Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Química

NEW ULTRASONIC-BASED METHODS IN PROTEOMICS

Hugo Miguel Baptista Carreira dos Santos

Lisboa 2010

Ph.D dissertation in Biochemistry by Hugo Miguel Baptista Carreira dos Santos

ISBN: 978-989-20-2273-4 (Hardcover)
Title: New ultrasonic-based methods in proteomics
Author: Santos, Hugo Miguel Baptista Carreira dos
Editor: Santos, Hugo Miguel Baptista Carreira dos
Printed by Lápices 4

Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Química

Hugo Miguel Baptista Carreira dos Santos

NEW ULTRASONIC-BASED METHODS IN PROTEOMICS

Dissertação apresentada para obtenção do Grau de Doutor em Bioquímica pela Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia.

A presente dissertação foi preparada no âmbito do convénio bilateral existente entre a Universidade Nova de Lisboa e a Universidade de Vigo.

Lisboa
2010

Ph.D dissertation in Biochemistry by Hugo Miguel Baptista Carreira dos Santos

Supervised by:

José Luis Capelo Martínez, Ph.D

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa – Portugal and

Science Faculty of Ourense, University of Vigo – Spain

José João Galhardas de Moura, Ph.D

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa – Portugal

Garry Leander Corthals

Turku Centre for Biotechnology, University of Turku and Åbo Akademi University – Finland

To myself…

Acknowledgements

Giving thanks because of the successful ending of a hard work is indeed something nice to be done. And this is because with my doctoral thesis finally completed, I can see the way I have done to get this moment. It has been, it is and I hope it will be a never ending way to wisdom.

This way has had moments of difficulty, moments I needed to overcome alone or most of the times, with the aid of my family, my supervisors and my friends. Joy has been also present during the making of the thesis. For instance, I have had the opportunity to know the way of living of many people of other countries, such as Finland or Spain. This has opened my mind to a better understanding of people and has taught me to be proud of my country and my family.

The acceptance of the manuscripts was always a moment of happiness because is the final act of this incredible history named science.

Below I will thanks to all those whom in a way or another have helped to bring this work to safe port. Therefore I would like to thanks

To Prof. José Luis Capelo, main supervisor of this thesis, and to Carlos Lodeiro for their help guidance through this work, and for his truly friendship. We have passed hard but fruitful moments together, and together we have done great things, such as one book; good research, that has been published in the top-ten of the journals of the analytical and proteomics areas; and also many courses through which we have formed more than 50 researchers and technicians from countries as different as Brazil and Sweeden.

Last but not least, together we have created two research labs, one in Portugal and a new one in Spain. Thanks for all. I will try to follow your counsel: “From your supervisor learn the better, to copy it and to make it best, but also learn the pits, to use them never”.

To Prof. Garry Corthals co-supervisor of this thesis, not only for his support and truly friendship but also for giving me the opportunity to work in his laboratory. I could learn a different but positive way of doing things.

To Prof. José Moura co-supervisor of this thesis, for his help, support and guidance.

To Prof. Mário Diniz for his always useful advices, support and because he was the first researcher who taught me how to do real research.

To "Fundação para a Ciência e Tecnologia de Portugal", for the financial support given to perform this doctoral thesis (grant SRF/BD/38509/2007).

To the New University of Lisbon and in particular to the Chemistry Department of the Faculty of Science and Technology, for giving me the possibility to use its facilities.

To the Turku Centre for Biotechnology – University of Turku and Åbo Akademi for provide me the opportunity to learn and work on their facilities.

To the University of Vigo for giving me the possibility for work on their facilities.

To my parents, Leonor and Armindo, and to my brother Ruben, for being like they are and for their constant support and help. Without their constant presence and perseverance in pushing me to give my best, I had probably never tried to do a doctorate.

To all my laboratory colleagues, especially, to Gonçalo, Marco and Ricardo for their support.

To the group of Garry Corthals, for their kindness and help during my research period in Turku. A special thanks to Petri Kouvonen, Anne Rokka and Anni Vehmas.

To Elisabete, more than a colleague a real friend, for her always useful help, advice and support.

To all the people that in a way or another have contribute to this work and that I had might forgotten of named here.

Finally, a special thanks to all that truly friends that are always with me…

Abstract

The present dissertation is devoted, essentially, to the development of new ultrasonic-based sample preparation methods for mass-spectrometry-based analysis of proteins.

This dissertation is introduced by a brief overview about general aspects in proteomics such as protein identification, quantification and finally, tissue analysis by mass spectrometry. Special attention is given to fast methods for protein identification, as for example those using tools such as microwave energy and ultrasonic energy to speed-up enzymatic digestion. In addition, a general overview of label-based quantification approaches, such as SILAC, iTRAQ, ICAT, or ^{18}O and, label-free methods are presented. Furthermore, stable ^{18}O-isotopic labeling is described in detail. The last topic covered in the first chapter is the analysis of tissue samples by mass spectrometry, a short overview on this subject is done to introduce the reader on the importance of tissue analysis by mass spectrometry.

As far as experimental concerns, different types of ultrasonic devices, such as ultrasonic bath, sonoreactor, ultrasonic probe and multiprobe were assessed in different steps of currents protocols for protein identification through mass spectrometry. Variables such as temperature, sonication time, amplitude and ultrasonic frequency were evaluated in order to identify the optimum conditions. On the first trial, ultrasonic energy was used to accelerate the in-solution enzymatic protein digestion. It was found that the application of the ultrasonic energy (ultrasonic probe and sonoreactor) to a liquid medium containing protein and trypsin speeds protein cleavage from overnight (12h) to 5min.

To make the handling easier and faster, the ultrasonic energy was also successfully applied to the protein reduction and protein alkylation steps. Overall, the time was reduced from 3h to 10min, and from many single cleaning steps to just one for reduction and another for alkylation.

The efficiency of ultrasonic energy to accelerate the sample preparation for protein identification by peptide mass fingerprint was evaluated by comparison with the overnight method.

Once all the experiments were done, it was completed a new sample treatment for the identification of proteins by mass spectrometry. For proteins separated by gel-based approaches the total sample treatment time was reduced from 12-24h to just 20min whilst for proteins separated by off-gel approaches the total time was reduced to just 8min.

Regarding protein quantification, we chose to work with isotopic labeling, as this is one of the methods most suitable for protein quantification by mass spectrometry. In this work proteins separated by gel electrophoresis were inverse labeled with ^{18}O and analysed by MALDI-TOF-MS. The inverse labeling approach, allows to select some peptides that have reproducible loses in different sets of a given entire proteomic workflow. Those selected peptides were used for subsequent protein quantification.

To help to handle in a simple way the data from inverse labeling experiments and to use such data for protein quantification, a bioinformatics tool called "Decision peptide-driven", DPD; was developed. The simultaneous use of the inverse labeling approach to select the peptides suitable for quantification and the DPD software allowed the accurate quantification of targeted proteins. The proposed quantification strategy was successfully compared with a classic quantification done with the ELISA method.

Finally, to get insight into the tissue proteomics, a method was developed to help to distinguish, after on-tissue protein digestion, in a rapid manner, and with the aid of mass spectrometry, the peptides present in a tissue from other tissue components with similar m/z values. The differentiation was achieved by labeling the peptides with ^{18}O. The results obtained shown that the labeling can be done decoupling the steps of protein digestion and peptide labeling. In addition, it has been also demonstrated that ultrasonication can boost the digestion of tissue's proteins in seconds, thus opening for the first time sample treatment of tissues to high throughput.

Resumo

O presente trabalho descreve a aplicação da energia de ultra-sons a diferentes etapas da preparação de amostra para análise de proteínas por espectrometria de massa. A dissertação apresentada encontra-se orientada numa perspectiva de desenvolvimento de novos métodos com aplicações em proteómica.

Em primeiro lugar, apresenta-se de forma geral uma introdução que pretende contextualizar de forma simples os diferentes temas abordados, por exemplo identificação de proteínas, a sua quantificação e a análise directa de proteinas em tecidos biológicos, tudo com recurso a espectrometria de massa. No entanto, temas relacionados com a aceleração de tratamentos de amostra para identificação de proteínas por espectrometria de massa, como por exemplo a utilização de energia de microondas e de ultra-sons serão abordados com algum detalhe.

Tendo em conta a grande importância da quantificação de proteínas para o entendimento dos sistemas biológicos foi efectuada uma pesquisa bibliográfica em temas chave de quantificação de proteínas por espectrometria de massa, nomeadamente, a utilização de marcadores isotópicos como por exemplo SILAC, iTRAC, ICAT e ^{18}O. Este ultima, será abordada de forma mais exaustiva.

O último tema sobre o qual recai a presente dissertação, encontra-se relacionado com a análise directa de proteinas em tecidos biológicos por espectrometria de massa. Nesta parte pretende-se elucidar o leitor acerca da importância da análise de tecidos por espectrometria de massa, bem como esclarecer sobre a forma como se processa a análise de tecidos biológicos por espectrometria de massa.

No que diz respeito à parte experimental da presente dissertação, foram efectuados estudos sobre a aplicabilidade de sistemas de ultra-sons, nomeadamente, banho, sono-reactor, sonda e multi-sonda de ultra-sons a procedimentos normalmente utilizados em proteómica.

Alguns parâmetros, como por exemplo, temperatura, tempo, amplitude e frequência de ultra-sons foram avaliados quanto à performance destes sistemas em protocolos de identificação de proteínas.

Numa primeira aproximação, foi testada a aplicação da energia de ultra-sons para acelerar a digestão enzimática de proteínas em solução. A aplicação desta energia a um meio liquido contendo proteína e enzima produz um efeito de cavitação aumentado os processos de transferência de massa, como consequência, a digestão de proteínas foi reduzida de 12h para 5 min. A energia de ultra-sons, foi aplicada com sucesso às etapas de redução e alquilação de proteínas, que foram reducidas de 3h para apenas 10min.

A capacidade da energia de ultra-sons para acelerar a preparação de amostra para identificação de proteínas por PMF foi avaliada por comparação com o método clássico, verificando-se a redução no total de 12-24h para apenas 20min para proteínas separadas em gel ou 8min para proteínas em soluçao

Relativamente ao tema de quantificação de proteínas com utilização de isótopos e análise por espectrometria de massa, foi efetuado um estudo de quantificação de proteínas separadas por eletroforese em gel. Deste modo as proteínas de interesse foram digeridas e marcadas com ^{18}O, seguindo uma estratégia de marcação inversa e análise por MALDI-TOF-MS.

A marcação inversa permite a identificação de péptidos que apresentam um comportamento reprodutível ao longo das etapas de preparação de amostra e análise. Estes péptidos sao seleccionados e são utilizados para a quantificação. Para facilitar o processo de quantificação, foi desenvolvida uma ferramenta bioinformática chamada "Decision Peptide-Driven", DPD, que se destina ao tratamento de dados provenientes das experiências de marcação inversa e analise por MALDI-TOF-MS. A utilização conjunta da metodologia proposta e do software permite seleccionar os péptidos adequados e como consequência a quantificação é efectuada de forma exacta e rápida.

Os resultados de quantificação obtidos com a abordagem proposta formam comparados com um método clássico de ELISA, utilizado normalmente para a quantificação de proteínas em misturas complexas.

Com o objectivo de diferenciar peptidos de outros componhentes presentes nos tecidos biológicos com recurso a espectrometria de massa, foi estudada a possibilidade da aplicação da marcação isotópica com ^{18}O de péptidos, em tecidos biológicos após digestão

enzymatica. Os resultados obtidos demonstram que quando as duas etapas, digestão e marcação, são efectuadas separadamente verifica-se a incorporação de ^{18}O nos péptidos formados. Esta metodologia apresenta uma grande aplicação para distinguir péptidos endógenos de outros componentes presentes no tecido. Finalemnte, foi demonstrado o potencial da aplicação da energia de ultra-sons na digestão enzimática de tecidos biológicos.

Abbreviations

1D-LC-MS	One dimension liquid chromatography
2D-GE	Two-dimensional gel electrophoresis
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
A&D	Aspirate and dispense
AmBic	Ammonium Bicarbonate
cICAT	Cleavable Isotope-code affinity tags
D&I	Direct and Inverse
DIGE	Diference gel electrophoresis
DNA	Deoxyribonucleic acid
DPD	Decision peptide-driven
DTT	DL-Dithiothreitol
ECDs	Endocrine disruptor compounds,
ESI	Electrospay ionization
FT-ICR- MS	Fourier transform ion cyclotron resonance mass spectrometry
GE	Gel elctrophoresis
HPLC	High pressure liquid chromatography
IAA	Iodoacetamide
ICAT	Isotope-code affinity tags
IEF	Isoelectric focusing
iTRAQ	Isobaric tag for relative and absolute quantification
LC	Liquid chromatography
MALDI-TOF-MS	Matrix Asisted laser/desorption ionization time-of-flight mass spectrometry
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MW	Molecular weight
PBS-Tween	Phosphate Buffered Saline with Polyethylene glycol sorbitan monolaurate

PMF	Peptide mass fingerprint
PNPP	p-nitro-phenylphosphate
RP-HPLC	Reverse phase high performance liquid chromatography
RSD	Relative Standard Deviation
SDS	Sodium dodecyl sulfate
SILAC	Stable isotopic labeling with acids in cell culture
SIM	Selected ion monitoring
TCA	Trichloroacetic acid
Vtg	Vitellogenin
α-CHCA	α-Cyano-4- hydroxycinnamic acid,

Table of Contents

Index of Figures

Index of Tables

Chapter I.

Introduction

Index

I.1 Proteomics, the new biology

Large-scale DNA sequencing has changed biological and biomedical research in a little period of time. As the DNA sequencing technique is spread among the scientific community the list of entire genomes decoded or at least entire sequences decoded, is growing exponentially. This decodification allows getting a better understanding in the insight of the operational parts of an organism [1]. Ultimately, our era has witnessed the answer to difficult biological questions thanks to the fact that DNA sequencing has allowed the understanding of intricate biological processes [2].

Genome, transcriptome, proteome and metabolome are names that generally refer to different compartments of biological complexity within a living organism. The part that studies the proteome is referred to as Proteomics.

The study of proteomics, defined as the entire complement of proteins, including the modifications made to a particular set of proteins, produced by an organism or system, presents many challenges. Nowadays it has been clearly established that a gen or group of gens can support the information to produce one protein or proteins, yet this does not necessarily means that such protein(s) is(are) presented (expressed) in the proteome of a living organism. Furthermore, once a protein has been produced it may suffer a series of modifications called post-translational modifications, and therefore the final protein does not correspond with the original design corresponding to the information contained in the gen or gens. Therefore much of the present attention devoted to proteomics relies in its capability to unravel the secrets of the biological systems whilst overcoming the limits of classical biochemical approaches [2-4].

Originally the word proteomics was associated to the study of a large number of proteins from a given organism or cell by means of two dimensional polyacrylamide gels, 2D-PAGE [5,6]. However, to determine the identity of the proteins was difficult by this method. In the 1990s mass spectrometry (MS) emerged as a powerful tool with which overcome the limitations inherent to 2D-PAGE and protein identification [2]. Nowadays MS-based protein identification by means of peptide mass fingerprint, PMF, or peptide fragment fingerprint [7-9] is broadly adopted, and the transition from analyzing one protein at a time to analyzing proteins in highly complex mixtures has been made [10,11].

Proteomics plays nowadays an important role in biological and biomedical research, because research developed under those areas of knowledge has demonstrated that it is possible to identify and quantify proteins, to determine the presence of post-translational

modifications, to elucidate protein structure and function, and to establish correlations between genes and proteins.

I.2 Sample preparation for mass spectrometry-based protein identification

The identification of proteins is becoming more important to the scientific community due to emerging issues related to proteins critically significant to the society. Many hereditary diseases, cancer and other common illnesses, such as diabetes, can be distinguished on the basis of the expression of certain proteins, known as biomarkers. As a result, disease screening and medical diagnosis take advantage of MS-based technological improvements that allow protein identification [12,13].

Gel-based and gel-free are the two main approaches used in proteomics research. The first one is based in the resolving power, semi-quantitative nature and instant visualization of proteins separated by gel electrophoresis in the first or in the second dimension. However, gel-based approaches lack in reproducibility, are expensive in the second dimension, and are considered labor intensive [14,15]. Gel-free approaches are mainly based in the coupling of high performance liquid chromatography, HPLC, and mass spectrometry. The proteins are digested into peptides either off- or on-line, and then the peptides are separated by HPLC. The identification of the proteins is done through the information retrieved from the peptides using mass spectrometry either off-line using matrix assisted laser desorption ionization, MALDI, or on-line using electrospray ionization, ESI [16].

General aspects of sample preparation steps for both, gel-based and gel-free proteomics are described on the following sections.

I.2.1 In-gel protein digestion

This is by far the most time-consuming and tedious approach. First, proteins must be separated and purified. Usually, a complex protein mixture is separated by 2D-GE but first, depending on the sample, it may be necessary to concentrate and eliminate interfering substances before electrophoresis. In such a case, protein precipitation with a mixture of trichloroacetic acid (TCA) / acetone is very useful to remove salts and detergents [17]. To guarantee reproducibility in the separation process, protein extracts must previously be denatured in urea/thiourea. In addition, thiourea increases the solubility of hydrophobic

proteins [18]. To run the first dimension of 2D-GE, samples are loaded into isoelectric focusing (IEF) gel strips containing ampholytes.

When a protein is placed in a medium with a pH gradient and then is subjected to an electric field, it will initially move toward the electrode with the opposite charge. During migration through the pH gradient, the protein will either pick up or lose protons. As it migrates, its net charge and mobility will decrease and the protein will slow down. Finally, the protein will arrive in the pH gradient at the point equal to its pI. There, being uncharged, it will stop migrating. Once the first dimension is completed, and to prevent further re-oxidation, protein-cystine residues must be reduced using dithiothreitol (DTT) and the resulting cysteines blocked with iodoacetamide (IAA).

After focusing, IEF strips are soaked in SDS to prepare proteins for separation by molecular weight (MW) in the second dimension. When the separation is complete, proteins are visualized in the gel by staining with Coomassie Blue, fluorescent dyes, or with MS-compatible silver staining [19]. After protein separation and visualization, protein spots are excised and de-stained and usually digested with specific proteases as trypsin or Lys-C. The resulting pool of peptides are extracted from the gel peace and analysed by mass spectrometry with either ESI or MALDI as ionization technique. The resulting peptide profiling is matched against the theoretical masses obtained from the *in silico* of all protein amino acid sequences in the database. Then the protein in the database is ranked according to the number of peptides matching their sequence within a given mass error tolerance [20].

The in-gel digestion protocol presents several drawbacks, namely (i) trapping protein substrates in the gel makes some peptide bonds inaccessible to the enzyme, and (ii) not all the peptides produced during digestion can diffuse freely from the gel.

I.2.2 In-solution protein digestion

In-solution protein digestion, usually applied to whole-protein extracts is the base of shotgun proteomics but is also widely used in the analysis of purified proteins. There are key steps related to sample preparation issues that guaranty the success of the in-solution protein digestion.

It is widely known that proteins are very heterogeneous in terms of physical and chemical properties. In addition the dynamic range of protein expression is also very broad and for that reason some general rules should be taken into consideration. For example, the presence of most abundant proteins that usually are easier to identify, might interfere with

the detection of less abundant proteins. So to facilitate the identification of less abundant proteins, some fractionation (e.g., organelle purification) or depletion to separate some proteins from another is needed [21].

As a general rule, protein extracts are first precipitated using cold acetone or TCA and then proteins are solubilizeded/denaturated in order to facilitate the action of proteolytic enzymes. To achieve this goal, several buffers have been described in literature, including some ones for specific types of proteins. As a general role, protein solubilization and denaturation are achieved by the addition of chaotropic agents such as urea or guanidine hydrochloride in order to break the intra-molecular forces that maintain secondary and tertiary structure of the proteins. Finally, to prevent protein renaturation before enzymatic cleavage, reduction and alkylation of protein disulfide bridges are frequently done using DTT and IAA [22].

As previously mentioned, trypsin is one of the enzymes used for protein digestion. However, care must be taken, when high urea concentration is used. Normally, trypsin retains most of its activity in 2M urea, 2M guanidine HCl or 0.1% SDS. Protein digestion is usually performed for 12–24 hours, due to protein heterogeneity in samples. Finally, the reaction is stopped by adding an acid, such as trifluoroacetic or formic, to a final pH of 2–3. Then, the digest can be analysed by bi-dimensional chromatography coupled to tandem mass spectrometry [22].

I.3 Accelerating protein digestion

In the latest years it has been reported in the literature promising results on the acceleration of enzymatic protein digestion since is one of the most time consuming steps needed for protein identification. The first attempts reported make use of basic knowledge in enzymology, such as the role of enzyme concentration and the temperature influence on the rate of the enzymatic reaction [23]. However, other approaches to proteomics sample preparation have been done in order to accelerate protein digestion. The application of the microwave energy and ultrasonic energy, the two most useful tools, will be commented below

I.3.1 Microwave energy

Microwave-assisted protein enzymatic digestion under controlled microwave irradiation was first reported in 2002 [24], and have been used for both in-solution [24] and in-gel digestion [25]. Using this methodology, digestion times required to get a complete digestion are lowered from 12-24h up to 20 min.

Critical parameters, as irradiation time, temperature and power require careful optimization in order to maintain the benefits of microwave energy. For example, radiation power should be maintained below 30% of the total nominal power of the microwave oven; otherwise, temperature control becomes difficult [22].

Microwaves are known to catalyze reactions by inducing molecular perturbation by a stimulation of ionic diffusion, and also, by enhancement of dipole rotation without causing any rearrangement of molecular structures. This mechanism differs from conventional heating due to the dipole rotation that constitutes an alternative efficient form of molecular agitation. It is due to this added molecular agitation that increased molecular catalysis is thought to occur.

I.3.2 Ultrasonic energy

I.3.2.1 Basics on ultrasonic energy

When ultrasonic waves cross through a liquid media, an effect known as cavitation occurs [26,27]. The name cavitation defines a physical process by which numerous tiny gas bubbles are generated. Those bubbles, grow, oscillate, split and finally implode, in such a way, that they can be considered as micro-reactors in which temperatures near to 5000ºC and pressures of 1000atm are reached.

Cavitation causes physical phenomena, such as pitting and mechanical erosion of solids, including particle rupture, leading to smaller particle size. In addition, chemical radicals are formed into the liquid media, which produces the oxidation of chemical species. As an example, energy oxidative radicals, such as hydroxyl radical or chemical compounds, such as hydrogen peroxide are formed when ultrasonic power is applied to water.

Nowadays ultrasonication can be applied in two main ways with different devices, direct ultrasonication can be applied with the ultrasonic probe whilst indirect sonication can be applied with the common ultrasonic bath or with the sonoreactor. Figure I.1 shows three examples of ultrasonic devices.

A **B** **C**

Figure I. 1. Example of ultrasonic devices: A- ultrasonic bath; B- ultrasonic probe and C- Sonoreactor

As a general role the influence of the following variables should it be studied when the optimization of a methodology of sample treatment is done through ultrasonication: (i) particle size, if a solid is studied; (ii) reagent(s) used in the treatment; (iii) sample volume; (iv) sonication time; (v) temperature; (vi) frequency of the ultrasonic energy; (vii) for the case of an ultrasonic bath, the position in which the sample container is situated inside the bath (vertical and horizontal position) and finally (viii) amplitude of sonication.

I.3.2.2 Indirect sonication: the ultrasonic bath and the sonoreactor

Indirect sonication means that the ultrasonic waves need to cross the wall of the sample container. The ultrasonic bath is not a powerful tool; the irradiation power given by a common ultrasonic bath is comprised between 1 and 5 Wcm^{-2}. When used for analytical tasks, the ultrasonic bath lacks in reproducibility. Finding the highest intensity place of sonication inside an ultrasonic bath is always a critical issue, to do so, the so-called "aluminum foil test", which is done to find the best place inside the bath in terms of sonication intensity [28].

The sonoreactor and their equivalent, the cup horn, can be compared to high intensity ultrasonic water baths. Samples can be processed in sealed tubes or vials eliminating aerosols and cross-contamination. Those devices are ideal for samples, such as

sterile or dangerous pathogenic ones. In the cup horn, the titanium probe is held within an acrylic cup filled with water. Samples are placed within the cup, above the probe. The cavitation produced in the immersed samples is higher than the one given by an ultrasonic bath but it is lower than the cavitation produced by direct immersion of the ultrasonic probe into the solution. Both systems allow refrigeration.

I.3.2.3 Direct sonication: the ultrasonic probe

The ultrasonic probe is directly immersed in the sample, giving direct in-sample sonication. Two are the main differences when comparing the ultrasonic probe with the ultrasonic bath. Firstly, the ultrasonic probe is immersed directly into the solution, where the sonication takes place, and secondly, the ultrasonic power provided by the probe is at least up to 100 times greater than the one supplied by the bath. Those major differences make each system devoted for a different set of applications. The probe is a powerful system for the solid–liquid extraction of analytes that can be extracted but can also be degraded. There are dedicated probes for a given range of volumes. It should be stressed that the amplitude control of the probes allows the ultrasonic vibrations at the probe tip to be set to any desired level. However, to achieve cavitation, normally it is not necessary to use high amplitude levels; otherwise the probe will deteriorate rapidly. Temperature is another factor that must be controlled. As the ultrasound is delivered into the solution, a slow but constant increase in the bulk temperature is achieved and, at one point, the physical characteristics of the liquid media changes so that a decoupling of the probe can occur and no more cavitation is achieved. At this point the procedure must be stopped and the solution refrigerate. If long sonication times are needed the “pulse” mode is recommended.

Probes are usually made of titanium alloy (titanium probes) and are thermo-resistant, can be treated in autoclaves and are resistant to corrosive media. The sample volume to be treated along with the sample type is crucial to determine the selection of unit and the type of probe. It must be always borne in mind that the higher the amplitude provided by the probe the more intense is the sonication.

I.3.2.4 Accelerating protein digestion with ultrasonic energy

Ultrasonic-assisted enzymatic digestion was first introduced in 2005 [29].This method was successfully applied to the in solution and in-gel protein digestion, thus effectively reducing the digestion time from overnight (12 h) to less than 5min [30,31]. In the in-solution protein digestion, the procedure entails sonication of small volumes of sample, typically 20–50μL while the trypsin digestion proceeds. The ultrasonic treatment probably boosts enzyme-substrate kinetics by the enhancement of mass transfers processes in the solution. In the in-gel protein digestion, slides of gel containing protein can be submitted to the same procedure: once excised, the gel piece is placed in an Eppendorf cup and a small amount of a buffer containing the enzyme is added. The liquid jets produced by the ultrasonication act as micro-syringes [32], delivering the enzyme into the gel and making the protein digestion faster. The mechanical erosion of the gel surface caused by the cavitation associated with ultrasonication enhances peptide release from the gel.

The recent advances in ultrasonic energy performance have led to the development of new and powerful devices, whose many possible applications has only become recently to be known [26,33].

I.4 Beyond protein identification - quantitative proteomics

Whilst no long ago qualitative protein identification was considered enough to describe a biological system, at the present time protein identification and quantification have become mandatory to address the same aims. All modern biosciences are given special attention to the changes that occur in the proteome as important information can be retrieved that may be linked to important medical and biological processes. The changes in the expressed proteins in a given type of cells, tissue or organism need to be detected with the highest accuracy and precision possible. These changes can be linked to cellular and tissue localization, to signaling cascades, and to changes due to disease or drug treatment.

The aim of quantitative proteomics is to obtain quantitative information about all proteins in a sample. Rather than just providing lists of proteins identified in a certain sample, quantitative proteomics yields information about differences between samples. For that purpose, quantitative proteomics was based essentially in 2D gel electrophoresis and in mass spectrometry. The gel-based method has been the method of choice for decades, and is still widely used, however, due to the extraordinary development of mass spectrometry-

based technologies and the development of different types of stable isotopic labeling reagents and label-free approaches, mass spectrometry-based quantitative proteomics is, nowadays, the ultimate technology used to quantify proteins. In the next sections we will briefly describe the above mentioned methods for protein quantification.

I.5 Gel-based quantification methods

Native or denatured proteins can be separated through GE in the first or in the second dimension. Once the proteins have been separated they are stained with different reagents such as coomassie blue, silver nitrate or fluorescent dyes.

A general method to quantify proteins separated by GE is densitometry. Densitometry is the quantitative measurement of optical density in light-sensitive materials. Optical density is expressed as the number of dark spots in a given area.

Gel densitometry is widely used for protein expression studies; however some drawbacks have been reported. For example, the low linearity ranges of work, usually of 1-2 orders of magnitude [34]. Nevertheless, and depending on the reagent used for staining, the order of magnitude can be varied, as it is showed in table I.1.

Table I. 1. Protein staining for gel electrophoresis-based quantification of proteins [35].

Protein Stain	Lower Limit of Sensitivity	Linear Range	Cost	Imaging System Requirements	MS Compatibility
Coomassie Blue R-250 stain	10–25 ng	2 orders of magnitude	+	Densitometer	+++++
Bio-Safe Commassie G-250 stain	5–10 ng	2 orders of magnitude	++	Densitometer	+++++
Silver stain kit - Merril method	0.5–1 ng	1 order of magnitude	+++	Densitometer	Not compatible
Silver Stain Plus kit	0.5–1 ng	1 order of magnitude	+++	Densitometer	++
Dodeca silver stain	0.5–1 ng	1 order of magnitude	+++	Densitometer	+++
SYPRO Ruby protein gel stain	1–10 ng	3 orders of magnitude	+++++	Fluorescent	++++
Flamingo fluorescent gel stain	0.25–0.5 ng	>3 orders of magnitude	++++	Fluorescent	+++++

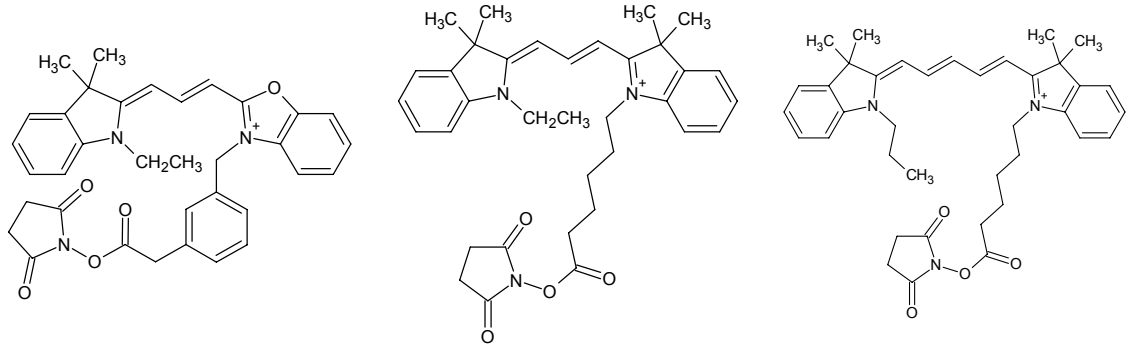
Another drawback is the low reproducibility of the separations done by gel electrophoresis in the second dimension. 2D-gel electrophoresis it requires to be done by skilled personnel otherwise many problems are encountered when the same sample is run several times.

I.5.1 Differential gel electrophoresis

Modern gel electrophoresis research offers powerful software-based image analysis tools primarily to analyze biomarkers by quantifying individual, as well as showing the separation between one or more protein spots on a scanned image of a 2D-GE.

Difference gel electrophoresis - DIGE can be used to highlight differences in the spot patterns. This quantification technique is done on intact proteins, and the differential expression determination is based on fluorescence.

In this technique, three different fluorescent labels (e.g. Cy2, Cy3, and Cy5, see figure I.2) with different absorbance and emission characteristics, are used to covalently modify the amino group of lysines in proteins via an amide linkage.



NSH-Cy2 **NSH-Cy3** **NSH-Cy5**

Figure I. 2. N-hydroxy-succinimidyl esters of Cy2, propyl Cy3 and methyl Cy5 used for DIGE labeling of lysine residues [37].

In a typical protocol, the control and target samples are separately labeled using different dyes (e.g., Cy3 and Cy5, respectively), while a mixture consisting of an equal amount of the control and target samples is labeled with Cy2. The labeled samples are combined and run in a single 2D gel to allow better spot matching and minimize gel-to-gel variations [36]. However, DIGE has some drawbacks that come from the 2D-GE, and fluorescence analysis.

The 2D-GE itself does not generally allow the resolution of proteins with high (150kDa) or low (10kDa) molecular weights, or very basic or hydrophobic proteins.

The fluorescence analysis itself comes with some problems such as high background, the detection of signals from non-protein sources and the overlap of signals from different fluorophores [37].

I.6 Mass-spectrometry based protein quantification

Proteomics aims to analyze as much proteins as possible in the same experiment. Until recently, mass spectrometry was used to characterize complex mixtures of proteins in a qualitative approach.

Mass spectrometry is nowadays a powerful tool for protein quantification, both in a relative or absolute manner, with a continuously growing interest among the research community, as it is showed by the large amount of reviews published regarding this item in the last years, 183 since 2005 (key searching words: protein quantification and mass spectrometry; searching engine: scopus). Those reviews cover a broad area of research activities, including proteomics, genomics and medical care or food chemistry.

The quantification of proteins can be done following two main ways. The first approach is named label-free quantification of proteins and it is based on the measurement of the area of the mass spectrometry peaks belonging to the specie of interest. Theoretically, this area is directly proportional to the concentration of the analyte of interest in a given range.

The second approach it takes advantage of the so called isotopic labeling; the basis of which is the increment in the mass of any chemical specie when one atom is substituted by one heavier isotope or when a chemical tag is added to the molecule of interest. The key of this method is that the mass increment can be easily followed through the mass spectrum and in some cases the intensity of the mass peak of interest can be linked to the concentration of the analyte studied.

I.6.1 Label-free methods for protein quantification

Label-free quantification seems at first to be “easy”, requiring no labeling steps, only LC-MS or LC-MS/MS. It has the advantage of eliminating the need for expensive labeling reagents, and does not require time-consuming steps as the labeling approach. In

addition, those steps can lead to irreproducibility and loss of target peptides. It also eliminates the need for the synthesis of expensive isotopically labeled reference peptides. However in this methodology there is a greater need for reducing potential interferences that could lead to suppression effects. This leads to an increased need for high-resolution or multidimensional chromatography to reduce suppression effects and to allow the detection of low-abundance proteins. In addition, there is the challenge of normalizing the data so that accurate quantification can be done across multiple samples and multiple analyses. Two approaches are currently being used to compare samples: normalizing and aligning the chromatograms (and using the peptide MS data), or using the MS and/or MS/MS spectra to normalize the data [38].

Label free methods are far of being fully accepted among the scientific community since a number of serious drawbacks seem to be difficult to overcome. As an example in a 1D-LC-MS it has been reported that only 66% of peptides which were present in one analysis were also present in the second LC/MS run, and that ten replicate analyses were needed before no new peptides were detected [39]. Furthermore, label-free approaches do not allow for sample multiplexing neither can avoid matrix-dependant suppression effects [38,40]. For the aforementioned reasons label-free techniques are generally considered inferior in their quantification accuracy when compared to methods relying on stable isotopes. Furthermore the accuracy and linearity of label free techniques is still in question [41].

I.6.2 Methods based on isotopic labeling

I.6.2.1 Monitoring

Quantification can be done by comparing targeted ions from labeled and unlabeled standards with respect to the peak heights or areas of labeled versus unlabeled analytes. Quantification of analytes based on isotopically labeled internal standards was first reported for GC-MS [42]. In selected ion monitoring, SIM, instead of scanning all of the possible m/z values, only selected ions are measured. This approach can improve the limits of detection for an analyte by several orders of magnitude.

The advent of tandem mass spectrometry has allowed the monitorization of parent/fragment ions. Thus, multiple reaction monitoring, MRM, has been regularly done in triple-quadrupoles for more than 30 years [43]. In a triple-quadrupole mass spectrometer,

high sensitivity and specificity are achieved by only allowing a selected peptide to pass through the first quadrupole and enter the collision cell (Q2). Inside the collision cell, the peptide dissociates into fragments specific to the amino acid sequence of the precursor peptide. A second stage of specificity is added in the second MS (Q3), and only a specific fragment is allowed to pass through and strike the detector. By repeatedly cycling through a list of selected reaction monitoring, SRM, ion pairs associated with a set of specific retention times, multiple peptides can be targeted in a single multiple reaction monitoring (MRM) experiment.

The use of MRM technique for the absolute quantification (AQUA) of proteins was first introduced by Steven Gygi in 2003 [44]. Absolute quantification workflows include stable isotope labeled (^{2}H, ^{13}C, or ^{15}N) versions of the targeted peptides. The endogenous and isotopically labeled peptides co-elute (or nearly co-elute in the case of deuterated standards) from reversed-phase high performance liquid chromatography (HPLC) columns, and will behave identically in the mass spectrometer with the exception of the introduced mass shift. Both versions are monitored and quantified, and their concentrations are determined from the relative responses (peak heights or peak areas) of the spiked-in compound to the target analyte [38]. Normally this approach is limited to a reduced number of proteins because suitable internal standards need to be identified and synthesized. The internal standard, because it is present in every sample, means that more accurate expression ratios can be calculated. It is not possible to calculate an accurate "treated *vs.* control" expression ratio if there is no peptide detected in the control sample. With an internal standard, if no signal is detected in the sample, and the internal standard is detected, then you can be certain that the level of that peptide in the sample is below the detection limit.

For an experiment with multiple treatments, the use of an internal standard ensures that all of the treatments can be compared with each other (e.g. treatment 1 *vs.* control, treatment 2 *vs.* control, allows the calculation of treatment1 *vs.* treatment 2).

In label-free quantification, the control peptide may not always be present or may not be detected. The internal standard also corrects for suppression effects or irreproducibility in sample processing, at least when these problems occur after the point where the standard is added to the sample. Because the internal standard is added after the digestion step, this method of quantification cannot correct for variable and unpredictable losses that occur during the sample digestion/processing steps that occur prior to the addition of the reference standards [45].

I.6.2.2 SILAC Labeling

SILAC is the acronym of stable isotope labeling with acids in cell culture. In this method the labeling of the proteins is done *in vivo*, since the incorporation of the normal or heavy amino acids is done during the grown of two cell populations, that are identical except that one of them is make to grown in a medium that contains a form of amino acids without heavy isotopes added whilst the other is make to grown in a medium that contains such amino acids [46,47]. These "heavy" amino acids have been prepared with stable isotopic nuclei, such as ^{2}H, ^{13}C and ^{15}N. The basis of this approach relies in that when the labeled analog of an amino acid is supplied to cell in culture it is incorporated into the new proteins that are synthesized during the cell growing. The amino acids generally involved in this labeling are arginine, leucine and lysine. The main advantage is that the labeling is uniformly done in every protein. As main disadvantages may be cited that sample complexity is not reduced, it is no applicable to human samples, and that arginine can be inter-converted to proline [38].

I.6.2.3 ICAT Labeling

ICAT is the acronym of isotope-coded affinity tags. It was first introduced by Gygi and Aebersold in 1999 [48].The reagent consists of three elements as it is showed in figure I.3.

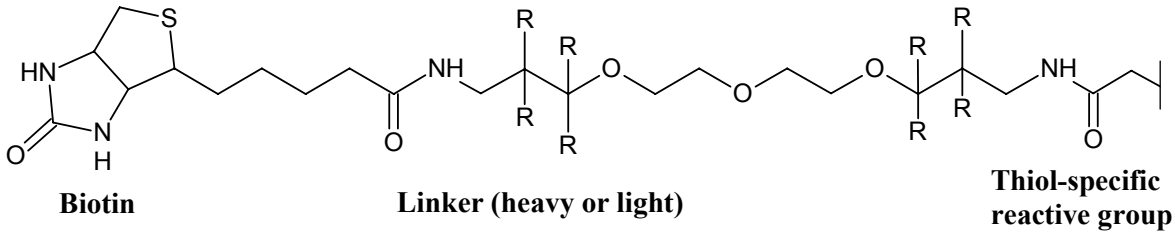


Figure I. 3. Chemical structure of the original ICAT label. Heavy reagent, d8-ICAT (R-deuterium); Light reagent, d0-ICAT (R-hydrogen) [38].

The original ICAT reagent is composed by a thiol reactive group, iodoacetamide, a spacer that contains either eight oxygen atoms (light) or eight deuterium atoms (heavy) and an affinity tag, biotin, this simplifies the analysis of the ICAT reagent-labeled peptides by

selecting and concentrating the cysteine-containing peptides, thereby reducing the complexity of the peptide mixture.

The cleavable ICAT reagent (cICAT) was introduced in 2003 [49,50]. This new version of ICAT retains the main structure of the original ICAT, but the oxygen/deuterium atoms are substituted by nine ^{12}C atoms (light) and nine ^{13}C atoms in the heavy version.

The heavy reagent with nine ^{13}C atoms is +9 Daltons heavier than the light reagent. Mass spectrometric comparison of peptides labeled with heavy and light reagents provides a ratio of the concentration of the proteins of interest. This new reagent eliminates the potential confusion of a double ICAT label and an oxidation, both of which would have lead to a +16Da mass shift. In addition deuterium causes a slight shift in retention time in RP-HPLC, with the heavy form eluting slightly earlier than the light form. Furthermore, cleavage of the biotin moiety after affinity purification and before the MS analysis helps to improve the quality of the CID spectra, leading to the identification of a larger number of proteins.

The labeling occurs at the protein level and it is a chemical labeling that it takes place in the cysteines, which is its main drawback, since only peptides with cysteines are labeled, yet this means that, in comparison with the SILAC method, sample complexity may be reduced. This is because cysteines are only the 1.42% of all amino acids [51]. It also makes impossible to detect changes in the 20% of proteins that do not contain cysteine residues [52]. Two disadvantages to be mentioned are that side reactions with methionine residues may occur and that only two labels are available, meaning only two samples may be compared.

I.6.2.4 iTRAQ labeling

The SILAC and ICAT methods allow only the comparison of two treatments in a single analysis. The need for the comparison of larger number of treatments led to the development of the 4 or 8 plex iTRAQ which may be used to compare up to four or eight samples in a single analysis. The iTRAQ technique was first described by Ross *et al.* in 2004 [53] and is an isobaric tagging compound consisting of a reporter group (variable mass of 114-117Da or 113-121Da) a balance group at lysine side chains and at peptide N-termini.

During the initial MS scan, labeled peptides appear as a single peak due to the isobaric masses, and in MS/MS the label releases the reporter groups as single charged

ions. The isobaric nature of the iTRAQ-labeled peptides allows the signal from all peptides to be summed in both MS and MS/MS modes thus enhancing the sensitivity of detection.

I.7 Isotopic labeling of proteins with ^{18}O

The use of ^{18}O as a label in isotopic studies in biochemistry can be trace back to the work done by Boyer *et al.* in 1956 [54]. However, the use of ^{18}O-labeling for protein quantification and protein differential expression has recently recalled the attention in research due to its almost perfect characteristics. Enzymatic labeling with ^{18}O-water is easy to do and all the peptides formed during the protein cleavage are labeled. The resulting mass shift from ^{18}O incorporation does not alter the chromatographic separation or the ionization efficiency of the labeled peptides. The characteristics, variables and applications of ^{18}O-labeling will be commented below.

I.7.1 Chemical reactions involved in the ^{18}O labeling of proteins

It is generally agreed that the ^{18}O-isotopic labeling of peptides can be considered as consisting in two different chemical reactions. The first one is an amide bond cleavage whilst the second reaction is a carboxyl oxygen exchange. Both chemical reactions can be written as follows [55]:

(i) First reaction: amide bond cleavage

$$RC^{16}ONHR' + H2^{18}O \rightarrow [RC^{16}O^{18}O]^{-} + [H_3NR\_]$$

(ii) second reaction: carboxyl oxygen exchange

$$[RC^{16}O^{18}O]^{-} + H_2{}^{18}O \rightarrow [RC^{18}O^{18}O]^{-} + H_2{}^{16}O$$

As may be seen peptides incorporate one ^{18}O in the first reaction, during the enzymatic cleavage. If this reaction is done in pure ^{18}O-water (normally 95% reach in $H_2{}^{18}O$) then it can be considered that almost 100% of the peptides will be labeled with one ^{18}O. If the reaction is done in mixed $^{16}O/^{18}O$-water, then the labeling ratio will be proportional to the percentage of $H_2{}^{18}O$ in solution: the higher the amount of $H_2{}^{18}O$ the higher the percentage of peptides labeled. This is the reason why accuracy is only guaranteed when the labeling is done in pure $H_2{}^{18}O$ (generally 95%).

In any case, the labeling cannot be studied without taking into consideration the second reaction, the carboxyl oxygen exchange. This reaction is an equilibrium, which ideally should be displaced towards the double oxygen incorporation, since the introduction of two oxygen in a peptide shifts its mass 4Da, facilitating the identification of the labeled peptide and its quantification by mass spectrometry-based techniques.

Both, the amide bond cleavage and the carboxyl oxygen exchange reactions, can take place at the same time, which is the principle of the direct ^{18}O-labeling procedure; or separately, in two steps, which is the basis for the decoupling method. In the next sections the differences between both procedures along with the variables affecting the performance of each one will be commented.

I.7.1.1 The direct labeling

Digestion of proteins in the presence of ^{18}O can isotopically label the resulting peptides. When the peptide bond is broken one atom of heavy oxygen is incorporated and further incorporation is done trough the carboxyl oxygen exchange at the same time.

Typically the protocol for ^{18}O-labeling is very similar to a normal in-solution digestion. First proteins are solubilized/denaturated then reduced and alkylated and finally digested with trypsin in the presence of heavy water. The direct labeling can be done accelerating each step with the aid of ultrasonic energy or not [55].

The main advantage of this procedure is simplicity and speed. In addition, it can be used in on-line approaches. The main drawback is the lack in labeling efficiency, which makes direct labeling suitable only for relative quantification and for studies regarding proteins over-expressed or under-expressed [56].

I.7.1.2 The decoupling procedure

The decoupling procedure was first proposed by Yao *et al.* who have demonstrated that the cleavage of the protein can be done separately from the labeling in such a way that almost total double oxygen incorporation can be achieved using the carboxyl oxygen exchange process. The carboxyl oxygen reaction is shown in figure I.4 [57, 58].

The decoupling procedure consists in a more elaborated, labor-intensive and time-consuming protocol.

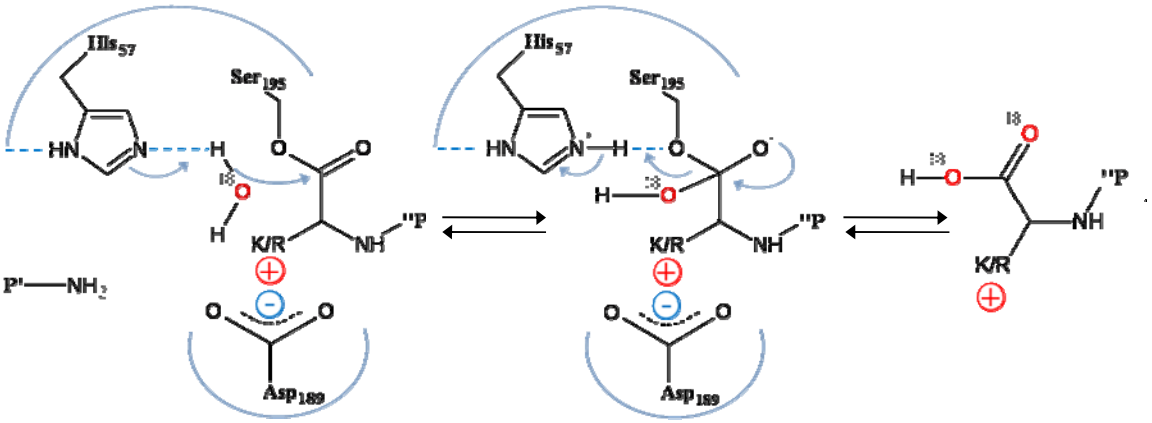


Figure I. 4. Mechanism of the double oxygen incorporation by carboxyl oxygen exchange.

