Università degli Studi di Firenze

DOTTORATO DI RICERCA IN

CHIMICA E TECNOLOGIA DEL FARMACO

CICLO XXV

COORDINATRICE Prof.ssa Elisabetta Teodori

Proteomic investigations on autoimmune diseases

Settore Scientifico Disciplinare CHIM/08

Dottoranda

Dott. Dioni Ilaria

Tutor

Prof. Rovero Paolo

Anni 2010/2012

Determination is often
the first chapter
in the book of excellence

TABLE OF CONTENTS

ABSTRACT

Defined as the protein complement of the genome, the proteome is a varied and dynamic repertoire of molecules that in many ways dictates the functional form that is taken by the genome. The importance of proteomics is a direct consequence of the central role that proteins play in establishing the biological phenotype of organisms in healthy and diseased states. As proteins are the principal targets of drug discovery, the evolution of proteomics techniques is increasingly relevant in the field of medicinal chemistry. Proteomics is a promising approach for the comprehension of fundamental cell biology, the elucidation of biochemical pathways involved in disease mechanisms, and the identification of new protein targets. Accordingly, this technique is gaining widespread use in each phase of drug development, including target identification and validation, investigations into mechanisms of drug action or toxicity, biomarker discovery and also toxicological profiling in pre-clinical and clinical settings.

Proteomics concretely underpin the research and the development of both therapeutic agents and diagnostic tools. In fact, the achievement of an early diagnosis, especially in the field of autoimmune diseases, is a mandatory task of medicinal chemistry, equally relevant as the development of new drugs.

Humoral responses represent a well-established component of several autoimmune pathologies where circulating autoantibodies address self-structures, frequently provoking severe damages. Some examples of pathogenic autoantibodies are represented by anti-myelin antibodies, contributing to the corrosion of the myelin sheath of multiple sclerosis patients, anti-nuclear antibodies, involved in the glomerulonephritis manifestations of systemic lupus erythematosus, and autoantibodies directed to insulin-secreting beta cells of the pancreas in type 1 diabetes.

The specific antigenic targets addressed by these autoantibodies are in most cases elusive and the underlying etiopathogenetic mechanisms remain largely unknown. Thus, unraveling the native antigens responsible for the production of autoantibodies endowed with diagnostic or prognostic value might be of invaluable interest, not only for a better comprehension of the disease mechanism(s), but also for the design and development of effective synthetic antigenic probes to be used to detect autoantibodies as diagnostic tools. By monitoring the presence of the autoantibodies, these probes can be used for an early detection of the disease, as well as for its follow-up.

In this context proteomics provides a technology platform able to address these issues from several fronts. Indeed beyond the identification of proteins involved in disease progression, it is often critical to explore protein post-translational modifications (PTMs). Several studies have highlighted the importance of identifying potential

modifications as these may be indicative of the disease state and may lead to the development of biomarker assays.
Proteomics is, by definition, an interdisciplinary science and derivatization reactions are frequently translated from chemistry to proteomic experiments as tagging reactions for proteins and peptides in the context of complex mixture of biological origin. The fields of application span from the enrichment and characterization of PTMs to the modification of functional groups to improve the mass spectrometric detection or to the introduction of labels for quantitative proteomic analysis. As far as quantitative proteomics is concerned, studying the change in the abundance of proteins within a cell over time or between different cellular states (normal versus diseased states) helps in providing further insights into the pathophysiological basis of protein target identification and validation for disease intervention and treatment.
On the whole, proteomics rises as an appropriate tool for a deeper insight into systems, be it cells or organisms, and beyond the basic research, the whole field is driven by a vision of fundamental amelioration within medicinal chemistry.

In this scenario, the first chapter of this work of thesis describes a project concerning the research of possible autoantigens recognized by anti-citrullinated proteins antibodies (ACPAs), biomarkers of rheumatoid arthritis (RA). The attention has been focused on histones, nuclear proteins whose presence has also been demonstrated in the extracellular environment. In particular citrullinated histones have been recognized as main actors in the production of neutrophil extracellular traps (NETs) by neutrophils. It has been described that NETs are formed in response to infective and inflammatory stimuli that bring to the extrusion of proteins with antimicrobial activity. However when NETs clearance is not correctly regulated, these structures may persist in the extracellular environment, leading to the presentation of antigens to the immune system and possibly supporting autoimmune responses toward self-antigens.
The target of this research was to confirm the reactivity of citrullinated histones from nuclear extracts and NETs toward RA patients' sera. A proteomic approach has been followed, based on the employment of gel separations, immunodetection, mass spectrometric analysis of trypsin digested bands and subsequent database searches.
ACPAs are a hallmark of RA patients and their specificity to citrullinated antigens has been clearly proved. Citrulline residues are the result of an enzymatic deimination of arginine residues and feature a slight mass increase of 0.984 Da for residue. The characterization of PTMs is a primary, very challenging task in the proteomic research; accordingly, only few approaches have been described for the detection of deimination.
In this work, a chemical derivatization strategy has been initially employed, using 2,3 butanedione and antipyrine that selectively tag the ureic group of citrulline residues

introducing a 238.11 Da mass increase. MALDI-TOF analyses were performed on the trypsin digest to highlight the presence of derivatized citrullinated peptides.
Subsequently, a LC-MS/MS strategy has been approached, aimed to more extensively characterize the deimination pattern of the identified protein by means of a direct detection of citrullines, testing two different stationary phases, namely reversed-phase and HILIC. LC-MS/MS data were acquired using a *nano*HPLC system coupled to a LTQ-Orbitrap mass spectrometer equipped with a *nano*ESI source.

The second chapter reports a research conducted on a reaction of stable isotope labeling, aiming at the relative quantification of the differential expression of proteins in two or more samples representing various conditions of biological systems.
In this project a reaction has been taken into account, based on a reductive amination with aldehydes and $NaCNBH_3$ and already reported as labeling strategy for quantitative experiments. It has been challenged considering that, beside primary amines which are the specific target of the reaction, the reactivity of aldehydes to other nucleophilic sites may impair the reliability of the reaction for quantitative purposes.
Side products have been observed using formaldehyde or acetaldehyde and several tests were performed using synthetic peptides to both elucidate the chemical nature of the rearrangements and to find the conditions to suppress their formation. Among different experimental conditions, a specific attention was set aside for the pH of the reaction environment in order to highlight its role in the progression of the reaction. Upon fixation of general rules for the use of these reagents in stable isotope peptides labeling, the occurrence of the side-products was monitored on a single protein digest and a mixture of proteins digest.
LC-MS/MS experiments were performed and quantitative reports were derived using a quantitation software able to extrapolate protein light/heavy ratios by processing the intensities of differentially labeled peptide precursors.

The third chapter reports a project which has focused on the research of native autoantigens recognized by specific autoantibodies detected in Multiple Sclerosis (MuS) patients' sera. Differently from the research of citrullinated histones recognized by RA patients' ACPAs, in this project the research of candidate antigenic proteins was guided by autoantibodies recognized by a specific peptide probe called CSF114(Glc).
In a previous work from this laboratory it was shown that CSF114(Glc) specifically identifies serum autoantibodies in a subset of MuS patients, representing approximately the 30% of the patient population. A "chemical reverse approach" allowed the identification of the N-glucosylated peptide CSF114(Glc) as the most efficient antigenic probe in detecting autoantibodies in relapsing remitting MuS patients.

Since serum autoantibodies ought to reflect the presence of corresponding antigens, peptides recognized by these autoantibodies might mimic structural features of neo-epitopes. Given this background CSF114(Glc) affinity-purified antibodies from MuS patients' sera were used to back-track CNS antigens implicated in the MuS-related autoimmune response.
Classical proteomic tools were employed starting from gel separation of rat brain proteins followed by western blot with anti-CSF114(Glc) antibodies purified from MuS patients' sera by means of an immunoaffinity chromatography. Mass spectrometric analyses of immunoreactive bands were performed and a validation of the identified proteins was fulfilled in order to confirm their reactivity.

In conclusion, the three chapters of this thesis highlight the versatility of proteomics within the biomedical research. The use of different, sophisticated techniques contributed to the fine characterization of elusive self-antigens, recognized by autoantibodies of diagnostic/prognostic relevance in two different autoimmune disease, i.e. AR and MuS. Moreover the relative quantification of the differential expression of proteins in various conditions represents a field of growing interest and a correct employment of labeling strategies was proved to be essential for providing reliable results.

CHAPTER 1

Antibodies from rheumatoid arthritis patients target deiminated H4 contained in NET

1.1 INTRODUCTION

1.1.1 Rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic inflammatory disorder with autoimmune features that primarily affects the small diarthrodial joints of hands and feet, lately involves the large joints and finally spreads systemically.[1] It is considered the most common autoimmune disease with an incidence range of 0,5-1,0% in the world's population; it affects women more frequently than men and preferentially occurs between the ages of 25 to 50.[2]

Despite several factors are thought to be involved in the etiology of RA, the underlying mechanisms had not still been pointed out; however genetic predisposition, environmental and autoimmune factors certainly represent a triad with a central role in the pathogenesis of the disease.[3]

The primary target of RA is the synovium that in physiological conditions is a relatively acellular site while in RA patients is interested by a massive recruitment of immuno competent cells. In fact T $CD4^+$ lymphocytes, activated B-cells and macrophages consistently infiltrate and, in the worst way, organize in lymphoid aggregates with germinal centers. At the same time an extensive network of pro-inflammatory autocrine and paracrine cytokines (IL-1. IL-6, IL-15, IL-18, TNF-α, GM-CSF) creates a persistent inflammatory status. In this situation resident synovial cells are activated to produce hydrolytic degradative enzymes such as metalloproteinases, serine proteases, collagenases which altogether start the digestion of the extracellular matrix and the formation of the synovial *pannus*. This process triggers a progressive and aggressive erosion of the articular structures (both cartilages and bones) that leads patients to suffer a substantial loss of mobility because of the joints destruction and the correlated pain.

Disease progression can be, generically, divided into three phases: disease initiation at peripheral lymphoid organs that starts the migration of activated immuno cells to the joint, disease propagation which leads the release of cytokines and tissue damages mediated by resident cells (Figure 1.1).

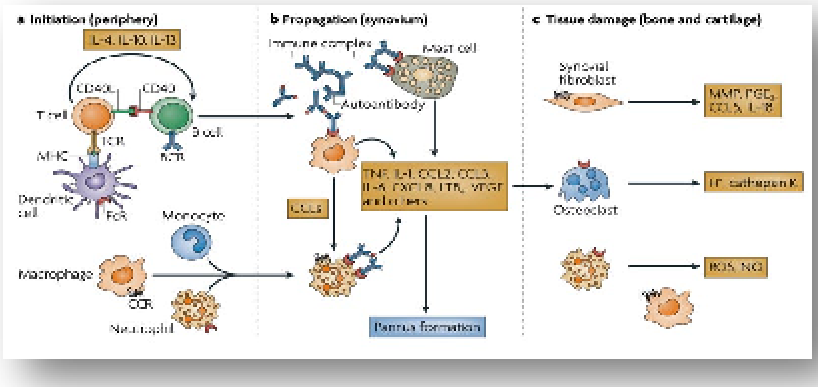


Figure 1.1. Stepwise progression in the development of RA: initiation, propagation and tissue damage

The development of RA reminds the progression of other autoimmune diseases for the presence of an asymptomatic onset followed by a time-variable subclinical phase that lately evolves into a claimed pathology. During the pre-clinical phase there are inflammatory changes in the synovium without any clinical sign or symptom; joints inflammation dominates the initial clinical phase and erosion phenomena joined to systemic manifestations characterize the chronic phase.

The very beginning symptoms of RA are often undifferentiable from those of other arthritis diseases such as systemic lupus erythematosus (SLE), psoriatic arthritis or seronegative spondylarthropathies, making difficult its discrimination. Morning stiffness, reduced grip strength, symmetric arthritis are only few examples of the most common symptoms. Their entity largely depends on the presence and the degree of phlogosis so when body tissues are inflamed the disease is active while in absence of inflammation the disease is inactive and in remission. In this phase, which can lasts for weeks, months or years, symptoms disappear and patients feel better; as soon as the symptoms reappear the disease is active again.

Since RA etiopathogenesis is widely undiscovered the therapeutic treatments are focused on controlling patient's signs and symptoms, preventing joints damages and maintaining patient's quality of life and ability to function. Three main classes of drugs are used in RA treatment: non-steroidal anti-inflammatory agents (NSAIDs), corticosteroids, and disease modifying anti-rheumatic drugs (DMARDs).

NSAIDs mitigate the pain and the inflammation but unfortunately they do not limit the long-term damaging effects of RA on the joints. They need to be taken continuously to achieve the effects but side events risk to emerge after a prolonged treatment. Corticosteroids are more effective and faster in reducing the main RA symptoms (pain, stiffness, joint swelling and tenderness) so they are frequently used to treat the severest disabilities caused by RA progression. Contraindications regarding the use of

steroids are well-known. DMARDs also work to decrease pain and inflammation by suppressing overactive body's immune system; they are not designed to provide immediate relief of symptoms because several weeks or months are needed to produce the effects. DMARDs include drugs such as methotrexate, hydroxychloroquine, sulfasalazine, and leflunomide but also biological agents like some recombinant antibodies (adalimumab, infliximab, rituximab as examples) that inhibit inflammatory cytokines or transiently deplete mature reactive B-cells.

It is well established that joint damages start during the asymptomatic phase; around the 90% of RA patients exhibit some forms of disability within two decades from the onset and early treatments decrease the rate of disease progression. An early diagnose is of the greatest importance to allow therapeutic agents to be given as soon as possible. Unfortunately it is still a difficult issue to address due to the long asymptomatic onset, the plurality of clinical aspects and the absence of clear specific RA symptoms.[4]

Up to now the diagnosis is performed by combining clinical symptoms with biological and radiographic indices which have been standardized by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) (Figure 1.2).

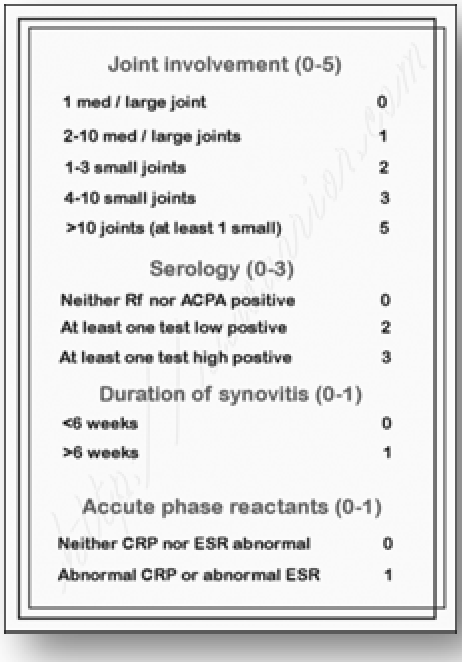


Figure 1.2. Classification criteria for RA according to the ACR/EULAR committee

These criteria have been developed starting from a previous classification criticized for its lack of sensitivity toward the early stage of the disease; it was revaluated trying to

identify factors that could better correlate with the persistent/erosive features of RA. With these new criteria, measured with specific scores, the diagnosis is now based on a confirmed presence of synovitis in at least one joint and with no other reliable explanations and a total score of 6, summing individual scores of the four groups. These groups take into consideration the number and the sites of involved joints, the presence of serological abnormalities and consistent acute-phase responses, and, finally, symptoms duration. Clinical analyses highlight the presence of an inflammatory event in progress however some parameters, such as the presence of acute-phase proteins, are not specific for RA if compared with other serological markers such as RF and ACPAs.

RF, Rheumatoid Factor, is an antibody that recognize the Fc portion of immunoglobulins G and high titers are, typically, correlated to chronic inflammatory conditions. RF is not an hallmark of RA, since its presence is correlated to other infective and inflammatory conditions, however it has been demonstrated that, in RA patients, occurs many years before the onset of the disease, justifying its inclusion within the classification criteria. The discovery of RF and its association to RA have been important steps in the investigation of the disease opening the possibility to search for biomolecules circulating in the asymptomatic phase of RA to use as early diagnostic markers of the disease.

1.1.2 Anti-citrullinated peptide antibodies

Observations made on RF have promoted the search for other antibodies to associate with RA and several ones have been found to be more specific and to have a higher positive predictive value than RF. The first antibodies described after RF were the anti-perinuclear factor (APF) and the anti-keratin factor (AKA) which were detected in RA patients sera using immunofluorescence techniques.[5,6] Both the antibodies recognize the same antigen, namely filaggrin, which is a protein involved in the cytokeratin filament-aggregation.[7] Filaggrin is not a joint-specific protein therefore the recognition is supposed to be the consequence of a cross-reaction against un unknown antigen expressed in the inflamed sites.

APF and AKA antibodies are included in the antifilaggrin autoantibodies (AFA) family which was proved to be directed against citrullinated epitopes.[3,8,9] This discovery represented a remarkable breakthrough toward the understanding of the reactivity of AFA since citrulline residues constitute the major antigenic determinant recognized by antibodies present in RA sera.

Besides AFA other antibodies were retrieved such as the so called anti-Sa (anti-Savoie) antibodies that showed a very high specificity (around 95%) for RA, a variable

sensitivity according to the stage of the disease and a strong association with the disease severity; it was also demonstrated a reactivity toward citrullinated vimentin.[10,11] Antinuclear antibodies have been detected in RA patients such as the anti-A2/anti-RA33 antibodies directed to the heterogeneous nuclear ribonucleoprotein A2 (hnRNP-A2) which is overexpressed in synovial membranes, where it may be targeted by autoreactive B and T cells.[12] During years others autoantibodies with different sensitivity and specificity toward RA have been described, all of them able to recognize citrullinated proteins. For this reason the acronym ACPAs, anti-citrullinated peptides antibodies, has been coined to describe this heterogeneous family of antibodies sharing reactivity toward citrullinated epitopes.

Based on these discoveries enzyme-linked immunosorbent assays (ELISA) have been developed using citrullinated antigens as immobilized probes able to specifically elicit ACPAs from RA patients' sera. Cyclic citrullinated peptides (CCPs) are synthetic filaggrin peptides developed to act as surrogate target antigens and used in the CCP-test that represents the gold standard assay for an high specific and sensitive detection of ACPAs in RA sera. Three generation of CCP-tests have been developed with increasing performances and the antigenic probes have been optimized until a mixture of citrullinated peptide sequences have been adopted as the best discriminating.

Clinical interest on ACPAs has become justified when they were found to be highly specific for RA and to have higher predictive value compared to RF since they were monitored in the very early stages of the disease.[13,14] In other words they represented an important biomarker of RA thanks to their presence in patients' sera several years before the symptomatic onset of the disease.

Biomarkers are measurable characteristic elements whose presence is indicative of a pathological status therefore they represent an important tool for the diagnosis or prognosis of a disease. In the field of autoimmunity the research of biomarkers is a stimulating topic because they are progressive diseases diagnosed only in the late phase when tissue damages became sadly evident. Disease outcome varies from mild to severe systemic symptoms when joint destruction is accompanied by extra-articular manifestations and an early intervention is crucial in preventing irreversible joint damages.[15] Therefore the availability of molecular markers to be monitored during the asymptomatic stage of the disease can improve the effectiveness of the treatments promoting, as example, a well-timed use of DMARDs able to produce a favorable effect on the course of the pathology.[16]

It is not surprising that for RA the biomarkers are represented by autoantibodies since autoimmune diseases are characterized by a strong activation of the humoral component of the immune system. The development of proper assays able to monitor their presence constitutes another topic of great interest that has led to the release in

the market of the CCP tests which represent easy to be performed, low cost and not invasive tests.

