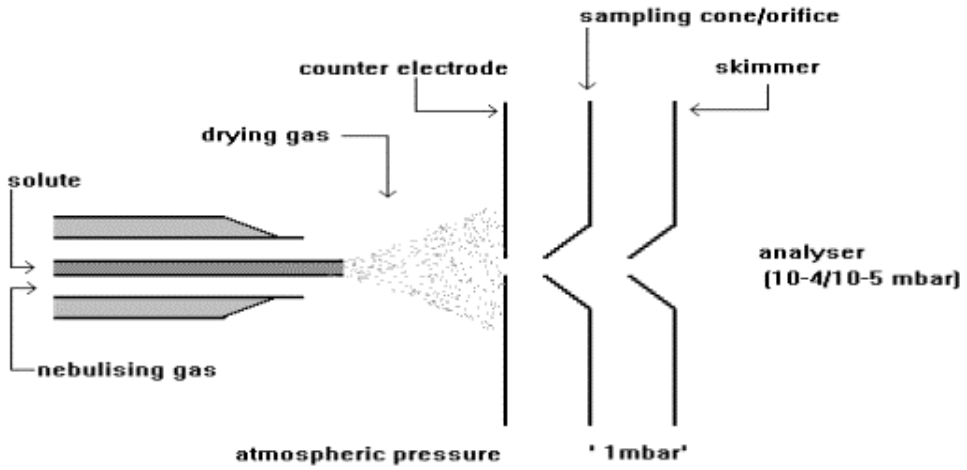


Figure 3. Standard electrospray ionisation source

During standard electrospray ionisation[15], the sample is dissolved in a polar, volatile solvent and pumped through a narrow, stainless steel capillary (75 - 150 micrometers i.d.) at a flow rate of between 1 μL/min and 1 mL/min. A high voltage of 3 or 4 kV is applied to the tip of the capillary, which is situated within the ionisation source of the mass spectrometer, and as a consequence of this strong electric field, the sample

emerging from the tip is dispersed into an aerosol of highly charged droplets, a process that is aided by a co-axially introduced nebulising gas flowing around the outside of the capillary (Figure 4). This gas, usually nitrogen, helps to direct the spray emerging from the capillary tip towards the mass spectrometer. The charged droplets diminish in size by solvent evaporation, assisted by a warm flow of nitrogen known as the drying gas which passes across the front of the ionisation source. Eventually charged sample ions, free from solvent, are released from the droplets, some of which pass through a sampling cone or orifice into an intermediate vacuum region, and from there through a small aperture into the analyser of the mass spectrometer, which is held under high vacuum. The lens voltages are optimised individually for each sample.

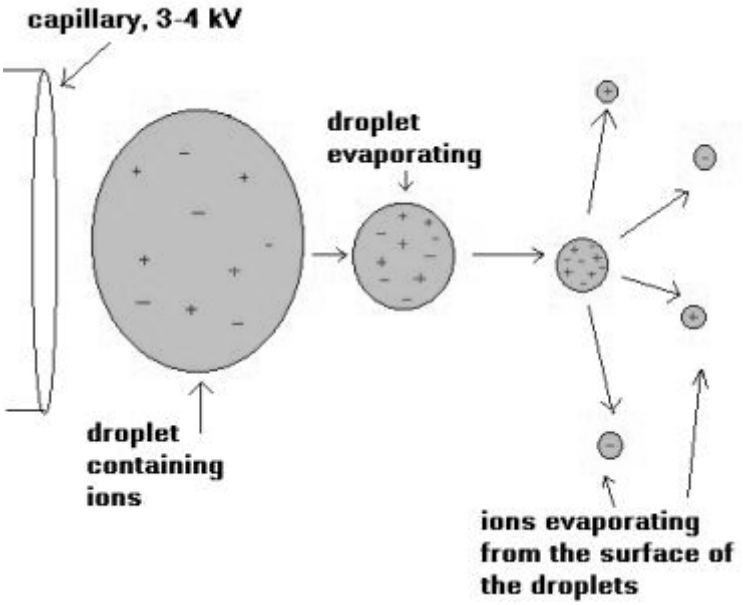


Figure 4. The electrospray ionisation process

1.4 Nanospray ionisation

Nanospray ionisation[16] is a low flow rate version of electrospray ionisation. A small volume (1-4 microL) of the sample dissolved in a suitable volatile solvent, at a concentration of ca. 1 - 10 pmol/microL, is transferred into a miniature sample vial. A reasonably high voltage (ca. 700 - 2000 V) is applied to the specially manufactured gold-plated vial resulting in sample ionisation and spraying. The flow rate of solute and solvent using this procedure is very low, 30 - 1000 nL/min, and so not only is far less sample consumed than with the standard electrospray ionisation technique, but also a small volume of sample lasts for several minutes, thus enabling multiple experiments to

be performed. A common application of this technique is for a protein digest mixture to be analysed to generate a list of molecular masses for the components present, and then each component to be analysed further by tandem mass spectrometric (MS-MS) amino acid sequencing techniques.

ESI and nanospray ionisation are very sensitive analytical techniques but the sensitivity deteriorates with the presence of non-volatile buffers and other additives, which should be avoided as far as possible.

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

1.5 Matrix assisted laser desorption ionisation (MALDI)

Matrix Assisted Laser Desorption Ionisation (MALDI)[17] (Figures 5 and 6) deals well with thermo labile, non-volatile organic compounds especially those of high molecular mass and is used successfully in biochemical areas for the analysis of proteins, peptides, glycoproteins, oligosaccharides, and oligonucleotides. It is relatively straightforward to use and reasonably tolerant to buffers and other additives. The mass accuracy depends on the type and performance of the analyser of the mass spectrometer, but most modern instruments should be capable of measuring masses to within 0.01% of the molecular mass of the sample, at least up to ca. 40,000 Da.

MALDI is based on the bombardment of sample molecules with a laser light to bring about sample ionisation. The sample is pre-mixed with a highly absorbing matrix compound for the most consistent and reliable results and a low concentration of sample to matrix work best. The matrix transforms the laser energy into excitation energy for the sample, which leads to sputtering of analyte and matrix ions from the surface of the mixture. In this way energy transfer is efficient and also the analyte molecules are spared excessive direct energy that may otherwise cause decomposition. Most commercially

available MALDI mass spectrometers now have a pulsed nitrogen laser of wavelength 337 nm.

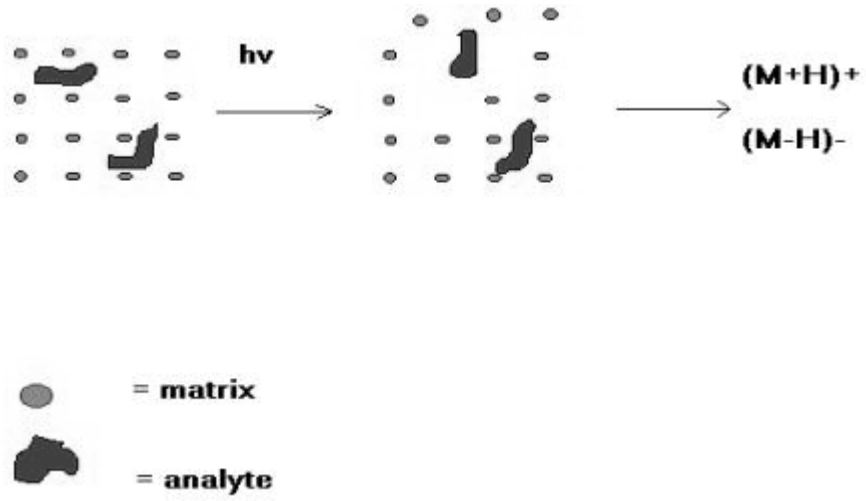


Figure 5. Matrix assisted laser desorption ionisation (MALDI)

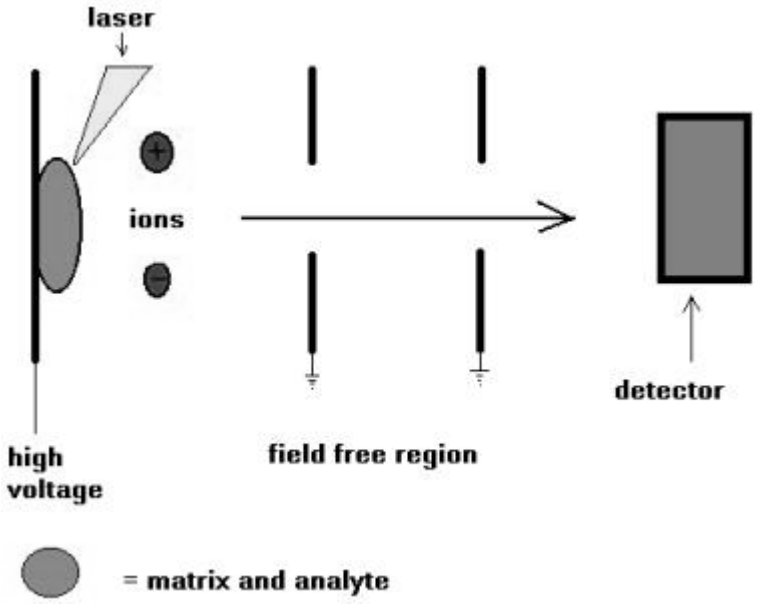


Figure 6. Simplified schematic of MALDI-TOF mass spectrometry (linear mode)

The sample to be analysed is dissolved in an appropriate volatile solvent, usually with a trace of trifluoroacetic acid if positive ionisation is being used, at a concentration of ca. 10 pmol/μL and an aliquot **(**1-2 μL**)** of this removed and mixed with an equal volume of a solution containing a vast excess of a matrix. A range of compounds is suitable for use as matrices: sinapinic acid is a common one for protein analysis while alpha-cyano-4-

hydroxycinnamic acid is often used for peptide analysis (Figure 7). An aliquot (1-2 μL) of the final solution is applied to the sample target which is allowed to dry prior to insertion into the high vacuum of the mass spectrometer. The laser is fired, the energy arriving at the sample/matrix surface optimised, and data accumulated until a *m/z* spectrum of reasonable intensity has been amassed. The time-of-flight analyser separates ions according to their mass(m)-to-charge(z) (*m/z*) ratios by measuring the time it takes for ions to travel through a field free region known as the flight, or drift, tube. The heavier ions are slower than the lighter ones.

The *m/z* scale of the mass spectrometer is calibrated with a known sample that can either be analysed independently (external calibration) or pre-mixed with the sample and matrix (internal calibration).

MALDI is also a "soft" ionisation method and so results predominantly in the generation of singly charged molecular-related ions regardless of the molecular mass, hence the spectra are relatively easy to interpret. Fragmentation of the sample ions does not usually occur.

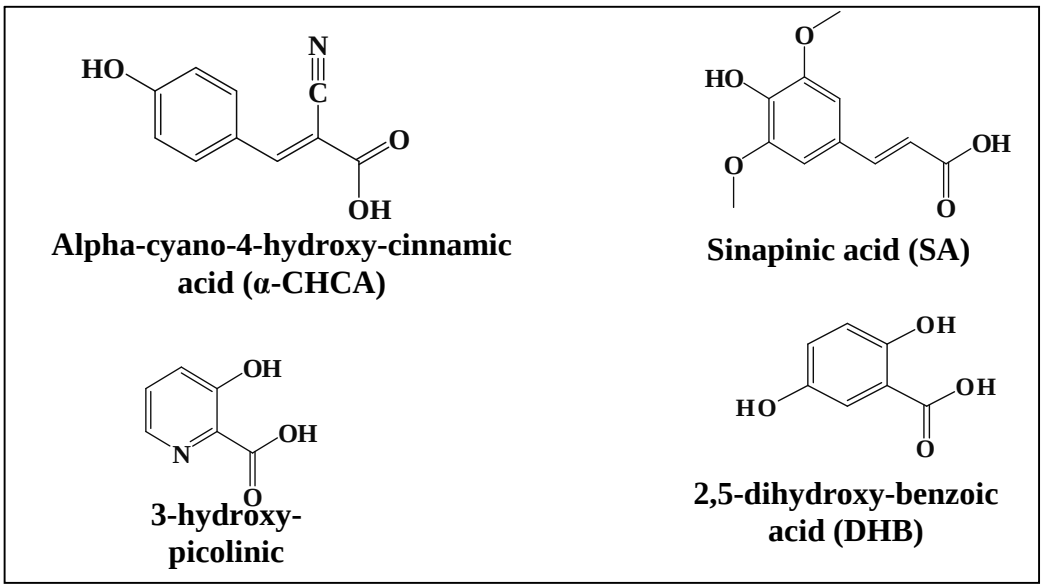


Figure 7. Common MALDI matrices

In positive ionisation mode the protonated molecular ions (M+H+) are usually the dominant species, although they can be accompanied by salt adducts, a trace of the doubly charged molecular ion at approximately half the *m/z* value, and/or a trace of a

dimeric species at approximately twice the *m/z* value. Positive ionisation is used in general for protein and peptide analyses (Figure 8).

In negative ionisation mode the deprotonated molecular ions (M-H-) are usually the most abundant species, accompanied by some salt adducts and possibly traces of dimeric or doubly charged materials. Negative ionisation can be used for the analysis of oligonucleotides and oligosaccharides.

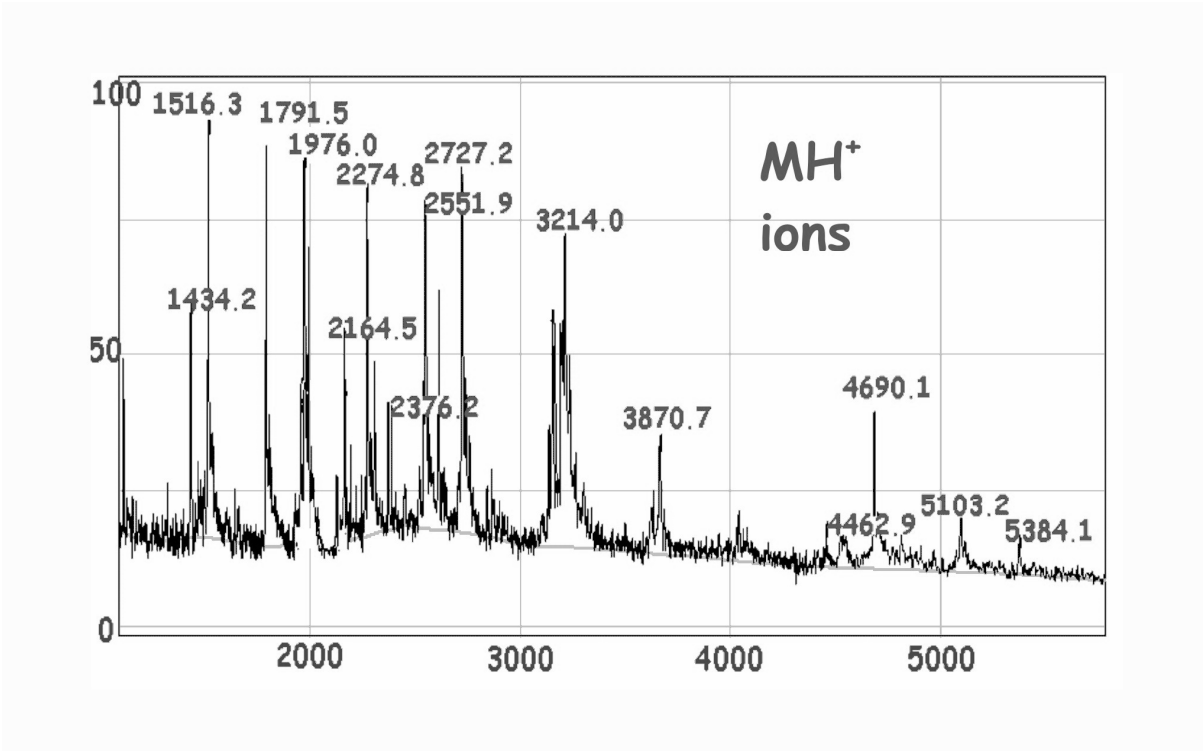


Figure 8. Positive ionisation MALDI m/z spectrum of a peptide mixture using alpha-cyano-4-hydroxy-cinnamic acid as matrix

1.6 Positive or negative ionisation?

If the sample has functional groups that readily accept a proton (H^+) then positive ion detection is used e.g. amines $R\text{-}NH_2 + H+ = R\text{-}NH_3+$ as in proteins or peptides.

If the sample has functional groups that readily lose a proton then negative ion detection is used e.g. carboxylic acids $R\text{-}CO_2H = R\text{-}CO_2-$ and alcohols R-OH = R-O- as in saccharides or oligonucleotides**.**

1.7 Tandem mass spectrometry (MS-MS)

Structural and sequence information from mass spectrometry.

1.7.1 Tandem mass spectrometry

Tandem mass spectrometry (MS-MS) is used to produce structural information about a compound by fragmenting specific sample ions inside the mass spectrometer and identifying the resulting fragment ions. This information can then be pieced together to generate structural information regarding the intact molecule. Tandem mass spectrometry also enables specific compounds to be detected in complex mixtures on account of their specific and characteristic fragmentation patterns.

A tandem mass spectrometer is a mass spectrometer that has more than one analyser, in practice usually two. The two analysers are separated by a collision cell into which an inert gas (e.g. argon, xenon) is admitted to collide with the selected sample ions and bring about their fragmentation. The analysers can be of the same or of different types, the most common combinations being:

- quadrupole - quadrupole
- magnetic sector - quadrupole
- magnetic sector - magnetic sector
- quadrupole - time-of-flight.

Fragmentation experiments can also be performed on certain single analyser mass spectrometers such as ion trap and time-of-flight instruments, the latter type using a post-source decay experiment to effect the fragmentation of sample ions.

Tandem mass spectrometry analyses.

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

1.7.2 Product or daughter ion scanning

The first analyser is used to select user-specified sample ions arising from a particular component; usually the molecular-related (i.e. $(M+H)^+$ or $(M-H)^-$) ions. These chosen ions pass into the collision cell, are bombarded by the gas molecules which cause fragment ions to be formed, and these fragment ions are analysed i.e. separated

according to their mass to charge ratios, by the second analyser. All the fragment ions arise directly from the precursor ions specified in the experiment, and thus produce a fingerprint pattern specific to the compound under investigation.
This type of experiment is particularly useful for providing structural information concerning small organic molecules and for generating peptide sequence information.

1.7.3 Precursor or parent ion scanning

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

1.7.4 Constant neutral loss scanning.

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

1.7.5 Selected/multiple reaction monitoring

Both of the analysers are static in this case as user-selected specific ions are transmitted through the first analyser and user-selected specific fragments arising from these ions are measured by the second analyser. The compound under scrutiny must be known and have been well-characterised previously before this type of experiment is undertaken. This methodology is used to confirm unambiguously the presence of a

compound in a matrix e.g. drug testing with blood or urine samples. It is not only a highly specific method but also has very high sensitivity.

1.8 Peptide Sequencing by Tandem Mass Spectrometry

The most common usage of MS-MS in biochemical areas is the product or daughter ion scanning experiment which is particularly successful for peptide and nucleotide sequencing.

1.8.1 Peptide sequencing: H_2N-CH(R')-CO-NH-CH(R")-CO_2H

Peptides fragment in a reasonably well-documented manner[18,19]. The protonated molecules fragment along the peptide backbone and also show some side-chain fragmentation with certain instruments

There are three different types of bonds that can fragment along the amino acid backbone: the NH-CH**,** CH-CO, and CO-NH bonds. Each bond breakage gives rise to two species, one neutral and the other one charged, and only the charged species is monitored by the mass spectrometer. The charge can stay on either of the two fragments depending on the chemistry and relative proton affinity of the two species. Hence there are six possible fragment ions for each amino acid residue and these are labelled as in the diagram, with the a, b, and c" ions having the charge retained on the N-terminal fragment, and the x, y", and z ions having the charge retained on the C-terminal fragment. The most common cleavage sites are at the CO-NH bonds which give rise to the b and/or the y" ions. The mass difference between two adjacent b ions, or y"; ions, is indicative of a particular amino acid residue (Figure 9).

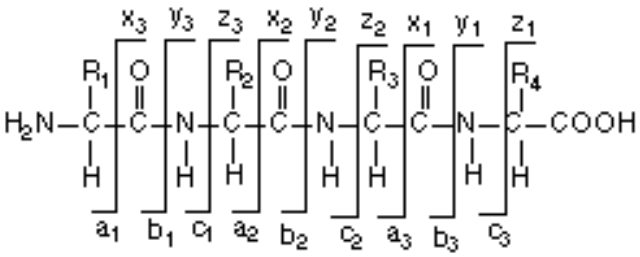


Figure 9. Peptide sequencing by tandem mass spectrometry - backbone cleavages

The extent of side-chain fragmentation detected depends on the type of analysers used in the mass spectrometer. A magnetic sector - magnetic sector instrument will give rise to high energy collisions resulting in many different types of side-chain cleavages. Quadrupole - quadrupole and quadrupole - time-of-flight mass spectrometers generate low energy fragmentations with fewer types of side-chain fragmentations.

1.8.2 Immonium ions (labelled "i") appear in the very low *m/z* range of the MS-MS spectrum. Each amino acid residue leads to a diagnostic immonium ion, with the exception of the two pairs leucine (L) and iso-leucine (I), and lysine (K) and glutamine (Q), which produce immonium ions with the same m/z ratio, i.e. *m/z* 86 for I and L, *m/z* 101 for K and Q. The immonium ions are useful for detecting and confirming many of the amino acid residues in a peptide, although no information regarding the position of these amino acid residues in the peptide sequence can be ascertained from the immonium ions.

1.8.3 Protein identification common procedure

A **protein identification study** would proceed as follows:

- a. The **protein** under investigation would be analysed by mass spectrometry to generate a molecular mass to within accuracy of 0.01%.
- b. The protein would then be **digested** with a suitable enzyme. **Trypsin** is useful for mass spectrometric studies because each proteolytic fragment contains a basic **arginine (R)** or **lysine (K)** amino acid residue, and thus is eminently suitable for positive ionisation mass spectrometric analysis. The digest mixture is analysed - without prior separation or clean-up - by mass spectrometry to produce a rather complex spectrum from which the molecular weights of all of the proteolytic fragments can be read. This spectrum, with its molecular weight information, is called a **peptide map**. (If the protein already exists on a **database**, then the peptide map is often sufficient to confirm the protein.) For these experiments the mass spectrometer would be operated in the **"MS" mode**, whereby the sample is sprayed and ionised from the nanospray needle and the ions pass through the

sampling cone, skimmer lenses, *Rf* hexapole focusing system, and the first (quadrupole) analyser. The quadrupole in this instance is not used as an analyser, merely as a lens to focus the ion beam into the second (time-of-flight) analyser which separates the ions according to their mass-to-charge ratio.

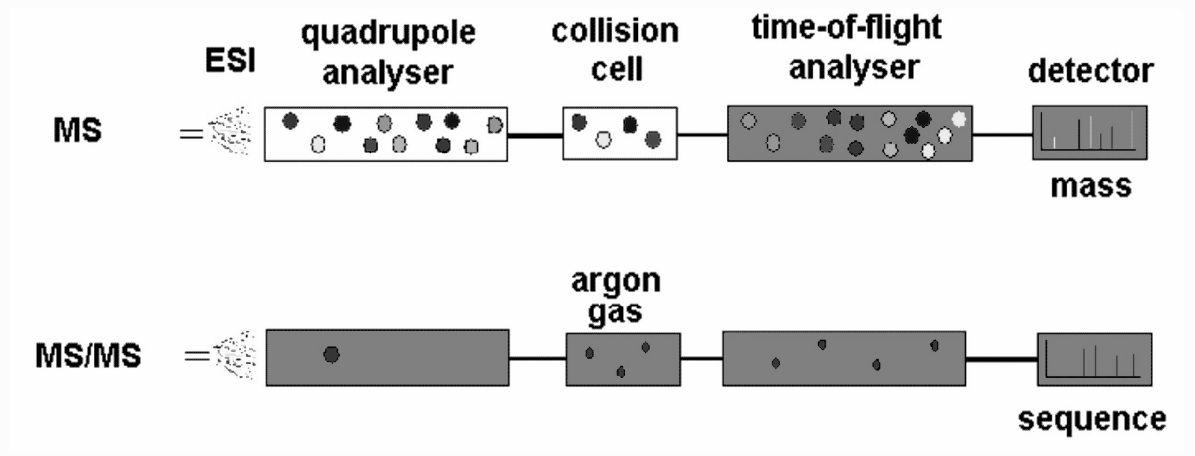


Figure 10. Q-TOF mass spectrometer operating in MS (upper) and MS/MS mode (lower) modes

- c. With the digest mixture still spraying into the mass spectrometer, the Q-TOF mass spectrometer is switched into **"MS/MS" mode** (Figure 10). The protonated molecular ions of each of the digest fragments can be independently selected and transmitted through the quadrupole analyser, which is now used as an analyser to transmit solely the ions of interest into the **collision cell** which lies inbetween the first and second analysers. An inert gas such as argon is introduced into the collision cell and the sample ions are bombarded by the collision gas molecules which cause them to fragment. The optimum collision cell conditions vary from peptide to peptide and must be optimised for each one. The **fragment (**or **daughter or product**) **ions** are then analysed by the second (time-of-flight) analyser. In this way an **MS/MS spectrum** is produced showing all the **fragment ions** that arise directly from the chosen **parent** or **precursor ions** for a given peptide component.

  An **MS/MS daughter** (or **fragment**, or **product**) ion spectrum is produced for each of the components identified in the proteolytic digest. Varying amounts of sequence information can be gleaned from each fragmentation spectrum, and the spectra need to be interpreted carefully. Some of the processing can be

automated, but in general the **processing** and **interpretation** of spectra will take longer than the data acquisition if accurate and reliable data are to be generated.

The amount of sequence information generated will vary from one peptide to another, Some peptide sequences will be confirmed totally, other may produce a partial sequence of, say, 4 or 5 amino acid residues. Often sequence "tag" of 4 or 5 residues is sufficient to search a protein database and confirm the identity of the protein.

1.8.4 Peptide sequencing in summary

Peptides fragment along the amino acid backbone to give sequence information. Peptides ca. 2500 Da or less produces the most useful data. The amount of sequence information varies from one peptide to another. Some peptides can generate sufficient information for a full sequence to be determined; others may generate a partial sequence of 4 or 5 amino acids.

A protein digest can be analysed as an entire reaction mix, without any separation of the products, from which individual peptides are selected and analysed by the mass spectrometer to generate sequence information. About 4 μL of solution is required for the analysis of the digest mixture, with a concentration based on the original protein of ca. 1-10 pmol/μL. MS/MS sequencing is a sensitive technique consuming little sample.

Sometimes the full protein sequence can be verified; some proteins generate sufficient information to cover only part of the sequence. 70 - 80% coverage is reasonable. Often a sequence "tag" of 4/5 amino acids from a single proteolytic peptide is sufficient to identify the protein from a database. The final point in this summary means that mass spectrometers have been found to be extremely useful for **proteomic** studies, as illustrated below.

The **proteomics procedure** usually involves excising individual spots from a **2-D gel** and independently **enzymatically digesting** the protein(s) contained within each spot and/or band, before analysing the digest mixture by mass spectrometer in the manner outlined above. Electrospray ionisation or MALDI could be used at this step.The initial **MS spectrum** determining the **molecular masses** of all of the components in the digest mixture can often provide sufficient information to search a **database** using just several of the molecular weights from this **peptide map.**

If the database search is not fruitful, either because the protein has not been catalogued, is previously uncharacterised, or the data are not accurate or comprehensive enough to distinguish between several entries in the database, then further information is required.This can be achieved by sample clean-up and then MS/MS studies to determine the amino acid sequences of the individual proteolytic peptides contained in the digest mixture, with which further database searching can be carried out.

1.8.5 Peptide Mass Fingerprinting (PMF)

PMF is an analytical technique for protein identification that was developed by John Yates and colleagues[20]. In this method, a "mass fingerprint" is obtained of a protein enzymatically degraded with a sequence-specific protease such as trypsin. This set of masses, typically obtained by MALDI-TOF, is then compared to the theoretically expected tryptic peptide masses for each entry in the database. The proteins can be ranked according to the number of peptide matches (Figure 11).