In addition to the four main steps of the direct procedure, as described above, it is necessary to dry the sample, once the protein has been digested. Then the sample is recomposed in a solution of acetonitrile/ammonium bicarbonate, that facilitates peptide solubilization and that lowered the pH of the solution to speed the subsequent carboxyl oxygen exchange process. Then the sample is dried again and finally recomposed in ^{18}O-water. Finally, the labeling takes place.

The advantage of this procedure is that a higher degree of peptides are double labeled in an efficient maner, which is extremely important for the accuracy of the quantification, as it will be shown in the next sections. Nevertheless this procedure takes longer than the direct method, usually 12h, if ultrasonication is used to speed the protein digestion step; otherwise it takes 24-48h. In addition, it cannot be done on-line, and consequently it is hard of automation.

I.7.1.3 The use of immobilized enzymes

Enzymes, such as trypsin, can be immobilized in solid supports that can be used either for protein digestion/labeling in a single step (direct labeling) or in different ones (decoupling procedure). The main advantage of using immobilized enzymes is that once the solution is separated from the solid support, then the protein digestion is stopped and the carboxyl oxygen exchanges is slowed down to almost a negligible effect. This is very important for online approaches, where the sample will be mixed with solvents containing ^{16}O-water. This means that some ^{16}O/^{18}O interchange will happen, thus interfering with the quantification process. If the enzyme is not present in the solution the interchange can be consider small to have a significant effect on the quantification process. In addition, the

presence of peptides coming from the autolysis of the enzyme is negligible and thus the interferences that they may cause in the subsequent measurements by mass spectrometry are avoided [59].

I.7.2 The problem of the ^{18}O-labeling efficiency

The ^{18}O-labeling of peptides following any of the approaches described previously, results in a mixture of non-labeled, single labeled or double labeled peptides. The efficiency in the labeling depends not only on the protocol followed, but also in different variables affecting each protocol, such as pH, type of enzyme used, peptide characteristics, labeling time and conditions of labeling. In addition, the contribution to the peak intensities from the naturally occurring isotopes, others than ^{18}O, such as ^{13}C and ^{15}N, needs also to be considered.

The variability in the labeling efficiency makes extremely difficult the interpretation of data from complex protein mixtures, such as proteomes, making hard to distinguish highly up-regulated from highly down-regulated proteins or C-terminal peptides, in studies regarding differential proteomics. This problem is even worst when the direct method is the choice to perform the labeling.

To overcome the labeling efficiency problem, different algorithms have been proposed [60]. Other approaches to solve this problem focus on the sample treatment itself, trying to make the labeling as complete as possible, thus minimizing the effects on the efficiency to a residual level.

I.7.3 Variables affecting ^{18}O labeling of proteins

I.7.3.1 The enzymes used for ^{18}O labeling

The enzyme generally used to perform the ^{18}O labeling of proteins is the enzyme trypsin, a serine protease that predominantly cleaves peptide chains at the carboxyl side of amino acids lysine and arginine, which is also the most common enzyme used to digest proteins in proteomic studies. Other enzymes can also be used such as the endoproteinase Lys-C, that specifically cleaves peptide bonds C-terminally at lysine; and the endoproteinase Glu-C, that cleaves peptide bonds C-terminally at glutamic and aspartic.

Different yields of single and double incorporation can be gotten with any of these enzymes, yet yields are highly dependent of the labeling conditions [61-67].

I.7.3.2 The influence of the pH

The influence of the pH on the labeling process depends on the way in which this is done. In the direct process, the protein is labeled at the same time that is cleaved.

To guarantee the efficiency of the process, the pH must be the optimum at which the enzyme cleaves the protein, for instance the optimum pH for the enzyme trypsin is about 8-9. If the cleavage is done in pure ^{18}O-water, then all peptides will have one ^{18}O at the C-termini at least. Interestingly, the double ^{18}O incorporation will depend on various factors, such as the type of peptide and the pH of the solution. For instance, lysine terminated peptides do not incorporate two oxygen efficiently.

In addition, Yao *et al.* have demonstrated that by lowering the pH of the solution, the double incorporation can be done faster and with higher efficiency [57]. Precisely, the possibility to separate the protein cleavage from the isotopic incorporation doing both separately at different pH is the basis of the decoupling procedure, as proposed by the former mentioned authors.

I.7.4 ^{18}O inverse labeling

Since the goal of a protein differential analysis is to extract and to identify the small number of proteins that deviate in expression level upon a perturbation, any method that enables subtractive analysis of protein signals with unaltered levels would be of great value. This is the base of the method called inverse ^{18}O-labeling as first reported by Wang *et al.* [62].

In this procedure, parallel inverse labeling experiments are performed where the labeling is reversed in the second experiment, i.e., the heavy isotope labeled pool in the first experiment is light isotope labeled in the second experiment and vice versa as shown in figure I.5.

The characteristic inverse labeling pattern (i.e., isotope intensity profile swap) presented by peptide signals between the two inverse labeling experiments indicate the differential expression of proteins from which the peptides are derived.

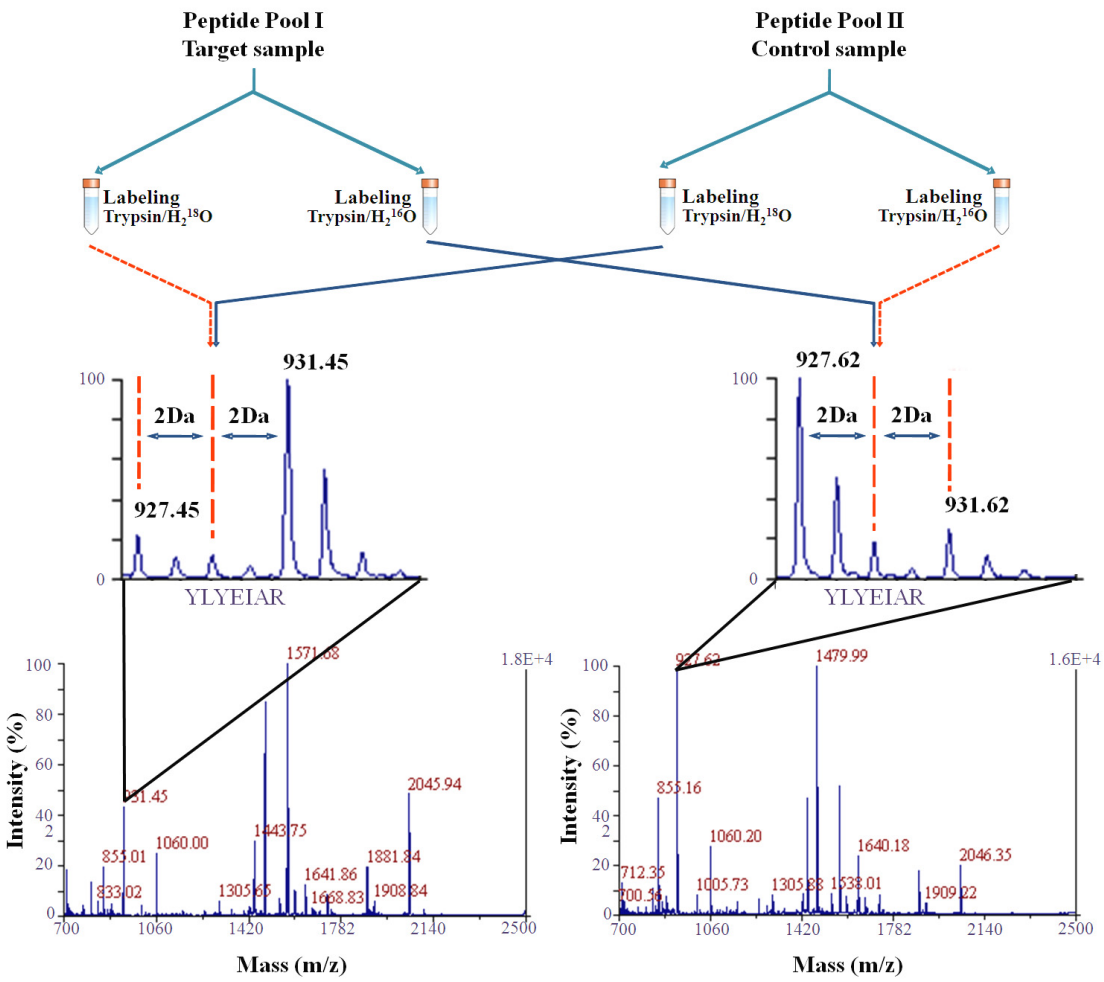


Figure I. 5. Direct and inverse labeling experiments of a target protein and the correspondent control sample.

The strategy works well with labeling methods that result in variable mass differences among peptides of different sequences. An MS instrument of high resolution is not required. The differentially expressed proteins and proteins of covalent changes resulting from an altered state are rapidly identified without ambiguity. The procedure presents a logical sequence in approaching the problem: data reduction of irrelevant signals, quick focus on signals of interest, detailed analysis on signals of interest only.

I.7.4.1 Applications to proteins separated by gel electrophoresis

The absolute quantification of proteins separated in polyacrylamide gels lacks in efficiency and accuracy, as it was demonstrated by Havlis and Shevchenko [67]. However this method is a powerful tool when applied for relative quantification. The best sample

treatment for relative quantification of proteins separated in polyacrylamide gels consists firstly in digest the gel spots containing the control sample, a certain amount of a known protein, and the target sample in ^{16}O-water, afterward the extracted pool of peptides is dried in a speed vacuum. Finally, the control sample is recomposed in ^{18}O-water whilst the target sample is recomposed in ^{16}O-water. The same amount of both samples is then mixed and submitted to mass spectrometry-based techniques. The signal intensities of the same peptides are then compared and the relative $^{16}O/^{18}O$ ratios are used to calculate the concentration of the unknown protein.

By carrying out ^{18}O-labeling post proteolysis, the amount of ^{18}O-water required is substantially reduced [64], since the labeling is performed on the extracted and dried peptides not on the excised gel pieces.

The ^{18}O-labeling of N-glycosylation sites during in-gel deglucosylation has been claimed as a way to extend the capabilities of MALDI peptide mass mapping and database searching for the identification of gel-separated proteins [68]. There are also some studies that report in-gel ^{18}O-labeling as a tool for protein quantification and characterization. For instance, Kosaka *et al.* [69] reported a method for the C-terminal characterization and identification of proteins separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), based on in-gel digestion/labeling of proteins. After 2D-PAGE of a cell lysate of rat liver, 10 different protein spots were in-gel digested during overnight with trypsin in a buffer solution of 50% (v/v) ^{18}O enriched water, and analyzed by nano-electrospary Fourier transform ion cyclotron resonance mass spectroscopy (nano-ESI FT-ICR MS). Due to the high resolution of FT-ICR MS and since the C-terminal peptides were the only unlabeled peptides in the protein digests, the authors were able to identify and characterize the C-termini of 8 proteins, from the initial 10 chosen, with a high level of confidence. On the other side, and following the same strategy, Larsen *et al.* [69] were unable to characterize and identify by MALDI the C-terminal peptides of different enolase-2-forms form stressed cells of *Saccharomyces cerevisiae*, when using trypsin to digest the separated proteins from 2D-PAGE. However, when the endoproteinase Asp-N was used to perform the in-gel digestion of proteins, in 50% of $H_2{}^{18}O$ buffer solution, they were able to fully characterize and identify five C-terminal peptides out of seven C-terminally processed enolase 2 forms. These results show how important it is the choice of the correct enzyme to perform protein digestion and labeling. In this case the authors were only concerned in distinguishing the unlabeled C-terminal peptide from the labeled internal peptides, so they use an enzyme, endoproteinase Asp-N that only incorporates one ^{18}O atom in the

carboxylic group of the peptide whilst trypsin incorporates one or two ^{18}O. This way, they were able to control the oxygen incorporation in the peptides, avoiding the presence of a mixture of single and double labeled peptides which would be responsible for the spreading of the peptide signal over more peaks, thus reducing the intensity in the mass spectra and compromising a correct identification.

The previous examples demonstrate the potential of ^{18}Oisotopic labeling in the detection of modified peptides by mass spectrometry thus enhancing the characterization of protein modifications, many of them implicated in a large number of disease states. On the other hand, there are also a few examples where in-gel ^{18}O-labelingwas successfully used for protein quantification and to identify potential biomarkers of disease states.

Using a combination between in-gel ^{18}O-labeling and nano-LC-MS/MS, Körbel *et al.* [71] were able to identify and quantify approximately ^{18}O low-expressed erythropoietin (EPO) receptor-dependent proteins, which are related with signaling pathways and EPO-dependent cellular responses like cell differentiation and proliferation. The developed method relies on immunoaffinity and 1D-PAGE for protein separation; in-gel protein digestion in the presence of in $H_2{}^{16}O$ or $H_2{}^{18}O$ (95% ^{18}O abundance) with non-methylated (nonstabilized) trypsin to avoid post-digestion 18O/16O back-exchange; and nano-LC–MS/MS techniques for the detection of low-fentomole proteins.

More recently, Chaerkady *et al.* [72] used lectin affinity enrichment and 1D-PAGE to purify and separate glycosylated proteins from liver tissues of healthy patients and patients with hepatocellular carcinoma. After in-gel digestion with trypsin in the presence of $H_2{}^{16}O$or $H_2{}^{18}O$ and LC–MS/MS analysis they were able to perform relative quantification and find differential expressed proteins from non-tumor/tumor samples.

Finally, the application of in-gel 18O-labeling has been used by Ang *et al.* [73] in the study of the *Porphyromonas gingivalis* proteome, an oral pathogen related with chronic periodontitis. Based on an inverse labeling strategy, where the control sample was first digested in $H_2{}^{16}O$ and the target sample digested in $H_2{}^{18}O$, whilst in the inverse experiment the control sample was digested in $H_2{}^{18}O$ and the target in $H_2{}^{16}O$, they found 24 proteins up-regulated and 18 proteins down-regulated in target sample. Once more, these examples show the reliability and applicability of the in-gel ^{18}O-labeling technique when applied to comparative proteomic studies.

I.7.4.2 Applications to proteins in solution

This is by far the most reported use given to ^{18}O-labeling. A combination of tandem mass spectrometry and ^{18}O-labeled internal standards provide absolute or relative quantification of simple or complex mixtures of proteins in solution. Internal standards could be directly obtained by digesting in ^{18}O-water a stock solution of known concentration of the protein of interest.

Since many peptides are quantified in parallel, the resulting estimate of the protein concentration is robust and would also be applicable to complex protein mixtures, when combined with MS/MS spectrometry, if powered by appropriate software.

The application of ^{18}O-labeling in assisting sequencing of peptides by helping to distinguish a-type ions from b-type ions in tandem mass spectrometry has been reported by different authors [74,75].

The isomerization of aspartic acid (Asp) to isoaspartic acid (IsoAsp) is a spontaneous reaction that can alter protein structure and function, and for this reason the identification of this phenomenon in proteins is a biochemical challenge. Xiao *et al.* [76] have proposed a method for the identification and quantification of the above-mentioned reaction based on the utilization of ^{18}O labeling and tandem mass spectrometry.

Wa *et al.* [76] have proposed a method for the study of the binding sites of proteins when they are immobilized in solid supports. The method is simple: the free protein is cleaved in ^{18}O-water whilst, after its immobilization, the protein is digested in ^{16}O-water. The two digests are mixed and then analyzed by mass spectrometry. Peptides having significantly higher ^{18}O/^{16}O ratios than other peptides in the same digest are involved in the immobilization.

Another interesting application consists in using the ^{18}O-labeling for the quantification of ribonucleic acids using ^{18}O. RNA labeled and non-labeled samples are mixed and then the intensities of the ^{16}O and ^{18}O monoisotopic peaks are then correlated through a specific formula. It must be taken into account that for this biomolecules, the labeling is only single [78].

I.8. Direct tissue mass spectrometry analysis

Tissues are one of the most challenging samples in proteomics. The possibility to follow and localize changes at the molecular level on a tissue sample, for example from a biopsy or a disease model organism is of vital importance in biomedical research.

The development of methodologies for the analysis of tissue sample at the protein level remains a rapidly growing field [79-81].

Histological and histochemical characterization of biological and medical tissue has been the method of reference for tissue analysis, allowing the discrimination of tissue types after specific staining. Histochemical staining techniques can only be used in a targeted manner for known compounds, and only a limited number of such targets can be visualized from a given sample at the same time [82].

Another approach to tissue analysis is the one using tissue homogenization and 2D-GE or multidimensional chromatography and protein identification by tandem mass spectrometry. One of the drawbacks of this type of approaches is that the sample preparation removes the relationship between morphology and proteins [83].

A further understanding of the role of the components of a biological system can be aided by the knowledge of the spatial localization of these components. This understanding can be achieved by performing direct tissue analysis by mass spectrometry. This technique avoids tissue homogenization and separation step, and it allows to maintain the spatial distribution of molecules within the tissue which is preserved. Thus, molecular analysis of this type allows for the direct targeting of diseased cells because it can be integrated with established histology protocols. Spectra generated directly from tissue sections can be highly enriched for a single cell type and thus contain higher concentrations of relevant proteins useful for medical purposes.

Since the development of MALDI ionization in the 80s, that this has become the method of choice to analyse biomolecules. With a growing need of molecular information on peptides and proteins and because of the characteristics of MALDI ion sources the direct tissue analysis was attempted.

In certain ways, MALDI-tissue analysis is a simple concept. On MALDI ion source, the ions are generated by laser irradiation of a solid surface containing matrix and analytes. Most of the irradiated energy is transferred in the depth of the irradiated sample with very little spreading of energy over the tissue surface. It means that if a sample is irradiated at a

precise coordinate the ions will directly come from the cross section between the tissue and the laser beam [84].

The analysis of a tissue surface is done by moving the sample under the laser beam in such a way to cover the whole tissue surface. Each acquired spectra represents the average of several laser shots in order to obtain a statistically representation of the analyzed area.

I.9 References

[1]Simpson, R.J. *Nature* 422 (2003) 473–473.

[2] Pandey, A., and Mann, M. *Nature,* 405 (2000) 837–846.

[3] Petricoin, E.F., Zoon, K.Z., Kohn, E.C., Barrett, J.C., *et al. Nat. Rev. Drug. Discov.* 1 (2002) 683–695.

[4] Sam, H. *Nature* 422 (2003) 226-231.

[5] Celis, J.E., Gromov, P., Ostergaard, M., Madsen, P., *et al. FEBS Lett.* 398 (1996) 129–134.

[6] Anderson, N.G., and Anderson, N.L. *Electrophoresis* 17 (1996) 443–453.

[7] Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C. *et al.*, *Proc. Natl. Acad. Sci.* 90 (1993) 5011–5015.

[8] Mann, M., and Wilm, M., *Anal. Chem.* 66 (1994) 4390–4399.

[9] Yates, J.R.III, McCormack, A.L., Schieltz, D., Carmack, E., Link A. *Anal. Chem.* 67 (1995) 3202–3210.

[10] Washburn, M. P., Wolters, D., Yates, J. R. III, *Nat. Biotechnol.* 19 (2001) 242–247.

[11] Wilkins, M.R.; Appel, R.D.; Van Eyk, J.E., et al. *Proteomics* 6 (2006) 4–8.

[12] M. Baker, *Nat. Biotechnol.* 23 (2005) 297–304.

[13] D.H. Chace*, Chem. Rev.* 101 (2001) 445–477.

[14] Fey, S.J., and Larsen, P.M., *Curr. Opin. Chem. Biol.* 5 (2001) 26–33.

[15] D. López-Ferrer, B. Cañas, J. Vázquez, C. Lodeiro, *et al. Trends in Analytical Chemistry*, 25 (2006) 996–1005.

[16] Nesvizhskii, A.I., Aebersold, R., *Mol. Cell. Proteomics* 4 (2005)1419–1440.

[17] Jiang, L., He, L., Fountoulakis, M., *J. Chromatogr. A* 1023 (2004) 317–320.

[18] Rabilloud, T. *Electrophoresis* 19 (1998) 758–760.

[19] Lauber, W.M., Carroll, J.A., Dufield, D.R., Kiesel, J.R., et *al. Electrophoresis* 22 (2001) 906–918.

[20] Cañas, B., López-Ferrer, D., Ramos-Fernández, A., Camafeita, E., *et al. Brief. Funct. Genomic. Proteomic.* 3 (2006) 1–26.

[21] Chang, J., Ruiz, V., Vancura, A., *Methods Mol. Biol.* 457 (2008) 141–149.

[22] Capelo, J.L., Carreira, R., Diniz, M., Fernandes, L. *et al. Anal. Chim. Acta*. 650 (2009) 151–159.

[23] Havlis, J., Thomas, H., Sebela, M., Shevchenko, A., *Anal. Chem*. 75 (2003) 1300–1306.

[24] Pramanik, B.N., Mirza, U.A., Ing, Y.H., Liu Y.H. *et al. Protein Sci.* 11 (2002) 2676–2687.

[25] Juan, H.F., Chang, S.C., Huang, H.C., Chen, S.T., *Proteomics* 5 (2005) 840-842.

[26] Santos, H.M., Capelo, J.L. *Talanta* 73 (2007) 795–802.

[27] Capelo, J. L. (Ed.). Ultrasound in Chemistry. Analytical Applications. Wiley-VCH, 2009.

[28] Mason, T.J., Sonochemistry, Oxford University Press, Oxford, UK, 1999. pp. 44–45.

[29] López-Ferrer, D., Capelo, J.L., Vázquez, J., *J. Proteome. Res.* 4 (2005) 1569–1574.

[30] Rial-Otero, R., Carreira, R.J., Cordeiro, F.M., Moro, A.J., *J. Chromatogr. A* 1166 (2007) 101–107.

[31] Hansen, K.C., Kiemele, L., Maller, O., O'Brien, J. *et al. Mol. Cell. Proteomics.* 8 (2009) 1648–1657.

[32] Postema, M. Van Wamel, A.,. Lancee, C.T., De Jong, N. *Ultrasound Med. Biol.* 30 (2004) 827–840.

[33] Klimek, J., Eddes, J., Hohmann, L., Jackson, J. *et. al. J. Proteome Res.* 7 (2008) 96–103.

[34] Fishbein, W.N. *Anal. Biochem.* 46 (1972) 388–401.

[35] http://www.bio-rad.com.

[36] Wu, W.W., Wang, G., Baek, S.J., Shen, R.F. *J. Proteome Res.* 5 (2006) 651–658.

[37] Timms. J.F., Cramer, R., *Proteomics* 8 (2008) 4886–4897.

[38] Elliott, M.H., Smith, D.S., Parker, C.E., Borchers, C. *J. Mass. Spectrom.* 44 (2009) 1637–1660.

[39] Durr, E., Yu, J., Krasinska, K.M., Carver, L.A. *et al*. *Nat. Biotechnol.* 22 (2004) 985–992.

[40] Nagalla, S.R., David, L.L. *J. Proteome Res*. 8 (2009) 1263–1270.

[41] Simpson, K.L., Whetton, A.D., Dive. C., *J. Chromatogr. B.* 877 (2009) 1240–1249.

[42] Van Langenhove, A., Costello, C.E., Biller, J.E. *Clin. Chim. Acta* 115 (1981) 263–275.

[43] Kitteringham, N.R., Jenkins, R.E., Lane, C.S., Elliott, V.L. *J. Chromatogr. B.* 877 (2009) 1229–1239.

[44] Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., Gygi. S.P. *PNAS* 100 (2003) 6940–6945.

[45] Kuzyk, M., Smith, D., Yang, J. *Mol. Cell. Proteomics* 8 (2009) 1860–1877.

[46] Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B. *et al. Mol. Cell. Proteomics* 1 (2002) 376–386.

[47] Chen, X., Smith, L.M., Bradbury. M.E., *Anal. Chem.* 72 (2000) 1134–1143.

[48] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F. *Nat. Biotechnol.* 17 (1999) 994–999.

[49] Li, J., Steen, H., Gygi, S.P, *Mol. Cell. Proteomics* 2 (2003) 1198–1204.

[50] Hansen, K.C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J. *et al. Mol. Cell. Proteomics* 2 (2003) 299–314.

[51] Gevaert, K., Impens, F., Ghesquiere, B., P. Van Damme, A. *et al. Proteomics* 8 (2008) 8 4873–4885.

[52] Y. K. Wang, D. F. Quinn, Z. Ma, E. W. Fu, *J. Chromatogr. B,* 782 (2002) 291–306.

[53] Ross, P. L., Huang, Y. N., Marchese, J. N.,Williamson, B. *et al. Mol. Cell. Proteomics* 3 (2004) 1154–1169.

[54] Boyer, P.D., Luchsinger, W.W. and Falcone, A.B. *J. Biol. Chem.* 223 (1956) 405–421.

[55] Capelo, J.L., Carreira, R.J., Fernandes, F., Lodeiro, C. *et al. Talanta*, 80 (2010) 1476–1486.

[56] Carreira, RJ; Rial-Otero, R; Lopez-Ferrer, D, Lodeiro, C. *et al. Talanta*, 2008 (76) 400–406.

[57] Yao, X., Freas, A., Ramirez, J., Demirev, P.A. *et al. Anal. Chem.* 73 (2001) 2836–2842.

[58] Peggi, M., Orlando, R., *Anal. Biochem.* 359 (2006) 26–34

[59] Sevinsky, J.R., Brown, K.J., Cargile, B.J., Bundy, J.L., *Anal. Chem.* 79 (2007) 2158–2162.

[60] Mason, C.J., Therneau, T.M., Eckel-Passow, J.E., Johnson, K.L. *et al. Bioinformatics* 22 (2006) 2739–2745.

[61] Reynolds, K., Yao, X., Fenselau, C. *J. Proteome Res.* (1) 2002 27–33.

[62] Wang, Y.K, Ma, Z., Quinn, D.F., Fu, E.W. *Anal. Chem.* 73 (2001) 3742–3750.

[63] Rose, K., Savoy, L.A., Simona, M.G., Offord, R.E., *et al. Biochem. J.* 250 (1998) 253–259.

[64] Schnolzer, M., Jedrzejewski, P., Lehmann, W.D. *Electrophoresis* 17 (1996) 945–953.

[65] Shevchenko, A. Shevchenko, *Anal. Biochem.* 296 (2001) 279–283.

[66] Regnier, F.E., Riggs, L., Zhang, R., Xiong, L. *et.al. J. Mass Spectrom.* 37 (2002) 133–145.

[67] Havlis, J., Shevchenko, A., *Anal. Chem.* 76 (2004) 3029–3036.

[68] Kuster, B., Mann, M., *Anal. Chem.* 71 (1999) 1431–1440.

[69] Kosaka, T., Takazawa, T., Nakamura, T. *Anal. Chem.* 72 (2000) 1179–1185.

[70] Larsen, R.M., Larsen, P.M., Fey, S.J., Roepstorff, P., *Electrophoresis* 22 (2001) 566–575.

[71] Körbel, S., Schümann, M., Bittorf, T., Krause, E., *Rapid Commun. Mass Spectrom.* 19 (2005) 2259–2271.

[72] Chaerkady, R., Thuluvath, P.J., Kim, M.S., Nalli, A. *et al. Clin. Proteomics* 4 (2008) 137–155.

[73] Ang, C.S., Veith, P.D., Dashper, S.G., Reynolds, E.C., *Proteomics* 8 (2008) 1645–1660.

[74] Uttenweiler-Joseph, S., Neubauer, G., Christoforidis, S., Zerial, M., M. Wilm, *Proteomics* 1 (2001) 668–682.

[75] Shevchenko, A., Chernushevich, I., Standing, K.G., Thomson, B., Wilm, M., Mann, M. *Rapid Commun. Mass Spectrom.* 11 (1997) 1015–1024.

[76] Xiao, G., Bondarenko, P.V., Jacob, J., Chu, G.C., *et al. Anal. Chem.* 79 (2007) 2714–2721.

[77] Wa, C., Cerny, R.L., Hage, D.S., *Anal. Chem.* 78 (2006) 7967–7977.

[78] Meng, Z., Limbach, P.A., *Anal. Chem.* 77 (2005) 1891–1895.

[79] Caprioli, R. M., Farmer, T. B., Gile, J. *Anal. Chem.* 1997 (69) 4751–4760.

[80]. Stoeckli, M., Chaurand, P., Hallahan, D. E., Caprioli, R. M. *Nat. Med.* 2001 (7) 493–496.

[81] Erin, H. and Caprioli, R.M. *PNAS* 2008 (25) 18126–18131.

[82] Römpp, A., Guenther, S., Schober, Y., Schulz, O. *et al. Angew. Chem*. 2010 (49) 3834–3838.

[83] Walsh, A., Rauser, S., Deininger, S.O. Histochem. Cell. Biol. 2008 (130) 421–434.

[84] Fournier, I., Marinach, C., Tabet, J.C., Bolbach, G. *J. Am. Soc. Mass Spectrom.* 2003 (14) 893–899.

Chapter II.

Objectives and working plan

This dissertation aims the development of new methods in proteomics, based on ultrasonic energy. The objectives covered by this thesis include:

(i) The application of ultrasonic energy to speed-up current sample preparation methods for protein identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry MALDI-TOF-MS;

(ii) The development of new approaches based on fast protein digestion and ^{18}O-labeling for protein quantification along with dedicated bio-informatics platform to manage the data generated by mass spectrometry (MALDI-TOF-MS);

(iii) The application of ultrasonic energy and stable ^{18}O-isotopic labeling for mass spectrometry analysis of tissue samples.

Working Plan

To develop the topics described above we chose to test and evaluate the effects of three ultrasonic devices, the ultrasonic probe, the ultrasonic bath and the sonoreactor, in the different steps of current sample treatments devoted to protein identification, protein quantification and mass spectrometry of tissues.

When possible, and depending on the topic, the following variables affecting the performance of ultrasonic-based treatments will be studied: ultrasonication time, ultrasonication amplitude, temperature of sonication and ultrasonication frequency.

Regarding the first topic, and as commented above, we will intend to apply ultrasonication in the main steps of current workflows for protein identification, namely protein denaturation/solubilisation, protein reduction, protein alkylation and protein digestion. The aim is to develop a sample treatment allowing a non intensive sample handling whilst speeding the entire procedure from overnight (12h) to minutes.

Concerning protein quantification, the second topic, we will try to develop a simply methodology joining the benefits of a well established technique for protein separation, namely gel electrophoresis, with the straightforwardness of ^{18}O labeling of proteins and with the simplicity of matrix assisted laser desorption/ionization time-of-flight-based mass-

spectrometry, the latest as the tool of quantification. To achieve the aforementioned objectives we will focus in several main aspects, as commented below. First, it will be revised the problems associated with protein quantification using MALDI and ^{18}O-labeling, and then it will be tried to find out an approach that might allow to overcome them.

Second to make things easier and faster, dedicated software to help in the quantification process will be developed whenever this would be required.

Third, on every occasion possible during the making of the experiments, the expertise of our research group in the simplification of sample treatment using ultrasonic energy will be used to achieve a better performance in the protein digestion or in the protein labeling or in both.

On the topic of mass spectrometry of tissues we will focus on two main limitations that up to date this technology needs to overcome: the bottle neck of sample throughput and the differentiation of the molecules present in the plume formed during the sample volatilization induced when the laser is shot over a tissue.

Chapter III.

Improving sample treatment for in-solution protein identification by peptide mass fingerprint using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

---

Published in:

Journal of Proteome Research, 6 (2007) 3393–3399

H. M. Santos, R. Rial-Otero, L. Fernandes, G. Vale, M. G. Rivas, I. Moura and J. L. Capelo

Index

III.1 Resumo

Foram estudadas três fontes de energia ultra-sónica com o objectivo de acelerar o tratamento de amostra para identificação de proteínas em solução por PMF *(peptide mass fingerprint)* utilizando MALDI-TOF-MS (*matrix assisted laser desorption/ionization time-of-flight mass spectrometry*). As etapas de redução, alquilação e digestão foram realizadas em 15min. Nove proteínas incluindo a proteína de resistência ao zinco da bactéria *Desulfovibrio desulfuricans* G20 e a proteína split-soret citocromo c da bactéria *D. desulfuricans* ATCC27774 foram identificadas com sucesso utilizando o novo protocolo.

Palavras-chave: Digestão em solução, identificação de proteínas, sonda de ultra-sons, sono-reactor, banho de ultra-sons, MALDI-TOFMS

A minha contribuição para este trabalho consistiu na optimização de todas as variáveis experimentais, análise de proteínas por MALDI-TOF-MS e o processamento e interpretação dos dados.

III.2 Abstract

Three ultrasonic energy sources were studied to speed up the sample treatment for in-solution protein identification by peptide mass fingerprint using matrix assisted laser desorption/ionization time-of-flight mass spectrometry. Protein reduction, alkylation, and enzymatic digestion steps were done in 15 min. Nine proteins; including zinc resistance-associated protein precursor from *Desulfovibrio desulfuricans* strain G20 and split-soret cytochrome c from *D. desulfuricans* ATCC27774 were successfully identified with the new protocol.

Keywords: in-solution protein digestion, protein identification, ultrasonic probe, sonoreactor, ultrasonic bath, MALDI-TOFMS

My contribution to this work was the optimization of all experimental variables, MALDI-TOF-MS analysis and data processing and interpretation.

III.3 Introduction

Protein identification is nowadays an issue of primary importance, since proteins are involved in virtually all biological processes, such as enzymatic catalysis, ion transport, immune protection, and signal transduction [1]. In addition, protein biomarkers are used for the identification of bacteria, disease screening, medical diagnosis, and virus identification, helping to reduce morbidity and mortality across the globe [2-6].

Different strategies can be found in literature for the identification of proteins, in which separation and purification steps, previous to protein identification using mass spectrometry, are mandatory. Three main approaches for protein identification throughout mass spectrometry can be found in literature [7] as follows:

The first approach is named in-gel protein digestion. In this sample treatment mixtures of proteins are first denaturized and then separated through 1D or 2D sodium dodecyl sulfate polyacrilamide gel electrophoresis, SDS-PAGE. The proteins are isolated in gel bands or spots that are excised. Each gel band or spot containing (theoretically) only one protein is mixed with a trypsin enzyme solution and incubated for a certain time. The enzyme trypsin is widely used to digest the protein into its peptides because it cleaves the protein exclusively after arginine or lysine residues in a reproducible manner; producing peptides of an average size of 800-2000Da, very adequate for mass spectrometry (MS) analysis. The set of peptides masses obtained is enough for the unambiguous identification of the protein. This method is known as peptide-mass fingerprint, PMF, and involves comparing the experimental masses from the peptides produced by the digested protein and those produced by in silico, theoretical digestion of all the proteins in a particular database [8, 9].

The second and third approaches use HPLC for the separation of peptides or proteins [7]. Regarding the second approach, called in-solution protein digestion, a complex mixture of proteins is first enzymatically digested into its peptides, and then the peptides are separated by HPLC and used for protein identification. Concerning the third approach, named in-column protein digestion, proteins are on-line: (i) first separated by HPLC; (ii) enzymatically digested in columns with immobilized trypsin to form peptides, and (iii) peptides are then separated by HPLC and used for protein identification.

The first approach previously described was drastically modified after the introduction of ultrasonic energy for the acceleration of the enzymatic protein digestion step. The total treatment time was reduced ca. 75% [10-12]. The mechanism that is

responsible for the enzymatic digestion enhancement using high-intensity ultrasound is not completely understood yet. It appears to be related to an increase in diffusion rates as consequence of the cavitation phenomena and heating. Although some previous experiments have shown that ultrasonic energy provided by an ultrasonic probe can be used for the acceleration of the in-solution digestion of proteins [12], it is not clear if all ultrasonic devices available on the market, such as sonoreactor or ultrasonic bath with dual frequency, can be used for the same purpose. In addition, no data is at present available in the literature regarding the effects of ultrasound on sample treatment previous to protein digestion with enzymes.

The aim of this work is to study the applicability of three ultrasonic energy sources, the sonoreactor, the ultrasonic probe, and the dual frequency ultrasonic bath, for the acceleration of the sample treatment for in-solution protein digestion for protein identification using PMF.

III.4 Experimental procedures

III.4.1 Apparatus

Protein digestion was done in safe-lock tubes of 0.5mL from Eppendorf (Hamburg,Germany). A vacuum concentrator centrifuge from UniEquip (Martinsried, Germany) model UNIVAPO 100H with a refrigerated aspirator vacuum pump model Unijet II was used for (i) sample drying and (ii) sample preconcentration. A minicentrifuge, model Spectrafugemini, from Labnet (Madrid, Spain), and a minicentrifuge-vortex, model Sky Line, from ELMI (Riga, Latvia) were used throughout the sample treatment, when necessary. A Simplicity 185 from Millipore (Milan, Italy) was used to obtain Milli-Q water throughout the experiments.

III.4.2 Ultrasonic devices

Different ultrasonic devices were tested as follows: (i) Ultrasonic bath, model Transsonic TI-H-5, from Elma (Singen, Germany). This device provides indirect sonication and is supplied with the last technological improvements regarding ultrasonic baths: dual frequency of sonication, for choosing between 35 or 130 kHz; power regulation, the ultrasonic power is variable from 10 to 100%; three operation modes, sweep, standard, and

degas; heating and timer. The highest intensity place of sonication inside the ultrasonic bath was located using the called "aluminum foil test".13 (ii) Sonoreactor, model UTR200, from Dr. Hielscher (Teltow, Switzerland). The sonoreactor provides indirect sonication and can be considered a high-intensity ultrasonic water bath. Samples can be processed in sealed tubes or vials eliminating aerosols and cross-contamination. The intensity of sonication produced by the sonoreactor is higher than the one given by an ultrasonic bath, ca. 50 times higher, but it is lower than the one produced by the ultrasonic probe, ca. 30 times. (iii) Ultrasonic probe, model UP 100H (Dr. Hielscher). The ultrasonic probe is immersed directly into solution, where the sonication takes place. This avoids indirect sonication, and the ultrasonic energy efficiency of this device is the highest of the three tested in this study.

III.4.3 Standards and reagents

The following protein standards were used: α-lactalbumin from bovine milk (g85%), BSA (>97%), and carbonic anhydrase (>93%) from Sigma (Steinheim, Germany) and albumin from hen egg white (>95%) from Fluka (Buchs, Switzerland). Chymotrypsinogen A, catalase bovine, and aldolase from rabbit were standards for gel filtration calibration kit high molecular weight from Amersham Biosciences (Piscataway, NJ). Trypsin enzyme, sequencing grade, was purchased from Sigma. All materials were used without further purification. α-Cyano-4-hydroxycinnamic acid (α -CHCA) puriss for MALDI-MS from Fluka was used as MALDI matrix. ProteoMass Peptide MALDI-MS Calibration Kit (MSCAL2) from Sigma was used as mass calibration standard for MALDI-TOF-MS. The following reagents were used for protein digestion: acetonitrile, iodoacetamide (IAA), and DL-dithiothreitol (DTT) (99%) were purchased from Sigma; formic acid and ammonium bicarbonate (Ambic) (>99.5%) were purchased from Fluka; trifluoroacetic acid (TFA, 99%) was from Riedel-de-Haën (Seelze, Germany); and urea (99%) was from Panreac (Barcelona,Spain).

III.4.4 Sample treatment

III.4.4.1 Classic method

Two different protein concentrations, 1 and 0.1µg/µL, prepared in 6.5M urea were used and treated according to the protocol schematized in figure III.1. Analyses were done by duplicate.

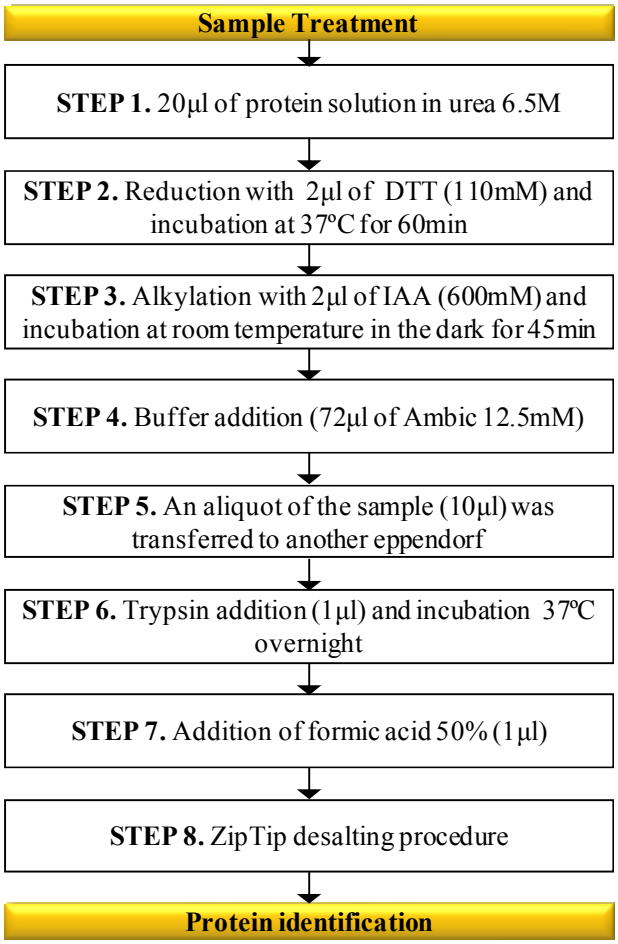


Figure III.1. Classic sample treatment for in-solution protein digestion.

To 20µL of protein solution, 2µL of a DTT solution (110mM in Ambic 12.5mM) were added, and the solutions were incubated at 37 °C for 1 h. Then, 2µL of an IAA solution (600mM in Ambic 12.5mM) were added. The solutions were further incubated at room temperature in the dark for 45 min. Solutions were diluted to 96µL with Ambic 12.5mM. Then, to 10µL of this solution, 1µL of the appropriate trypsin solution, in TFA 0.01%, was added to digest the protein. The protein/trypsin ratio was always 20:1 (w/w) as

recommended by the manufacturer [14]. So, amounts of proteins of 2 and 0.2µg were digested with 0.1 and 0.01µg of trypsin, respectively. For the classic method, overnight digestion at 37 °C was done. Then, 1µL of formic acid 50% (v/v) was added to stop the trypsin activity. Finally desalting using the ZipTip procedure was done, to avoid high saline concentration in the MALDI, as follows: (a) activation: aspirate and dispense, A&D, 10µL of acetonitrile (1x), then A&D 10µL of [50% acetonitrile + 0.1% TFA] (1x), and then A&D 10µL of 0.1% TFA (2x); (b) peptide binding: 10µL of sample (A&D the sample 20 cycles), (c) washing: A&D 10µL of 0.1% TFA (3x) and (d) peptide elution: 10µL of [90% acetonitrile + 0.1% TFA] (A&D the sample 20 cycles).

III.4.4.2 Protein digestion accelerated method

In this method, protein reduction and protein alkylation were done using the classic times of 60 and 45 min, respectively. The protein enzymatic digestion was reduced from overnight to 5 min, and it was performed under the effects of ultrasonic energy, as depicted in figure III.2. Ten microliters of sample were sonicated, by duplicate, using the ultrasonic bath or the sonoreactor, while 100µL of sample were sonicated with the ultrasonic probe. The reason for the difference in the volume of sonication lies with the minimum volume for which sonication with an ultrasonic probe can be done without extensively sample handling. Nevertheless, it must be pointed out that the protein concentration and the protein/enzyme ratio were the same for the different volumes sonicated. The sonication time, sonication amplitude, sonication frequency, and sonication temperature for each ultrasonic treatment are shown in detail in figure III.2, in which can be seen that short times were chosen for the sonoreactor and for the ultrasonic probe (5min), while for the ultrasonic bath, the less efficient ultrasonic device, longer times were also tried (15min). In addition, for the ultrasonic bath, the two frequencies of sonication were also investigated. After sonication, 1µL of formic acid 50% (v/v) was added to the samples to stop trypsin activity. ZipTip were used thorough the sample treatment to avoid high saline concentration in the MALDI as described above.

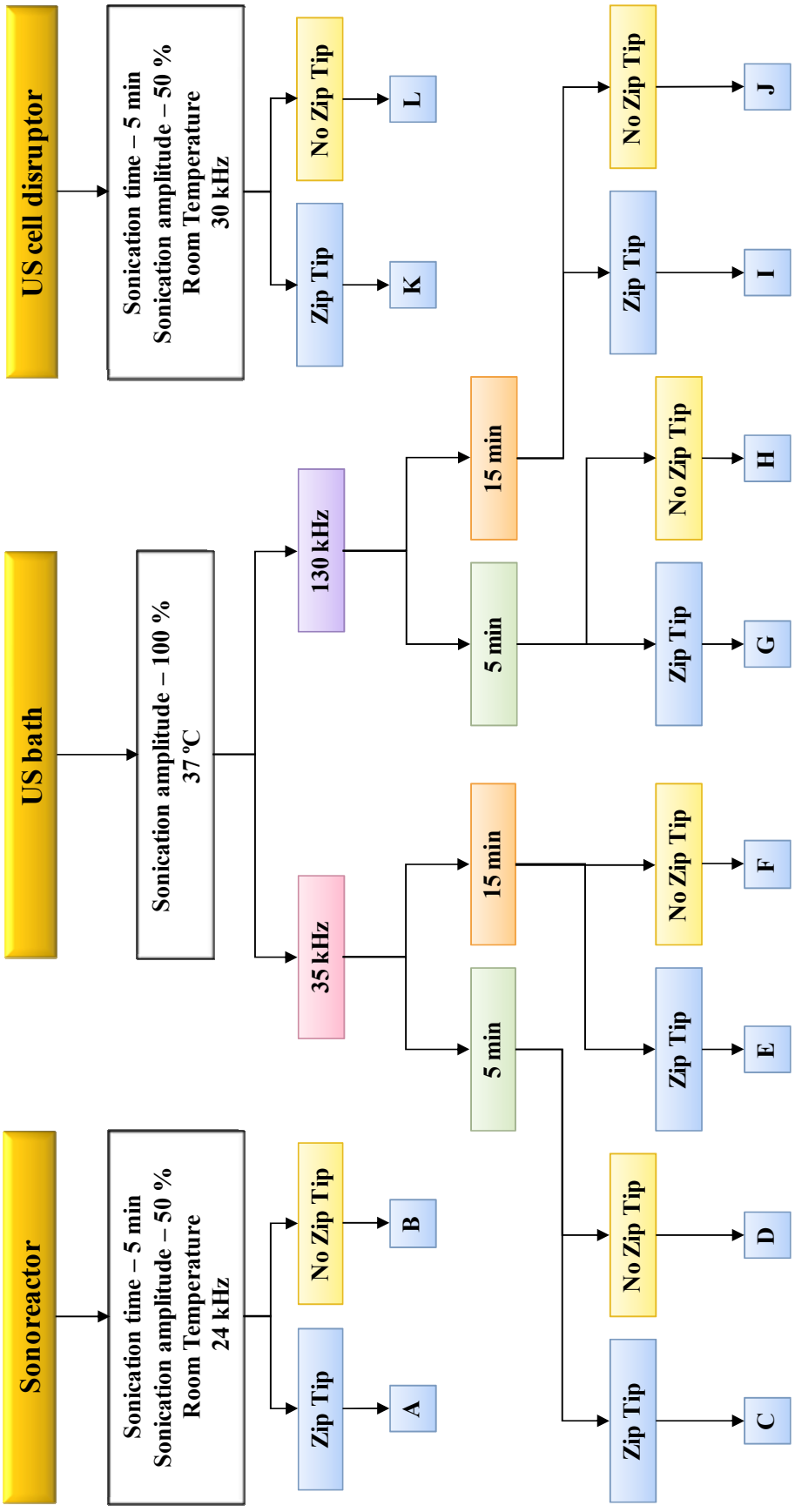


Figure III. 2. Sonication time, sonication amplitude, frequency of sonication, and temperature of the different sample treatments investigated.