The presence of ACPAs in the serum is one of the prominent criteria to be considered for a correct RA diagnosis but itself it is not sufficient because a small percentage of healthy individuals (about 2%) and a more consistent fraction of patients with other systemic inflammatory disease (<10%) are ACPAs positive, too. In addition there are also ACPAs negative individuals who are affected by RA, that suffer different risk factors and have different progression rate of the disease but that share similar clinical characteristics with the ACPAs positive ones.[17]

The correlation between the presence of ACPAs in RA patients' sera and a more severe and erosive progression of the disease has been demonstrated fueling the hypothesis that ACPAs themselves may be involved in the pathophysiology of RA and even because their presence is detected before a clear manifestation of disease symptoms.[18,19]

However ACPA recognition of citrullinated proteins had stimulated a lot of interest around these antibodies and several studies had been led in order to disclose the auto-antigens that can function as real immunogens for the anti-citrulline responses of this disease. Many citrullinated proteins have been suggested as candidate autoantigens responsible of the induction and/or perpetuation of ACPAs responses but only for few of them the pathogenetic mechanism has been confirmed (Figure 1.3).

Candidate citrullinated antigens*	
Histones	Components of chromatin. Act as spools, around which the DNA winds. Histones also play a role in gene regulation.
Nucleophosmin/B23	Functions in ribosome assembly, nucleocytoplasmic transport and centrosome duplication.
Sa antigen/vimentin	Role in supporting and anchoring the position of the organelles in the cytosol.
Fibronectin	Mediates several physiological processes through cell-surface integrin receptor interaction and growth factors.
Collagen type I + II	Functions as a structural protein in connective tissue and comprises the main extracellular support system.
Alpha-enolase	Plays a role in many processes, such as glycolysis, growth control, hypoxia control, and allergic responses.
Eukaryotic translation initiation factor 4G1	Component of the protein complex eIFF4F, which is involved in the recognition of the mRNA cap, ATP-dependent unwinding of 5′-terminal secondary structure and recruitment of mRNA to the ribosome.
Fibrin(ogen)	Fibrin and its precursor fibrinogen are involved in the process of blood clotting.
Epstein-Barr Virus nuclear antigen 1	Nuclear protein encoded by the Epstein-Barr virus (EBV). EBV is known to infect human B lymphocytes and epithelial cells creating a reservoir in these cells.
Cap-Z alpha 1	F-actin capping proteins bind in a Ca^{2+}-independent manner to the fast-growing ends of actin filaments, thereby blocking the exchange of subunits at these ends.

Figure 1.3. Candidate citrullinated (auto)antigens that have been associated with RA[18]

The synovium has always been considered the principal source of autoantigens because it is the primary site of disease manifestations where immuno competent cells

massively migrate creating follicular structures that act as a sort of secondary lymphoid organ. Since ACPAs are also produced by the local plasma cells of the *pannus* it is presumable that the affinity maturation of the infiltrating B cells is guided by local citrullinated proteins.[20] Therefore the characterization of potential citrullinated autoantigens can help a better understanding of the role of the ACPAs.

An important point to be clarified is that citrullinated proteins occur in different types of joint inflammation, not only in RA, therefore the presence of ACPAs rather than the presence of such proteins is a characteristic of this disease. Moreover it has been proved that not all citrullinated proteins are specifically recognized by ACPAs so some unknown mechanisms are supposed to be responsible for the restriction of the immune response toward some citrullinated antigens.

A first consideration to take is that antibodies responses to proteic antigens require a cooperation between B and T lymphocytes. As consequence a citrullinated protein recognized by a B cell and containing the correct deiminated B cell epitope but lacking an appropriate T cell epitope (deiminated or not) will not provoke the B cell complete activation, preventing antibodies' release. In the anti-citrulline reactivity observed in RA autoimmune responses may widely vary between different citrullinated epitopes and the presence of specific autoepitopes promoting reciprocal B cell-T cell interactions may be the crucial aspect to initiate the ACPA response.

A second important aspect to take into account is the influence of the genetic background that in autoimmunity plays a crucial role since specific poligenic profiles enhance the probability to develop the pathology. In particular the MHC class II alleles have been found associated with autoantibodies specificity, and involved in the etiology and pathogenesis of autoimmune disease according to the "shared epitope hypothesis".[21] The better-known *HLA-DRB1* alleles include members of the group **04*, **0101*, **0102, *1402* and **1001* which code for variants of the five amino acid sequence 70–74 of the region of the DRB chain responsible for the peptide binding and T cell recognition. It is assumed that structural differences between class II molecules can influence this interaction and the higher ACPAs levels found in HLA shared-positive RA patients suggest that a specific MHC background may contribute to the antibody response restriction. The mentioned HLA alleles showed a higher affinity for citrullinated peptides demonstrating the targeted recognition of such epitopes. They also proved that a B-cell response directed to citrullines and involving the T cell compartment is at the centre of RA pathology.

As mentioned before, citrullination is typically seen as a symptom of inflammation however it is described not only in inflammatory arthritis but also in other autoimmune diseases (i.e. Multiple Sclerosis, psoriasis), in neurological pathologies (Alzheimer's disease), in cancer, in glaucoma.[22,23] As example citrullinated Myelin Basic Protein (MBP) has been associated to Multiple Sclerosis correlating a reduced myelin's

compactness with the secondary structure loss of the protein, as consequence of the citrullination. An enhanced digestion by cathepsin D, an increased release of antigenic peptides and a propagated autoimmune response were consequences of the modification, too.[24] This PTM has been also described in some physiological conditions such as brain development, female fertility, hair formation, plasticity of the central nervous system, terminal differentiation of the epidermis and also regulation of gene expression.[25]

1.1.3 Deimination and proteomic methodologies for the characterization of deiminated antigens

The research of citrullinated proteins, as possible candidate autoantigens involved in the pathological mechanisms of RA, represents an intriguing topic and many articles have been published both using classical biochemical approaches and innovative mass spectrometric strategies. Such methodologies rely on the specific chemical properties of the citrullines which are quite different from the originating arginines. Citrulline is a non standard amino acid since there are neither a corresponding mRNA codon nor a tRNA. It is not incorporated into proteins during translation but it is the result of a PTM, namely citrullination or arginine deimination (Figure 1.4).

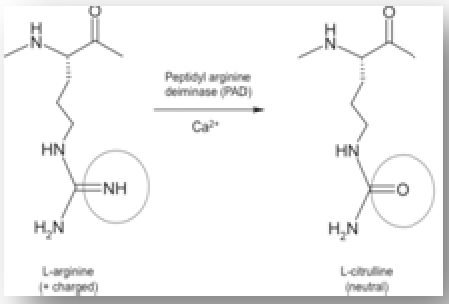


Figure 1.4. Deimination of arginine into citrulline by means of the PAD

This PTM essentially consists in the conversion of the guanidinium group of the arginine's side chain into an ureic group through an enzymatic reaction catalyzed by the peptidylarginine deiminases (PAD), calcium-dependent metalloenzymes that belong to the guanidinium-group modifying enzymes superfamily. Five PAD isotypes have been described, sharing consistent sequence homology among them (50-55%) and mammals (70-95%).[2] The main difference concerns their anatomic distribution since each of them has a specific tissue localization: PAD1 is mainly expressed in the epidermis and uterus, PAD2 is more ubiquitous even if it is more expressed by the

neural tissue and the macrophages, PAD3 has an epidermic expression too, PAD4 is expressed by the white blood cells, mainly granulocytes, neutrophils and eosinophils and PAD6 has been recently identified within mammalian oocytes.[26] Among the five isotypes PAD4 is the only one that has a nucleic localization, compared to the cytosolic distribution of the others, and, together with PAD2, its presence has been demonstrated in rheumatoid synovial membranes, synovial fluid cells and extracellular synovial fluid.[3]

PADs aroused interest within RA research starting from the discovery that RA patients owned autoantibodies to citrullinated peptides and that PADs genes conferred susceptibility for the disease but it increased once found a strong association between a specific PAD4 haplotype and RA.[27] Four single nucleotide polymorphisms (SNPs) found in the PAD4 gene provoke three amino acid substitutions supposed to be responsible for a more stable transcript and an increased activity of the enzymes which, ultimately, result into an enhanced production of citrullinated proteins acting as autoantigens. Such association has been demonstrated for Japanese and Korean population but was not confirmed for Caucasian people probably for the occurrence of different polymorphism or haplotypic combinations.

The hydrolytic deimination activity of PADs depends on several factors including the sequence of substrate proteins, the presence of other PTMs and even the type of residues flanking the targeted arginine so it seems that some amino acid such as tryptophan favor the modification while other such as proline has a negative effect.

The conversion of an arginine into a citrulline involves some critical chemical modifications such as loss of a positive charge, decreased isoelectric point and mass increase of 0.984 Da which altogether are responsible for specific modifications on proteins (Figure 1.5; on the left). A positive charged arginine side chain converted into a neutral group produces relevant effects on the protein structure because the hydrogen bonding properties are lost changing the overall conformation (Figure 1.5; on the right).[2]

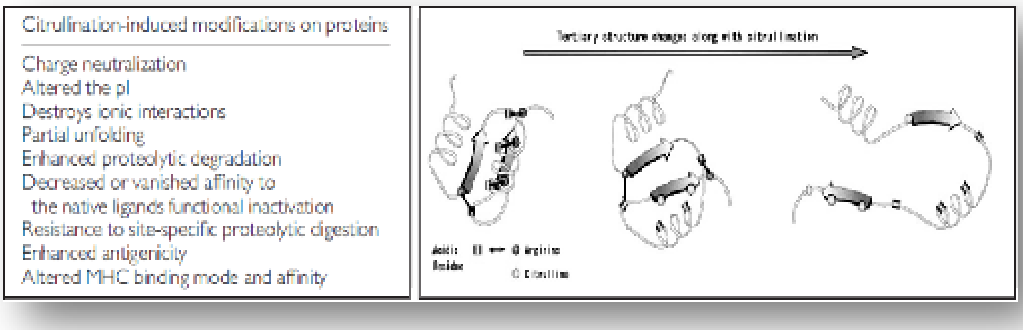


Figure 1.5. Possible modifications induced by citrullination (on the left); as example the change of the tertiary structure is shown (on the right)

Taking into account these features during years several methods have been developed for the detection of citrullinated proteins and peptides in mixture: immunodetection, mass spectrometric (MS) analyses with modifications and MS analyses without modifications.

The most classical approach regards 2D-SDS-PAGE followed by specific immunodetection of citrullinated proteins, as originally described by *Senshuo et al*.[28] Basically, proteins transferred into a membrane are incubated at low pH with 2,3-butanedione and antipyrine which selectively tag citrulline residues; then the immunodetection step is performed using commercially available antibodies able to recognize citrulline modified with such reaction. The method is sensitive and specific thanks to the use of antibodies and is exploitable for *in vivo* samples citrulline detection but it does not provide information neither about the citrulline-bearing proteins nor about the exact location of the modification.

Preliminary information can be extrapolated by a 2D-gel carefully observing the migration of the proteins: since PTMs can provoke changing at mass and/or isoelectric point, trains' of spots on the gel are presumably formed by multiple versions of the same protein. Citrullination do not provoke a consistent change in the mass but in the isoelectric point greater the greater is the number of deiminated arginine therefore an horizontal cluster of spots with highest immunoreactivity at the more acidic side of the train is typical of proteins bearing an increased status of citrullination. According to previous considerations it is evident that the primary purpose of this technique is the identification of citrullinated proteins in a target tissue and other proteomic strategies have to be used in order to provide more detailed molecular information.

Different strategies based on specific modifications of citrulline residues have been developed, using chemical tags which facilitate its MS detection or its enrichment from a protein digest.[29]

Derived from the original reaction developed by *Senshuo et al* a labeling strategy with 2,3-butanedione and antipyrine can be applied directly to a protein digest, selectively introducing into citrullines a chemical tag that produces a mass increase of 238.11 Da.[30] Such citrulline adduct can be detected by MS analysis and, in particular, using MALDI-TOF-MS it can be also easily recognized thanks to a characteristic isotopic pattern; moreover it acquires chromophoric features that allow a specific UV-absorption at 464 nm. The potential of this strategy has been further enhanced after the discovery that citrulline-modified peptides exhibit a specific 201.1 Da loss after collision induced dissociation (CID) on LC-MS/MS analysis, representing an important signature ion.[31] Due to the difficulty of identifying derivatized citrullinated peptides by CID spectra, the employment of a different fragmentation technology has been proposed, namely electron transfer dissociation (ETD). It showed a nearly complete coverage with all fragment ions shifted of 238.11 Da indicating that a soft fragmentation can preserve the

adduct in the peptide. This labeling strategy was proved to be selective for citrullines and highly efficacious but such results came from standard peptides or standard protein deiminated *in vitro*, both samples available in large amount. The efficacy on real samples containing a small fraction of citrullinated peptides is not known. Even ETD is not a routinely technology nor it is widely available on the most common mass spectrometers.

A variation of the described method has been proposed without the use of antipyrine and increasing the incubation time with 2,3-butanedione; LC-MS data were processed using an in-house search application that performed comparison of the peptide mass fingerprinting.[32] In this case no complex MS instrumentation was required but an hinged sample pre cleaning and fractioning was necessary, with evident difficulties if applied to low concentrated samples.

Another labeling strategy based on the specific reaction of glyoxal derivatives with the citrulline ureido group under acidic conditions was proposed using derivatized beads exposing 4-hydroxyphenylglyoxal to react with citrulline.[33] This way only citrullinated peptides were stopped on the beads and, after a proper washing step and a strongly basic cleavage, citrullinated peptides were released carrying a modified ureido group.

In a recently published paper it was proposed the use of a citrulline-specific chemical probe, namely rhodamine-phenylglyoxal, to selectively tag citrulline residues and furthermore allowing a fluorescent detection on SDS-page thanks to the presence of the linked rhodamine dye.[34] This methods are all very specific and contribute to a reduction of the sample complexity but they require a large amount of sample.

It is interesting to note that all the labeling strategies that have been developed start from considering the reactivity of arginine with dicarbonyl compounds including α-α'-diketones, such as 2,3-butanedione, and α-keto aldehydes, such as glyoxal and substituted glyoxals, which are known to react with an amidine group to form an addiction product.[35] The breakthrough has been represented by the use of acidic conditions that prevent the reactants to attack the protonated guanidinium group of arginines and instead promote the selectively addition to the ureic group of citrullines.

Additionally a stable isotope labeling strategy has been proposed based on the introduction of the ^{18}O heavy isotope of the oxygen atom by means of PAD in presence of $H_2{}^{18}O$ so that citrullinated peptides acquire a 3 Da mass shift and a characteristic isotope distribution different from natural abundance.[36] The big deal of this strategy is its exclusive employment for the detection of proteins citrullinated *in vitro*.

Unlike the strategies based on labeling reactions the approaches that rely on the direct detection of citrullines do not use any tagging reaction but exploit the both chemical and mass spectrometric properties of this residue, starting from the 0.984 Da mass shift from the unmodified arginine. Methods based only on the detection of such small difference anticipate several problems: the presence of the unmodified peptides, the

difficulty to discriminate arginine deimination from the isobaric asparagine/glutamine deamidation, the incomplete fragmentation of citrulline-bearing peptides, the controversy regarding trypsin ability to cleave citrullines or not, and the need for an high performing LC-MS/MS instrumentation.

Two interesting approaches have been described: one based on both the accurate mass and retention times of specific citrullinated peptides and the other based on a diagnostic mass loss from citrullinated peptides submitted to CID. In the first case some citrullinated peptides from fibrinogen deiminated *in vivo* were used as reference (in terms of mass and retention times) for the identification of the corresponding *in vivo* citrullinated fibrinogen obtained from synovial tissue samples. It is evident that the analysis of the biological sample is limited to the peptides which can be detected in the analysis of the *in vitro* citrullinated protein.[37] Instead in the second approach MS/MS analysis of citrullinated peptides highlighted a specific neutral loss of 43 Da, representing the detachment of isocyanic acid from the ureido group.[38] This mass signature is very interesting but, once again, it was proved on standard abundant peptides and no information regarding a low concentration range are available.

An integrated strategy has been proposed combining the diagnostic isocyanic acid departure obtainable from CID with the higher identification power of ETD; basically ETD is performed only on those peptides that after CID fragmentation exhibited the typical neutral loss of 43 Da.[39] The need for an MS instrument able to perform both of the fragmentations in a single analysis is evident therefore the real applicability is limited.

A very recent and pioneering approach has been proposed for the MS detection of citrullines; it compares the experimental isotopic pattern of a mono charged citrullinated peptide with its expected theoretical one.[40] If the corresponding unmodified peptide is present, the citrullinated peptide would exhibit a skewed isotopic distribution, due to the partial overlapping between the two forms. Using statistical models to calculate the deviation it is possible to quantify citrullinated peptides. However the decisive confirmation regarding the presence of citrullines is fulfilled by means of labeling strategies since a deamidated asparagine/glutamine provokes an analogues mass shift of 0.984 Da.

The major consideration that emerges from the overview of the MS currently available methods for the detection of citrullines is that the great part of them has been proved only on standard peptides or at least on standard proteins deiminated *in vitro*. Both cases represent ideal conditions for the low complexity and the large amount available for the analysis; only few methods have been tested on biological samples. In addition some strategies rely on complex instrumentations and other present limited applications.

In conclusion the best approach is still not there since each of them presents positive and negative aspects. It is also evident the huge demand to improve the sensitivity of these methods to be applied to biological samples in order to map the citrullination status of *in vivo* deiminated proteins and do advancing the correlated biomedical research.

1.1.4 Deiminated histones and neutrophil extracellular traps

Histones are hydrophilic and basic proteins characterized by the presence of several positively charged amino acids, namely lysines and arginines; these residues are involved in interactions with the negatively charged phosphate groups of DNA. Indeed histones participate in the organization of nucleosomes which are the structural core units responsible for the chromatin packaging.

Five canonical histone isotypes, namely H1, H2A, H2B, H3 and H4, are known and many other subtypes for each of them have been described. Histones are typically nuclear proteins but evidences have been accumulated proving their presence in different sub cellular compartments. As example in the cytosol they can form complexes with other histones or other proteins.[41,42] Additional features have been described such as the ability to cross lipid bilayers through an endocytosis-indipendent pathway based on a direct translocation and to mediate penetration of macromolecules covalently attached to them.[43,44] The antimicrobial activity of histones is not a recent finding and three histonic peptides have been described to be active against bacteria. They are buforin I (39 amino acids),[45] parasin I (21 amino acids),[46] and hipposin (51 amino acids),[47] all derived from the N-terminus domain of histone H2A. However histones themselves have an antimicrobial activity suggesting that they could actively take part in the innate immunity responses.[48] All these data prove that histonic proteins do not exclusively act as scaffold for the nucleosome core but they are possibly involved in many other functions, largely unknown.