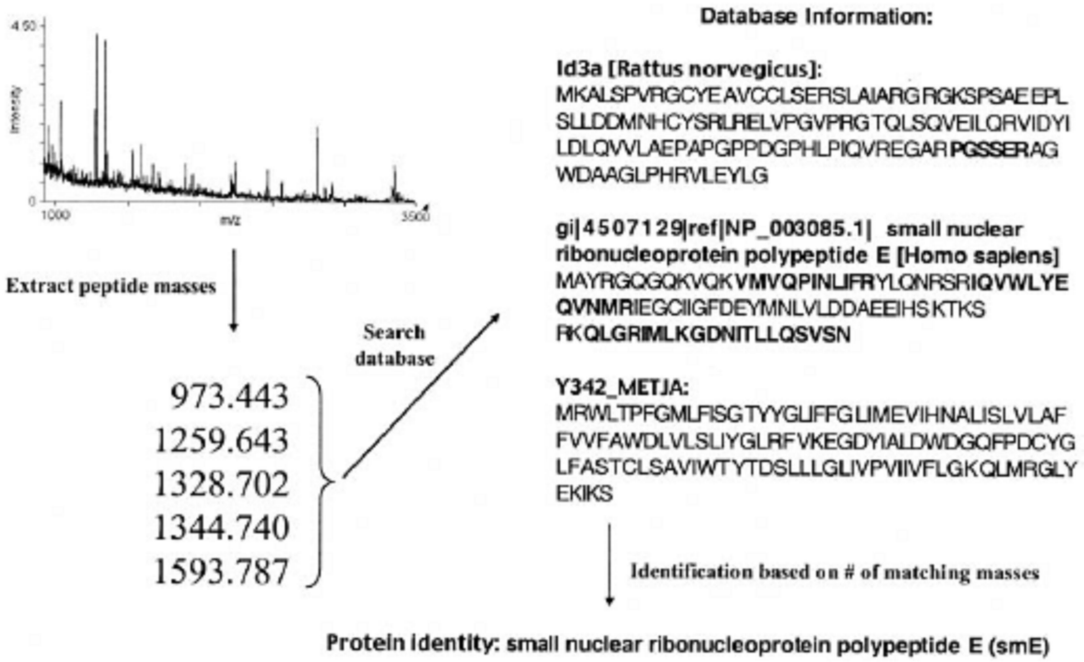


Figure 11. Peptide mass fingerprinting

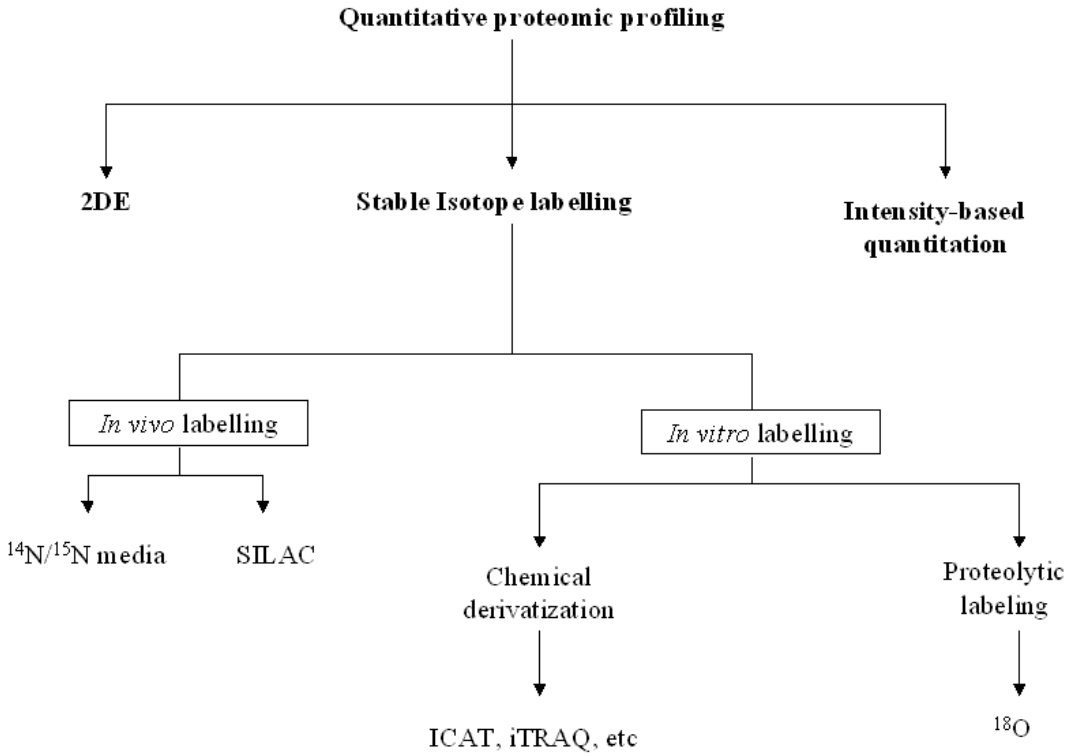
More sophisticated scoring algorithms take the mass accuracy and the percentage of the protein sequence covered into account and attempt to calculate a level of confidence for the match[21, 22, 23]. Other factors can also be included, such as the fact that larger peptides are less frequent in the database and should therefore count more when matched. The accuracy obtained in the measurement of peptide mass strongly influences the specificity of the search[24, 25]**.** When high mass accuracy (10 to 50 ppm) is achieved, as a rule at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification. After a match has been found, a second-pass search is performed to correlate remaining peptides with the database sequence of the match, taking into account possible modifications. Mass fingerprinting can also resolve simple protein mixtures, consisting of several proteins within a roughly comparable amount. For example, databases can be searched iteratively by removing the peptides associated with an unambiguous match[26]. Generally, peptide mass fingerprinting is used for the rapid identification of a single protein component. Protein sequences need to be in the database in substantially full length. Isoforms can be differentiated from each other, if peptides covering the sequence differences appear in the peptide map. If proteins from organisms with fully sequenced genomes can be identified with a 50–90% success rate when at least a few hundred femtomoles of gel-separated protein are present[26].

2

Quantitative Proteomics

2. Quantitative Proteomics

One of the primary goals of proteomics is the description of the composition, dynamics and connections of the multiprotein modules that catalyze a wide range of biological functions in cells. Mass spectrometry (MS) has proven to be an extremely powerful tool for characterizing the composition of purified complexes. However, because MS is not a quantitative technique, the usefulness of the data is limited. For overcoming this limitation various quantitative methods for proteomic studies have been reported[27].



Scheme 2. General scheme of the strategies for quantitative proteomic profiling. a) *In vivo*: 2DE (two-dimensional electrophoresis), SILAC. b) *In vitro*: Chemical derivatization, Proteolytic labeling. c) Intensity-based quantitation (quantitation without stable isotopes)

2.1 Quantitation via 2DE and MS

An older and still commonly used approach to proteomic quantitation is the combination of 2DE and MS. In this procedure, proteins are separated by 2DE and quantified based on the intensity of the protein spots of individual gels. Following proteolysis, the derived peptides are extracted from the gel spots and subjected to MS analysis for identification[28]. A potential problem with 2DE-based quantitation is that the gel spots corresponding between different experiments can be difficult to measure reproducibly. This has been significantly improved by the recent development of a two-colour fluorescence labeling system that allows parallel comparison of two protein samples within the same gel[29]. Nevertheless, global proteomics analysis by 2DE/MS is still limited for complex protein mixtures.

2.2 Quantitation without stable isotope tags

Isotopic labelling of proteins is not always practical and has several disadvantages. For example, labelling with stable isotopes is expensive, and sometimes the isotopic labels exhibit chromatography shifts that can make quantitation of differentially labelled peptides computationally difficult. Moreover, there may not be enough different isotopes to allow for simultaneous quantitation of proteins from multiple samples. As an alternative, several groups have presented methods of peptide and protein quantitation without isotopic tags by comparing peptide signal intensities measured in sequential MS analyses.

For example, signal intensity-based quantitation has been applied to quantify differentially expressed proteins from samples analysed by LC-MS. Pasa-Tolic et al.[30] have presented a system that uses the mass of a peptide coupled with its corresponding chromatographic elution time as peptide properties that uniquely define a peptide sequence, a method termed ‘the accurate mass and time (AMT) tag approach’[30]. Using LC coupled with Fourier transform ion cyclotron resonance (LC-FTICR) MS to obtain the chromatographic and high mass accuracy information, they identified peptide

sequences by matching the AMT tags to previously acquired LC-MS/MS sequence information stored in a database.

By taking advantage of the observed linear correlation between peak area of measured peptides and their abundance[31, 32] these peptides were relatively quantified by the signal intensity ratio of their corresponding peaks compared between MS runs. The primary advantage of this method is that by not selecting masses for CID, as is carried out with traditional LC-MS/MS, one can obtain higher run to run identification reproducibility, consequently allowing for accurate quantitation of more peptides.

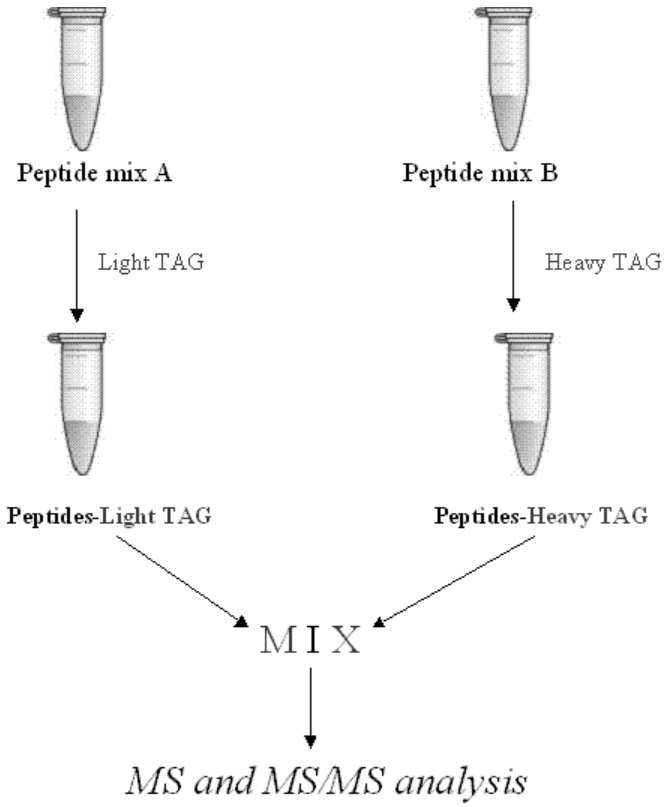
A major disadvantage of peptide quantitation by the signal intensity is that it often includes experimental variation and signal noise, which can affect the quantitative value and accuracy. To circumvent this problem, Wiener and colleagues at Merck Research Laboratories applied statistics tools such as the Student's t-test to analyze data from multiple LC-MS runs (i.e. ten times in their experiment) for each sample[45]. At each point of acquisition time and m/z, the amplitudes of signal intensities from multiple LC-MS runs were compared between two samples to detect peptides with statistically significant differences in abundance between samples. This statistically validated approach of signal intensity comparison, focusing on signal variability between samples, appears to be more sensitive and robust than the traditional approach focusing on absolute signal intensity acquired from background/noise subtraction. The major benefit of this approach is its ability to detect low-abundance peptides that are significantly different between samples but are of such low abundance that they would be ignored by traditional intensity-dependent acquisition. As a consequence, more peptides can be identified and quantified by this approach.

The use of signal intensities for quantitation is of particular interest for cases in which isotopic labelling is impractical or not feasible.

2.3 Stable Isotope Labelling

Recently, stable-isotope labelling was introduced into the field of proteomics as a means to measure global differences in protein expression between proteins from different cellular conditions. This exciting the use of stable-isotope labelling provided internal standards for the relative quantitation of multiple proteins simultaneously and

thus allowed a global-type experiment to be performed for measuring protein expression in parallel. Gygi et al.[34] introduced the new term isotope-coded affinity tag (ICAT) and incorporated heavy isotopes *in vitro*. In the ICAT approach the Cys-containing peptides from two samples (e.g. proteins expressed by a cell under two different physiological conditions) are differentially labelled with light (isotopic normal regent containing hydrogen atoms on the carbon backbone) and heavy (where the hydrogen atoms have replaced with deuterium atoms) using a cleavable, biotinylated reagent, and these peptides then affinity purified using immobilized avidin as a result the selectively isolation from non-Cys-containing peptides. The identification was made my LC-ESI-MS.



Scheme 3. General workflow of quantitative proteomic strategies using stable isotopes

Since then several strategies have been reported (Figure 12, Table 1) that target also the N-terminal group of peptides and the side chain amino group of Lysines and Arginines. The most important is the isobaric tag for relative and absolute quantification (iTRAQ)[35]. iTRAQ used NHS (N-hydroxy-succinimide) active esters as a regent and adds an innovative concept, namely a tag that generates a specific reporter ion for

quantification in MS/MS spectra (mass 114,115,116,117) but with isobaric mass at MS level. Therefore, mass spectra are relatively simple and differential behavior is only reported after fragmentation. Moreover, multiplexing (currently 8-plex) is an interesting feature as it allows comparing more than two conditions.

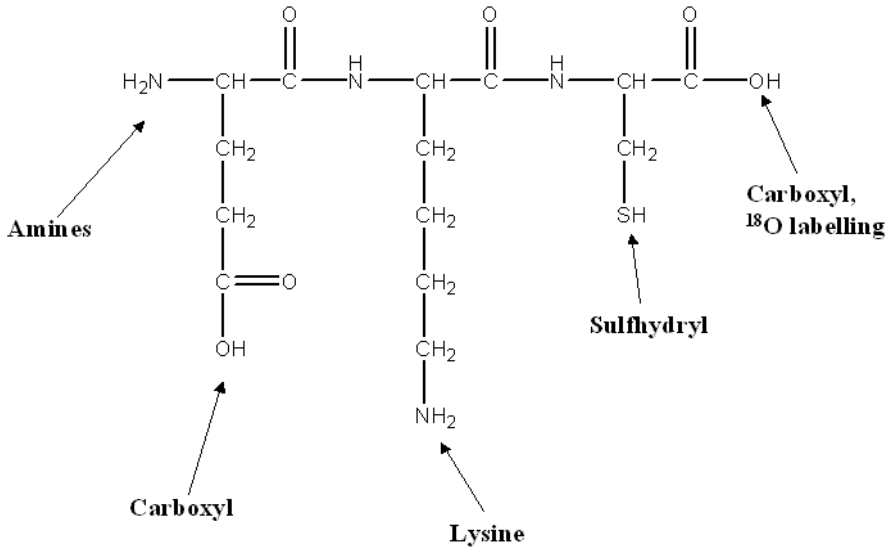


Figure 12. Generic summary of coding strategies currently employed to exploit various reactive centers in peptides

2.3.1 O^{18} Labelling

The tryptic digestion of peptides is a hydrolysis reaction, requiring the incorporation of water molecules. The use of O^{18} labelled water in the digestion buffer will result in the incorporation of a single O^{18} labelled carboxyl group at the carboxyl end of each tryptic peptide[70, 71]. Thus a control sample may be digested with normal water and a test sample may be digested in the presence of O^{18} labelled water, resulting in peptides differing in mass by 2 Da. These may be quantified in the same manner as ICAT peptides. The main drawback to this type of quantitation is that the labelled peptides will occur at the same point as +2 and subsequent isotopic peaks of the unlabelled peptide. This significantly limits the dynamic range of the technique, although attempts have been made to improve quantitation through the use of isotope matching software[56].

2.3.2 Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC)

The technique of SILAC is based on the addition of isotopically labelled amino acids to culture medium. Cells are grown in medium containing ^{13}C and/or ^{15}N labelled amino acids (commonly lysine or arginine). Carbon starvation is sometimes employed to assist uptake of labelled amino acids. These labelled amino acids are taken up and incorporated into cellular proteins. Once cells are lysed, the lysates mixed together, and analysed by MS, it is possible to observe the relative abundances of a given labelled and unlabelled peptide or protein pair in terms of the spectral abundance of their individual mass peaks. The principle drawback of SILAC is that the technique works only with cells that can be grown on medium containing the labelled amino acid and that are auxotrophic for the relevant amino acid. The most commonly used amino acids for SILAC labelling are lysine and arginine, based on the principle that after a tryptic digestion every peptide (excepting the C terminal peptide) can be used for quantitation. Other amino acids suitable for SILAC labelling include leucine (which was used in the original report of the technique), and labelled methionine[60] and tyrosine are commercially available. Lysine is commonly used in $^{12}C6$ and $^{13}C6$ versions for a mass shift of 6, but can be supplemented by $^{14}N2$ and $^{15}N2$ labels for a total mass shift of 8. Triplet labels can be implemented with the use of labelled arginine, as the 6 and 10Da mass shifts between 12C6 14N4 arginine, 13C6 14N4 arginine and $^{13}C6$ $^{15}N4$ arginine can be clearly visualised on most MS instruments.

Target	Name of method or reagent	Isotopes	References
Amines	Tandem mass tag (TMT)	D	36
	Succinic anhydride	D	37
	N-acetoxysuccinamide	D	38
	Acetic anhydride	D	39
	Propionic anhydride	D	40
	Nicotinoyloxy succinimide (Nic-NHS)	D	41
	Phenyl isocyanate	D or ^{13}C	42
	Suldo-NHS biotin and ^{13}C, CD_3I	D and ^{13}C	43
	Formaldehyde	D	44
	Isobaric tag for relative and absolyte quantification (iTRAQ)	^{13}C, ^{15}N, ^{18}O	35
	Benzoic acid lebeling (BA of ANIBAL)	^{13}C	45
Lysines	Guanidination (O-methyo-isourea)	^{13}C, ^{15}N	46,47,48
	Quantitation using enhanced sequence tags (QUEST)	No isotope	49
	2-Methoxy-4,5-1H-imidazole	D	50
Sulfhydryl	Isotope-coded affinity tagging (ICAT)	D	34
	Cleavable ICAT	^{13}C	51,52,53
	Catch and release (CAR)	^{13}C	54
	Acrylamide	D	55
	Isotope-coded reduction off of a chromatographic support (ICROC)	D	56
	2-vinyl-pyridine	D	57
	N-t-butyliodoacetamide	D	58
	Iodoacetanilide	D	58
	HysTag	D	59,60
	Visible isotope-coded affinity tags (VICAT)	^{13}C, ^{14}C, ^{15}N	61
Carboxyl	Methyl esterification	D	67
	Ethyl esterification	D	68
	C-terminal isotope-coded tagging using sulfanilic acid (SA)	^{13}C	69
	Aniline labeling (ANI of ANIBAL)	^{13}C	67
	Proteolytic ^{18}O labeling ($H_2{}^{18}O$)	^{18}O	70,71,72,73
	Quantitative cysteinyl-peptide enrichment technology (QCET)	^{18}O	74
Solid Phase	Acid-labile isotope-coded extractants (ALICE)	D	62
	Solid phase mass tagging	^{13}C	63
	Solid-phase ICAT	D	64
	A novel class of iodo containing resins	D	65
	Design and synthesis of a solid-phase fluorescent mass tag	^{13}C, ^{15}N	66

Table 1. Summary of currently strategies for quantitative proteomics (with refs)

2.4 Solid-Phase Chemical tools for proteomics

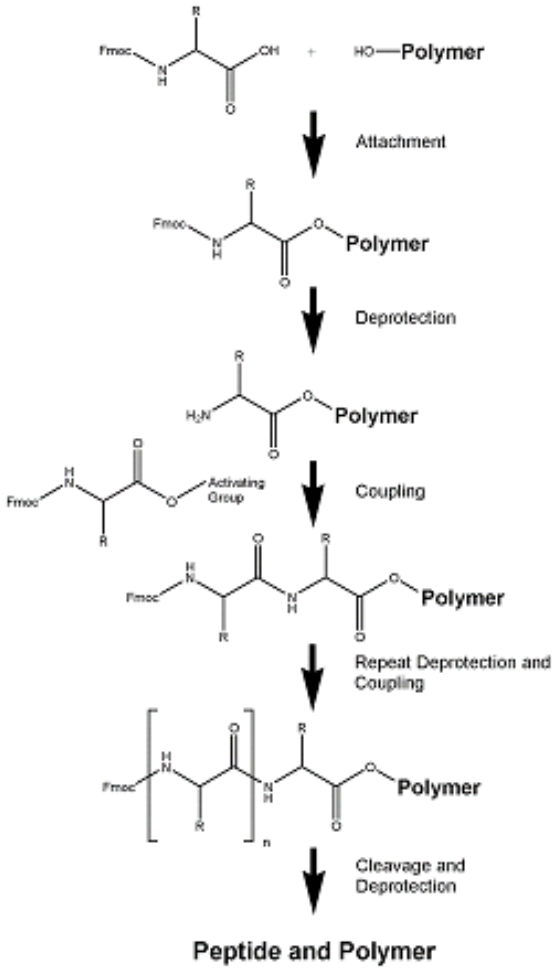
Techniques involving solid supports have played crucial roles in the development of synthetic peptides & proteins, genomics, proteomics and the molecular biology in general. Similarly, methods for immobilization or attachment to surfaces and resins have become ubiquitous in sequencing, synthesis, analysis and screening of oligonucleotides, peptides and proteins.

2.4.1 Solid Phase Synthesis

Solid-phase synthesis is a method in which molecules are bound on a bead (a resin) and synthesized step-by-step in a reactant solution. Compared with traditional liquid phase synthetic techniques, it is easier to remove excess reactant or byproduct from the product. In this method, building blocks are protected at all reactive functional groups. The two functional groups that are able to participate in the desired reaction between building blocks in the solution and on the bead can be controlled by the order of deprotection. This method is used for the synthesis of peptides, deoxyribonucleic acid (DNA), and other molecules that need to be synthesized in a certain alignment. Recently, this method has also been used in combinatorial chemistry.

Solid-phase synthesis is the most common method for peptide synthesis. Usually, peptides are synthesized from the carbonyl group side to amino group side of the amino acid chain in this method, although peptides are synthesized in the opposite direction in cells. In peptide synthesis, an amino-protected amino acid is bound to a bead (a resin), forming a covalent bond between the carbonyl group and the resin. Then the amino group is deprotected and reacted with the carbonyl group of the next amino-protected amino acid. The bead now bears two amino acids. This cycle is repeated to form the desired peptide chain. After all reactions are complete, the synthesised peptide is cleaved from the bead.

The protecting groups for the amino groups mostly used in this peptide synthesis are 9-fluorenylmethyloxycarbonyl group (Fmoc) and t-butyloxycarbonyl (Boc). The Fmoc group is removed from the amino terminus with basic conditions while the Boc group is removed with acid conditions.



Scheme 4. General scheme of Fmoc strategy for peptide synthesis using solid phase supports

2.4.2 Solid phase support (resin)

Small solid beads, insoluble yet porous, are treated with functional units ('linkers') on which peptide chains can be built. The peptide will remain covalently attached to the bead until cleaved from it by a reagent such as trifluoroacetic acid. The peptide is thus 'immobilized' on the solid-phase and can be retained during a filtration process, whereas liquid-phase reagents and by-products of synthesis are flushed away. The physical properties of the solid support, and the applications to which it can be utilized, vary with the material from which the support is constructed, the amount of crosslinking, as well as the linker and handle being used.

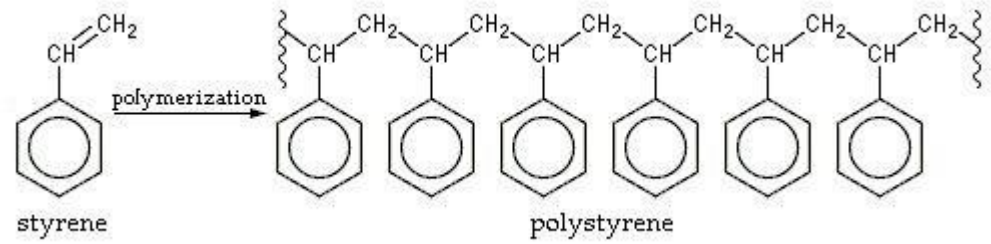


Figure 13. Polystyrene synthesis

Merrifield[75] firstly introduced solid phase strategies. His resin (Figure 14, 1) is a polystyrene resin based on a copolymer of styrene and chloromethylstyrene. In addition this polymer is also cross-linked with divinylbenzene present in the monomer composition up to 5%. Merrifield resin is named after its inventor. The cleavage of the products from the resin can be done by treatment of the resins with strong acid, such as HF or TFMSA, or by hydrogenolysis.

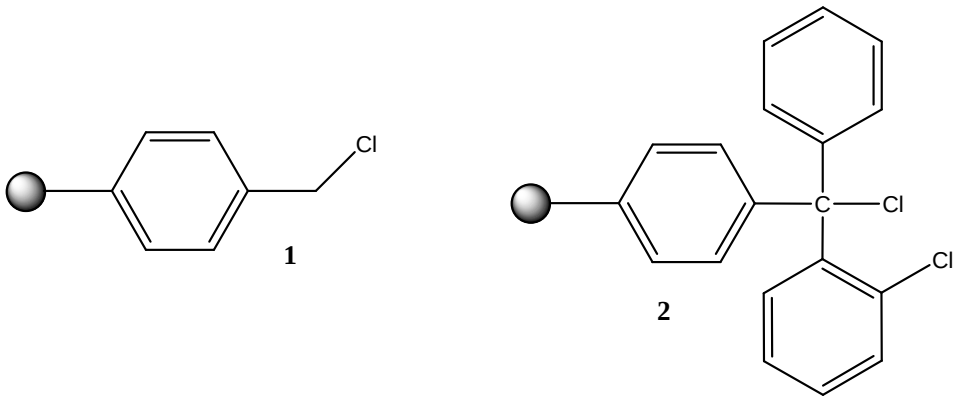


Figure 14. Common polystyrene resins for Solid Phase Synthesis

The handling of these acids is the main disadvantage of Merrifield's resin. Barlos introduced[76, 77] an easy-cleavable resin (Figure 14, 2), which is a copolymer of styrene and 2-Chlorotrityl. The cleavage can be done using mild conditions (0.5% TFA). These types of resins are used for automated peptide synthesizers.

2.4.3 TentaGel resins

Tentagel resins are grafted copolymers consisting of a low crosslinked polystyrene matrix on which polyethylene glycol (PEG or POE) is grafted. As PEG is a "cameleon type" polymer with hydrophobic and hydrophilic properties, the graft copolymer shows modified physico chemical properties. There are in principle two ways to introduce PEG onto the modified polystyrene matrix. The simplest immobilization procedure is to couple PEG via one of its terminal hydroxyl groups to chloromethylated polystyrene according to the classical ether synthesis or to use other bifunctional PEG's for coupling onto the solid support. The hydrophilic properties of TentaGel resins make them suitable for Proteomics applications because of the aqueous solutions that are used in the proteins and peptides sample preparation.

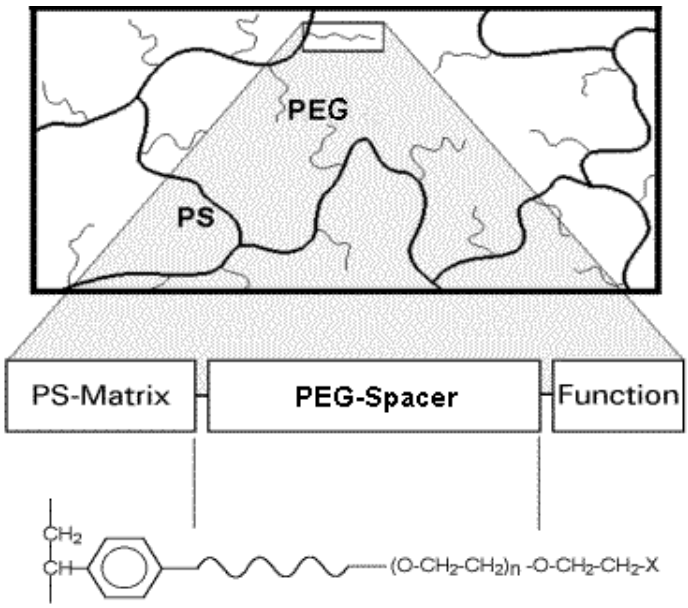


Figure 15. TentaGel Resins

2.5 Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry

Recently, solid-phase stable-isotope labelling methods emerged. Zhou et al.[64] (Figure 16) reported the photo cleavable iodo-containing glass beads method. Briefly, Cysteinyl peptides from two samples were covalently captured on the solid phase containing isotopically heavy or normal (in the leucine residue, d_0/d_7) tag. The beads were combined, washed and exposed to UV light (360nm). This resulted the photocleavage of the linker and the transfer of isotope tags from the solid phase onto the side chain of cysteine residues. Finally, recovered tagged peptides were analyzed by LC-MS/MS to determine the sequence and relative abundance of each peptide.