III.4.4.3 Accelerated method

In this method, protein reduction and protein alkylation were done reducing the times from 60 and 45min to 5 and 5min, respectively, using the ultrasonic probe and the sonoreactor. The protein enzymatic digestion was reduced from overnight to 5min. ZipTip were used thorough the sample treatment to avoid high saline concentration in the MALDI as described above. Duplicate analyses for each protein were done.

III.4.5 MALDI-TOF-MS Analysis

A MALDI-TOF-MS system model Voyager DE-PRO Biospectrometry Workstation equipped with a nitrogen laser radiating at 337nm from Applied Biosystems (Foster City, CA) was used to obtain the PMF. MALDI mass spectra were acquired as recommended by the manufacturer and treated with the Data Explorer software version 4 series.

Prior to MALDI TOF-MS analysis, the sample was mixed with the matrix solution. α-CHCA matrix was used throughout this work and was prepared as follows: 10mg of α -CHCA was dissolved in 1mL of Milli-Q water/acetonitrile/ TFA (1mL+ 1mL + 2μL). Then, 2μL of the aforementioned matrix solution were mixed with 2μL of sample, and the mixture was shaken in a vortex for 30s. Finally, 1μL of the sample/matrix mixture was spotted on a well of a MALDI-TOF-MS sample plate and was allowed to dry. The estimated final amount of protein deposited in each spot of the MALDI plate was 0.1 or 0.01μL depending on the initial protein concentration used, as written above.

Measurements were done in the reflector positive ion mode, with a 20kV accelerating voltage, 75.1% grid voltage, 0.002% guide wire, and a delay time of 80ns. Two close external calibrations were performed with the monoisotopic peaks of the Bradykinin, Angiotensin II, P14R, and ACTH peptide fragments (*m*/*z*: 757.3997, 1046.5423, 1533.8582, and 2465.1989, respectively). Monoisotopic peaks were manually selected from each of the spectra obtained. Mass spectral analysis for each sample was based on the average of 500 laser shots. Peptide mass fingerprints were searched with the MASCOT [http://www.matrixscience.com/search_form_select.html] search engine with the following parameters: (i) Swiss-Prot, 2006 Database; (ii) molecular weight (MW) of protein, all; (iii) one missed cleavage; (iv) fixed modifications, carbamidomethylation (C); (v) variable modifications, oxidation (M); (vi) peptide tolerance up to 150 ppm. A match

was considered successful when the protein identification score was located out of the random region and the protein analyzed scores first.

III.4.6 Protein samples from complex mixtures

As a proof of the procedure, two different proteins, zinc resistance associated protein (Zrap) from *Desulfovibrio desulfuricans* strain G20 and split-soret cytochrome *c* from *D. desulfuricans* ATCC27774 were digested according to the accelerated method described above and identified by the PMF procedure by MALDI-TOF-MS. *D. desulfuricans* ATCC27774 and *D. desulfuricans* G20 cells were cultured in sulfate-lactate medium [15,16]. Cells were collected by centrifugation (8000 *g* during 15min at 4 °C), resuspended in 10mM Tris-HCl buffer, and ruptured in a French press at 9000 psi. After centrifugation (10 000*g*, 45min) and ultracentrifugation (180 000*g*, 60min), the supernatant was dialyzed against 10mM Tris-HCl buffer.

Both proteins were isolated from the soluble extract using chromatographic columns (anionic exchange, hydroxyapatite column, and molecular exclusion chromatography). The purity of the proteins was evaluated by SDS-PAGE and UV-visible spectroscopy. All purification procedures were performed under aerobic conditions at 4 °C and pH 7.6.

III.5 Results and discussion

III.5.1 Protein digestion accelerated method

The traditional sample treatment for in-solution protein digestion used in this work is presented in figure III.1. As can be seen, the number of total steps is 8, lasting for a total time of 2h from step 1 to step 4, and overnight for step 6. Our first approach to speed up the traditional procedure for fast protein identification was focused on the enzymatic protein digestion process, step 6, which takes as long as 12h. Two proteins with different masses, ovalbumin, 45kDa, and α-lactalbumin, 14.4kDa, were chosen as initial targets, in order to know whether the size of the protein could be also a variable to take into account. The three ultrasonic systems described in the apparatus section were tested with conditions as detailed in figure III.2. Two different levels of protein concentration, 1 and 0.1µg/µL, were assayed, as described in sample treatment, always maintaining the protein/enzyme ratio in 20:1 (w/w), as recommended by the enzyme manufacturer [14].

As far as higher concentration concerns, protein identification was successfully achieved in all conditions described in figure III.2 (data not shown). As expected, when ZipTip desalting was not used, the MALDI analysis was troublesome due to the high saline content of the sample, being necessary to increment twice the MALDI laser energy to achieve successful protein identification (data not shown). In addition, the sequence protein coverage and the number of peptides identified were higher when the sample desalting procedure with ZipTip was used (data not shown).

Regarding the lower protein concentration tested, results showed the effectiveness of ultrasonic energy for speeding up the protein digestion process, as it is shown in figure III.3, where panels A-C correspond to the protein α-lactalbumin, while panels 3D-F correspond to the protein ovalbumin.

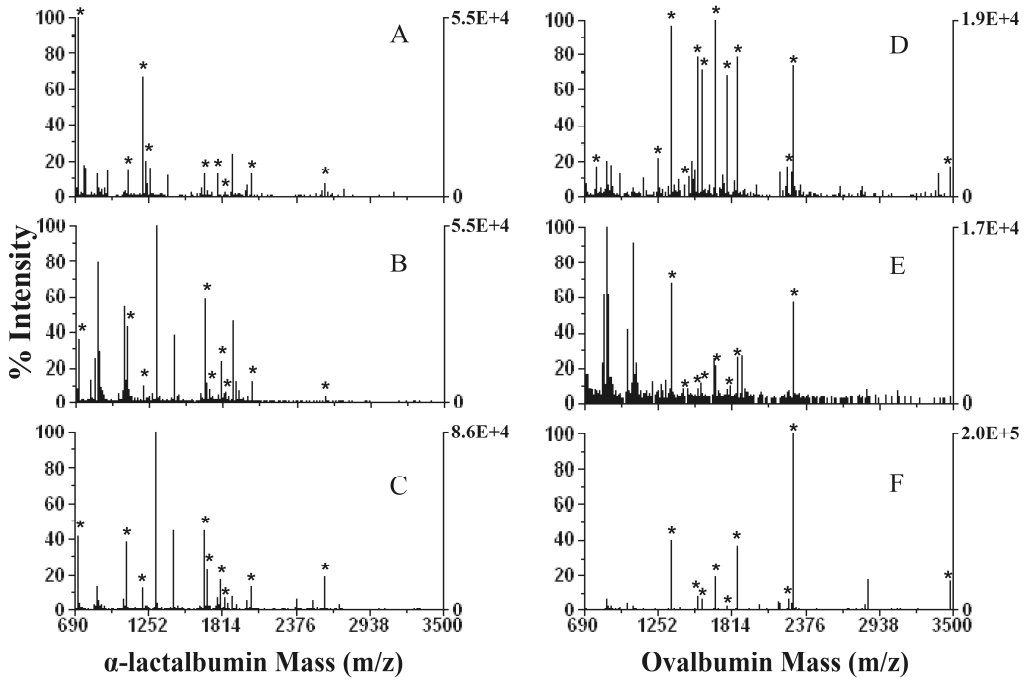


Figure III. 3. MALDI-TOF-MS spectra for α-lactalbumin and ovalbumin. (A and D) Classic treatment; (B and E) protein digestion accelerated method using ultrasonic probe; (C and F) protein digestion accelerated method using sonoreactor. For ultrasonic conditions refer to figure III.2. Protein/enzyme ratio was 20:1. Expected protein in the MALDI spot: 0.01μg.

Thus, unambiguous protein identification was obtained when the digestion process was accelerated with the sonoreactor and ultrasonic probe devices, with conditions corresponding to letters A and K in figure III.2. Figure III.3, panels B and E, shows the

MALDI spectra obtained after protein digestion accelerated with ultrasonic probe, 5min, for α-lactalbumin and ovalbumin, respectively, and panels C and F show the MALDI spectra obtained after protein digestion accelerated with sonoreactor, 5min, for α-lactalbumin and ovalbumin, respectively.

It can be noted that the number of peptides obtained was similar for the acceleration of the digestion process with the sonoreactor or with the ultrasonic probe, thus, indicating that both sonic devices can be used to speed up the enzymatic kinetics of the digestion process. In addition, the number of peptides and the standard deviations (8±1 peptides for α-lactalbumin and ovalbumin) were similar and comparable to the ones obtained with the classic method (12h) (9±1 peptides for α-lactalbumin and 13±1 peptides for ovalbumin), shown in figure III.3, panels A and D, for α-lactalbumin and ovalbumin, respectively. Moreover, the peptides obtained for the three protocols tested were the same as confirmed by their *m*/*z* ratio. It must be also remarked that to obtain positive identification was necessary to do desalting with ZipTip. It must be also pointed out that the background obtained was higher for the spectra obtained using ultrasonic probe, and that peak intensities were, as a general rule, higher for the method using sonoreactor, figures III.3 C and F. On the other hand, sample treatments using ultrasonic bath, corresponding to letters C-J in Figure III.2, did not speed the digestion process for low protein concentrations, and protein identification was not possible, even for the longer time, 15min (spectra are presented in figure III.1.SM of Supporting Information). These data could suggest that to speed up the kinetic process for enzymatic protein digestion, there is a minimum threshold of ultrasonic cavitation efficiency required, that an ultrasonic bath cannot provide.

In summary, the best results were obtained with the sonoreactor. In addition, the sonoreactor allows higher throughput, six samples at once, while the ultrasonic probe can be used only for one sample at a time. Moreover, sample handling is easier with the sonoreactor. So far, it was decided to use the sonoreactor instead of the ultrasonic probe for accelerating the enzymatic digestion of proteins for further experiments.

III.5.2 Accelerated method

The sample treatment was improved using ultrasonic energy in the step in which enzymatic digestion is done (step 6, figure III.1), as demonstrated in section III.5.1; however, two other steps of the procedure, namely, steps 2 and 3 (see figure III.1) are time-

consuming. Thus, the protein reduction and protein alkylation steps need 60 and 45min, respectively, to be done.

To investigate the applicability of ultrasonication to speed up these processes, R-lactalbumin and ovalbumin protein samples (0.1μg) were treated with three different sample treatments as follows:

(i) reduction was done with a sonication time of 2 or 5min, while alkylation was done for 45min according to the classic method (step 3, figure III.1);

(ii) reduction was done for 60min according to the classic method (step 2, figure III.1), but alkylation was done with a sonication time of 2 or 5min; and finally,

(iii) alkylation and reduction were both accelerated using ultrasound with a sonication time of 2 or 5min, each one.

In these experiments, for accelerating the steps 2 and 3, the ultrasonic probe and the sonoreactor were tested, but for all cases, enzymatic digestion was developed in the sonoreactor (50% amplitude and 5min of sonication time). Regarding 2min sonication time, protein identification was not possible, neither with sonoreactor nor with ultrasonic probe, for any of the three approaches described above (data not shown). However, when the sonication time was increased up to 5min, protein identification was possible for the three approaches described above, for both proteins and with both sonic devices. The sonoreactor was chosen for further experiments because it allows higher throughput and lower sample handling when comparing it with the ultrasonic probe. MALDI spectra of the fastest procedure, employing 5min for alkylation, 5min for reduction, and 5min for protein digestion with the sonoreactor, are shown in figure III.4, where it can be seen that the number of peptides obtained is similar for α-lactalbumin and ovalbumin proteins, and comparable to the classic approach, also shown in figure III.4. It must be remarked that the peptides obtained were the same as confirmed by their *m*/*z* ratio.

To avoid high saline contents and the utilization of ZipTip in the sample treatment, some experiments, with the fast ultrasonic method and the classic protocol, were developed in which the concentration of the reagents used in the procedure, as described in Figure III.1, were 10 times lower, namely, in steps 2 and 3. Results showed, however, that protein identification was not possible.

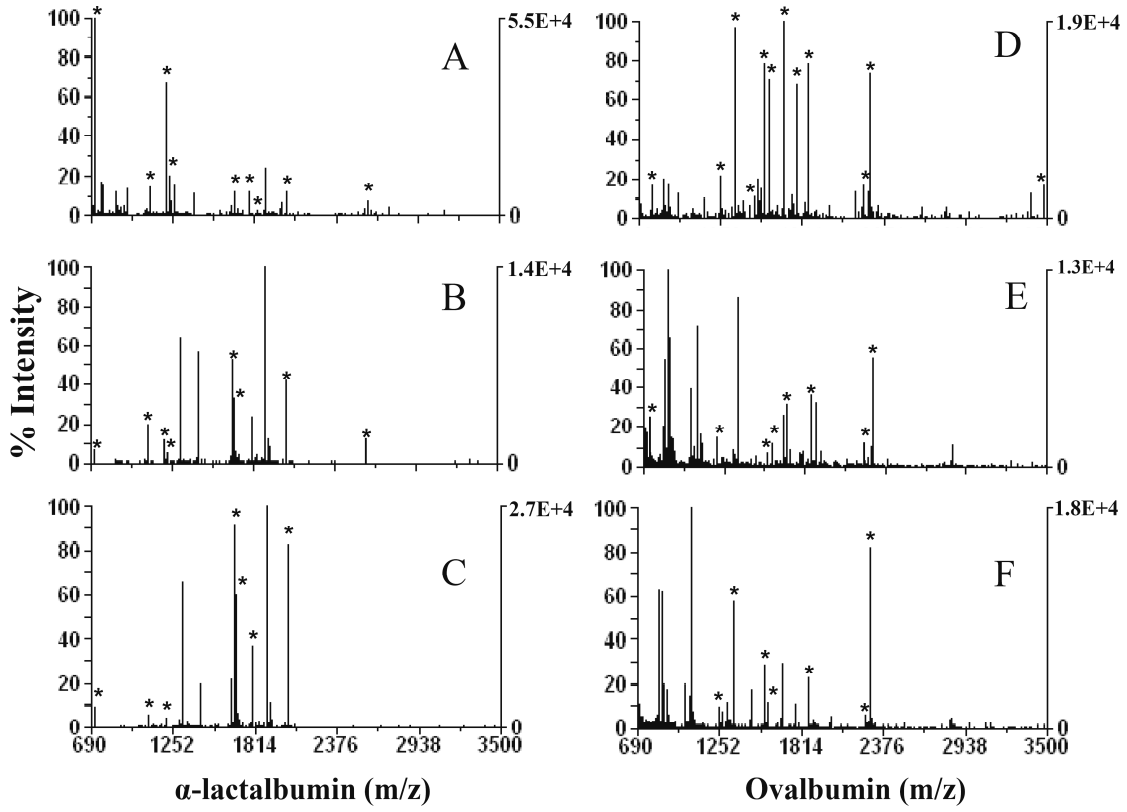


Figure III. 4. MALDI-TOF-MS spectra for α-lactalbumin and ovalbumin. (A and D) Classic treatment; (B and E) accelerated method using ultrasonic probe for protein alkylation, protein reduction, and protein digestion; (C and F) accelerated method using sonoreactor for protein alkylation, protein reduction, and protein digestion. For ultrasonic conditions refer to the text. Protein/enzyme ratio was 20:1 (w/w). Expected protein in the MALDI spot: 0.01μg

The new fast sample treatment for protein identification was further applied to the identification of the following proteins: aldolase, BSA, catalase, carbonic anhydrase, chymotrypsinogen A, α-lactalbumin and ovalbumin. The sonoreactor was the sonic device used to accelerate the alkylation (5min), reduction (5min), and enzymatic digestion processes (5min). The traditional protocol was also done for comparative purposes. Results are presented in table III.1, where it can be seen that protein identification was possible for all proteins using the accelerated protocol. In addition, the number of peptides matched and the protein sequence coverage were similar for the fast sample treatment and the classic one.

Table III. 1. Number of Peptides Matched and Protein Sequence Coverage (%) for the Classic and Accelerated Methods. Initial protein concentration: [a] 2 µg/µl; [b] 10 µg/µl; [c] 20 µg/µl.

Protein name	Theor. Mr (kDa)	Overnight			Accelerated method		
		Mascot score	Sequence Coverage (%)	No. of unique peptides identified	Mascot score	Sequence Coverage (%)	No. of unique peptides identified
α-Lactalbumin [a]	16.7	115±7	44±8	9±1	84±1	43±3	8±1
Chymotripsinogen A [a]	26.2	115±1	50±0	8±0	70±1	33±3	5±0
Carbonic anhydrase [a]	29.1	101±1	50±0	8±0	71±1	60±1	13±1
Ovalbumin [a]	43.2	106±6	47±2	13±1	71±1	24±2	7±1
BSA [a]	71.2	137±1	56±4	25±6	139±2	36±1	15±0
Aldolase rabit [b]	39.8	103±4	33±4	10±1	104±7	36±7	9±4
Catalase bovine [b]	60.1	152±4	40±1	17±1	216±5	44±1	20±1
Zinc resistance-associated protein precursor *Desulfovibrio desulfuricans* (strain G20) [c]	17.9	89±2	52±1	7±0	88±9	39±7	8±0
Split-Soret cytochrome c *Desulfovibrio desulfuricans* [c]	27.8	132±4	53±4	11±1	157±1	55±0	12±0

III.5.3. Proof of the method

As a proof of the method, two different proteins, zinc resistance-associated protein precursor from *D. desulfuricans* strain G20 and split-soret cytochrome *c* from *D. desulfuricans* ATCC27774 were digested according to the classic and accelerated method described above and identified by the PMF procedure by MALDI-TOF-MS.

Three different concentrations of proteins, 1, 0.5, and 0.1µg/µL, were tested. The classic method was done using the protocol depicted in figure III.1. The accelerated method was done as follows: proteins were reduced and alkylated in the sonoreactor (5min + 5min). Then the samples were diluted with Ambic, and 10µL of the sample was taken and digested in the sonoreactor (50% amplitude and 5 min of sonication time). The protein/trypsin ratio equal to 20:1 was constant for all cases. Positive identification was only possible for the higher protein concentration using any of the methods described above. Results are reported in table III.1. As can be seen, positive identification was always achieved, with similar results for both approaches. Furthermore, the number of peptides matched and the protein sequence coverage were also comparable. In addition, equivalent signal-to-noise ratios and intensities were observed, as it is shown in the figure III.5.

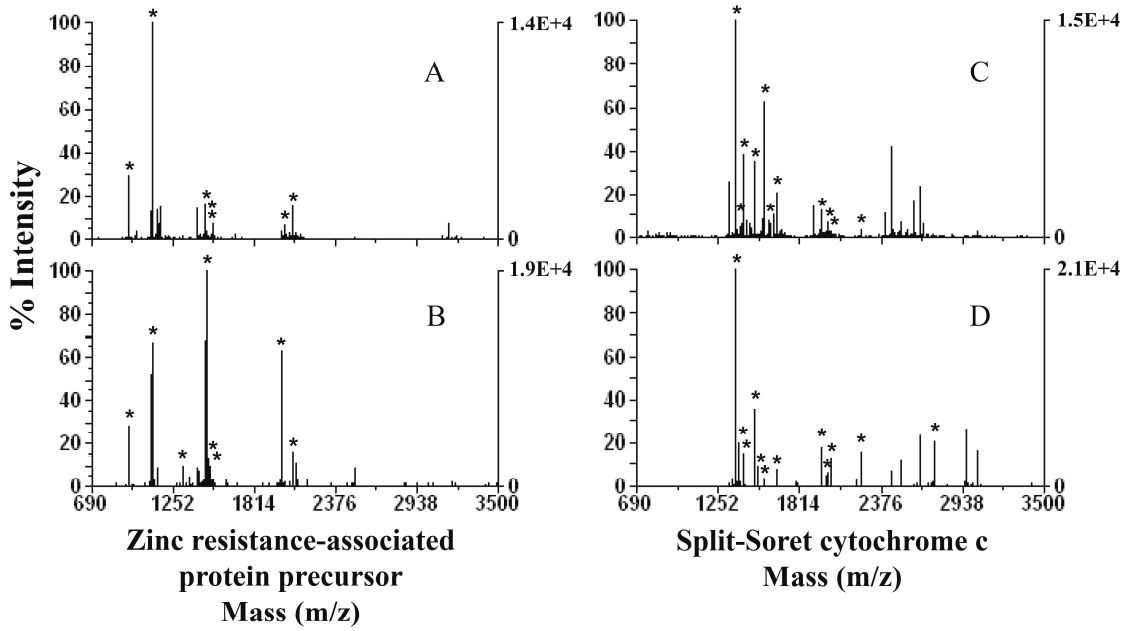


Figure III. 5. MALDI-TOF-MS spectra of zinc resistance-associated protein precursor from *D. desulfuricans* strain G20 and split-soret cytochrome c from *D. desulfuricans* ATCC27774 obtained after in-solution protein digestion according to the classic treatment (A and C, respectively) or to the accelerated method using sonoreactor for protein alkylation, protein reduction, and protein digestion (B and D, respectively). For ultrasonic conditions refer to the text. Protein/enzyme ratio was 20:1 (w/w). Expected protein in the MALDI spot: 0.1 μg.

The zinc resistance associated protein precursor from *D. desulfuricans* strain G20was identified using MSDB or NCBInr databases, because no conclusive result can be obtained from the Swiss-Prot database. This fact is related with the presence or absence of the protein in the database used. Thus, zinc resistance-associated protein precursor from *D. desulfururicans* is not included in the Swiss-Prot database, but it is in the MSDB or NCBInr databases. These results confirm that comparable digestion yields to those obtained by previous time-consuming published methods could be attained in only 15 min with the new sample treatment.

III.6 Conclusions

It has been demonstrated that ultrasonic probe and sonoreactor can be used for accelerating the sample treatment for protein digestion for protein identification by PMF using MALDI-TOF-MS from 24h to 15min, without compromising the number of peptides

matched or the protein sequence coverage obtained. The protein alkylation, protein reduction, and protein enzymatic digestion times were reduced to 5min for each step, respectively. Therefore, the new methodology represents a good alternative to the classic protocol. In addition, with the new proposed methodology, sample handling has been enormously simplified. The following proteins, in the mass range 16.7-71.2kDa, α-lactalbumin, ovalbumin, aldolase, BSA, catalase, carbonic anhydrase, and chymotrypsinogen A were correctly identified with the new accelerated method using the sonoreactor device. The identification of the zinc resistance-associated protein precursor from *D. desulfuricans* strain G20 and split-soret cytochrome *c* from *D. desulfuricans* ATCC27774 demonstrates that (a) parameters tested on standard samples can be also applied to complex biologic samples and (b) the method provides important advances in fast protein recognition. The findings reported in this work open a new way of sample treatment for in-solution protein digestion, of easy implementation, for on-line procedures, and tandem mass spectrometry. Thus, the advances here described are at present being implemented in our laboratory for on-line high-throughput protein identification from complex mixtures.

III.7 Acknowledgments

R. Rial-Otero and M. G. Rivas acknowledge the postdoctoral grants SFRH/BPD/23072/2005 and SFRH/BPD/29577/2006, respectively, of FCT (Science and Technological Foundation) from Portugal. Dr. J. L. Capelo acknowledges the MALDI-TOF-MS service of the Chemistry Department of the New University of Lisbon (http://www.dq.fct.unl.pt/MALDI/) for their helpful assistance and valuable suggestions.
The research findings here reported are protected by international laws under patent pending PCT/IB2006/052314 and PT 103 303.

III.8 References

[1] Doonan, S. Peptides and Proteins. In *Tutorial Chemistry Texts*; RSC: London, 2002.

[2] Chace, D. H. *Chem. Rev.* **2001**, *101*, 445.

[3] Demirev, P. A.; Feldman, A. B.; Lin, J. S. *John Hopkins APL Tech. Dig.* **2004**, *25*, 27.

[4] Stump, M. J.; Black, G.; Fox, A.; Fox, K. F.; Turick, C. E.; Matthews, M. J. *Sep. Sci.* **2005**, *28*, 1642.

[5] Drake, R. R.; Deng, Y. P.; Schwegler, E. E.; Gravenstein, S. *Proteomics* **2005**, *2*, 203.

[6] Yao, Z. P.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2002**, *74*, 2529.

[7] Lopez-Ferrer, D.; Canas, B.; Vazquez, J.; Lodeiro, C.; Rial-Otero, R.; Moura, I.;. Capelo, J. L. *TrAC, Trends Anal. Chem.* **2006**, *25*, 996.

[8] Henzel, W. J.; Watanabe, C.; Stults, J. T. *J. Am. Soc. Mass. Spectrom.* **2003**, *14*, 931.

[9] Pappin, D. J.; Hojrup, P.; Bleasby, A. *J. Curr. Biol.* **1993**, *3*, 327.

[10] Rial-Otero, R.; Carreira, R. J.; Cordeiro, F. M.; Moro, A. J.; Rivas, M. G.; Fernandez, L.; Moura, I.; Capelo, J. L. *J. Proteome Res.* **2007**, *6*, 909.

[11] Carreira, R. J.; Cordeiro, F. M.; Moro, A. J.; Rivas, M. G.; Rial- Otero, R.; Gaspar, E. M.; Moura. I.; Capelo, J. L. *J. Chromatogr.. A* **2007**, 1153, 291 .

[12] López-Ferrer, D.; Capelo, J. L.; Vázquez, J. *J. Proteome Res.* **2005**, *4*, 1569.

[13] Mason, T. J. *Sonochemistry*; Oxford University Press; Oxford, U.K., 1999, pp 44-45.

[14] Sigma-Aldrich Technical Bulletin of trypsin proteomics grade (http://www.sigmaaldrich.com/sigma-aldrich/bulletin/t6567bul.pdf, last accessed March 2007).

[15] Devreese, B.; Costa, C.; Demol, H.; Papaefthymiou, V.; Moura, I; Moura, J. J.; Van Beeumen, J. *Eur. J. Biochem.* **1997**, *248*, 445.

[16] Feio, M. J.; Zinkevich, V.; Beech, I. B.; Llobet-Brossa, E.; Eaton, P.; Schmitt, J.; Guezennec, J. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 1747.

III.9 Supporting information

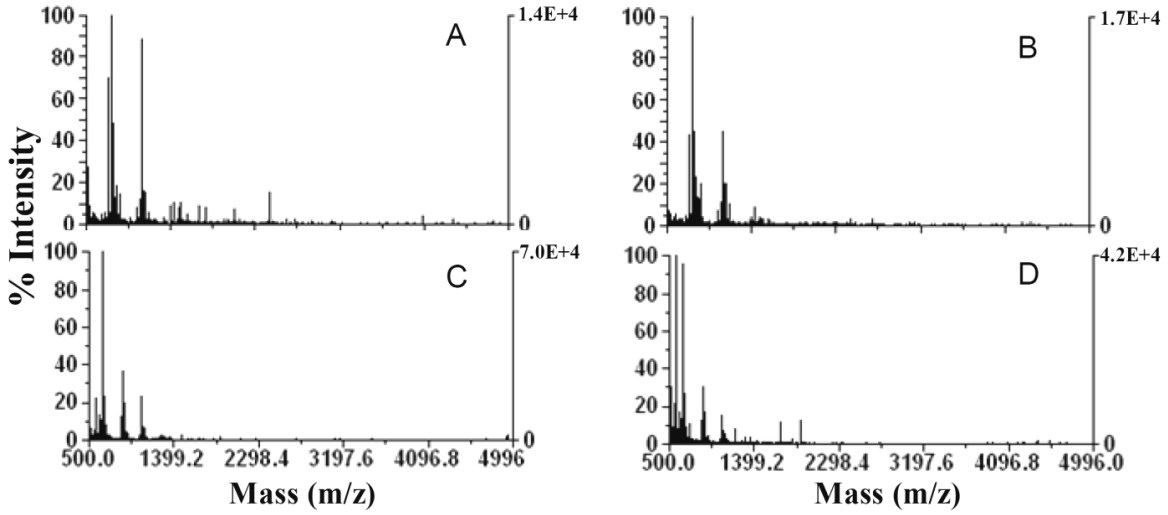


Figure III. SM. 1. MALDI-TOF-MS spectra for α-lactalbumin after protein digestion in the ultrasonic bath: A: frequency 35KHz, 100% amplitude and 5min sonication time; B: frequency 35KHz, 100% amplitude and 15min sonication time; C: frequency 130KHz, 100% amplitude and 5 min sonication time; D: frequency 130KHz, 100% amplitude and 15min sonication time.

Chapter IV.

Ultrasonic multiprobe as a new tool to overcome the bottleneck of throughput in workflows for protein identification relaying on ultrasonic energy

---

Published in:

Talanta, 81 (2010) 55–62

Hugo M. Santos, Ricardo Carreira, Mario S. Diniz, M.G. Rivas, Carlos Lodeiro, José J.G. Moura, José L. Capelo

Index

IV.1 Resumo

No presente trabalho realizamos um estudo acerca das capacidades de uma nova multi-sonda de ultra-sons para utilização em proteómica. Foram avaliados parâmetros como robustez facilidade de manuseamento reprodutibilidade. O estudo foi realizado aplicando a multi-sonda a duas velocidades diferentes fluxos de trabalho proteómica. O método "clássico" de digestão de proteínas (12h) foi utilizado como procedimento padrão. Este trabalho demonstra claramente a importância da optimização prévia de parâmetros como amplitude de ultra-sons e o tempo sonicação quando se pretende utilizar um determinado protocolo com um novo sistema de ultra-sons. Os resultados apresentados demonstram e confirmam as vantagens da utilização conjunta da energia de ultra-sons e uma placa de 96 poços para o tratamento em grande escala de amostras para identificação de proteínas. Os métodos estudados e comparados apresentam resultados semelhantes em termos de robustez, mas o método "*desalting-free*" provou ser o mais rápido, requerendo apenas 2min/amostra. Este método demonstrou também, ser o mais simples em termos de manuseamento, uma vez que não requer nenhum passo para remoção de sais. As proteínas, albumina de plasma bovino, α-lactalbumina, ovalbumina, anidrase carbónica, fructose-bisfososfatase aldolase A, catalase e quimotripsinogen A foram identificadas com sucesso utilizando os métodos estudados. Adicionalmente a proteína Split-Soret citochromo c isolada da bactéria *D. desulfuricans* ATCC 27774 foi também identificada com sucesso.

Palavras-chave: "*Desalting free*"**,** PMF**,** Ultrasonicação, MALDI**,** Split-Soret citocromo c

A minha contribuição para este trabalho consistiu na optimização de todas as variáveis experimentais, análise de proteínas por MALDI-TOF-MS e o processamento e interpretação dos dados.

IV.2 Abstract

We studied in this work the performance of the new ultrasonic multiprobe in terms of throughput, handling and robustness. The study was conducted using the multiprobe to speed two different proteomics workflows. The "classic" method relaying on overnight protein digestion (12h), was used as the standard procedure. This work clearly shows the importance of testing variables such as ultrasonic amplitude and ultrasonic time when adapting an ultrasonic-based treatment to a new ultrasonic device. The results here presented also shown and confirm the advantage of speed up sample treatment workflows with the aid of ultrasonic energy in combination with a 96-well plate. The methods compared were similar in terms of robustness, but the desalting free method was the fastest, requiring only 2 min/sample for completion. In addition it was also the simplest in terms of handling, since no desalting step was needed. The following standard proteins were successfully identified using the methods studied: bovine serum albumin, α-lactalbumin, ovalbumin, carbonic anhydrase, fructose-bisphosphate aldolase A, catalase, chymotrypsinogen A. As case study, the identification of the protein Split-Soret cytochrome c from *D. desulfuricans* ATCC 27774 was carried out.

Keywords: Desalting free**,** PMF**,** Ultrasonic, MALDI**,** Split-Soret cytochrome c

My contribution to this work was the optimization of all experimental variables, MALDI-TOF-MS analysis and data processing and interpretation.

IV.3 Introduction

The enzymatic cleavage of proteins is regularly used in proteomics for protein identification through peptide mass fingerprint and mass spectrometry-based techniques [1].

Nowadays high throughput in sample treatment is generally recognized for the scientific community as a priority demand in proteomics approaches. Over the last years we have witnessed the reporting of different strategies to (i) reduce the time needed to perform protein digestion and to (ii) simplify the handling for protein identification [2]. Thus, warming [3], ultrasonic energy [4–6], infrared radiation [7,8], high pressure [9] or spinning [10] are recent strategies that allow performing sample treatment for protein identification of complex mixtures in a fast, efficient and reproducible manner. From the strategies mentioned above, the utilization of ultrasonic energy as a way to speed the enzymatic kinetics of protein cleavage from overnight (hours) to minutes was first reported in 2005 [4] and it was soon validated by different research groups [11,12]. Later, the use of ultrasonic energy was successfully extended to the different steps of the sample handling for protein identification, namely protein solubilization/denaturation, protein reduction and protein alkylation [13]. The most recent sample treatment reported in literature making use of ultrasonic energy to accelerate sample handling for protein identification claims a time to complete the process of 8min in a clean method that avoids desalting procedures [14].

The present work shows a step forward in the application ultrasonic energy in proteomics workflows, since high sample throughput is obtained by jointing for the first time an ultrasonic multiprobe, allowing the treatment of four samples at once, with a 96-well plate. The comparison study was conducted through the identification of seven target proteins by mass spectrometry and peptide mass fingerprint using three different sample treatment workflows. In addition, as a case study, the identification of Split-Soret cytochrome c from *D. desulfuricans* ATCC 27774 was carried out using the three methods compared in this work.

IV.4 Experimental

IV.4.1 Apparatus

Protein digestion was done in a 96-well plate (Digilab-Genomic Solutions, USA). A vacuum concentrator centrifuge from UniEquip (Martinsried, Germany) model UNIVAPO 100H with a refrigerated aspirator vacuum pump model Unijet II was used for (i) sample drying and (ii) sample pre-concentration. A minicentrifuge, model Spectrafuge-mini, from Labnet (Madrid, Spain), and a minicentrifuge-vortex, model Sky Line, from ELMI (Riga, Latvia) were used throughout the sample treatment, when necessary. A SimplicityTM 185 from Millipore (Milan, Italy) was used to obtain Milli-Q water throughout the experiments.

IV.4.2 Ultrasonic devices

An ultrasonic multiprobe from Branson Ultrasonics Corporation (USA), model SLPe (150 W, 40 kHz ultrasonic frequency, 1mm diameter probe). The ultrasonic generator SLPe is equipped with a multiprobe detachable horn (model 4c15), with four tips for simultaneous ultrasonication of four samples and it was used in conjunction with a 96-well plate, as it is depicted in figure IV1.

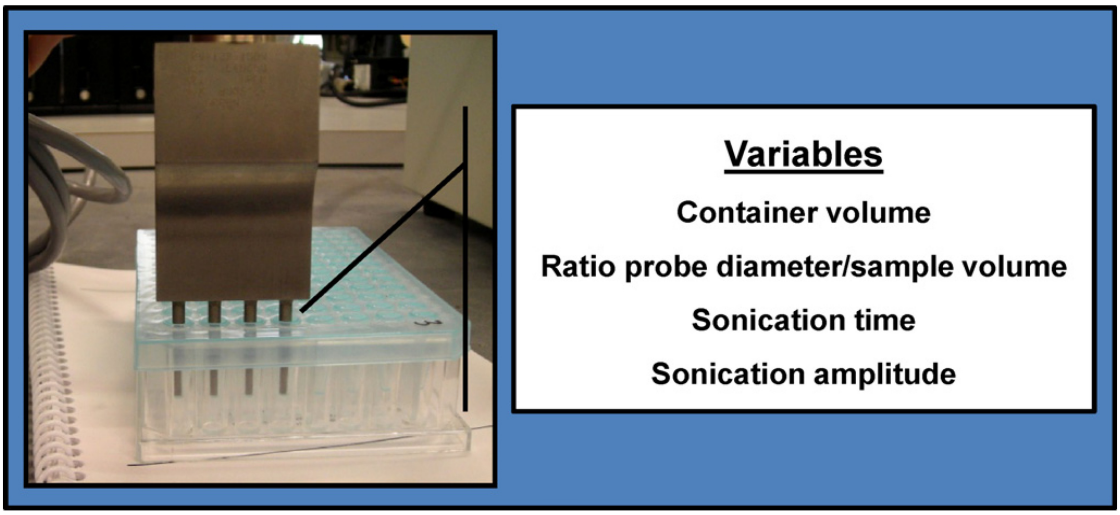


Figure IV. 1. This figure shows the new ultrasonic multiprobe coupled to the 96-well plate.

IV.4.3 Standards and reagents

The following protein standards were used: α-lactalbumin from bovine milk (≥85%), bovine serum albumin (>97%) and carbonic anhydrase (>93%) from Sigma (Steinheim, Germany), albumin from hen white (>95%) from Fluka (Buchs, Switzerland). Chymotrypsinogen A, catalase bovine and fructose-bisphosphate aldolase from rabbit were standards for gel filtration calibration kit high molecular weight from Amersham Biosciences (Piscataway, USA). Trypsin enzyme, sequencing grade was purchased from Sigma. All materials were used without further purification. α-Cyano-4-hydroxycinnamic acid (α-CHCA) puriss for MALDI-MS from Fluka was used as MALDI matrix. ProteoMassTM Peptide MALDI-MS Calibration Kit (MSCAL2) from Sigma was used as mass calibration standard for MALDI TOF-MS.The following reagents were used for protein digestion: acetonitrile, iodoacetamide (IAA) and dl-dithiothreitol (DTT) (99%) were purchased from Sigma; formic acid and ammonium bicarbonate (>99.5%) were from Fluka; trifluoroacetic acid (TFA, 99%) were from Riedel-de Haën (Seelze, Germany); and urea (99%) was from Panreac (Barcelona, Spain).

IV.4.4 Sample treatments

The sample treatments summarized below are based in works previously developed in our laboratory. However, they are applied for the first time with an ultrasonic multiprobe in a 96 well plate.

IV.4.4.1 Classic method

The overnight digestion was performed after protein denaturation with 6.5M urea and reduction with 10mM DTT in 12.5Mm ammonium bicarbonate at 37ºC for 1h. Iodoacetamide was then added to a final concentration of 50mM. The resulting mixture was incubated at room temperature in darkness for 45 min. The mixture was then diluted 4-fold to reduce urea concentration and an aliquot of 50µL was taken to perform the enzymatic digestion.

After trypsin addition (1:20, w/w trypsin-to-protein ratio), all samples were incubated at 37ºC overnight (12h). Then, 1µL of formic acid 50% (v/v) was added to stop

the trypsin activity (final formic acid concentration in sample 5%). Finally, to avoid high saline concentration in the MALDI, desalting using ZipTips® was done as follows:

(a) Activation: aspirate and dispense, A&D, 10µL of acetonitrile (×1), then A&D 10µL of [50% acetonitrile/0.1% TFA] (×1), and then A&D 10µL of 0.1% TFA (×2),

(b) Peptide binding: 10µL of sample (A&D the sample 20 cycles),

(c) Washing: A&D 10µL of 0.1% TFA (×3),

(d) Peptide elution: 10µL of [90% acetonitrile+0.1% TFA] (A&D the sample 20 cycles).

IV.4.4.2 Accelerated urea method

In brief, the method described above and referred as "classic method" was followed but protein alkylation, protein reduction, and protein digestion were done in 10, 10, and 4min respectively, under the effects of an ultrasonic field [15]. ZipTip® were used thorough the sample treatment to avoid high saline concentration in the MALDI as described above.

IV.4.4.3 Accelerated clean method

This method was recently reported by our group [14]. In brief, the proteins were dissolved in mixed acetonitrile/aqueous solutions, and (i) denaturation, (ii) reduction and (iii) alkylation of proteins were done in steps of 1min whilst protein digestion was done during 4min. Ultrasonic energy was used in all steps.

IV.4.5 A case study

IV.4.5.1 D. desulfuricans ATCC 27774

D. desulfuricans ATCC 27774 cells were cultured in sulfate–lactate medium. Cells were collected by centrifugation (8000×g during 15min at 4ºC), resuspended in 10mM Tris–HCl buffer and ruptured in a French press at 9000 psi. After centrifugation (10,000×g, 45min) and ultracentrifugation (180,000×g, 60 min) the supernatant was dialyzed against 10mM Tris–HCl buffer. The soluble extract was loaded in a DEAE-CelluloseTM and then

in a Q-SepharoseTM column both equilibrated with 10Mm Tris–HCl and eluted with a linear gradient to 250mM Tris–HCl.

The third purification step included a hydroxyapatite column equilibrated with 100mM Tris–HCl and eluted with a potassium phosphate linear gradient from 1 to 200mM. Finally, the fraction containing the haemic-protein was concentrated in a diaflow system (membrane YM 10) and loaded in a Superdex 200 column (Pharmacia) equilibrated with 300mM Tris–HCl. The purity of the proteins was evaluated by SDS-PAGE and UV–vis spectroscopy. All purification procedures were performed under aerobic conditions at 4ºC and pH 7.6.

IV.4.6 MALDI-TOF-MS analysis

A MALDI-TOF-MS model Voyager DE-PRO Biospectrometry Workstation equipped with a nitrogen laser radiating at 337nm from Applied Biosystems (Foster City, USA), was used to obtain the PMF. MALDI mass spectra were acquired as recommended by the manufacturer and treated with the Data ExplorerTM software version 4 series. Prior to MALDI-TOF-MS analysis, the sample was mixed with the matrix solution. α-CHCA matrix was used throughout this work and was prepared as follows: 10mg of α-CHCA was dissolved in 1mL of Milli-Q water/acetonitrile/TFA (1ml+1ml + 2µL). Then, 2µL of the aforementioned matrix solution was mixed with 2_l of sample and the mixture was shaken in a vortex for 30s. Finally, 1µL of the sample/matrix mixture was spotted on a well of a MALDI-TOF-MS sample plate and was allowed to dry.

Measurements were done in the reflector positive ion mode, with a 20kV accelerating voltage, 75.1% grid voltage, 0.002% guide wire and a delay time of 100ns. Two close external calibrations were performed with the monoisotopic peaks of the Bradykinin, Angiotensin II, P14R and ACTH peptide fragments (m/z: 757.3997, 1046.5423, 1533.8582 and 2465.1989, respectively). Monoisotopic peaks were manually selected from each of the spectra obtained. Mass spectral analysis for each sample was based on the average of 500 laser shots. Peptide mass fingerprints were searched with the MASCOT[http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF] search engine with the following parameters: (i) SwissProt. 2006 Database; (ii) molecular weight (MW) of protein: all; (iii) one missed cleavage; (iv) fixed modifications: carbamidomethylation (C); (v) variable modifications: oxidation (M); (vi) peptide tolerance up to 150 ppm.

A match was considered successful when the protein identification score is located out of the random region and the protein analyzed scores first.

IV.4.7 Statistical analysis

The t-test was used to determine differences between the ultrasonic-based methods and the classic overnight methods. Statistical analysis was performed with the significance level of 5%, using the software Statistica 6.0 (StatSoft, Inc. 1984–2001, Tulsa, OK, USA).

IV.5 Results and discussion

IV.5.1 Classic method

Table IV.1 show that all the target proteins were identified when the classic method was carried out. The target proteins scored always first and out of the random region in all replicates done in this study (n = 4). The number of peptides matched and the sequence coverage obtained for each protein were used to compare performance with the other two methods studied in this work.

IV.5.2 Accelerated urea method

Next, we carried out a set of experiments to compare the performance of the ultrasonic multiprobe in the acceleration of the classic method. Previous research developed in our group has shown that the classic method can be accelerated using ultrasonic energy in the different steps of its workflow [5,6,13–15]. A pitfall of this procedure when the treatment is done with single probe is that samples must be handled one by one. In other words, sample throughput is still a bottleneck in the application of ultrasonic-based approaches for the acceleration of methods commonly used for protein identification. Nevertheless, the ultrasonic-probe-based devices have evolved in such a way that multiprobes for the simultaneous treatment of samples ranging from 4 to 12 are nowadays available [16]. Their performance for proteomics applications has not been tested yet, to the best of our knowledge.

Table IV.1. Protein sequence coverage and number of peptides matched for the in-solution protein digestion: overnight method, accelerated method with urea and ZipTip® and clean method in H2O/acetonitrile (n=4, pt=0.05[a]).

Protein*	Theor. Mr (kDa)	Classic method x±S.D.			accelerated urea method x±S.D. (p_{ex}[b])			accelerated clean method x±S.D. (p_{ex}[b])		
		Mascot score	Sequence coverage (%)	Peptides matched	Mascot score	Sequence coverage (%)	Peptides matched	Mascot score	Sequence coverage (%)	Peptides matched
α-lactalbumin *B. taurus*	16.7	137±7	47±2	11±1	74±2	49±7 (0.65)	9± (0.10)	89±8	46±4 (0.71)	9±1 (0.07)
Chymotrypsinogen A *B. taurus*	26.2	139±8	61±8	10±2	76±2	51±7 (0.10)	8±2 (0.26)	118±6	51±7 (0.08)	9±2 (0.39)
Carbonic anhydrase II *B. taurus*	29.1	104±8	53±5	13±3	107±7	52±4 (0.66)	11±2 (0.41)	178±4	59±4 (0.12)	16±4 (0.16)
Fructose-bisphosphate aldolase A *O. cuniculus*	39.8	164±9	48±4	13±2	104±15	47±2 (0.92)	12±1 (0.51)	90±7	44±6 (0.41)	11±2 (0.19)
Ovalbumin *G. gallus*	43.2	121±3	50±9	15±2	134±6	43±5 (0.25)	15±2 (0.77)	147±7	46±9 (0.56)	15±4 (1.00)
Catalase *B. taurus*	60.1	159±6	43±4	16±1	129±7	39±7 (0.32)	14±4 (0.44)	254±5	43±2 (0.14)	17±4 (0.18)
Serum Albumin *B. taurus*	71.2	288±10	61±5	39±4	307±10	62±5 (0.39)	37±4 (0.47)	207±9	56±12 (0.22)	34±7 (0.24)
Split-Soret cytochrome c *D. desulfuricans*	27.8	128±8	62±8	18±3	119±9	53±2 (0.09)	15±2 (0.10)	107±8	65±12 (0.70)	15±2 (0.10)

* Initial Protein concentration: 0.3µg/µl. Accelerated method with urea: protein reduction and protein alkylation were done with 10 min ultrasonication time and 25% ultrasonication amplitude each one, whilst protein digestion was done with 4 min ultrasonication time and 10% ultrasonication amplitude. Accelerated clean method with acetonitrile: protein reduction and protein alkylation were done with 1 min ultrasonication time and 25% ultrasonication amplitude each one, whilst protein digestion was done with 5 min ultrasonication time and 10% ultrasonication amplitude. [a]p_t= theoretical significance level. [b]experimental significance level.