Strictly correlated to the mentioned antimicrobial activity of histones is the recent finding that they are also involved in a form of innate response carried out by neutrophils. In presence of infectious and inflammatory stimuli these cells extrude structures named neutrophil extracellular traps (NETs) which exhibit an antibacterial activity.[49] This mechanism ensures a high local concentration of antimicrobial molecules able to disarm and kill bacteria. It is very interesting to note that the key event for mediating NETs release is just the histones citrullination mediated by enzyme PAD4. This event provoke the chromatin unfolding so that web-like structures form and trap granule and nuclear constituents of neutrophils.[50]

Histones bearing citrullinated residues do not represent a novelty because they have been already described in the epigenetic regulation of gene transcription where deimination antagonize arginine residues methylation. With the discovery of NETs a further functionality of deiminated histones has been discovered, that is chromatin unfolding for NETs release. The presence of PTMs on histones able to modulate their activity is a known event and there are so different occurring PTMs (acetylation, phosphorilation, ubiquitination, methylation, ADP-ribosylation, biotinylation etc) that an "histone code" hypothesis has been proposed, predicting that a single or a combinations of PTMs may work to regulate histones structure and to guide distinct biological functions.[51]

Histone deimination and the production of reactive oxygen species (ROS) are crucial events to initiate the NETosis which is a cell-death program different from the well known necrosis and apoptosis mechanisms. Basically, stimulation of neutrophils results in the activation of NADPH oxidases and in the formation of ROS. How ROS directly promote NETs formation remains unclear but their signaling is required to start the subsequent molecular events: disruption of the nuclear membrane, PAD4 activation, histones deimination, chromatin decondensation, mixing of nuclear contents with cytoplasmic and granular proteins (Figure 1.6).

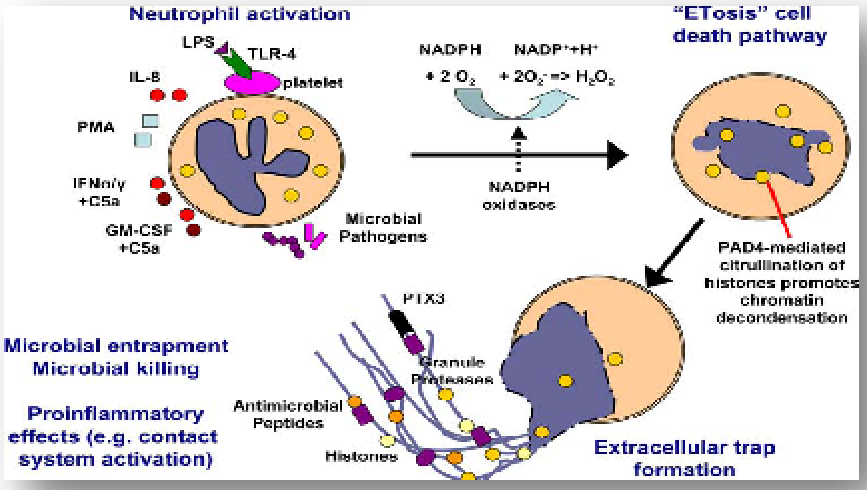


Figure 1.6. Model for the formation of neutrophil extracellular traps[52]

As a final step, traps formed by DNA fibers trapping antimicrobial components and nucleic proteins are released in the extracellular medium thus blocking and killing different microbes.[53] Elastase, cathepsin G, myeloperoxidase, lactoferrin, gelatinase B, pentraxin 3 and even histones are examples of proteins that NETs carry outside the cell.

A wide panel of proinflammatory stimuli have been proved able to initiate the NETs program: bacterial lipopolysaccharide (LPS), the mitogen phorbol myristate acetate (PMA), the interleukin 8 (IL-8), hydrogen peroxide and so on. In few cases a direct exposure is sufficient to trigger NETs formation like in the case of some pathogens (S. aureus, Streptococcus pyogenes, Mycobacterium tuberculosis, Escherichia coli, Candida albicans). In other case a priming step must have been done to stimulate NETs formation like for the chemotactic complement-derived peptide C5a while no such priming step is required for LPS, PMA, or IL-8.

As evident both infectious and not-infectious stimuli can generate NETs and several evidences had accumulated that NETs may be involved in various clinical settings, in particular in the autoimmunity field. Indeed the release of PAD4 into the extracellular space may generate modified self-antigens with potential immunogenic features. Also the exposure of intracellular proteins by means of NETs could plays a role during the development of autoimmune diseases.

An example of NETs occurring in autoimmune disorders is proved by the small-vessel vasculitis, a chronic auto inflammatory condition, where it was shown that NETs contained two known autoantigens, proteinase-3 and myeloperoxidase, normally contained in the neutrophils granules.[54] Instead in systemic lupus erythematosus (SLE) a subset of patients were found to have an impaired NETs degradation which was correlated to an higher anti-NET Abs deposition at kidney level that contributed to the renal damages.[55]

These last evidences are very interesting because SLE patients own a type of autoantibodies called anti-neutrophil cytoplasmic antibodies (ANCA) which are mainly directed against DNA, histones, riboproteins and other nucleoproteins. Immuno complexes formed by ANCA and relative antigens may be pathogenic in the development of gromerulonephritis which is a major organ manifestation of SLE. Authors propose that NETs may act as a source of autoantigens by exposing chromatin and neutrophil proteins in an immunostimolatory context. Moreover a persistent exposure can exacerbate the autoimmune response leading to the formation of immuno complexes.

The role of NETs in the etiopathogenesis of autoimmune diseases is largely unknown but these preliminary studies demonstrated that they constitute an interesting and innovative scenario to explore.

1.2 OBJECTIVES

The research of citrullinated proteins with a potential role in the pathogenesis of rheumatoid arthritis (RA) started many years ago, when anti-citrullinated proteins antibodies (ACPAs) were recognized as biomarkers of the disease.

It is now accepted that ACPAs represent a family of antibodies with overlapping specificities that are directed to a variety of citrullinated substrates, of exogenous or endogenous origin.

In this work of thesis the research of deiminated proteins recognized by ACPAs has been conducted focusing the attention toward histone proteins, which have been studied in the context of neutrophils extracellular traps (NETs).

Histones are nuclear proteins described to occur even in other sub cellular compartments or extracellularly and the recent discovery that neutrophils extrude deiminated histones outside the cell to form NETs has attracted attention.

Several evidences have proved that NETs exert an antimicrobial function and their release occur in case of infections. When NETs clearance is not correctly regulated, they may persist in the extracellular environment, leading to the presentation of antigens to the immune system and possibly supporting autoimmune responses toward self antigens.

Therefore the aim of this project was to investigate the presence in neutrophils nuclear extracts and NETs of deiminated histones recognized by RA sera and to characterize their deimination pattern using proteomic strategies.

In the first part of the research we evaluated the presence of deiminated histones within nuclei and NET of purified neutrophils, using a classical approach that employed SDS-PAGE followed by immunodetection with RA sera. Mass spectrometric analyses have been conducted to identify the proteins present in the bands and a derivatization technique has been used to characterize the presence of the deimination.

In the second part of the work we investigated more in depth the deimination pattern of histone H4 by means of an LC-MS/MS strategy based on the direct detection of citrulline residues with two different stationary phases, namely reversed-phase and HILIC.

1.2.1 Scheme of the work

The first part of the project was developed as follows:

- Neutrophils isolated from healthy blood donors were cultured and incubated with or without a calcium ionophore able to activate PAD4-mediated deimination. Acid extracted proteins were separated on Tris-Tricine-SDS PAGE and subjected to immunodetection with RA sera to check the reactivity. Corresponding reactive gel bands were trypsin-digested and submitted to MALDI-TOF analysis to achieve protein identification.
- A derivatization technique with 2,3-butanedione and antipyrine, specific for citrulline residues, was optimized on different standard samples including citrullinated and not citrullinated peptides combined in binary and more complex mixtures.
- The derivatization technique was applied to the trypsin-digested bands and analyzed by MALDI-TOF MS to search for possible deiminated peptides.
- Neutrophils were used again and activated to produce NETs in presence of PMA which is a stimulus known to prime the respiratory burst by enhancing superoxide release and promoting the generation of ROS. NETs' proteins were gel separated and immunoblotted to confirm the reactivity toward RA sera. The proteolytic digest of the gel band was analyzed by MALDI-TOF MS. The same labeling strategy was employed to highlight the presence of deiminated peptides.

The second part of the project was developed as follows:

- A direct detection of citrullines was attempted with the NETs digest trying to obtain more detailed information regarding the deimination pattern of the identified protein by means of LC-MS/MS analysis. Taking into account all the intrinsic critical aspects of this approach reversed-phase and HILIC were considered as the most appropriate chromatographic supports.
- Each stationary phase, equipped on a *nano*LC-LTQ-Orbitrap system, was tested with standard peptides to verify the selectivity toward a pair of deiminated and not deiminated peptides.
- Tryptic NETs digest was separated on the reverse-phase and MS/MS spectra of citrullinated peptides were analyzed in order to deduce the exact deiminated position by looking for fragment ions bearing the 0.984 Da increment.
- Similar analysis were performed on HILIC highlighting the major problems that emerged from such technology; a targeted MS/MS analysis was repeated focusing the attention on the peptide of interest.

1.3 RESULTS AND DISCUSSION

1.3.1 Identification of antigens expressed in nuclei of activated neutrophils

A density gradient fractionation of blood is a common procedure employed to isolate mononuclear cells and granulocytes by means of polysucrose media that allow the separation of the cells on the basis of their size, shape and density. A density gradient is typically created by layering media of different density in a centrifuge tube. When a sample is layered on top of this density gradient and centrifuged, the various particles move through the gradient at different rates. According to the protocol described by *English et al* centrifugation of heparinized human blood on a discontinuous gradients created by sucrose media results in the simultaneous separation of mononuclear leukocytes, granulocytes, and erythrocytes. [56]

In this project granulocytes were purified starting from anticoagulated blood of healthy subjects which was layered on a double gradient of two different Histopaque® solutions. After centrifugation granulocytes were recovered from the ring formed at the interphase between the two solutions while PBMCs stopped at the plasma interphase and erythrocytes sedimented at the bottom of the tube. Washed granulocytes were cultured and activated by A23187 stimulation for 15 min at 37°C. Cells were lysed in H_2SO_4 and acid extracted proteins were precipitated in TCA. After an appropriate re-suspension proteins were separated on acrylamide gels, transferred to PVDF and subjected to Western blotting using RA sera and normal healthy subjects. Blots showed a specific reaction of the pathological sera with an antigen of 11 kDa while no reactivity was observed with control sera (Figure 1.7).

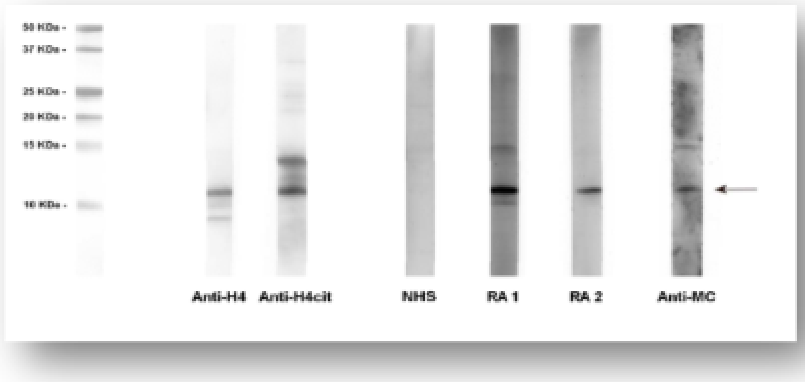


Figure 1.7. RA sera bind to citrullinated H4 from stimulated neutrophils. RA1 and RA2 =sera of two RA patients; NHS=sera of a normal blood donor

Immunodetection was also performed with a specific anti-histone H4 antibody (anti-H4), an anti-citrullinated-histone H4 polyclonal antibody (anti-H4cit) and an anti-modified citrulline antibody (anti-MC); a reaction occurred at the same position in the blot.

Altogether these data suggested that the protein at 11 KDa recognized by RA sera, by anti-H4, by anti-H4cit and by anti-MC antibodies might be histone H4, probably carrying deimination sites. To confirm such hypothesis an MS analysis of the corresponding gel band was performed.

The band was excised from two lines of the gel where the same protein extract was initially loaded so as to have the result in duplicate. The enzymatic in-gel digestion protocol was used to convert the protein contained in the bands in peptides that could be extracted and analyzed by MS to achieve the identification. The procedure involved consecutive washes to remove the Coomassie staining, a step of reduction and carbamidomethylation of cysteine residues to destroy disulfide bonds and an overnight trypsin digestion. Tryptic peptides were analyzed by MALDI-TOF MS and the corresponding peptide mass fingerprintings (PMFs) were acquired (Figure 1.8).

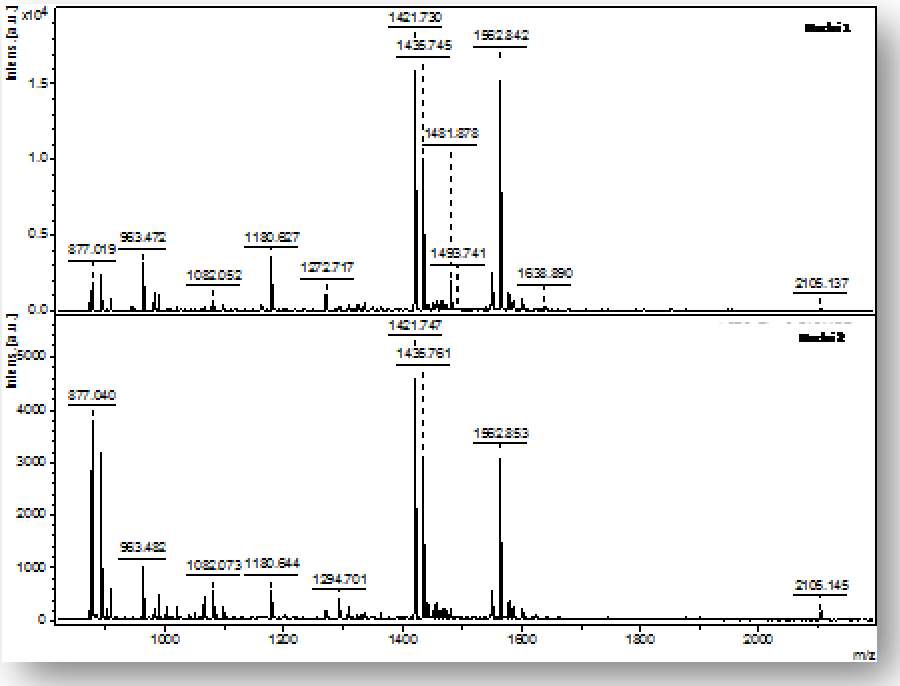


Figure 1.8. PMFs from MALDI-TOF spectra obtained from the excised bands from nuclear extracts after tryptic digestion. Nuclei 1= digest from lane 1 from the gel of nuclear extract; Nuclei 2= digest from lane 2 from the gel of nuclear extract

The spectra were sent to databank search within the Swiss-Prot human proteome using Mascot as search engine and, in both cases, two proteins were significantly

identified, namely protein S100-A8 and histone H4. For both samples Mascot query results are shown (Figure 1.9).

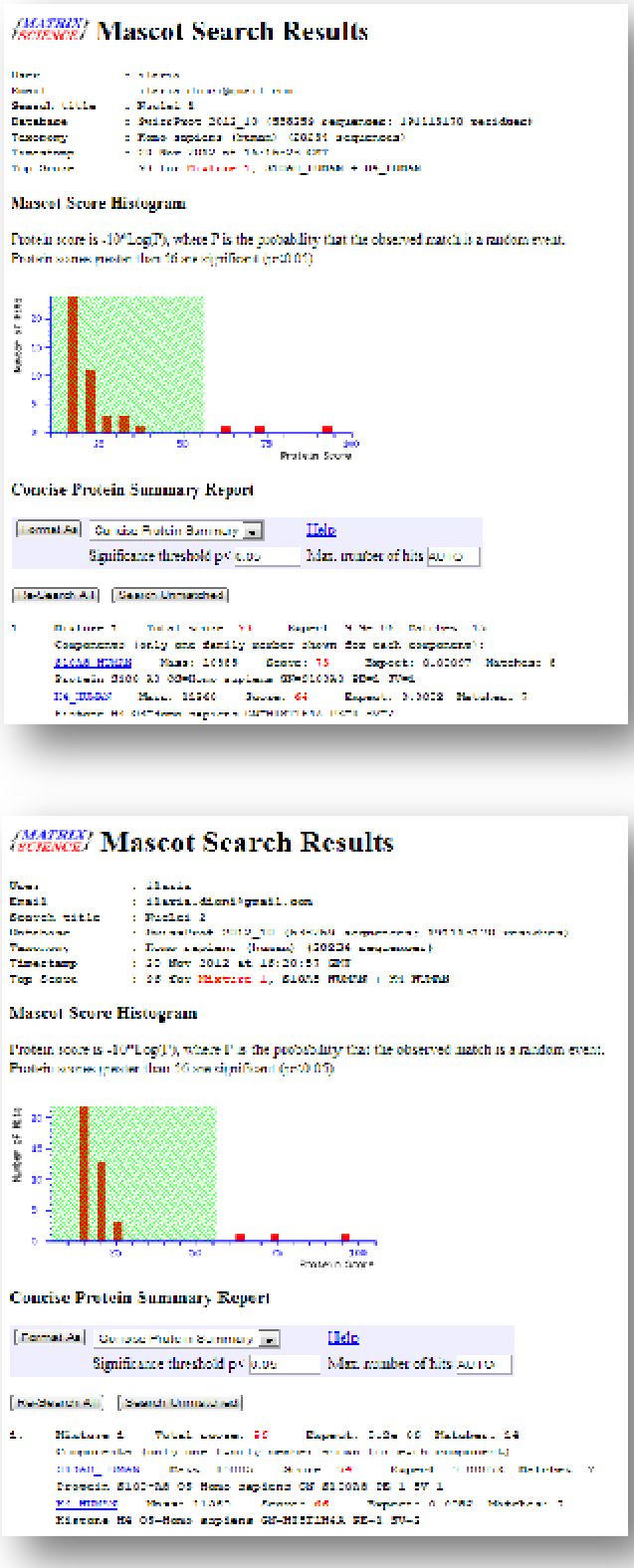


Figure 1.9. Databank search of nuclei sample number 1 (above) and number 2 (below)

Both proteins returned a score above the threshold set at 56 by the software and a sequence coverage higher than the 50% was calculated for all the proteins.
For nuclei sample number 1, 8 peptides were recognized for protein S100-A8 (Figure 1.10; on the left) and 7 for histone H4 (Figure 1.10; on the right).

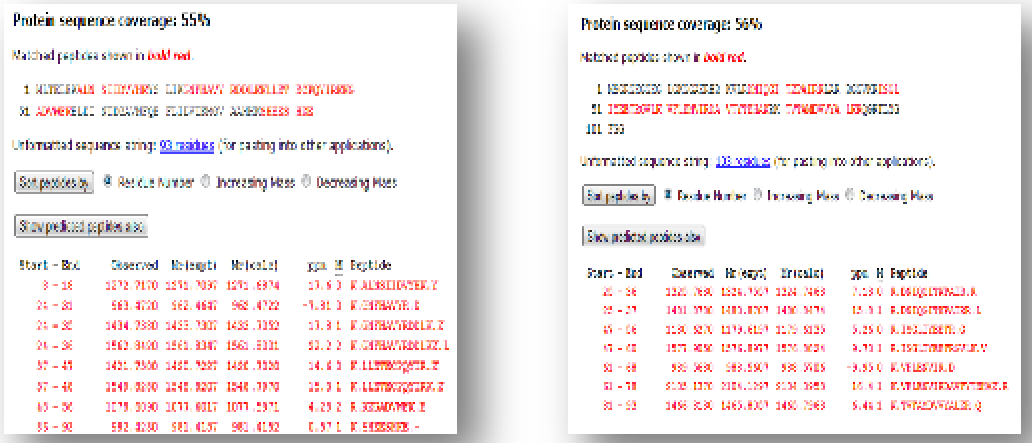


Figure 1.10. Protein view of S100-A8 protein (on the left) and histone H4 (on the right), for nuclei sample number 1

For nuclei sample number 2, 7 peptides were retrieved for both protein matches (Figure 1.11).