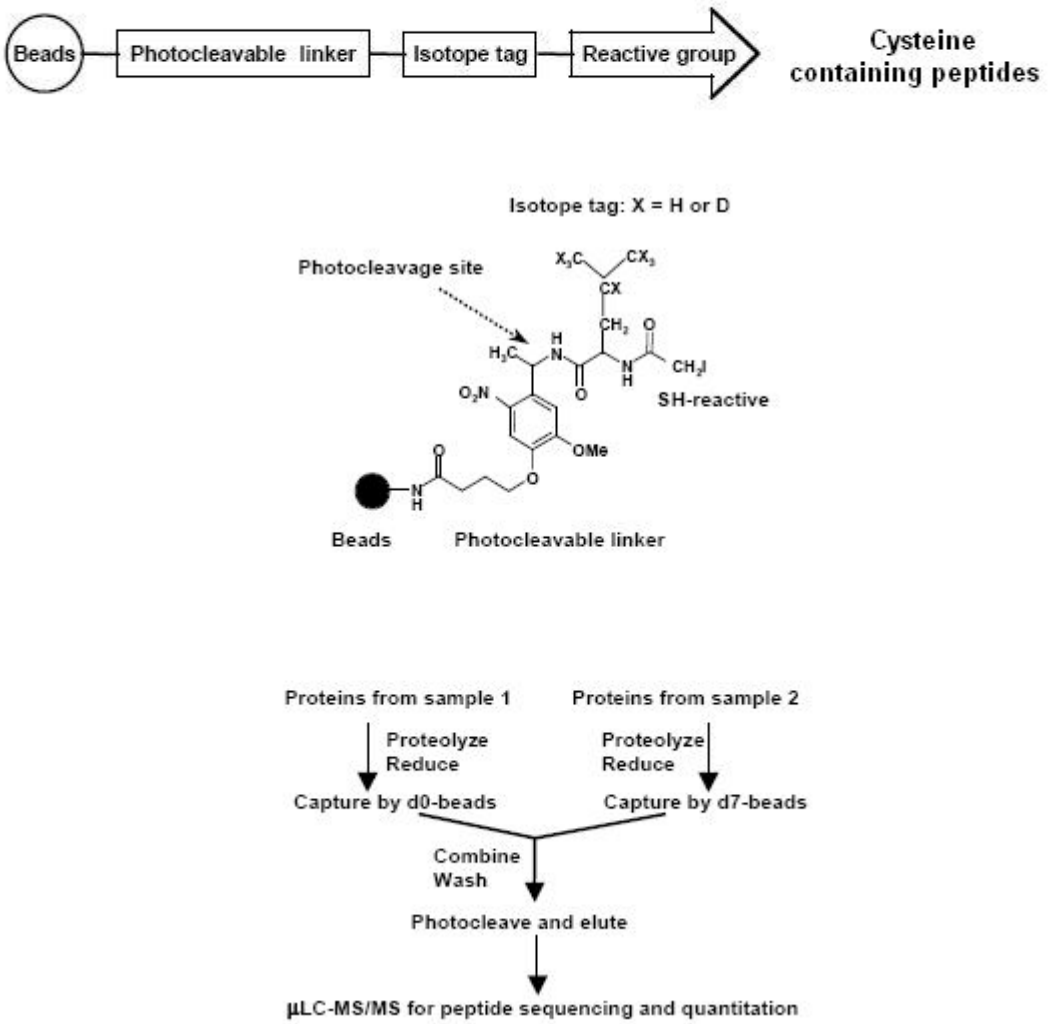


Figure 16. Schematic representation of the solid phase isotope tagging method.[64]

Two years later, Li Zhang et al.[65] (Figure 17) reported an iodo-containing resin with a thiol-reactive group that is used to capture the Cys-containg peptides from peptide mixtures using an aminomethyl polystyrene resin with Ring Amide Linker. The "tag" amino acid was designed to be leucine (d_0/d_{10}). The Cys-containing parts were cleaved from the resin with trifluoroacetic acid (TFA)-ethanedithiol(EDT)-thioanisole-phenol-water (81,5:2,5:5:5:6) followed by ì LC-MS analysis.

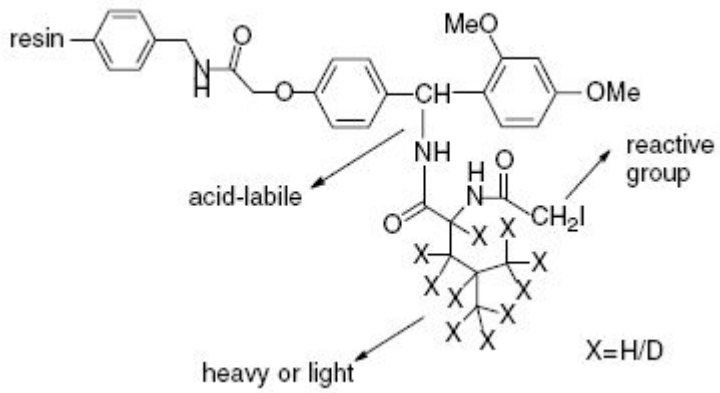


Figure 17. Structure of the iodo-containing resins H/D of Li Zhang et al.[65]

The same period Yang Shi et al.[63] (Figure 18) reported another iodo-containing resin using a polymethacrylate/PEG resin with a Ring Amide Linker. The mass tag was a trialanine peptide containing ^{12}C or ^{13}C producing a 9 mass unit difference. The Cys-containing peptides were cleaved with a 50% TFA in DCM solution. The analysis was made using MALDI-TOF and LC/NamoESI-MS mass spectrometry.

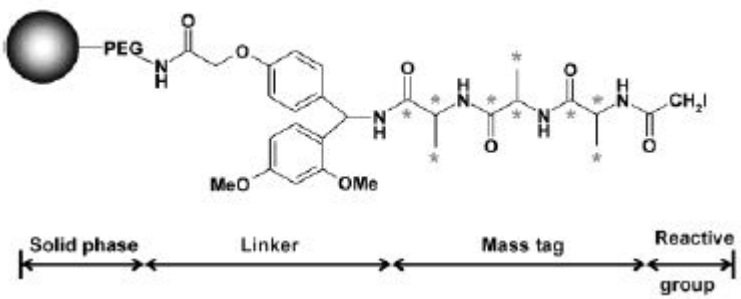


Figure 18. Structure of the iodo-containing resins ^{12}C or ^{13}C of Yang Shi et al.[63]

The same group reported[66] (Figure 19) a solid phase fluorescent iodo-containing mass tag. The fluorescent molecule (7-methoxycoumarin-3-carboxylic acid) was attached to the ε amino group of a lysine (^{12}C,^{14}N / ^{13}C,^{15}N) and the product was attached to a polymethacrylate/PEG resin with a Ring Amide Linker. The thiol-reactive group was coupled in the α amino group using succinimidyl iodoacetate. The Cys-containing peptides were cleaved with TFA-ACN-H_20 (95:2,5:2,5) and analyzed MALDI-TOF and RP HPLC (LIF detector).

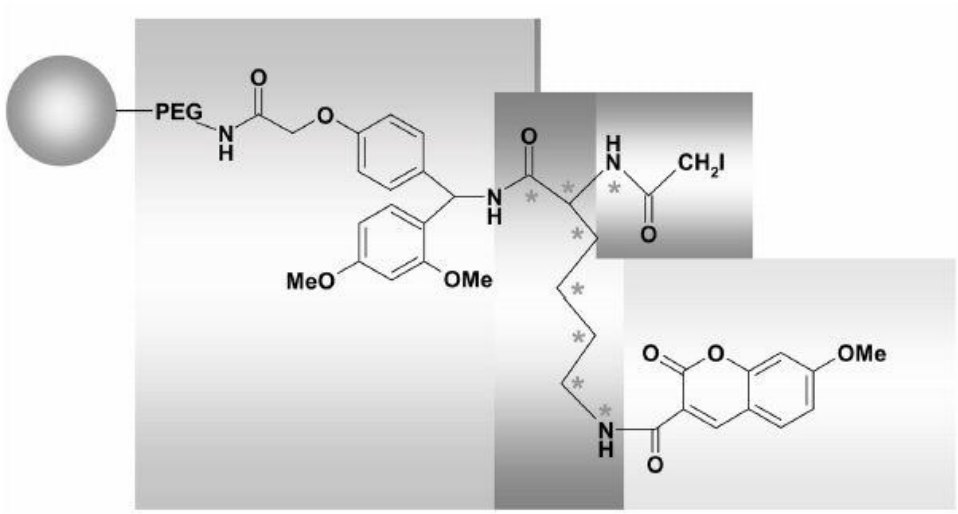


Figure 19. Schematic diagram of the solid phase fluorescent lysine (^{12}C,^{14}N / ^{13}C,^{15}N) mass tag by Yang Shi et al.[66]

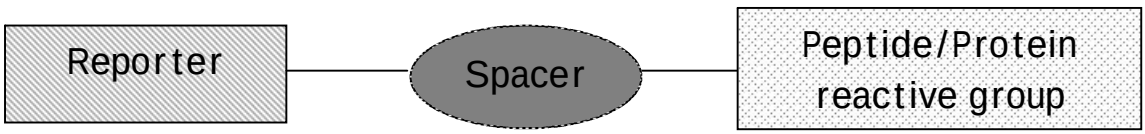
3

Results and Discussion

3. Results and discussion

Proteomics, the analysis of the protein complement of a cell or an organism has grown rapidly as a sub discipline of the life sciences. The quantitative information is a necessity for understanding the protein world. Quantitative proteomics involves the identification and quantitation of protein components in various biological systems. Stable isotope labelling technology, by chemical derivatization, has been the most commonly used approach for global proteome profiling. The ion intensity-based quantitative approach has progressively gained more popularity as mass spectrometry performance has improved significantly.

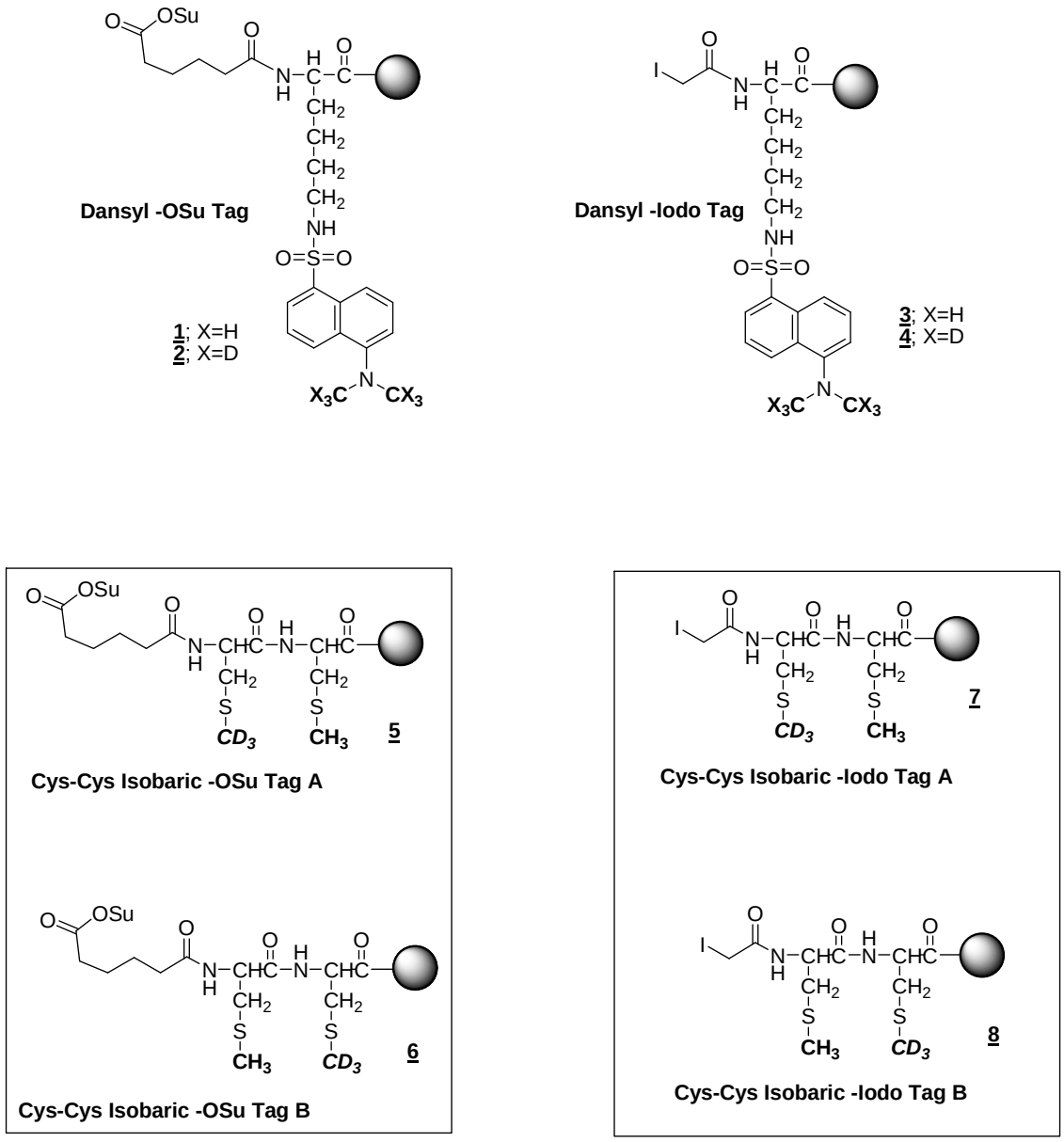
Our approach to the quantitative proteomics field was to develop various types of solid phase-based reagents that could be easily synthesized using commonly available techniques and reagents at reasonable cost. The reagents are trifunctional molecules composed of a "reporter", a "spacer" group and a specific reactive group capable of reacting easily with the N-terminal group of the peptides or the Thiol side chain group of Cysteins.



Scheme 5. General structure of reagents needed for quantitative proteomics

The area of solid phase peptide synthesis provides a rich source of options for creation of such reagents. In this study, I describe the synthesis of mass tags based on the d_0, d_6 Dansyl group (Scheme 6) compounds as well as the synthesis of Isobaric mass tags based on the d0, d3 S-Methylated thiol group of Cysteine (Scheme 6). Two different peptide/protein reactive groups have been tried on their N-terminal. Thiol-reactive group can catch the cysteine containing peptides while the succinimidyl tag catches the amine groups. The C-terminal is linked with an amide bond through an acid-labile handle to a TentaGel resin. After cleavage, these reagents provides a 6 (Dansyl) or 3 (Cys-Cys Isobaric) mass unit difference between heavy and light labelled peptides

that can be used to accurately measure the ratios of proteins in mixtures using MALDI TOF , TOF/TOF or ESI-QTOF techniques.



Scheme 6. Solid Phase isotope-label Tags

3.1 Mass spectrometry equipment

MALDI-TOF experiments were performed using a 4700 Proteomics Analyzer mass spectrometer, from Applied Biosystems (Foster City, CA), equipped with a 200-Hz Nd:YAG laser at 355- nm wavelength. The MS spectra were acquired in reflectron mode (20-keV accelerating voltage), with 400-ns delayed extraction, averaging 2000 laser shots with a mass accuracy of 20 ppm. MS/MS experiments were performed at a collision energy of 1-2 kV, defined by the potential difference between the source acceleration voltage (8 kV) and the floating collision cell (7-6 kV); 3000 laser shots were averaged, while the pressure inside the collision cell was 8 X 10-7 Torr. Linear MALDI MS spectra were acquired averaging 2500 laser shots with a mass accuracy of 500 ppm in default calibration mode that was performed using the following set of standards: insulin (bovine, $[M + H]^{+}_{avg}$) 5734.59 Da), apomyoglobin (horse, $[M + H]^{2+}_{avg}$) 8476.78 Da, $[M + H]^{+}_{avg}$) 16952.56 Da), and thioredoxin (*Escherichia coli*, $[M+ H]^{+}_{avg}$) 11674.48 Da).

For each MALDI analysis, a 1 ìL portion of a premixed solution of each sample and α-CHCA (0.3% in TFA) (analyte/matrix 1:5) was spotted on the matrix target, dried at room temperature and directly analyzed by MALDI mass spectrometry.

Proteins were identified by searching a comprehensive protein database using Mascot programs (www.matrixscience.com). Peak harvesting was done automatically using Data Explorer software (Applied Biosystems). Peak resolution was calculated using the Data Explorer software.

High-resolution ESI experiments were carried out in a hybrid Q-Star Pulsar-i (MDS Sciex Applied Biosystems, Toronto, Canada) mass spectrometer equipped with an ionspray ionization source. Samples were introduced by direct infusion (5ì L/min) of the solution at the optimum ion spray voltage of 4800 V. The nitrogen gas pressure was set at 30 psi and the declustering and the focusing potentials were kept at 70 and 140V relative to ground, respectively. MS/MS experiments were performed in the collision cell on the isotopically pure (^{12}C) peak of the selected precursor ions by keeping the first

quadrupole analyzer at 20V relative to ground and operating at unit resolution, and scanning the time-of-flight (TOF) analyzer. The collision energy was set from 15 to 35 eV, depending on the compound, while the gas pressure of the collision chamber was set at the instrumental parameters CAD 5. All the acquisitions were averaged over 60 scans at a TOF resolving power of 8000. The software used was AnalystTM QS software (MDS Sciex).

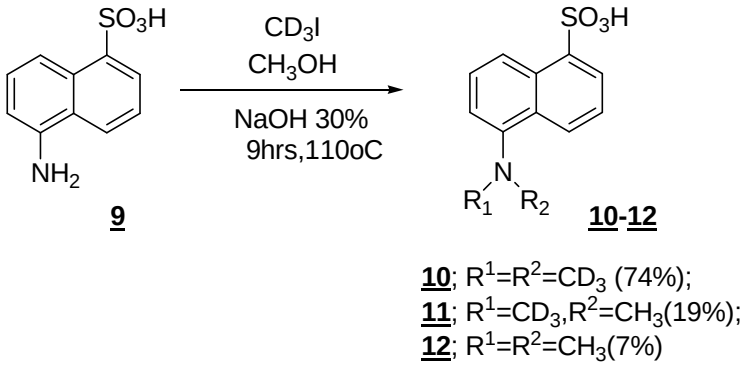
3.2 Dansyl Solid Phase Tag

3.2.1 Synthesis of d6-Dansyl Chloride

Dansyl chloride (*Dns-Cl*) is a widespread used reagent in biochemistry for its fluorescent properties. It was introduced several years ago in peptide chemistry[78] as a molecular device for driving the gas-phase sequencing of peptides by electron ionization mass spectrometry (EIMS). An increase in BSA protein sequence coverage[79] was quite recently reported when Dns-derivatized tryptic peptides were analyzed by matrix-assisted laser desorption ionization (MALDI), whereas a selective detection and identification of phosphopetides was achieved by electrospray ionization (ESI) of Dns derivatized peptide digests[80]. The rediscovering of Dansyl peptide chemistry has prompted the development of new deuterium labelled tags to be used in modern proteomic applications for identification and assay of protein digests[81].

The most accessible way to synthesize the deuterium labelled Dansyl derivative was thought to be the alkylation of 5-aminonaphthalene-1-sulfonic acid (**9**, Scheme 7) using a reliable very old method[82]. Our first attempt to prepare compound **10** was represented by the bis *N*-methylation of **9** replacing methyl iodide with its trideutero analogue. Thus, **9** (2.4mmol), d_3-methyliodide (5.4mmol), NaOH (0.6mL 30% aq.) in 2mL of methanol were stirred in a schlenk flask at 110^{o}C, using the same experimental conditions reported[82] for the synthesis of the unlabelled analogue **12**. After 6hrs a fresh portion of d_3-methyliodide was added and the reaction mixture kept at the same experimental conditions for further 3hrs. The crude reaction mixture was then treated with HCl 6N

solution and the resulting gray crystals (m.p.>300°C, 85% yield) were collected by filtration.



9 **10-12**

10; $R^1=R^2=CD_3$ (74%);
11; $R^1=CD_3, R^2=CH_3$(19%);
12; $R^1=R^2=CH_3$(7%)

Scheme 7. Trideuteromethyl iodide mediated alkylation[82] of **9**

The structure of the compounds was easily ascertained by ESIMS. The rationale for the observed behavior was considered to be the establishment of an equilibrium between the reagents, which under the adopted strong basic conditions, promoted the formation of methyl iodide *via* an intermediate methoxyde species produced from methanol. This hypothesis is confirmed by the observation that the analysis of the reaction mixture by ESIMS did not show any presence of partially labelled molecules due to an unlikely isotope isomerization (Figure 20).

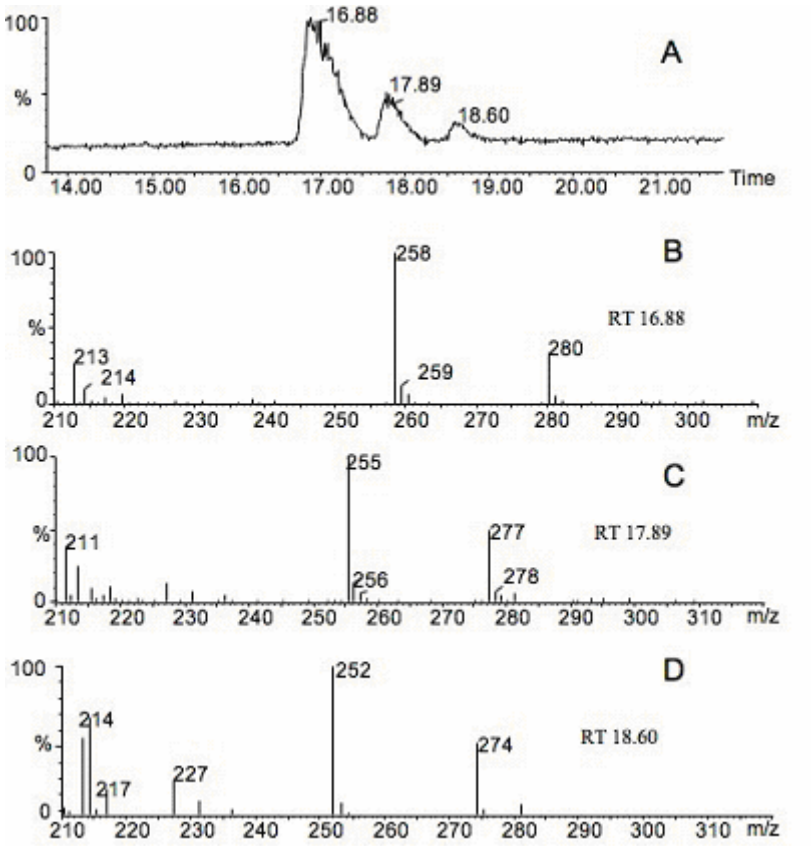
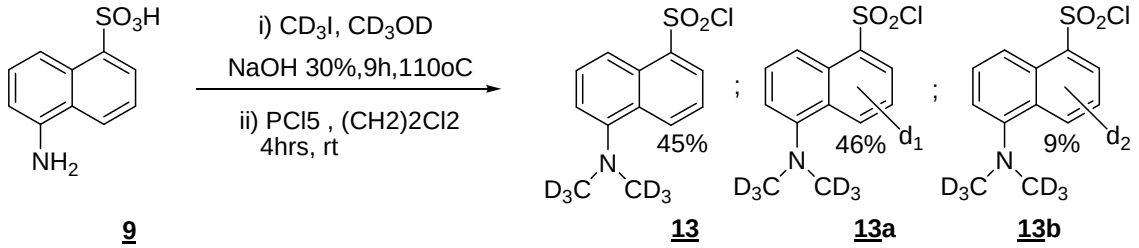


Figure 20. (A) ion chromatogram of the specie **10** (16.88), **11** (17.89) and **12** (18.60) obtained in the preparation reported in scheme 1; (B), (C) and (D) ESIMS spectra of **10**, **11** and **12**, respectively

Accordingly, the reaction was repeated under the same experimental conditions by using tetradeuteromethanol as solvent (Scheme 8). This time only labelled Dansyl derivatives where present in the final crude product. ESIMS analysis of the latter, yielded in the $[M+H]^+$ region a complex ion cluster (*m/z* range 276?282) which was clearly due to protonated species formed by the expected Dns derivative **13**, overlapping with other protonated molecular ions formed by its d_7 and d_8 isotopic analogues. The deconvolution of the cluster indicates that the labelled Dansyl derivatives **13**, **13a** and **13b** were approximately formed with 45, 46 and 9% yields, respectively.



9 **13** **13a** **13b**

Scheme 8. Alkylation by trideuteromethyl iodide in fully deuturated methanol

The results strongly suggest that the classic thermal approach to the alkylation of the amino moiety of **9** does not correspond, at least in this particular case, to a straightforward SN_2 displacement reaction. Different mechanisms can be envisaged for the formation of the d_7 and d_8 analogues[83] which should bear the extra deuterium on both aromatic rings, as shown by 500 MHz NMR spectrum (Figure 21) of the above mixture, which displayed the reduction of the intensity of the aromatic H_2 and H_6 proton signals centered at 8.92 ppm and 7.76 ppm respectively. Nevertheless it has to be considered that all the problems encountered when deuterium labelled reagents were used, are not present in the synthesis of the unlabelled Dansyl sulphonic acid **12** which, as described in the early report[82] provides gray crystals directly from the reaction mixture.

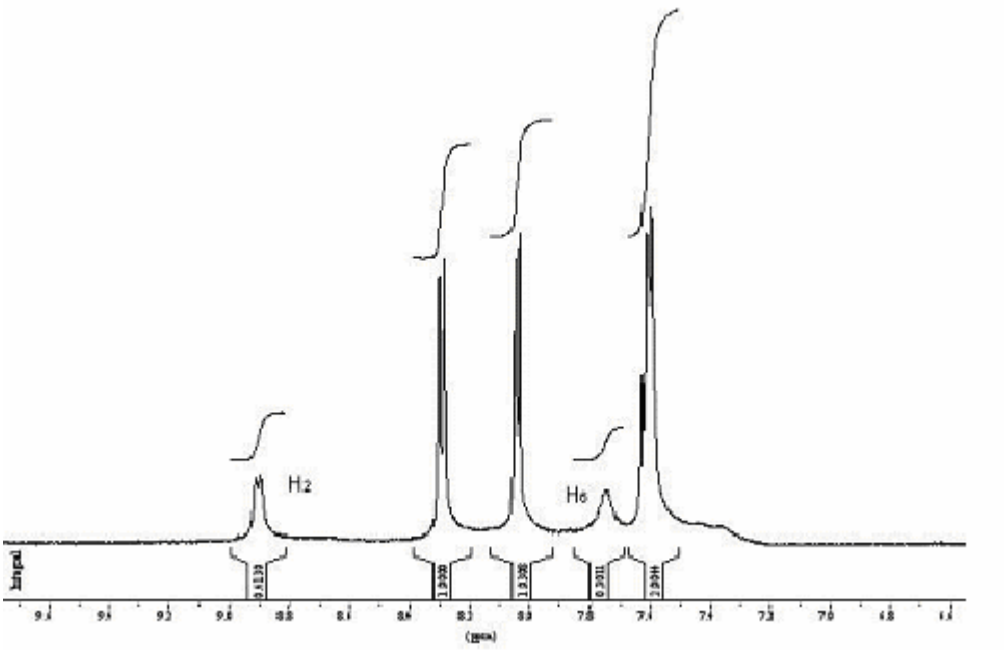
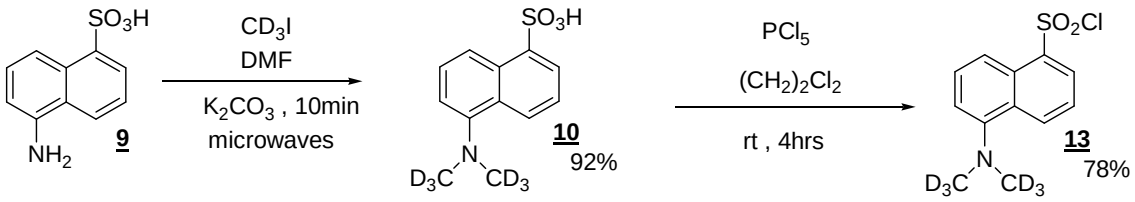


Figure 21. Reduction of the intensity of aromatic proton signals at 8.92 and 7.76 ppm (proton ratio 0.41:1:1:0.30:2 instead of the expected 1:1:1:1:2).