It must be stressed that it is necessary to adequately address the challenges of high sample throughput while maintaining data quality. Therefore, the most important variables affecting ultrasonic-based sample treatments, namely amplitude and time of ultrasonication [16] were studied. This set of experiments was carried out in a 96-well plate which is the regular plate used in robotic platforms. It must be noteworthy that when a single probe is used to speed proteomics workflows, the diameter of the tip regularly used is 0.5mm, whilst the new multiprobes are manufactured with a diameter of 1mm [16]. This difference is important since the lower is the sample container and the ratio sample volume/probe diameter, the higher is the risk of to lose sample by aerosol formation and subsequent sample spreading out of the sample container. In addition, aerosol formation can led to cross-contamination among the wells of the 96-well plate. For those reasons, testing the changes in sample treatment performance any time that a new ultrasonic device is tested is

very important. Figure IV.2 shows the results obtained for the identification of BSA and α-lactalbumin under different conditions of ultrasonication.

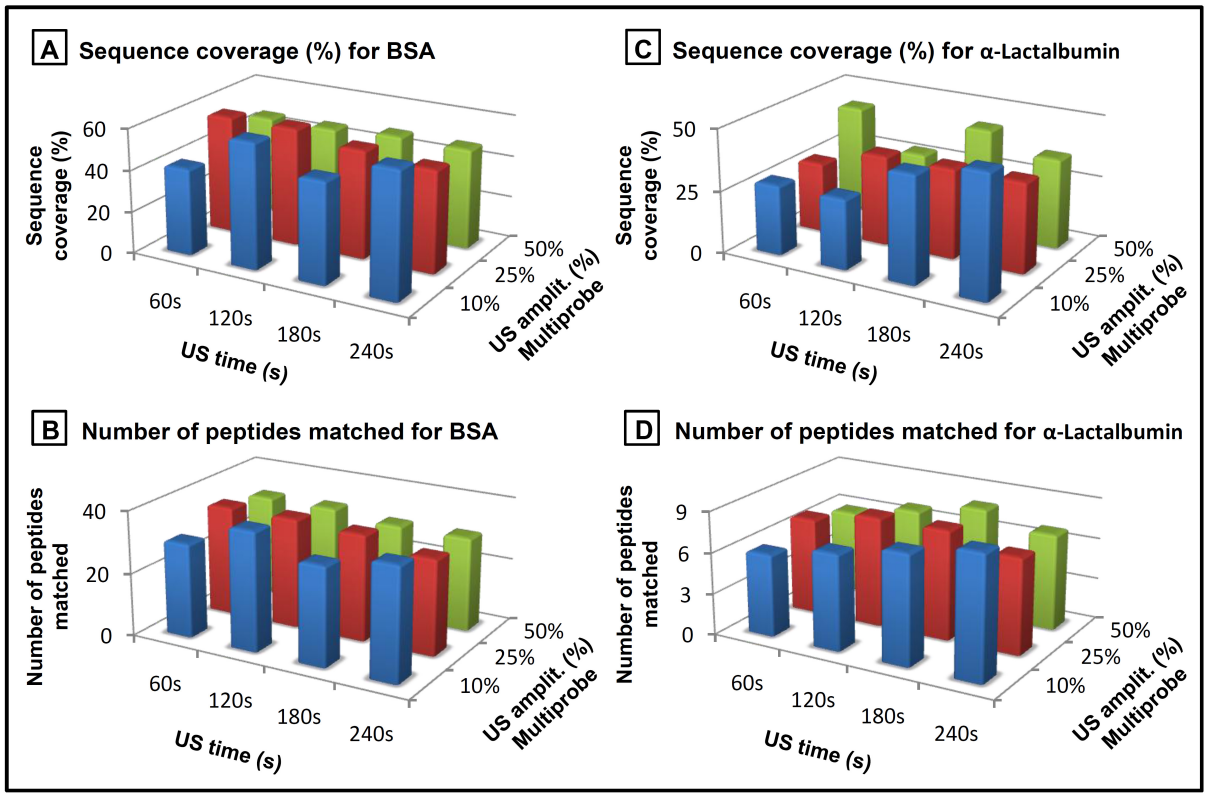


Figure IV. 2. Number of peptides matched and sequence coverage for bovine serum albumin and α-lactalbumin as a function of time, and amplitude of ultrasonication. Sequence coverage and peptides matched for the overnight method was 61±5 and 39±4 for bovine serum albumin and 47±2 and 11±1 for α-lactalbumin, respectively.

For both proteins studied, the results suggest that jointing long times of ultrasonication (i.e. 240 s) with low ultrasonication amplitudes, the number of peptides matched and the protein coverage obtained allow protein identification at the same confidence level that with the classic protocol, as it is also showed in table IV.1. This conclusion may be also observed for higher amplitudes. This result suggests that amplitude of ultrasonication is not an important variable in order to accelerate the in-solution digestion of proteins, at least for the ultrasonic multiprobe here assessed. This finding is important since high ultrasonic amplitudes can degrade the protein with the result of failing in protein identification.

More important, to use amplitudes as low as possible when working with a 96-well plate, is also advantageous because it avoids the spreading out of the sample through drops

or aerosol formation, overcoming cross-contamination. As an example, it is noteworthy that when we assayed an ultrasonic amplitude of 75%, sample was lost, being withdrawing from the well as an aerosol due to the effects of the high ultrasonic amplitude used. As consequence, cross-contamination caused by the aerosol formation was also verified, being detected peptides of BSA protein in well plates that in theory only contained protein α-lactalbumin and vice versa. Consequently the highest amplitude used in further experiments was 10%. Moreover, the maximum volume of sample recommended to work with is 50µl, higher volumes can lead to random cross contamination.

Regarding ultrasonication time, as may be seen in figure IV.2, this variable was found to affect the results; since the number of peptides matched and their respective sequence coverage for either BSA or α-lactalbumin were slightly improved when this variable was increased in the digestion step from 60 to 240s. These findings are consistent with data previously reported in the literature [14,15]. We hypothesize that a complex relation between the type of enzyme, the type of substrate, the ultrasonic amplitude and the ultrasonic time, influences the efficiency of the enzymatic process when it is carried out under the effects of an ultrasonic field. Thus, Sakakibara *et al*. have shown an enhancement in the reaction's kinetic for the hydrolysis of sucrose, when it was used the enzyme invertase in conjunction with ultrasonication [17]. Nevertheless, other authors have pointed out, that ultrasonic energy can inactivate enzymes. Thus, Bracey *et al*. have reported an inhibitory effect in the activity of the enzyme subtilisin when the subtilisin catalyzed interesterification reaction in an organic solvent was studied under the effects of ultrasonication [18]. As further example, although the enzyme protease XIV was inactivated towards casein substrate after 2 min of ultrasonication with probe, the same enzyme in the same conditions was active towards mussel tissue substrate after 4min of ultrasonication [19].

Next, with the best conditions found in the set of experiments above described, we proceed to identify other proteins, as showed in table IV.1, The number of peptides matched and the protein coverage were statistically compared with those obtained with the classic method and no differences were found at a significance level of $p > 0.05$ (test t, $n1 = 4$, $n2 = 4$). This result indicates that with the right conditions chosen the multiprobe can be used in conjunction with a 96-well plate to obtain fast and high throughput sample treatment for protein identification.

IV.5.3 Accelerated clean method

A drawback of the classic method is that, prior to MS analysis, ZipTip® tips or other kind of home-made mini-columns containing C18 beads are often used as peptide micro extraction and purification columns. It has been demonstrated; however, that sample loss can be as high as 90%, when ZipTips® columns are used [20]. This loss depends on the absolute concentration of the initial peptide digest loaded into the ZipTips® and is peptide type-dependant [20]. In addition, using ZipTips® sample handling becomes time consuming, labor intensive and expensive. Therefore we also tested the performance of the multiprobe in a clean method relaying in the use of a mixture of water/acetonitrile to solve the sample. The initial trials were done applying ultrasonic energy in the reduction, alkylation and digestion steps of our proteomic workflow (25% ultrasonic amplitude and 5min ultrasonic time in each step). Results, however, were unexpectedly low in terms of protein sequence coverage and peptides matched, especially for protein α-lactalbumin. This can be explained because amino acid residues valine and isoleucine has the potential to sterically hinder trypsin binding when an incomplete protein denaturation has been done. Therefore, a step was added in which protein denaturation was done by applying ultrasound to the sample before proceed with the subsequent protein reduction. Remarkably, after this change, the results obtained in terms of peptides matched and protein coverage for BSA, α-lactalbumin and for the other proteins used in this study, as showed in table IV.1, were as good as for the classic or the accelerated urea methods. The number of peptides matched and the protein coverage obtained were statistically compared with those of the classic method and differences were not found (test t, $p > 0.05$, $n1 = 4$, $n2 = 4$).

IV.5.4 MALDI spectra

MALDI spectra of the sample treatments here compared are presented in figure IV.3 for BSA and α-lactalbumin. For both proteins, the spectrum belonging to the classic method shows a different pattern of peak intensities, when compared to the ultrasonic-based ones. This could suggest that under the effects of an ultrasonic field some peptides are preferentially formed. Furthermore, when the spectra of the accelerated method are compared with the ones of the clean method, some differences in peak intensities are also observed that can be attributed, in this case, to the differences between both sample treatments: the use of urea/Zip Tips or organic solvents respectively. It is also possible that

the different reagents used in the sample treatments might influence the peptide distribution within the matrix spot. It must be pointed out that, despite of the differences in peak intensities, when the optimum conditions find out for each method were used, protein identification was always possible.

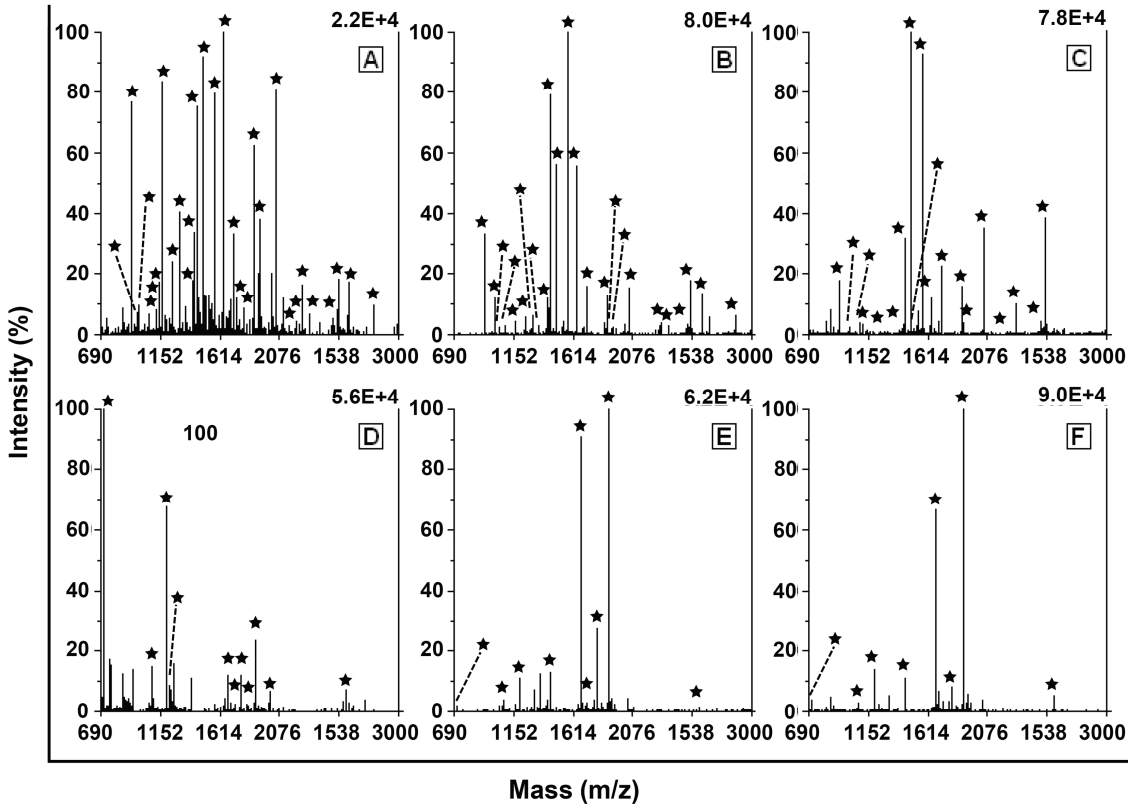


Figure IV. 3. MALDI spectra obtained of in-solution digestion of 3µg/µl of bovine serum albumin and α -lactalbumin. Panel A, B and C, spectrum of digested bovine serum albumin with classic method (A), accelerated urea method (B) and (C) accelerated clean method. Panel D, E and F, spectrum of digested α-lactalbumin with classic method (D), accelerated method (E) and (F) accelerated clean method.

IV.5.5 Application to a case study

To compare the sample treatments studied in this work, we test the identification of a cytochrome produced by D. desulfuricans ATCC 27774. This organism, which is a facultative nitrate/sulfate bacterium, considerably expresses a protein named Split-Soret [21] in the presence of nitrate. This fact suggests that this protein can be involved in the nitrate metabolism [22]. The production and purification of this protein was explained in section IV.4.5.1. A sample containing 0.3µg/µl of Split-Soret cytochrome c was used in this

set of experiments. The samples containing the protein were then submitted to the three methods studied in this work and the results presented in table IV.1 clearly demonstrate that the classic method and the accelerated classic method provided protein coverage and peptides matched that were not found statistically different (test t, $p > 0.05$, $n1 = 4$, $n2 = 4$). Figure IV.4 shows the spectra of the Split-Soret cytochrome c for the three sample methods used.

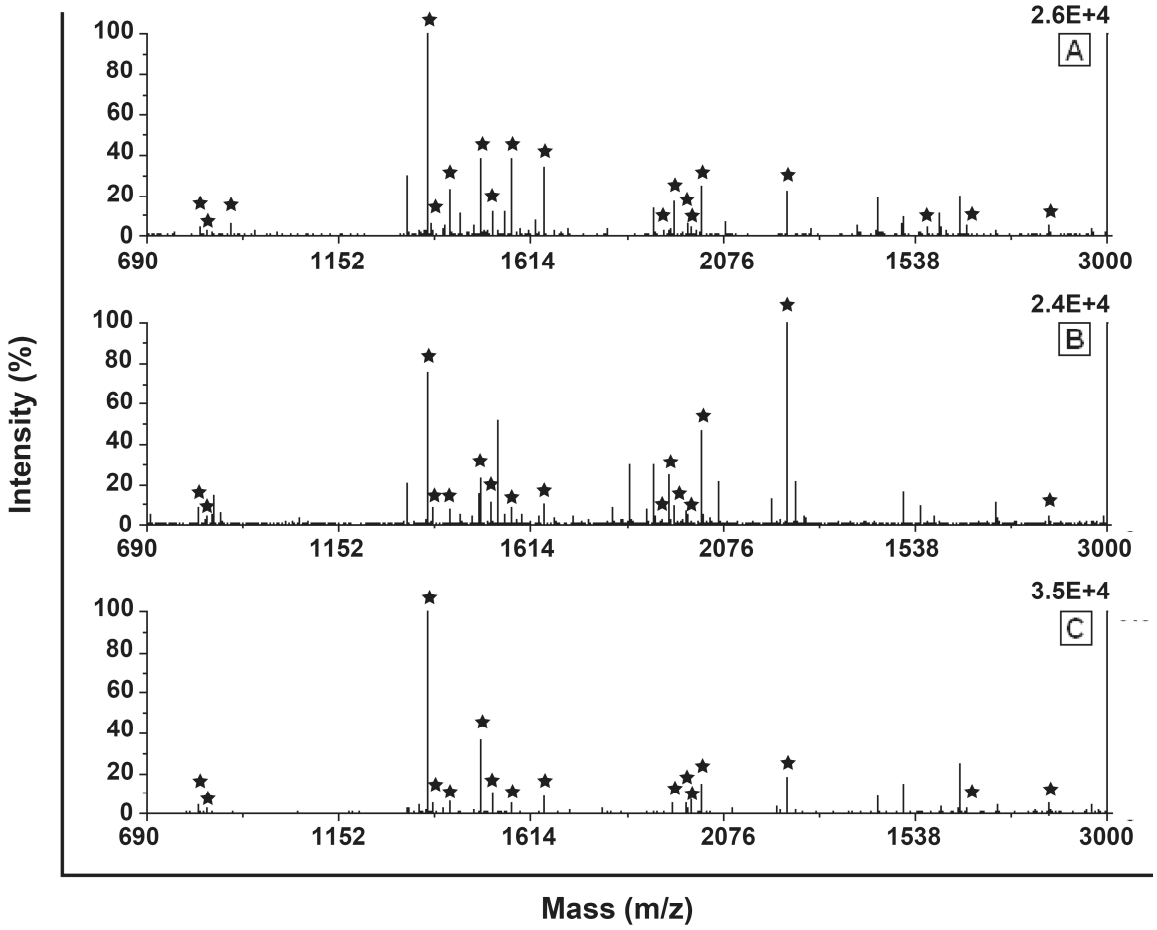


Figure IV. 4. MALDI spectra of Split-Soret cytochrome c from *D. desulfuricans,* (A) classic method, (B) accelerated urea method and (C) accelerated clean method.

IV.5.6 Final remarks

Table IV.2 shows the differences in time and handling for the three sample treatments compared in this work. As may be seen, if we consider the workflow as composed of five main steps as follows: (i) denaturation, (ii) reduction, (iii) alkylation, (iv) digestion and (v) desalting, handling is not the same. The clean method avoids the use of

ZipTips or any other kind of desalting processes. This means that it becomes also economically cheaper, since ZipTips are an expensive reagent. In terms of time consumed, the best is again the clean method, being necessary only 2min/sample to complete the workflow. The new ultrasonic multiprobe-device can work as efficiently as the single probe. In terms of throughput, however, it allows to work 4–12 times faster, depending on the multiprobe chosen to work with.

Table IV. 2. Comparison of handling and time consumed for the three methods studied in this work.

Method	Denaturation	Reduction	Alkylation	Digestion	Desalting	Total steps	Total time(h)[a]
Overnight (urea)	Urea	60min	45min	12h	yes	4	14
Ultrasonic (urea)	Urea	10min US	10min US	4min US	yes	4	6.5
Ultrasonic (H_2O/acetonitrile)	1min US	1min US	1min US	5min US	no	4	3

[a] Total time needed to complete the analysis for 96 samples (96-well plate).

It must be stressed that whilst an analysis runned using all the wells of a 96-well plate takes 12 h with the classic protocol, it last only for 3 h with the ultrasonic clean method. Due to simplicity of use and high throughput, it may be advanced that the ultrasonic multiprobe-device will be implemented in robotic platforms.

IV.6 Conclusions

The new ultrasonic multiprobe-device has been studied in conjunction with a 96-well plate in the acceleration of two different proteomic workflows in terms of speed, throughput, handling and robustness. We have demonstrated that to avoid cross contamination between samples in this approach, low amplitudes must be used. For the six standard proteins studied, the two workflows accelerated with ultrasound give results that were found similar in terms of robustness, as their utilization provide results comparable with a classic non-ultrasonic method. The clean fast method has the best performance in terms of speed and handling since only 2min/sample are necessary to complete it, being desalting not necessary, thus diminishing the total number of steps.

Regarding throughput, it has been proven that the combination of a 96-well plate and an ultrasonic multiprobe is a potential powerful tool in sample treatment for

proteomics, allowing high sample throughput. The methods proposed allow for rapid processing, minimizing the risk of contamination and reducing the chance of application errors. In addition, a potentially enormous number of different proteomics applications are advanced, such as fast and high throughput protein quantification using isotopic labeling [23]. Sample preparation steps, including reduction and alkylation, digestion, spotting on MALDI targets or transfer to LC/MS input plates can potentially be combined on a single automated platform making use of ultrasonic energy provided by ultrasonic multiprobes.

IV.7 Acknowledgments

H. M. Santos and R. Carreira acknowledge the doctoral grants SRFH/BD/38509/2007 and SFRH/BD/28563/2006, respectively, from FCT (Science and Technological Foundation) from Portugal. The research findings here reported are protected under patent pendings PCT/IB2006/052314 and PT 103 303. The University of Vigo is acknowledged for financial support under projects INOU/UVIGO/K919/2009 and INOU/UVIGO/K914/2009. J. L. Capelo and C. Lodeiro acknowledge the Isidro Parga Pondal Program from Xunta de Galicia, Spain.

IV.8 References

[1] D. Lopez-Ferrer, B. Canas, J. Vazquez, C. Lodeiro, R. Rial-Otero, I. Moura, J.L. Capelo, *Trac-Trends Anal. Chem.* 25 (2006) 996–1005.

[2] J.L. Capelo, R. Carreira, M. Diniz, L. Fernandes, M. Galesio, C. Lodeiro, H.M. Santos, G. Vale, *Anal. Chim. Acta* 650 (2009) 151–159.

[3] J. Havlis, H. Thomas, M. Sebela, A. Shevchenko, *Anal. Chem.* 75 (2003) 1300–1305.

[4] D. Lopez-Ferrer, J.L. Capelo, J. Vazquez, *J. Proteome Res.* 4 (2005) 1569–1574.

[5] M. Galesio, D.V. Vieira, R. Rial-Otero, C. Lodeiro, I. Moura, J.L. Capelo, *J. Proteome Res.* 7 (2008) 2097–2106.

[6] R. Rial-Otero, R.J. Carreira, F.M. Cordeiro, A.J. Moro, M.G. Rivas, L. Fernandez, I. Moura, J.L. Capelo, *J. Proteome Res.* 6 (2007) 909–912.

[7] H. Bao, T. Liu, X. Chen, G. Chen, *J. Proteome. Res. 7* (2008) 5339–5344.

[8] S. Wang, T. Liu, L. Zhang, G. Chen, *J. Proteome Res.* 7 (2008) 5049–5054.

[9] D. Lopez-Ferrer, K. Petritis, N.M. Lourette, B. Clowers, K.K. Hixson, T. Heibeck, D.C. Prior, L. Pasa-Tolic, D.G. Camp, M.E. Belov, R.D. Smith, *Anal. Chem*. 80 (2008) 8930–8936.

[10] S.P. Mirza, A.S. Greene, M. Olivier, *J. Proteome Res.* 7 (2008) 3042–3048.

[11] H. Lam, E.W. Deutsch, J.S. Eddes, J.K. Eng, S. King, E. Stein, R. Aebersold, *Proteomics* 7 (2007) 655–667.

[12] K.J. Eddes, J.S.L. Hohmann, J. Jackson, A. Peterson, S. Letarte, P.R. Gafken, J.E. Katz, P. Mallick, H. Lee, A. Schmidt, R. Ossola, J.K. Eng, R. Aebersold, D.B. Martin*, J. Proteome Res.* 7 (2008) 96–103.

[13] F.M. Cordeiro, R.J. Carreira, R. Rial-Otero, M.G. Rivas, I. Moura, J.L. Capelo, *Rapid Commun. Mass. Spectrom*. 21 (2007) 3269–3278.

Chapter V.

Decision peptide-driven: a free software tool for accurate protein quantification using gel electrophoresis and matrix assisted laser desorption ionization time-of-flight mass spectrometry

---

Published in:

Talanta, 82 (2010) 1412 – 1420

Hugo M. Santos, Miguel Reboiro-Jato, Daniel Glez-Peña, J. D. Nunes-Miranda, Florentino Fdez-Riverola, R. Carvallo, J. L. Capelo

Index

V.1 Resumo

A ferramenta *"Decision Peptide-Driven"* consiste numa aplicação bioinformática que tem como principal função assistir o utilizador no tratamento de dados para quantificação de proteínas seguindo os seguintes passos: (1) separação de proteínas por eletroforese em gel; (2) digestão em gel; (3) marcação isotópica directa e inversa com ^{18}O; (4) A análise das amostras marcadas por MALDI-TOF-MS. O software DPD, compara os resultados experimentais de marcação directa e inversa e identifica os péptidos que apresentam perdas reprodutíveis durante os diferentes passos do tratamento de amostra.

Os péptidos previamente seleccionados pelo programa DPD são posteriormente utilizados para quantificação exacta de proteínas. A interpretação manual dos dados de marcação directa e inversa obtida por MALDI-TOF-MS requer muito tempo. O software DPD reduz e simplifica significativamente o tempo necessário para a interpretação dos dados. Deste modo, com a introdução de vários espectros de MALDI no programa, o investigador é conduzido na comparação, automática, de dados de marcação isotópica directa e inversa calculando os rácios correspondentes. Numa segunda fase os rácios calculados são comparados de modo a identificar os péptidos que apresentam perdas paralelas. Os péptidos identificados podem posteriormente ser utilizados como padrão interno para quantificar proteínas de forma exacta.

Neste trabalho, o software DPD, é apresentado e explicado através da quantificação da proteína anidrase carbónica.

Palavras-chave: Quantificação de proteínas, eletroforese em, MALDI-TOF-MS, marcação isotópica com ^{18}O, DPD software.

A minha contribuição para este trabalho consistiu na execução da parte experimental, análise de MALDI-TOF-MS, processamento de dados, interpretação e desenvolvimento/teste do software.

V.2 Abstract

The Decision Peptide-Driven tool implements a software application for assisting the user in a protocol for accurate protein quantification based on the following steps: (1) protein separation through gel electrophoresis; (2) in-gel protein digestion; (3) direct and inverse ^{18}O-labeling and (4) Matrix assisted laser desorption/ionization time-of-flight mass spectrometry, MALDI. The DPD software compare the MALDI results of the direct and inverse ^{18}O-labeling experiments and quickly identifies those peptides with paralleled loses in different sets of a typical proteomic workflow. Those peptides are used for subsequent accurate protein quantification. The interpretation of the MALDI data from direct and inverse labeling experiments is time-consuming requiring a significant amount of time to do all comparisons manually. The DPD software shortens and simplifies the searching of the peptides that must be used for quantification from a week to just some minutes. To do so, it takes as input several MALDI spectra and aids the researcher in an automatic mode (i) to compare data from direct and inverse ^{18}O -labeling experiments, calculating the corresponding ratios to determine those peptides with paralleled losses throughout different sets of experiments; and (ii) allow to use those peptides as internal standards for subsequent accurate protein quantification using ^{18}O-labeling. In this work the DPD software is presented and explained with the quantification of protein carbonic anhydrase.

Keywords: Protein quantification, gel electrophoresis, MALDI-TOF-MS, ^{18}O, labeling, DPD software.

My contribution to this work was the execution of the experimental, MALDI-TOF-MS analysis, data processing, interpretation and software development/testing.

V.3 Introduction

Protein quantification methods based on protein separation by 1D gel electrophoresis and matrix assisted laser desorption/ionization time-of-flight mass spectrometry, MALDI, have a number of drawbacks that make reliable quantification difficult. The differences in the yields of protein digestion obtained in different sets of in-gel digestions, the variation observed in the recovery yields of individual peptides within a set of in-gel digest, the biased losses of peptides that might occur during the post-digestion sample processing of in-gel digests when are used ZipTip pipette tips to clean the sample or the speed vacuum pump to dry down and preconcentrate the sample. The aforementioned drawbacks can cause loss of peptides ranging in between 30% to 90% depending on the amount of sample loaded in the gel and on the type of peptide studied [1-3].

If for a given protein was possible to identify a certain number of peptides that had low and paralleled loses through a typical proteomic workflow entailing 1D-gel protein separation and in-gel protein digestion, then such peptides would allow robust and accurate protein quantification. The experimental method that could allow to extract and to identify the number of peptides that remains constant in expression level through a typical in-gel digestion workflow should be based in a peptide differential analysis. A variation of the method proposed by Wang *et al.* and called “inverse labeling” can be used to do such analysis [4]. With this procedure it is easily detected if a peptide is randomly loosed, or the observed yields of individual peptides vary strongly within a set of in-gel digests. This methodology can be used to unambiguously verify the yield of peptides obtained during in-gel protein digestion at different concentrations, and thus clearly illustrates which peptides can be used for quantification through a given dynamic range of differential quantification.

The application of the “inverse labeling” methodology requires the use of mass spectrometry. MALDI can be used for this purpose. The comparison of MALDI spectra makes this approach for protein quantification tedious and time-consuming.

To speed the treatment of data the software “Decision Peptide Driven”, DPD, has been developed based on previous software developed for medical applications [5] as a computer tool to extract and to identify the peptides that remains constant in expression level through different sets of a typical in-gel digestion workflow.

The present manuscript described in detail the software tool DPD, explaining through a real example how to use it. This software is freely source code available, and it can be run as a multiple platform.

V.4 Materials and Methods

V.4.1 Apparatus

Gel Electrophoresis was performed with an electrophoresis system, model Mini-PROTEAN Tetra Cell, from Biorad (Hercules, CA, USA), following the manufacturer instructions. Protein digestion and labeling were done in safe-lock tubes of 0.5 ml from Eppendorf (Hamburg, Germany).

A minicentrifuge, model Spectrafuge-mini, from Labnet (Madrid, Spain), and a minicentrifuge-vortex, model Sky Line, from ELMI (Riga, Latvia) were used throughout the sample treatment, when necessary. A vacuum concentrator centrifuge from UniEquip (Martinsried, Germany) model UNIVAPO100H with a refrigerated aspirator vacuum pump model Unijet II was used for (i) sample drying and (ii) sample pre-concentration.

Milli-Q natural abundance ($H_2{}^{16}O$) water was obtained from a SimplicityTM from Millipore (Milan, Italy). An ultrasonic bath, model Transsonic TI-H-5, from Elma (Singen, Germany) with control of temperature and amplitude was used to speed up the gel washing, the protein reduction and the protein alkylation steps, and a sonoreactor model UTR200, from Dr. Hielscher (Teltow, Switzerland), was used to accelerate the enzymatic digestion step.

All materials were used without further purification. α-Cyano-4- hydroxycinnamic acid, α-CHCA, *puriss* for MALDIfrom Fluka (Buchs, Switzerland) was used as MALDI matrix. ProteoMass Peptide MALDI-MS Calibration Kit (MSCAL2) from Sigma was used as mass calibration standard for MALDI-TOF-MS.

V.4.2 Standards and reagents

Reduction and alkylation were carried out, respectively, with D,L-dithiothreitol (DTT, 99%) and iodoacetamide (IAA) from Sigma. The following reagents were used during sample digestion: ammonium bicarbonate buffer (AmBic, pH 8.5, ≥99.5%) and formic acid (FA, ~98%) from Fluka (Buchs, Switzerland); and labeling: ammonium acetate (>99.0%) from Fluka, calcium chloride (~97%, anhydrous) from Sigma, Mag-Trypsin from Clontech (USA) and $H_2{}^{18}O$ (97 atom%) from Isotec^{TM} (Miamisburg, USA). Trifluoroacetic acid (TFA, 99%) was obtained from Riedel-de Haäen (Seelze, Germany).

V.4.3 In-gel protein digestion

Ultrasonic in-gel enzymatic digestion was done according to the ultrafast proteolytic digestion protocol previously developed in our laboratory [6, 7]. Protein bands were manually excised from the gel and placed in safe-lock tubes of 0.5 mL. Gel pieces were washed, first with AmBic 25mM/acetonitrile (100µL) and then with acetonitrile (100µL), in an ultrasonic bath operating at 35kHz (60% amplitude) for 5min for each step. Then, the gel pieces were dried in a vacuum concentrator centrifuge for 5min. Protein reduction and alkylation steps were done as follows: disulfide bonds from cystine residues were reduced with DTT in an ultrasonic bath operating at 35kHz (60% amplitude) for 5 min at room temperature, and then, the reduced cysteines were blocked with IAA in an ultrasonic bath operating at 35kHz (60% amplitude) for 5min at room temperature. After reduction and alkylation steps, the gel was submitted again to the washing procedure in the same way as described above, followed by another dry step of 10min. Afterward, the dried gel pieces were incubated with trypsin (375 ng in 25µL) in an ice bath for 60min to rehydrate the gel and to allow enzyme penetration into it. Subsequently, in-gel protein digestion was performed in a sonoreactor operating at 50% amplitude for 4min. Next, trypsin activity was stopped by the addition of 20µL of formic acid 5% (v/v).

V.4.4 ^{18}O labeling: the decoupled procedure

For the ^{18}O-labeling, the digested peptides were reconstituted with 10µL of 25mM calcium chloride and 10µL of (acetonitrile 20% v/v + 50 mM ammonium acetate, pH 6.75) Then the samples were vacuum dried again, and after evaporation the dried samples were reconstituted in 5µL of natural abundance water or 97% ^{18}O-enriched water and 5µL of a 5% suspension of Mag-Trypsin in $H_2{}^{16}O$ or $H_2{}^{18}O$ were added. The digested peptides were labeled during 15 min of vortexing and centrifugation and finally trypsin were removed by a magnetic separation. A detailed explanation of this procedure can be found elsewhere [8].

V.4.5 Inverse ^{18}O labeling of peptides

Proteins were separated by 1D-PAGE and then submitted to the protocols described in V.4.3 and 2.4 and then the inverse 18O labeling protocol as described by Wang *et al.* [4] was then used.

V.4.6 Quantification of peptides

Quantification of peptides through ^{18}O was done with the mathematical algorithm for deconvolution described by Yao *et al.* [8] Eq. 1, Reduction of the spectra to a centroided plot was done using the centroiding option function of the Data ExplorerTM software (version 4.0) from Applied Biosystems. This function is an advanced peak filtering method that improve mass spectral data quality and reduce data file size. Profile data, in which many points are used to delineate a mass spectral peak, is converted into mass-centroided data by a data compression algorithm. The centroided mass peak is located at the weighted center of mass of the profile peak. The normalized area of the peak provides the mass intensity data.

$$\left(\frac{^{16}O}{^{18}O}\right) = \frac{I_0}{I_4 - \frac{M_4}{M_0} I_0 + \left(1 - \frac{M_2}{M_0}\right) I_2 - \left(1 - \frac{M_2}{M_0}\right)\frac{M_2}{M_0} I_0} \quad \text{(Eq. 1)}$$

where M_0, M_2 and M_4 correspond to the theoretical relative intensities of the monoisotopic peak and the monoisotopic peaks with masses 2Da and 4Da higher, respectively; and I_0, I_2 and I_4 are the measured relative intensities of the first, the third and the fifth peaks in the isotopic cluster.

V.4.7 Case study

To explain how to work with the DPD program we have follow a real example based in the standard protein carbonic anhydrase.

V.4.8 MALDI analysis

Prior to MALDI analysis, the sample was mixed with the matrix solution. α-CHCA matrix was used throughout this work and was prepared as follows: 10mg of α-CHCA was dissolved in 1mL of Milli-Q water/acetonitrile/TFA (1mL/1mL/2μL). Then, 4μL of the aforementioned matrix solution was mixed with 4μL of sample and the mixture was shaken in a vortex for 30s. One micro liter of each sample was hand-spotted on a well of a MALDI sample plate and was allowed to dry. A MALDI system model Voyager DE-PRO Biospectrometry Workstation equipped with a nitrogen laser radiating at 337nm from Applied Biosystems (Foster City, CA) was used to acquire the PMFs.

Measurements were done in the reflector positive ion mode, with a 20kV accelerating voltage, 75.1% grid voltage, 0.002% guide wire and a delay time of 100ns. Two close external calibrations were performed with the monoisotopic peaks of the bradykinin, angiotensin II, P14R, and ACTH peptide fragments (m/z: 757.3997, 1046.5423, 1533.8582, and 2465.1989, respectively). Monoisotopic peaks were manually selected from each of the spectra obtained. Mass spectral analysis for each sample was based on the average of 500 laser shots.

V.4.9 Software distribution

Current versions of the software and their supporting user manuals are freely available for downloaded and use, without restriction, via the internet at http://sing.ei.uvigo.es/DPD. This program operates on excel comma-separated values, CSV, files with centroid mass and relative intensity data extracted from the Data ExplorerTM. The program was developed based on previous work related to cancer diagnosis [23] and on the suggestions given by the Bioscope group.

V.4.10 Decision peptide-driven experimental workflow

A schematic diagram illustrating the sequential steps of the sample treatment workflow is presented in Figure. V.1. In brief, 1μg and 2μg, four replicates each, are loaded and separated by 1D-SDS-PAGE. The bands are then excised from the gel and the proteins are in-gel tryptic digested with the aid of ultrasonic energy as reported by Galesio et al. [6]. The pool of peptides thus obtained is then dried and finally, reconstituted in normal water

or in 97% ^{18}O water [7, 10-14]. Following the pipeline of figure V.1, after protein separation, in-gel protein digestion and peptide labeling, the next step is to perform the so-called inverse labeling protocol [4]. With this procedure two converse labeling experiments are performed in parallel as follows. In the "direct" labeling, the sample is reconstituted in normal water whilst its counterpart of higher amount (i.e. 1:2) is reconstituted in ^{18}O-water as described in 2.4. In the «inverse» method the labeling is done conversely. Finally, an equal sample volume of non-labeled and its labeled counterpart are mixed and analysed through MALDI-TOF-MS. The $^{16}O/^{18}O$ peak ratios (MALDI relative peak intensity) are used then in the final step of the workflow, as it is shown in figure. V.1. In this step, the software DPD (Decision Peptide Driven) is used to find out which peptides are adequate for protein quantification within a given accuracy. In brief, this software compares the labeled to unlabeled ratios of the same peptides obtained in the "direct" and "inverse" methods. Only those peptides having the "direct" and "inverse" ratios within a given *p* significance level (*t*-test) are selected for quantification.

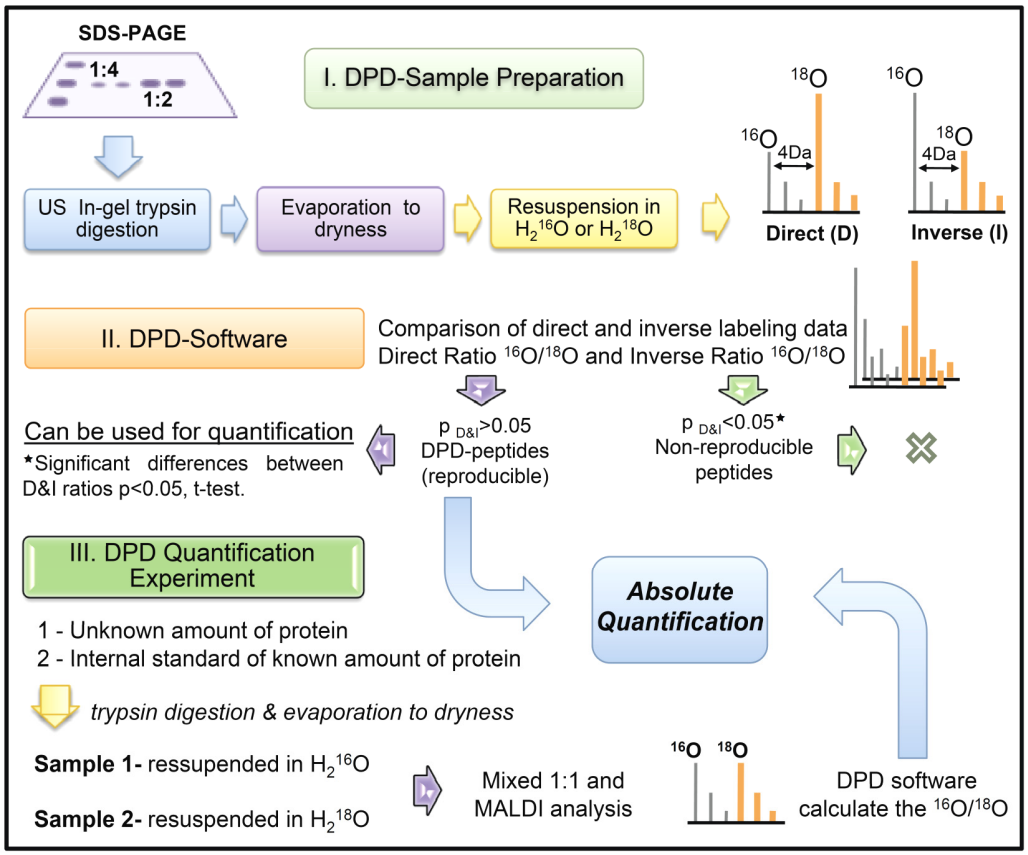


Figure V. 1. Experimental workflow for the identification of DPD (decision peptide driven) peptides and subsequent protein quantification.

V.5 Results and Discussion

The DPD software is explained in detail in the following sections. The data which is needed to enter in each step of the software, the reasons why, and the information which is obtained once each step is completed is described below. To run the program, real data is provided in the supplementary material, V.9.SM.1. This example is based in the quantification of carbonic anhydrase.

V.5.1 Preparing theoretical data

V.5.1.1 How to obtain the in-silico protein digestion

After 1D-Gel electrophoresis separation, the protein is in-gel digested with the aid of an enzyme, generally trypsin. This process will generate a pool of peptides which can be previously known, because the enzymes used to digest proteins do the cleavage always in the same residues of the peptide chain. For instance, the enzyme trypsin cleavages the proteins in the amino acids arginine and lysine, if they are not followed by a proline residue. In other words, if the protein to be quantified and the enzyme to be used are known, the pool of peptides expected can be obtained in advance. Nowadays there are powerful software tools that can provide the above mentioned theoretical pool of peptides (http://www.expasy.org/sprot/). When the DPD program is used, the first step consists in the introduction of this theoretical pool of peptides. This is necessary because the program will compare the masses of those theoretical peptides with the masses of the peptides obtained using MALDI. This comparison has the goal to assign the experimental masses obtained with the MALDI with their corresponding (theoretical) peptides. A step-by-step description of how to obtain the *in-silico* protein digestion is provided in supplementary material V.9.SM.2.

V.5.1.2 How to obtain the isotopic mass distribution of the peptides

The *in-silico* digestion of the protein provides a list of theoretical peptides. The IMD, isotopic mass distribution, of those peptides, M_0, M_2, M_4, is required for the subsequent calculation of the ratios between the ^{18}O-labeled and the non-labelled peptides in following parts of the workflow, as it is shown in Eq. 1 (see above).

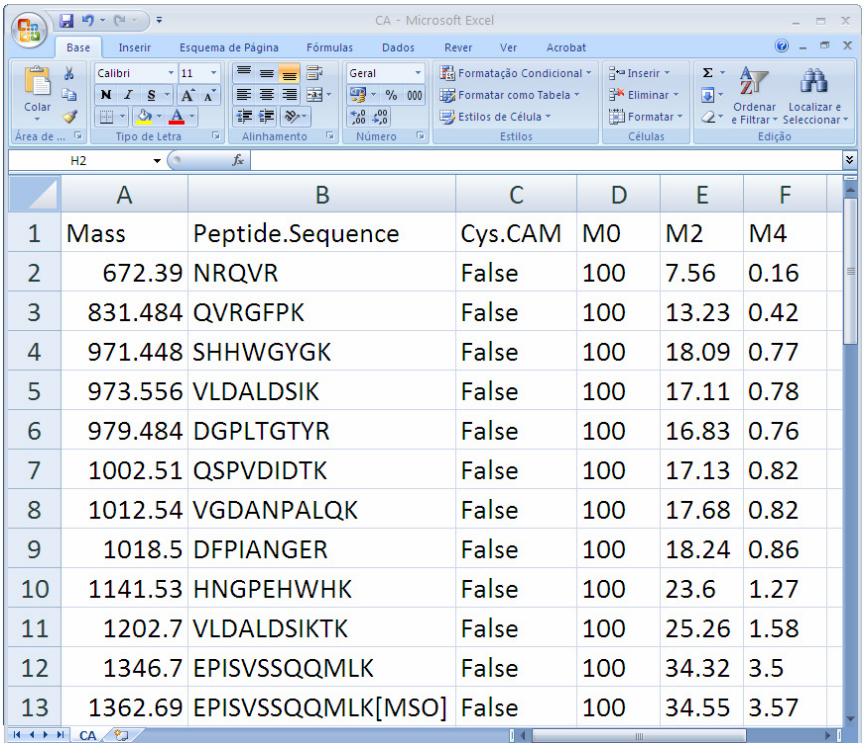
For each peptide the IMD is calculated in an automated mode in the following webpage http://prospector.ucsf.edu/cgi-bin/msform.cgi?form=msisotope.

A step-by-step description of how to obtain the isotopic mass distribution of the peptides is provided in supplementary material, V.9.SM.3.

V.5.2 Finding out the DPD peptides

V.5.2.1 First step: loading the in-silico data

With the data obtained in V.5.1.1 and V.5.1.2 a CSV excel file is created containing six columns. The first column has the theoretical masses of the peptides, the second column contains the theoretical peptide amino acid sequences, and the third column contains the possibility of carboxyamidomethilations on cysteine due to alkylation. The fourth, fifth and sixth columns, contains the IMD of the M0, M2 (M0+2Da) and M4 (M0+4Da) masses respectively obtained for each peptide. Figure V.2 shows an example of an *in-silico* file from carbonic anhydrase.



	A	B	C	D	E	F
1	Mass	Peptide.Sequence	Cys.CAM	M0	M2	M4
2	672.39	NRQVR	False	100	7.56	0.16
3	831.484	QVRGFPK	False	100	13.23	0.42
4	971.448	SHHWGYGK	False	100	18.09	0.77
5	973.556	VLDALDSIK	False	100	17.11	0.78
6	979.484	DGPLTGTYR	False	100	16.83	0.76
7	1002.51	QSPVDIDTK	False	100	17.13	0.82
8	1012.54	VGDANPALQK	False	100	17.68	0.82
9	1018.5	DFPIANGER	False	100	18.24	0.86
10	1141.53	HNGPEHWHK	False	100	23.6	1.27
11	1202.7	VLDALDSIKTK	False	100	25.26	1.58
12	1346.7	EPISVSSQQMLK	False	100	34.32	3.5
13	1362.69	EPISVSSQQMLK[MSO]	False	100	34.55	3.57

Figure V. 2. *In-silico* file -excel CSV- containing peptide masses, the theoretical peptide sequence assigned to each mass, carboxyamidomethilations present (yes-true or no-false) and the isotopic mass distribution for the protein Carbonic Anhydrase.

When the DPD program is started the interface shown in figure V.3 appears in the screen. By clicking in the "*load in-silico data*" button we are asked to introduce the CSV excel file containing the *in-silico* information.

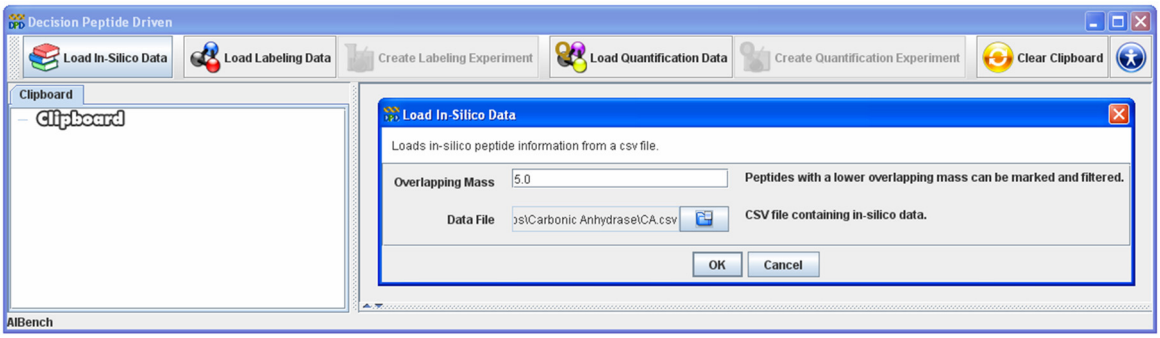


Figure V. 3. By clicking in "load *in-silico* data" the data created as described in sections V.5.1.1 and V.5.1.2 is introduced in the form of a CSV excel sheet.