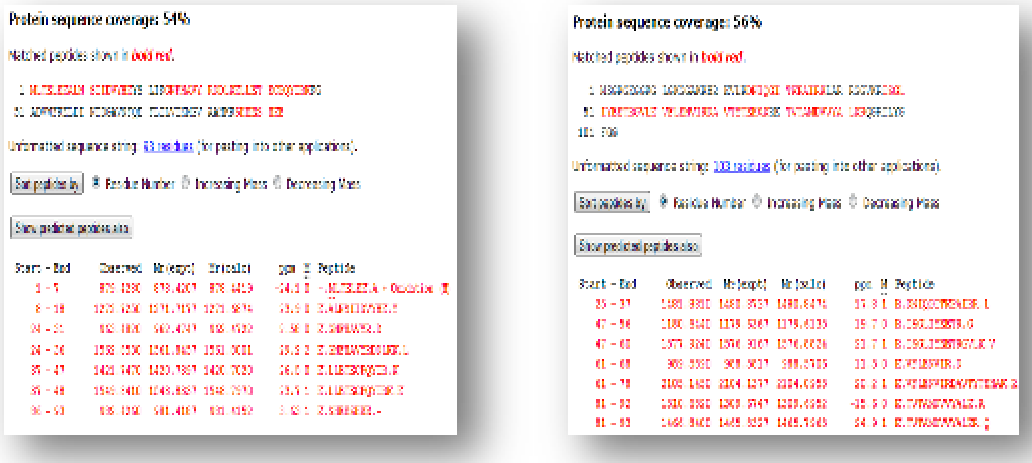


Figure 1.11. Protein view of S100-A8 protein (on the left) and histone H4 (on the right), for nuclei sample number 2

To confirm proteins identification tandem mass spectrometry analyses were performed on selected peptides from the two proteins of the two bands and MS/MS spectra acquired to achieve the peptide sequencing. As example are reported the MS/MS

spectra of two peptides, one for each protein: precursor MH^+=1180.600 corresponding to the sequence ISGLIYEETR contained between residues 47-56 of histone H4 (Figure 1.12) and precursor at MH^+=1421.700 corresponding to the sequence LLETECPQYIR (bearing a carbamidomethylation on the cysteine residue) contained between residues 37-47 of protein S100-A8 (Figure 1.13).

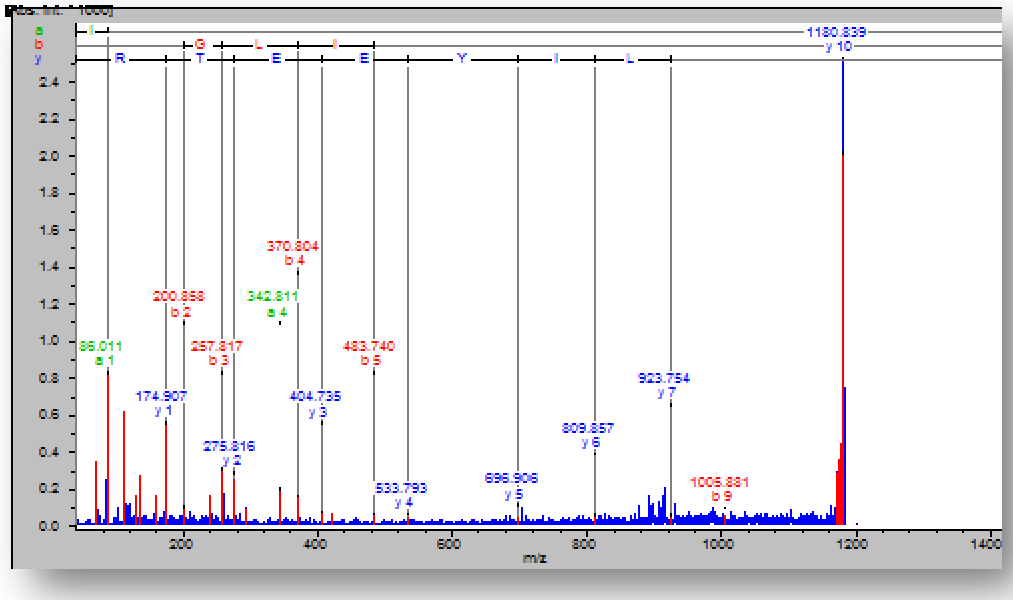


Figure 1.12. Precursor MH^+=1180.600 Da corresponding to ISGLIYEETR sequence contained between residues 47-56 of histone H4

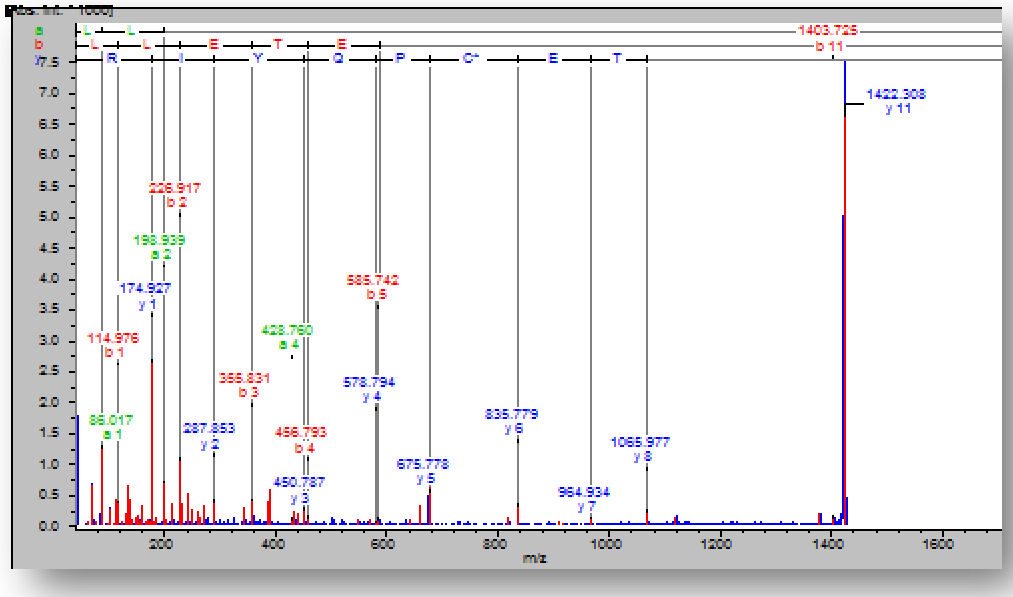


Figure 1.13. Precursor MH^+=1421.700 Da corresponding to LLETECPQYIR sequence contained between residues 37-47 of protein S100-A8

After the identities were confirmed for both samples, the searches were repeated including deimination of arginine as variable modification but any citrullinated peptide was retrieved. This was not surprising since arginine deimination causes a 0.984 Da mass increase and, if both modified and modified peptides are present, MALDI analysis is not able to separate them and search engine to discriminate their presence.
These preliminary results revealed the specific recognition of a protein at around 11 kDa by the sera of RA patients. Furthermore Western Blots performed with anti-H4 antibodies suggested that the protein recognized might be histone H4 and the results obtained with anti-H4cit and anti-MC strongly supported the hypothesis that not only histone H4 was recognized by that it was carrying deimination sites.
MS analysis confirmed in duplicates the presence of histone H4 but also highlighted the presence of protein S100-A8. It is a calcium binding protein endowed with pro-inflammatory and regulatory properties, highly abundant in cytosol of neutrophils and monocytes.[57] Although its levels are increased in RA sera, it has never been described as target of autoantibodies and it might be assumed that its presence in the band was due to a co-migration with histone H4 since they both share a similar molecular weight.[58]

1.3.2 Derivatization of citrullines with 2,3-butanedione and antipyrine

MS analyses performed on a MALDI-TOF instrument clearly identified two proteins but no information regarding their deimination status were obtained. Comparing the different strategies described in literature for the detection of citrulline residues the derivatization technique with 2,3-butanedione and antipyrine was considered the most appropriate for our samples. The reaction involves two consecutive steps: first the 2,3-butanedione reacts with the ureido group to form an intermediate which then, in presence of antipyrine, evolves into a final adduct having a mass increase of 238.11 Da (Figure 1.14).

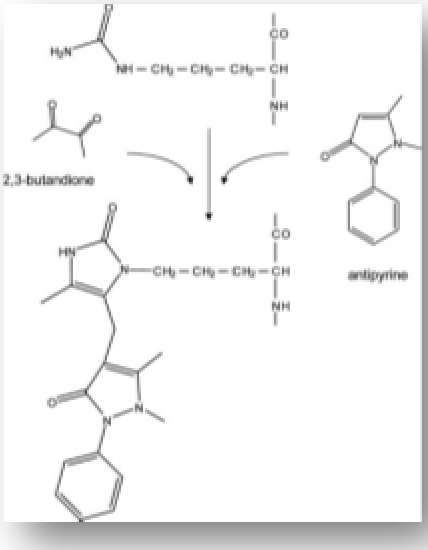


Figure 1.14. Modification of the ureido group of a citrulline residue by 2,3-butanedione and antipyrine[31]

The reaction is extremely specific for citrulline residues when led in acid conditions that block the reactivity on the guanidinium group of arginine since in basic conditions the reaction is described specific for arginine. In the original work by *Holm et al*, it was observed that no side products were generated except for two signals at minus 1 Da and minus 2 Da from the derivatized mass, in the MALDI-TOF spectra of derivatized citrullinated peptides.[30] Employing LC-ESI-MS analysis it was proved that such signals were artifacts generated during the MALDI ionization process since no corresponding chromatographic peaks were found, evidence that they were not chemical species produced during the reaction. Such signals however contribute to a characteristic isotopic peak of a derivatized citrullinated peak serving as an interesting marker.

In a further study it was demonstrated that citrullinated peptides, derivatized with 2,3-butanedione and antipyrine, once submitted to CID produce a worst fragmentation that impair the quality of the signals in the spectrum. In fact only a major signal is observed at 201.1 Da corresponding to an antipyrine fragment detached from the peptide.[31] This might be another specific ion signature of citrullinated derivatized peptides but its generation results in a lack of b- and y-ions, greatly compromising peptide identification.

In this thesis, in order to confirm the efficacy and specificity of the reaction the same experimental conditions described by *Holm et al* were used during an optimization phase with three citrullinated and not citrullinated standard peptides.

Corresponding MALDI-TOF spectra were acquired. For all the citrullinated peptides tested, the reaction was proved to be effective and the typical isotope pattern previously described was observed. As example is shown a citrullinated peptide with

MH^+=2218.730 that after derivatization acquired the increment of 238.11 Da showing a signal at 2456.590 m/z and a minus 2 Da signal at 2454.570 m/z (Figure 1.15).

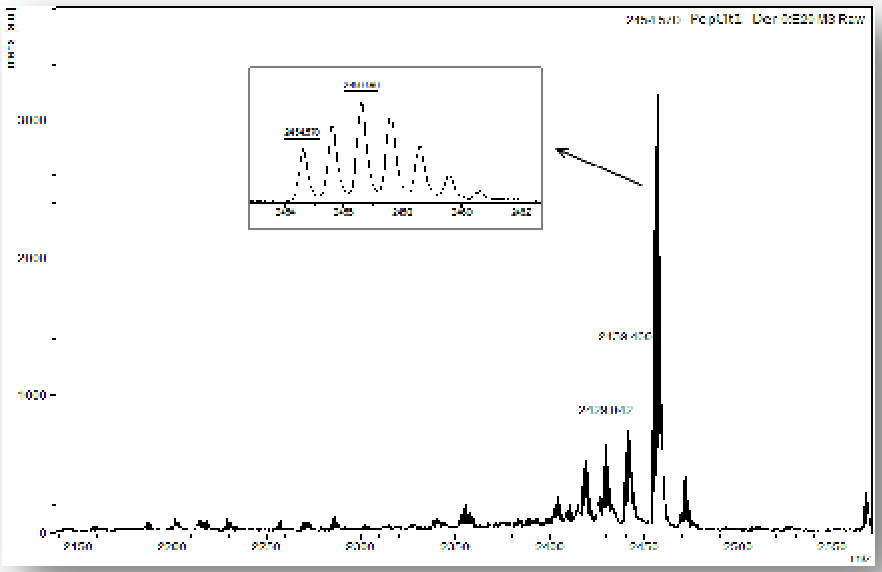


Figure 1.15. MALDI-TOF spectrum of a derivatized citrullinated peptide with original MH^+=2218.730

The reaction was proved to be specific when tested on several not citrullinated peptides primary on those containing arginine confirming that the experimental conditions were reliable to tag citrulline residues. The reaction was, furthermore, performed with decreasing concentrations of peptides from 1 nmol until 10 pmol and the MALDI signals were still clearly visible at this final concentration. It was important to check the detection limit of the instrument because the *in vivo* citrullinated peptides are supposed to be much less concentrated than 10 pmol and the results suggested that it was possible to detect even less concentrated peptides.

It was interesting to note that the shape of the derivatizated citrullinated peptide slightly changed if decreased concentrations were used. The minus 1 and 2 Da signals exhibited different intensities, in some cases higher than the plus 238.11 Da signal. As example is reported the spectra of a citrullinated peptide that acquires a mass at 2537 m/z after derivatization, at different concentrations. At 10 or 100 µM the pattern of the signals was as expected while at 1 µM it was almost covered by the minus 1 Da signal (Figure 1.16). As consequence the isotopic shapes of the same peptide were completely different.

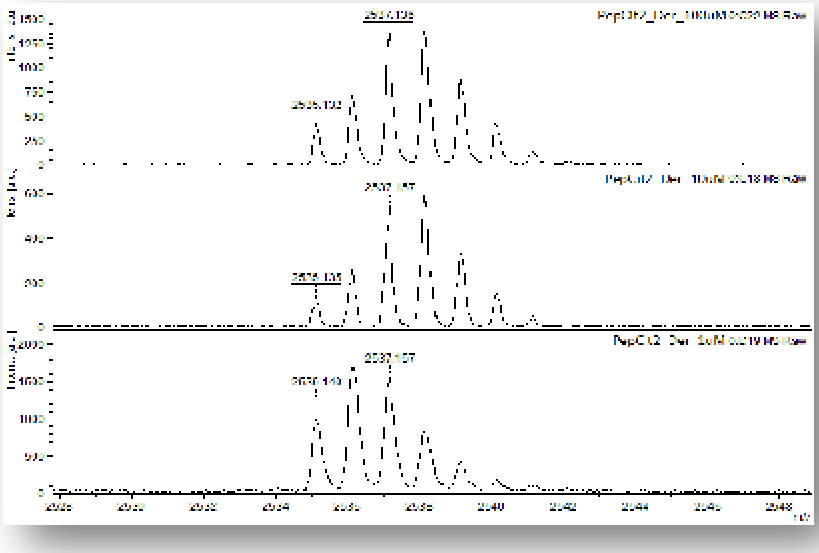


Figure 1.16. Different shape of the isotopic pattern of a derivatizated citrullinated peptide at different concentrations

Several mixtures of citrullinated and not citrullinated peptides at different concentrations were also tested confirming that the reaction was effective and specific even in more complex conditions.

Additional optimization experiments were performed spiking 10 pmol of citrullinated peptide inside a protein tryptic digest trying to model a situation as much similar as possible to the expected in biological samples. Once MALDI spectrum was acquired, the low concentrated derivatizated citrullinated peptide became hardly visible, lost in the midst of a large amount of different and high intense peptides.

It was evident that at low concentrations a purification step was necessary before the MS analysis. At first a SpeedVac concentration was performed reducing the volume up to 10 µl and then a purification step with Ziptip® was done to efficiently remove the reagents and concentrate the sample. This time, once the spectrum was acquired, the overall intensity of the peptides was enhanced and the signal of the derivatizated citrullinated peptide was tri-fold increased.

All the reactions performed on the standard peptides confirmed the reliability of 2,3-butanedione and antipyrine as reagents able to selectively and efficiently tag citrulline residues. An optimization of the experimental conditions, including a proper sample preparation for the subsequent MALDI-TOF analysis, was found to be crucial in order to delineate the better conditions for the analysis of the tryptic digests of the neutrophils nuclear extracts.

1.3.3 Identification of a citrullinated histone H4 peptide in nuclear extracts

The derivatization technique with 2,3-butanedione and antipyrine was applied to the two samples from the reactive bands of neutrophils' nuclear extracts and MALDI-TOF acquisitions were performed for both of them (Figure 1.17).

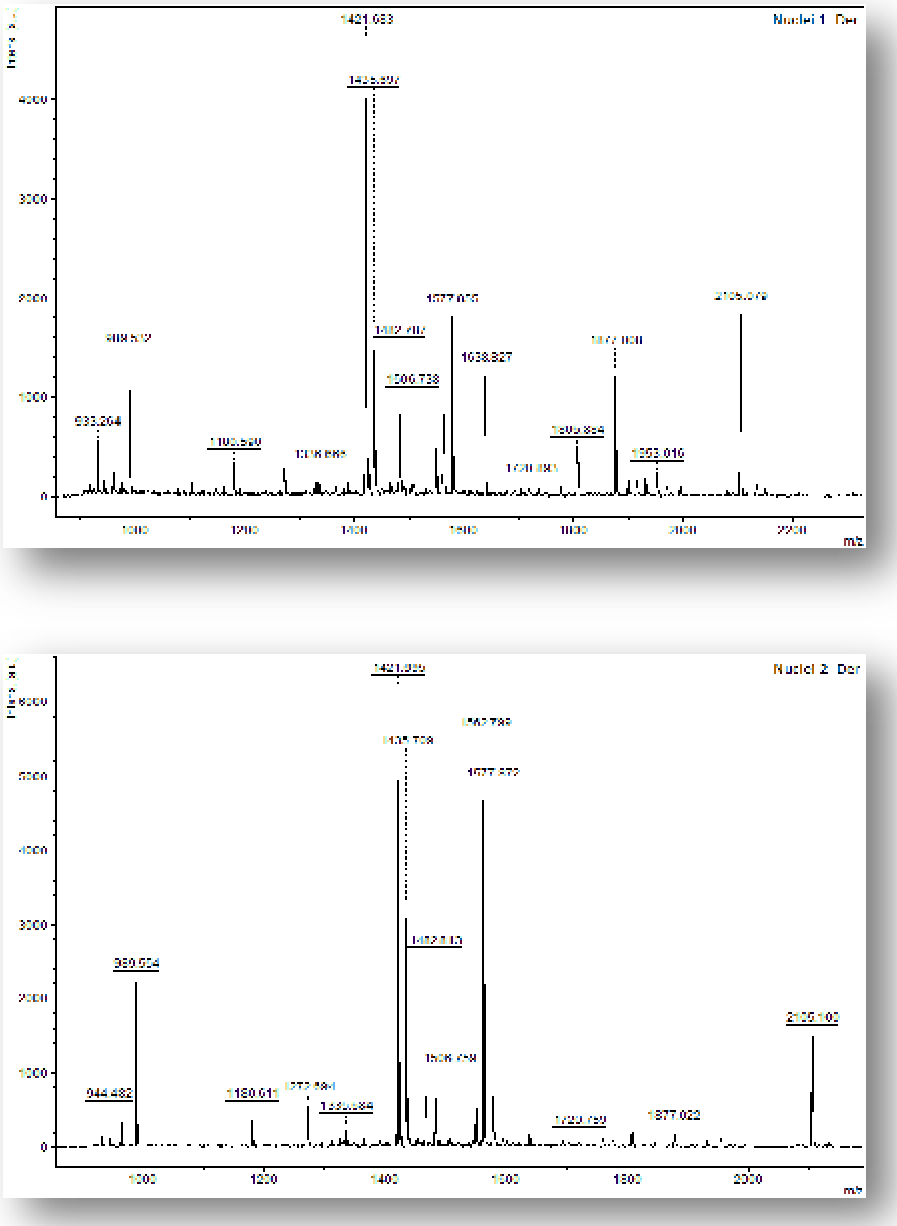


Figure 1.17. MALDI-TOF spectrum of the tryptic digest from sample "nuclei 1" (above) and sample "nuclei 2" (below)

After a careful manual investigation of both the spectra, signals were observed at 1720.893 m/z for one sample and at 1720.759 m/z for the other, that were absent before the derivatization (Figure 1.18).