Microwave assisted alkylation of amines has been recently reported[84]. It is well recognized that very smooth conditions are often needed when these type of procedures are chosen, so we thought of interest to examine the possibility to synthesize the desired d_6-labelled sulphonic acid **10** using microwave chemistry (Scheme 9).



Scheme 9. Novel synthesis of d_6-Dns-Cl using MAOS.

Accordingly a mixture of compound **9** and 2.5 equiv. excess of each K_2CO_3 and d_3-methyliodide in DMF were placed in the sealed reaction vessel of a Synthos 3000 (Anton Paar GmbH) multimode microwave synthesizer and stirred at 110 ^oC (direct

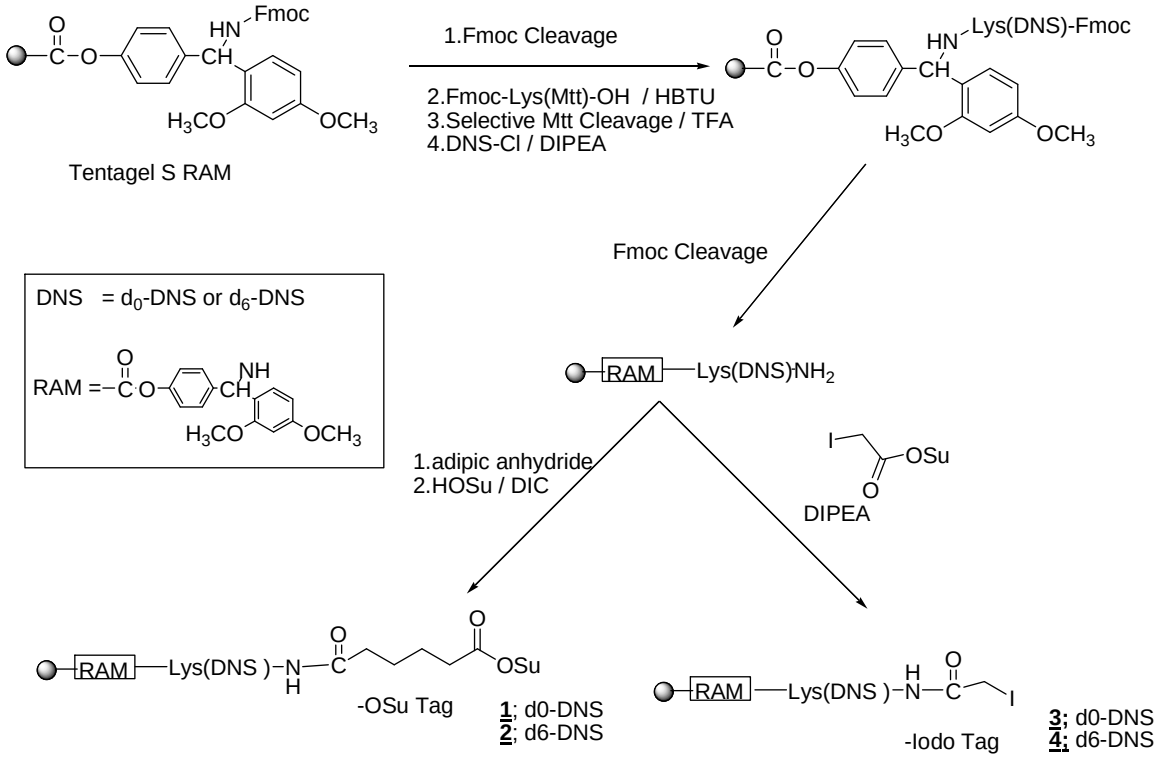
temperature control, T-probe sensor) for 10 min (3 min ramp , 7 min hold at 110^{o}C) under microwave irradiation. After acidification with HCl 6N, followed by filtration of the precipitated gray crystals, we obtained in 92% yield the d_6-Dansyl sulphonic acid **9**. The latter was subsequently chlorinated using PCl_5 in dry 1,2-dichloroethane for 4hrs at r.t., followed by aqueous work-up and evaporation of the solvents to afford the expected 5-*N*,*N*-trideutero methylaminonaphthalene-1-sulphonyl chloride **13** in 78% yield as yellow crystals.

3.2.2 Development of Dansyl-tag containing resins

Dansyl-tag was loaded in the resin in the way described in Scheme 10. Detailed synthetic procedures are available in the Experimental Part. The resin used for our synthesis was Tentagel S RAM (0,24 mmol/g) bought from Rapp Polymere (Germany). Briefly, we used 0,074mmol of resin placed in a 5mL syringe equipped with a frit. Then followed Fmoc cleavage using a solution of 25% piperidine in DMF , in 3 cycles for 20 min each. Attachment of the "tag" was performed through amide bond formation using 2-fold excess of Fmoc-Lys(Mtt)-OH in DMF and HBTU as the coupling reagent. Then Mtt group was cleaved using TFA 10% in DCM for 1h and the ε-amino group of Lys was dansylated after treatment with Dns-Cl (2-fold excess) (or d6-Dns-Cl, **13**) and DIPEA (2,5-fold excess) in DCM for 2h. A second cycle followed by adding a fresh portion of the same reagents and let the mixture to react for additional 2h. After Lysine Fmoc protecting group cleaved with 25% piperidine in DMF the resulted resin was spitted in two new syringes.

The first part was derivatized using 2-iodoacetic N-Hydroxysuccinimidyl ester (4-fold excess) and DIPEA (4-fold excess) in DMF for 30min. (-Iodo Tag , Scheme 10)

The second part was derivatized using adipic anhydride (4-fold excess) and DIPEA (4-fold excess) in DMF for 1h. Finally the resin was treated with N-hydroxy succinimide (4-fold excess) carried out for 5h using DIC (4-fold excess) in DMF. (-OSu Tag , Scheme 10)



Scheme 10. Synthesis of Dansyl Solid Phase Tag

3.2.3 Reactivity test of Dansyl –OSu Tags

To test the reactivity of the Amine Group Selective Tag (-OSu , **1** and **2**) towards amino groups in *N*-terminal and/or side chain position, a mixture of amino acids plus a small tripeptide were selected (Figure 22).

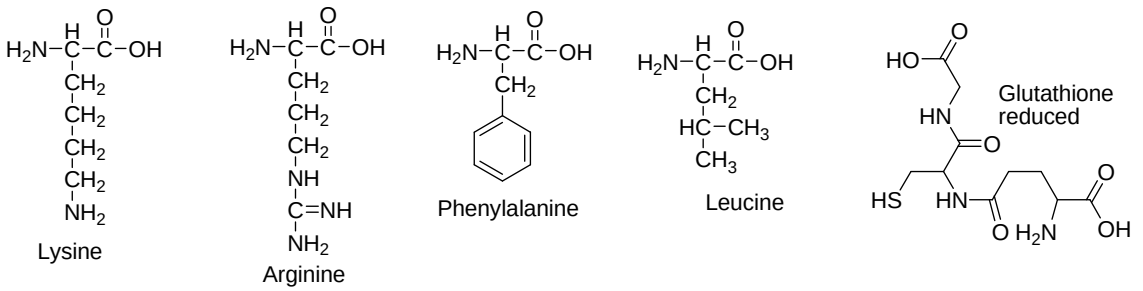


Figure 22. Mixture of AA plus Glutathione used for the reactivity test of Dansyl –OSu (**1, 2**) Tags (**Standard Mixture**)

Our first attempt was to perform derivatization reactions in pure organic environment. Thus, d_0 / d_6 resins (**1**, **2**) and standard mixture were incubated overnight in DMF at pH=10 using DIPEA as the base. After washing the resin (2 x DMF , 2 x ACN / H_20 6:4 , 2 x DMF, 1 x DCM) the tagged products were cleaved from the resin by treatment with 30% TFA in DCM for 1h. All our products were peptide amides because of the Ring Amide Tentagel resin and expected to provide a mass shifting of 488 Da (light resin **1**) or 494 Da (heavy resin **2**). The increment of the molecular weight of the small peptides is crucial for analysis with MALDI because it shifts their peak to the zone above 700 Da which is clear from matrix clusters.

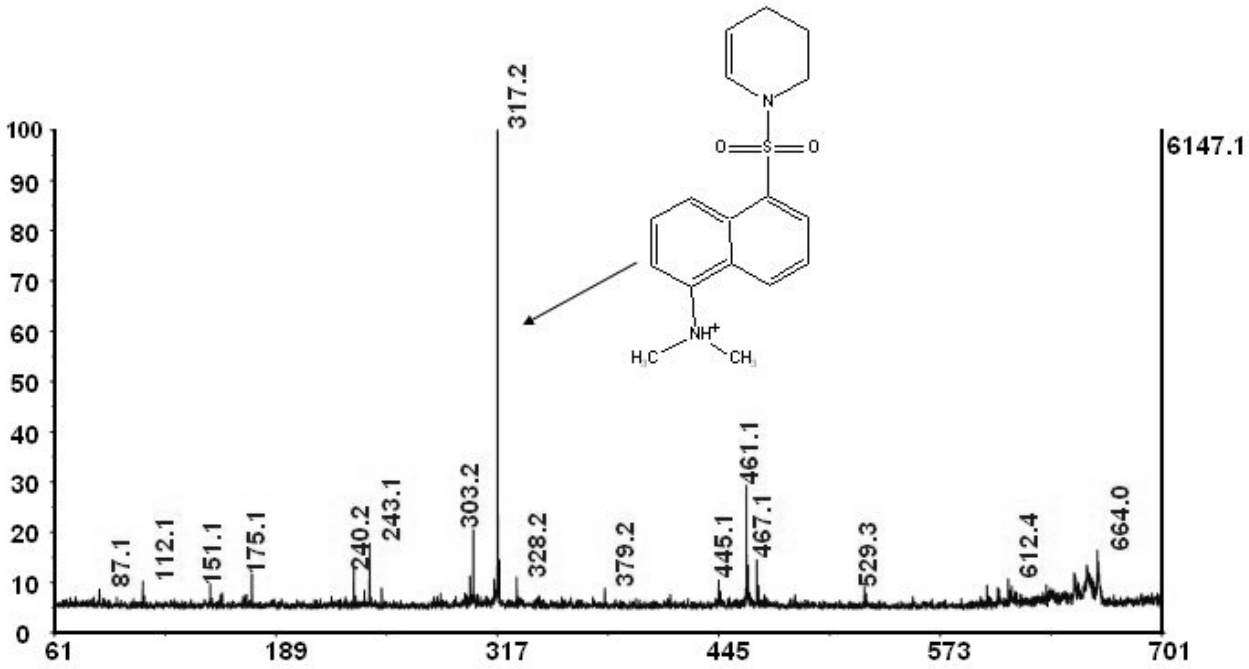


Figure 23. MALDI MS/MS of Arginine derivatized by light Dansyl –OSu (**1**). (CID window -3,+6)

Indeed in pure organic medium derivatization of all AA plus the *N*-terminal amino group of Glutathione occured. Then we examined the reaction in aqueous medium using Tris buffer plus KOH to adjust pH=10. Also in this we have managed to derivatize all AA plus the *N*-terminal of Glutathione. In Figure 23 can been seen the MALDI MS/MS experiment of the derivatized Arginine by the light Dansyl –OSu (**1**) tag. The fragment m/z = 317.2 comes from Dansyl which protonates first and drives the fragmentation pathway to products that are not easily recognized by the conventional MS software.

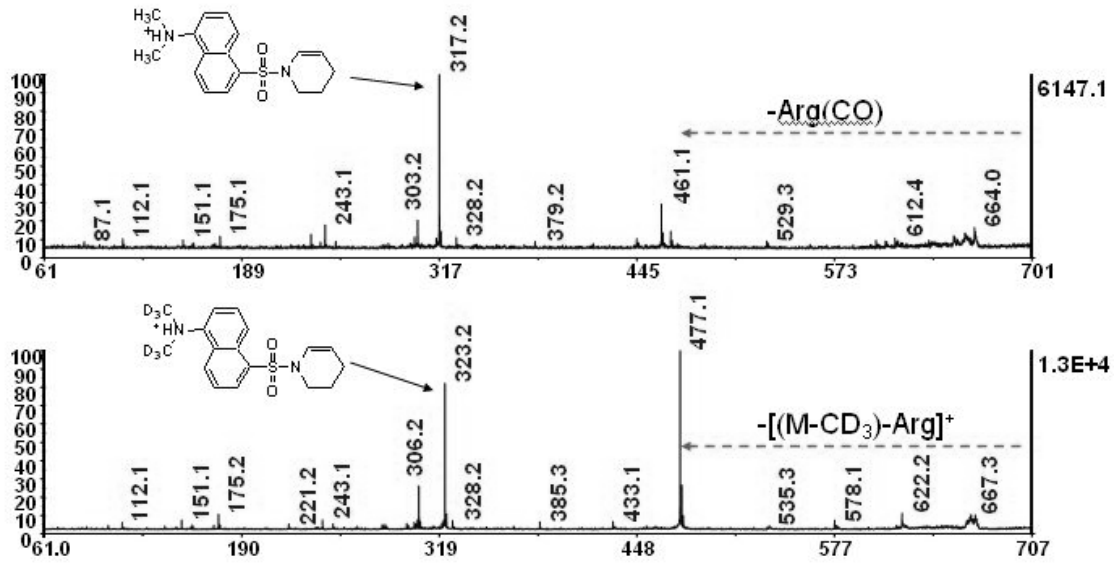


Figure 24. MALDI MS/MS of Arginine derivatized by heavy(2) and light(1) Dansyl –OSu tag. (CID window -3,+6)

Finally we performed the same reaction in Tris buffer (pH=8,8) in which we have selectively derivatized only the *N*-side chain amine group of Lysine (Figure 25). This interesting result will be used for the determination of Lysines or other primary amines in foodstuff.

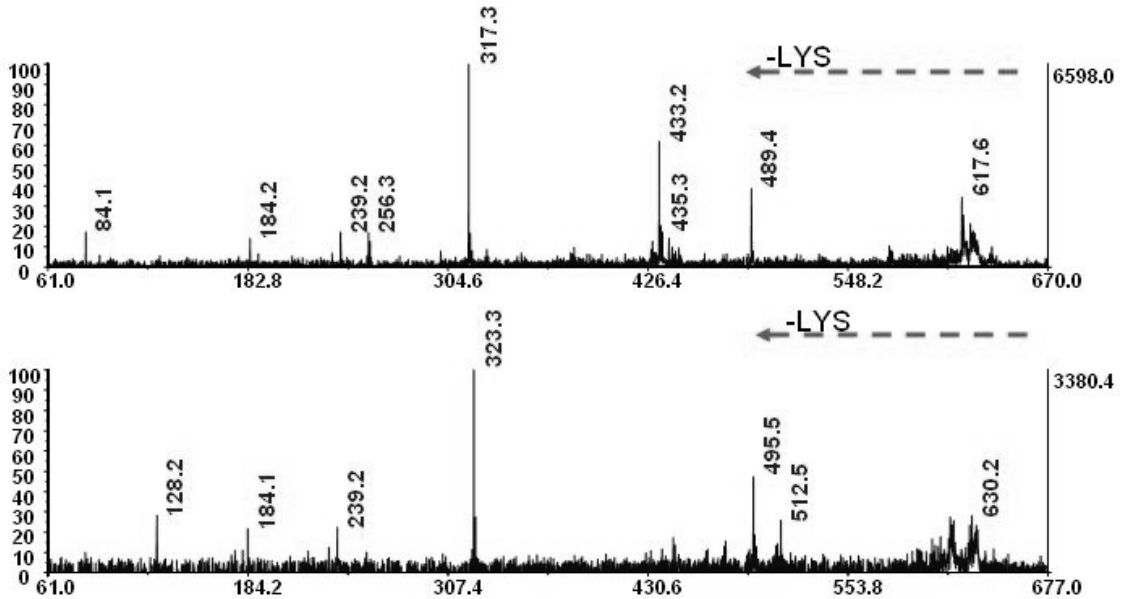


Figure 25. MALDI MS/MS of Lysine derivatized heavy(**2**) and light(**1**) Dansyl –OSu tag in aqueous medium and pH=8.8. (CID window -3,+6)

3.2.4 Reactivity test of Dansyl –Iodo Tags

The reactivity of Dansyl-Iodo Tags (**3** and **4**) was tested using the Standard Mixture (Figure 22). In fact in organic medium, as expected, only the thiol group of Glutathione was attached to the resin. In aqueous medium the reaction didn't take place probably because of the poor solubility of Glutathione in water. Thus, d_0 / d_6 resins (**3**, **4**) and standard mixture were incubated overnight in DMF in presence of DIPEA as the base (pH=10). Then the resin washed twice subsequently with DMF, ACN / H_20 6:4, DMF, DCM and the tagged products were cleaved-off the resin using a solution of 30% TFA in DCM for 1h. In the following Figure 26 are shown the corresponding ESI spectra. The peak m/z = 726,35 represents the GSH tagged by the d0 Dansyl-Iodo Tag (**3**) and the ion with m/z = 740,36 corresponds to the oxidized GSH tagged by the same reagent.

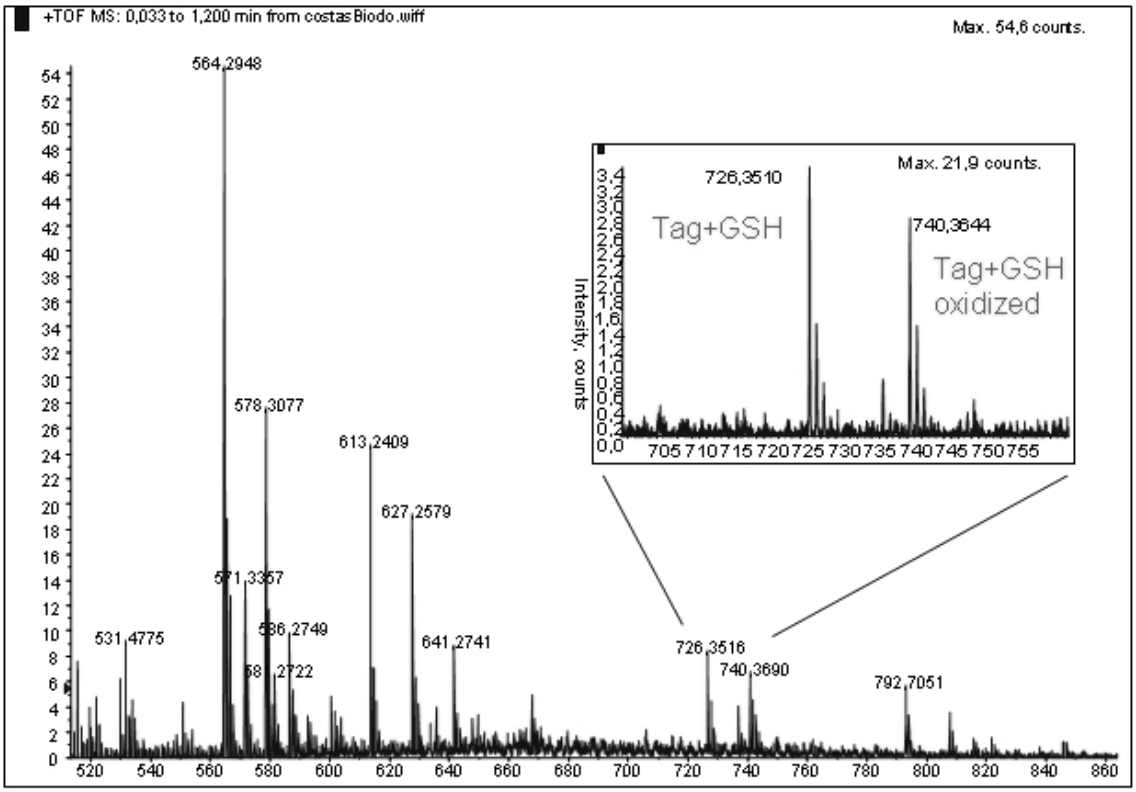


Figure 26. ESI-MS of GSH derivatized be d_0 Dansyl-Iodo Tag (**3**)

This methodology has been successfully applied to a standard protein mixture (BSA, Casein and Transferrin) as shown in Figure 27.

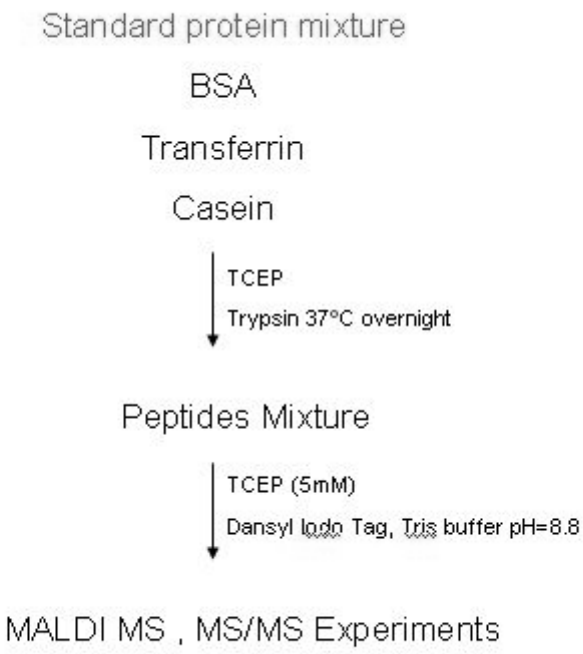


Figure 27. Application of Dansyl-Iodo Tags in a standard protein mixture

The tryptic digests of proteins were prepared as follows. Solutions of each bovine serum albumin (BSA), Transferrin and Casein at concentration of 1mg/mL in 10mM Tris buffer plus 5mM TCEP. Then were heated at 100 ^{o}C for 5 min to denaturate the proteins. The solutions were cooled to room temperature, followed by trypsin addition at the ratio of 1:50 (w/w) in respect to the protein. After overnight incubation at 37 ^{o}C the digestion of the proteins, occured. A second treatment of the tryptic digests with TCEP 5mM followed in order to reduce the disulfide bonds of cysteins. The thus prepared tryptic digests were treated overnight with the resins (**3**, **4**) in a vortex device. The day after the beads were washed subsequently twice with DMF, ACN / H_20 6:4 , DMF, DCM and the tagged products were cleaved-off the resin after the treatment with a solution of 30% TFA in DCM for 1h.

In the Figure 28 can be easily observed the difference of 6 Da between some peptides from the two spectra (d0 (**3**) and d6 (**4**) Dansyl-Iodo Tag resins, respectively).

The limitation of this method is the derivatization of very few peptides only. That probably has to do with the swelling conditions of the resins as well as the 3D stereochemical structure of the polymer. Our first hypothesis was that only small peptides can enter in the cavities of the resins and is based to the fact that in the corresponding spectra we observe peaks at maximum m/z ~900 (Figure 28).

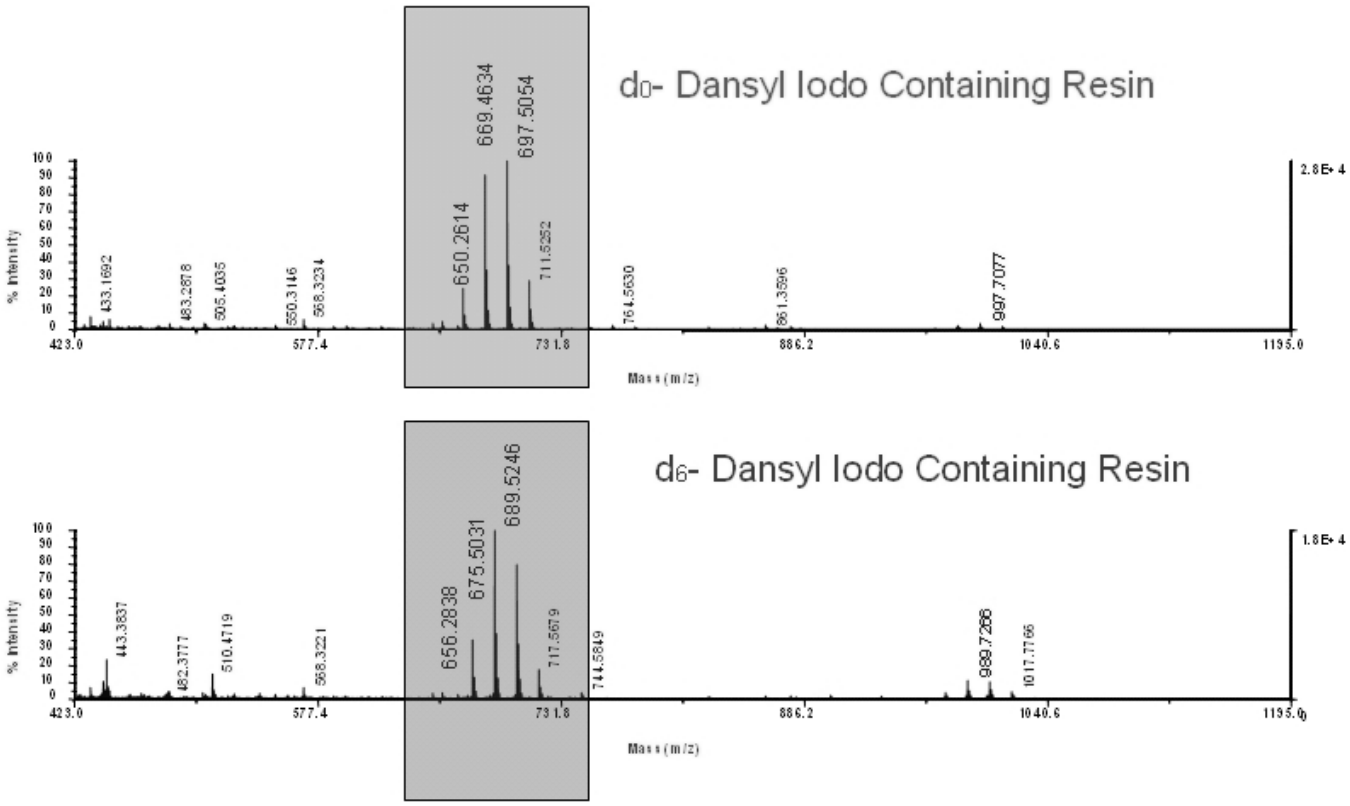


Figure 28. (Up) peptides derivatized by the d_0 Dansyl-Iodo Tag (**3**). (Down) derivatized by the d_6 Dansyl-Iodo Tag (**4**)

MS/MS experiments were performed as well. In Figure 29 there can be seen the 6Da difference between the two spectra. The peak at m/z = 170 derives from the d_0 Dansyl fragment while the ion m/z = 176 derives from the d_6 Dansyl fragment.