In our example the file to input is the one named as "*in silico* carbonic anhydrase" given in the supplementary material, and prepared as described in sections V.5.1.1 and V.5.1.2. In addition, the user must introduce an "overlapping mass" value that indicates the mass difference considered critical between two peptides whose isotopic mass distribution can potentially be overlapped, in our case 5Da is the value chosen. This concept can be explained with the simplest case of overlapping as follows: the *in-silico* digestion of carbonic anhydrase predicts the occurrence of YGDFGTAAQQPDGLAVVGVFLK (2253.16 m/z) and RMVNNGHSFNVEYDDSQDK (2254.98m/z), their isotopic mass distribution are overlapped, after isotopic labeling, if both are present in the experimental data . Nevertheless, could also happen that the peak 2253.16 m/z is present in the experimental data but not the peak 2254.98 m/z or vice versa. In this case, the peptide virtually could be used for quantification. Therefore those peaks will be assigned as potential overlapped ("true") by the program and they must be checked in the spectrum to assess whether overlapping occurs or not.

V.5.2.2 Second step: loading experimental data

It was explained in the experimental section that two converse experiments are done to identify the peptides that have paralleled loses in different sets of experiments. Of each set of direct (n=4) and inverse (n=4) labeling experiments, MALDI spectra are obtained, showing the typical pattern of labeled and non-labeled peptides (see figure V.4). Those

spectra are converted in a CSV excel file in which one column contents the peak mass values whilst other contents their corresponding intensities, as shown in figure V.5. By clicking in "load labeling data", see figure V.6, the program asks for the introduction of the files corresponding to the direct labeling. In our example, those files are named in the supplementary material as direct 1, direct 2, direct 3 and direct 4.

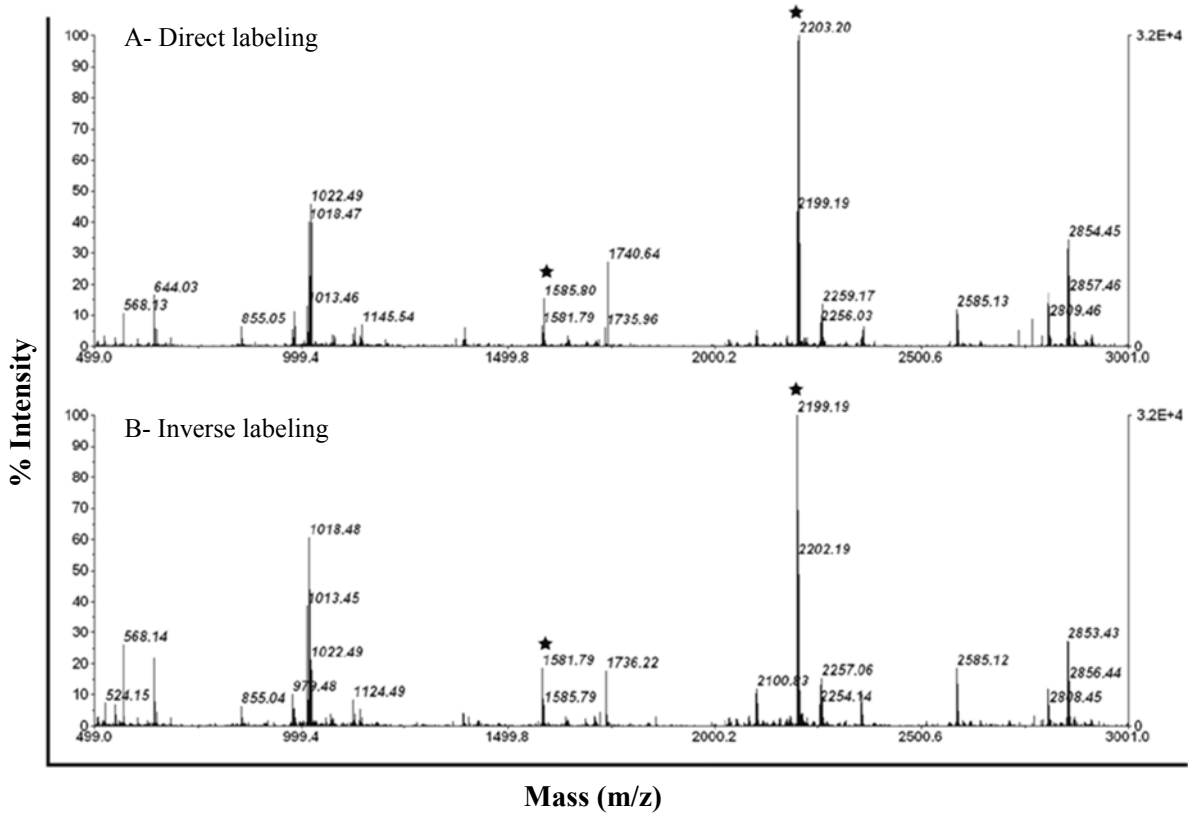


Figure V. 4. MALDI spectra of carbonic anhydrase obtained for the direct (1µg of unlabeled protein and 2 µg of ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis) and inverse labeling (2µg of ^{18}O-labeled protein and 1µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis).The symbol("★") denotes the peptides assigned by the DPD software

Once the direct files have been loaded, then the files corresponding to the inverse labeling are also loaded. In supplementary material the corresponding files are named as inverse 1, inverse 2 inverse 3 and inverse 4. The DPD program has an algorithm that allows the user to choose the minimum peak intensity from which the masses will be considered different of instrumental noise. In other words, it is possible to select the minimum intensity from which a peak will be considered as belonging to a peptide. This value will depend on the quality of the MALDI spectra. In our example the recommended value is 3%, which represents the percentage of the maximal relative peak intensity.

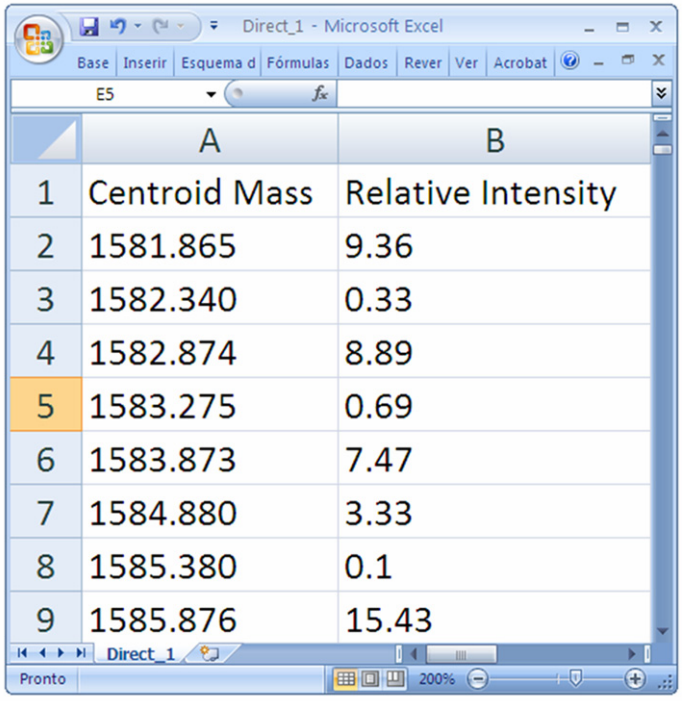


Figure V. 5. CSV excel sheet in which one column contents the peak mass values whilst other contents their corresponding MALDI intensities.

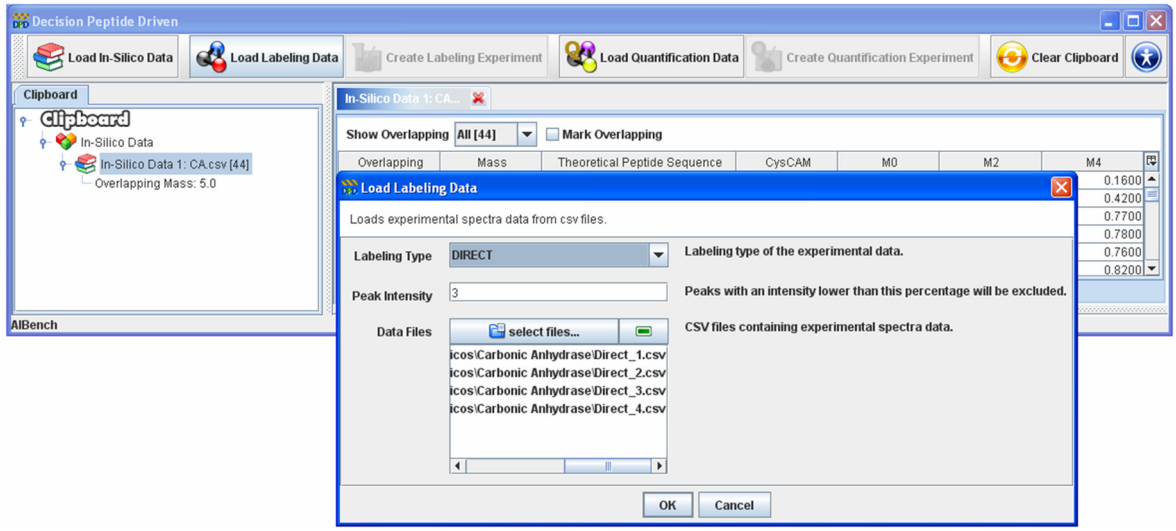


Figure V. 6. By clicking in "load labeling data" the data created as described in sections 3.3 for the direct (n=4) and inverse (n=4) experiments are introduced in the form of a CSV excel sheet. The user is also asked to introduce a value for the minimum peak intensity for which a peak will be considered as different from noise.

V.5.2.3 Third step: creating a label experiment

At this stage, the list of *in-silico* peptide masses and the list of experimental m/z values (corresponding most likely to peptides) of the direct and inverse experiments have been loaded in the program. In the next step, by clicking in "create labeling experiment" a dialog box will appear and the user is asked to define an error margin via the "peptide tolerance" box. This parameter is instrument dependent and is directly linked to the accuracy that it is expected to obtain in the MALDI system. In our conditions, a typical value to be assigned to this parameter is 0.25 (corresponding to an accuracy of 0.25Da). Once a experimental peptide mass is matched with its theoretical value, M0, the system checks if the spectrum also contains that mass plus 2Da , M2, and 4Da, M4, corresponding to one and two ^{18}O incorporations, respectively. The ratio between the intensities of M0 and M4 (Ratio I4/I0 in the dialog box) is used to discharge natural occurring peptides. This is because the peptides that we are comparing correspond to mixtures of labeled and non labeled peptides, the intensity ratios between M0 and M4 are different of the natural occurring M0 and M4 ratios. A typical value assigned to this parameter is 0.15. Every time that M0, M2 and M4 are found within the given values of 0.25 for peptide tolerance and 0.15 for ratio M0/M4, the peptide is selected as a candidate to be considered a reproducible peptide. Otherwise, the mass is discharged.

V.5.2.4 Fourth step: intersect peptides

At this stage, the lists of masses from the direct and inverse experiments include only the peptide masses that (i) have matched the corresponding *in-silico* mass, (ii) that have the masses corresponding to the 2 and 4Da shift caused by the single and double ^{18}O incorporation, within a given peptide mass tolerance of 0.25Da and a experimental overlapping of 1.00Da and (iii) that have a I4/I0 ratio over a given threshold, 0.15 in this case. Now the direct and inverse lists are compared to select the common masses, this is, the masses corresponding to peptides that are found in both direct and inverse experiments. In addition, a labeling ratio ($\frac{^{16}O}{^{18}O}$) is calculated as explained in Eq. 1 for these peptides taking into account the isotope mass distributions M0, M2 and M4, and the corresponding intensities measured in the experimental data I0, I2 and I4, respectively. By following this criterion, DPD software generates a list of common peptides for both direct and inverse

samples, along with their corresponding non-label to label ratios. The ratio must reflect the relation between the amount of label and non-label protein as established at the beginning of the experiment. In our example, the amount of protein labeled was twice the amount of protein non-labeled, therefore the direct ratio is 0.5 (1/2), whilst the inverse ratio is 2 (2/1). To facilitate the comparison the DPD program shows the inverse ratio as (inverse ratio)$^{-1}$, thus the expected values in our case are 0.5 for both ratios. In addition, the average amount of protein ratios is given with their corresponding relative standard deviations, RSDs. Finally, to find the peptides that have similar ratios in the direct and in the inverse method, it is needed to click in the "execute experiment" tool bar. Now the program asks which threshold level of RSD is required. The program compares the medium values and provides a relative standard deviation, $RSD_{D\&I}$, that arise the difference in % between the medium values. Thus, the peptides with direct and inverse values within a chosen difference (for instance less than 10%) can be easily selected for quantification, as showed in figure V.7.

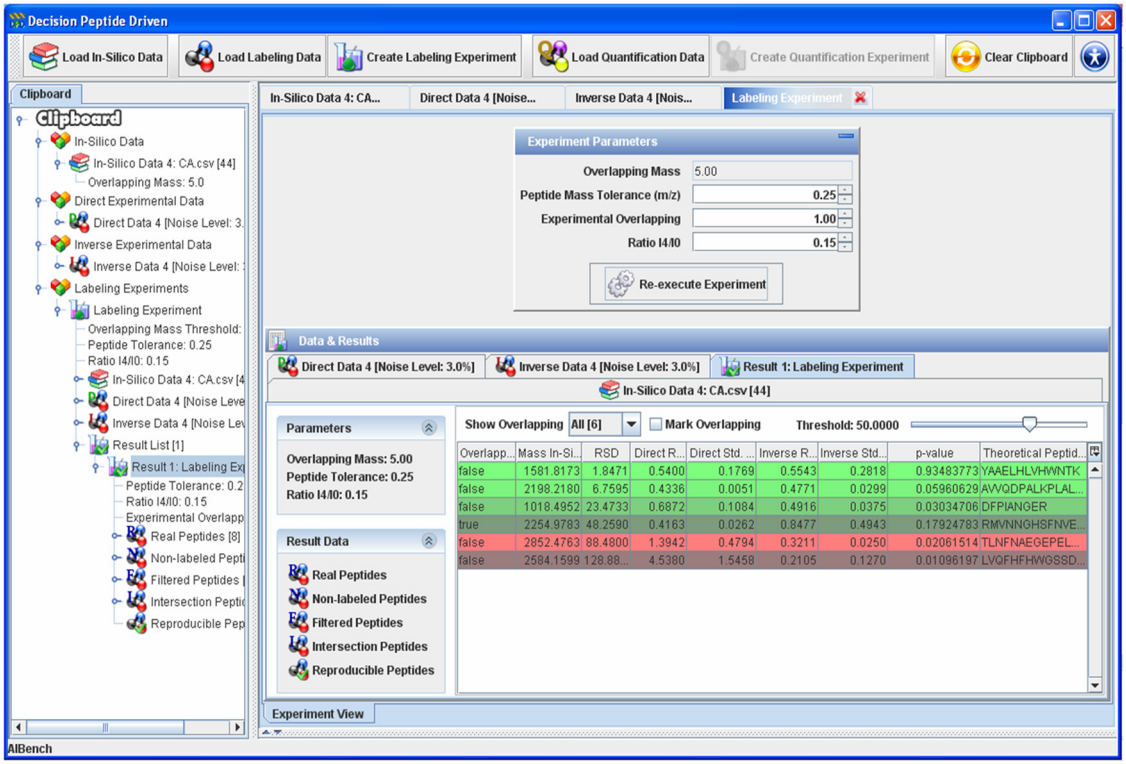


Figure V. 7. This figure shows the result of the execution which is a report containing the values of the input parameters and the generated output data including: (*i*) a list of peptides identified in the raw data input files, (*ii*) a list of non-labeled peptides, (*iii*) a list of filtered peptides, (*iv*) a list of intersection peptides and finally (*v*) a list of reproducible peptides, those that can be used for quantification.

In addition, the program also provides a *p* value. This *p* is obtained by comparing the direct and inverse ratios for each peptide. If $p>0.05$, then both values can be considered statistically non-different, and the peptide can be considered valuable for protein quantification.

The software, allows the user to change the experiment parameters in order to perform multiple analyses on the same data. Every time the user selects the "Execute Experiment" button, a new result is added to the experiment containing all the information generated during the process. In this manner the variables "peptide mass tolerance", "experimental overlapping" and ratio I4/I0 can be changed as desired. Before starting the execution, the user must specify the RSD threshold parameter used to highlight those peptides that are useful (under the statistical threshold) and those that are invalid (over the threshold). In our example the peptides recovered with a RSD between direct and inverse methods below 10% correspond to peptides YAAELHLVHWNTK and AVVQDPALKPLALVYGEATSR. If the RSD chosen is changed to 50% the peptides are now YAAELHLVHWNTK, AVVQDPALKPLALVYGEATSR, DFPIANGER and RMVNNGHSFNVEYDDSQDK. It is noteworthy that only two peptides can be used to accurately (below 10%) quantify the protein.

V.5.3 Quantification experiment

The user can proceed to load the data to be used for protein quantification through the "Load Quantification Data" toolbar button (please be sure that the file containing the *in-silico* data of the protein has been introduced). The *noise level* (as peak intensity) parameter and the files containing the spectra are introduced in this step. Following our example, the recommended peak intensity value is 3 and the files to be introduced are named as Direct_1, 2, 3 and 4 corresponding to MALDI data of four independent samples of carbonic anhydrase provided in supplementary material. Now, by clicking in "create quantification experiment" the *amount of internal standard* used for quantification, *peptide tolerance*, *ratio I4/I0* and the *experimental overlapping* are introduced as displayed in figure V.8. In our case, the following parameters were introduced: Peptide tolerance, 0.25; Ratio I4/I0, 0.15; amount of internal standard 2.08μg.

The detailed description of how this sample was treated is explained in the experimental section. Once the quantification experiment was done, a list of peptides and

the corresponding calculated amount of protein are generated by clicking in "execute experiment".

As in the labeling experiment, the application allows the user to change the experiment parameters in order to perform multiple analyses using the same data. Every time the user selects the "Execute Experiment" button, a new result is added to the experiment containing all the information generated during the process.

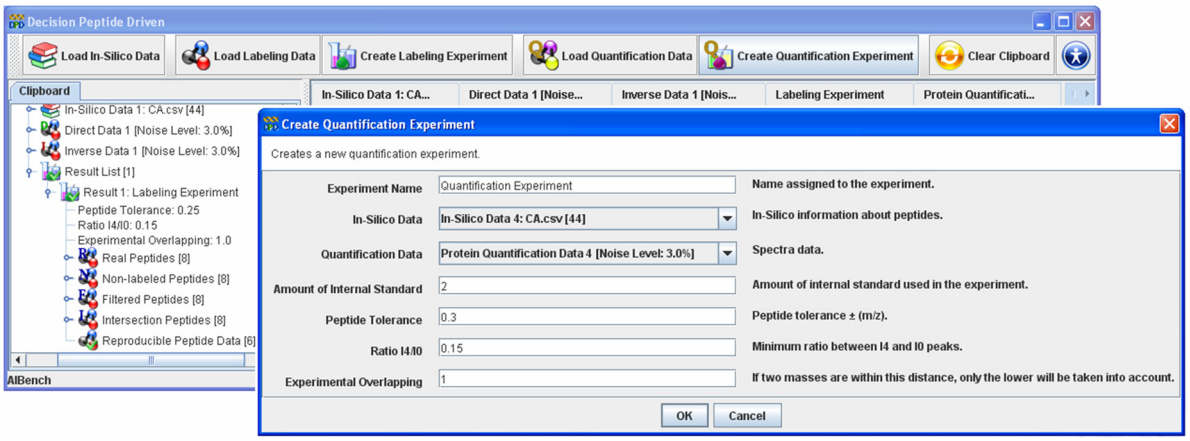


Figure V. 8. DPD application showing a ready-to-run quantification experiment where the user-defined parameters can be specified.

The result of the execution is a report containing the values of the input parameters and the generated output data including: (*i*) a list of peptides identified in the raw data input files, (*ii*) a list of non-labeled peptides, (*iii*) a list of filtered peptides, (*iv*) a list of intersected peptides and (*v*) the protein quantification ratios. Figure V.9 shows the results of the experiment in the working zone panel.

Now it is necessary only to check the peptides that have been previously identified as the DPD peptides, this is, the peptides that can be used for quantification. In our example using peptides YAAELHLVHWNTK (1581.82 m/z) and, AVVQDPALKPLALVYGEATSR (2198.21 m/z) the amounts of protein calculated are 1.12µg and 0.90 µg respectively, corresponding to the amount of protein loaded into the gel.

Table V.1 shows the amount of protein found for the different peptides selected in this work.

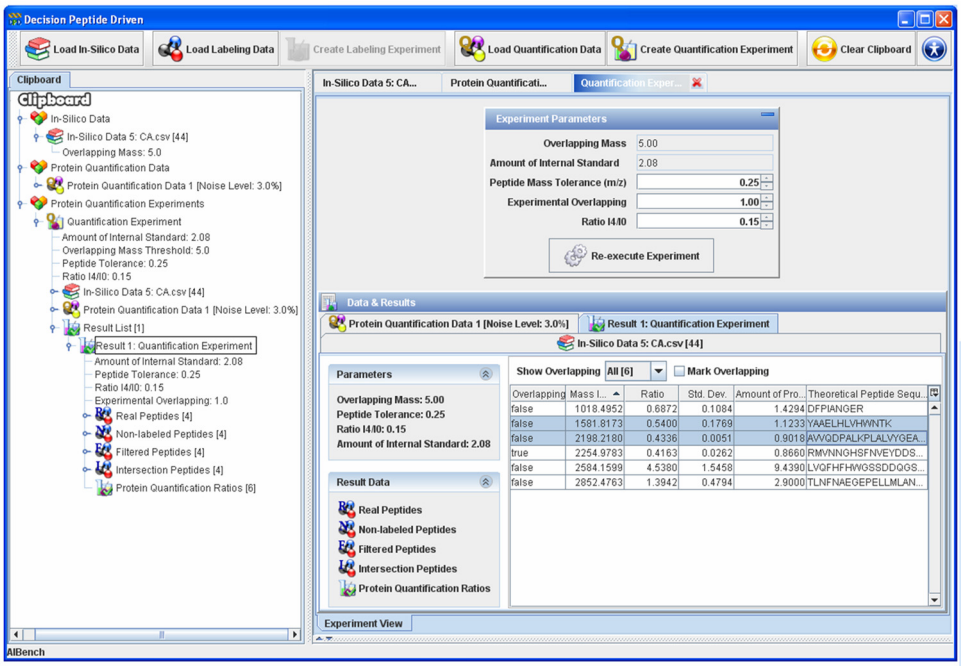


Figure V. 9. Protein quantification results computed by the DPD application once an experiment is executed

Table V. 1. Peptides assigned by the DPD software as candidates for quantification of carbonic anhydrase. Direct labeling: 1.04µg of unlabeled protein and 2.08µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1.04/2.08). Inverse labeling: 1.04µg of ^{18}O-labeled protein and 2.08µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: Peak intensity - 3; Peptide mass tolerance ±0.25Da; experimental overlapping - 1; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test); ‡ Expected values.

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$ value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (1.04)‡	Inverse - µg of protein (2.08)‡
1581.82	0.8	0.54±0.18	0.55±0.28	0.97	YAAELHLVHWNTK	2-10%	1.0±0.2	2.0±0.2
2198.22	5.8	0.43±0.01	0.48±0.03	0.18	AVVQDPALKPLALVYGEATSR			
1018.50	23.6	0.69±0.11	0.49±0.04	0.03	DFPIANGER	20-40% ⋆	1.1±0.4	1.7±0.6
2098.88	48.3	0.42±0.03	0.9±0.5	0.18	MVNNGHSFNVEYDDSQDK			
2852.48	91.1	1.4±0.5	0.3±0.1	0.02	TLNFNAEGEPELLMLANWRPAQPLK	> 90% ⋆	6±5	4.2±1.1
2584.16	128.9	4.5±1.6	0.2±0.1	0.01	LVQFHFHWGSSDDQGSEHTVDR			

V.5.4 Merits and limitations

The present program has been developed specifically for proteins separated by 1D-gel electrophoresis. However the program can be potentially used for proteins separated

through 2D-gel electrophoresis or for proteins separated by HPLC. Another merit is the possibility to adapt the program to other type of labeling. The main limitation of this program is that it has been developed for MALDI ionization systems.

V.6 Conclusions

We have developed a friendly software to help in an automated mode to identify those peptides that have paralleled loses through a typical proteomic workflow. The use of such peptides allow robust and accurate quantification of proteins using 1D-gel electrophoresis an matrix assisted laser desorption ionization time of flight mass spectrometry. Those peptides have been named as "decision peptide driven", DPD peptides.

The software presented in this work allows for the identification of DPD peptides and is based in a series of steps entailing different algorithms that perform in an automated mode a peptide differential analysis to extract and to identify the number of peptides that remains constant in expression level through different sets of a typical in-gel digestion workflow as the one described in this work. The DPD software saves times, allowing the user to accurate quantify proteins in an automated mode, overcoming the long time needed when the treatment of data is done manually. In addition the DPD software has a wizard easy to follow for its installation. Furthermore, the interface has been done in an easy-to-follow mode, and therefore the skills required for any potential operator are reduced to know how to apply the sample treatment procedure.

The installation wizard is available from the DPD web site as an executable file that depends on the final user operating system: Windows, Linux or MAC. By executing the setup file, the installation wizard will be automatically launched. If the user does not have the required Java Runtime Environment (JRE) installed in the computer, the installation wizard will first install this component, and then it will continue with the DPD installation. The user has to simply follow the instructions on the screen to successfully complete the installation.

V.7 Acknowledgements

H. M. Santos acknowledges the doctoral grant SRFH/BD/38509/2007 provided by FCT (Fundação para a Ciência e a Tecnologia - Portugal). This work was partially funded by the project *Research on Translational Bioinformatics* (08VIB6) from University of

Vigo. D. Glez-Peña was supported by a María Barbeito contract from Xunta de Galicia. M. Reboiro-Jato was supported by a PhD grant from Xunta de Galicia. Dr. José-Luis Capelo-Martínez and Dr. Carlos Lodeiro are grateful to the Xunta de Galicia (Spain) for the program Isidro Parga Pondal and the University of Vigo for financial support under projects InOu-Univ.Vigo 2009-K915 and K914. Xunta de Galicia (Spain) is also acknowledged for financial support under project 09CSA043383PR-2009. The research findings here reported are protected under patent pending PCT/IB2006/052314 and PT 103 303.

V.8 References

[1] J. Havlis, A. Shevchenko, *Anal. Chem.* 76 (2004) 3029–3036.

[2] I. I. Stewart, T. Thomson, D. Figeys, *Rapid Commun. Mass Spectrom.* 15 (2001) 2456–2465.

[3] K. Speicher, O. Kolbas, S. Harper, D. Speicher, *J. Biomol. Tech.* 11 (2000) 74–86.

[4] Y. K. Wang, Z. Ma, D. F. Quinn, W. F. Emil, *Anal. Chem.* 72 (2001) 3742–3750.

[5] The Artificial Intelligence workBench [http://www.aibench.org/].

[6] M. Galesio, D.V. Vieira, R. Rial-Otero, C. Lodeiro, I. Moura, J. L. Capelo, *J. Proteome Res.* 7 (2008) 2097–2106.

[7] R. Carreira, C. Lodeiro, M. S. Diniz, I. Moura, J. L. Capelo, *Proteomics* 9 (2009) 4974–4977.

[8] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, *Anal. Chem.* 73 (2001) 2836–2842.

[9] D. Glez-Pena, F. Diaz, J.M. Hernandez, J.M. Corchado, F. Fdez-Riverola BMC Bioinformatics, In press. doi:10.1186/1471-2105-10-187

[10] C. A. White, O. Nicomedus, A. Emili, *J. Proteome Res.* 7 (2009) 3653–3665.

[11] X. D. Yao, C. Afonso, C. Fenselau, *J. Proteom. Res.* 2 (2003) 147–152.

[12] J. L. Capelo, R. J. Carreira, L. Fernandes, C. Lodeiro, H. M. Santos, J. Simal, *Talanta* 80 (2010) 1476–1486.

[13] R. J. Carreira, R. Rial-Otero, D. López-Ferrer, C. Lodeiro, J. L. Capelo, *Talanta,* 76 (2008) 400–406

[14] H. M. Santos, C. Mota, C. Lodeiro, I. Moura, I. Isaac, J. L. Capelo, *Talanta* 77 (2008) 870–857.

V.9 Supplementary material

V.9.SM.1 Step by step description for the preparation of the data to use in the DPD software.

Files description:

The file named as CA, it contains (i) the theoretical list of peptides of carbonic anhydrase after the *in-silico* digestion with trypsin and (ii) the theoretical isotopic distribution of those peptides.

The files named as direct 1, direct 2, direct 3 and direct 4 corresponds to 4 different sets of carbonic anhydrase digestions (direct method 1:2).

The files named as inverse 1, inverse 2, inverse 3 and inverse 4, corresponds to different sets of carbonic anhydrase digestions (inverse method 2:1).

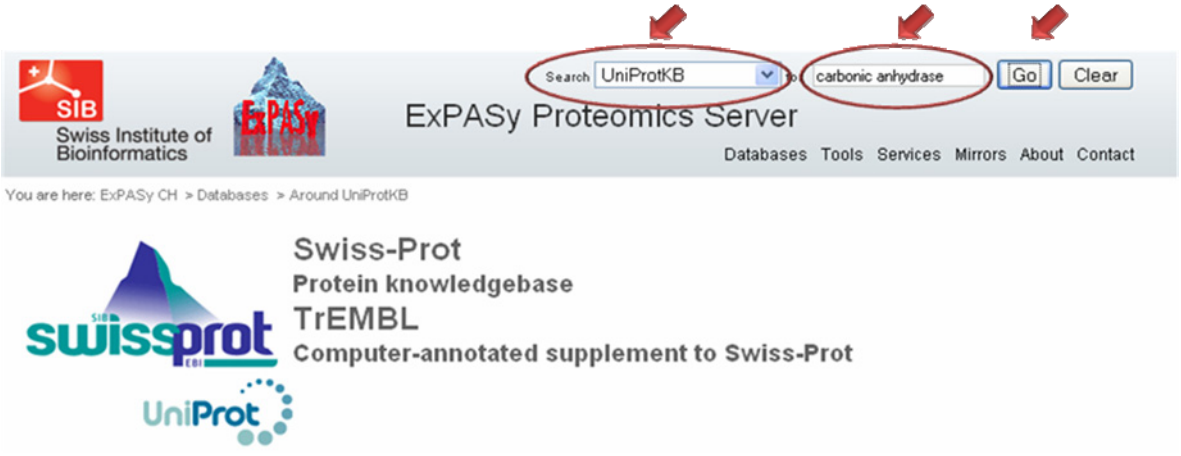
This information is free available to download on http://sing.ei.uvigo.es/DPD

V.9.SM.2 How to obtain the *in-silico* digestion of a target protein

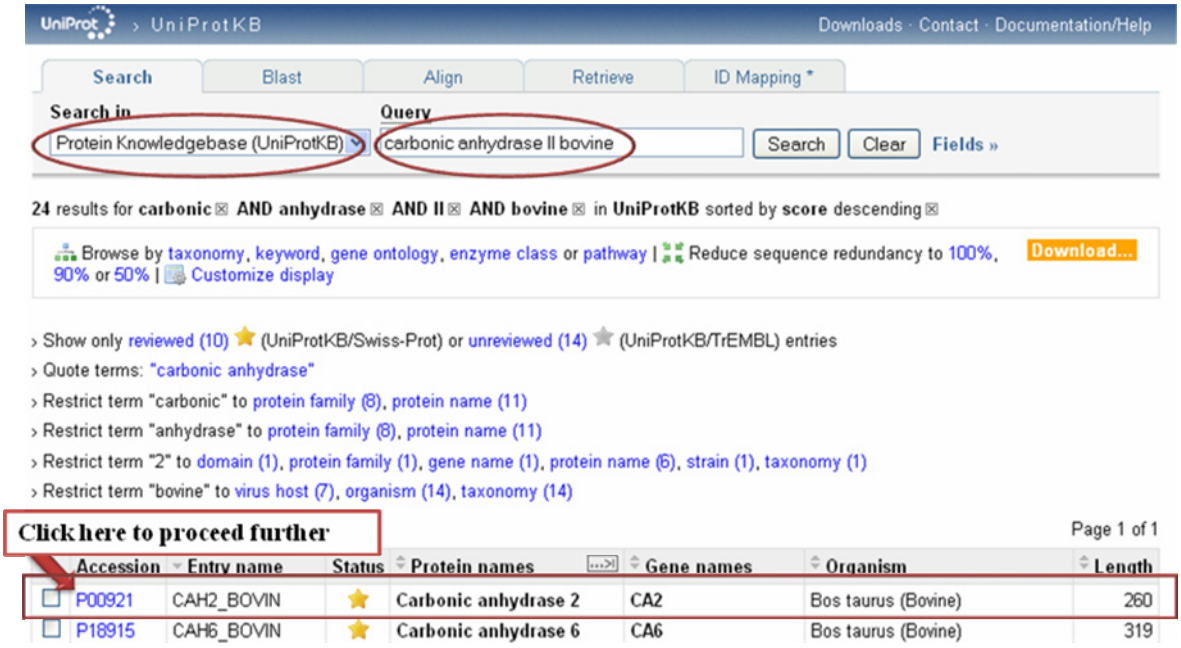
The program you are going to work with has been specially designed for the quantification of target proteins through MALDI-TOF-MS. The following steps given below must be followed to obtain a maximum performance of the program.

The first thing to be done is to obtain the *in-silico* digestion of the targeted protein. To proceed do as follows:

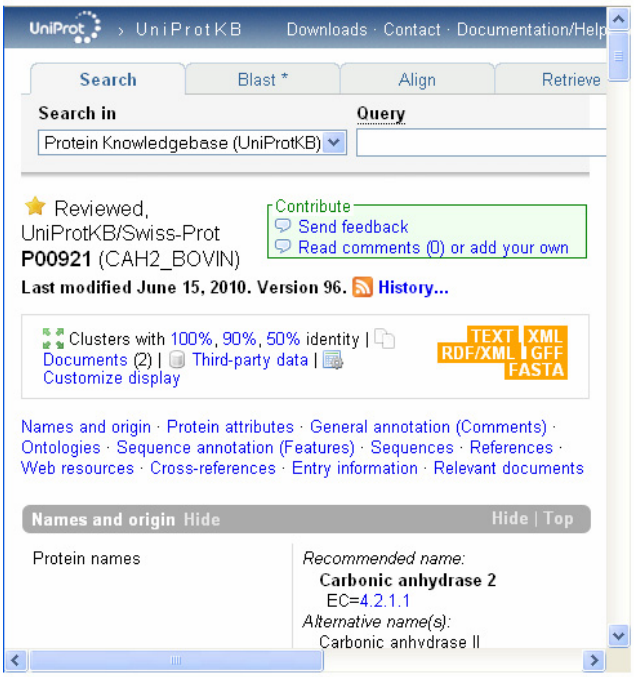
(i) Go to swissprot data base: http://www.expasy.org/sprot/ and introduce the name of your protein (carbonic anhydrase II bovine) and then click go.



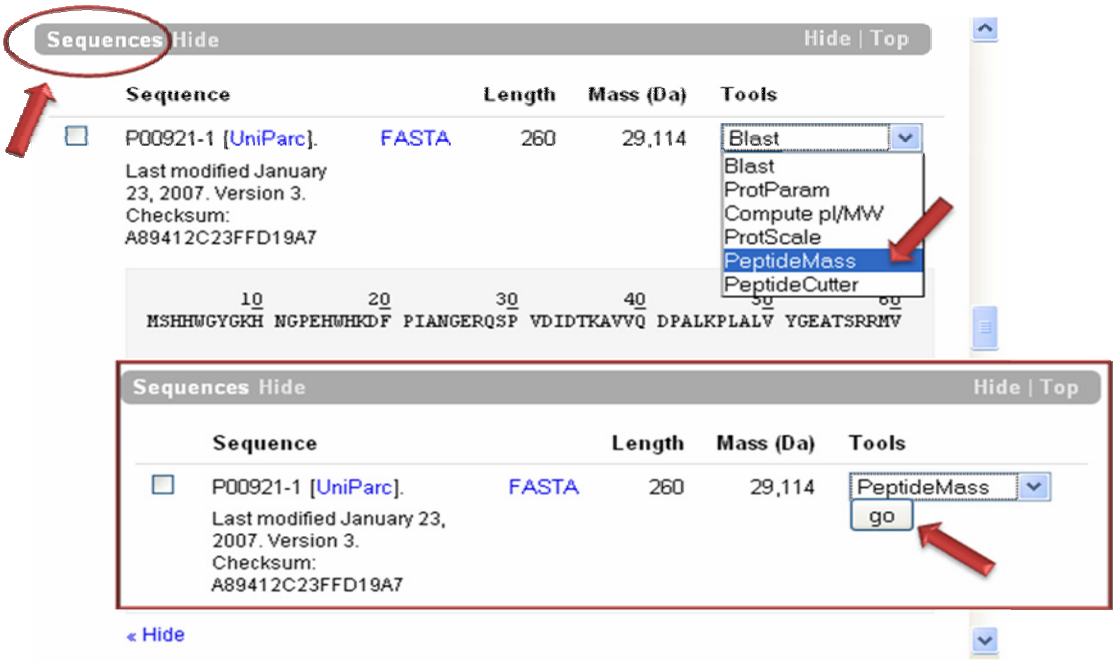
(ii) Then a new interactive menu appears, click on the accession number of the target protein.



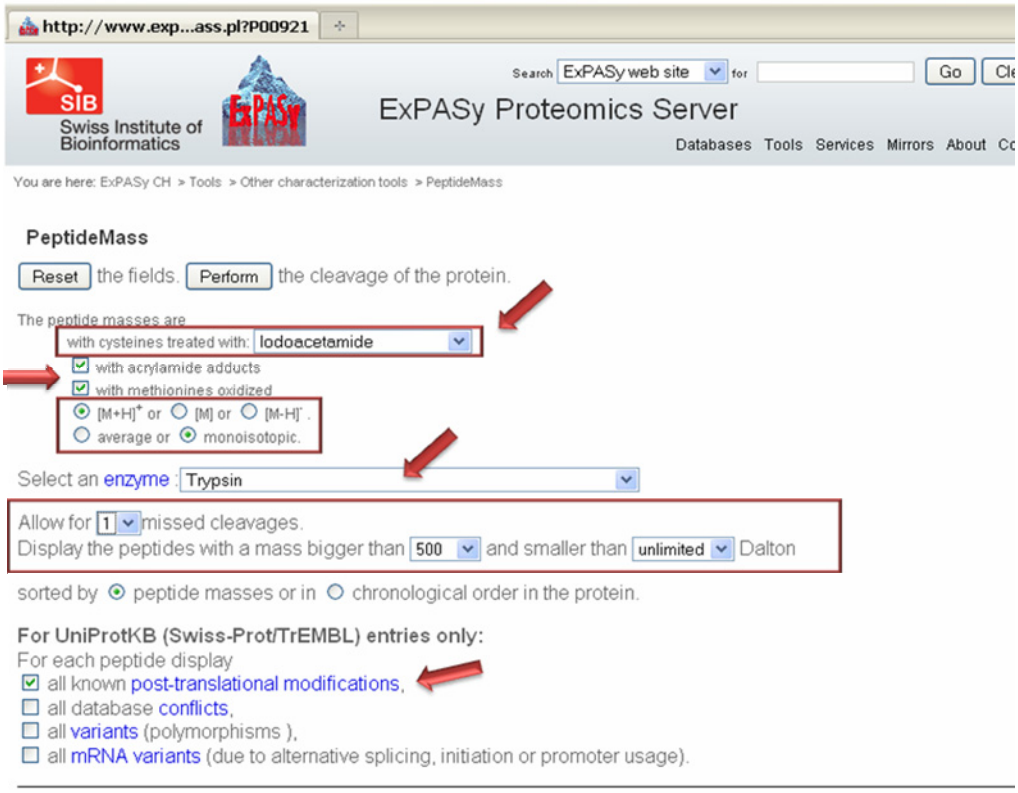
(iii) Once you had clicked in the accession number you will be prompted to the following menu.



(iv) Scroll down, until you reach "sequences", then chose "PeptideMass" in tools and click "go"

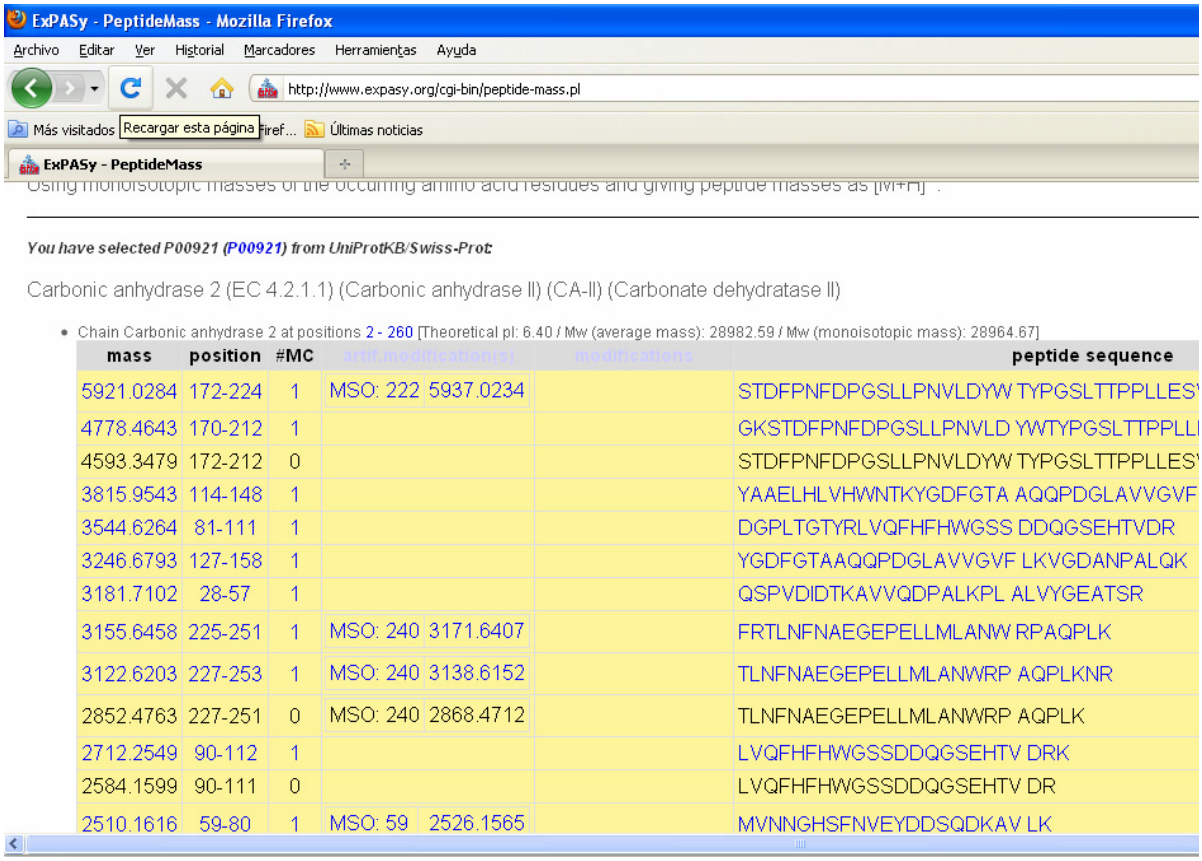


(v) Once you are in the page named "PeptideMass", the following parameters are chosen:



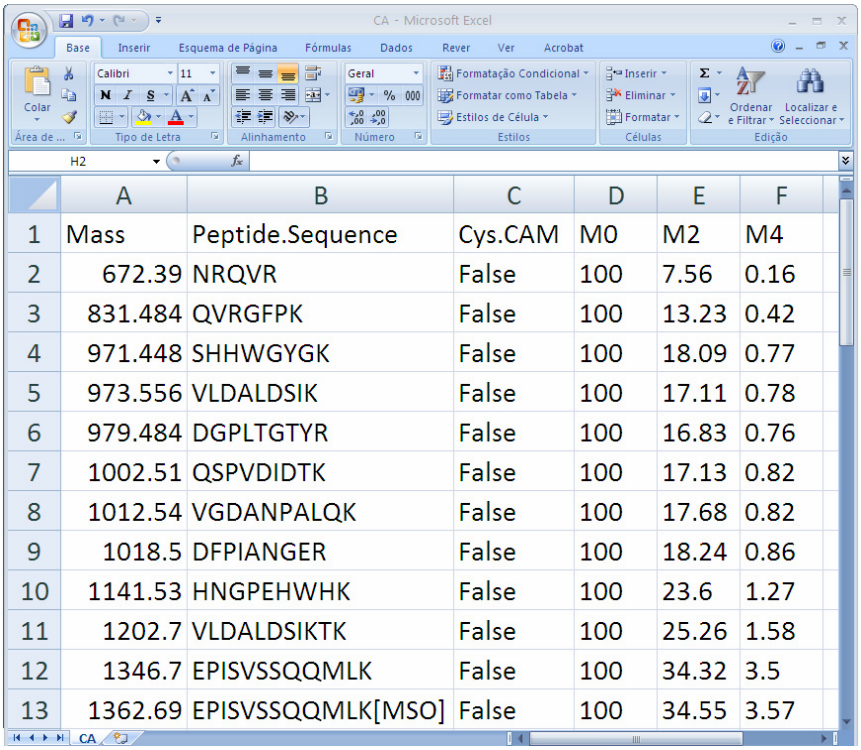
The criteria to select each option, it will depend on the sample preparation used. In this case, protein reduction and alkylation were done so "peptide masses are with cysteines treated with": Iodoacetamide. The presence of acrylamide adducts may result because of protein separation by SDS-PAGE. The methionines can be oxidized as a result of the presence of oxygen in the sample/lab atmosphere. Because we are working with MALDI ionization with positive polarity, the peptide peaks are mainly formed with $[M+H]^+$ forms. We want to obtain the peaks as monoisotopic. The enzyme used to do the protein digestion was trypsin. The number of missed cleavages allowed is 1 and the preferred peptides masses to be obtained are bigger than 500 Da. Finally we also ask for all known translational modifications. Once you have introduced all the parameters of your searching, click in "perform".

The system gives you back the *in-silico* digestion of the protein you are interested in. Copy the information that is needed to create the "*in-silico* file", namely the masses and sequences of all potential peptides formed after trypsin digestion.



mass	position	#MC	artif. modification(s)	modifications	peptide sequence
5921.0284	172-224	1	MSO: 222 5937.0234		STDFPNFDPGSLLPNVLDYW TYPGSLTTPPLLES
4778.4643	170-212	1			GKSTDFPNFDPGSLLPNVLD YWTYPGSLTTPPLL
4593.3479	172-212	0			STDFPNFDPGSLLPNVLDYW TYPGSLTTPPLLES
3815.9543	114-148	1			YAAELHLVHWNTKYGDFGTA AQQPDGLAVVGVF
3544.6264	81-111	1			DGPLTGTYRLVQFHFHWGSS DDQGSEHTVDR
3246.6793	127-158	1			YGDFGTAAQQPDGLAVVGVF LKVGDANPALQK
3181.7102	28-57	1			QSPVDIDTKAVVQDPALKPL ALVYGEATSR
3155.6458	225-251	1	MSO: 240 3171.6407		FRTLNFNAEGEPELLMLANW RPAQPLK
3122.6203	227-253	1	MSO: 240 3138.6152		TLNFNAEGEPELLMLANWRP AQPLKNR
2852.4763	227-251	0	MSO: 240 2868.4712		TLNFNAEGEPELLMLANWRP AQPLK
2712.2549	90-112	1			LVQFHFHWGSSDDQGSEHTV DRK
2584.1599	90-111	0			LVQFHFHWGSSDDQGSEHTV DR
2510.1616	59-80	1	MSO: 59 2526.1565		MVNNGHSFNVEYDDSQDKAV LK

(vi) Now you should be able to prepare you excel CSV file with the columns as follows (this is going to take some time). You will use this file any time you need to quantify the targeted protein.