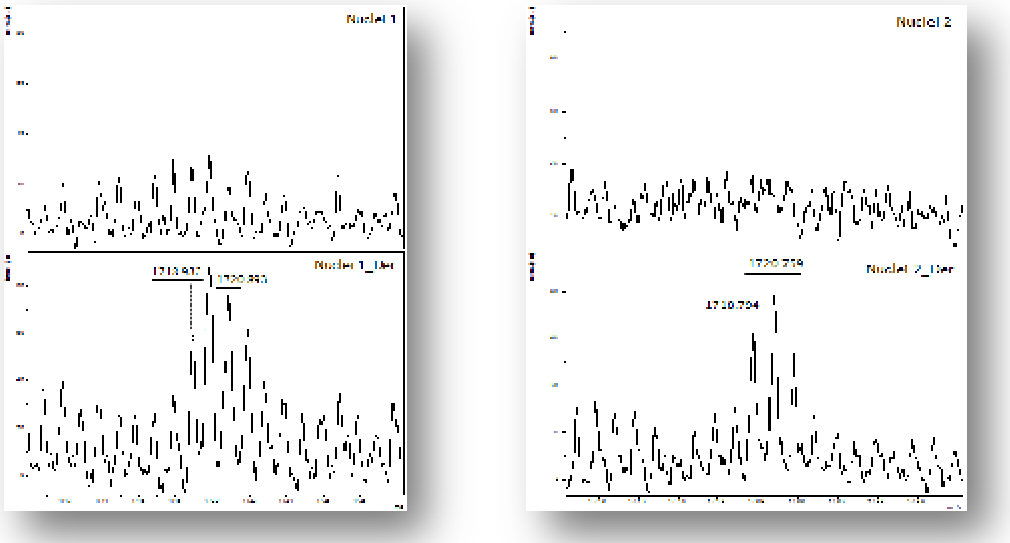


Figure 1.18. MALDI-TOF spectra of the tryptic digest from "nuclei 1" (on the left) and from "nuclei 2" (on the right). For each sample a comparison is shown between the zoomed spectra in correspondence of the signal at 1720 m/z, before(above) and after(below) the derivatization

These signals were compatible with a citrullinated peptide of mass 1482.703 m/z that achieves a mass of 1720 m/z after the derivatization step responsible of the 238.11 Da mass increase. The signals at 1720 m/z exhibited the typical isotope pattern of a derivatized citrullinated peptide with the minus 2 Da signal at 1718.933 m/z and 1718.794 m/z.

Through theoretic digestion of histone H4 it was possible to discover the existence of a possible citrullinated peptide with sequence DNIQGITKPAIRR and mass of 1482.814 Da which might bear one deiminated arginine. Unfortunately the very low intensity of the signal prevented the precursor to be selected for subsequent fragmentation and confirmation of the peptide sequence was not obtained. The peptide sequence showed two arginine residues liable to deimination and simply from the mass value of the precursor it was not possible to assess which one contained the modification, requiring other strategies to be used.

On the whole these MS data confirmed that H4 was deiminated, as initially highlighted by the Western Blot analysis and supported the hypothesis that RA serum had recognized this deiminated protein, among the neutrophils' nuclear proteins. Indeed even if protein S100-A8 contains two arginines susceptible of deimination corresponding citrullinated peptides were not detected after derivatization.

1.3.4 Identifications of deiminated histone H4 contained in NETs of activated neutrophils

After the preliminary results achieved with the nuclear extracts of neutrophils activated to produce deiminated histones, the attention was focused on the NETs that are structures containing a fraction of all the nuclear proteins.

Granulocytes were isolated using the same approach previously described based on a gradient separation of blood on polysucrose solutions. Granulocytes cultures were *in vitro* stimulated with PMA to induce NETosis; proteins were isolated by DNase I digestion that allowed the release of the neutrophils proteins after having minced the DNA filaments of NETs. Proteins were acid precipitated, fractionated on TRIS Tricine SDS-PAGE and transferred to PVDF to perform the immunodetections with RA and control sera and purified IgG from RA and control sera.

Interestingly, RA immunoglobulins as well as RA sera decorated bands around 11kDa and anti-deiminated H4 antibodies recognized a band at the same level; normal subject sera and immunoglobulins used as negative control were not reactive (Figure 1.19).

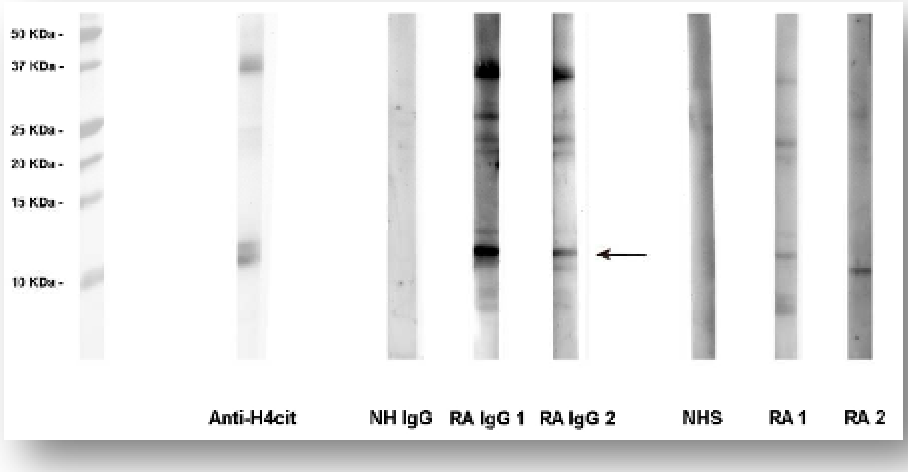


Figure 1.19. RA sera and purified RA IgG binding to proteins from NETs of stimulated neutrophils. RA1 and RA2=sera from RA patients; NHS=normal healthy subject; RA IgG1 and RA IgG2= IgGs purified from RA patient's blood by means of affinity chromatography on sepharose coupled to protein-G

The band of interest was excised from different lines in order to enhance the protein quantity available for the trypsin digestion; the derived peptide mixture was analyzed by MALDI-TOF MS to acquire the corresponding PMF spectrum (Figure 1.20).

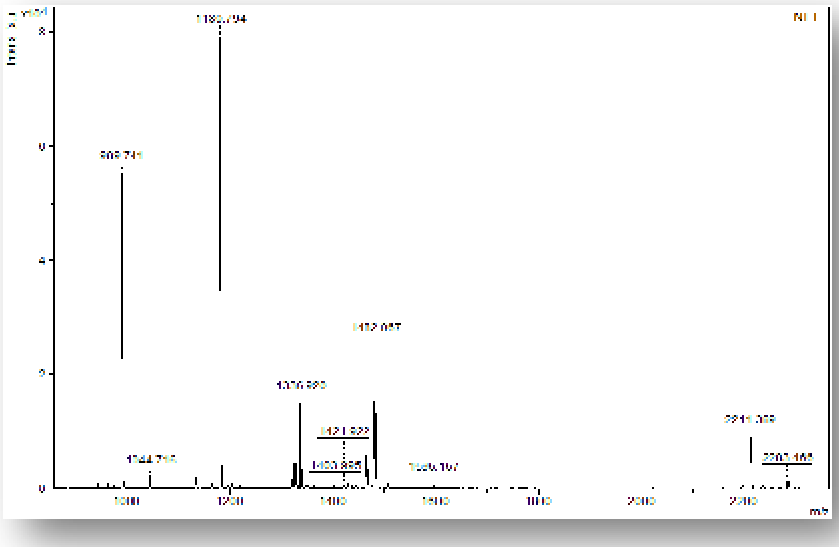


Figure 1.20. PMF MALDI-TOF spectrum obtained from the excised pooled bands from NETs after tryptic digestion

A databank search was submitted setting the same parameters used in the previous search. It clearly returned the identification of histone H4 with a score of 76 which was widely superior than the threshold of 56 (Figure 1.21) and 6 peptides were correctly identified allowing a protein coverage of 40 % (Figure 1.22).

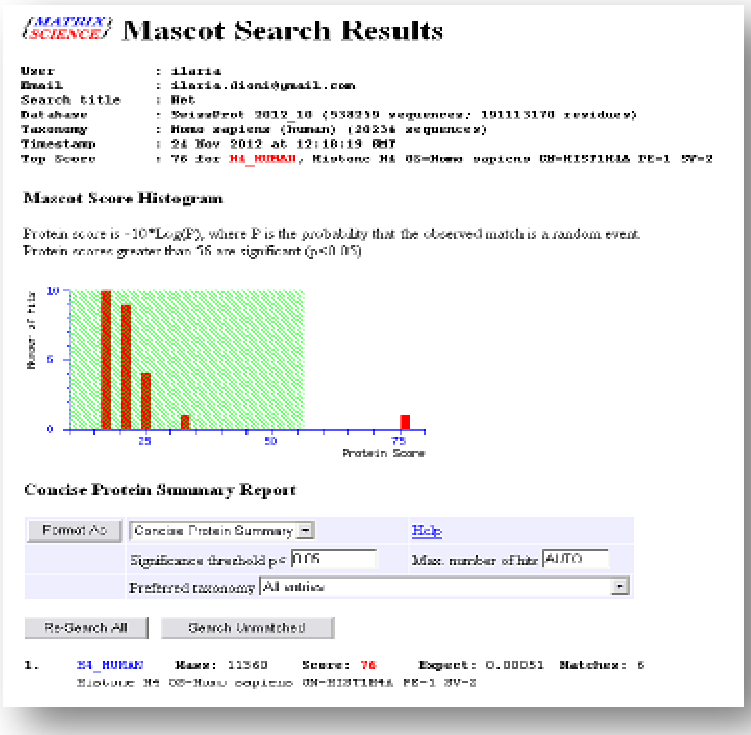


Figure 1.21. Databank search of NETs sample

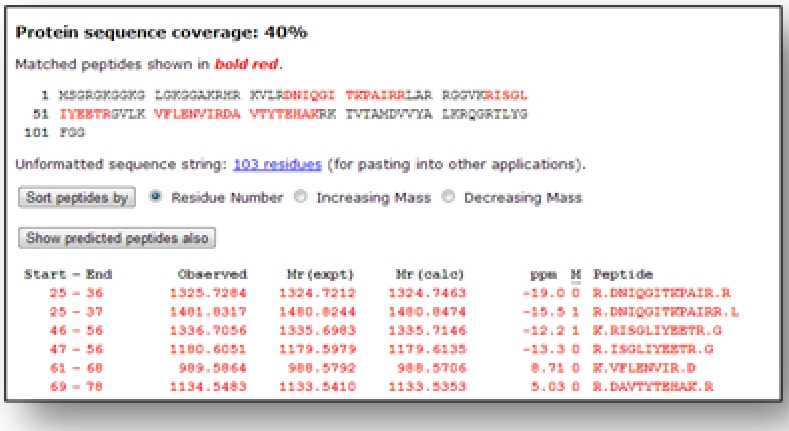


Figure 1.22. Protein view of histone H4, for NETs sample.

Even in this case some peptides have been selected for the fragmentation even if the PMF was clearly compatible with histone H4; as example is reported the MS/MS spectrum of precursor MH^+= 989.700, corresponding to the sequence VFLENVIR included between residue 61 and 68 (Figure 1.23).

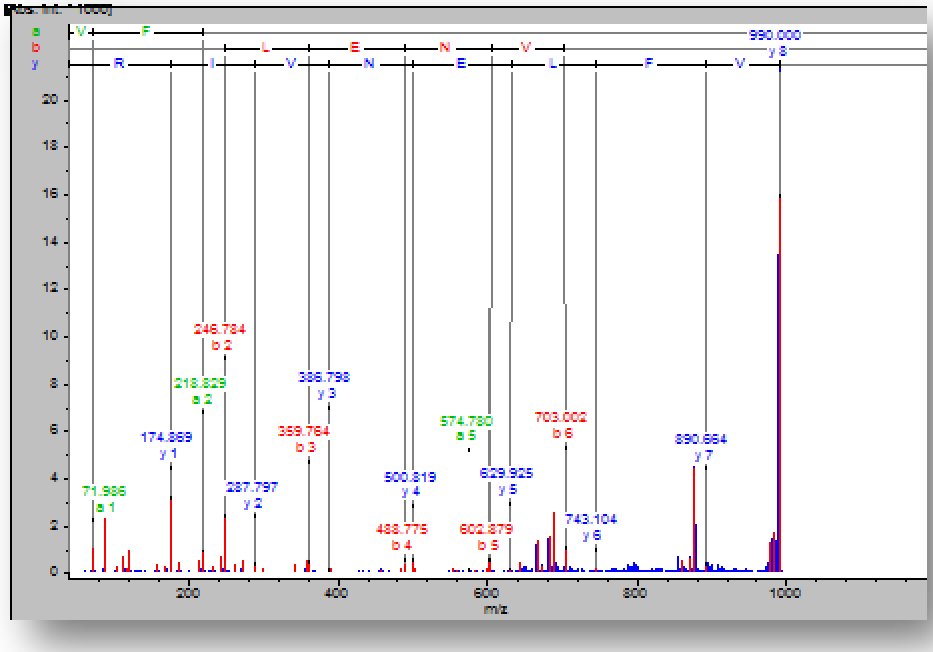


Figure 1.23. Precursor at 989.700 Da corresponding to VFLENVIR sequence contained between residues 61-68 of histone H4

In order to clearly demonstrate the presence of deiminations in the histone H4 recognized among the NETs proteins, the same derivatization strategy with 2,3-butanedione and antipyrine was applied to the corresponding tryptic digest.

The sample, prepared in the proper optimized way, was analyzed through MALDI-TOF MS (Figure 1.24).

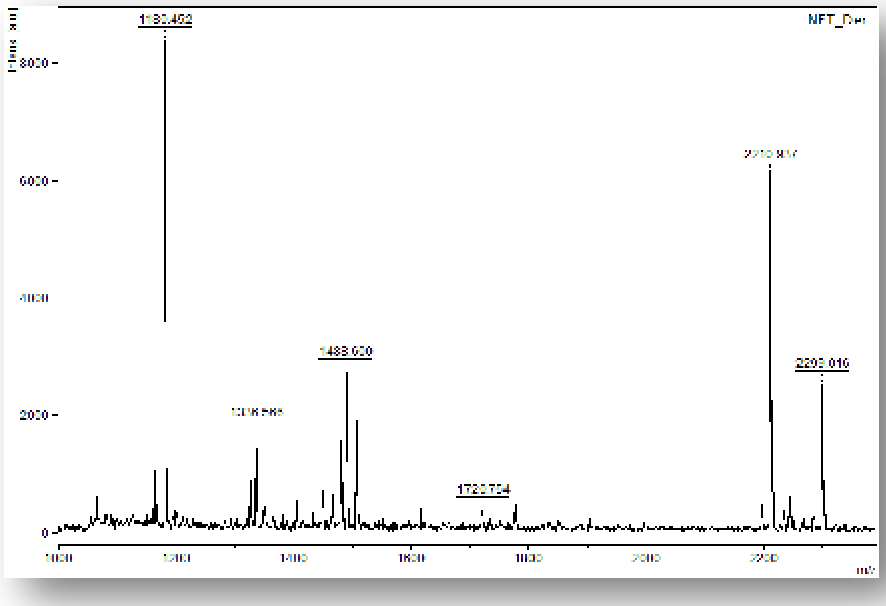


Figure 1.24. MALDI-TOF spectrum of the tryptic digest from NETs sample

Once again a signal at 1720.754 compatible with the peptide of sequence DNIQGITKPAIRR was found and the characteristic isotopic pattern with a signal at 1718.875 was observed (Figure 1.25).

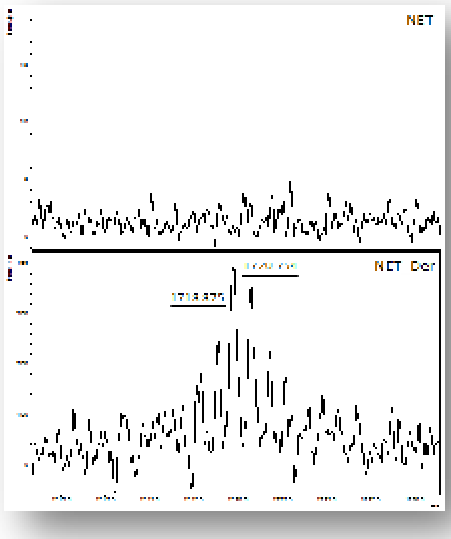


Figure 1.25. Comparison of the MALDI-TOF spectra of the tryptic digest from NETs, zoomed at 1720.754 m/z, before(above) and after(below) the derivatization

The MS data confirmed that deiminated histone H4 was the protein contained in the band from NETs sample, recognized by RA patients' sera, purified IgGs from RA sera and commercial anti-cit H4 antibodies. In order to validate the reactivity of RA sera with H4, human purified H4 was *in vitro* deiminated by PAD4 and used as an antigen to measure IgG antibodies by ELISA assay. The binding of RA sera to deiminated H4 was significantly higher than the binding of NHS ($p<0.005$). On the contrary, no difference was observed with unmodified H4 tested with RA and NHS sera.

Concerning the identification achieved by the processing of mass spectrometry data it is interesting to note that with NETs proteins no contamination from protein S100-A8 was observed. NETs proteins constitute a subgroup of all the nuclear proteins and the acid extraction had provided a still more histones-enriched sample, reducing the contamination from other proteins.

It is also noteworthy that the proteins obtained from NETs produced slightly Coomassie-stained bands, difficultly able to produce reliable MS spectra; therefore several co-localized bands from different lanes of the same sample were needed to be pooled, ensuring a better yield of tryptic peptides.

The detection of citrulline residues by means of the chemical labeling strategy had allowed the individuation of a deiminated peptide from both of the samples, nuclei and NETs. This peptide had a mass of 1482.8 m/z that resulted compatible with the tryptic peptide DNIQGITKPAIRR included in the histone H4 sequence flanked by residues 25 and 37. The derivatization technique had not distinguished which was the carrier of the modification, between the two arginines, therefore other strategies were required to find it out.