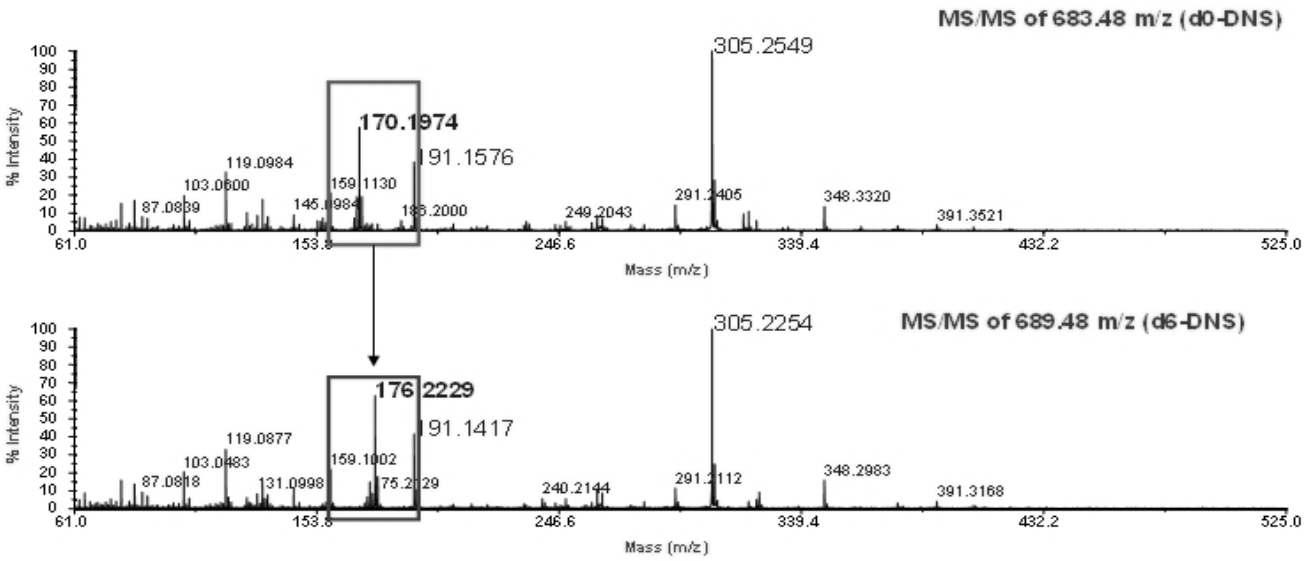


Figure 29. MALDI MS/MS experiments of Dansyl d0 and d6 peptides. (CID window -3,+6)

Another limitation of the Dansyl-tag method is that the fragment at m/z = 176 (d6 analogue) overlaps with the characteristic ion of arginine (m/z = 175) which is very often present, considering that trypsin cleaves exclusively C-terminal to arginine and lysine residues.

Concluding the Dansyl Solid Phase Tag work, our experiments stated that we have developed a new solid phase methodology using Dansyl chemistry for derivatizing BSA tryptic peptides. Under specific conditions (Tris, pH=8,8) we managed to derivatize only the ε-side chain group of Lysines. This observation can be extended in other research projects. However the fragment m/z = 176 from the d6-DNS analogue overlaps with the fragment of Arginine (175 m/z) and the available software can not recognize easily the fragmentation pathway. Another disadvantage of the Dansyl method is the 6 Da difference of the derivatized peptides. The CID window should be opened very much (-3 to +6, total 9 m.u) as a result the entering of undesirably ions in the collision cell that make the MS/MS spectra more complex. Finally this method derivatizes only small peptides (AA residues < 7). This is probably due to the decreased affinity of the polar amino acids towards the TentaGel resins. This subject will be discussed in the chapter. 3.4.

3.3 Cysteine-Cysteine Isobaric Solid Phase Tag

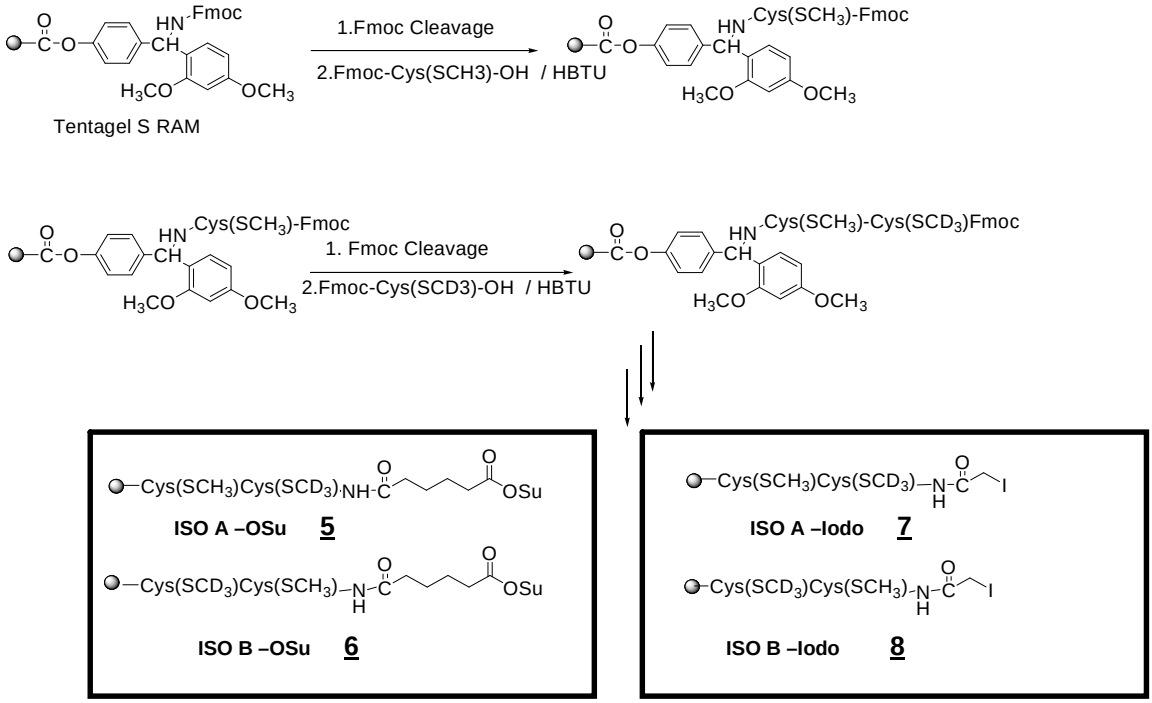
3.3.1 Development of Cys-Cys Isobaric resins

Considering the results from the Dansyl-tag containing resins we have designed and developed a new solid phase tag replacing the Dansyl-Lysine by two Cysteine residues which are methylated in the Thiol group (light *CH_3*, Heavy **CD_3**) but keeping the same reactive groups (Adipic -OSu and Iodo Tag). The inversion of the reagents during the solid phase synthesis of the resins, gives 2 different Isobaric tags (A(*CH3*)-A(**CD3**) and A(**CD3**)-A(*CH3*)) which under normal MS/MS fragmentation produce "y" fragments.

We started with the synthesis of Fmoc-Cys(CH_3)-OH (**14**) and Fmoc-Cys(CD_3)-OH (**15**) in one pot reaction through Cysteine methylation with CH_3I or CD_3I respectively in the presence of MeONa, followed by the Fmoc-protection of the *N*-terminal by Fmoc using Fmoc-OSu. Detailed synthetic procedure is described in the Experimental Part. Again the resin used was the Tentagel S RAM (0,24 mmol/g) bought from Rapp Polymere (Germany). Two different Tag-resins were synthesized (ISO A and ISO B). 0,090mmol of resin were placed in two different 5mL syringes equipped with a frit used for solid phase synthesis. Then followed Fmoc cleavage using a solution of 25% piperidine in DMF, in 3 cycles for 20 min each. Attachment of the "tag" was performed through amide bond formation, using a 2-fold excess of Fmoc-Cys(CH_3)-OH (ISO A) and Fmoc-Cys(CD_3)-OH (ISO B) in DMF and HBTU as the coupling reagent in the presence of DIPEA. Then Fmoc protecting group was cleaved as mentioned above. The second amide bond coupling was performed by the inversion of the two Cys analogues in each syringe. So 2-fold excess of Fmoc-Cys(CD_3)-OH (ISO A) and Fmoc-Cys(CH_3)-OH (ISO B) respectively dissolved in DMF were used in the presence of the system HBTU/DIPEA. The quantity of each of the two resulted resins (ISO A and ISO B) was splitted in two new syringes.

The first one was derivatized using the 2-iodoacetic N-Hydroxysuccinimidyl ester (4-fold excess) and DIPEA (4-fold excess) in DMF for 30min. (-Iodo Tag , Scheme 11)

The second one was derivatized using adipic anhydride (4-fold excess) and DIPEA (4-fold excess) in DMF for 1h. Finally the resin was treated for 5h with N-hydroxy succinimide (4-fold excess) and DIC (4-fold excess) in DMF. (-OSu Tag, Scheme 11)



Scheme 11. Synthesis of Cys-Cys Isobaric Solid Phase Tags

3.3.2 Reactivity test of Cys-Cys Isobaric Tags

The reactivity of Cys-Cys Isobaric –OSu Tags (**5**, **6**) was tested using the Standard Mixture (Figure 22). Both in aqueous or organic medium the results were comparable with those of Dansyl-tag –OSu resins. In this case there is a mass increment of 365 Da for each derivatized analyte (Figure 29). The tagging protocol involves overnight incubation of ISO A (**5**) / ISO B (**6**) resins with our standard AA mixture in DMF and DIPEA as the base (pH=10). After washing the resin twice subsequently with DMF,

ACN / H_20 6:4, DMF, DCM the tagged products were cleaved-off the resin by treatment with a solution of 30% TFA in DCM for 1h. The resulted solutions were mixed in various ratios. As shown in the ESI spectrum (Figure 29), there is a mass increment of 365 Da for each derivatized analyte.

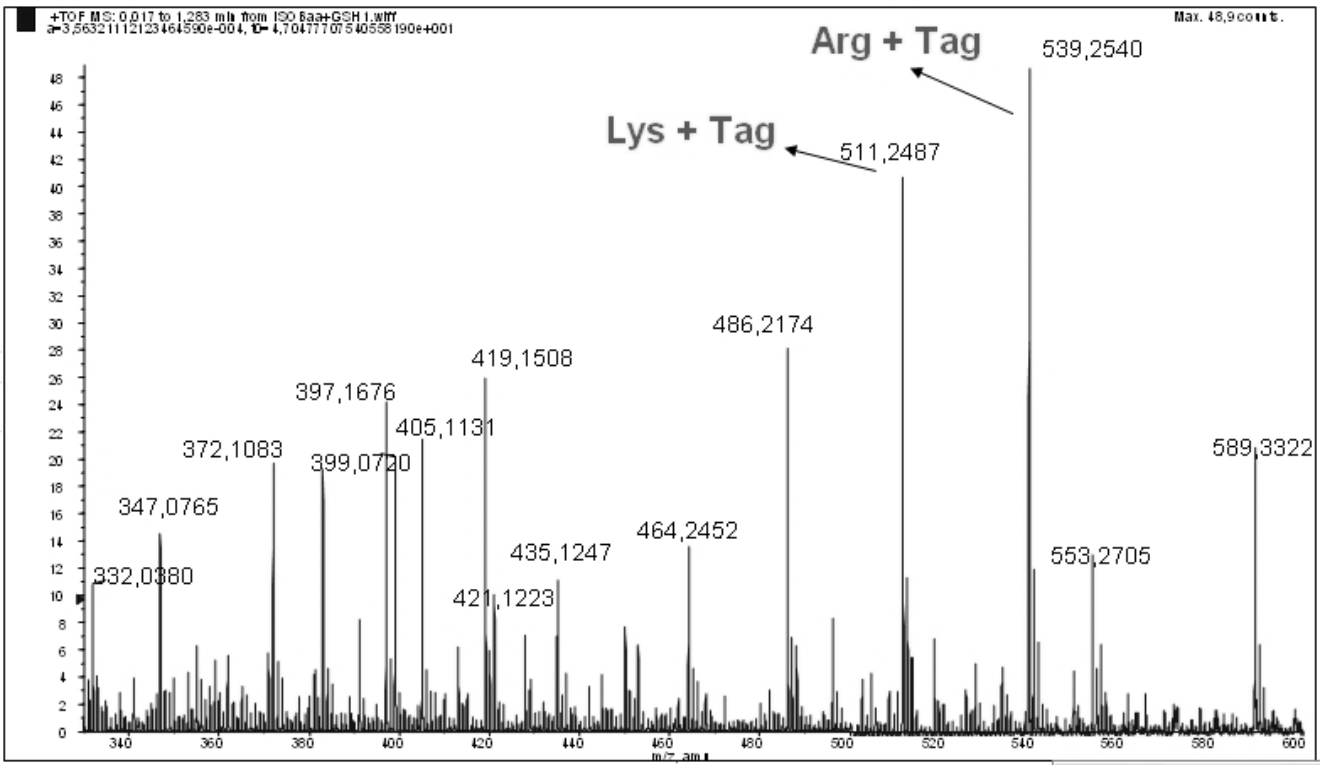


Figure 29. ESI-MS spectra of the derivatized Lysine and Arginine

In the MS spectrum appears only one peak for each derivatized analyte, because the solid phase tagging reagents have equal molecular weight (Isobaric). Thus the relative quantitative results can be obtained only by MS/MS experiment (Figure 30).

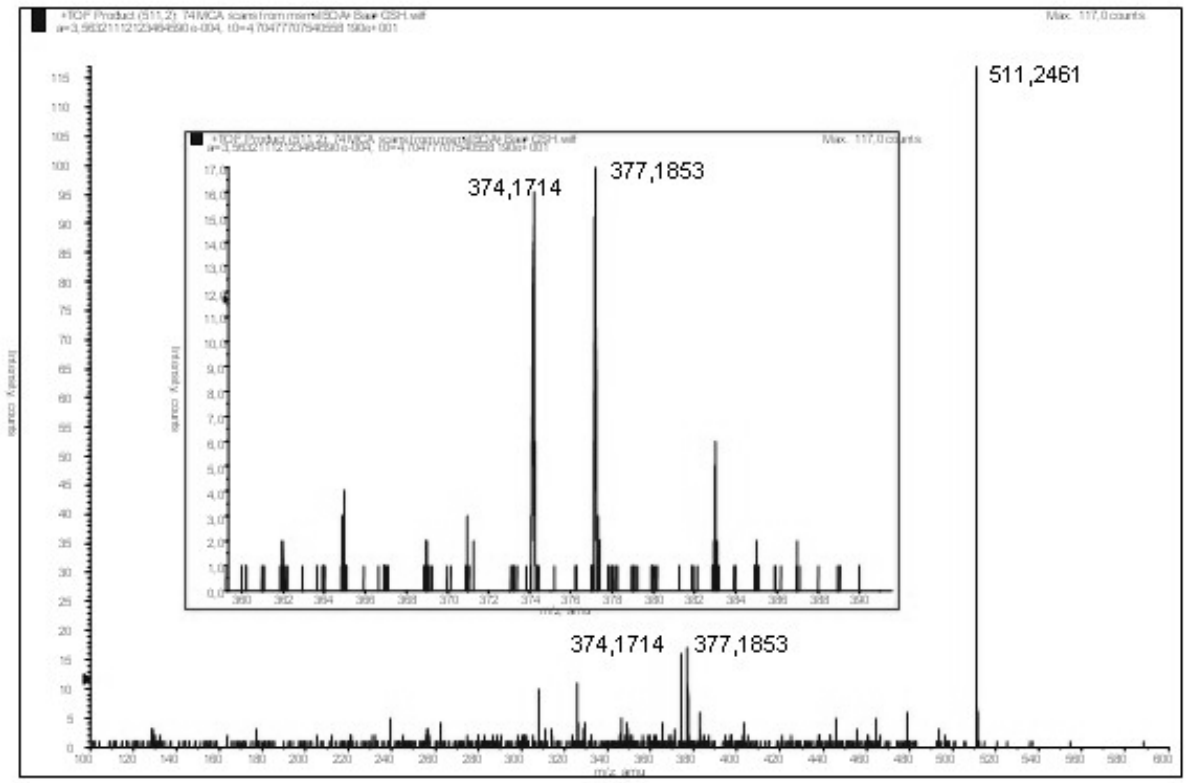


Figure 30. ESI-MS/MS of derivatized Lysine by ISO A -OSu (**5**) and ISO B –OSu (**6**) in 1:1 ratio

The fragments at m/z = 374 and m/z = 377 are produced from the ISO B-lysine and ISO A-lysine respectively (Figure 31). These are typical product-ions of peptide “y fragmentation” and thus can be easily recognised from the available MS analysis software.

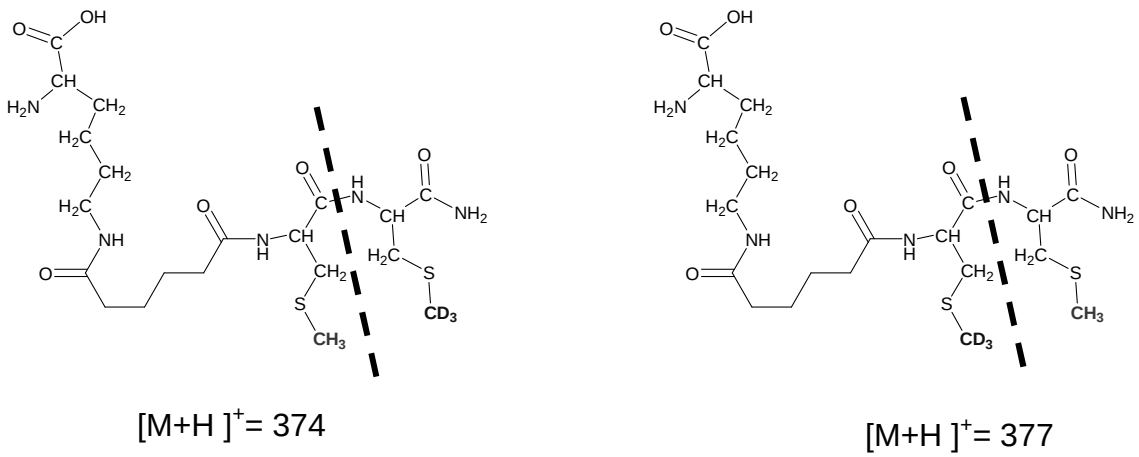


Figure 31. Fragments produced in the MS/MS experiments of the Isobaric tag-reagents

Furthermore this methodology was successfully applied to a standard protein (Transferrin, Figure 32).

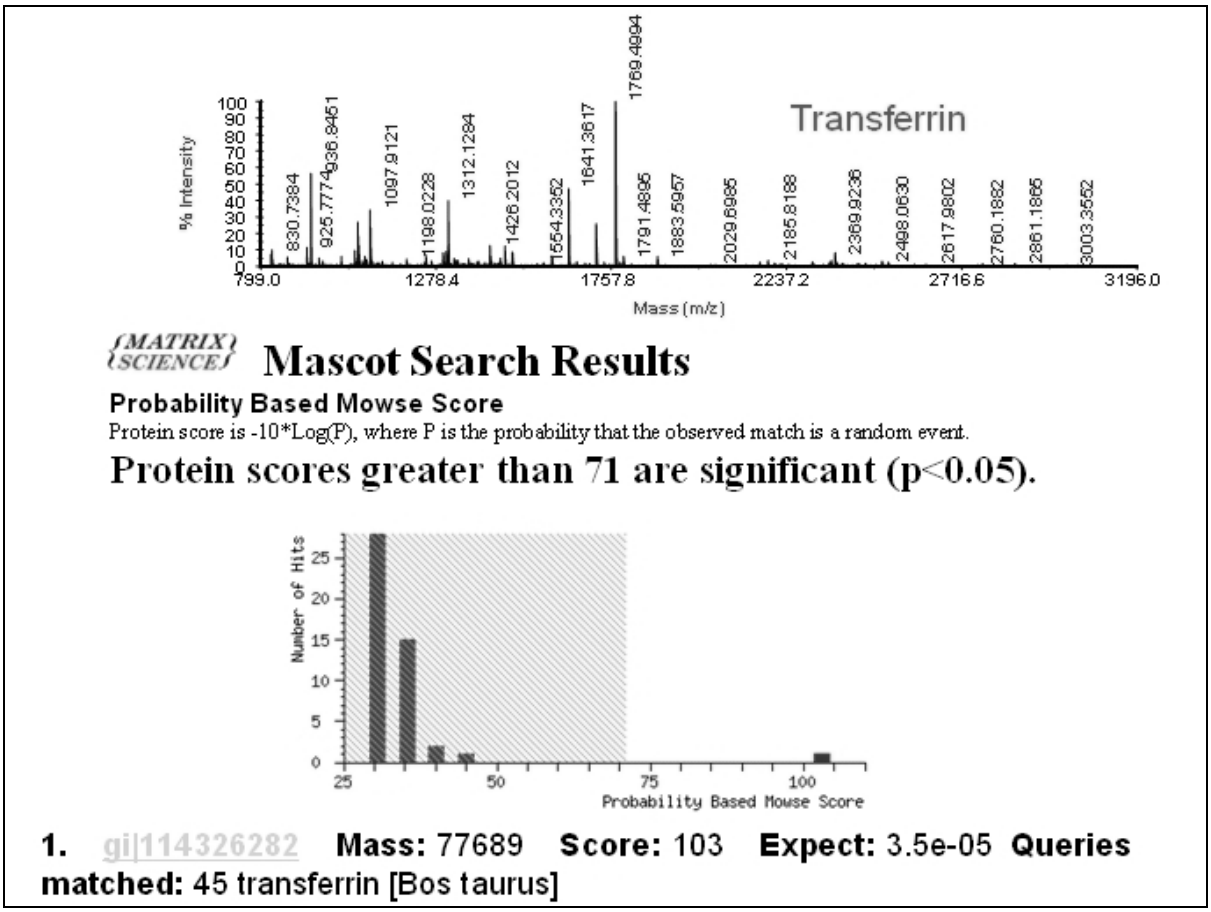


Figure 32. MALDI MS and Mascot response of the Transferrin Tryptic digest

The tryptic digest of Transferrin was prepared as follows. To a solution of Transferrin at concentration of 1mg/mL each in 10mM Tris buffer plus KOH to adjust pH=10, was added Trypsin at ratio 1:50 (w/w) in respect to the protein. After overnight incubation at 37 ^{o}C the digestion of the protein occured. The thus prepared tryptic digests were treated overnight with the resins ISO A -OSu (**5**) and ISO B –OSu (**6**) in a vortex device. The day after the beads were washed subsequently twice with DMF, ACN / H_20 6:4 , DMF, DCM and the tagged products were cleaved-off the resin after the treatment with a solution of 30% TFA in DCM for 1h. The resulted solutions were lyophilized and then mixed in various ratios.

The MS/MS spectra of the peptide TSHMDCIK (m/z = 1296,57) can be seen in Figure 33.

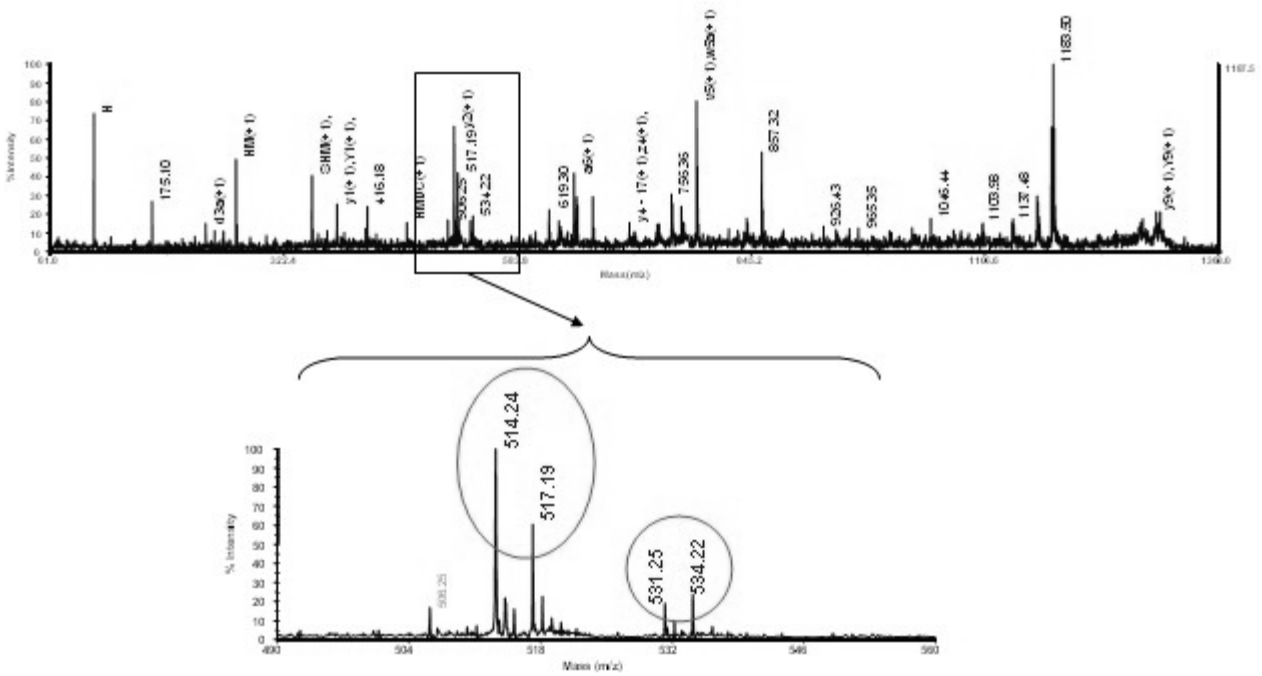


Figure 33. MALDI MS/MS spectra of TSHMDCIK (m/z = 1296,57) from Transferrin in 1:1 ratio of ISO A (**5**) and ISO B (**6**)

In conclusion the idea to use Cys-Cys Isobaric tags gave us the advantage to overcome the necessity to use large CID window width. Indeed in this case we could allow only one specific ion to pass to collision cell, because this one contains both heavy and light tags. The quantitative results are then obtained from the typical peptide "y fragments" after MS/MS analysis. Furthermore these fragments are easily recognised from the available MS analysis software, which is important for the application of the method in protein identification. Unfortunately the limitation of this tagging protocol is the derivatization of low mass peptides like in the case of the Dansyl-tag resins.

3.4 Glycine-Alanine Isobaric Isotopic Label-Free Solid Phase Tag

3.4.1 Development of Glycine-Alanine Isobaric Isotopic Label-Free resins

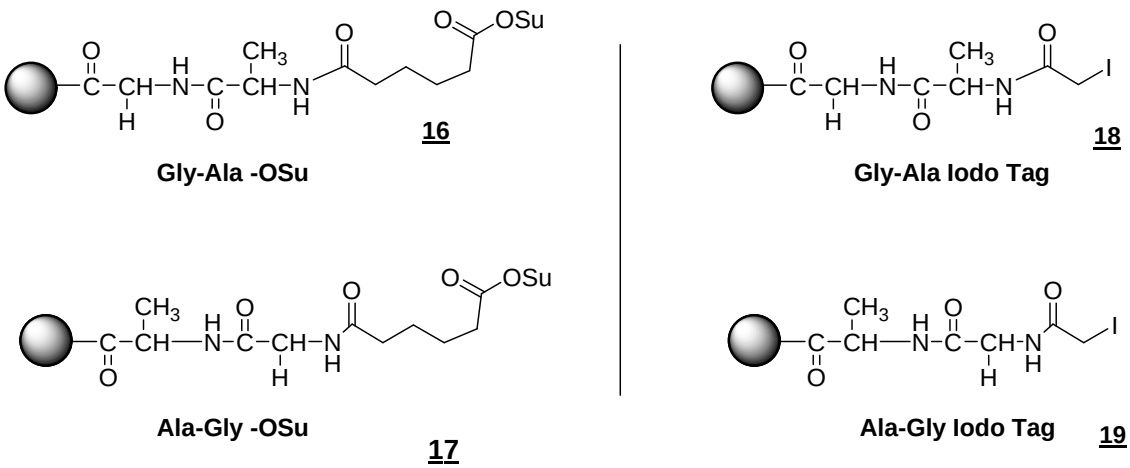
Focusing our research in the elimination of the isotopic effect in chromatography and in MS that creates the presence of stable isotopes we have developed a non-isotopic Isobaric Tag. The stable isotope techniques in literature provide an error variable from 8% to 20%. This was also confirmed in our case (Dansyl), where the observed error was 20%. Having in our mind the simplicity and the ease of synthesis we designed our isotope label-free resin using the simplest amino acids in nature: Glycine and Alanine. Lagana et al[85] demonstrated that quantitation can be performed using matrix-matched calibration curves (due to a matrix effect) of two synthetic peptides without Stable Isotopes but only changing an amino acid in the sequence.

We have prepared a new Isobaric Tag without Stable Isotopes using only Glycine and Alanine. The inversion of the sequence during the solid phase synthesis of the tag-resins produces 2 Isobaric tag-reagents. Under MS/MS experiment the typical peptide "y fragments" differ by 14 mass units and allow us to obtain quantitative results.