	A	B	C	D	E	F
1	Mass	Peptide.Sequence	Cys.CAM	M0	M2	M4
2	672.39	NRQVR	False	100	7.56	0.16
3	831.484	QVRGFPK	False	100	13.23	0.42
4	971.448	SHHWGYGK	False	100	18.09	0.77
5	973.556	VLDALDSIK	False	100	17.11	0.78
6	979.484	DGPLTGTYR	False	100	16.83	0.76
7	1002.51	QSPVDIDTK	False	100	17.13	0.82
8	1012.54	VGDANPALQK	False	100	17.68	0.82
9	1018.5	DFPIANGER	False	100	18.24	0.86
10	1141.53	HNGPEHWHK	False	100	23.6	1.27
11	1202.7	VLDALDSIKTK	False	100	25.26	1.58
12	1346.7	EPISVSSQQMLK	False	100	34.32	3.5
13	1362.69	EPISVSSQQMLK[MSO]	False	100	34.55	3.57

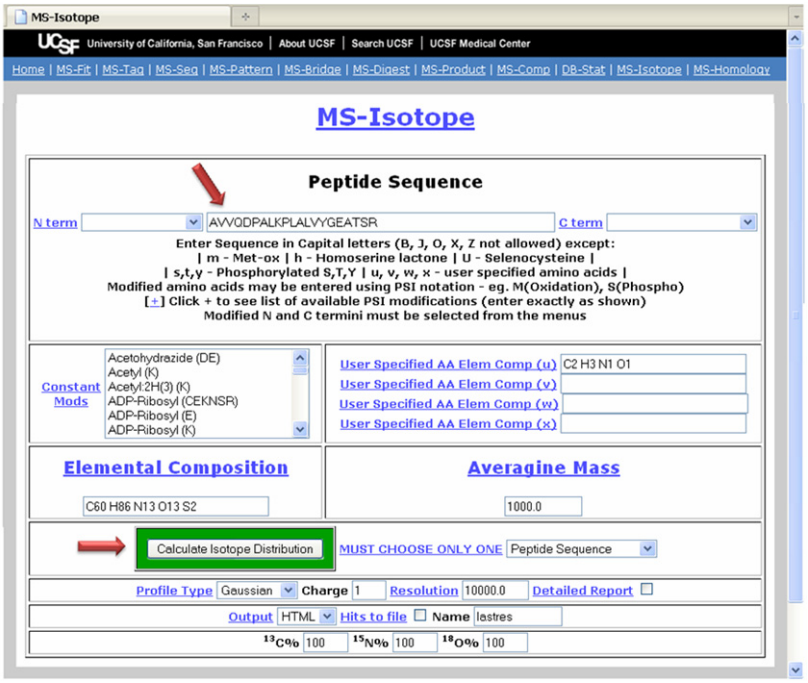
Column A contains all the theoretical m/z values originated from the peptides obtained after a complete digestion of carbonic anhydrase with trypsin (including those with 1 missed cleavage). Column B contains the *in-silico* peptide sequence and Column C contains, as a true or false entry values, the presence (true) or not (False) of carboxyamidomethilations.

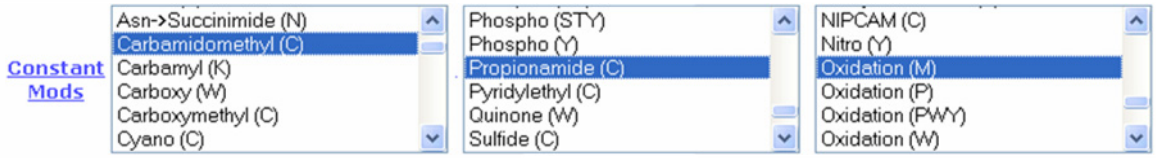
(vii) As you may see, the excel CSV file, with the *in-silico* digestion also includes the predicted intensities of the isotopic distributions M_0, M_2 and M_4 corresponding to each peptide. This information in needed for the algorithms that will calculate the isotope ratios of labeled to non labeled peaks.

V.9.SM.3 How to obtain the predicted isotopic mass distribution of the peptides

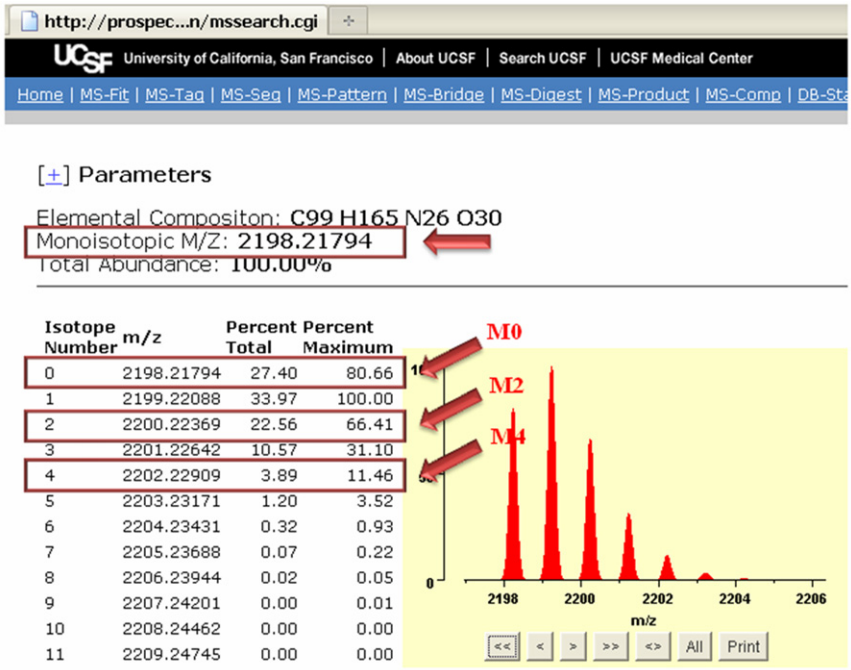
To calculate the predicted isotopic distribution for each peptide sequence, go to: http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msisotope.

For each peptide sequence there are different possibilities: normal peptide; peptide with carbamidomethylations (carbamidomethyl C), peptides with methionines oxidized (oxidation M) and peptides with acrylamide adducts (propionamide C). Do not forget, to indicate in the constant mode of the web page (see below) the presence of modifications.

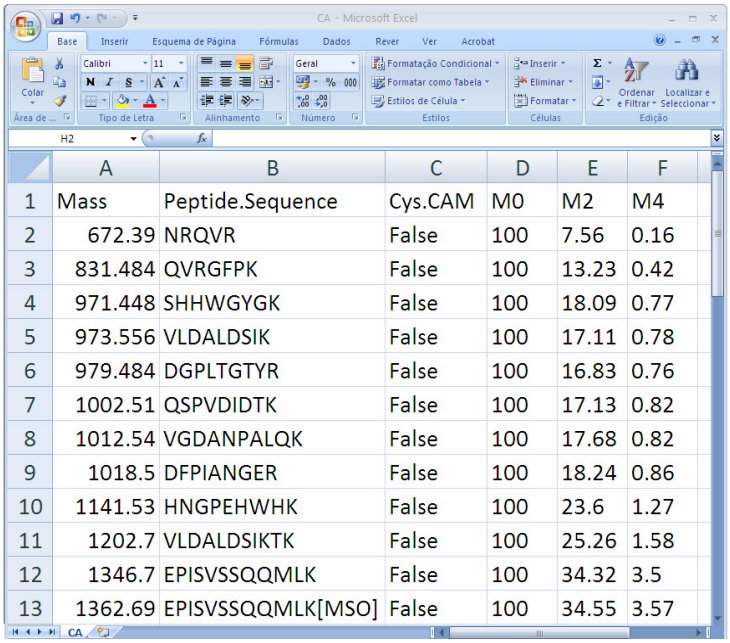




Remember to check that the monoisotopic m/z value is in agreement with the one obtained in the *in-silico* digestion for the target peptide. Copy the Values for M0, M2 and M4 and paste them next to the corresponding peptide in the excel CSV file of the *in-silico* digestion.



The above mentioned process to obtain the M0, M2 and M4 masses must be done for all peptides and for their respective modifications. The final *in-silico* file you should look like the print screen present below.



	A	B	C	D	E	F
1	Mass	Peptide.Sequence	Cys.CAM	M0	M2	M4
2	672.39	NRQVR	False	100	7.56	0.16
3	831.484	QVRGFPK	False	100	13.23	0.42
4	971.448	SHHWGYGK	False	100	18.09	0.77
5	973.556	VLDALDSIK	False	100	17.11	0.78
6	979.484	DGPLTGTYR	False	100	16.83	0.76
7	1002.51	QSPVDIDTK	False	100	17.13	0.82
8	1012.54	VGDANPALQK	False	100	17.68	0.82
9	1018.5	DFPIANGER	False	100	18.24	0.86
10	1141.53	HNGPEHWHK	False	100	23.6	1.27
11	1202.7	VLDALDSIKTK	False	100	25.26	1.58
12	1346.7	EPISVSSQQMLK	False	100	34.32	3.5
13	1362.69	EPISVSSQQMLK[MSO]	False	100	34.55	3.57

Chapter VI.

A novel ^{18}O inverse labeling-based workflow for accurate bottom-up mass spectrometry quantification of proteins separated by gel electrophoresis

---

Published in:

Electrophoresis, 31 (2010) 3407-3419

Hugo M. Santos, Daniel Glez-Peña, Miguel Reboiro-Jato, Florentino Fdez-Riverola, Mário S. Diniz, Carlos Lodeiro, J. L. Capelo.

Index

VI.1 Resumo

No presente trabalho apresentamos um novo protocolo para quantificação de proteínas separadas por eletroforese em gel e análise por MALDI-TOF-MS. O protocolo proposto supera os inconvenientes da digestão em gel e analise por MALDI-TOF-MS, mantendo os seus benefícios. Baseando-se na separação de proteínas por eletroforese em gel, digestão acelerada de proteínas, marcação com ^{18}O, quantificação com os péptidos previamente seleccionados através de marcação directa e inversa e por último como mais valia apresenta um software dedicado à escolha dos péptidos que vão guiar a quantificação de proteínas de modo automático

As seguintes proteínas foram quantificadas de forma precisa, albumina de plasma bovino, ovalbumina, anidrase carbónica, inibidor de tripsina e α-lactalbumina. Como aplicação da metodologia proposta procedeu-se à quantificação de vitelogenina de *Ciprinus carpio* após exposição a níveis elevados de estrogénio. O protocolo proposto foi validado por comparação com ELISA, ambos apresentaram resultados comparáveis (teste não paramétrico *Mann-whitney U*).

Palavras-chave: Quantificação de proteínas, software DPD, exactidão, vitelogenina, eletroforese em gel, biomarcador.

A minha contribuição para este trabalho consistiu na elaboração do design experimental, execução experimental da separação de proteínas por eletroforese em gel, digestão gel, análise de MALDI-TOF-MS, processamento de dados, interpretação e desenvolvimento/teste do software.

VI.2 Abstract

In the present work we report on a novel and fast protocol for accurate bottom-up protein quantification that overcomes the drawbacks of in-gel digestion and MALDI analysis, whilst maintaining their benefits. It relies on the following steps: (i) gel electrophoresis separation of proteins, (ii) fast in-gel protein digestion with trypsin, (iii) ^{18}O-labeling through the decoupled method, (iv) quantification through selected peptides previously chosen using the ^{18}O-inverse labeling approach and that, finally, (v) it takes advantage of software specifically developed to select the peptides that will drive the quantification of the protein in an automated mode. We have accurately quantified the following six proteins: glycogen phosphorylase, bovine serum albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor, and α-lactalbumin. As a case study we have quantified the protein vitellogenin in plasma of *Ciprinus carpio* exposed to high levels of estrogens. The proposed new protocol was validated against the traditional ELISA method; both were found to provide comparable results (non-parametric Mann-Whitney U-test).

Keywords: Protein quantification, DPD software, accuracy, Vitellogenin, gel electrophoresis, biomarker.

My contribution to this work was the elaboration of the experimental design, experimental execution of gel electrophoresis, in-gel digestion, MALDI-TOF-MS analysis, data processing, interpretation and software development/testing.

VI.3 Introduction

Since the sequencing of the human genome, quantification of proteins through mass spectrometry, MS,-based approaches has become an important research topic in chemistry, biology, medicine and even chemical engineering [1-7]. The sample treatments currently used for protein quantification can be classified into those relying on the isotopic labeling of proteins and peptides or others termed as label-free approaches [4,8]. Concerning labeling strategies, different methods can be found in the literature. The ^{18}O-labeling for quantitative proteomics is a method that has refocused the attention of the scientific community over recent years [9-12]. Some of the benefits claimed for ^{18}O-labeling in protein quantification are the following: (i) it is a universal tag, since (theoretically) all peptides present in a sample are labeled; (ii) post-translational modifications remain unaltered; (iii) reagents involved in the labeling are not expensive; (iv) the marked peptides and their non-marked pairs are eluted at the same time in HPLC. However some drawbacks are reported: (i) the method can compare only pairs of samples, e.g. normal state versus disease state; (ii) the labeling cannot be done at the protein level; (iii) the degree of labeling depends on the reagents present in the sample and on the type on labeling done and (iv) back exchange with ^{16}O may occur when labeled peptides are mixed with their unlabeled counterparts [13,14].

Recently, the use of an ^{18}O-labeled reference sample in conjunction with high resolution liquid chromatography - MS has been proposed as a universal internal standard for quantitatively analyzing proteins in large sample sets [15]. Protein quantification with ^{18}O is regularly done through MS with either ESI (Electrospay Ionization) or MALDI (Matrix Assisted Laser Desorption Ionization) as the ionization methods [16,17]. One method often employed uses gel electrophoresis as protein fractionation step. After in-gel protein digestion the obtained peptides are labeled and used for protein identification and quantification It has been claimed in the literature that in-gel digestion of proteins compromises the accuracy of quantification by affecting the recovery of individual peptides and, therefore, protein estimates might be strongly influenced by the selection of the peptides used in the quantification process [18]. In a typical LC-ESI-MS/MS quantification experiment peptides are detected, selected and fragmented in real time. In the most common instruments MS and MS/MS scan times are selected at the beginning of a run as a compromise between peak intensity and the capability to analyze as many peptides as

possible in a chromatographic window [19,20]. In an opposite way, MALDI data acquisition of MS and MS/MS data is decoupled from the separation steps (2D-gel electrophoresis or peptide separation by Liquid Chromatography - LC). Typically, after a MS scan, the precursor ions are selected for MS/MS based on signal intensity, signal-to-noise ratio or resolution, minimizing the possibility to collect redundant data [19,20]. However manual examination of MS spectra is recommended to improve the quantification accuracy [21]. Whilst literature reports that both MALDI-MS and ESI-MS produce consistent results, it has been also reported that ESI produce more accurate results and allow relatively straightforward quantification [18]. Despite some drawbacks of MALDI ionization, related to sample/matrix deposition and crystallization [22], preferent ionization of R-terminated peptides [23] and arginine-containing peptides [24], more recently it has been reported that LC-MALDI-MS/MS is quite accurate for quantitative proteomics using iTRAQ labeling in addition to be a complementary strategy to LC-ESI-MS/MS. [20]

In the present work we report on a novel workflow for protein quantification that overcomes the drawbacks of in-gel digestion and MALDI analysis, whilst maintains their benefits. Our new method relies on the following steps: (i) SDS-PAGE separation of proteins, (ii) ultrasonic in-gel tryptic digestion, (iii) ^{18}O-labeling through the decoupled method, (iv) quantification through selected peptides previously chosen using the ^{18}O inverse labeling approach [25] and that, finally, (v) it takes advantage of software specifically developed to select the peptides that will drive the quantification of the protein in an automated mode. The proposed quantification method is presented here in a case-study. For validation purposes a non-MS-based approach, ELISA, was also used.

VI.4 Materials and Methods

VI.4.1 Apparatus

Gel Electrophoresis was performed with an electrophoresis system, model Mini-PROTEAN Tetra Cell, from Biorad (Hercules, CA, USA), following the manufacturer's instructions. Protein digestion and labeling were carried out in safe-lock tubes of 0.5 ml from Eppendorf (Hamburg, Germany). A minicentrifuge, model Spectrafuge-mini, from Labnet (Madrid, Spain), and a minicentrifuge-vortex, model Sky Line, from ELMI (Riga, Latvia) were used throughout the sample treatment, when necessary. A vacuum

concentrator centrifuge from UniEquip (Martinsried, Germany) model UNIVAPO100H with a refrigerated aspirator vacuum pump model Unijet II was used for (i) sample drying and (ii) sample pre-concentration. Milli-Q natural abundance ($H_2{}^{16}O$) water was obtained from a SimplicityTM from Millipore (Milan, Italy).

An ultrasonic bath, model Transsonic TI-H-5, from Elma (Singen, Germany) with control of temperature and amplitude was used to accelerate the gel washing, the protein reduction and the protein alkylation steps, and a sonoreactor model UTR200, from Dr. Hielscher (Teltow, Germany), was used to accelerate the enzymatic digestion step. All materials were used without further purification. α-Cyano-4- hydroxycinnamic acid, α-CHCA, *puriss* for MALDI-TOF-MS from Fluka (Buchs, Switzerland) was used as MALDI matrix. ProteoMass Peptide MALDI-MS Calibration Kit (MSCAL2) from Sigma was used as mass calibration standard for MALDI-TOF-MS.

VI.4.2 Standards and reagents

Standard protein mixtures of glycogen phosphorylase 97kDa; bovine serum albumin, 66kDa; ovalbumin, 45kDa; carbonic anhydrase, 30kDa; trypsin inhibitor, 20.1kDa; and α-lactalbumin, 14.4kDa; were purchased from Amersham Biosciences (Piscataway, USA, part number 17-0446-01). α-lactalbumin from bovine milk (≥85%) and trypsin, sequencing grade, were purchased from Sigma (Steinheim, Germany). Reduction and alkylation were carried out, respectively, with D,L-dithiothreitol (DTT, 99%) and iodoacetamide (IAA) from Sigma. The following reagents were used during sample digestion: ammonium bicarbonate buffer (AmBic, Ph 8.5, ≥99.5%) and formic acid (FA, ~98%) from Fluka (Buchs, Switzerland); and labeling: ammonium acetate (>99.0%) from Fluka, calcium chloride (~97%, anhydrous) from Sigma, Mag-Trypsin from Clontech (USA) and $H_2{}^{18}O$ (97 atom%) from Isotec (Miamisburg, USA). Trifluoroacetic acid (TFA, 99%) was obtained from Riedel-de Haäen (Seelze, Germany).

VI.4.3 Gel electrophoresis

Amounts of protein ranging from 0.8 to 3.6μg were dissolved in 5μL of water plus 5μL of sample buffer (5mL of Tris-Base 0.5M, 8mL of SDS 10% w/v, 1mL of β-mercaptoethanol, 2mL of glycerol, and 4mg of bromophenol blue in a final volume of 20mL in water) and then boiled for 5 min to denature the proteins for sodium dodecyl

sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The denatured proteins were loaded in 10% polyacrylamide gels with 0.75mm thickness. Proteins were separated at 120V and 400mA for 65 min.

VI.4.4 Stain and image analysis

The stain solution was prepared dissolving 1g of Coomassie blue R-250 in 200mL of a mixture of 45% (v/v) methanol, 7.5% (v/v) glacial acetic acid and 47.5% (v/v) water. After gel electrophoresis, gels were rinsed with Milli-Q water and then incubated for 1 h in the staining solution at room temperature. After incubation, the staining solution was removed, and the gel was rinsed with Milli-Q water and then incubated for 1.5 h at room temperature in the destaining solution (45% v/v methanol, 7.5% v/v glacial acetic acid, and 47.5% v/v water). Gel imaging was carried out with a ProPicII-robot (Digilab-Genomic Solutions, USA), and when necessary, the robot was also used to excise the spots containing the proteins.

VI.4.5 In-gel protein digestion

Ultrasonic in-gel enzymatic digestion was done according to the ultrafast proteolytic digestion protocol previously developed in our laboratory [26-28]. Protein bands were manually excised from the gel and placed in safe-lock tubes of 0.5mL. Gel pieces were washed, first with AmBic 25mM/acetonitrile (100µL) and then with acetonitrile (100µL), in an ultrasonic bath operating at 35 kHz (100% amplitude) for 5 min for each step. Afterwards, the gel pieces were dried in a vacuum concentrator centrifuge for 5 min.

Protein reduction and alkylation steps were included in the protocol to facilitate the enzymatic action and to increase the protein sequence coverage (%). To do so, disulfide bonds from cysteine residues were reduced with DTT in an ultrasonic bath operating at 35kHz (100% amplitude) for 5min at room temperature, and then, the reduced cysteines were alkylated with IAA in an ultrasonic bath operating at 35kHz (100% amplitude) for 5min at room temperature.

After the reduction and alkylation steps, the gel was submitted again to the washing procedure in the same way as described above, followed by another drying step of 10min. Afterwards, the dried gel pieces were incubated with trypsin (375ng in 25µL) on ice for 60min to rehydrate the gel and to allow enzyme penetration into it. Subsequently, in-gel

protein digestion was performed in a sonoreactor operating at 50% amplitude for 4min. Next, trypsin activity was stopped by the addition of 20µL of formic acid (5% (v/v)).

VI.4.6 ^{18}O labeling: the decoupled procedure

For the ^{18}O-labeling, the digested peptides were reconstituted with 10µL of 25mM calcium chloride and 10µL of acetonitrile 20% v/v + 50mM ammonium acetate pH 6.75. Then the samples were vacuum re-dried, and after evaporation the dried samples were reconstituted in 5µL of natural abundance water or 97% ^{18}O-enriched water and 5µL of a 5% suspension of Mag-Trypsin in $H_2{}^{16}O$ or $H_2{}^{18}O$ were added. The digested peptides were labeled during 15min of vortexing and centrifugation, and finally trypsin was removed by magnetic separation. A detailed explanation of this procedure can be found elsewhere [28].

VI.4.7 Inverse ^{18}O labeling of peptides

Proteins were separated by 1D-PAGE and then submitted to the protocols described in VI.4.5 and VI.4.6. The inverse ^{18}O labeling protocol as described by Wang et. al [25] was then used.

VI.4.8 Analytical and biological sample replicates

Eight commercial vials containing 6 standard proteins were used and two gels were run. For each gel, eight lines were run (one line for vial). For each protein a total of sixteen bands were obtained that were excised and treated. Eight bands were labeled and mixed with their non-labeled counterparts, giving a total of eight mixtures of label to non label proteins, which were grouped in two sets of four direct and four inverse samples as depicted in figure VI.1. Each biological replicate corresponds to nine different *Cyprinus carpio* specimens as follows: 3 control males named as Control male I, II and III; 6 exposed males named as Exposed male I, II, III, IV, V and VI. Plasma collected from each specimen was prepared in triplicate.

VI.4.9 Quantification of peptides

Quantification of peptides through ^{18}O was done with the mathematical algorithm for deconvolution described by Yao et al. [29]. Reduction of the spectra to a centroided plot was done using the centroiding option function of the Data ExplorerTM software (version 4.0) from Applied Biosystems. This function is an advanced peak filtering method that improves the quality of mass spectral data and reduce data file size. Profile data, in which many points are used to delineate a mass spectral peak, is converted into mass-centroided data by a data compression algorithm. The centroided mass peak is located at the weighted center of mass of the profile peak. The normalized area of the peak provides the mass intensity data.

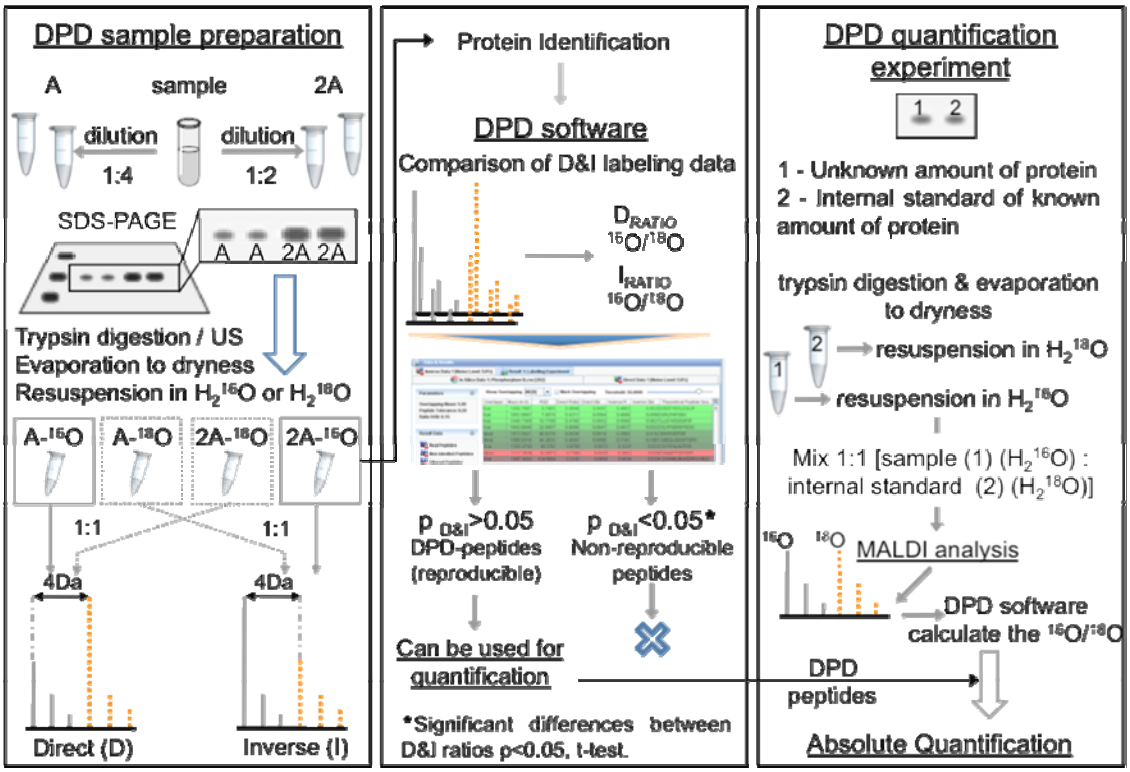


Figure VI. 1. Schematic diagram illustrating steps of the sample treatment workflow including gel electrophoresis, labeling, protein identification, DPD analysis of the D&I data that allows the best choice of peptides for quantification and finally the DPD quantification experiment.

VI.4.10 Case study

VI.4.10.1 Plasmatic vitellogenin from Cyprinus carpio

Vitellogenin (Vtg) is an egg yolk phospholipo-glycoprotein that is secreted by the liver of oviparous vertebrates. It is usually restricted to females but it can be induced in males and juveniles by exposure to estrogens. Vtg is a large (300 to 600kDa native or 160 to 200kDa subunit) serum phospholipo-glycoprotein that serves as the major precursor to the egg-yolk proteins of oviparous vertebrates [30].

Carp (*Cyprinus carpio)* males were placed indivudually in 15L polyestyrene tanks supplied with de-chlorinated tap water (temperature: 15±2ºC; pH 7.4±0.2; dissolved oxygen >6mgL^{-1}). Vtg synthesis was induced by a single intraperitoneal injection of 17β-estradiol (Sigma, USA),dissolved in PBS (1% absolute ethanol) at a concentration of 5mg/Kg body weight (bw; positive control), and vehicle (negative control), respectively. After 48hours, blood was taken from the caudal vein of each treated fish using a heparinized syringe, treated with 4TIU/mL of Aprotinin (Sigma, USA), and centrifuged immediately for 15min at 4,000xg (4ºC) to obtain the plasma. Samples were stored at –80ºC until further analysis of Vtg. 80µL of plasma from *Cyprinus carpio* was diluted to 100µL with cold PBS (Phosphate Buffer Solution). 300µL of -20ºC cold acetone was added to the diluted plasma solution and kept overnight on ice. The sample was centrifuged at 10000g, 4ºC for 30min. The supernatant was removed and the pellet was suspended in 100µL of buffer (10mM Tris-HCl pH 7.4; 2% of SDS; 1% of β-mercaptoethanol).

Amounts of precipitated plasma (5µL) were mixed with 5µL of sample buffer (5mL of 0.5M Tris-Base + 8mL of 10% SDS + 1mL of β-mercaptoethanol + 2mL of glycerol + 4mg of bromophenol blue in a final volume of 20mL in water) for analysis by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (7.5% 0.75mm thickness). After gel electrophoresis (65min, 120V, 400mA), the gel was stained with Coomassie blue R-250 and destained in order to visualize the proteins bands. Vitellogenin was in-gel digested according to the accelerated method described in VI.4.5.

VI.4.10.2 ELISA vitellogenin quantification in Cyprinus carpio fish plasma

The synthesis of vitellogenin (Vtg) was confirmed and quantified by direct enzyme linked immunosorbent assay (ELISA) as described in Diniz et al. [31]. Briefly, following dilution, at least 1:200 for males and 1:1000 for females, the samples were pipetted (50µL) in triplicate into a Maxisorp F96 microtiter plate (Nunc-Roskilde, Denmark) and incubated overnight at 4ºC. After that, the microplate was washed with PBS-Tween and 280µl of blocking buffer (1% BSA in PBS with 0.02% sodium azide) was added to microplate wells to block free binding sites. The microplate was incubated overnight (4ºC) and then re-washed. The microplate was coated (50µl) with a monoclonal antibody against carp Vtg (Biosense, Bergen, Norway), diluted to an appropriate concentration (0.1-5µg/mL), and then incubated overnight at 4ºC. After plate washing, 100µL of a secondary antibody (goat-antimouse immunoglobulin-IgG conjugated to alkaline phosphatase, Sigma, USA), diluted in a blocking buffer (1:1000), was added to each well and incubated for a further 60 min. at 38ºC.

After a final washing step, 100µL of the substrate (p-nitro-phenylphosphate – PNPP, SIGMA-Aldrich) was added to each microplate well and incubated at room temperature (10-30min) in the dark. A standard carp Vtg (Biosense, Norway) was diluted to give a range from 15 to 1000ngmL^{-1} and a calibration curve was constructed allowing the quantification of Vtg in plasma samples. Then the enzyme reaction was stopped by adding 50µL of stop solution (3N NaOH). The plates were read in a 96-well microtiter plate reader (BioRad-Benchmark, USA) at 405nm.

VI.4.11 MALDI-TOF-MS analysis

Prior to MALDI-TOF-MS analysis, the sample was mixed with the matrix solution. α-CHCA matrix was used throughout this work and was prepared as follows: 10mg of α-CHCA was dissolved in 1mL of Milli-Q water/acetonitrile/TFA (1mL/1mL/2µL). Then, 4µL of the aforementioned matrix solution was mixed with 4µL of sample and the mixture was shaken in a vortex for 30s. One microlitre of each sample was hand spotted in a well of a MALDI-TOF-MS sample plate and allowed to dry.

A MALDI-TOF-MS system model Voyager DE-PRO Biospectrometry Workstation equipped with a nitrogen laser (λ= 337 nm) from Applied Biosystems (Foster City, CA) was used to acquire the PMFs. Measurements were done in reflector positive ion mode, with a 20kV accelerating voltage, 75.1% grid voltage, 0.002% guide wire and a delay time of 100ns. Two close external calibrations were performed with the monoisotopic peaks of the bradykinin fragment 1-7, angiotensin II, $P_{14}R$, and ACTH peptide fragments (*m*/*z*: 757.3997, 1046.5423, 1533.8582, and 2465.1989, respectively). Monoisotopic peaks were manually selected from each of the spectra obtained. Mass spectral analysis for each sample was based on the average of 500 laser shots.

VI.4.12 Software distribution

Current versions of the software and their supporting user manuals are freely available for downloading and use with Windows, Mac or Linux software, without restriction, via internet at http://sing.ei.uvigo.es/DPD. This program operates on MS Excel CSV files with centroid mass and relative intensity data extracted from the Data ExplorerTM. The program was developed based on previous work related to cancer diagnosis [32] and on suggestions given by the Bioscope group.

VI.5 Results and Discussion

VI.5.1 Sample treatment workflow

A schematic diagram illustrating the sequential steps of the sample treatment workflow is presented in figure VI.1. Proteins were separated through 1D gel electrophoresis. Samples were loaded into the gel at different concentrations covering a range inclusive of the commonly accepted thresholds for defining biological significance and they were 1:2, 2:1 and 1:5 [33]. After ultrasonic protein digestion [26,27] the pool of peptides was dried and reconstituted in normal water or in 97% ^{18}O water[13,34]. Labeling of peptides can be done faster by doing the so called proteolytic labeling [35]. However, work previously developed in our laboratory has shown that the post-proteolysis labeling of peptides performs better since both the ratio of total peptides labeled (single+doubled) and the ratio of peptides double labeled are higher [28,35]. Finally the labeled samples are mixed as described in figure VI.1 to perform the so-called inverse labeling [25]. With this

procedure two converse labeling experiments are performed in parallel as follows. In the "direct" labeling, the sample is reconstituted in normal water whilst its counterpart of higher amount (i.e. 1:2, 1:5) is reconstituted in ^{18}O-water. In the "inverse" method the labeling is done conversely. Finally, an equal sample volume of non-labeled and its labeled counterpart are mixed and analysed through MALDI-TOF-MS. The ratios of the relative intensities of the ^{16}O peaks and the ^{18}O peaks are used then in the final step of the workflow, as it is showed in figureVI.1. In this step, the software DPD (Decision Peptide Driven) compares the labeled to unlabeled ratios of the same peptides obtained in the "direct" and "inverse" (D&I) methods. Only those peptides having the D&I ratios within a given RDS and *p* significance level (*t*-test) are used for protein quantification.

VI.5.2 Finding out the best peptides for protein quantification

Protein quantification which relies on the in-gel digestion of proteins presents several pitfalls. Thus, it has been demonstrated that the recovery of in-gel digestion was rather low when compared to in-solution digestion and enabled no accurate absolute or relative estimates of proteins [18]. Furthermore, the observed recoveries of individual peptides vary strongly within a set of in-gel digests, leading to a high RSD of the measured amounts of proteins. In addition to these drawbacks, it has been stressed in the literature that biased losses of peptides might occur during the post-digestion sample processing of in-gel digests. [18,36,37]. Despite these problems and based on our own work and that from other laboratories we hypothesised that, for each protein, it might be possible to identify a certain number of peptides that might have low and paralleled loses through a typical proteomic workflow, and that therefore, using these, a robust and accurate protein quantification might result.

Since the goal of our approach was a peptide differential analysis and to extract and to identify the number of peptides remaining constant in expression level through a typical in-gel digestion workflow, any method that enables analysis of peptide signals within replicates would be of great value.

The approach that best fits the goal described above is the method proposed by Wang et al. and called "inverse labeling" [25]. With this procedure it is easily detected if a peptide is randomly lost, or the observed recoveries of individual peptides vary strongly within a set of in-gel digests. In addition, further information is also provided with this

approach as follows. One underlying assumption of ^{18}O labeling is that the digestion efficiency remains constant regardless of the amount of protein loaded in the gel. Otherwise, the amount of protein derived from different samples would be falsified.

Comparing the media value of the "direct" and "inverse" (D&I) experiments this methodology can be used to unambiguously verify the recovery of peptides obtained during in-gel protein digestion at different concentrations, and thus it can clearly illustrate which peptides can be used for quantification through a given dynamic range of differential quantification.

To facilitate the comparison of the D&I labeling, a dedicated software, named Driven Peptide Decision, DPD, was developed as a computer tool to extract and to identify the peptides which remain constant in expression level throughout a typical in-gel digestion work-flow. It allows rapid identification of target marked peptides and unambiguous identification of peptides that exhibit the characteristic inverse labeling pattern of a 4Da mass shift. The DPD software analyzes the experimental data to identify the peaks that match with the *in-silico* (theoretical) digestion of the protein within a given m/z tolerance and intensity threshold. The program also uses an algorithm to find out the peaks corresponding to unlabeled peptides, which are consequently discharged. Another algorithm also rejects overlapping peptides resulting from the 4Da gained in the labeling process.

In the following example we demonstrate how our sample treatment and the associated DPD software work together. Glycogen phosphorylase was submitted to the workflow depicted in figure VI.1, and the obtained MS-data was used as input to the DPD program. The complete list of peaks and their intensities obtained with the MALDI for this set of experiments is given in the web page http://sing.ei.uvigo.es/DPD (supplementary material VI.9.SM.1, n=4, files are named as "inverse" and "direct").

The list of peaks (a total of 13) belonging to peptides that might be used for quantification by the DPD program, are shown in table VI.1. Since the ratio used when the protein was loaded in the gel was 1:2 for the "direct" experiment and 2:1 for the "inverse" experiment, the expected ratios should be 0.5 and 2 respectively. For simplicity, the "inverse" ratio is given in table VI.1 as $(\text{inverse ratio})^{-1}$. In this way both ratios are expected to be the same, 0.5, and therefore the comparison between the D&I methods is more easily carried out. We have defined a, $RSD_{D\&I}$, that expresses (in %) the difference between the values obtained for the D&I ratios.

As can be seen in Table VI.1, the relative standard deviation of the ”direct” or “inverse”, $RSD_{D\&I}$, ratios can be used to track the robustness of the sample treatment for each peptide.

Table VI. 1. Peptides assigned by the DPD software as candidates to be used for quantification of glycogen phosphorylase, muscle from *Oryctolagus cuniculus*. Direct labeling: 0.84µg of unlabeled protein and 1.68µg of ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (0.84/1.68). Inverse labeling: 0.84µg of ^{18}O-labeled protein and 1.68µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected inverse ratio is 2 corresponding to the unlabeled/labeled ratio (1.68/0.84). (note: the inverse ratio is expressed in the table as (inverse ratio)$^{-1}$ for simplicity. In this way the direct and inverse ratios may be compared easily since both have the same value of 0.5). For the direct method the amount of protein calculated was the lower of 0.84µg whilst for the inverse method was the higher of 1.68µg. DPD software parameters: Peak intensity - 5; Peptide mass tolerance ± 0.25Da; experimental overlapping - 2; Ratio I_4/I_0 - 0.15. ⋆Significant differences between D&I ratios (p<0.05, t-test). n=4; ‡ Expected values.

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (0.84)‡	Inverse - µg of protein (1.68)‡
1550.77	0.2	0.48±0.05	0.49±0.05	9.71E-01	IGEEYISDLDQLR	0-15%	0.79±0.06	1.7±0.2
1053.57	7.5	0.42±0.06	0.47±0.06	2.93E-01	QRLPAPDEK			
1689.89	7.5	0.50±0.04	0.45±0.03	8.54E-02	ARPEFTLPVHFYGR			
1440.74	10.8	0.48±0.07	0.56±0.06	1.32E-01	LLSYVDDEAFIR			
1678.86	32.1	0.60±0.02	0.38±0.02	4.07E-06	IGEEYISDLDQLRK	30-45%⋆	0.9±0.2	1.9±0.6
1886.9	32.5	0.44±0.02	0.71±0.02	2.01E-06	GYNAQEYYDRIPELR			
1072.54	40.6	0.64±0.04	0.36±0.02	2.55E-04	EIWGVEPSR			
1580.83	44.4	0.40±0.07	0.8±0.2	2.84E-02	QIIEQLSSGFFSPK	45-50%⋆	0.9±0.4	1.7±0.7
1357.77	45.9	0.34±0.02	0.67±0.03	1.30E-05	GLAGVENVTELKK			
1874.91	48.2	0.8±0.1	0.41±0.02	7.91E-03	TCAYTNHTVLPEALER[Cys CAM]			
1355.67	49.3	0.67±0.07	0.32±0.02	9.57E-04	DYYFALAHTVR			
1117.56	62.9	0.74±0.04	0.28±0.02	2.01E-05	VAAAFPGDVDR	>60%⋆	0.8±0.6	2±2
1687.8	124.6	0.22±0.05	3.4±0.6	1.91E-03	DIVNMLMHHDRFK[MSO]			

Thus the peptides IGEEYISDLDQLR, QRLPAPDEK, ARPEFTLPVHFYGR, and LLSYVDDEAFIR are recovered with a reproducibility of within a relative standard deviation, $RSD_{D\&I}$, of less that 15% (n=4) and no statistical differences were found between both D&I experiments (p>0.05, t-test.). Those results indicate that these peptides have losses and average yields of digestion that are paralleled through different sets of in-gel digestions and that the yields of digestion are the same within ratios of 1:2 or 2:1. Peptides IGEEYISDLDQLRK, GYNAQEYYDRIPELR and EIWGVEPSR are recovered with a

$RSD_{D\&I}$ of between 30% and 45% and significant differences were found from the D&I labeling experiments ($p_{D\&I}$<0.05), whilst peptides QIIEQLSSGFFSPK, GLAGVENVTELKK, TCAYTNHTVLPEALER, DYYFALAHTVR are recovered with a $RSD_{D\&I}$ of between 45% and 50% and significant differences were also found from the D&I labeling experiments ($p_{D\&I}$<0.05). Finally, peptides VAAAFPGDVDR and DIVNMLMHHDRFK are recovered with a $RSD_{D\&I}$ higher than 60% and significant differences were found from the D&I labeling experiments ($p_{D\&I}$<0.05).The high RSDs and the low *p*-values obtained for the aforementioned peptides reveal that they cannot be used as potential peptides for the quantification of glycogen phosphorylase, as shown in Table VI.1. The advantage of the proposed approach is easily shown by comparing protein quantification by the DPD method, with the quantification commonly done by approaches relying on the peptides with the highest intensities in the MALDI spectra. Thus, taking the four peptides showing the most intense peaks in the MALDI spectrum of glycogen phosphorylase, see figure VI.2.

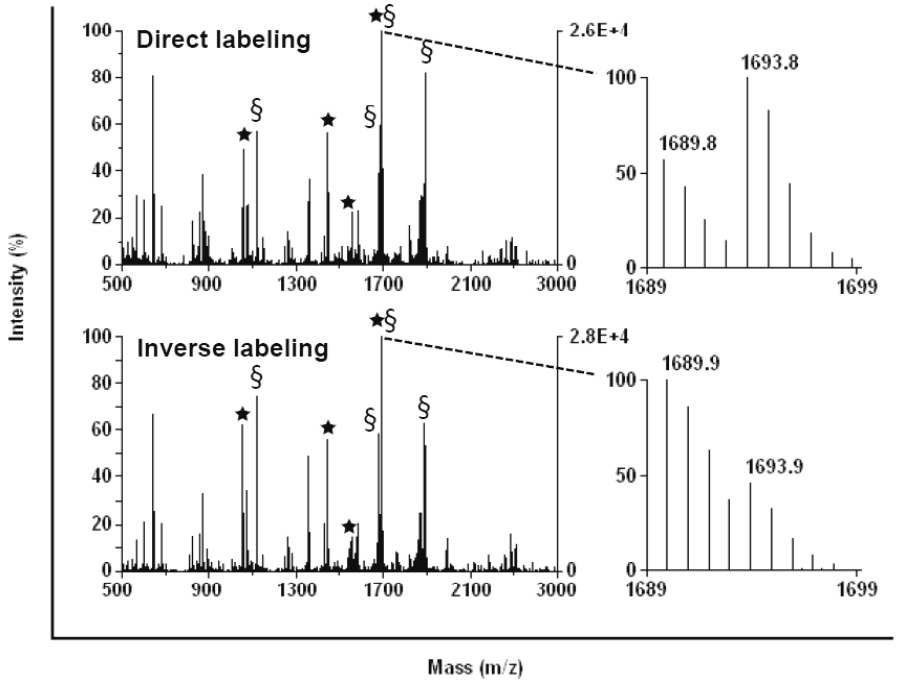


Figure VI. 2. MALDI-TOF-MS spectra of glycogen phosphorylase obtained for the direct (0.84µg of unlabeled protein and 1.68µg of ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis) and inverse labeling (0.84µg of ^{18}O-labeled protein and 1.68µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis). ★peptides assigned by the DPD software, § indicates the peptides used for the quantification with the most intense peaks and ★§ indicates peptides that were simultaneously assigned by DPD and are the most intense peaks.

The experimental values calculated for 0.84µg and 1.68µg of protein are 1.0±0.2µg (RSD 20%), and 2.1±0.7µg (RSD 33%) respectively whilst using the DPD program the values obtained are 0.79±0.06 (RSD 8%) and 1.7±0.02 (RSD 12%).

As a further example, Table VI.2 shows the differences in the accuracy obtained calculating the amount of six different proteins using (i) the peptides addressed by the DPD approach or (ii) the four peptides with the most intense peaks in the MALDI spectrum. Data shown in table VI.2 further confirms that only the DPD approach can guarantee accuracy. In addition table VI.9.SM.2 of supplementary material, shown the peptides used for quantification for each protein for the DPD and the most intense peaks-based method respectively, along with their relative intensities. In supplementary material VI.9.SM.3 shown the peptides assigned by the DPD software as candidates of bovine serum albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor and α-lactalbumin.

Table VI. 2. Amounts of different proteins loaded into the 1D-gel for the direct and inverse labeling and the respective quantification using (i) the DPD methodology, (ii) the most intense peak of the mass spectra and (iii) the four most intense peaks of the mass spectra. For the quantification done with the DPD methodology the number of peptides used is indicated between branches.

Protein	Direct labeling			Inverse labeling		
	Amounts of protein loaded (µg)	Quantification		Amounts of protein loaded (µg)	Quantification	
		DPD peptides	4 most intense peak		DPD peptides	4 most intense peak
Glycogen phosphorylase *Oryctolagus cuniculus*	0.84	0.79±0.06 (4) RSD 8%	1.0±0.2 RSD 20%	1.68	1.7±0.2 (4) RSD 12%	2.1±0.7 RSD 33%
Serum Albumin *Bos taurus*	1.04	1.01±0.09 (4) RSD 9%	1.00±0.09 RSD 9%	2.08	2.13±0.26 (4) RSD 12%	2.3±0.2 RSD 9%
Ovalbumin *Gallus gallus*	1.84	1.77±0.12 (2) RSD 7%	2.2±0.8 RSD 37%	3.68	3.83±0.15 (2) RSD 4%	4±2 RSD 50%
Carbonic anhydrase II *Bos taurus*	1.04	1.00±0.2 (2) RSD 20%	3.2±4.1 RSD 128%	2.08	2.0±0.2 (2) RSD 10%	2.8±1.4 RSD 50%
Trypsin inhibitor A *Glycine max*	1	1.0±0.2 (1) RSD 19%	1.2±0.9 RSD 75%	2	2.1±0.3 (1) RSD 14%	1.8±0.6 RSD 33%
α-lactalbumin *Bos taurus*	2.18	2.2±0.2 (2) RSD 9%	2.3±0.2 RSD 9%	2.9	3.4±0.3 (2) RSD 9%	3.7±0.6 RSD 16%
Vitellogenin *Cyprinus carpio*	1	1.09±0.06 (5) RSD 6%	1.2±0.1 RSD 8%	2	2.0±0.2 (5) RSD 10%	2.2±0.2 RSD 9%

The strength of the DPD procedure lies in that, instead of looking for the ±2/4Da isotope pair and quantitatively calculating the ratio ^{16}O to ^{18}O signals for every peptide, one only needs to compare the two data sets and identify those peptides whose ratios ”direct” and ”inverse” remains unaltered. This is achieved rapidly and automatically using the DPD approach. In addition, this experiment is required only once for a given protein, since the selected peptides will drive the quantification of the protein from which they originated.