1.3.5 Direct detection of citrulline residues using LC-MS/MS strategies

Peptide DNIQGITKPAIRR from histone H4 bears two arginines liable of deimination, representing the C-terminal of the peptide and its adjacent residue. It would be quite easy to discriminate the modification site simply considering the ability of trypsin to cleave deiminated arginines. Indeed if the cleavage was not allowed, the C-terminal position would be immediately discarded as possible modification site while in the contrary case deeper investigations would be necessary.

Concerning this topic an explorative investigation was led performing a trypsin digestion on two couples of standard peptides containing a citrulline residue or not and in both cases the modified arginines were not cleaved.

The signals at 1189.629 m/z and 844.470 m/z (Figure 1.26; top spectra) belonged to the two unmodified peptides cleaved at the arginine residue. Signals at 1190.613 m/z

and 845.454 m/z, expectable from their corresponding citrulline-containing peptides after the trypsin digestion, were not present in the spectra (Figure 1.26; low spectra).

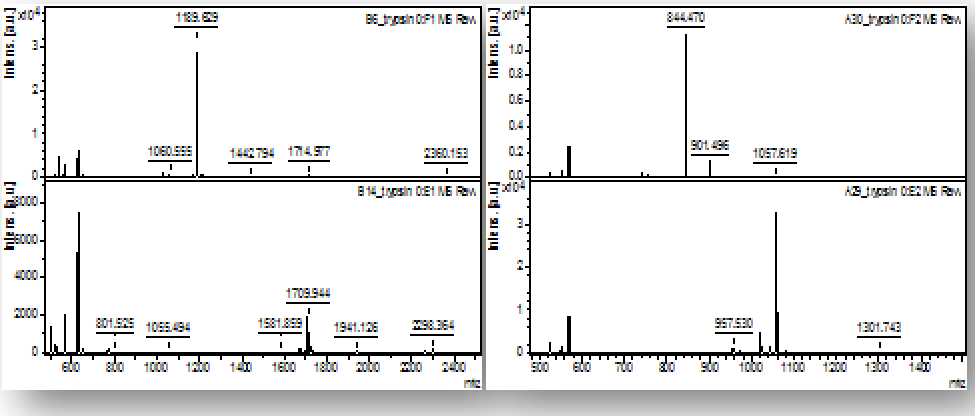


Figure 1.26. MALDI spectra of the two pairs of trypsin-digested peptides bearing a citrulline and not. Both of the spectra at the top belong to an unmodified peptide and the underlying spectra to their corresponding citrullinated form

These results suggested that trypsin does not cleave deiminated arginines therefore C-terminal citrullines should not exist. However these considerations would require the support of more systematic analyses, to be reliable. Unfortunately there is no trace of scientific literature systematically facing this topic but there are, on the contrary, conflicting opinions regarding it so that this shortcut cannot be used to map the citrullination sites and other strategies need to be explored.

The scenario of the available proteomic technologies developed to highlight the presence of citrulline residues has been described in the introduction of this thesis. For our purpose LC-MS/MS analyses appeared as the most immediate and the most proper approach to characterize the deimination site of a peptide because the peptide sequencing allows a direct localization of the expected increment at amino acidic level.

Anyway also the problematic aspects that could rise from such strategy were considered and weighted.

Firstly, if considering that the two arginines liable of deimination are adjacent residues in the peptide sequence, the localization of the 0.984 Da increment at amino acidic level appears not so easy because the discrimination is limited on two single fragment ions. If position 36 represents the deimination site the b_{12} ion will be the discriminating signal whereas if residue 37 is the citrulline then the y_1 ion will be the signal to consider.

The second crucial aspect to be evaluated is the presence of the unmodified peptide and its possible interference in the characterization of the deimination. It has to be considered that PTMs usually spread into a very narrow fraction of the entire protein

population. Also for the deimination, the *in vivo* incidence of citrullinated proteins can be assumed to be low and, consequently, unmodified peptides arising from not citrullinated proteins not only are supposed to exist but also to be the widest fraction. Having established this aspect it is, now, important to understand if their presence could be a problem, above all considering that they are tryptic peptides. Whether trypsin proteolysis does not occur at the C-terminal of citrullines, deiminated peptides shall be completely different from their unmodified version because they will always be longer peptides due to the presence of the missed cleaved deiminated arginine. The sequence of unmodified peptide will be different from the citrullinated one, not representing a problem. If, instead, the cleavage of citrullines is allowed, two forms of a same peptide sequence could be awaited and this constitutes an aspect to be taken into account.

Considering a real situation such as the one offered by the samples of interest in this project, namely the tryptic digests of the gel bands of the nuclear extracts and the NETs extracts, it became evident that the experimental conditions were even more complex than the theoretical ones.

Indeed MALDI spectra of the samples and databank searches previously performed on corresponding PMFs, had disclosed the presence of the unmodified form of the deiminated peptide DNIQGITKPAIRR having the exact amino acidic sequence of its modified version and a decreased mass of only 0.984 Da. Worthy to note was that the unmodified peptide DNIQGITKPAIRR was the result of a missed cleavage of trypsin whose under 100% proteolytic efficiency is a well documented feature.

The presence of the two peptides differing for that slight mass were argued by the skewed shape of the corresponding integrated MALDI signal. Indeed the monoisotopic peak corresponded to the monoisotopic of the unmodified peptide and the M+1 peak was the result of the M+1 peak of the unmodified summed up to the monoisotopic of the citrullinated one. According to the natural abundance of the elemental isotopes, each peptide isotope pattern is mass-dependent and the deriving peak shape is predictable. The peaks observed were symptomatic of overlapped signals because the M+1 peak was more intense than the monoisotopic one and this pattern was anomalous for that mass range.

Summing up, in the spectrum of the digested band coming from neutrophils' nuclear extracts the M+1 peak contained contributions from both the monoisotopic peak of the citrullinated peptide (1482.8) and the M+1 peak of the unmodified (1481.8 + 1) (Figure 1.27). In the spectrum of the NETs the situation was even more complicated due to the presence of an unexpected signal at 1481.072 that summed up to the other two and that did not correspond to any histone H4 theoretical tryptic peptides.

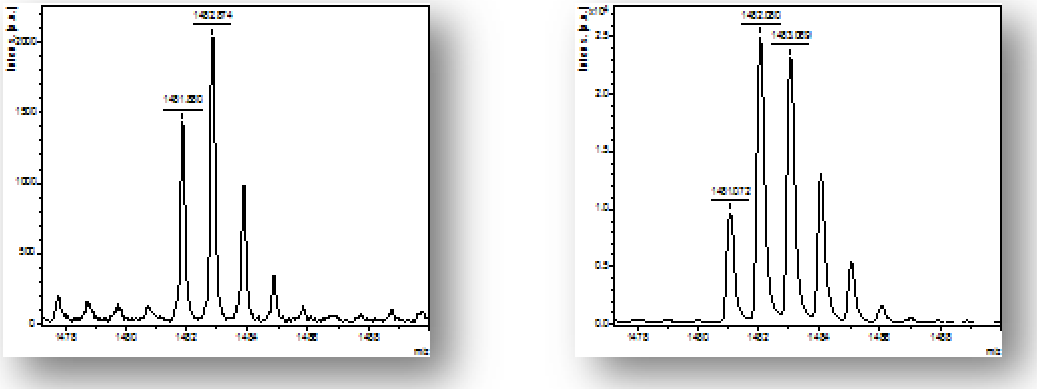


Figure 1.27. MALDI spectra zoom around the peak of the peptide of sequence DNIQGITKPAIRR from digested band of one of the two samples from the nuclear extract (on the left) and NETs (on the right)

These evidences pointed out that a collateral event, trypsin missed cleavages, was greatly enhancing the complexity of the topic by having produced peptides with similar chemical properties and slight mass difference.

Such situation imposed new consideration to be formulated while approaching the LC-MS/MS characterization of citrullines. Firstly a complete chromatographic separation of the two peptide forms was mandatory otherwise the mass spectrometer would have operated the co-selection of both precursors, for the subsequent fragmentation. Indeed the m/z window used for data-dependent analysis is not narrow enough to exclude one isotopic pattern from the other.[29] Since peptides bearing the citrullination are supposed to be a minority fraction compared to the unmodified ones, if no separation is achieved the MS/MS spectra of the citrullinated would be contaminated by the presence of signals of the unmodified peptide. If there are misleading signals, besides more intense and able to cover the citrullinated-peptides' ones, the deimination site's characterization by means of a direct detection of the 0.984 Da increment would risk to be heavily undermined. Given that the MS/MS spectra represent the final tool to extract the information to discriminate the mass increment, the entire procedure would be dramatically invalidated.

Altogether these considerations highlight that the characterization of the modification site is a rather challenging issue; an accurate work of spectral interpretation is mandatory and a specialized equipment able to ensure high analytical sensitivity and high data confidence is fundamental.

Another topic interesting to be more extensively analyzed in our samples was the citrullination pattern of histone H4 in order to discover other possible modification sites whom the derivatization techniques did not point out. This aim is of great importance because the discovery of other possible deiminated peptides within histone H4 means

pointing out other possible citrullinated epitopes within the protein and such information can help a better qualification of the interaction with RA antibodies to be delineated.
LC-MS/MS analyses provide more information than MALDI analyses for what concerns the complexity of a sample, by allowing the detection of low abundant proteins or low abundant peptides as the citrullinated are, because the methodology itself is more sensitive.
In proteomics there is a huge need to reach a high analytical sensitivity in order to cope with quite frequent problematic aspects: limited sample amounts to handle, wide concentration variations of the molecules to analyze and specific targets to detect among an high complex sample (i.e low abundant deiminated peptides as in this context). A *nano*LC-*nano*ESI configuration of the LC-MS/MS analyses is an important tool to fulfill this request since the electrospray ionization mechanism is concentration dependent.[59] This means that reducing flow rates and the dimension of the system dramatically improve the sensitivity and, consequently, the detection range, that can be pushed up until sub-femtomole concentrations (Figure 1.28).[60]

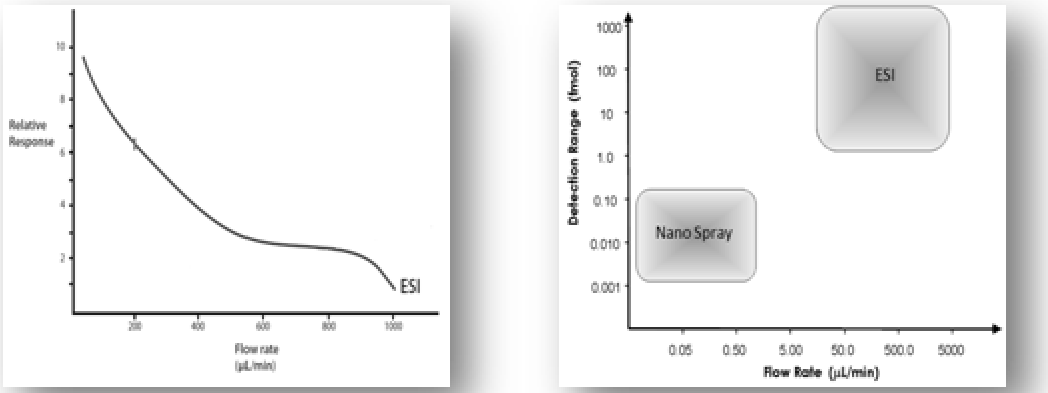


Figure 1.28. Change in response with incoming sample flow rate for LC/MS interface (on the left); flow and sensitivity gap between *nano* ESI and ESI (on the right)

Indeed the charged droplets produced by the spraying process pander to subsequent fission events producing always smaller offspring droplets and the fission goes forward until a lowly charged residue is formed.[61] At each fission a fraction of the ionizated analytes are set free but, unfortunately, another is lost in the residue of the initial droplet therefore the total ion yield is heavily affected. Instead in presence of flow rates smaller droplets are produced, the number of fission processes required to form mass spectral ions is reduced ("predominant fission pathway" theory) and so the loss of analytes is eased. Furthermore this process reduces the concentration of salts in the primary droplets available for MS analyses thus reducing the ionic suppression effect

they are responsible of.[62,63] These aspects explain the very higher sensitivity of a *nano* configuration compared to standard analytical experimental conditions.

Toward the purposes of this project a high performing instrumentation was required and an UltiMate® 3000 *nano*HPLC system coupled to a LTQ-Orbitrap mass spectrometer equipped with a *nano*ESI source constituted a proper platform. The *nano* flow configuration and the precursors accurate mass detection provided by the Orbitrap analyzer ensure the system to be sensitive enough to supplement the information gained through the chemical labeling and hopefully deepen the characterization of the histone H4 citrullination pattern.

LC-MS/MS analyses have been performed on a standard sample represented by a couple of a citrullinated and an unmodified peptide and on the tryptic digest coming from the NETs extracts that constitute the specific biological sample to be investigated. Two different set-up have been tested, namely reversed-phase liquid chromatography (RP-LC) and hydrophilic interaction liquid chromatography (HILIC), in order to evaluate their ability to provide the best experimental conditions to achieve the predetermined aims.

1.3.5.1 Analysis on standard peptides

Since a direct detection of deimination using the mentioned instrumental configuration strongly recommends a good chromatographic separation of citrullinated peptides from the quite similar unmodified versions an optimization of the method is necessary to maximize the selectivity. In order to be as much selective as possible several factors should be handled such as the nature and the composition of the eluent, the gradient used for elution and the temperature.[64] However these factors require a systematic investigation to reach the ideal experimental conditions and a great amount of sample to be kept available for the analyses but enzymatic digestions of bands usually provides few tens of sample.

Nevertheless, the selectivity is largely based on the stationary phases; each phase exhibits an own chemistry that makes possible interactions to occur on its surface. Since the sort of bond that analytes establish is dependent on both their reactive groups and the stationary phase's ones, different supports can provide significantly different chromatographic behaviors by stimulating specific chemical reactivity and evoking different mechanisms of separation. These aspects are relevant because they finally indicate that a phase can be more selective than others for specific molecules, inherently being more proper to separate them.

Taking into account all these considerations, the separation of deiminated peptides from their corresponding unmodified version had firstly implied an evaluation of their different chemical properties in order to find the way to amplify them as much as

possible. Since arginine residues exhibit a positively charged guanidinium group that, after deimination, becomes a neutral ureido group, the challenge was to find the condition to maximize such a slight difference.

RP-LC and ZIC-HILIC represents two orthogonal chromatographic separation methods (Figure 1.29) due to the opposed characteristics of their stationary phases that allow an inverted order of elution of hydrophobic/hydrophilic molecules. In RP the more hydrophilic components are the first eluted while in HILIC the opposite happens. Actually, in RP-LC analytes are separated by means of hydrophobic interactions according to combined processes of partitioning and adsorption of the solute molecules at the surface of the hydrophobic solid support, in presence of a mobile phase with increasing concentrations of organic modifier for the gradient elution.

HILIC technology, instead, combines a normal-phase type of separation and uses an increasing water gradient for the elution. The surface of the HILIC hydrophilic phases, in presence of a water-deficient mobile phase, forms a water-rich layer where the analytes can absorb; this way a sort of liquid/liquid extraction system is created so that analytes are distributed between the two layers. In addition to these partitioning events important contributions are constituted by hydrophilic interactions arising from hydrogen bonding as well as dipole-dipole attractions and ionic forces. Compound's polarity joined to the degree of solvatation of the surface of the phase constitutes the crucial aspects of the retention mechanisms.

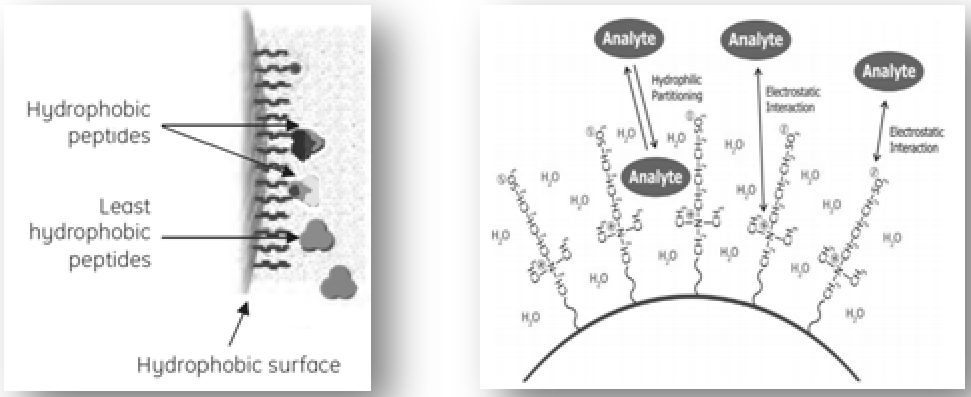


Figure 1.29. Interactions on RP stationary phase (on the left) compared to HILIC's ones (on the right)

For HILIC separations different types of polar stationary phases are available, each chemically bounded with specific polar functionalities which provide various retention capacities and selectivity to polar compounds. Basically three groups can be defined: neutral, charged and zwitterionic phases.

The first type does not engage electrostatic interactions, turning out to be as the less selective because of the failed exploitation of such forces that, instead, play an important role in the separation by contributing to the retention of polar analytes.
As opposite, the charged and the zwitterionic phases make use of strong and weak electrostatic interactions, respectively. The charged ones require high ionic strength to elute the anchored compounds but the presence of high concentration of salt ions weakens the following MS detection reducing its sensitivity. Instead zwitterionic phases are characterized by the presence of positive and negative charges placed in close proximity each other and at a stoichiometric ratio.[65] That chemical design allows weaker attractions to realize between the groups displayed on the material surface and the oppositely charged analytes so that an elution less invasive for the MS analyses is possible. Different types of zwitterionic phases exist, bearing distinctive selectivities toward polar compounds, according to the type and the spatial arrangements of the groups. These factors are responsible for the surface charge distribution on the phase and regulate the different interaction toward zwitterionic, neutral and charged compounds.
For our purposes, the attractive feature of HILIC technology was its polar selectivity that made it suitable for the separation of molecules differing for the nature of a polar group.
All these premises seemed to well fit with the intent of separation of a peptide from its deiminated form, since both the guanidinium and the ureic group had polar characteristics. Moreover the guanidinium shows a permanent positive charge that can be the anchor to selectively retain the unmodified peptide to the HILIC support through a ionic interaction that the citrullinated peptide is, instead, not able to support. This was the rationale of having chosen a HILIC stationary phase as alternative to a reversed-phase, in order to separate citrullinated peptides from the corresponding non-citrullinated form.
The zwitterionic HILIC stationary phases appeared as the most appropriate for the aim of this project; in particular, ZIC®-HILIC columns, bearing a betaine-sulfonate function with a positively charged quaternary ammonium and a negatively charged sulfonic group (Figure 1.30), were chosen for the analyses.

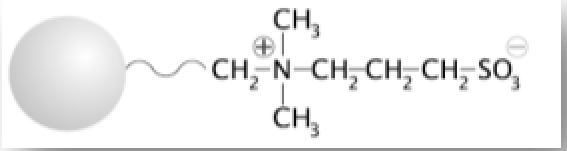


Figure 1.30. ZIC-HILIC bonded with sulfobetaine group

Taking into account all the previous considerations, RP and ZIC-HILIC® stationary phases were tested using a pair of a citrullinated and an unmodified peptide, as model for evaluating their behavior in presence of chromatographic orthogonal conditions.