Two different tag-resins were synthesized (Gly-Ala and Ala-Gly). Again the resin used was the Tentagel S RAM (0,24 mmol/g). Briefly, we used 0,080mmol of resin placed in two different 5mL syringes equipped with a frit used for solid phase synthesis. Then followed Fmoc cleavage using a solution of 25% piperidine in DMF, in 3 cycles for 20 min each. Attachment of the "tag" was performed through amide bond formation using 2-fold excess of Fmoc-Gly-OH (Gly-Ala resin) and Fmoc-Ala-OH (Ala-Gly resin) in DMF and HBTU as the coupling reagent. Then Fmoc protecting group was cleaved as mentioned above. The second amide bond coupling was performed by the inversion of the two reagents in each syringe. So 2-fold excess of Fmoc-Ala-OH (Gly-Ala resin) or Fmoc-Gly-OH (Ala-Gly resin) respectively dissolved in DMF were used in the presence of the system HBTU /DIPEA. The quantity of each of the two resulted resins (Gly-Ala and Ala-Gly) was splitted in two new syringes.

The first one was derivatized using the 2-iodoacetic N-Hydroxysuccinimidyl ester (4-fold excess) and DIPEA (4-fold excess) in DMF for 30min. (-Iodo Tag , Scheme 12)

The second one was derivatized using adipic anhydride (4-fold excess) and DIPEA (4-fold excess) in DMF for 1h. Finally the resin was treated for 5h with N-hydroxy succinimide (4-fold excess) and DIC (4-fold excess) in DMF. (-OSu Tag, Scheme 12)



Scheme 12. Gly-Ala isobaric Label-Free resins

3.4.2 Reactivity test of Glycine-Alanine Isobaric Isotopic Label-Free resins

Investigating the limited reactivity of peptides generated by trypsin, we have tried to generate smaller peptides (containing less than 10 AA residues) by digesting the BSA with pepsin for 4h followed by overnight digestion with trypsin. In this case the pepsin hydrolyzes BSA on lipophilic sites (Phe, Leu) and after trypsin hydrolyzes on Lys and Arg. The reactivity of Glycine-Alanine Isobaric –OSu Tags (**16**, **17**) was tested using this low molecular weight peptide mixture. The derivatization was performed overnight in Tris (pH=10) buffer. The observed derivatized peptides this time were up to 1800 Da. These results prompted us to review our initial hypothesis that the poor reactivity towards peptides with molecular mass > 900 Da which we have observed before with the Dansyl-tag resin and Cys-Cys isobaric resin is directly connected to the dimension

of the resin cavity. We suppose then that lipophilic peptides have a major affinity towards the TentaGel resin.

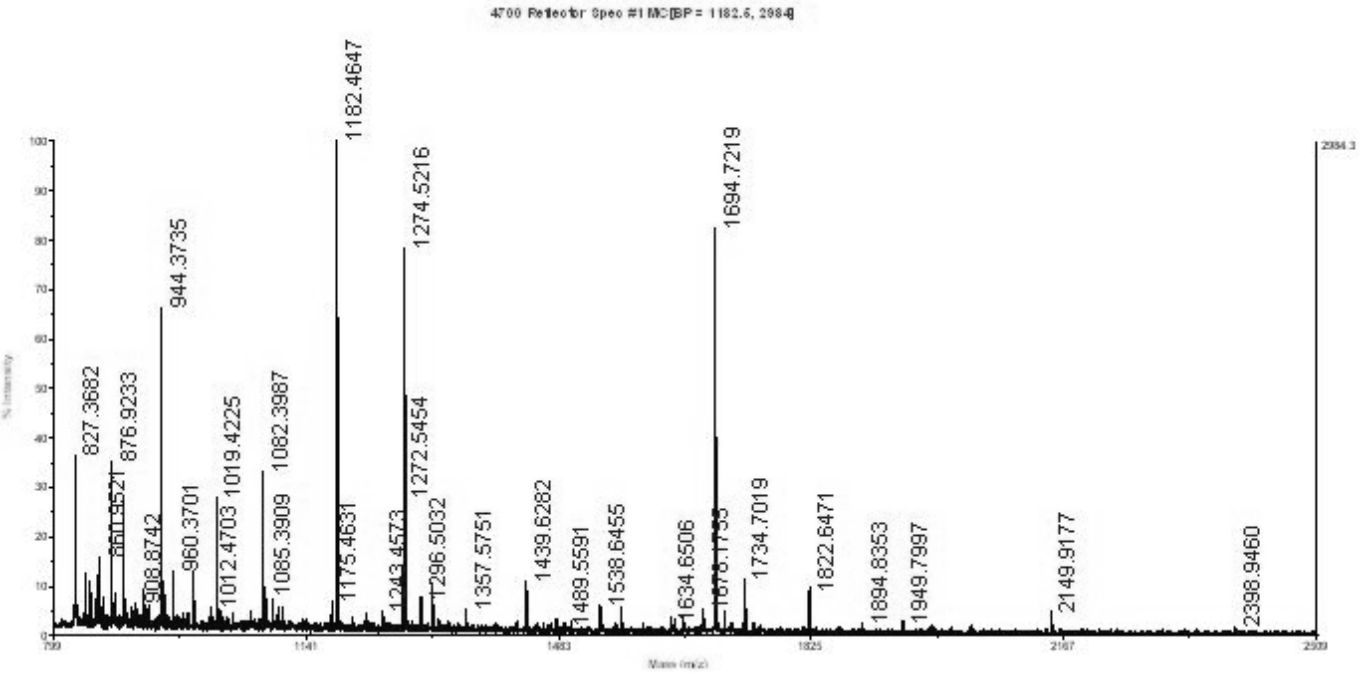


Figure 34. Derivatized BSA peptides by **Gly-Ala**-OSu isobaric resin (**16**) digested first with Pepsin followed by Trypsin

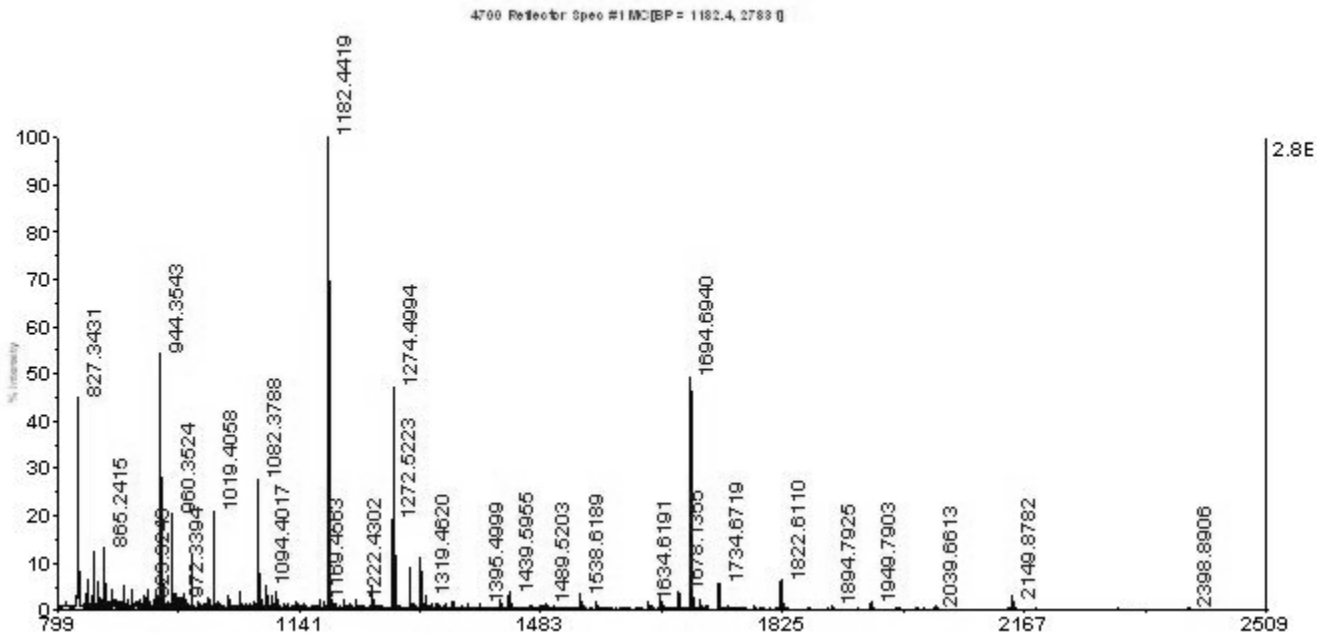


Figure 35. Derivatized BSA peptides by **Ala-Gly**-OSu isobaric resin (**17**) digested first with Pepsin followed by Trypsin

In the Figures 34 and 35 can easily be observed that the MS spectra of the derivatized peptide mixtures by the resins Gly-Ala-OSu and Ala-Gly-OSu are equal. The difference can only be valuated in the MS/MS experiments where the fragmentation changes due to the different sequence of the tagged peptide.

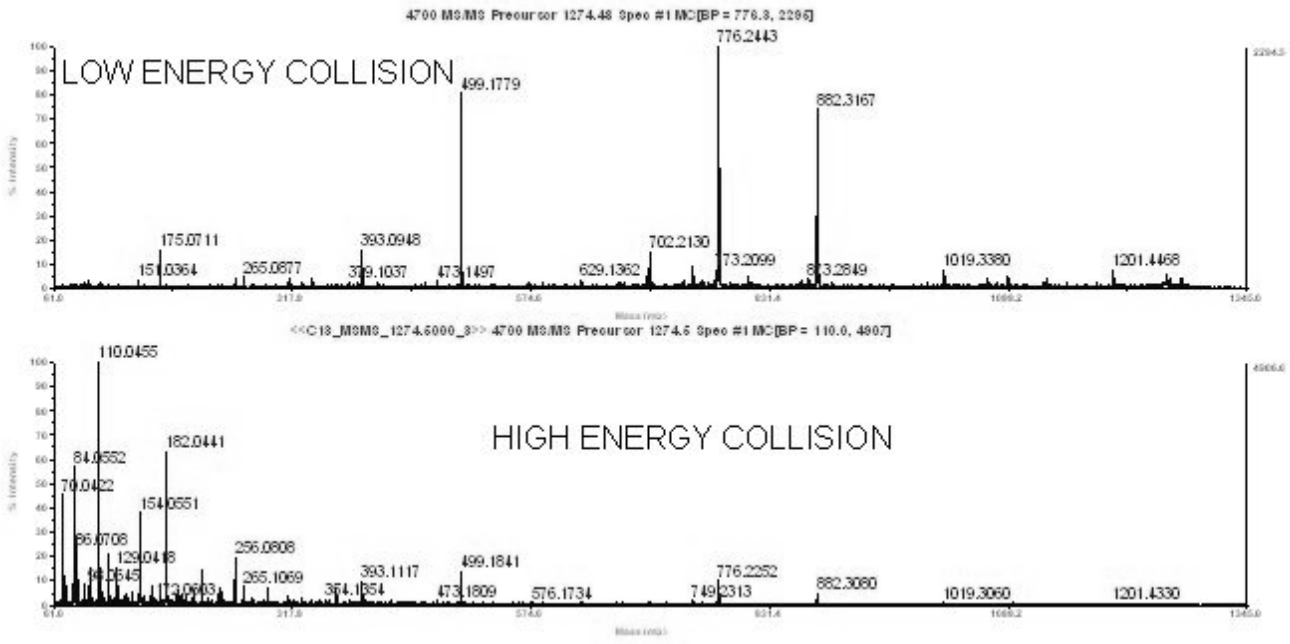


Figure 36. MS/MS spectra of m/z = 1274.6 (Gly-Ala-OSu Tag)

In Figure 36 is shown the MS/MS fragmentation of the peptide m/z = 1274.6 both in low collision energy (CID OFF) and in high collision energy (CID ON). The specific fragment of this tag is generated only in the high energy experiment (Figure 37).

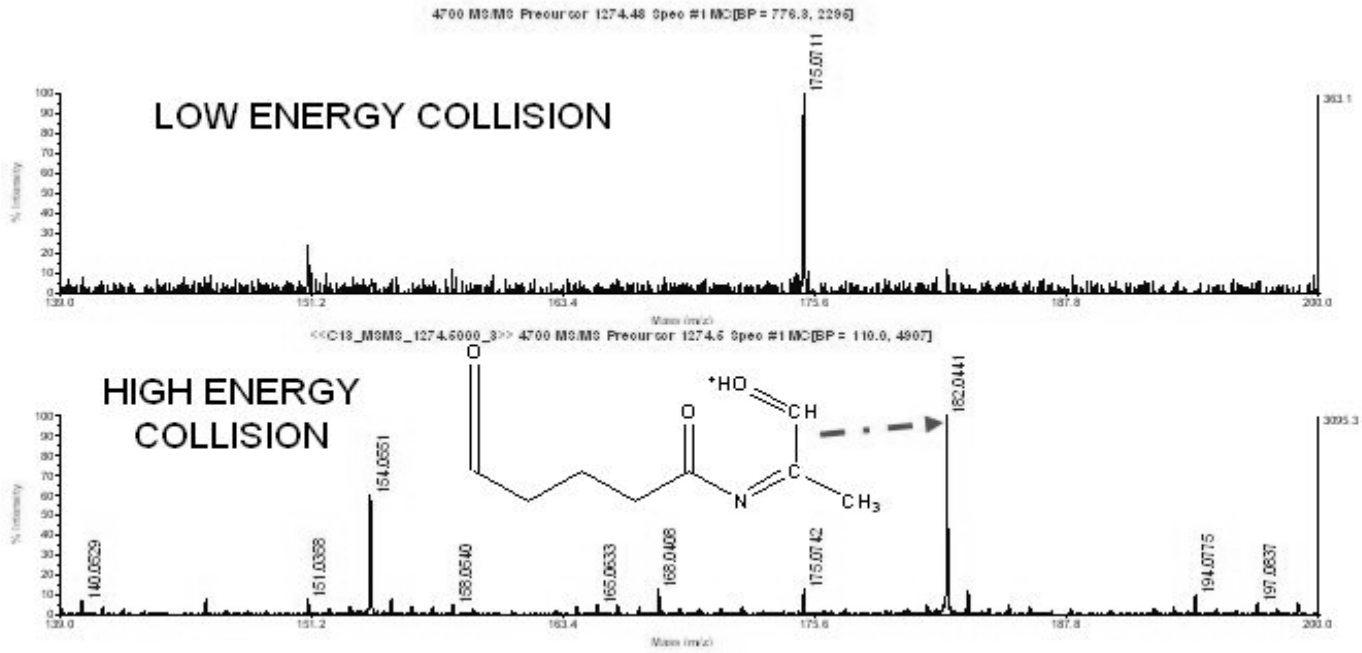


Figure 37. The specific-fragment (m/z = 182) of MS/MS spectra of m/z = 1274.6 (Gly-Ala-OSu Tag)

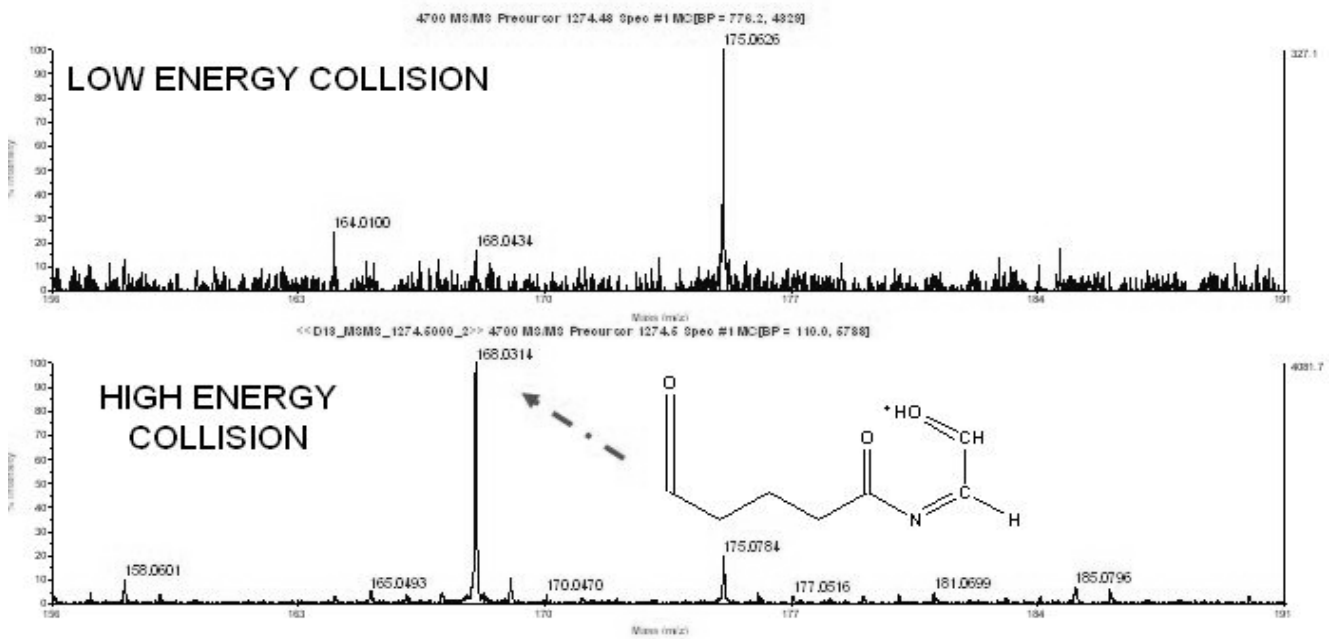


Figure 38. The specific-fragment (m/z = 168) of MS/MS spectra of m/z = 1274.6 (Ala-Gly-OSu Tag)

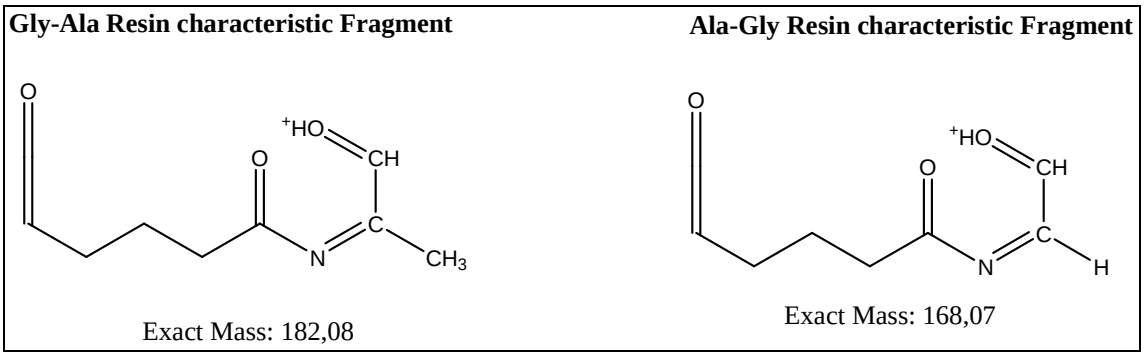


Figure 39. Characteristic fragmentation of the isobaric resins (Gly-Ala and Ala-Gly)

In the Figures 37, 38, 39 are shown the specific fragments that are generated from the two different isobaric resins only in high energy MS/MS experiment.

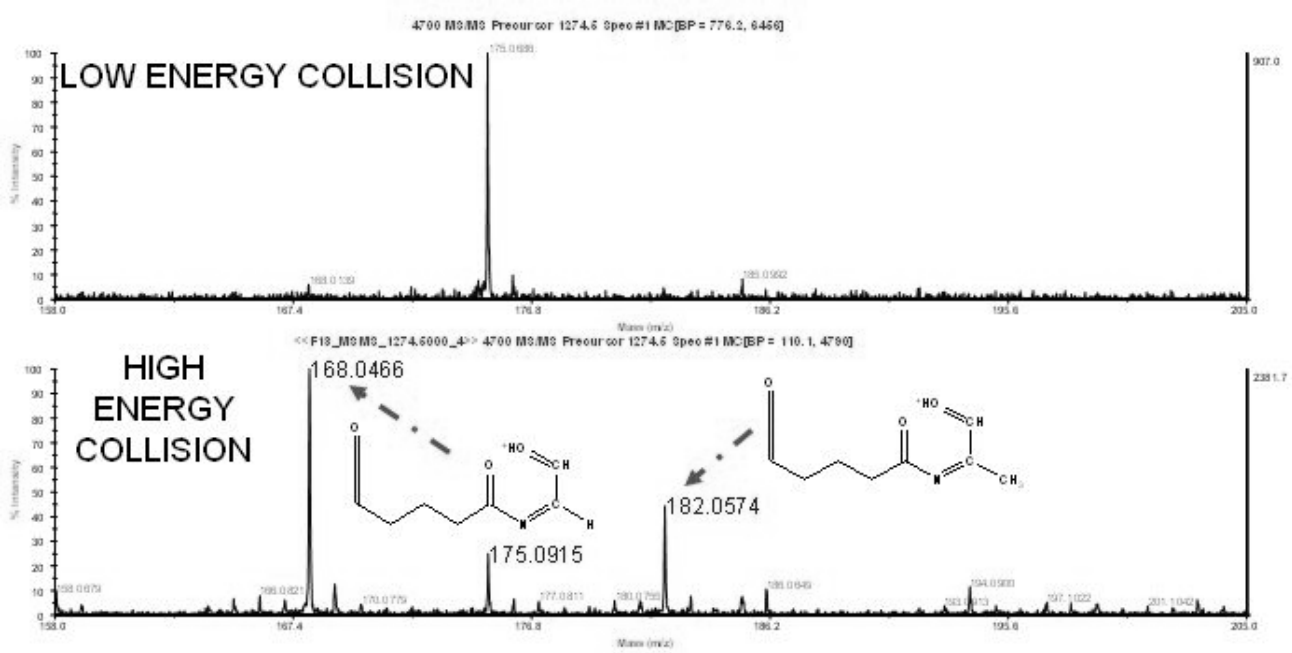


Figure 39. MS/MS spectra of m/z = 1274.6 in Mix 1:1 of Gly-Ala and Ala-Gly resins

The relative quantitative analysis of proteins can be made calculating the relative intensity of the peaks corresponding to the characteristic fragments of the two isobaric resins.

For the confirmation of our theory that the lipophilic peptides are shown an increased affinity towards the type of the TentaGel-resin, we have digested BSA (Figure 40) with pepsin in order to produce only lipophilic peptides. These peptides were derivatized using the Gly-Ala isobaric resins.

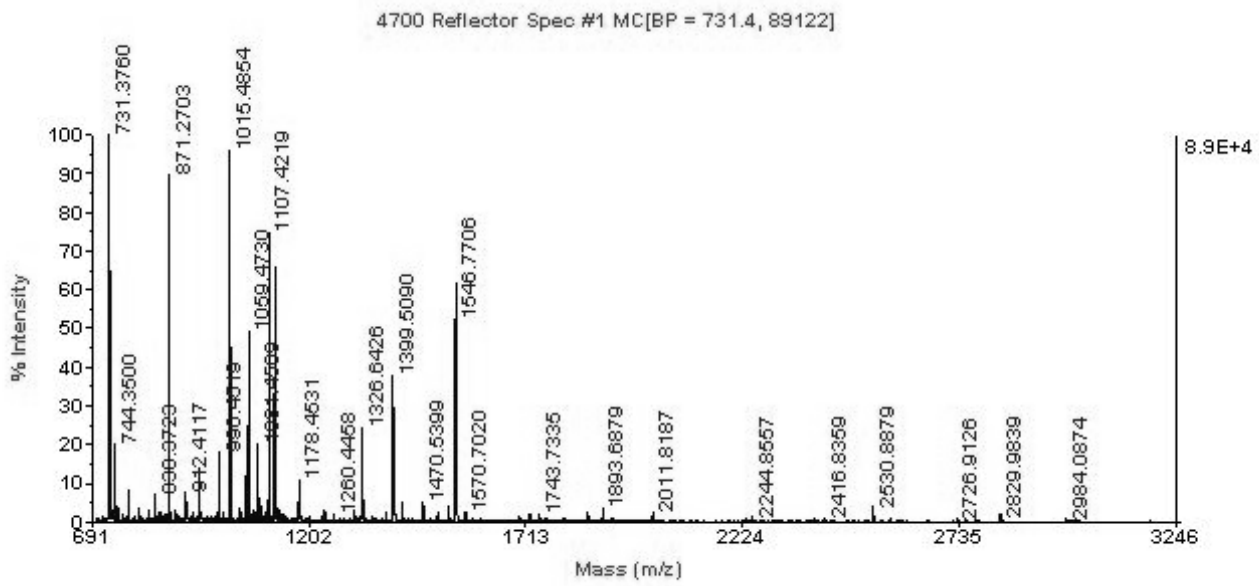


Figure 40. BSA peptides digested with Pepsin

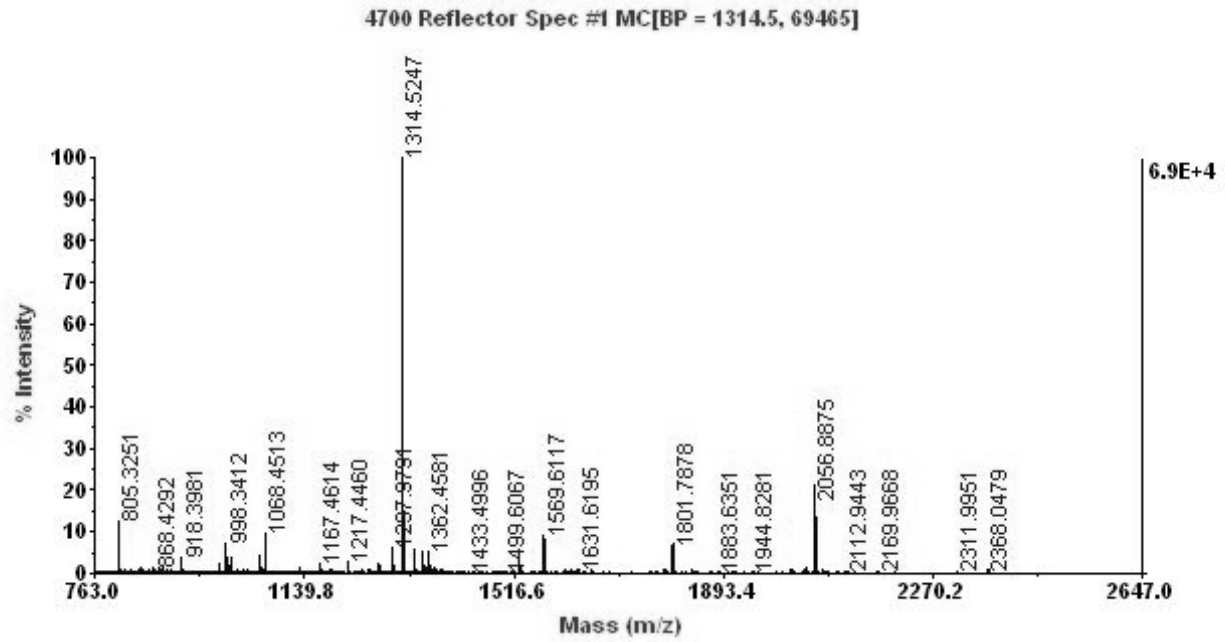


Figure 41. Derivatized BSA peptides digested with Pepsin using Ala-Gly isobaric Resin

Indeed in the Figure 40 is clearly shown that several peptides with mass up to 2000 Da have been derivatized when treated with the Ala-Gly isobaric Resin (**17**).