Figure VI.3 shows the number of peptides obtained with the DPD approach for the proteins bovine serum albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor and α-lactalbumin. The accuracy and precision of the quantification was excellent for all cases, when the peptides with $p>0.05$ (those included in groups named as I) were used, indicating the robustness of the procedure. One finding was that, as the size of the protein increases, so does the number of “DPD peptides” as well.

VI.5.3 The use of “DPD peptides” as internal standards

As explained above, the DPD method allows for fast and unambiguous identification of peptides that can be used for the accurate quantification of proteins. Such peptides have losses and yields of digestion that are paralleled in different sets of the same sample treatment workflow within a given range amount of protein loaded in the gel.

Once the peptides are known, they can be used as internal standards in further quantifications of the protein. This means that the DPD method is required only once for each protein of interest. It should be noted that an internal standard must have some crucial intrinsic properties, such as that it behaves in a similar manner to the authentic sample although the conditions of sample preparation may change over a wide range [38]. This is the case for the “DPD peptides”. They can be used as internal standards because their recoveries after digestion, their losses and their sample deposition and crystallization behaviour in MALDI remain constant over a wide range of protein concentration.

To demonstrate this, 0.4, 1 and 2μg of BSA, carbonic anhydrase and α-lactalbumin were loaded into a 1D-gel. 2μg of each protein was also loaded to act as internal standard in the labeling experiments. Then proteins were separated by electrophoresis, the gel bands were excised and in-gel digested. A three-fold or greater difference in expression levels is generally considered to be statistically significant [39].

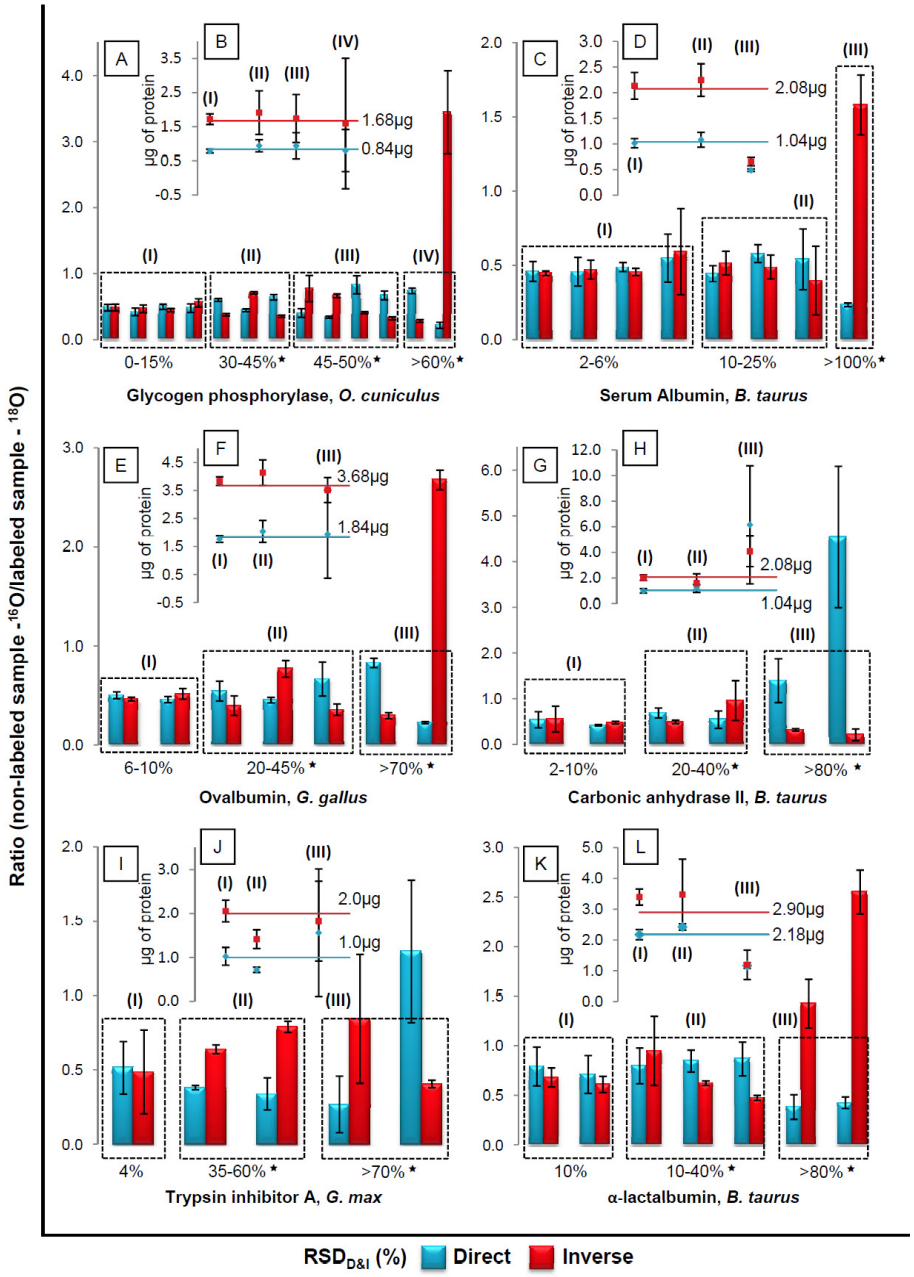


Figure VI. 3. Results obtained with the DPD approach for glycogen phosphorylase, bovine serum albumin, ovalbumin carbonic anhydrase, trypsin inhibitor and α-lactalbumin. Panels A, C, E, G, I and K: x-axes shows the RSD of the direct (blue colour) and inverse experiments (red colour), $RSD_{D\&I}$. Peptides are grouped as a function of the $RSD_{D\&I}$ and only those included in the group named (I), can be used for accurate quantification. As an example in panel A, (I) means that four peptides can be used for accurate quantification. The expected direct and inverse ratios are 0.5 for all proteins except for α-lactalbumin for which the expected ratio is 0.75. Insets B, D, F, H, J and L show the quantification results obtained for each protein using the peptides included in each group. For instance, in inset B, the amount of protein is calculated four times. Each value is named as (I), (II), (III) and (IV) and correlates with the group of peptides with the same letters that

were used for the calculation.

We chose, however, a higher difference. Equal volumes of digested protein (0.4, 1 and 2µg) and their corresponding internal standard (2µg) were mixed obtaining 1:5, 1:2 and 1:1 mixtures of protein/^{18}O-labeled internal standard. As can be seen in figure VI.4 the calculated amounts of protein were in excellent agreement with their expected values.

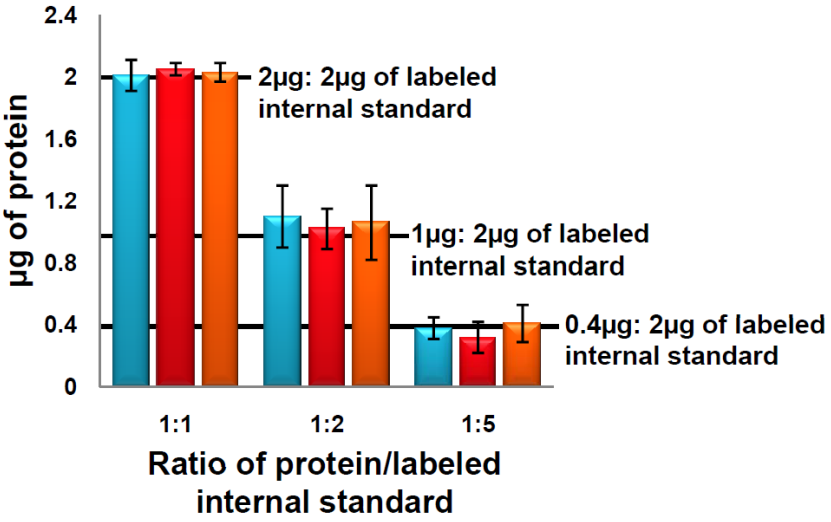


Figure VI. 4. DPD protein quantification of 2, 1 and 0.4µg of (1) bovine serum albumin, (2) carbonic anhydrase and (3) α-lactalbumin. The quantification was carried out with different ratios of protein/^{18}O-labeled internal standard, 1:1, 1:2, 1:5. No statistical differences were found between the calculated value and the amount of protein loaded on the 1D-gel (p>0.05, t-test, n=4).

VI.5.4 A case study: quantifying 17β-estradiol-induced vitellogenin in fish

Endocrine disruptor compounds, EDCs, are a group of chemicals present in sewage effluents [31,40]. EDCs can simulate the action of estrogens in some living organism, such as fish. One of the most sensitive responses to estrogens in fish is the induction of vitellogenin (Vtg), which is used as a biomarker of exposure to estrogenic chemicals in an aquatic environment [41]. Vtg is a phospholipoglycoprotein synthesized in all oviparous liver vertebrates. It serves as the major precursor to the egg yolk proteins [31]. In other words, this protein is expected to be at high concentration in mature females. The gene for Vtg is found in the liver of females but also in males, so Vtg production can be activated when males are exposed to EDCs [30].

The effects of the estrogen in the exposed males can be seen in figure VI.5, as an increment in the intensity of the gel band corresponding to a plasma sample of exposed males (lanes G and H), if the Vtg band is compared to its counterpart in non-exposed males (lanes E and F). This indicates a clear Vtg over-expression response to a single intraperitoneal injection of 17β-estradiol.

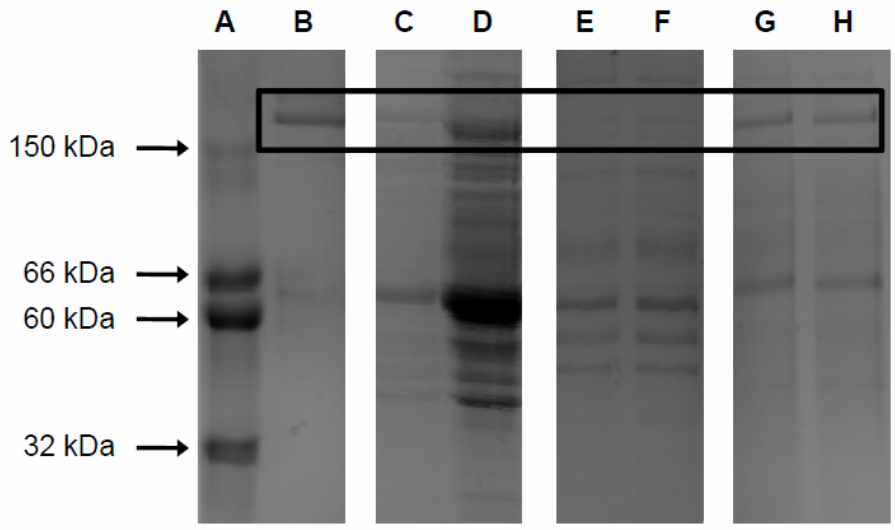


Figure VI. 5. 1D-gel electrophoresis of A- molecular weight marker; B-1µg of *Cyprinus carpio* vitellogenin standard; C- Precipitated plasma with $MgCl_2$/EDTA method (250µL of plasma were precipitated and the resulting pellet was reconstituted in 5µL of sample buffer and loaded on the gel); D- Plasma precipitated with the acetone method (80µL of plasma were precipitated and the resulting pellet was reconstituted in 100µL of sample buffer and diluted 1:2 before loading on the gel); E and F- Plasma from control males (non-exposed) precipitated by the acetone method; G and H- Plasma from exposed males precipitated by the acetone method (dilution 1:50).

To show the great potential of the proposed DPD method, we studied the effects of EDCs in fish, focusing on the identification and quantification of the Vtg protein. Plasma from male *Cyprinus carpio* exposed and non-exposed to EDCs was collected as described in the ample treatment section. Since plasma is a complex protein mixture, we followed two protocols of protein precipitation, one with Cl_2Mg/EDTA and other with acetone, to obtain the best subsequent separation of the target protein by 1D-PAGE. Both methods are described in detail in the supplementary material section (VI.9.SM.4). 1D-PAGE of C*yprinus carpio* plasma submitted to both protein precipitation protocols is shown in figure VI.5, where it may be seen that the most intense bands of Vtg were obtained after protein precipitation with the acetone protocol, which was consequently chosen for further

experiments. Next, the DPD approach was run using standards of Vtg to identify the peptides that could be used to perform subsequent accurate Vtg quantification. The results of the DPD software for Vtg are shown in table VI.3.

Table VI. 3. Peptides assigned by the DPD software as candidates for quantification for vitellogenin from *Cyprinus carpio*. Direct labeling: 1µg of unlabeled protein and 2µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1/2). Inverse labeling: 1µg of ^{18}O-labeled protein and 2µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected inverse ratio is 2 corresponding to the unlabeled/labeled ratio (2/1). (note: the inverse ratio is expressed in the table as (inverse ratio)-1 for simplicity. In this way the direct and inverse ratios may be compared easily since both have the same value of 0.5). For the direct method the amount of protein calculated was the lower of 1µg whilst for the inverse method was the higher of 2µg. DPD software parameters used: Peak intensity - 5; Peptide mass tolerance ±0.25Da; experimental overlapping - 2; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test).

Mass In-silico	RDS(%) D&I	Direct ratio	Inverse ratio	p-value D&I	Peptide sequence – In-silico	$RSD_{D\&I}$(%) $p_{D\&I}$	Direct µg of protein	Inverse µg of protein
1588.87	1.51	0.51±0.01	0.52±0.04	6.82E-01	AYLAGAAADVLEIGVR			
1139.64	2.25	0.58±0.03	0.56±0.07	6.92E-01	LELEVQVGPR			
821.46	4.43	0.56±0.09	0.53±0.06	6.09E-01	ALHPEVR	2-15%	1.09±0.06	2.0±0.2
1570.91	13.17	0.54±0.03	0.45±0.06	8.27E-02	FLGNAVPPVFAVIAR			
1129.71	13.30	0.53±0.02	0.44±0.09	2.10E-01	FVQLIQLLR			
1646.83	16.63	0.56±0.04	0.44±0.06	5.46E-02	NDPLYYIIGQHSAR			
1287.67	24.62	0.63±0.05	0.44±0.03	9.21E-03	AEAGVLGEFPAAR	15-30% ⋆	1.21±0.08	2.31±0.09
1677.94	28.28	0.62±0.09	0.41±0.04	4.52E-02	VFAPAGVSPTVLNLHR			
1044.45	31.01	0.44±0.08	0.68±0.08	2.18E-02	DLSHCQER[CysCAM]			
1354.75	37.58	0.66±0.05	0.38±0.05	3.28E-03	ICVDGALLSKHK [Cys PAM]	30-50% ⋆	1.3±0.3	2.4±0.7
1065.54	41.45	0.9±0.2	0.44±0.05	3.12E-02	TATIMEPFR			
1425.72	49.37	0.67±0.04	0.32±0.07	5.09E-03	EVVMLGYGSMIAR			
1610.79	52.11	0.4±0.9	0.93±0.02	3.66E-02	GCVELHSHNAAFIR[CysCAM]			
1463.89	54.34	0.8±0.1	0.37±0.05	8.79E-03	LVQPVALQLVLDR	>50% ⋆	1.0±0.7	1.4±1.2
976.48	119.28	0.20±0.01	2.30±0.47	1.63E-02	EILETEDK			

Peptides with $RSD_{D\&I}$ below 15% used for quantification were AYLAGAAADVLEIGVR, LELEVQVGPR, ALHPEVR, FLGNAVPPVFAVIAR and FVQLIQLLR. Finally, we proceeded to quantify the Vtg protein in exposed and non-

exposed males using the peptides obtained through the DPD method as internal standards. Therefore, we run in the same gel a Vtg standard of known concentration to be used as internal standard, along with plasma samples of exposed and non-exposed males. Gel pieces of the same size were then excised and treated as described in the sample treatment section. Samples were reconstituted in ^{16}O water, unless the Vtg standard that was reconstituted in ^{18}O-water. Then, the ^{18}O-label Vtg peptides were used as internal standards, by mixing them with the same volume of unlabeled samples. The MALDI mass spectrum thus obtained (figure VI.6), was used for Vtg quantification. This can be automatically done with the DPD program.

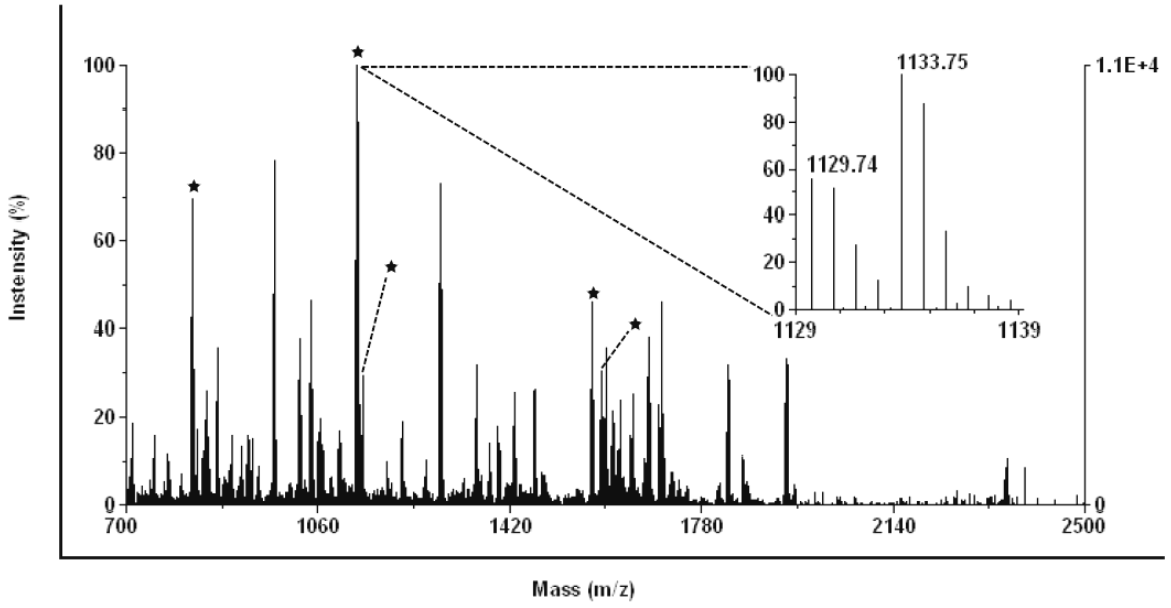


Figure VI. 6. MALDI mass spectrum obtained for the quantification of Vtg (0.5µg of ^{18}O-labeled internal standard).* DPD-peptides

The advantage of this procedure, as explained throughout all this manuscript, is that using the "DPD peptides" any selective peptide extraction, different yields of digestion or biased losses of peptides due to digestion or post-digestion sample processing are paralleled for all samples. To further validate the proposed methodology, we developed in parallel the analysis of the same samples through the traditional ELISA procedure. Results of Vtg quantification for exposed and non- exposed males using both DPD and ELISA procedures are shown in Table VI.4.

Table VI. 4. Quantification of Vtg from *Cyprinus carpio* fish plasma by enzyme-linked immunosorbent assay (ELISA) and by DPD. Nonparametric Mann-Whitney *U*-test: no statistical differences were found between both quantitation methods.

	ELISA (µg/µL) (n=3)	DPD (µg/µL) (n=3)	Mann-Whitney U-test
Control male I	0.081±0.001	0.09±0.04	p>0.05
Control male II	0.070±0.004	0.06±0.03	p>0.05
Control male III	0.0603±0.0005	0.08±0.05	p>0.05
Exposed male I	0.17±0.02	0.14±0.01	*p*>0.05
Exposed male II	0.178±0.002	0.26±0.03	*p*=0.05
Exposed male III	0.389±0.03	0.51±0.05	*p*=0.05
Exposed male IV	0.108±0.001	0.28±0.03	*p*=0.05
Exposed male V	0.117±0.005	0.13±0.01	*p*>0.05
Exposed male VI	0.216±0.008	0.28±0.05	*p*=0.05

In literature has been described non-parametric statistics tests-based approaches to compare results from differential protein abundance by MS and liquid chromatography, [42,43]. In our case, a nonparametric Mann-Whitney U-test shows that no statistical differences between both quantitation methods were present. As a further example, if the Paired t-test is here applied considering each *Cyprinus carpio* as a test sample, we may think in nine independent samples with 3 replicates for each one, and therefore we may use the paired t-test to compare the methods. The paired t-test described above does not require that the precision of the two methods are equal, as it is our case. Results reveals a calculated *t* value of 0.088 well below the critical value of 4.30 (P = 0.05). This means that the methods do not give statistically different media for the vitellogenin concentrations. The above cited results are excellent and reveal the great potentiality of the DPD approach. One further point worth remembering - the ELISA method can only be performed if the antibody of the protein to be quantified exists. The DPD approach is without such a limitation. The results confirm that the amount of protein obtained using the two methodologies are of the same magnitude.

VI.5.5 Merits and limitations

The method described in this work, named as DPD, allows the identification of peptides for which selective extraction, different recoveries after digestion or biased losses due to digestion or post-digestion sample processing of gel bands are paralleled in different experimental sets. In addition, as demonstrated by the experimental data, the DPD peptides behave in the same manner in MALDI, so shot-to-shot, region-to-region and sample-to-sample reproducibility are guaranteed. Those peptides allow for accurate quantification of a target protein in an automatic mode. The DPD strategy can be extended to other labeling methods to achieve the same goal of fast and accurate protein quantification. Although ^{18}O-labeling has the drawbacks that isotope enrichment may not be 100%, and that incorporation of the isotope may never go to completion. In our laboratory, however, we have demonstrated [35] that the decoupled labeling approach offers a unique method to obtain almost complete labeling. This is that the case since labeling is carried out in an environment 97% rich in ^{18}O-water, and that when the peptides are reconstituted in this environment, the double oxygen incorporation reaches a value higher than 90% for almost all peptides, as shown in table VI.5. Therefore, the concentration of peptides that are not labeled or that only incorporate one oxygen may be considered residual, and therefore the accuracy of the method does not suffer. In addition, the utilization of algorithms to correct failed or incomplete labeling, necessary when the oxygen incorporation is done directly, are not necessary here.

A further advantage of this protocol is the total automation of the quantification process, once the “DPD peptides” have been selected. A minor but ever present constraint is decidedly the 2-fold reduction in sensitivity, since two experiments (direct and inverse) are necessary thus halving the available sample. However, the advantages cited above overwhelmingly outweigh this loss in sensitivity. Furthermore, in typical proteomic studies the limiting factor is often handling capacity rather than material supply.

VI.6 Conclusions

A new method for bottom-up protein quantification termed decision peptide driven, DPD, has been established and experimentally validated. The methodology, which employs the combination of in gel-protein separation, in-gel proteolytic digestion, ^{18}O peptide labeling and the inverse labeling strategy, has been demonstrated successfully in the fast

and accurate quantification of seven different proteins. Furthermore, the DPD approach has been applied with success to study the effects of high concentration levels of estrogens on *Ciprinus carpio* males. To do so, protein Vtg, a biomarker of exposure to high levels of estrogens, was accurately quantified using the DPD method. The DPD method necessitates only the identification of peptides with paralleled losses and similar digestion yields through different in gel-digestion sets. Using these peptides accurate quantification of the protein is achieved. Suitable software has been also developed to automate the task. The methodology described here can be used to provide a quick and automated quantification of proteins that are differentially expressed, to elucidate drug action mechanisms or to study drug toxicity. Proteins that are differentially expressed upon a drug treatment are also potentially quantified. Finally, this approach can be also applied to other bottom-up strategies with off-gel based protein separation.

Table VI. 5. Relative labeling efficiency of peptides formed by the tryptic digestion of vitellogenin and postdigestion labeling in the presence of 97% $H_2{}^{18}O$.

Mass In-silico	**Peptide sequence - In-silico**	**$^{16}O/^{18}O$**	**$^{18}O_1/^{18}O_2$**
1588.8693	AYLAGAAADVLEIGVR	<0.1	0.10±0.03
1139.6418	LELEVQVGPR	<0.1	0.11±0.01
821.4628	ALHPEVR	<0.1	0.09±0.02
1570.9103	FLGNAVPPVFAVIAR	<0.1	0.24±0.03
1129.7092	FVQLIQLLR	<0.1	0.12±0.03
1646.8285	NDPLYYIIGQHSAR	<0.1	0.07±0.02
1287.6691	AEAGVLGEFPAAR	<0.1	0.05±0.01
1677.9434	VFAPAGVSPTVLNLHR	<0.1	0.08±0.01
1044.4528	DLSHCQER[CysCAM]	<0.1	0.18±0.02
1354.7511	ICVDGALLSKHK[Cys PAM]	<0.1	0.09±0.04
1065.5397	TATIMEPFR	<0.1	0.20±0.04
1425.7228	EVVMLGYGSMIAR	<0.1	0.13±0.04
1610.7856	GCVELHSHNAAFIR[CysCAM]	<0.1	0.15±0.05
1463.8944	LVQPVALQLVLDR	<0.1	0.25±0.12
976.4833	EILETEDK	<0.1	0.07±0.02

VI.7 Acknowledgements

H. M. Santos acknowledges the doctoral grant SRFH/BD/38509/2007 provided by FCT (Science and Technological Foundation) of Portugal. D. Glez-Peña acknowledges the Xunta de Galicia (Spain) for the program María Barbeito. Dr. José-Luis Capelo-Martínez and Dr. Carlos Lodeiro are grateful to the Xunta de Galicia (Spain) for the program Isidro Parga Pondal and for financial support given under project 09CSA043383PR. University of Vigo is thanked for financial support under projects InOu-Univ.Vigo 2009-K915 and K914. M. Diniz acknowledges financial support given by REQUIMTE (Portugal). The research findings here reported are protected under patent pending PCT/IB2006/052314 and PT 103 303.

VI.8 References

[1] Abbott A. *Nature* 1999, 402, 715.

[2] Quadroni M., James P. *Electrophoresis* 1999, 20, 664.

[3] Domon B., Abersold R. *Science* 2006, 312, 212.

[4] Lange V., Picotti P., Domon B., Aebersold R. *Mol. Sys. Bio.* 2008, 4, 1.

[5] Karbassi, I.D., Nyalwidhe, J. O., Wilkins, C. E., Cazares, L., Lance, S. R., Semmes, J., Drake, R. *J. Proteom. Res*. 2009, 8, 4182.

[6] Jahn, O., Tenzer, S., Werner H. B. (2009) *Mol. Neur.* 2009, 40, 55.

[7] Bao, H.M., Song, P.M., Liu, Q.P., Liu, Y.K., Yun, D., Saiyin, H., Du, R.Y., Zhang, Y., Fan, H.Z., Yang, P.Y., Chen, X. *Proteomics Clin. Appl.*, 2009, 6, 705.

[8] Ong, S.E., Mann, M., *Nature Chemical Biology*, 2005, 1, 252.

[9] Sakai, J.; Kojima, S.; Yanagi, K.; Kanaoka, M. *Proteomics.* 2005, 5, 16

[10] Niles, R., Witkowska, H.E., Allen, S., Hall, S.C., Fisher, S.J., Hardt, M. *Anal. Chem*. 2009, 81, 2804.

[11] Liu, Z., Cao, J., He, Y., Qiao, L., Xu,C., Lu, H., Yang, P., *J. Proteome. Res.* 2010, 9, 227.

[12] Rachdaoui, N., Austin, L., Kramer, E., Previs, M.J., Anderson, V.E., Kasumov, T., Previs, S.F. *Mol Cell Proteomics* 2010, 8, 2653.

[13] Capelo, J. L., Carreira, R. J., Fernandes, L., Lodeiro, C., Santos, H. M., Simal, J. *Talanta* 2010, 80, 1476.

[14] Fenselau, C., Yao, X. *J. Proteom. Res.* 2009, 8, 2140.

[15] Qian, W.J., Liu, T., Petyuk, V.A., Gritsenko, M.A., Petritis, B.O., Polpitiya, A.D., Kaushal, A., Xiao, W., Finnerty, C.C., Jeschke, M.G., Jaitly, N., Monroe, M.E., Moore, R.J., Moldawer, L.L., Davis, R.W., Tompkins, R.G., Herndon, D.N., Camp, D.G., Smith, R.D. *J. Proteome Res.* 2009, 8, 290.

[16] Lopez-Ferrer, D., Canas, B.,Vazquez, J., Lodeiro, C., Rial-Otero, R., Moura, I., Capelo, J. L. *Trac-trends Anal. Chem.* 2006, 2, 996.

[17] Wisniewski, J.R., Zougman, A., Nagaraj, N., Mann, M. *Nat. Methods* 6 (2009) 359

[18] Havlis, J., Shevchenko, A. *Anal. Chem*. 2004, 76, 3029.

[19] Heller, M., Mattou, H., Menzel, C., and Yao, X. *J. Am. Soc. Mass. Spectrom.* 2003, 14, 704.

[20] Sally L. Shirran, Catherine H. Botting, *J. Proteomics.* 2010, 73, 391.

[21] Wu, W.W., Wang, G.H., Baek, S.J., Shen,R.F. *J. Proteome Res.* 2006, 5, 651

[22] Nicola A.J., Gusev A.I., Proctor A., Jackson, E.K., Hercules, D.M. *Rapid Commun. Mass Spectrom.* 1995, 9, 1164

[23] Duncan, M. W., Roder, H., Hunsucher, S. W. *Brief Funct. Genomic. Proteomic*. 7 (2008) 355

[24] Krause, E., Wenschuh, H., Jungblut, P.R. Anal. Chem. 1999, 71, 4160.

[25] Wang Y.K., Ma Z.X., Quinn D.F., Fu E.W. Anal. Chem. 2001, 72, 3742.

[26] Cordeiro, F. M., Carreira, R. J., Rial-Otero, R., Rivas, M. G., Moura, I., Capelo, J. L. *Rapid Commun. Mass Spectrom*. 2007, 21, 3269.

[27] Galesio, M., Vieira, D.V., Rial-Otero, R., Lodeiro, C., Moura, I., Capelo, J.L. *J. Proteome Res.* 2008, 7, 2097.

[28] Carreira, R., Lodeiro, C., Diniz, M. S., Moura, I., Capelo, J. L. *Proteomics*, 2009, 9, 1.

[29] Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C. *Anal. Chem.* 2001, 73, 2836.

[30] Denslow, N.D., Chow, M.C., Kroll, K.J., Green, L. *Ecotoxicology*, 1999, 8, 385.

[31] Diniz, M.S., Peres, I., Magalhaes-Antoine, I., J. Falla, J., Pihan, J. C. *Ecotoxicology and Environmental Safety* 2005, 62, 427.

[32] Glez-Pena, D., Diaz, F., Hernandez, J.M., Corchado J.M., Fdez-Riverola F. *BMC Bioinformatics*, 10 (2009) doi:10.1186/1471-2105-10-187.

[33] White, C. A., Nicomedus, O., Emili, A. *J. Proteome Res*. 2009, 8, 3653.

[34] Yao, X. D., Afonso, C., Fenselau, C. *J. Proteom. Res.* 2003, 2, 147.

[35] Carreira, R. J., Rial-Otero, R., López-Ferrer, D., Lodeiro, C., Capelo, J. L. *Talanta*, 2008, 76, 400.

[36] Stewart, I. I., Thomson, Ty., Figeys, D. *Rapid Commun. Mass Spectrom.* 2001, 15, 2456.

[37] Speicher, K., Kolbas, O., Harper, S., Speicher, D. *J. Biomol. Tech.* 2000, 11, 74.

[38] Harris, D. C., in Quantitative Chemical Analysis, W. H. Freeman and Company, New York, Chap. 5, 91-92, 2006

[39] Page, M. J.; Amess, B.; Rohlff, C.; Stubberfield, C.; Parekh, R. *Drug Discovery Today*, 1999, 4, 55.

[40] Alda, M.J.L., Barcelo, D. *Fresenius J. Anal. Chem.* 2001, 371, 437.

[41] Jobling, S., Casey, D., Rodgers-Gray, T., Oehlmann, J., Schulte-Oehlmann, U., Pawlowski, S., Baunbeck, T., Turner, A.P., Tyler, C.R. *Aquat. Toxicol.* 2003, 65, 205.

[42] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B. *Anal. Bioanal. Chem.* 2007, 389, 1017.

[43] Anderle, M, Roy, S., Lin, H., Beckerand, C., Joho, K. *Bioinformatics*, 2004, 20, 3575.

VI.9. Supplementary material

VI.9.SM.1 Glycogen phosphorylase muscle from *Oructolagus cuniculus*

Data to run the DPD software for glycogen phosphorylase. The in-silico data (mass, *in-silico* peptide sequence, presence of carbamidomethylation of cysteines, and values of the isotopic distribution of the peptide I0, I2 and I4 used to calculate the ratios).
This information is available in the web page http://sing.ei.uvigo.es/DPD

VI.9.SM.2. Comparison of the DPD peptides and the 4 most intense peaks in the mass spectra for all the proteins studied. Peak intensities of each peak are shown in branches

Protein	DPD peptides	4 most intense peak
Glycogen phosphorylase *Oryctolagus cuniculus*	1689.89 (99±1) 1053.57 (64±1) 1440.74 (56±3) 1550.77 (18±2)	1689.89 (99±1) 1117.56 (87±10) 1886.9 (66±1) 1678.86 (60±2)
Serum Albumin *Bos taurus*	1479.80 (100) 1639.94 (38±3) 927.49 (31±2) 1439.81 (18±3)	1479.80 (100) 1567.74 (36±4) 1880.92 (34±3) 927.49 (31±2)
Ovalbumin *Gallus gallus*	1687.84 (100) 1773.9 (17±1)	1687.84 (100) 1858.97 (60±2) 2284.15 (45±6) 1555.72 (20±1)
Carbonic anhydrase II *Bos taurus*	2198.22 (98±2) 1581.82 (16±2)	2198.22 (98±2) 1018.5 (47±10) 2584.16 (19±1) 1581.82 (16±2)
Trypsin inhibitor A *Glycine max*	1200.71 (24±7)	997.42 (80±24) 1163.64 (58±25) 1200.71 (24±7) 1230.53 (15±2)
α-lactalbumin *Bos tauruS*	1200.62 (69±10) 710.31 (32±4)	1200.62 (69±10) 710.31 (32±4) 1091.52 (26±3) 1779.84 (10±1)
Vitelogenin *Cyprinus carpio*	1129.71 (98±4) 821.42 (82±15) 1570.91 (51±5) 1588.87 (30±1) 1139.64 (33±3)	1129.71 (98±4) 974.49 (90±11) 1287.67 (84±14) 821.42 (82±15)

VI.9.SM.3 Peptides assigned by the DPD software as candidates of bovine serum albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor and α-lactalbumin

VI.9.SM.3. 1. Peptides assigned by the DPD software as candidates for quantification of BSA. Direct labeling: 1.04µg of unlabeled protein and 2.08µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1.04/2.08). Inverse labeling: 1.04µg of ^{18}O-labeled protein and 2.08µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: Peak intensity - 3; Peptide mass tolerance ±0.25Da; experimental overlapping - 2; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test); ‡ Expected values

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (1.4)‡	Inverse - µg of protein (2.08)‡
1639.94	1.8	0.46±0.07	0.45±0.01	7.61E-01	KVPQVSTPTLVEVSR	2-6%	1.0+0.1	2.1+0.3
1479.80	2.2	0.5±0.1	0.5±0.1	8.10E-01	LGEYGFQNALIVR			
927.49	4.8	0.49±0.03	0.46±0.02	1.51E-01	YLYEIAR			
1439.81	5.4	0.5±0.2	0.6±0.03	8.04E-01	RHPEYAVSVLLR			
1880.92	10.6	0.4±0.1	0.5±0.1	1.92E-01	RPCFSALTPDETYVPK[CysCAM]	10-25%	1.1±0.1	2.2±0.3
1894.94	11.9	0.6±0.1	0.5±0.1	1.27E-01	RPCFSALTPDETYVPK[CysPAM]			
1567.74	21.5	0.5±0.2	0.4±0.2	3.91E-01	DAFLGSFLYEYSR			
1482.80	104.8	0.23±0.01	1.6±0.2	8.79E-04	LCVLHEKTPVSEK	>100% ⋆	0.49±0.02	0.7±0.1

VI.9.SM.3. 2. Peptides assigned by the DPD software as candidates for quantification of ovalbumin. Direct labeling: 1.84µg of unlabeled protein and 3.68µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1.84/3.68). Inverse labeling: 1.84µg of ^{18}O-labeled protein and 3.68µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: Peak intensity - 5; Peptide mass tolerance ±0.25Da; experimental overlapping - 2; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test); ‡ Expected values

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (1.84)‡	Inverse - µg of protein (3.68)‡
1687.84	5.7	0.50±0.03	0.46±0.02	1.04E-01	GGLEPINFQTAADQAR	6-10%	1.77±0.12	3.8±0.2
1773.90	8.4	0.46±0.03	0.52±0.05	1.24E-01	ISQAVHAAHAEINEAGR			
1581.72	22.1	0.5±0.1	0.40±0.10	8.24E-02	LTEWTSSNVMEER	20-45% ⋆	2.0±0.4	4.1±0.5
1858.97	36.5	0.45±0.03	0.77±0.08	2.84E-03	ELINSWVESQTNGIIR			
1555.72	42.8	0.7±0.2	0.36±0.06	3.09E-02	AFKDEDTQAMPFR			
2284.15	67.0	0.83±0.05	0.30±0.03	5.90E-06	VTEQESKPVQMMYQIGLFR	> 70% ⋆	2±2	3.52±0.45
2281.18	119.4	0.23±0.01	2.7±0.1	1.57E-05	DILNQITKPNDVYSFSLASR			

VI.9.SM.3. 3. Peptides assigned by the DPD software as candidates for quantification of carbonic anhydrase. Direct labeling: 1.04µg of unlabeled protein and 2.08µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1.04/2.08). Inverse labeling: 1.04µg of ^{18}O-labeled protein and 2.08µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: Peak intensity - 2; Peptide mass tolerance ±0.25Da; experimental overlapping - 2; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test); ‡ Expected values

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (1.04)‡	Inverse - µg of protein (2.08)‡
1581.82	1.9	0.5±0.2	0.55±0.3	9.35E-01	YAAELHLVHWNTK	2-10%	1.0±0.2	2.0±0.2
2198.22	9.2	0.42±0.01	0.48±0.03	2.30E-02	AVVQDPALKPLALVYGEATSR			
1018.50	23.5	0.67±0.11	0.49±0.04	3.03E-02	DFPIANGER	20-40% ⋆	1.3±0.2	1.6±0.7
2098.88	39.2	0.5±0.2	0.96±0.4	1.52E-01	MVNNGHSFNVEYDDSQDK			
2852.48	88.5	1.4±0.5	0.32±0.03	2.06E-02	TLNFNAEGEPELLMLANWRPAQPLK	> 80% ⋆	6±5	4.1±1.2
2584.16	128.9	4.5±1.5	0.21±0.13	1.10E-02	LVQFHFHWGSSDDQGSEHTVDR			

VI.9.SM.3. 4. Peptides assigned by the DPD software as candidates for quantification of trypsin inhibitor. Direct labeling: 1.0µg of unlabeled protein and 2.0µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1.0/2.0). Inverse labeling: 1.0µg of ^{18}O-labeled protein and 2.0µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: Peak intensity - 2; Peptide mass tolerance ±0.3Da; experimental overlapping - 2; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test); ‡ Expected values

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (1)‡	Inverse - µg of protein (2)‡
1200.71	3.82	0.5±0.1	0.49±0.1	6.65E-01	NKPLVVQFQK	4%	1.0±0.2	2.1±0.2
997.42	35.82	0.4±0.1	0.6±0.1	9.24E-03	DAMDGWFR	35-60% ⋆	0.7±0.1	1.4±0.2
1013.41	56.65	0.3±0.1	0.8±0.1	1.89E-03	DAMDGWFR[MSO]			
1163.64	73.29	0.7±0.1	0.8±0.1	1.04E-03	GIGTIISSPYR	> 70% ⋆	1.6±1.5	1.8±0.9
1230.53	74.12	1.3±0.2	0.40±0.04	1.04E-03	VSDDEFNNYK			

VI.9.SM.3. 5. Peptides assigned by the DPD software as candidates for quantification of α-lactalbumin. Direct labeling: 2.18µg of unlabeled protein and 2.90µg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.75 corresponding to the unlabeled/labeled ratio (2.18/2.90). Inverse labeling: 2.18µg of ^{18}O-labeled protein and 2.90µg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: Peak intensity - 2; Peptide mass tolerance ±0.3Da; experimental overlapping - 2; Ratio I4/I0 - 0.15. ⋆Significant differences between D&I ratios ($p<0.05$, t-test); ‡ Expected values

Mass In-silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)‡	Inverse ratio (0.5)‡	$p_{D\&I}$value	Peptide sequence - In-silico	$RSD_{D\&I}$ (%)	Direct - µg of protein (2,18)‡	Inverse - µg of protein (2.90)‡
710.329	10.679547	0.8±0.2	0.7±0.1	3.61E-01	CEVFR[CysCAM]	10%	2.2±0.2	3.4±0.3
1200.6523	10.827975	0.7±0.2	0.6±0.1	3.86E-01	VGINYWLAHK			
653.3075	12.230605	0.8±0.2	0.9±0.3	4.81E-01	CEVFR			
1091.519	21.589453	0.8±0.1	0.62±0.02	2.43E-02	LDQWLCEK[CysCAM]	10-40% ⋆	2.4±0.1	3.5±1.2
1779.8405	41.37589	0.9±0.2	0.47±0.03	1.80E-02	ALCSEKLDQWLCEK[CysCAM]			
620.3225	81.74236	0.38±0.12	1.4±0.2	1.07E-03	IWCK[Cys PAM]	> 80% ⋆	1.2±0.1	1.2±0.5
549.2853	101.1064	0.4±0.1	2.55±0.22	1.50E-04	IWCK			

VI.9.SM.4 Protein precipitation methods

VI.9.SM.4.1 Precipitation with Cl_2Mg/EDTA

A 250 µl of plasma sample were gently mixed with 1 ml of 20 mM Na_2EDTA and 80 µl of 0.5 M of $MgCl_2$ and centrifuged at 2500g for 15 min. The supernatant was discarded and the pellet containing vitellogenin was suspended in 150 µl of 1M NaCl, 50 mM TriS-HCl (pH 7.5) and then centrifuged at 2500g for 30 min. The supernatant was removed to another centrifuge tube, and the vitellogenin was precipitated with 1.24 ml of water and centrifuged at 2500g for 15 min. The resulting pellet of purified vitellogenin was suspended in 5 µl of buffer (10 mM Tris-HCl pH 7.4; 2% of SDS; 1% of β-mercaptoethanol). (Wiley, H. S., Opresko, L., Wallace, R. A. (1979) New methods for the purification of vertebrate vitellogenin, Anal. Biochem. **97**,145-152)

VI.9.SM.4.2 Acetone precipitation

80 µl of plasma was diluted to 100 µl with cold PBS. 300 µl of -20 ºC cold acetone were added to the diluted plasma solution and kept overnight on ice. The sample was centrifuged at 10000g, 4 ºC for 30 min. The supernatant was removed and the pellet was suspended in 100 µl of buffer (10 mM Tris-HCl pH 7.4; 2% of SDS; 1% of β-mercaptoethanol).

Chapter VII.

On-tissue ^{18}O labeling for mass spectrometry-based applications

---

Manuscript in preparation:

H. M. Santos, J.L. Capelo, G. Corthals

Index

VII.1 Resumo

Neste trabalho, foi efectuado um estudo sobre macacão isotópica com ^{18}O em amostras de tecidos biológico após digestão. A metodologia proposta apresenta-se como uma forma fácil de distinguir péptidos de outros componentes presentes num tecido. A digestão enzimática foi alcançada ao fim de 3h após aplicação manual de tripsina na superfície do tecido. Após a digestão, a marcação isotópica foi efectuada em 20min de incubação com ^{18}O. O método proposto, é rápido e reprodutível, e abre uma nova linha de trabalho em espectrometria de massa aplicada a tecidos biológicos. Adicionalmente, apresenta-se pela primeira vez, a aplicação da energia de ultra-sons para acelerar a digestão enzimática de tecidos biológicos.

Palavras-chave: Espectrometria de massa aplicada a tecidos, marcação isotópica com ^{18}O, aceleração por ultra-sons.

A minha contribuição para este trabalho consistiu na execução da parte experimental, incluindo preparação dos tecidos e analise por MALDI-TOF/TOF-MS

VII.2 Abstract

In this work we report the on-tissue 18-O labeling of peptides after tryptic digestion as a powerful tool for the rapid mass spectrometry differentiation of peptides from other tissue components. Successful digestion was achieved in 3 h after the on-tissue pipette spotting of trypsin, following 18-O labeling of peptides for 20 min. The method is fast and reproducible and opens a new line of work in mass spectrometry of tissues because the principle can be potentially expanded to other type of labeling protocols. In addition we report for the first time on the ultrafast digestion, 30 s, of tissues using ultrasonication. Taken together this new findings will allow higher throughput and faster sample treatment for any tissue mass spectrometry-based application.

Keywords: Tissue mass spectrometry, 18-O labeling, ultrasonic enhancement.

My contribution to this work was the execution of the experimental including tissue preparation and direct tissue MALDI-TOF/TOF-MS analysis.

VII.3 Introduction

When applied over a sample as complex as a tissue, laser energy promotes the formation of a plume of molecules including, but not limited to, peptides and proteins [1,2]. Molecule identification after laser desorption from a tissue section through mass spectrometry, MS, is a task that presents many challenges because the MS tissue analysis is severely affected by a great number of factors that require careful optimization to obtain reliable results. The way by which the tissue is attached to the MALDI glass slide (including tissue thickness, chemical tissue coating, and temperature of the cryostat microtome chamber); the type of tissue cleaning previous matrix coating (including composition of the cleaning solution); the type of tissue (fatty versus non fatty); the type of matrix (including matrix concentration, matrix solution composition and matrix application) and finally the adequate tuning of the instrument settings to maximize signal quality across the entire spectrum are variables that need to be carefully optimized for each tissue section [3-5].

There has been substantial effort directed towards optimizing the sample preparation methods, including washing protocols, matrix application and software development [6-9]. However, even for the case of an optimized tissue analysis, the ions reaching the detector cannot be directly assigned to a given type of molecule, being not possible to unequivocally assign a given ion mass to, for instance, a peptide.

In the present work we report, to the best of our knowledge for the first time, on the direct on-tissue ^{18}O labeling of proteins as a fast, simple and reproducible manner to unequivocally assign ions reaching the detector as peptides in MS tissue analysis.

VII.4 Materials and Methods

VII.4.1 Material

All reagents used were HPLC-grade or higher. CHCA was purchased from Sigma–Aldrich. Sequence Grade Modified Trypsin from Promega, TFA, ammonium bicarbonate and ammonium acetate from Fluka Bio-Chemika, $H_2{}^{18}O$ (97 atom%) from Isotec™. Water was purified in Millipore's Milli-Q Synthesis system.

Tissue sections, 12μm thick, were cut on a Leica CM 3000 cryostat (Leica Microsystems, Nussloch, Germany) at -20 °C.