For RP-LC an equimolar mixture of a deiminated peptide (Mr=2297.2088) and the corresponding not modified (Mr=2296.2248) was prepared in an aqueous buffer at a final concentration of 4.3 pmol/µl for each peptide. The miniaturized configuration of the system allowed the injection of just 1 µl of sample which first flowed to a C18 guard cartridge for the concentration step and then to the C18 column for the separation in a 30 minutes gradient. The LTQ-Orbitrap mass spectrometer was set up to work with the high resolution (HR) of the Orbitrap analyzer for the full scan acquisition while the precursor selection and the detection of the ion fragments were committed to the LTQ analyzer. The instrument operated in data-dependent mode performing the fragmentation of the three most intense precursors from each full scan.

The base peak chromatogram of the peptide binary mixture (Figure 1.31) showed a consistent chromatographic separation of the two molecules that exhibited retention times of 33.24 and 34.06 min. This result highlighted that the hydrophobic selectivity of the C18 stationary phase had been able to retain the citrullinated peptide for a sufficient time to realize a separation.

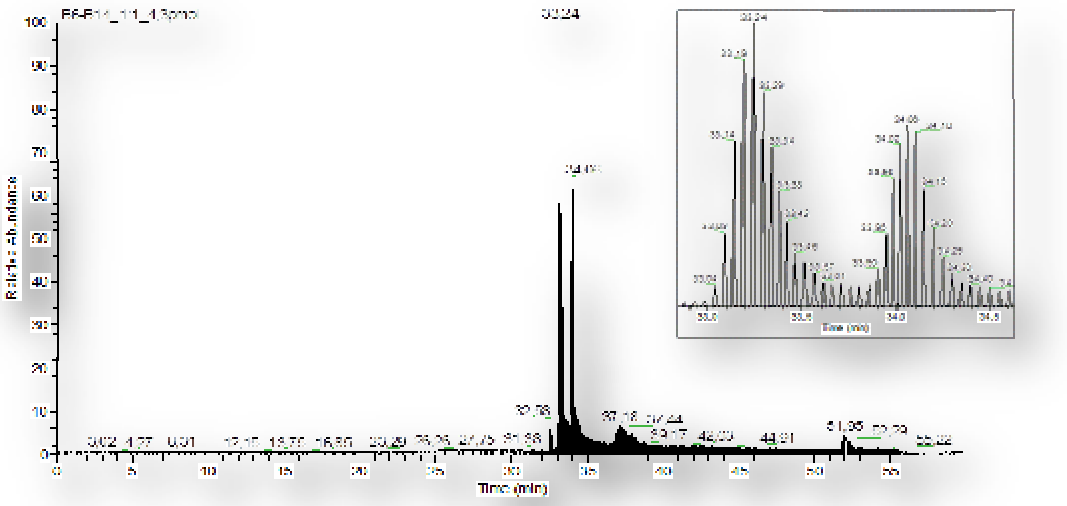


Figure 1.31. Base peak chromatogram of the two peptides under RP-LC and a zoom around their elution time

The HR mass detection of precursors (Figure 1.32) clearly assigned the first peak to the unmodified peptide (M^{3+}=766.41 and M^{2+}=1149.12) with the citrullinated one (M^{3+}=766.74 and M^{2+}=1149.61) eluted on tail of the other.

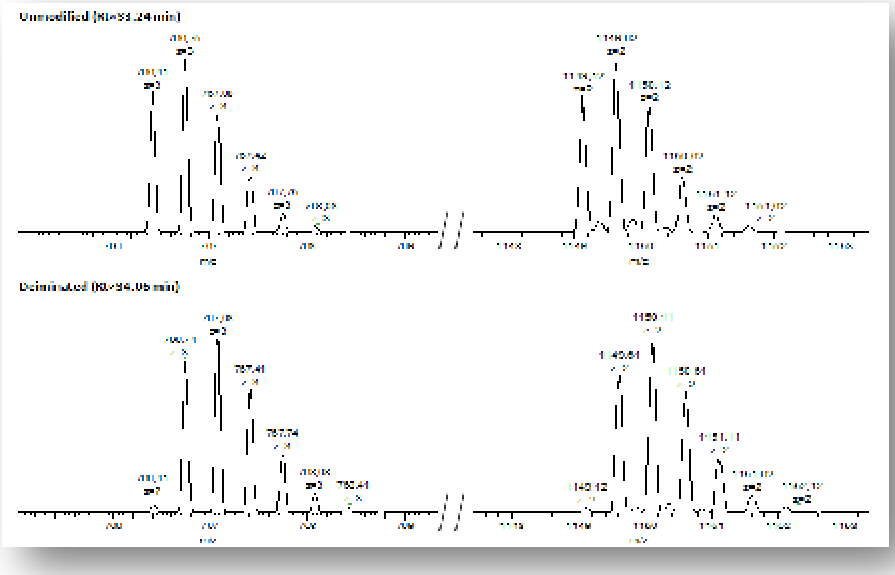
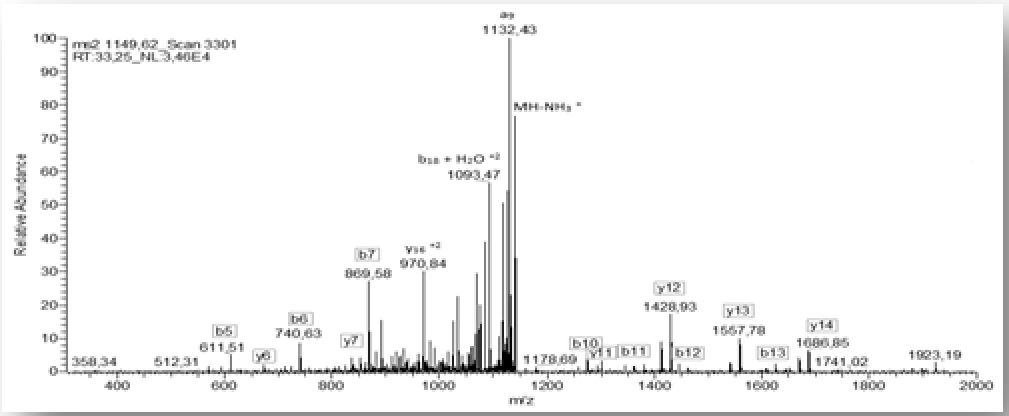


Figure 1.32. The HR-MS scan of the triple and double charged unmodified peptide (upper spectra) and the triple and double charged of the citrullinated (lower spectra)

The elution order agreed with the slightly lower hydrophobicity of arginine-containing peptides compared to the citrullinated ones resulting in a weaker interactions with the stationary phase and a faster elution from the column. However in the scans of the citrullinated peptide there was still a contribution from the unmodified peptide even if at a very low intensity, as evident from the presence of the signals at 766.41 Da and 1149.12 Da in the peak of the M^{3+} and M^{2+} of the citrullinated peptide. This indicated that the delay of around 40 seconds between the retention times of the two peptides was not enough to ensure a complete chromatographic separation.

Analyzing the CID spectra of two double-charge precursors of the two peptides (Figure 1.33) it was evident that citrullinated peptide had gone through a minimum backbone cleavage, with the major part of fragment signals mingling in the background noise.



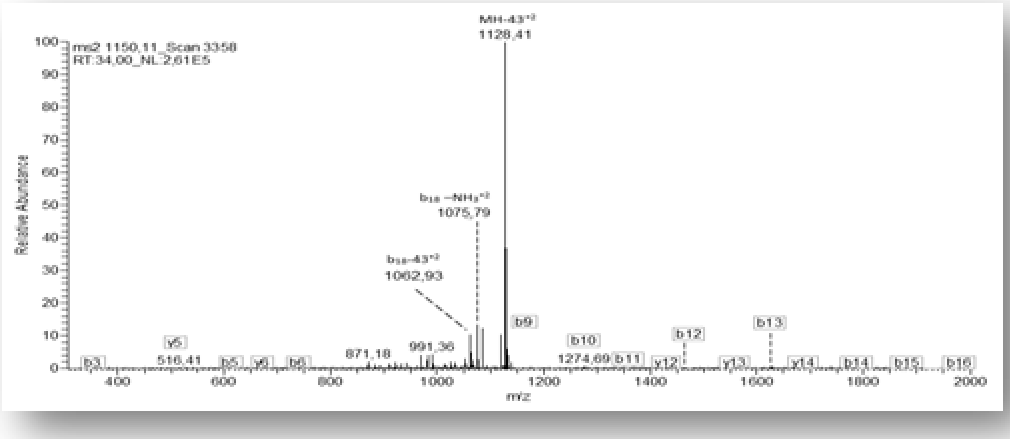


Figure 1.33 CID spectra of a doubly-charged precursor of the unmodified peptide (above) and of the citrullinated one (below)

This behavior aligned with *Hao*'s findings who first observed this phenomenon analyzing synthetic peptides by ESI tandem MS, as mentioned in the introduction of this thesis.[66]

He also discovered a specific rearrangement that takes place within citrullinated peptides during CID fragmentation. That rearrangement guides the release of isocyanic acid from the side chain of citrullines, through a process that competes with backbone fragmentation and that accounts for a specific neutral loss of 43 Da. In accordance to this evidence, CID spectra of our standard citrullinated peptide exhibited exclusive signals ascribable to b and y fragments with a 43 Da loss but they were really a few.

After RP analysis the same peptides were submitted to the HILIC analysis using a different preparation of the sample due to the opposite features of the method. Peptides were prepared in organic buffer containing 90% acetonitrile at the same final concentration of 4.3 pmol/µl.

HILIC columns require a quite long conditioning phase with high percentage of organic solvents because hydrophilic surfaces are proved to hold the water when exposed to mixtures of high concentrated organic solvent and low concentrated water. And the creation and maintenance of this water-enriched layer partially immobilized on the surface is essential because it constitutes the place where interactions with molecules realize and separation happens. Therefore an extended chromatographic method has been created including a conditioning phase of 50 min at the end of each run, to re-establish the initial condition of the column.

Peptide mixture analyzed this way produced a good separation of the two forms as evident from the corresponding base peak chromatogram (Figure 1.34) where the two main peaks at 64.86 and 71.10 minutes corresponded to the two forms.

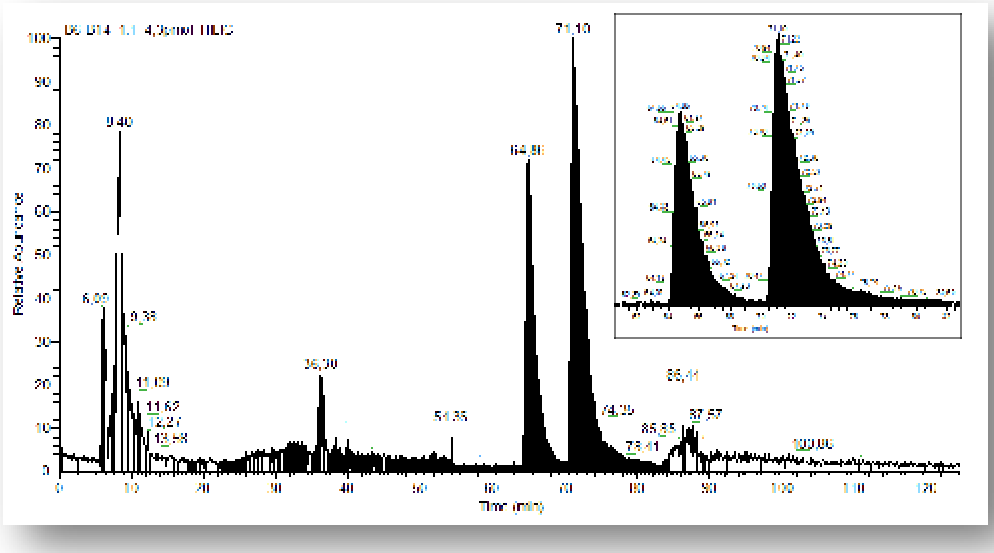


Figure 1.34. Base peak chromatogram of the two peptides under HILIC conditions and a zoom around their elution time

This stationary phase was able to produce a wider separation compared to the RP, with a temporal delay of more than 6 minutes between the first eluting peak that belonged to the citrullinated peptide and the second to the unmodified. Both of the signals presented a marked peak tailing which is a common drawback of HILIC technology probably due to the double mechanism of retention (partitioning and electrostatic interactions) which analytes are submitted to and which is responsible for a non homogeneous behavior toward the stationary phase. CID spectra of both peptides resembled those obtained in RP with the citrullinated ones showing a minimal fragmentation, confirming that the loss of the positive charge has a pronounced effect on both the ionization and the fragmentation, as already demonstrated by others.[66,67]
The analyses performed with RP-LC and HILIC on the couple of a citrullinated and the corresponding arginine-containing peptide revealed a good separation, especially on the hydrophilic surface, representing an interesting starting point for the analyses of the NETS digest. It has to be noticed that standard peptides represent an ideal sample for the low complexity and the great amount of the analytes available for the analysis while typical proteomic sample present opposite characteristics.

1.3.5.2 RP-LC analysis of NETs digest

The same instrumental set-up used for the analysis of standard peptides was applied to the analysis of the NETs digest and the corresponding LC-MS/MS data were processed using the software Xcalibur 2.0.7. As previously described the first aim of

the RP-LC analysis on this sample was to achieve a chromatographic separation of the citrullinated peptide DNIQGITKPAIRR (MH^+=1482.80) from its unmodified version (MH^+=1481.80), in order to characterize the exact deimination position. MALDI analysis had anticipated the presence of a contaminating peptide (MH^+=1480.80) having a minus 1 Da mass than the unmodified, thus suggesting to take it into account, too.
Therefore, in order to evaluate the respective retention times of the three peptides, three extracted ion chromatograms (XICs) corresponding to the double-charged unmodified peptide (M^{2+}=741.42), the citrullinated (M^{2+}=741.92) and the contaminating peptide (M^{2+}=740.92) were extrapolated.
Due to the close proximity of the three mass values, for each XIC a narrow window of 0.2 Da was selected trying to exclude the contribute coming from the other two peptides. Unfortunately the mass differences among them were not wide enough to prevent the algorithm to extract the second or third isotopic peak of the others when these converged with the monoisotopic peak of the precursor of interest (Figure 1.35).

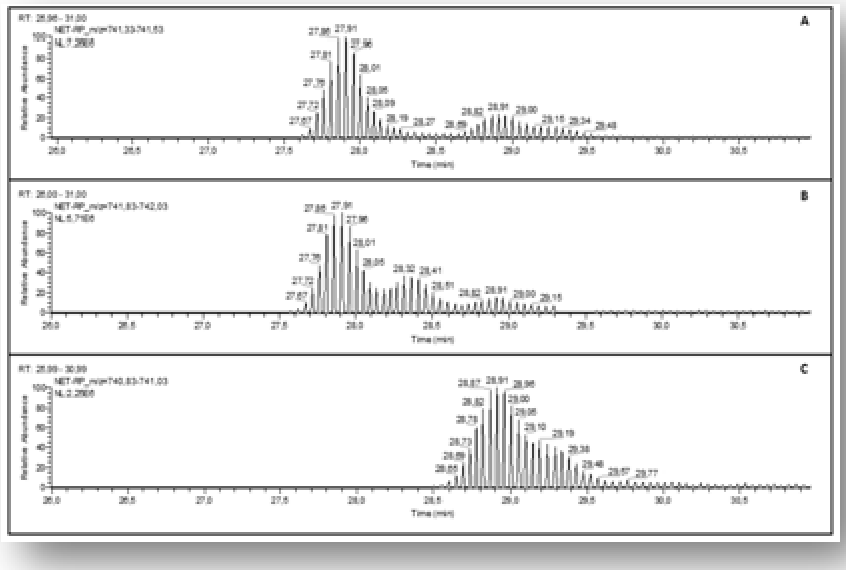


Figure 1.35. XICs of the unmodified peptide with M^{2+}=741.42 (panel A), the citrullinated with M^{2+}=741.92 (panel B) and the contaminant with M^{2+}=740.94 (panel C)

Therefore the XIC of the unmodified peptide (Figure 1.35, panel A) contained also the contribute coming from the ^{13}C isotopic peak of the contaminant peptide; the XIC of the citrullinated peptide (Figure 1.35, panel B) extracted also the second isotopic peak of the unmodified peptide and the third peak of the contaminant. The XIC of the contaminant peptide (Figure 1.35, panel C) was the only "pure" XIC due to the fact that its monoisotopic mass did not overlay to any isotopic peak of the others.

Through comparison of the three XICs it was possible to associate the peak with retention time 27.91 to the unmodified peptide, the one at 28.32 to the citrullinated and that at 28.91 to the contaminant peptide. From the XIC of the citrullinated peptide which contained contributes from all the three peptides it was evident the higher amount of the unmodified peptide compared to the other, as should be expected.
From the analysis of the three XICs one major consideration emerged: not only the citrullinated peptide was not separated from its unmodified version as good as had happened to the standard peptides but it also eluted in the middle of the unmodified and the contaminant peptide. MS full scan acquired at the beginning, at the middle and at the end of the citrullinated peptide's peak showed that it never eluted alone but in presence of the unmodified until the apex of the chromatographic peak and together with the contaminant at the end (Figure 1.36).