In order to perform the relative quantitative analysis of BSA, we have prepared solutions of different ratios of Gly-Ala and Ala-Gly derivatized peptides. We have chosen a peptide with average mass for performing the MS/MS experiments. The selected peptide was of m/z = 1314,5

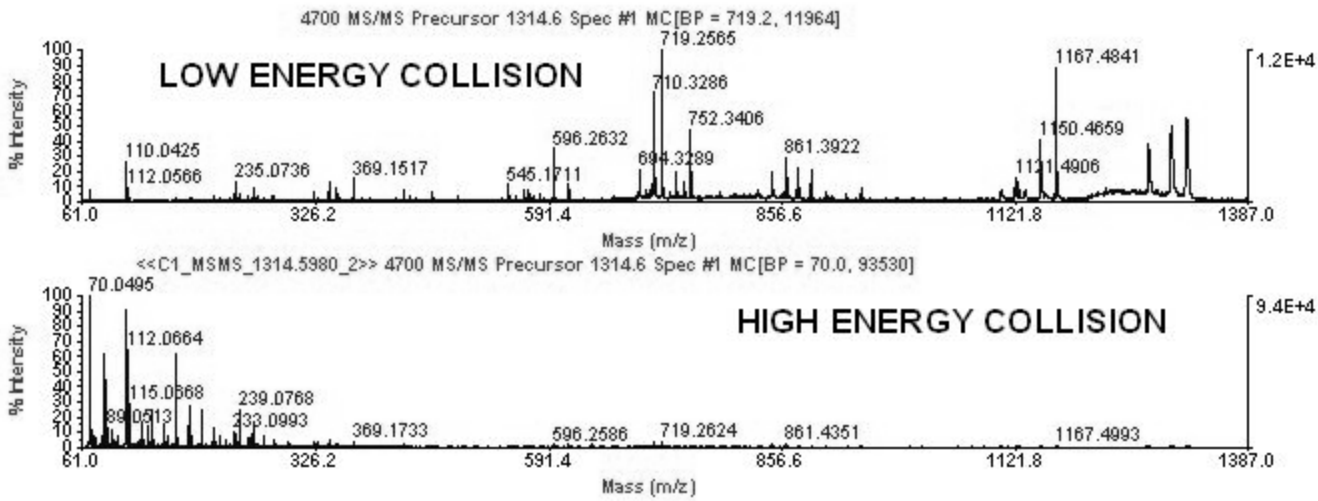


Figure 42. MS/MS experiment of the ion at m/z = 1314.5 in a mixture of Gly-Ala / Ala-Gly resins in ratio 1:2

The pair of fragments, that can be calculated the quantity of the BSA, is created only in high energy collision (Figure 39). The spectra of the various mixtures are shown in the Figure 43.

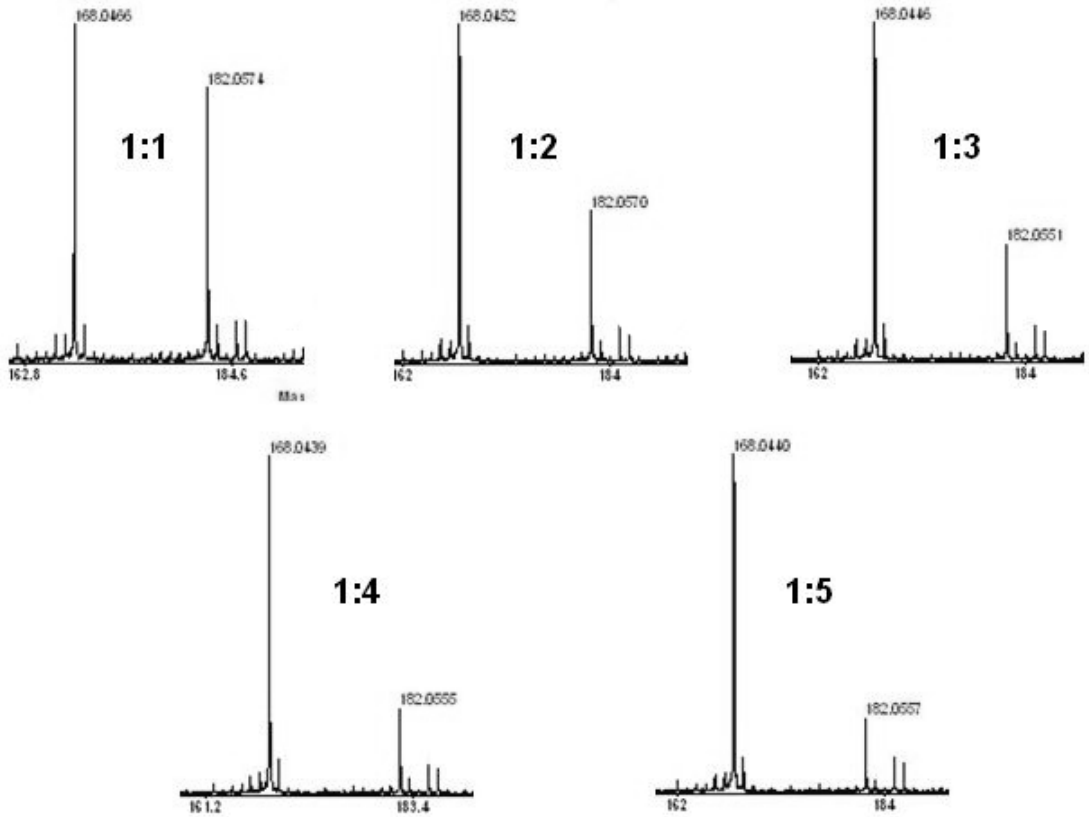


Figure 43. Pair of peaks (168 *vs* 182) generated by the MS/MS experiments of ion at m/z = 1314.5 in different ratios.

Ions/Ratios	*1:1*	*1:2*	*1:3*	*1:4*	*1:5*
168 m/z	34,38	55,57	51,27	54,47	56,81
182 m/z	26,67	23,96	17,57	13,49	11,64
168/182	**1,3**	**2,3**	**2,9**	**4,0**	**4,9**

Table 2. Relative intensities of fragments at m/z = 168 and at m/z = 182 given from Data Explorer Software and the experimental 168/182 peak ratio

In order to demonstrate the linearity of our method we have drawn the calibration curve (Figure 44) of the derivatized BSA peptides (Table 2). In this table there are the values from the Data Explorer Software in 1:1, 1:2,1:3, 1:4, 1:5 ratio of the fragments at m/z = 168 and m/z = 182 which derives from the Ala-Gly and Gly-Ala isobaric resins respectively.

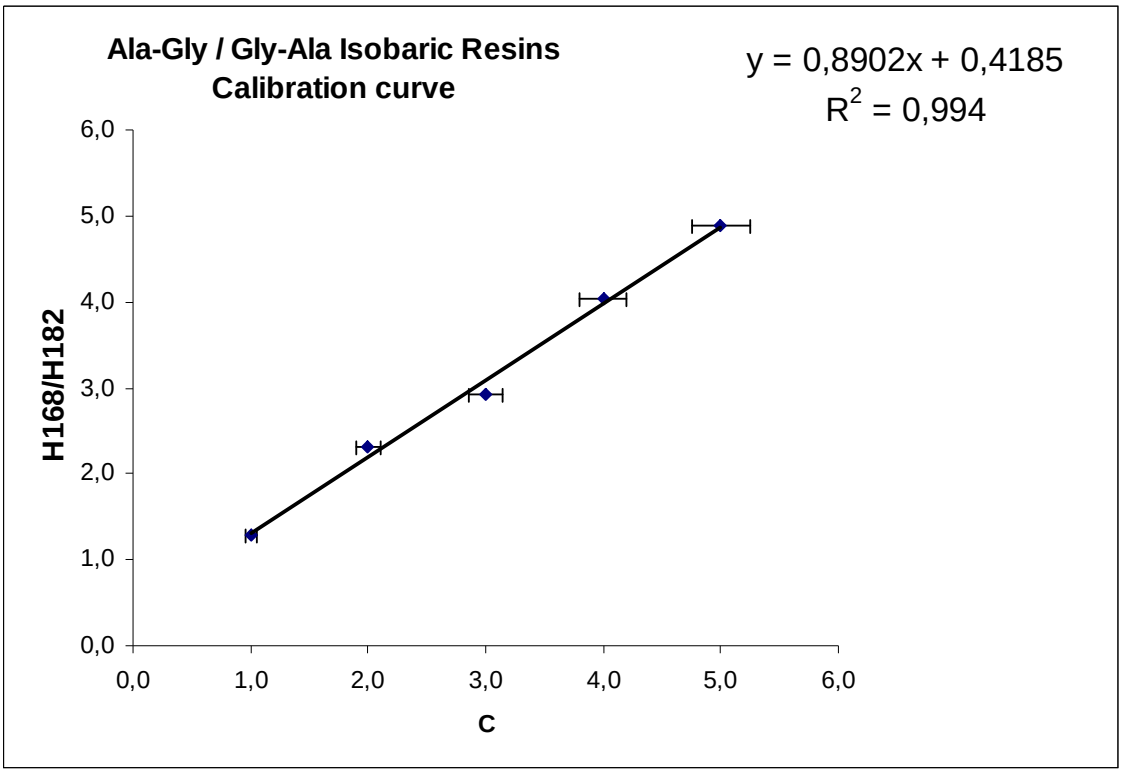


Figure 44. Calibration curve of Gly-Ala and Ala-Gly Isobaric resins

In conclusion the results of our work with the isobaric Gly-Ala and Ala-Gly-tag resins, have confirmed our hypothesis that lipophilic peptides derived from the digestion of BSA with pepsin show increased affinity towards the TentaGel resin. We have also demonstrated the linearity of our quantitative analysis method in MALDI MS/MS experiments by use of different ratios of derivatized peptides.

4

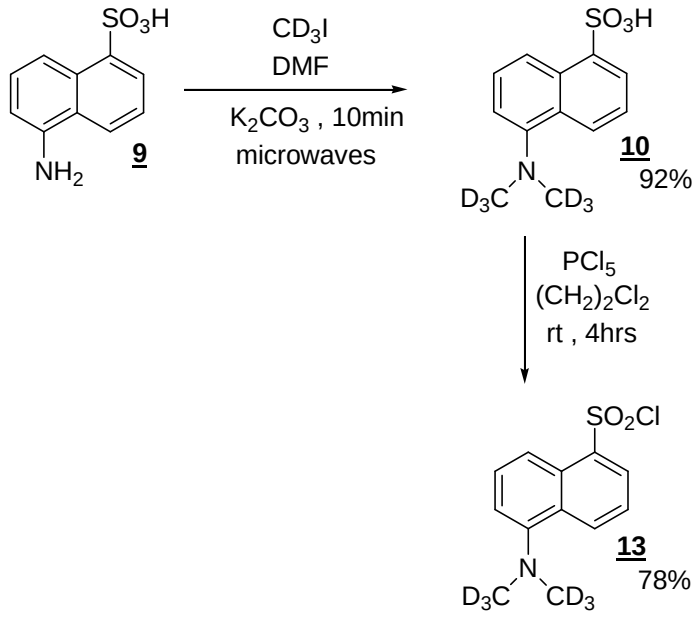
Experimental Part

4. Experimental part

Chemicals

General chemicals were obtained from Sigma (Milano) at Analar or better purity grade, unless otherwise stated.

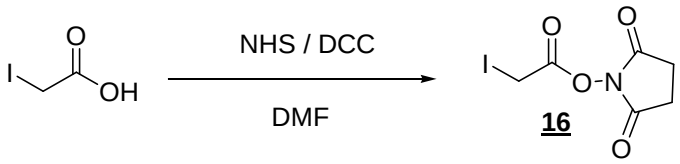
Synthesis of 5-*N*,*N*-trideuteromethylaminonaphtha-lene-1-sulphonic acid (10) using microwave assisted synthesis



To a suspension of **9** (5mmol) in DMF (10mL) placed in a sealed reaction vessel of a Synthos 3000 (Anton Paar GmbH) multimode microwave synthesizer were added K_2CO_3 (25mmol), d_3-methyliodide (25mmol) and stirred at 110 ^oC (direct temperature control, T-probe sensor) for 10 min (3 min ramp, 7 min hold at 110^oC) under microwave irradiation. After acidification with HCl 6N, 1.18g (gray crystals, m.p > 300 ^oC) of 5-(d_6-dimethylamino)naphthalene-1-sulfonic acid (d_6-Dansyl sulphonic acid, **10**) were obtained in 92% yield, after filtration. ^{1}H-NMR (300 MHz, d_6-DMSO) data: (δ) 8.97 (1H, d, J = 8.4 Hz, H_2), 8.27 (1H, d, J = 8.7 Hz, H_8), 8.08 (1H, d, J = 7.2 Hz, H_4),7.85 (1H, d, J = 7.5 Hz, H_6) 7.75 – 7.42 (2H, m, H_3+H_7), 4.45 (1H, s_b, -SO_3H).
ESI-MS: m/z= 258,25 $(M+H)^+$

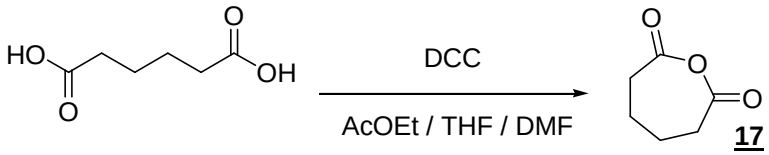
Preparation of 5-*N*,*N*-trideuteromethylaminonaphtha-lene-1-sulphonyl chloride (13): Phosphorus pentachloride (5,41 mmol, 1,13g) was suspended in dry 1,2-dichloro-ethane (10 mL) at 0 ^{o}C under nitrogen atmosphere. d_6-Dansyl sulphonic acid **(10)** (3,25mmol , 835mg) was added in small portions and after the completion of the addition the reaction mixture left to be stirred at r.t. for further 4 hrs. Then 5 mL of cold water were added, followed by 1N aqueous $NaHCO_3$ addition until pH=8. The mixture was rapidly extracted with $CHCl_3$ and the organic phase dried with Na_2SO_4, filtered and the solvent removed under reduced pressure to afford 685mg yellow crystals of d_6-Dansyl chloride (**13**) (78 % yield). Also 0,17mmol (44,5mg) of the starting material was recovered after the acidification of the water phase with HCl.
ESI-MS: *m/z*= 275,90 $(M+H)^+$

Synthesis of 2-iodoacetic N-Hydroxysuccinimidyl ester.



To a soln of 2-iodoacetic acid (10mmol) and N-hydroxy succinimide (11mmol) in 27,5 mL of DCM/DMF (10:1) cooled at 0^{o}C was added DCC (12mmol) and the thus resulting mixture stirred for further 2h in r.t. The DCU was filtered off and the filtrate was diluted with $CHCl_3$ and washed once with 5% $NaHCO_3$ and H_2O. After drying (Na_2SO_4), filtration and evaporation of the solvent under vaccum the 2-iodoacetic N-Hydroxysuccinimidyl ester was isolated as yellow flakes in 89% yield.
ESI-MS: *m/z*= 283,90 $(M+H)^+$

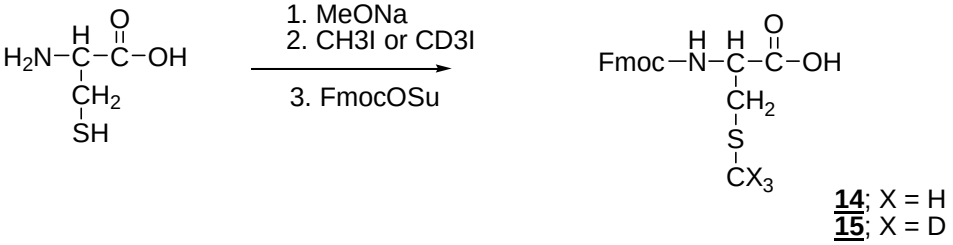
Synthesis of adipic anhydride.



17

To a soln of adipic acid (5 mmol) in 40mL AcOEt/THF (7:1) and 1 mL DMF cooled at 0°C was added dropwise a solution of DCC (5,25 mmol) in 5mL AcOEt for a period of 10min. The mixture stirred at 0°C for further 30min and then left at r.t. overnight. DCU was filtered off and washed with cold AcOEt. Then solvents evaporated under vaccum till to approximately 30 mL volume and refrigerated overnight. The mixture was filtered again and the filtrate crystallized from Et_2O to give the anhydride as white flakes in 72% yield.

Synthesis of Fmoc-S(CH_3) Cysteine (14)



14; X = H
15; X = D

To a soln of cysteine (3mmol) in MeOH was cooled at 0°C, was added dropwise a solution of 2M MeONa in MeOH (6mmol). After the completion of the addition the mixture was stirred for 5min at 0°C followed by dropwise addition of MeI (3,3mmol) in a period of 5min. The thus resulting mixture was stirred at r.t for further1h. Then the solvents were evaporated under vaccum and the residue diluted with 10% Na_2CO_3 in dioxane followed by the addition of Fmoc-OSu (3,1mmol). After the completion of the reaction (TLC) the mixture was extracted with Et_2O and the water phase was acidified by 5% aq. citric acid (pH=4) and extracted twice with AcOEt. The combined organic phases were dried (Na_2SO_4), filtrated and evaporated under vaccum. The resulting residue was subjected to flash column chromatography (FCC) using

$CHCl_3$/MeOH/AcOH=95:5:0,1 as eluant. Fmoc-S(CH_3) Cysteine was finally isolated as foam in 68% yield.
ESI-MS: *m/z*= 358,10 $(M+H)^+$

Synthesis of Fmoc-S(CD_3) Cysteine (15)

To a soln of cysteine (3mmol) in CD_3OD cooled at 0^oC was added dropwise a solution of 2M CD_3ONa in CD_3OD (6mmol). After the completion of the addition the mixture was stirred for 5min at 0^oC, followed by dropwise addition of CD_3I (3,3mmol) in a period of 5min. The thus resulting mixture was stirred at r.t. for further 1h. Then the solvents were evaporated under vaccum and the residue diluted with 10% Na_2CO_3 in dioxane, followed by the addition of Fmoc-OSu (3,1mmol). After completion of the reaction (TLC) the mixture was extracted with Et_2O and the water phase was acidified by 5% aq citric acid (pH=4) and extracted twice with AcOEt. The combined organic phases were dried (Na_2SO_4), filtrated and evaporated under vaccum. The resulting residue was subjected to flash column chromatography (FCC) using $CHCl_3$/MeOH/AcOH=95:5:0,1 as eluant. Fmoc-S(CD_3) Cysteine was finally isolated as foam in 75% yield.
ESI-MS: *m/z*= 361,17 $(M+H)^+$

Solid Phase Dansyl Lysine Tags

331mg (0,080 mmol) of the Tentagel S RAM (0,24mmol/g) resin, placed in 5mL plastic syringe equipped with a frit was swelled using 1mL DCM. Then the Fmoc group was cleaved using 1mL a solution of 25% piperidine in DMF in three cycles for 20min each. The thus resulting resin was washed thrice sequentially with 1mL of each DMF, iPrOH, DMF and iPrOH.. This latter washing procedure is the “standard resin washing protocol” and so will be referred from now on.

The coupling reaction of the resulted resin with Fmoc-Lys(Mtt)-OH was performed using HBTU (2 excess fold) and DIPEA (4 excess fold) in 600ì L DMF for 1h and then a second cycle followed using the same quantities of reagents and 1h reaction time.

The Mtt group was cleaved selectively using 1mL TFA 10% in DCM for 1h. The resulting resin was washed using the standard washing protocol mentioned above.

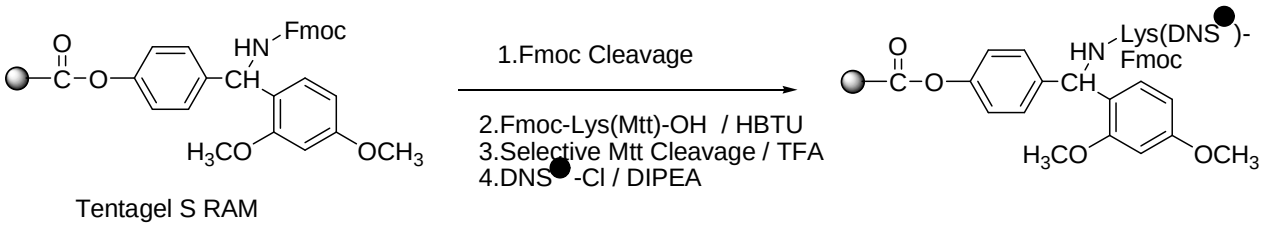
The ε-amino group of Lys was dansylated using Dns-Cl (2 excess fold) (or d6-Dns-Cl) and DIPEA (2,5 excess fold) in 750ì L DCM for 2h, followed by a second 2h reaction cycle with addition of a fresh portion of the same reagents.

The Fmoc protecting group was cleaved as mentioned above. The resulted resin was splitted in two new syringes.

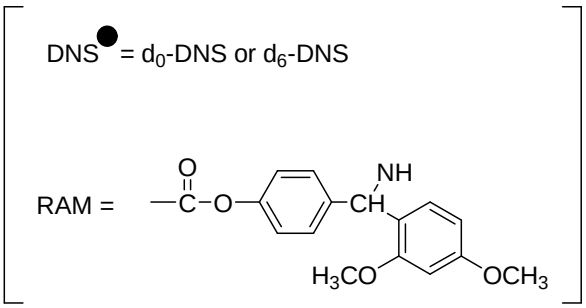
The first one was derivatized by treatment with 2-iodoacetic N-Hydroxysuccinimidyl ester (4 excess fold) and DIPEA (4 excess fold) in 700ì L DMF for 30min. The resulted resin was washed using the standard protocol and stored at 4^{o}C.

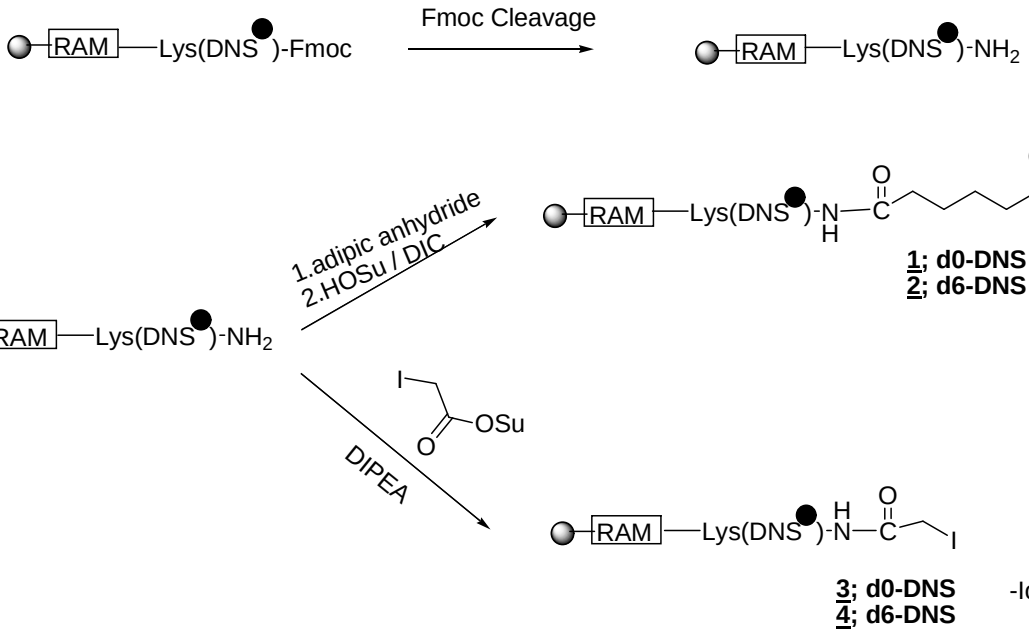
The second one was derivatized by treatment with adipic anhydride (4 excess fold) and DIPEA (4 excess fold) in 700ì L DMF for 1h. The resulted resin was washed as mentioned above and the activation reaction with N-hydroxy succinimide (4,5 excess fold) was carried out for 5h using DIC (3 excess fold) in 700 ì L DMF. The resulted resin was washed using the standard protocol and stored at 4^{o}C.

Dansyl Lysine Tag



Tentagel S RAM





Cysteine-Cysteine Isobaric Tags

306mg (0,074mmol) of the Tentagel S RAM (0,24mmol/g) resin, placed in 5mL plastic syringe equipped with a frit was swelled using 1mL DCM. Then the Fmoc was cleaved using a solution of 25% piperidine in DMF in three cycles for 20min each. The thus resulting resin was washed using the standard resin washing protocol that mentioned before.

Two coupling reactions of the resulted resin were performed using Fmoc-S(CH_3)-Cys-OH (**14**) and Fmoc-S(CD_3)-Cys-OH (**15**). The first resin was developed by coupling firstly the Fmoc-S(CH_3)-Cys-OH (**14**) followed by the coupling with Fmoc-S(CD_3)-Cys-OH (**15**). The second resin was developed by reversing the reagents: Firstly the heavy compound (**15**) and after the light one (**14**).

The coupling reaction of the resulted resin with Fmoc-S(CH_3)-Cys-OH ,resins **5**, **6** (or Fmoc-S(CD_3)-Cys-OH, resins **7**, **8**) was performed using HBTU (2 excess fold) and DIPEA (4 excess fold) in 600ì L DMF for 1h and then a second cycle followed using the same quantities of reagents and 1h reaction time.

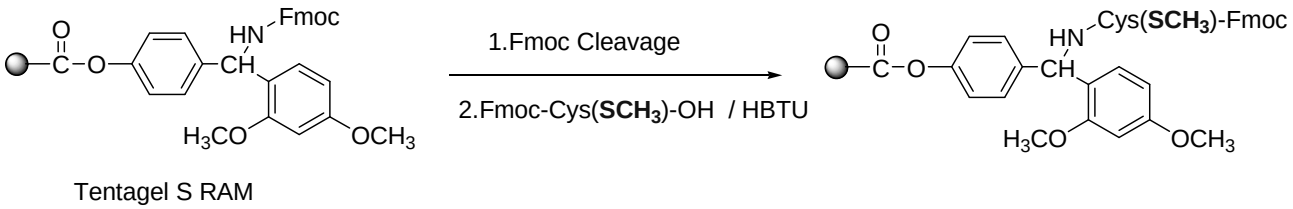
The Fmoc group was cleaved as mentioned above. The second coupling reaction was performed with Fmoc-S(CD_3)-Cys-OH for the synthesis of resins **5, 6** (or Fmoc-S(CH_3)-Cys-OH, resins **7, 8**) using HBTU (2 excess fold) and DIPEA (4 excess fold) in 600ì L DMF for 1h and then a second cycle followed using the same quantities of reagents and 1h reaction time.

The Fmoc protecting group was cleaved as mentioned above. The resulted resins were spitted in two new syringes.

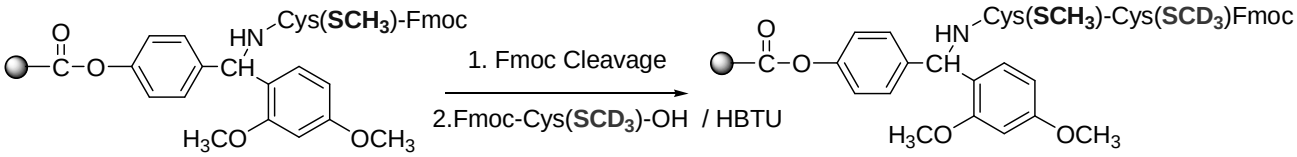
For the synthesis of **6, 8**: the two resins (isobaric to each other) were derivatized by treatment with 2-iodoacetic N-Hydroxysuccinimidyl ester (4 excess fold) and DIPEA (4 excess fold) in 700ì L DMF for 30min. The resulted resin was washed using the standard protocol and stored at 4^oC.

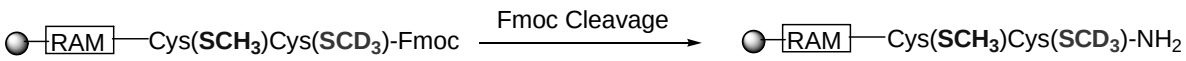
For the synthesis of **5, 7:** the two resins (isobaric to each other) were derivatized by treatment with adipic anhydride (4 excess fold) and DIPEA (4 excess fold) in 700ì L DMF for 1h. The resulted resin was washed as mentioned above and the activation reaction with N-hydroxy succinimide (4,5 excess fold) carried out for 5h using DIC (3 excess fold) in 700 ì L DMF. The resulted resin was washed using the standard protocol and stored at 4^{o}C.