VII.4.2 Tissue preparation

Livers from adult *Mus musculus* were dissected and stored at – 80°C until analysis. Thin (12μm) tissue sections were prepared and thaw mounted onto a conductive, transparent glass slides (75x25mm) coated with an indium tin oxide (ITO) layer (Bruker Daltonics, Bremen, Germany) and frozen on dry-ice and stored at -80 °C. Tissue section were dried and equilibrated to room temperature during 2h in a desiccators followed by a series of ethanol/water washes. The glass slide was dipped in ethanol/ water (70:30% v/v) and gently agitated for 30s followed by a 30s wash in ethanol. Excess solvent were removed by a gentle blow of N_2.

VII.4.3 On-tissue enzymatic digestion

A solution containing 125ng/μl of trypsin in 12.5mM of ammonium bicarbonate was manually spotted onto the liver sections by pipette deposition to obtain 40ng of trypsin/mm^2. Two interactions of 500nL were used. Each digests spot appeared to dry completely between each successive spotting iteration. The trypsin spotting proceeded at room temperature (21 °C). After trypsin application the sample were placed in a humidified chamber at 37ºC and digested for 3h.

VII.4.4 On-tissue ^{18}O labeling and matrix coating

For the ^{18}O-labeling, to the digested peptides 1μL of $H_2{}^{18}O$/ 50mM of ammonium acetate pH 6.75 was added to the array of tryptic spots and incubated at room temperature for 20min. Following digestion and labeling, a solution containing 20 mg/ml of CHCA in 1:1 acetonitrile/0.5% TFA was spotted directly onto the array of tryptic spots over 1x500nL.

VII.4.5 Mass spectrometry of peptides

The tissue profiles were acquired using an Ultraflex II MALDI-TOF TOF instrument (Bruker-Daltonics Billerica, MA) equipped with a LIFT cell and Smartbean laser. For peptides, the mass spectrometer was operated with positive polarity in reflectron mode and spectra were acquired in the range of *m*/*z* 1000-4000. A total of 3000 spectra were acquired at each spot position at a laser frequency of 100 Hz.

VII.5 Results and Discussion

VII.5.1 Preliminary experiments

The digestion of tissue proteins through the addition of a solution containing an enzyme has been previously reported, and it is generally referred as on-tissue enzymatic digestion [10-17]. In order to verify if the working conditions were adequate in our workflow, as depicted in figure VII.1, including the correct tuning of the instrument settings, we devised a set of experiments in which we added a standard intact protein, carbonic anhydrase, CA, in different quantities over tissue sections taken from the same sample, liver from adult *Mus musculus*.

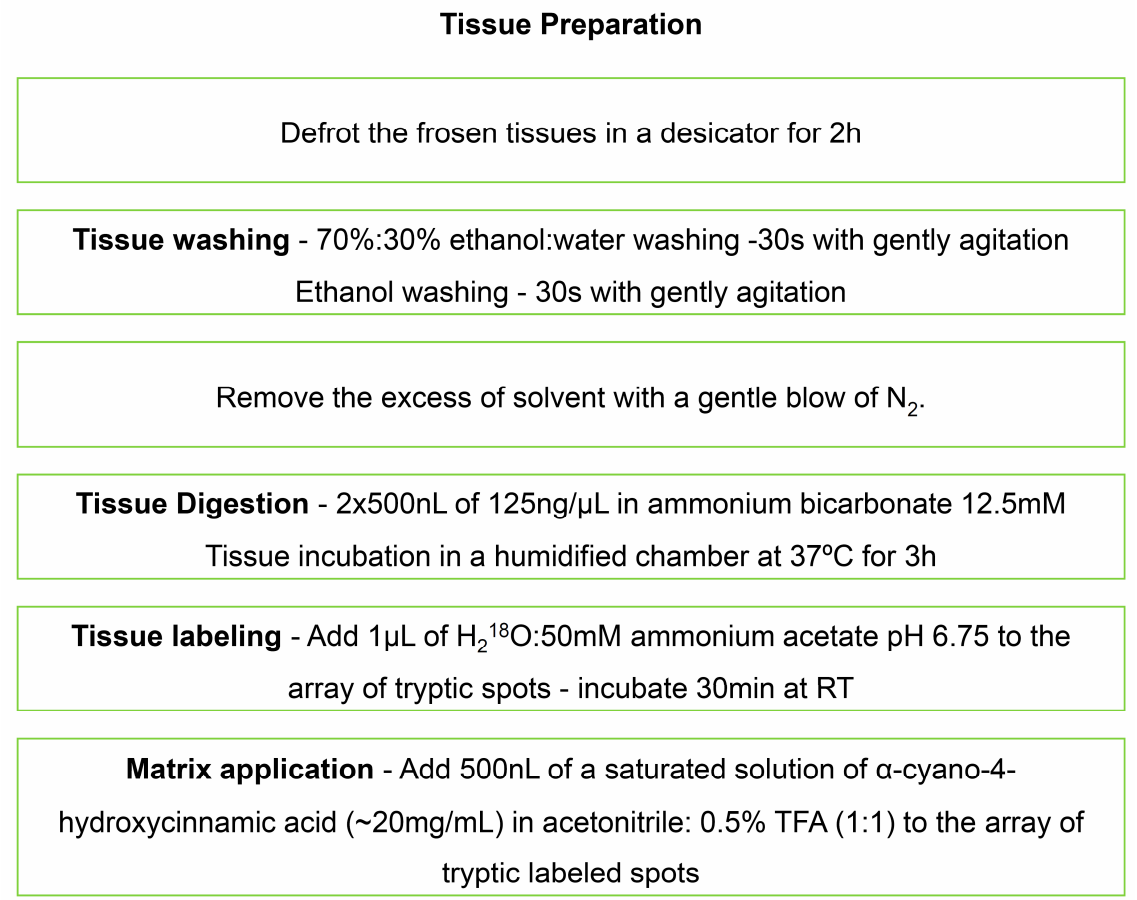


Figure VII. 1. Schematic diagram of the sample treatment workflow including: tissue washing, on-tissue trypsin digestion, ^{18}O labeling and matrix application.

Then a trypsin solution was pipette spotting over the tissue twice as described in the experimental section. It was allowed the trypsin to cleavage the tissue´s proteins during 3 h, in a humidified chamber, under controlled temperature conditions, 37 ºC. The reaction chamber was continuously monitored to control the drying process, to ensure that the enzyme solution was not immediately dried after deposition, what would hamper the protein cleavage [10].

As may be seen in figure VII.2, the peptides belonging to carbonic anhydrase can be only detected for protein amounts deposited over the tissue higher than 0.5µg.

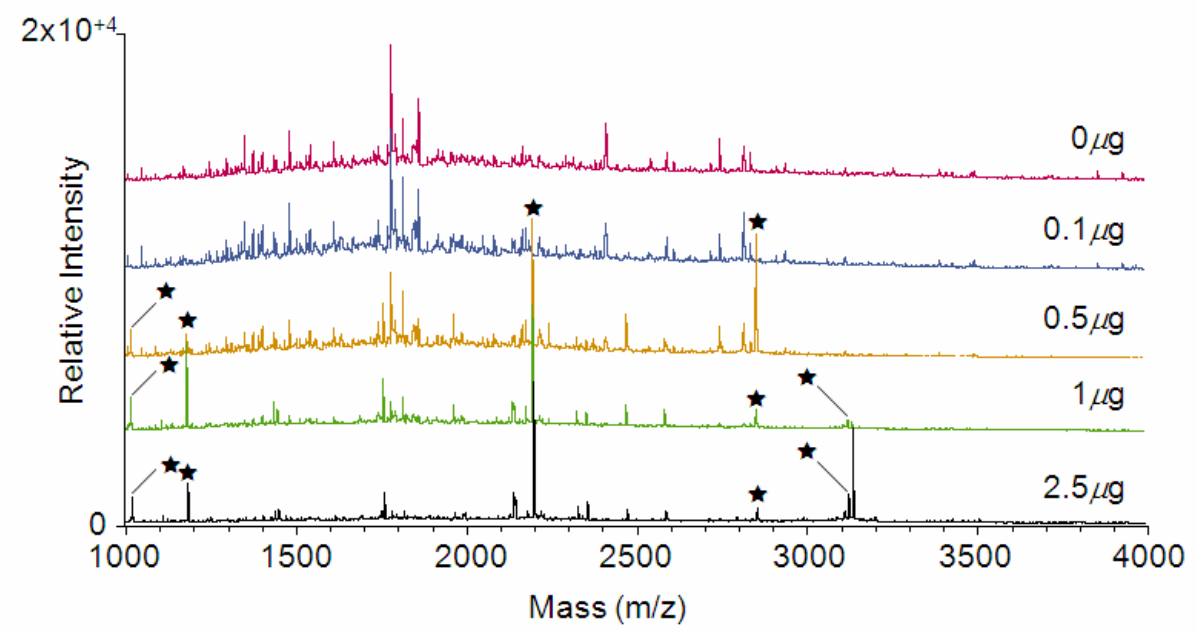


Figure VII. 2. MALDI spectra of a liver tissue from *Mus musculus* spiked with carbonic anhydrase (0, 0.1, 0.5, 2.5µg) and digested with 40ng/mm^2 of trypsin in a humidified chamber for 3h at 37ºC. ★ Indicate the m/z peaks from carbonic anhydrase.

The changes observed in the signal intensities as a function of the amount of protein deposited can be easily followed comparing the ratios for m/z peaks 2198, from carbonic anhydrase, and tissue´s m/z peak 1759, which are 40%, 30% and 20% for 0.5 µg, 1 µg and 2.5 µg respectively. This experiment allowed us to conclude that with our conditions effective digestion was achieved not only for the standard protein but also for the tissue´s proteins. In addition, this experiment clearly showed the ion suppression effect in MALDI, because, as it is noted following all spectra from figureVII.2, as the amount of CA is increased so the intensity of the peak ions belonging to the tissue is decreased.

VII.5.2 Influence of the digestion time

Once it was verified that the working conditions were appropriated we devised a set of experiments to study the influence of the digestion time. Digestion time is a variable that must be carefully optimized for each type of tissue. Therefore we tested two different digestion times 60min and 180min. The results showed that to obtain a good quality digestion the tissue needs to be in contact with the enzyme for at least 180min (data not shown). This results is in agreement with previously published data, where digestion times varying among 2h and 4h have been reported [12,14].

VII.5.3 Direct on-target proteolytic labeling

Once it was verified that the working conditions were adequate, we devised a set of experiments to proof that the labeling of the peptides formed after protein cleavage could be used to distinguish them of other tissue components. From the many methods available to label peptides we have chosen ^{18}O due to a number of benefits that we have described elsewhere [18]. It is well known that ^{18}O labeling can be accomplished following two different ways. The fastest and simplest consists in to do the protein digestion and the labeling at the same time [18]. This method lacks in labeling efficiency. The second way is done in two steps. First the proteins are digested, and then are labelled. Although it allows a labeling efficiency for most peptides higher than 90%, it is a labour intensive and time consuming approach [18]. Due to simplicity we chose to develop the labeling with the direct method. Therefore, we deposited spots of CA through the slides typically used for mass spectrometry of tissues, and they were submitted to trypsin digestion as follows: (i) with the enzyme dissolved in ^{16}O water, (ii) dissolved in a 1:1 $^{16}O/^{18}O$ water mixture and (iii) dissolved in pure ^{18}O water. The results shown in figure VII.3 demonstrate that the protein cleavage was successfully achieved, yet the peptide labeling was not obtained. As an example of the clearly resolved molecular species in the digest spectra, it is shown in figure VII.3, the isotopic distribution of m/z peaks 1013 and 1018, respectively; those indicating that digestion were well done in the three cases studied. However, it can be also seen that the isotopic distribution has not been altered with the presence of ^{18}O, therefore, and after careful inspection of all the spectra, it was concluded that under the conditions we used the labeling was not achieved.

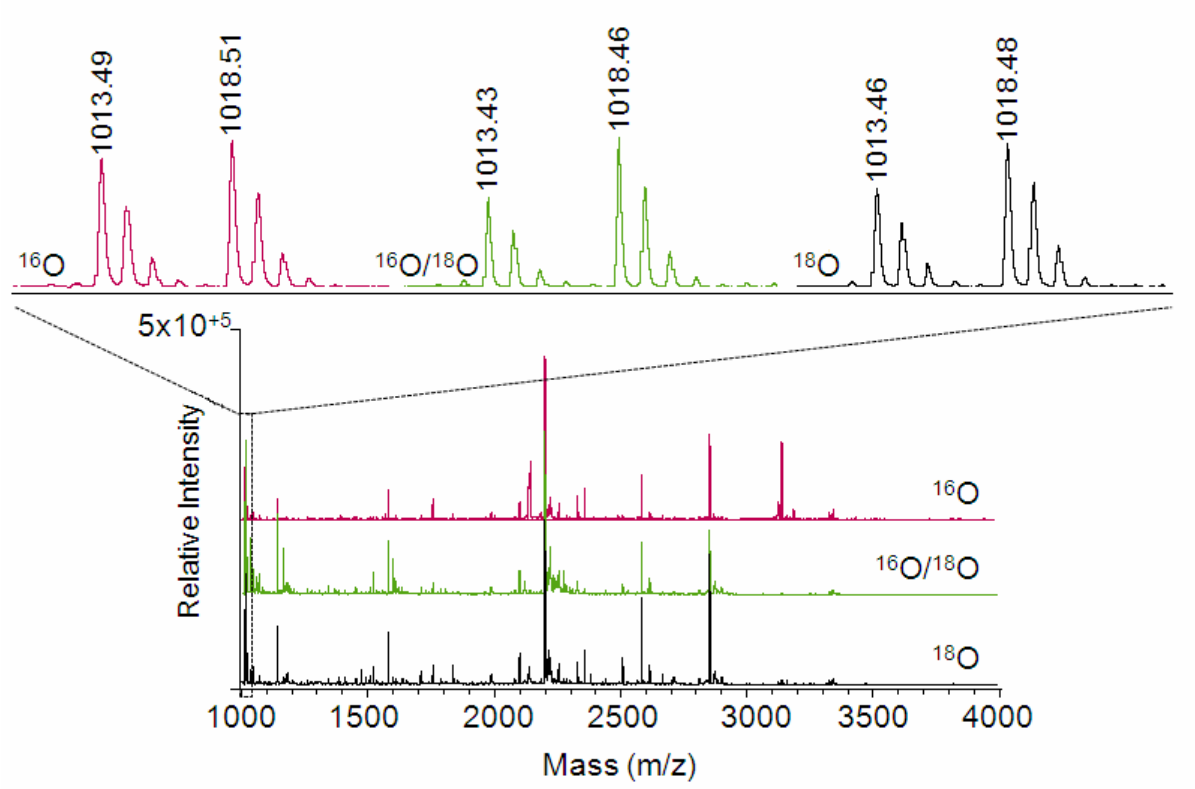


Figure VII. 3. MALDI spectra of 0.5µg of carbonic anhydrase digested over ITO-coated glass slide (trypsin 40ng/mm^2) in $H_2{}^{16}O$, 1:1 $H_2{}^{16}O$/ $H_2{}^{18}O$ and $H_2{}^{18}O$.

We hypothesise that the ^{16}O water used to maintain the constant conditions of humidity inside the chamber, where the digestion was done, was interchanged with the 18-O of the enzyme solution during the digestion time (3h) and so this could be the cause of the lack in labeling efficiency. Another explanation to the lack in labeling could be the long time contact between the trypsin and the ^{18}O water, because trypsin has been described as the primary mechanism by which the isotopic label is lost [19].

VII.5.4 Decoupled on-target labeling

Due to the lack in labeling using the direct approach, the next set of experiments was conducted doing the protein cleavage and the ^{18}O labeling in different steps. This approach has previously been successfully described in literature [18-20] for the labeling of proteins separated in-gel or through off-gel approaches [21,22]. Therefore the sample treatment was divided in two parts. The first part was focused in the enzymatic digestion of proteins whilst the second one was addressed to study the ^{18}O labeling of the peptides. This

set of experiments was done on an array of CA spots prepared in a glass slide. Since we have obtained, as explained above, good results in the digestion of tissues´ proteins, this step was done with exactly the same digestion conditions than in the direct approach. Once the digestion was considered completed, a solution of ^{18}O water was then pipette spotted over the CA spots as described in the experimental section. The results of this approach may be seen in figure VII.4, where the peptides from CA with m/z peaks 1013, 1018 and 1198 are presented.

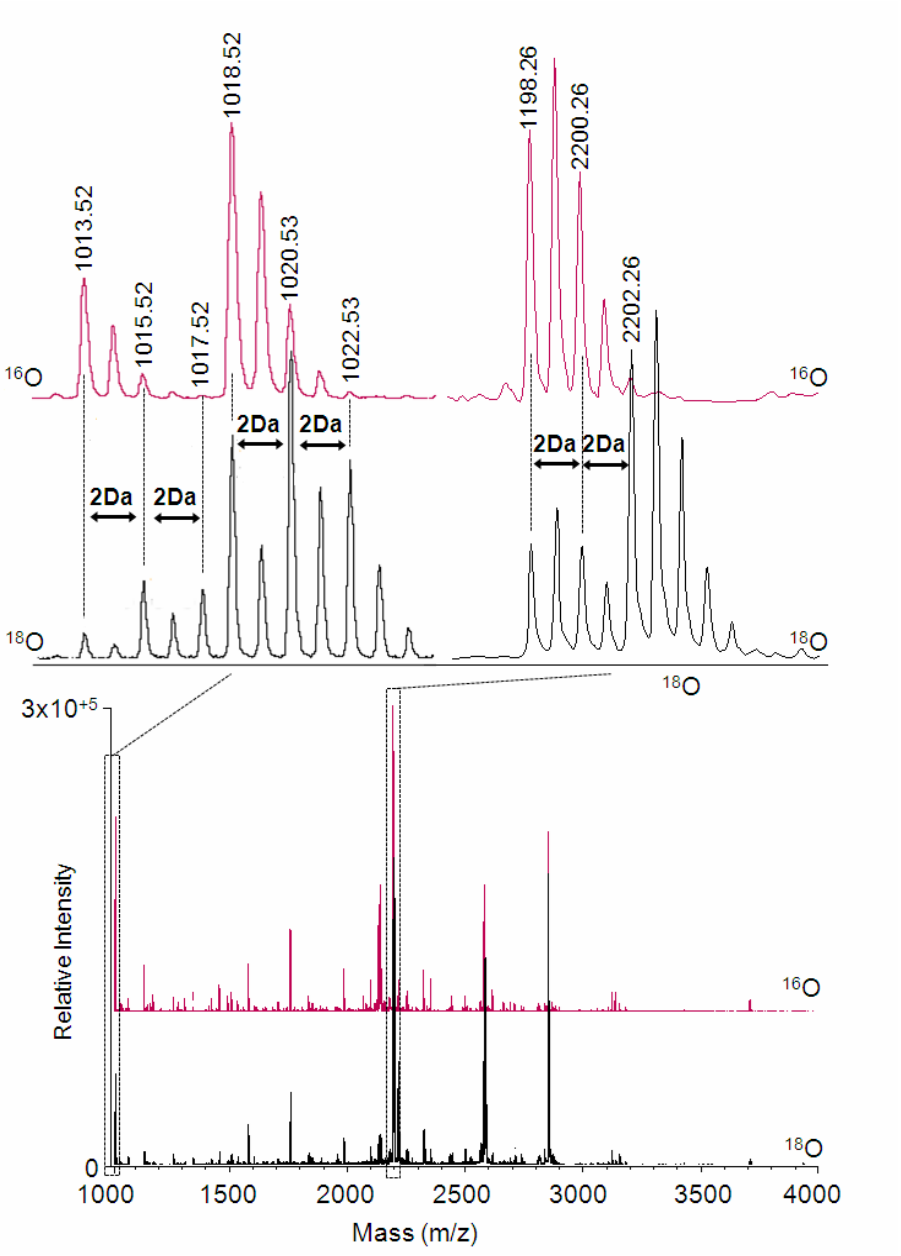


Figure VII. 4. MALDI spectra of 0.5µg of carbonic anhydrase digested over ITO-coated glass slide (trypsin 40ng/mm^2) followed by 20min of labeling with $H_2{}^{16}O$/trypsin 125ng/µL/ammonium acetate 50mM and $H_2{}^{18}O$/ trypsin 125ng/µL/ ammonium acetate 50mM.

The characteristic mixture pattern of single ($^{18}O_1$) and double ($^{18}O_2$) labelled peptides were observed when ^{18}O water was used. For peptide with m/z peak circa 1013, the main peaks corresponding to a single label (m/z *circa* 1015) and double label (m/z *circa* 1017) have almost the same intensity, thus indicating that the double labeling was not complete. Yet this result is useful if we consider that the identification of a peptide will be done easiest if an isotopic distribution as the one here reported is obtained. This is because in terms of identification a complete labeling will not help to distinguish peptides from other tissue´ components.

VII.5.5 The influence of trypsin concentration in the labeling efficiency

Many researchers have claimed that the presence of trypsin even in trace amounts can negatively affect the double labeling efficiency. The presence of any remaining trypsin after the digestion step will accelerate the $^{16}O/^{18}O$ carboxyl oxygen exchange. To avoid this problem some authors have suggested lowering the pH of the solutions below 2, since the activity of trypsin and other proteases at that pH is negligible [18,23,24]. By this way the problem of oxygen interchange is reduced to a minimum. However, other authors have claimed indeed that lowering pH, even below 2, is not enough to overcome this problem. Thus, Staes *et al.* [25,26] have found necessary to destroy the trypsin structure by covalent modification of its thiol group using trypsin inhibitors, to avoid any re-naturation at every step of the subsequent peptide isolation procedure. To complicate things further, Sevinsky *et al.* reported the failure of covalent trypsin inhibitors in their approach to quantitative proteomics making use of ^{18}O labeling [27,28]. In fact the aforementioned authors have proposed the use of immobilized trypsin in the digestion step as the best method to avoid the subsequent chemical exchange, due to residual trypsin. Therefore to investigate the effects of trypsin in the labeling of peptides, a set of experiments was done in which ^{18}O water-based solution of trypsin was spotted on arrays of CA prepared in glass slide. Results showed in figure VII.5 demonstrate that trypsin has no influence in the labeling step, at least for the amounts assayed in this work. This fact is highlighted trough the m/z peak 1013 and 1018 included in figure VII.5, where is clearly seen that no influence is observed neither in the ratios of the peaks corresponding to isotopic distributions nor in their respective intensities. Therefore for further experiments no trypsin was used during the labeling.

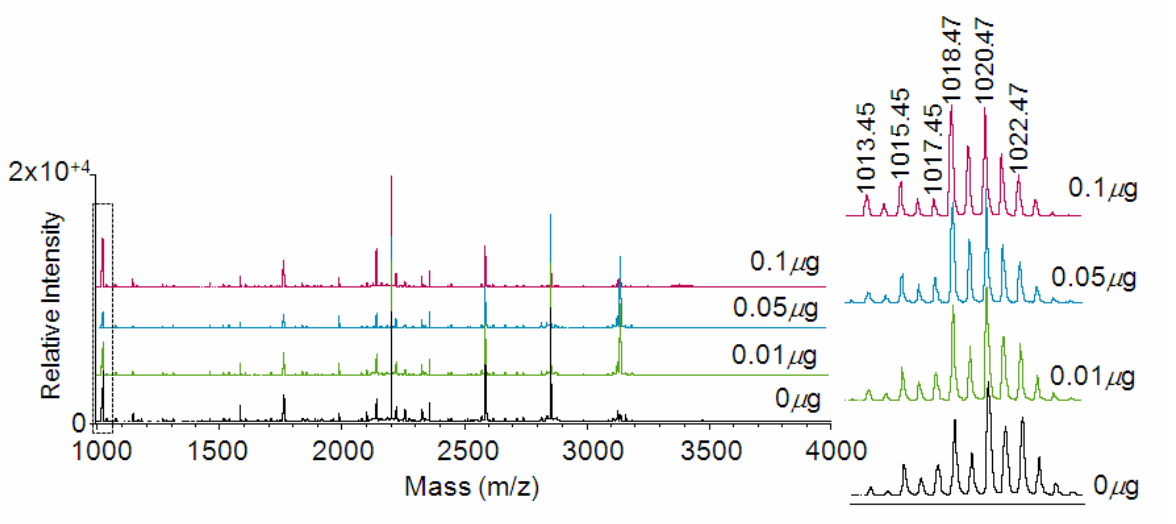


Figure VII. 5. MALDI spectra of 0.5µg of carbonic anhydrase digested over ITO-coated glass slide (trypsin 40ng/mm^2) followed by 20min of labeling with $H_2{}^{18}O$/ ammonium acetate 50mM with 0.1µg, 0.05µg, 0.01µg and 0µg of trypsin.

VII.5.6 On-tissue digestion and post-proteolytic labeling

The next step was to apply the treatment to several tissue sections that were first digested following the above described process, and then treated with (i) ^{16}O water, (ii) a mixture of $^{16}O/^{18}O$ water and (iii) pure ^{18}O water. The results are presented in figure VII.6.

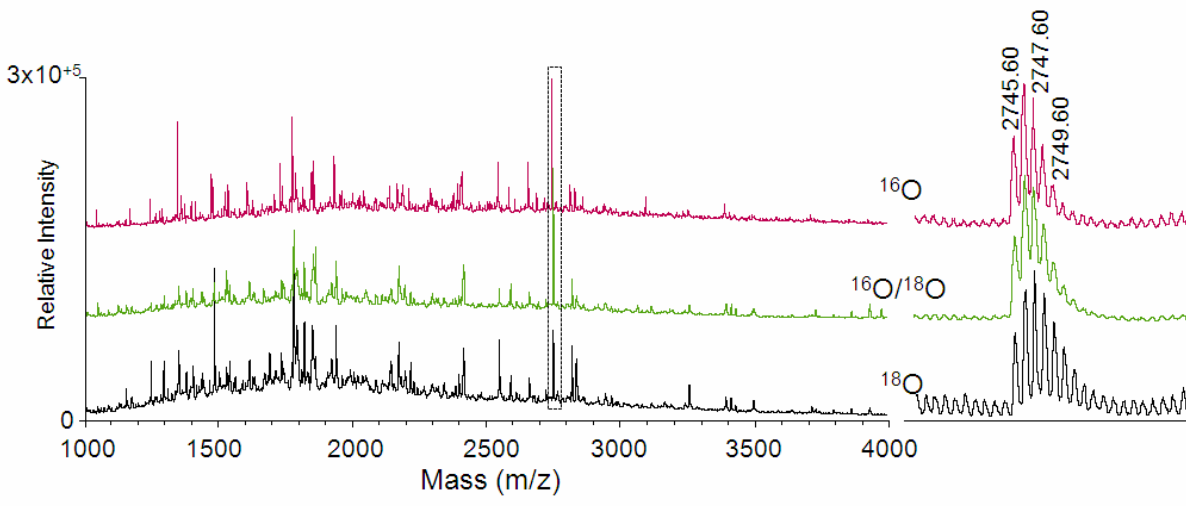


Figure VII. 6. MALDI spectra of a liver tissue from *Mus musculus* digested with 40ng/mm^2 of trypsin in a humidified chamber for 3h at 37ºC followed by 20min of labeling at room temperature with $H_2{}^{16}O$/ 50mM ammonium acetate, $H_2{}^{16}O/H_2{}^{18}O$/ ammonium acetate 50mM and $H_2{}^{18}O$/ 50mM ammonium acetate.

As may be seen the patterns are difficult to distinguish to naked eye, as they seem the same. However, and thanks to the labeling, a closest view of the single peaks obtained allows to clearly differentiating the peaks corresponding to peptides. As an example the peptide with m/z peak 2745 is also shown in figure VII.6. As for the case of the CA the typical pattern of single and double labeled peaks were obtained.

VII.5.7 The influence of time in the on-tissue post-digestion labeling

Time of labeling is an important factor that influences the efficiency in the single and double oxygen incorporation [18] and therefore must be carefully controlled to obtain reproducible results. In our case two times were studied 10min and 20min. figure VII.7 shows the results of this experiments and may be seen that for longer times the incorporation of ^{18}O (single and double) is higher. A time of 20 min was chosen for further experiments.

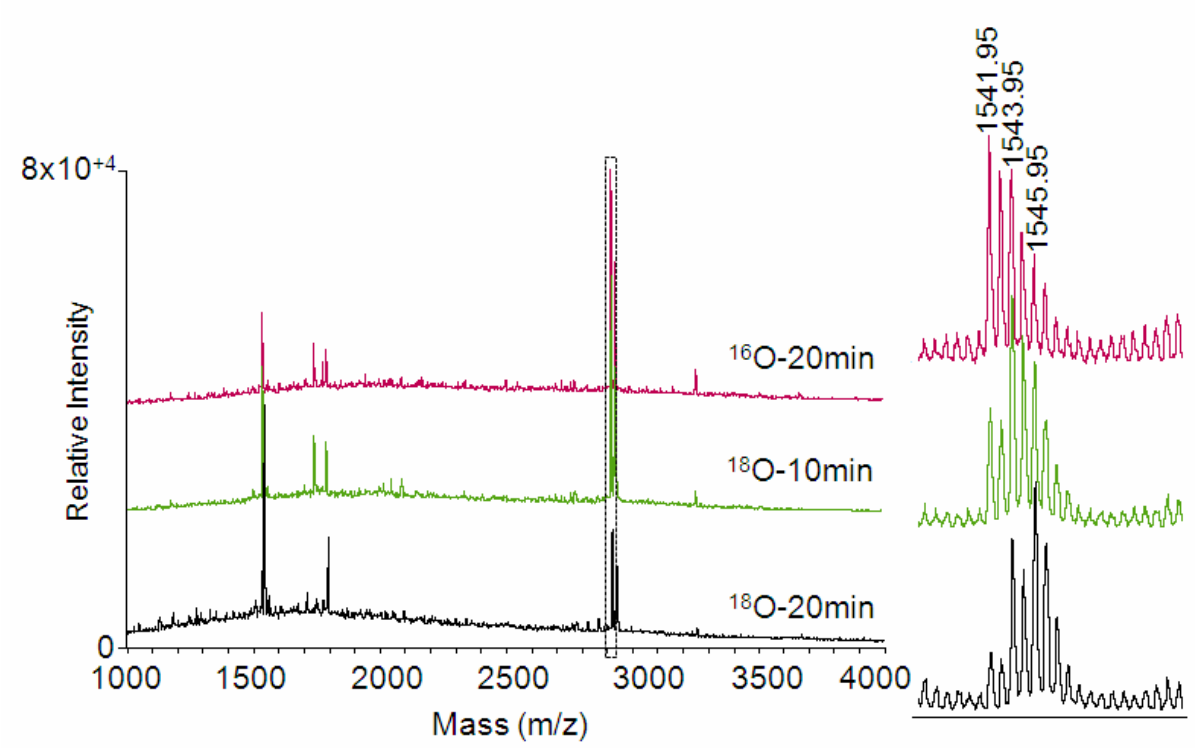


Figure VII. 7. MALDI spectra of a ovary tissue from *Mus musculus* digested with 40ng/mm^2 of trypsin in a humidified chamber for 3h at 37ºC followed by 20 min of labeling at room temperature with $H_2{}^{16}O$/ 50mM ammonium acetate, 10 min of labeling with $H_2{}^{18}O$/ ammonium acetate 50mM and 20min of labeling with $H_2{}^{18}O$/ 50mM ammonium acetate.

VII.5.8 Effect of ultrasonic energy on tissue digestion

It is well known that ultrasonic energy can accelerate the enzymatic digestion of proteins [18,24]. Furthermore, it also has been demonstrated that it can also speed the isotopic labeling of peptides in the direct method [21,26]. Therefore we tried to speed the tissue´ digestion with the aid of ultrasonic energy in a trial to overcome the bottleneck of low sample throughput in tissues´ handling. The first problem to be overcome was the technical design in order to apply ultrasonication over the tissue. After careful consideration we decide to use ultrasonic energy in an indirect way. Based on our expertise with ultrasonication for sample treatment [28], the sonoreactor was the device chosen. The arrangement used is shown in the video provided in the supplementary material. This arrangement was done to allow indirect ultrasonication over the tissue. The sonoreactor acts as a small yet powerful ultrasonic bath. The ultrasonic energy is transmitted from the sonoreactor´ base, where the ultrasonic transducer is situated, to the water, and from here to the face of the slide in contact with the water. The ultrasonic energy is then transmitted through the slide to the tissue situated in the opposite size of that one in contact with the water. A video showing the effects of ultrasonic energy in the degasification of some drops of coca-cola in the same arrangement is showed in supplementary material as a proof-of-the concept.

To ensure that water was not in contact with the tissue during the ultrasonication a set of experiments was done where the amplitude of the sonoreactor was varied among 10 and 100%. Amplitudes higher than 60% produced too much agitation as consequence of which the slide was cover completely with water. Therefore, ultrasonication amplitude of 50% was set in further experiments. Once the technical method was optimized to ensure that no water was in contact with the tissue, the digestion was done at different times ranging from 30s to 300s under the effects of an ultrasonic field. As may be seen in figure VII.8, excellent results in terms of digestion were obtained in just 30s whilst when no ultrasonication was used the cleavage was almost negligible, even for the longest time assayed (300s).

It was therefore concluded that ultrasonication indeed can boost the digestion process. This set of experiments clearly showed that ultrasonic energy can act promoting the contact between the solution containing the enzyme and the tissue. This effect is probability due to the vibration caused by the ultrasonic waves crossing the glass slide and finally reaching the tissue and the solution containing the enzyme. To this effect it must be

added the cavitation phenomena caused by the ultrasonic wave in the solid-liquid interface. The tiny bubbles created by the ultrasonic energy act as micro-reactors with cumulative effects such as increasing chemical reaction rates, increasing bulk temperature and increasing the surface of contact available to the enzyme by disrupting the tissue surface. Although this was a remarkably finding, its applicability needs to be further investigated. As an example, although ultrasonication can be used to boost the digestion of tissues for biomarker discovery, the viability of this method for imaging, may become useless because ultrasonication will probably lack in maintain spatial distribution.

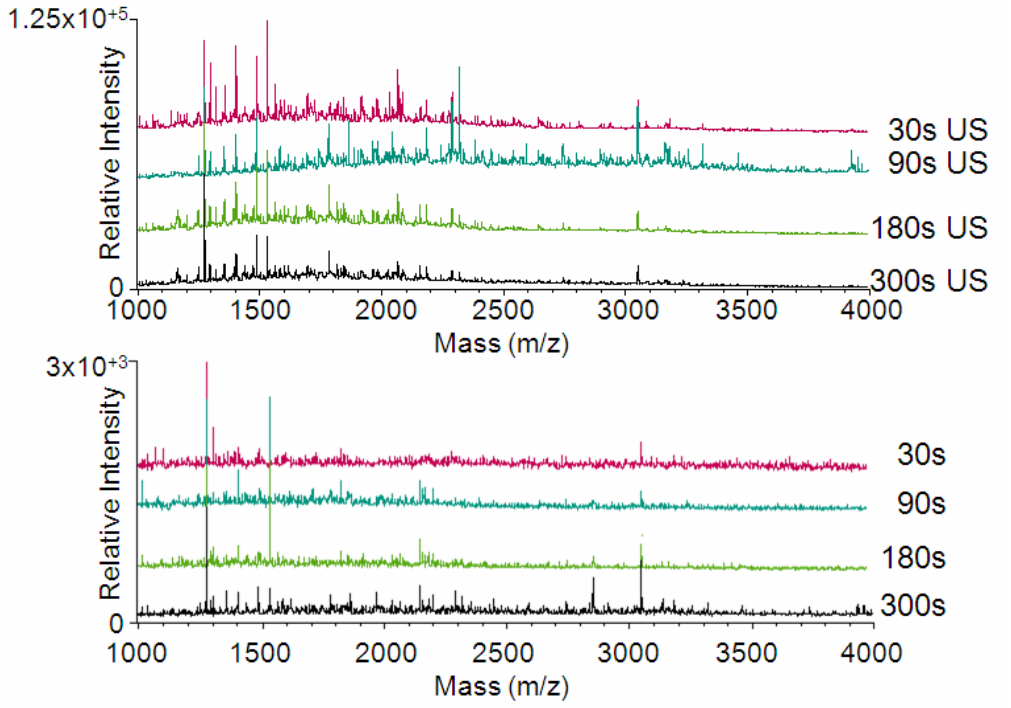


Figure VII. 8. MALDI spectra of a liver tissue from *Mus musculus* digested with 40ng/mm^2 of trypsin in sonoreactor with 50% of Sonication amplitude for 30s, 90s, 180s, 300s and for comparative purposes the same experiments were carried out without ultrasonication.

VII.6 Conclusions

It has been demonstrated for first time the possibility to perform the 18-O labeling of peptides directly on a tissue. The labeling can be done decoupling the steps of protein digestion and peptide labeling. It was found that no complete labeling was possible, which can facilitate the automation (software) recognition of peptides from other tissue components. In addition it has been also demonstrated that ultrasonication can boost the

digestion of tissue´s protein in seconds, thus opening for the first time sample treatment of tissues to high throughput.

VII.7 Acknowledgements

H. M. Santos acknowledges the doctoral grant SRFH/BD/38509/2007 provided by FCT (Science and Technological Foundation) of Portugal. Dr. José-Luis Capelo-Martínez is grateful to the Xunta de Galicia (Spain) for the program Isidro Parga Pondal. Xunta de Galicia, Spain, is acknowledge by the Isidro Parga Pondal Program (J. L. Capelo) and for financial support given under project 09CSA043383PR. The research findings here reported are protected under patent pending PCT/IB2006/052314 and PT 103 303.

VII.8 References

[1] Caprioli, R. M., Farmer, T. B., Gile, J. *Anal. Chem.* 1997 (69) 4751–4760.

[2] Chaurand, P., Stoeckli, M., Caprioli, R. M. *Anal. Chem.* 1999 (*71*) 5263–5270.

[3] Stoeckli, M., Chaurand, P., Hallahan, D. E., Caprioli, R. M. *Nat. Med.* 2001 (7) 493–496.

[4] Chaurand, P., Schwartz, S. A., Caprioli, R. M. *Curr. Opin. Chem. Biol.* 2002 (6) 676–681.

[5] Chaurand, P., Caprioli, R. M. *Electrophoresis* 2002, (23) 3125–3135

[6] Xu, B. J., Caprioli, R. M., Sanders, M. E., Jensen, R. A. *J. Am. Soc. Mass Spectrom.* 2002 (13) 1292–1297.

[7] Chaurand, P., Schwartz, S. A., Billheimer, D., Xu, B.J., Crecelius, A., Caprioli, R.M., *Anal. Chem.* 2004 (76) 1145–1155.

[8] Aerni, H. R., Cornett, D. S., Caprioli, R. M. *Anal. Chem.* 2006 (78) 827–834.

[9] Guo, J. Z., Colgan, T. J., DeSouza, L. V., Rodrigues, M. J. *et al. Rapid Commun. Mass Spectrom.* 2005 (19) 2762–2766.

[10] Richard J. A. Goodwin, Stephen R. Pennington and Andrew R. Pitt. *Proteomics* 2008 (8) 3785–3800.

[11] Petyuk. V.A., Qian, W.J. Chin, M.H., Wang, H.X. *et al. Genome Res.* 2007 (17) 328–336.

[12] Djidja, M.C., Francese, S., Loadman, P.M., Sutton, C.W. *et al. Proteomics* 2009 (9) 2750–2763.

[13] Lemaire, R., Desmons, A., Tabet, J.C., Day, R. *et al. J. Proteome Res.* 2007 (6) 1295–1305.

[14] Groseclose, M.R., Andersson, M., Hardesty, W.M., Caprioli, R.M. *J. Mass Spectrom.*2007 (42) 254–262.

[15] Altelaar, A.F.M., Luxembourg, S.L., McDonnell, L.A., Piersma, S.R. *et al. Nat. Protoc.* 2007 (2) 1185–1196.

[16] R. Lemaire, M. Wisztorski, A. Desmons, J.C. Tabet, R. *et al. Anal. Chem.* 2006 (78) 7145–7153.

[17] Stauber, J., Lemaire, R., Franck, J., Bonnel, D., *et al. J. Proteome Res.* 7 (2008) 969–978.

[18] Capelo, J.L., Carreira, R.J., Fernandes, L., Lodeiro, C. *et al. Talanta*, 2010 (80) 1476–1486.

[19] Angel, P. M., Orlando, R. *Anal. Biochem,* 2006 (359) 26–34

[20] Yao, X. D., Afonso, C., Fenselau, C. *J. Proteome Res.* 2003 (2) 147–152.

[21] Carreira, R.J., Lodeiro, C., Diniz, M.D., Moura, I. *Proteomics* 2009 (9) 4974–4977.

[22] Santos, H.M., Glez-Peña, D., Reboiro-Jato, M., Fdez-Riverola, F., *et al. Accepted in Electrophoresis*, 2010.

[23] Stewart, I.I., Thomson, T., Figeys, D., O-18 Labeling: a tool for proteomics. *Rapid Commun. Mass Spectrom*. 2001 (15) 2456–2465.

[24] Santos, H.M., Mota, C., Lodeiro, C., Moura, I. *Talanta* 2008 (77) 870–875.

[25] Staes, A., Demol, H., Van Damme, J., Martens, L. *et al. J. Proteome Res*. 2004 (3) 786–791.

[26] Carreira, R.J., Rial-Otero, R., López-Ferrer, D., Lodeiro, C. *et al. Talanta* 2008 (76) 400–406

[27] Sevinsky, J.R., Brown, K.J., Cargile, B.J., Bundy, J.L. *et al. Anal. Chem.* 2007 (79) 2158–2162.

[28] Santos, H.M., Capelo, J.L. *Talanta* 2007 (73) 795–802

Chapter VIII.

Conclusions and future prospects

The undertaken research presented in this dissertation involves, in its totality, some alternative approaches to the typical workflows for the identification and quantification of proteins. The link between the different chapters of this thesis, the ultrasonic energy applied to proteomics.

Ultrasonic energy was used in some of the different steps of the regular in-solution and in-gel protein digestion protocols used to indentified proteins by PMF of PFF. In addition to the new approaches for protein identification, a novel workflow based on ^{18}O-isotopic labeling was also developed for protein quantification. Furthermore dedicated software to deal with this type of data sets was also created. Finally, a new method for the fast in-tissue ^{18}O-isotopic labeling of peptides was created. As a general role the new sample treatment protocols are not only faster than the ones they come to replace but they are also less labor intensive.

Below are going to be presented in more detail the main goals accomplished during the making of this research work.

Chapter III. It has been demonstrated that ultrasonic probe and sonoreactor can be used to accelerate the sample preparation for protein identification by PMF using MALDI-TOF-MS. The workflow including protein reduction, alkylation and tryptic digestion was reduced from 24h to 15min, without compromising the number of peptides matched or the protein sequence coverage obtained. Each step, reduction and alkylation, usually done in 60min and 45min were reduced to 5min for reduction plus 5 for alkylation. Concerning tryptic digestion the classic overnight incubation at 37ºC was reduced to bust 5min under the effects of ultrasonic field applied by an ultrasonic probe or the sonoreactor.

The new proposed method represents and good alternative to the classic method, especially when a fast protein screening is needed. In addition, with the new proposed methodology, sample handling has been enormously simplified. Proteins in the mass range 16.7-71.2kDa including α-lactalbumin, ovalbumin, aldolase, BSA, catalase, carbonic anhydrase, and chymotrypsinogen A were successfully identified with the new accelerated method using the sonoreactor device. In addition the zinc resistance-associated protein precursor isolated from *D. desulfuricans* strain G20 and split-soret cytochrome *c* purified from *D. desulfuricans* ATCC27774 demonstrates that the parameters optimized on standard can be also applied to complex biologic samples. The new ultrasonic-based method provides itself, important advances in fast protein identification for protein screening. The

findings reported open a new way in sample treatment for in-solution protein digestion, of easy implementation, for on-line procedures, and tandem mass spectrometry.

Chapter IV. Following the previously described research and having in mind the goal of automation, a new ultrasonic multiprobe-device was studied in conjunction with a 96-well plate in the acceleration of two different proteomic workflows. The method was evaluated in terms of speed, throughput, handling and robustness.

It was demonstrated that, in order to avoid cross contamination between samples, low amplitudes should be used to avoid aerosol formation and sample spreading. For the two methods using ultrasonic multiprobe and a 96-well plate, the accelerated urea method and the accelerated clean method, all the standards proteins were identified and the results were found similar in terms of robustness in comparison with the classic overnight method.

The clean method has the best performance in terms of speed and handling since only 2min/sample are required to complete it. Since the clean method do not use urea, desalting is not required thus diminishing the total number of steps.

Regarding throughput, it has been proven that the combination of a 96-well plate and an ultrasonic multiprobe is a potential powerful tool in sample treatment for proteomics, allowing high sample throughput. Sample preparation steps, including reduction and alkylation, digestion, spotting on MALDI targets or transfer to LC/MS input plates can potentially be combined on a single automated platform making use of ultrasonic energy provided by ultrasonic multiprobes.

Chapter V. A new method for bottom-up protein quantification termed decision peptide driven, DPD, has been established and experimentally validated. The proposed method which combines SDS-PAGE separation of proteins, in-gel proteolytic digestion, ^{18}O peptide labeling and the inverse labeling strategy, has been demonstrated successfully in the fast and accurate quantification of seven different proteins. Furthermore, the DPD approach has been applied with success to study the effects of high concentration levels of estrogens on *Ciprinus carpio* males. To do so, protein vitellogenin, a biomarker of exposure to high levels of estrogens, was accurately quantified using the DPD method. The DPD method onllyneeds the identification of peptides with paralleled losses and similar digestion yields through different in gel-digestion sets. Using these peptides accurate quantification of the protein is achieved. Suitable software has been also developed to automate the task.

Chapter VII. A friendly software to help in an automated mode to identify those peptides that have paralleled loses through a typical proteomic workflow was described. The use of such peptides allows robust and accurate quantification of proteins using 1D-gel electrophoresis and MALDI-TOF-MS. The software presented, allows for the identification of reproducible peptides and is based in a series of steps entailing different algorithms that perform in an automated mode a peptide differential analysis to extract and to identify those peptides that systematically remains constant in expression level through different sets of a typical in-gel digestion workflow. The DPD software saves times, allowing the user to accurate quantify proteins in an automated mode, overcoming the long time needed when the treatment of data is done manually. In addition the DPD software has a wizard easy to follow for its installation.

Chapter VII. It has been demonstrated for the first time the possibility to perform the ^{18}O-labeling of peptides directly on tissue sections. The labeling can be done decoupling the steps of protein digestion and peptide labeling. It was found that no complete labeling was achieved, which can facilitate the automation (software) recognition of peptides from other tissue components. In addition it has been also demonstrated that ultrasonication can boost the digestion of tissue´s protein in seconds, thus open for the first time sample treatment of tissues to high throughput.

Future prospects

Regarding the use of ultrasonic energy to enhance the cleavage of proteins it is expected soon the development of on-line applications in shot-gun proteomics. In addition, the application of this methodology to robotic platforms is already a reality in the PropecII from Digilab and it is expected to be implementing in platforms from other companies. It remains unclear the mechanism by which UE can boost the enzymatic digestion of proteins. Therefore research in this point is also expected. Furthermore, the comparison of ultrasonication with other tools to speed protein cleavage, such as microwave energy, infrared irradiation or high pressure will surely soon appear in literature.

As far as the protein quantification concerns, the methodology developed in this thesis is easily adaptable to electrospray ionization and therefore this approach will be soon developed, including also new software. Furthermore, off-gel approaches accomplished

with shotgun proteomics will also take advantage of our quantification approach, opening the possibility to quantify hundreds of proteins from complex mixtures at the same time.