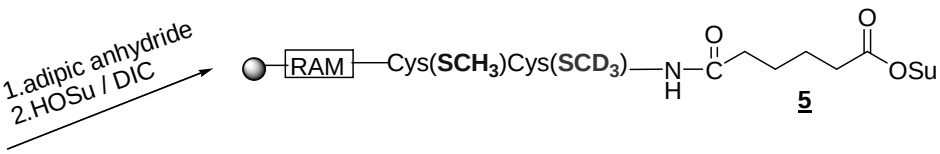
Cysteine Isobaric Tag



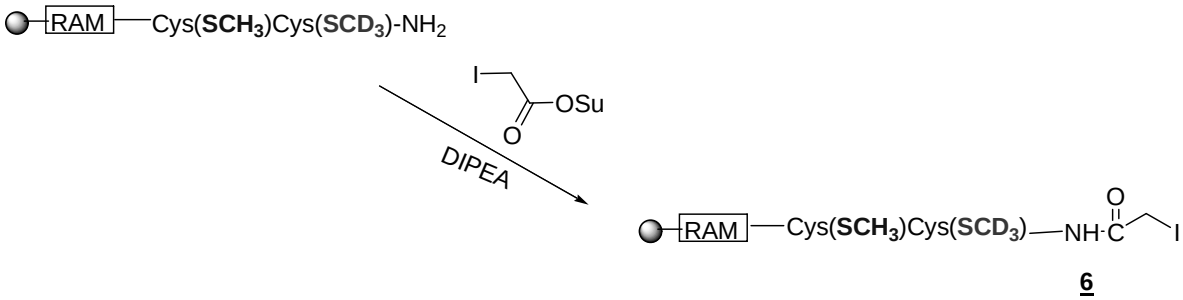
Tentagel S RAM





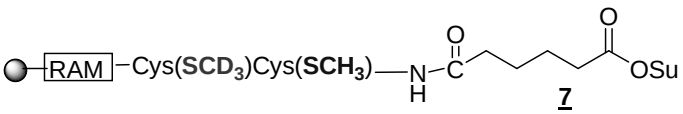


5

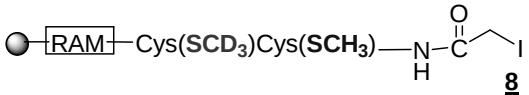


6

Following the same procedure coupling first Cys(**SCD3**)-OH and after Cys(**SCH3**)-OH we synthesised the following isobaric tags which are expected to give interesting results by MS/MS experiments



7



8

Gly-Ala Isobaric Label-Free Tags

The synthesis of the resins was made following the same procedure as the Cys-Cys isobaric resins. Briefly, 354mg (0,085mmol) of the Tentagel S RAM (0,24mmol/g) resin placed in 5mL plastic syringe equipped with a frit was swelled using 1mL DCM. Then the Fmoc was cleaved using a solution of 25% piperidine in DMF in three cycles for 20min each. The thus resulting resin was washed using the standard resin washing protocol that mentioned before.

Two coupling reactions were made using the low cost commercially available Fmoc-Gly-OH and Fmoc-Ala-OH. The first resin was developed by coupling firstly the Fmoc-Gly-OH followed by Fmoc-Ala-OH. The second resin was developed by reversing the reagents: Firstly Fmoc-Ala-OH and after Fmoc-Gly-OH

The coupling reaction of the resulted resin with Fmoc-Gly-OH, resins **16**, **18** (or Fmoc-Ala-OH, resins **17**, **19**) was performed using HBTU (2 excess fold) and DIPEA (4 excess fold) in 600ì L DMF for 1h and then a second cycle followed using the same quantities of reagents and 1h reaction time.

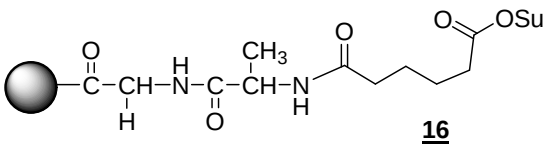
The Fmoc group was cleaved as mentioned above. The second coupling reaction was carried out with Fmoc-Ala-OH for the synthesis of resins **16, 18** (or Fmoc-Ala-OH, resins **17, 19**) using HBTU (2 excess fold) and DIPEA (4 excess fold) in 600ì L DMF for 1h and then a second cycle followed using the same quantities of reagents and 1h reaction time.

The Fmoc protecting group was cleaved as mentioned above. The resulted resins were spitted in two new syringes.

For the synthesis of **18, 19**: the two resins (isobaric to each other) were derivatized by treatment with 2-iodoacetic N-Hydroxysuccinimidyl ester (4 excess fold) and DIPEA

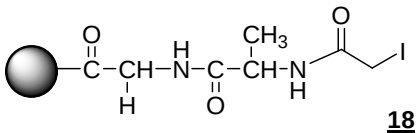
(4 excess fold) in 700ì L DMF for 30min. The resulted resin was washed using the standard protocol and stored at 4 ^{o}C.

For the synthesis of **16, 17:** the two resins (isobaric to each other) were derivatized by treatment with adipic anhydride (4 excess fold) and DIPEA (4 excess fold) in 700ì L DMF for 1h. The resulted resin was washed as mentioned above and the activation reaction with N-hydroxy succinimide (4,5 excess fold) carried out for 5h using DIC (3 excess fold) in 700 ì L DMF. The resulted resin was washed using the standard protocol and stored at 4 ^{o}C.



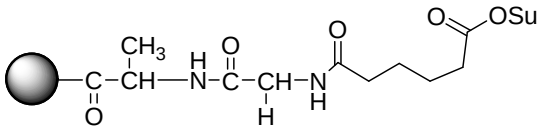
16

Gly-Ala -OSu



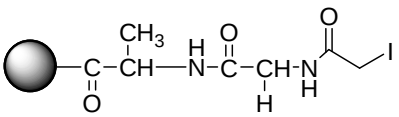
18

Gly-Ala Iodo Tag



Ala-Gly -OSu

17



Ala-Gly Iodo Tag **19**

Derivatization Protocols

Amino selective group reactive resin tags (-OSu Tags, **1**, **2**, **5**, **7**, **16**, **17**)

BSA was digested with a proteolytic enzyme. In case of Trypsin the ratio Trypsin:Protein was 1:100, while in case of Pepsin the ratio was 1:50.

1mg of BSA was dissolved in 1mL solution 50mM NH_4HCO_3 and the stechiometrical quantity of enzyme was added. The digestions (unless otherwise noted) were performed overnight at 37 ^{o}C. The resulted peptides were lyophilized and

dissolved again in 1mL of the desirable solvent (Tris pH=8.8, Tris pH=10 or DMF/DIPEA pH=10).

Approximately 2 mg of each resin (loading 0,24 mmol/g) was weighted in a 1mL vial with Teflon tap. The resins were firstly swelled adding 30ì L of DMF followed by the addition of 200ì L of the peptide solution. The peptides were incubated overnight with the resin in a vortex apparatus in 210 rpm. The resin was decanted in a 1mL syringe equipped with a frit and washed twice subsequently with DMF, ACN/H_2O 6:4, DMF,ACN/H_2O 6:4 and DCM. The derivatized peptides were cleaved from the resin after treatment with 200ì L of a 30% TFA in DCM solution for 1h. The resin was washed twice with ACN/H_2O 6:4 and the filtrate were collected in an eppendorf vial. The organic solvents were evaporated using a speedvac device and the resulted solution was lyophilized. The product was then diluted with 150ì L ACN/H_2O 6:4.

1ì L of the resulted solution was mixed with 5ì L of matrix α-CHCA (0.3% in TFA) and 1ì L of thus resulting mixture was placed on the MALDI plate for the MALDI experiments. For the ESI experiments, the product was diluted with water (1:1) and injected.

Thiol selective group reactive resin tags (-Iodo Tags, **3**, **4**, **7**, **8**, **18**, **19**)

Solutions of bovine serum albumin (BSA), Transferrin and Casein were prepared at concentration of 1mg/mL each in 10mM Tris buffer plus 5mM TCEP. The solutions were heated to 100 ^oC for 5 min to denature the proteins. The solutions were cooled to room temperature and trypsin was added at ratio 1:50 (w/w) in respect to the protein. The solutions were incubated overnight at 37 ^oC for the digestion of the proteins. Tryptic digests were treated again with TCEP 5mM (like above) for the reduction of the disulfide bonds of cysteins. The derivatization and the sample preparation procedure is the same as mentioned above in the case of –OSu Tags.

5

References

5. References

1. Sengupta, N. and Seto, E. *Journal of Cellular Biochemistry*. **2004**. 93, 57-67
2. Villar-Garea, A. and Imhof, A. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. **2006**. 1764, 1932-1939.
3. Haslam, S. M., North, S. J. and Dell, A. *Current Opinion in Structural Biology*. **2006**. 16,584-591.
4. Spickett, C. M., Pitt, A. R., Morrice, N. and Kolch, W. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2006**. 1764, 1823-1841.
5. Chen, F., David, D., Ferrari, A. and Gotz, J. *Current Drug Targets*. **2004**. 5, 503-515
6. Hsu, W., Rosenquist, G. L., Ansari, A. A. and Gershwin, M. E. *Autoimmunity Reviews*. **2005**. 4, 429-435.
7. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. *Analytical and Bioanalytical Chemistry*. **2007**. 389, 1017-1031.
8. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A. and Mann, M. *Nature*. **2003.** 426, 570-574.
9. Broadhead, R., Dawe, H. R., Farr, H., Griffiths, S., Hart, S. R., Portman, N., Shaw, M. K., Ginger, M. L., et al. *Nature*. **2006**. 440, 224-227.
10. Angers, S., Li, T., Yi, X. H., MacCoss, M. J., Moon, R. T. and Zheng, N. *Nature*. **2006**. 443, 590-593.
11. Rose, J.K.C., Bashir, S., Giovannoni, J.J., Jahn, M.M. and Saravanan, R.S. *Plant J*. **2004**, 39, 715-733.
12. Heazlewood, J.L. and Millar, A.H. *Funct. Plant Biol*. **2003**, 30, 471–482.
13. Lopez, M.F. *Electrophoresis*. **2000**, 21, 1082–1093.
14. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. *Science*. **1989**, 246, 64–71.
15. Yamashita, M., Fenn, J.B. *J. Phys. Chem*. **1984**, 88, 4451-4459.
16. Wilm, M., Mann, M. *Anal Chem*. **1996**, 1;68, 1–8.
17. Hillenkamp, F. M., Karas, R.C., Beavis and Chait, B. *Anal. Chem*. **1991**, 63, 1193-1203.
18. Roepstorff, P., Fohlman, J. *Biomed Mass Spectrom*. **1984**, 11,601.
19. Johnson, R.J., Biemann, K. Biomed. *Environ. Mass Spectrom*. **1989**, 18, 945-957.

20. Griffin, P.R., MacCoss, MJ., Eng, J.K., Blevins, R.A., Aaronson, J.S., Yates, J,R. *Rapid Commun Mass Spectrom*. **1995**, 9(15), 1546-51.
21. Berndt, P., Hobohm, U., Langen, H. *Electrophoresis*. **1999**, 20, 3521–3526.
22. Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. *Electrophoresis*. **1999**, 20, 3551-67.
23. Eriksson, J., Chait, B.T., and Fenyo, D. *Anal., Chem*. **2000**, 72, 999–1005.
24. Jensen, O.N., Podtelejnikov, A.V. and Mann, M. *Rapid Commun. Mass Spectrom*. **1996**, *10*, 1371 –1378.
25. Clauser, K.R., Baker, P., Burlingame, A.L. *Anal. Chem*. **1999**, 71, 2871-2882.
26. Jensen, O.N., Podtelejnikov, A.V. and Mann, M. *Anal Chem*. **1997**, 69, 4741-4750.
27. Fanselau, C. *Journal of Chromatography B*. **2007**, 855, 14-20.
28. Yan, W., Chen, S.S. *Briefings in functional genomics and proteomics*. **2005**, 4, 27-38
29. Patton, W.F. Beechem, J.M. *Curr. Opin. Chem. Biol.* **2002**, 6, 63-69
30. Pasa-Tolic, L., Masselon, C., Barry, R. C. et al. *Biotechniques*. **2004**, 37, 621–624, 626–33,636.
31. Tang, K., Page, J. S. and Smith, R. D. *J. Am. Soc. Mass Spectrom*. **2004**, 15, 1416–1423.
32. Chelius, D. and Bondarenko, P. V. *J. Proteome Res*. **2002**, 1, pp. 317–323.
33. Wiener, M. C., Sachs, J. R., Deyanova, E.G. and Yates, N. A. *Anal. Chem*. **2004**, 76, 6085–6096.
34. Gygi, S. P., Rist, B., Gerber, S. A. et al. *Nat. Biotechnol*., **1999**, 17, 994–999.
35. Ross, P.L, Huang, Y.N, Marchese, J.N. et al. *Mol. Cell Proteomics*. **2004**, 3, 1154-1169.
36. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. *Anal Chem*. **2003,**75, 1895–904.
37. Wang, S., Regnier, F.E. J *Chromatogr A*. **2001**, 924,345–57.
38. Ji, J., Chakraborty, A., Geng, M., Zhang, X., Amini, A., Bina, M., et al. *J Chromatogr B*. **2000**, 745,197–210.
39. Che, F.Y., Fricker, L.D. *Anal Chem*. **2002**, 74, 3190–8.
40. Zhang, X., Jin, Q.K., Carr, S.A., Annan, R.S. *Rapid Commun Mass Spectrom*. **2002**, 16, 2325–32.

41. Munchbach, M., Quadroni, M., Miotto,G., James, P. *Anal Chem*. **2000**, 72, 4047–57.
42. Mason, D.E., Liebler, D.C. *J Proteome Res*, **2003**, 2, 265–72.
43. Hoang, V.M., Conrads, T.P., Veenstra, T.D., Blonder, J., Terunuma, A., Vogel, J.C., et al. *J Biomol Tech*. **2003**, 14, 216–23.
44. Hsu, J.L., Huang, S.Y., Chow, N.H., Chen, S.H. *Anal Chem*. **2003**, 75, 6843–52.
45. Panchaud, A.,Hansson, J.,Affolter, M., Bel Rhlid, R., Piu, S.,Moreillon, P. et al. *Mol Cell Proteomics*. **in press**; M700216–MCP200.
46. Cagney, G., Emili, A. *Nat Biotechnol*. **2002**, 20,163–70.
47. Brancia, F.L., Openshaw, M.E., Kumashiro, S. *Rapid Commun Mass Spectrom*. **2002**,16,2255–9.
48. Brancia, F.L., Montgomery, H., Tanaka, K., Kumashiro, S. *Anal Chem*. **2004**, 76, 2748–55.
49. Beardsley, R.L., Reilly, J.P. *J Proteome Res*. **2003,** 2, 15–21.
50. Peters, E.C., Horn, D.M., Tully, D.C., Brock, A. *Rapid Commun Mass Spectrom*. **2001**,15, 2387–92.
51. Hansen, K.C., Schmitt-Ulms, G., Chalkley, R.J., Hirsch, J., Baldwin, M.A., Burlingame, A.L. *Mol Cell Proteomics*. **2003**, 2, 299–314.
52. Li, J., Steen, H., Gygi, S.P. *Mol Cell Proteomics*. **2003**, 2, 1198–204.
53. Oda, Y., Owa, T., Sato, T., Boucher, B., Daniels, S., Yamanaka, H., et al. *Anal Chem*. **2003**, 75, 2159–65.
54. Gartner, C.A., Elias, J.E., Bakalarski, C.E., Gygi, S.P. *J Proteome Res*. **2007**, 6, 1482–91.
55. Sechi, S., Chait, B.T. *Anal Chem*. **1998**, 70, 5150–8.
56. Shen, M., Guo, L., Wallace, A., Fitzner, J., Eisenman, J., Jacobson, E., et al. *Mol Cell Proteomics*. **2003**, 2, 315–24.
57. Sebastiano, R., Citterio, A., Lapadula, M., Righetti, P.G. *Rapid Commun Mass Spectrom*. **2003**, 17, 2380–6.
58. Pasquarello, C., Sanchez, J.C., Hochstrasser, D.F., Corthals, G.L. *Rapid Commun Mass Spectrom*. **2004**,18, 117–27.
59. Olsen, J.V., Andersen, J.R., Nielsen, P.A., Nielsen, M.L., Figeys, D., Mann, M., et al. *Mol Cell Proteomics*. **2004**, 3, 82–92.

60. Nielsen, P.A., Olsen, J.V., Podtelejnikov, A.V., Andersen, J.R., Mann, M., Wisniewski, J.R. *Mol Cell Proteomics*. **2005**, 4, 402–8.
61. Lu, Y., Bottari, P., Turecek, F., Aebersold, R., Gelb, M.H. *Anal Chem.* **2004**, 76, 4104–11.
62. Qiu, Y., Sousa, E.A., Hewick, R.M., Wang, J.H. *Anal Chem*. **2002**, 74, 4969–79.
63. Shi, Y., Xiang, R., Crawford, J.K., Colangelo, C.M., Horvath, C., Wilkins, J.A. *J Proteome Res*. **2004**, 3, 104–11.
64. Zhou, H., Ranish, J.A.,Watts, J.D., Aebersold, R. *Nat Biotechnol*. **2002**, 20, 512–5.
65. Zhang, L., Guo, Y.L, Liu, H.Q. *Journal of Mass Spectrometry*. **2004**, 39, 447-457.
66. Shi, Y. Xiang, R., Horvath, C., Wilkins, J.A. *J. Sep. Sci*. **2005**, 28, 1812-1817.
67. Goodlett, D.R., Keller, A., Watts, J.D., Newitt, R., Yi, E.C., Purvine, S., et al. *Rapid Commun Mass Spectrom*. **2001**, 15, 1214–21.
68. Syka, J.E., Marto, J.A., Bai, D.L., Horning, S., Senko, M.W., Schwartz, J.C., et al. *J Proteome Res*. **2004**, 3, 621–6.
69. Panchaud, A., Guillaume,E., Affolter, M., Robert, F., Moreillon, P., Kussmann, M. *Rapid Commun Mass Spectrom*. **2006**, 20, 1585–94.
70. Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I., Korner, R., Sonksen, C.P., Roepstorff, P. *Rapid Commun Mass Spectrom*. **2000,** 14, 1226–32.
71. Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C. *Anal Chem*. **2001**,73, 2836–42.
72. Miyagi, M., Rao, K.C. *Mass Spectrom Rev*. **2007**, 26, 121–36.
73. Rose, K., Simona, M.G., Offord, R.E., Prior, C.P., Otto, B., Thatcher, D.R. *Biochem J*. **1983**, 215, 273–7.
74. Liu, T., Qian, W.J., Strittmatter, E.F., Camp, D.G., Anderson, G.A., Thrall, B.D., et al. *Anal Chem*. **2004**, 76, 5345–53.
75. Merrifield, R. B. *Journal of the American Chemical Society*.**1963**, 85, 2149.
76. Barlos, K., et al. *Tetrahedron Lett.* **1989**, 30, 3943.
77. Barlos, K., et al. *Tetrahedron Lett*. **1989**, 30, 3947.
78. Addeo, F., Malorni. A., Marino, G., Randazzo, G. *Biomedical Mass Spectrometry*. **1974**, 1, 363-364
79. Park, S.J., Song, J.S., Kim, H.J. *Rapid Comm. Mass Spectrom*. **2005**, 19, 3089-3096.

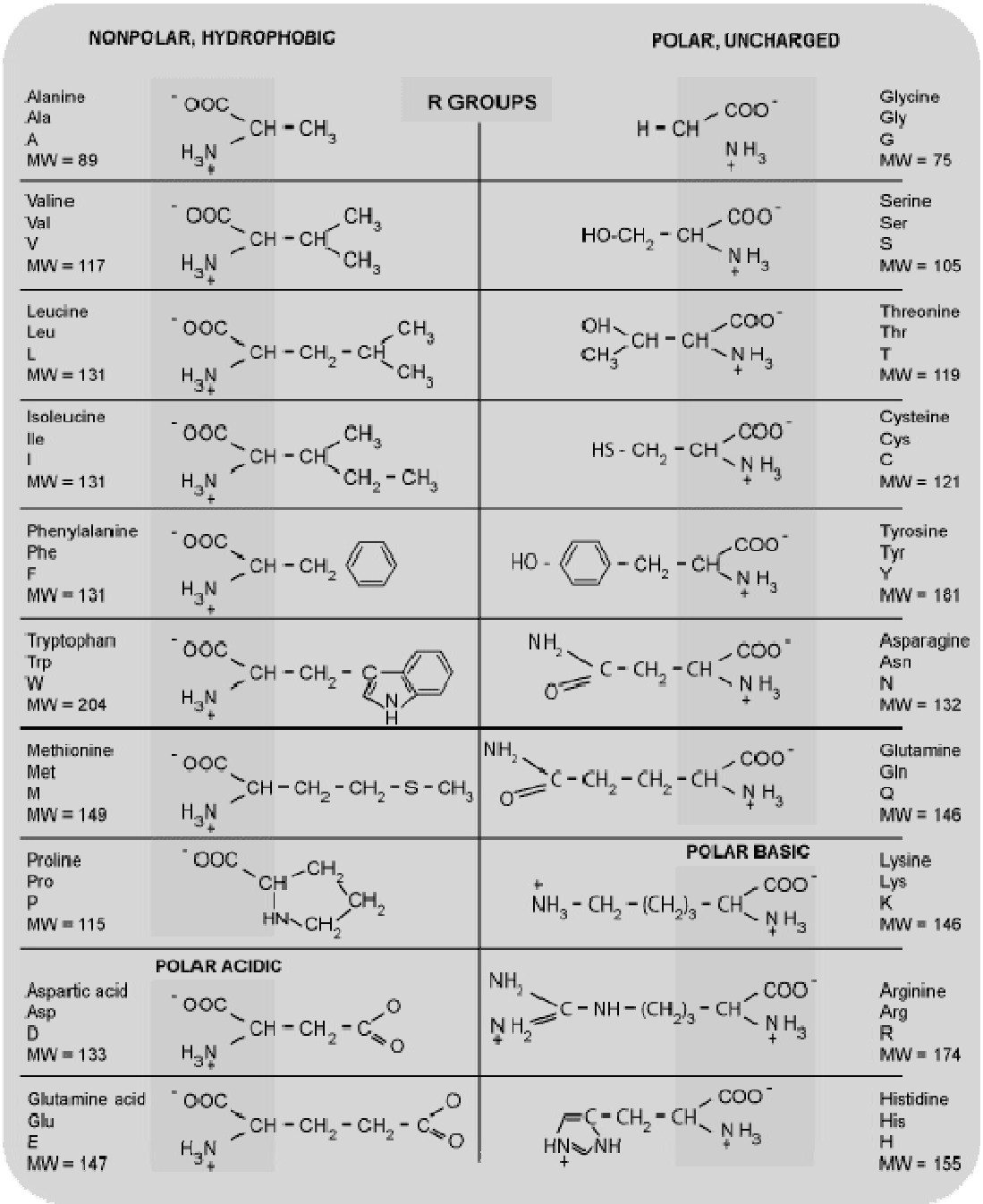
80. Amoresano, A., Monti, G., Citrulli, C., Marino, G. *Rapid Comm Mass Spectrom*. **2006**, 20, 1400-1404.
81. Moschidis, P., Aiello, D., Di Donna, L., Napoli, A., Athanassopoulos, C., Sindona, G.. *Poster presentation,XXII Italian National Conferense*, Italian Chemical Society, Florence,10-15 Sept **2006**.
82. Fussganger, V. *Chem Ber*. **1902**, 976-985.
83. Junk, T., Catallo, W.J., *Chem.Soc.Rev*. **1997,** 26, 401-406.
84. Romera, J.L., Cid, J.M., Trabanco, A.A. *Tetrahedron Letters*. **2004**, 45, 8797-8800.
85. Bizzarri, M., Cavaliere, C., Foglia, P., Guarino, C., Sampieri, R., Lagana' ,A. *Anal Bioanal Chem*. **2008**, 391, 1969-1976.

6

Appendix

6. Appendix

Amino Acid Table



Common Fragment Ions and Neutral Fragments

Common Small Ions	
m/z	composition
15 amu	CH_3
17	OH
18	H_2O
19	H_3O, F
26	C_2H_2, CN
27	C_2H_3
28	C_2H_4, CO, H_2CN
29	C_2H_5, CHO
30	CH_2NH_2
31	CH_3O
33	SH, CH_2F
34	H_2S
35(37)	Cl
36(38)	HCl
39	C_3H_3
41	C_3H_5, C_2H_3N
42	C_3H_6, C_2H_2O, C_2H_4N
43	C_3H_7, CH_3CO
44	C_2H_4O
46	NO_2
56	C_4H_8
57	C_4H_9
60	CH_4CO_2
79(81)	Br
80(82)	HBr
91	C_7H_7
127	I
128	HI

Common Neutral Fragments	
mass loss	composition
1 amu	H
15	CH_3
17	OH
18	H_2O
19	F
20	HF
27	C_2H_3, HCN
28	C_2H_4, CO
30	CH_2O
31	CH_3O
32	CH_4O, S
33	CH_3 + H_2O, HS
33	H_2S
35(37)	Cl
36(38)	HCl
42	C_3H_6, C_2H_2O, C_2H_4N
43	C_3H_7, CH_3CO
44	CO_2O, $CONH_2$
45	C_2H_5O
55	C_4H_7
57	C_4H_9
59	$C_2H_3O_2$
60	$C_2H_4O_2$
64	SO_2
79(81)	Br
80(82)	HBr
127	I
128	HI

Ιθάκη (Ithaka)

Σα βγεις στον πηγαιμό για την Ιθάκη, να εύχεσαι νά 'ναι μακρύς ο δρόμος, γεμάτος περιπέτειες, γεμάτος γνώσεις. Τους Λαιστρυγόνες και τους Κύκλωπας, τον θυμωμένο Ποσειδώνα μη φοβάσαι, τέτοια στον δρόμο σου ποτέ σου δεν θα βρεις, αν μέν' η σκέψης σου υψηλή, αν εκλεκτή συγκίνησης το πνεύμα και το σώμα σου αγγίζει. Τους Λαιστρυγόνες και τους Κύκλωπας, τον άγριο Ποσειδώνα δεν θα συναντήσεις, αν δεν τους κουβανείς μες στην ψυχή σου, αν η ψυχή σου δεν τους στήνει εμπρός σου.

Να εύχεσαι νά 'ναι μακρύς ο δρόμος. Πολλά τα καλοκαιρινά πρωία να είναι που με τι ευχαρίστησι, με τι χαρά θα μπαίνεις σε λιμένας πρωτοειδωμένους· να σταματήσεις σ' εμπορεία Φοινικικά, και τες καλές πραγμάτειες ν' αποκτήσεις, σεντέφια και κοράλλια, κεχριμπάρια κ' έβενους, και ηδονικά μυρωδικά κάθε λογής, όσο μπορείς πιο άφθονα ηδονικά μυρωδικά· σε πόλεις Αιγυπτιακές πολλές να πας, να μάθεις και να μάθεις απ' τους σπουδασμένους.

Πάντα στον νου σου νάχεις την Ιθάκη. Το φθάσιμον εκεί είν' ο προορισμός σου. Αλλά μη βιάζεις το ταξίδι διόλου. Καλλίτερα χρόνια πολλά να διαρκέσει· και γέρος πια ν' αράξεις στο νησί, πλούσιος με όσα κέρδισες στον δρόμο, μη προσδοκώντας πλούτη να σε δώσει η Ιθάκη.

Η Ιθάκη σ' έδωσε το ωραίο ταξίδι. Χωρίς αυτήν δεν θάβγαινες στον δρόμο. Άλλο δεν έχει να σε δώσει πια.

Κι αν πτωχική την βρεις, η Ιθάκη δεν σε γέλασε. Έτσι σοφός που έγινες, με τόση πείρα, ήδη θα το κατάλαβες η Ιθάκες τι σημαίνουν.

***Κ.Π Καβάφης** (1863-1933)*

As you set out for Ithaka hope the voyage is a long one, full of adventure, full of discovery. Laistrygonians and Cyclops, angry Poseidon—don't be afraid of them: you'll never find things like that on your way as long as you keep your thoughts raised high, as long as a rare excitement stirs your spirit and your body. Laistrygonians and Cyclops, wild Poseidon—you won't encounter them unless you bring them along inside your soul, unless your soul sets them up in front of you.

Hope the voyage is a long one.May there be many a summer morning when, with what pleasure, what joy,you come into harbors seen for the first time;may you stop at Phoenician trading stations to buy fine things,mother of pearl and coral, amber and ebony,sensual perfume of every kind as many sensual perfumes as you can;and may you visit many Egyptian citiesto gather stores of knowledge from their scholars.

Keep Ithaka always in your mind.Arriving there is what you are destined for.But do not hurry the journey at all.Better if it lasts for years,so you are old by the time you reach the island,wealthy with all you have gained on the way,not expecting Ithaka to make you rich.

Ithaka gave you the marvelous journey.Without her you would not have set out.She has nothing left to give you now.

And if you find her poor, Ithaka won't have fooled you.Wise as you will have become, so full of experience,you will have understood by then what these Ithakas mean.

Constantine P. Cavafy

Greek Poet, 1863-1933