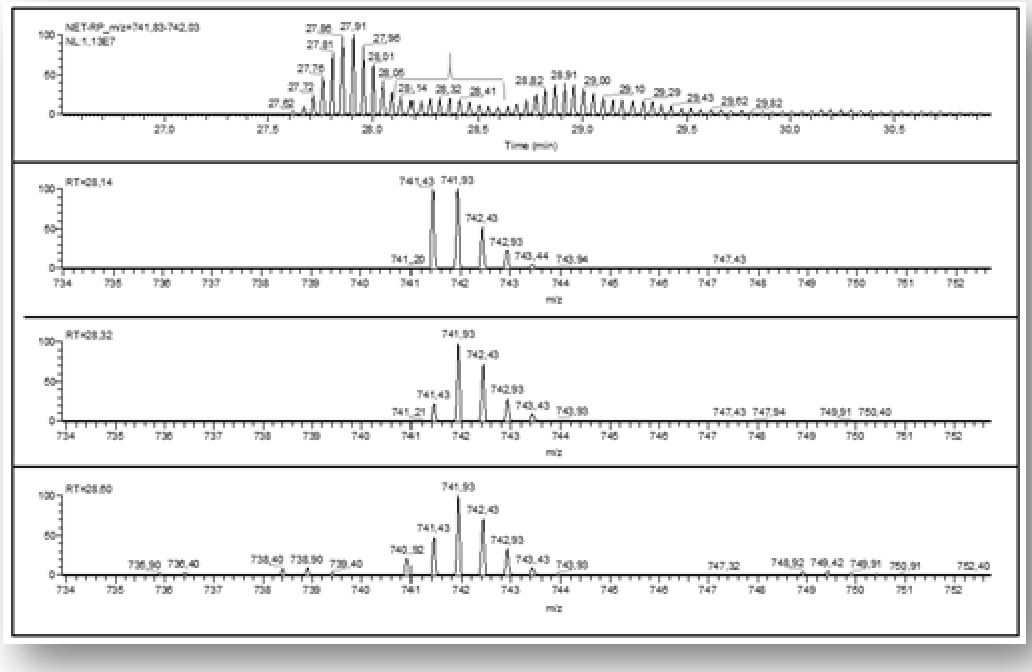
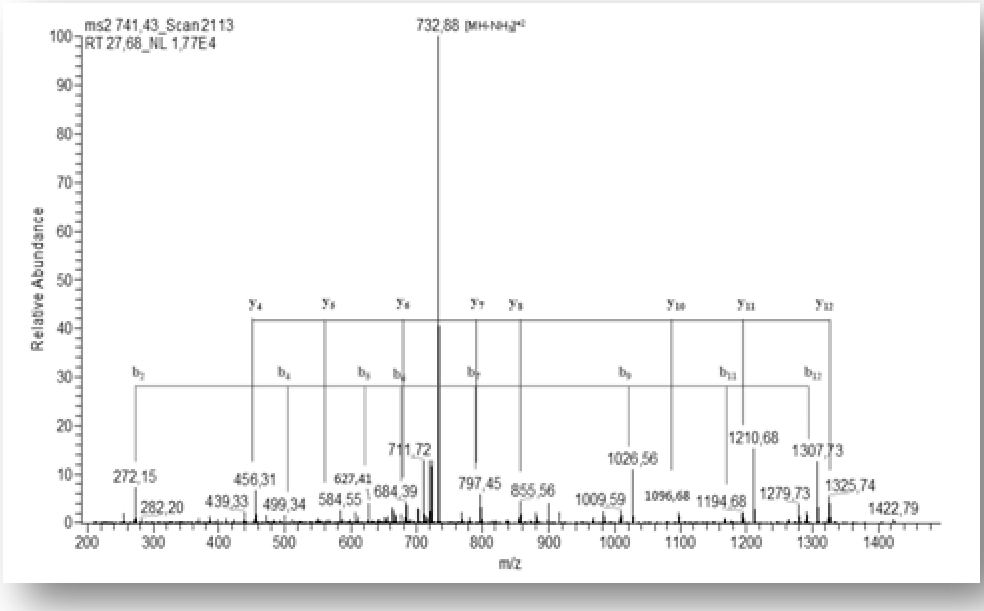


Figure 1.36. XIC of the citrullinated peptide. Full scans, zoomed on the mass of interest, acquired at the beginning, at the middle and at the end of the peak of the citrullinated peptide

The further step involved the analysis of the MS/MS spectra of the three precursors but a really unpredictable situation was unraveled. Indeed there was a complete disagreement between the theoretic fragment ions expectable from the peptide DNIQGITKPAIRR and the experimental signals observed which, on the contrary, perfectly fitted with the peptide RDNIQGITKPAIR (Figure 1.37, above).



b series

b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
D	N	I	Q	G	I	T	K	P	A	I	R	R
	230.0771	343.1612	471.2198	528.2413	641.3253	742.3730	870.4680	967.5207	1038.5578	1151.6419	1307.7430	
R	D	N	I	Q	G	I	T	K	P	A	I	R
	272.1353	386.1783	499.2623	627.3424	684.3424	797.4264	898.4741	1026.5691	1123.6218	1194.6589	1307.7430	

y series

Y_{13}	Y_{12}	Y_{11}	Y_{10}	Y_9	Y_8	Y_7	Y_6	Y_5	Y_4	Y_3	Y_2	Y_1
D	N	I	Q	G	I	T	K	P	A	I	R	R
	1366.8277	1252.7848	1139.7008	1011.6422	954.6207	841.5366	740.4890	612.3940	515.3412	444.3041	331.2201	175.1190
R	D	N	I	Q	G	I	T	K	P	A	I	R
	1325.7836	1210.7566	1096.6837	983.5996	855.5411	798.5196	685.4355	584.3879	456.2929	359.2401	288.2030	175.1190

Figure 1.37. MS/MS spectra fitting with the sequence RDNIQGITKPAIR (above); theoretic b and y ion series of the two sequences (below)

It meant that the sequence who had attracted all the attention until now was not DNIQGITKPAIRR identified after databank search of the previous MALDI spectra but corresponded to the isobaric RDNIQGITKPAIR.

The two peptides mostly covered the same region within histone H4 except that the first was included between residues 25-37 and the second between 24-36. However, even if the sequence was almost similar, the resulting fragmentation pattern was completely different (Figure 1.37, below), making undoubted the interpretation of the MS/MS spectrum.

Furthermore they were isobaric peptides due to the presence of the same amino acids located in a slight different position. Moreover the sequence RDNIQGITKPAIR was not a tryptic peptide and thus it could not be predicted through a theoretical trypsin digestion of histone H4.

Arginine 24 wss preceded by a leucine residue which is not recognized by trypsin but by chymotrypsin which mainly recognizes hydrophobic sites. The trypsin preparation used for the digestion had an extractive origin since it comes from porcine pancreas and residual chymotrypsin is known to occur, acting as a contaminant agent. Therefore RDNIQGITKPAIR peptide could be the result of a combined activity of both of the enzymes, producing a semitryptic peptide exhibiting C and N-terminal Arginines. The presence of non-specific peptides after tryptic digestion seems to be not unusual [68,69] and a specific bioinformatic tool called FindPept [70] that works on peptide mass fingerprinting data to search peptides rising from unspecific cleavages, had been developed by Expasy.[71,72]

The three peptides shared the same fragment ions indicating they all had the same amino acidic sequence; however the contaminant peptide brought a modification responsible for the 1 Da mass decrease compared to the mass of the unmodified. Its MS/MS spectra (Figure 1.38) showed that all b series ions were 1 Da lower than those corresponding to the unmodified while the y series was the same. This behavior suggested that the modification responsible for the 1 Da decrease was located at the Aspartate residue thus provoking a shift in all the b series but not affecting the y series; an amidation of the residue could be an explanation.

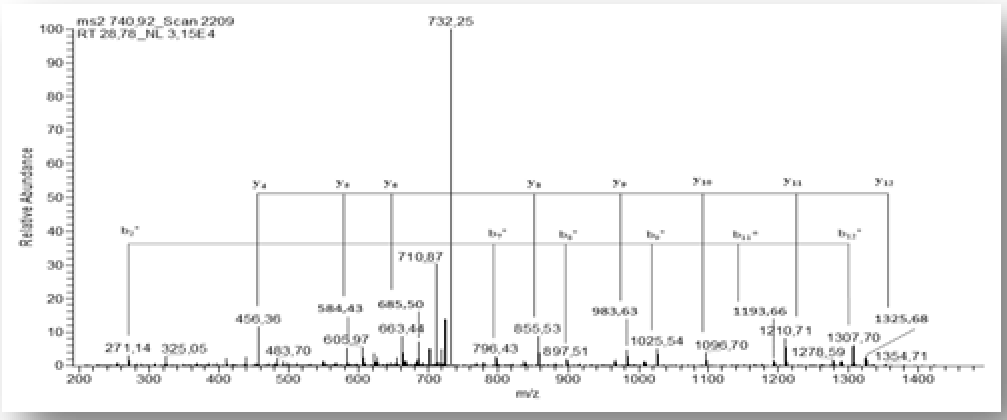


Figure 1.38. MS/MS spectrum of precursor 740.92 Da. Signal tagged with * are decreased of 1 Da

Whatever the modification this peptide was bringing the point was that it behaved as an interfering molecule because its elution time was very close to the citrullinated's one and the negative consequences rising from that situation has been previously depicted. In order to face the real purpose of these LC-MS/MS analyses all the MS/MS spectra of precursors at 741.9 m/z (namely the citrullinated peptide) were extrapolated and an accurate spectral interpretation was led, trying to establish the deimination site. As starting point double or triple charged fragment ions were not considered reliable since

fractions of 0.984 Da, which was itself a slight variation, may not be well appreciated nor distinct from the experimental bias of the low resolution analyzer. For this reason only the MS/MS spectra of double-charged precursors were extracted and only mono charged ions were evaluated (two spectra corresponding to triple-charged precursors were not considered).

With these restraints only three spectra were available for the analyses probably for the close elution of deiminated peptides with the others and for their low concentration, that had reduced the probability for such precursors to be selected for the fragmentation.

Each scan was analyzed in order to find fragment signals to be considered diagnostic of the presence of the citrulline. These did not correspond any more to the ones anticipated in the introduction since the amino acidic sequence has been proved to be different. The sequence RDNIQGITKPAIR brought the arginine residues at opposite positions so the presence of citrulline at N-terminus should produce a b-series completely shifted 0,984 Da while if present at C-terminus the same shift should involve the y-series. For the scan 2161 and 2157 a great part of the b series was increased of 0.984 Da and for scan 2097, even if the coverage was slightly lower, a similar trend was observed (Figure 1.39).

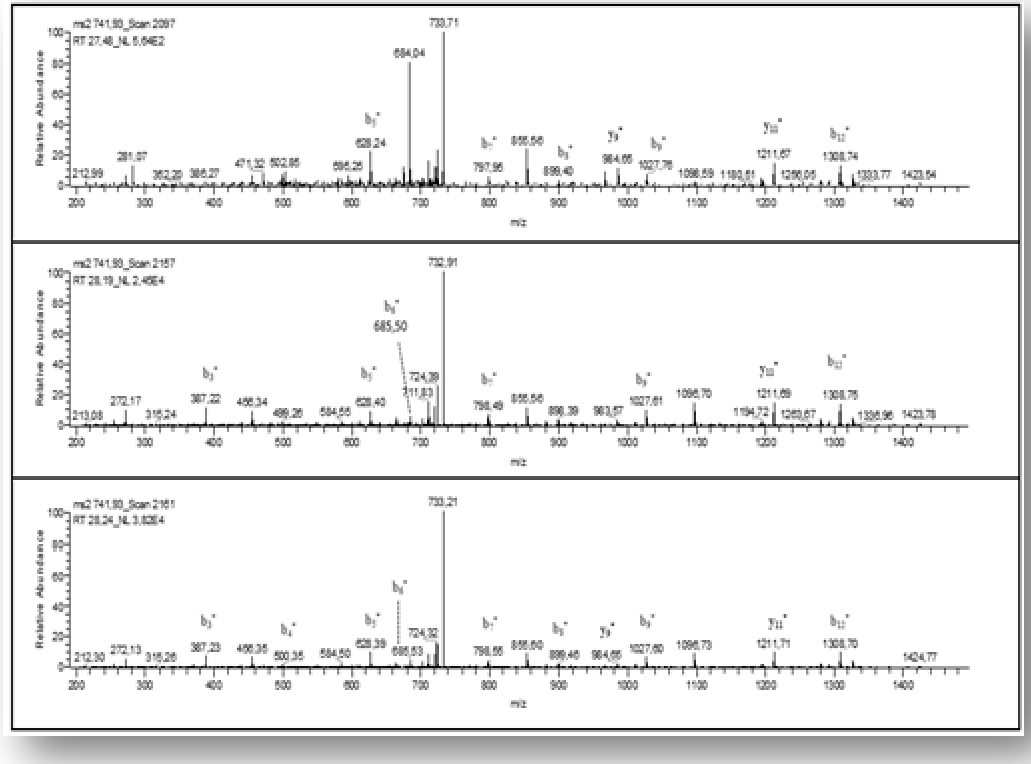


Figure 1.39. MS/MS spectra of the citrullinated peptide. Signal tagged with * represent the diagnostic peptides

However it should be noted that that for all the spectra the fragment ions presented signals of quite low intensity, some of which slightly higher from the background noise. This was partly due to the low amount of citrullinated peptides that reflected the low sampling rate of the fragment ions by means of the analyzer and partly due to the resistance of citrullinated peptide to go under fragmentation.
More carefully looking at the spectra it was also noticed that there were some b fragment ions that did not contain the increment and there were also one or two y ions increased of 0.984 Da. In order to find possible explanations for such behavior MS/MS spectra of the standard citrullinated peptide (Mr=2297.2088), previously analyzed by means of RPLC-MS/MS run, were re-analyzed since they represented a good model to extrapolate useful information. The peptide, indeed, contained, two arginines in the amino acidic sequence and the deimination site was known therefore the b and y fragment ions with the expected increment were predictable. Despite this, it could possible to highlight in the spectra both signals with an increment when not expected (expected, instead, if the other arginine was deiminated) and few signals without the increment when awaited.
Having found a similar behavior for both the standard citrullinated peptide and the one from the NETs tryptic digest was a confirmation that the low concentration and the contamination from the unmodified peptide were factors able to create tangible complications in the final interpretation of the spectrum.
Taking into account all the considerations exposed in this discussion it was concluded that the RP-LC analysis of the NETs digest had not been able to provide completely satisfactory results, concerning the unraveling of the exact deimination site of peptide RDNIQGITKPAIR. However in all citrullinated spectra the b series showed the majority of signals bearing the 0.984 Da increment, perhaps indicating that the deimination site was located on arginine 24. Other analysis were absolutely required, using the HILIC stationary phase firstly trying to reduce the negative contribution rising from the interfering peptides.

Before approaching other strategies LC-MS/MS data were analyzed to search for other possible deiminated peptides; in particular the Xcalibur raw data was processed using the software Raw2MSM which creates MGF peak list files, ready to be directly submitted to data bank search. This was led within the human proteome of the SwissProt database, imposing semitrypsin as enzyme, deamidation (R) and oxidation (M) as variable modifications and carbamidomethylation (C) as fixed modification.
The first result was the identification of histone H4 with 15 peptides as top scoring match and then other proteins identified by means of at least 2 peptides such as histone H3.1t and S100-protein isoforms. The identification of proteins that had not

emerged by the peptide mass fingerprinting strategy is not surprising considering the higher analytical sensitivity of the LC-MS/MS system (Table 1-I).

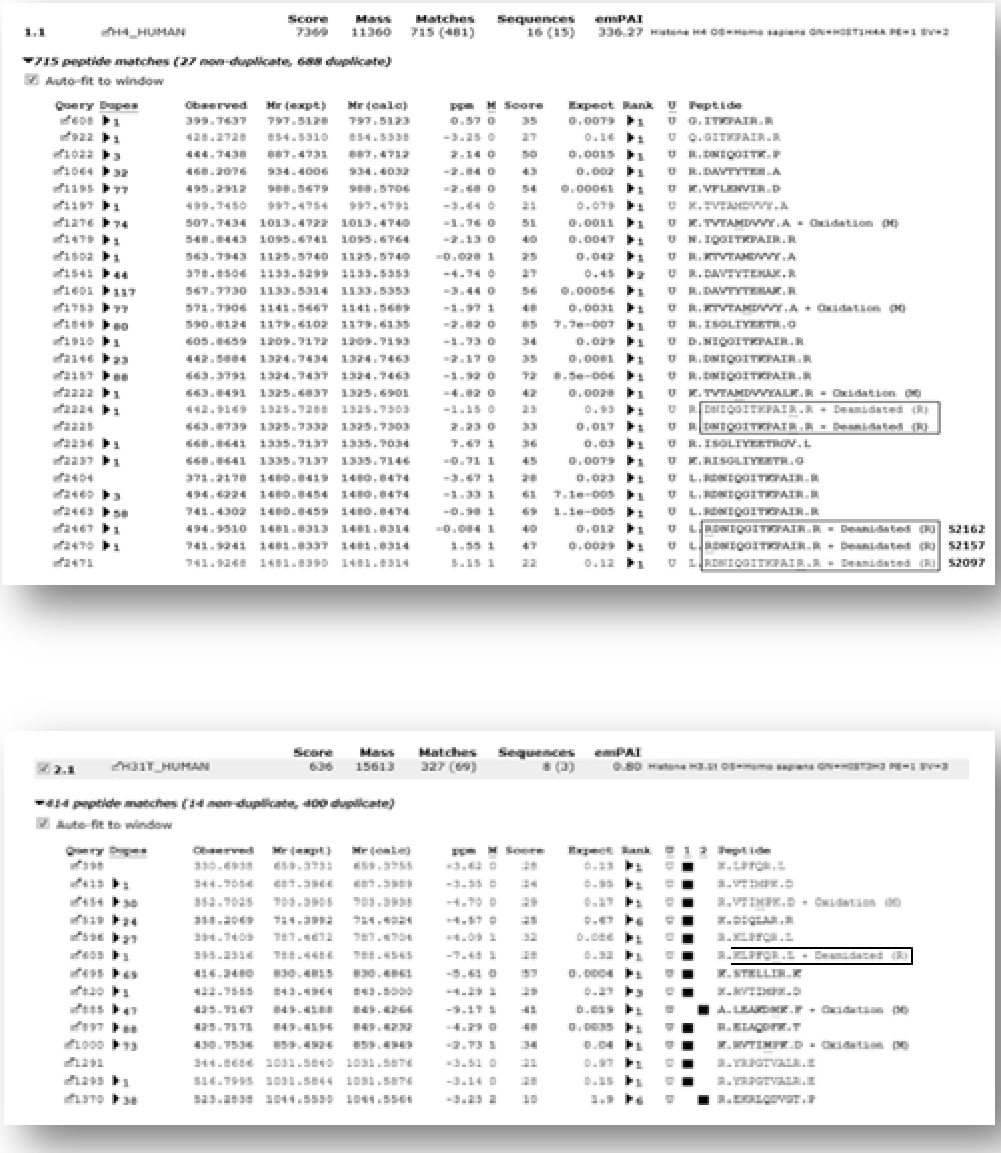
Family	Accession	Score	Mass	Matches	Pep(sig)	Sequences	Seq(sig)	emPAI	Description
1	H4_HUMAN	7369	11360	715	481	16	15	336.27	Histone H4 OS=Homo sapiens GN=HIST1H4A PE=1 SV=2
2	H31T_HUMAN	636	15613	327	69	8	3	0.80	Histone H3.1t OS=Homo sapiens GN=HIST3H3 PE=1 SV=3
3	S10A8_HUMAN	119	10885	21	14	6	4	2.00	Protein S100-A8 OS=Homo sapiens GN=S100A8 PE=1 SV=1
4	H2AZ_HUMAN	111	13545	10	5	3	2	0.56	Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2
5	S10AC_HUMAN	84	10569	14	6	3	2	0.76	Protein S100-A12 OS=Homo sapiens GN=S100A12 PE=1 SV=2

Table 1-I. Identified proteins from NETs digest after database search

Histone H4 sequence coverage was elevated but not complete (Figure 1.40). There were no information regarding the N-terminal region of the protein, where the close presence of arginine and lysine had caused a massive cleavage by trypsin, releasing peptides too short to be detected by the instrument.

Figure 1.40. Histone H4 sequence coverage. Underlined residues are those identified

Database search of citrullinated peptides confirmed the identification of the citrullinated peptide RDNIQGITKPAIR; it is interesting to note that all but one of the MS/MS spectra previously extrapolated from the raw data were found by the search algorithm, correctly identified and able to address Arg25 (N-terminal of the peptide) as the modification site. The exception was the scan 2097 that was identified below the score indicating a difficulty of the search engine to correctly assign the modification. Nevertheless the scan 2097 had been previously found to be the most problematic for the presence of a lower number of b ions signals containing the increment compared to the y ions, therefore the underscored identification might be the result of such unclear spectrum.
Within all the identified proteins the algorithm highlighted the presence of two further putative candidates citrullinated peptides: DNIQGITKPAIR (M^{2+}=663.865) from histone H4 and KLPFQR (M^{2+}=395.231) from histone H3.1t (Figure 1.41).



		Score	Mass	Matches	Sequences	emPAI	
1.1	♂H4_HUMAN	7369	11360	715 (481)	16 (15)	336.27	Histone H4 OS=Homo sapiens GN=HIST1H4A PE=1 SV=2

▼*715 peptide matches (27 non-duplicate, 688 duplicate)*

☑ Auto-fit to window

Query	Dupes	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide	
♂608	▶1	399.7637	797.5128	797.5123	0.57	0	35	0.0079	▶1	U	G.ITKPAIR.R	
♂922	▶1	428.2728	854.5310	854.5338	-3.25	0	27	0.16	▶1	U	Q.GITKPAIR.R	
♂1022	▶3	444.7438	887.4731	887.4712	2.14	0	50	0.0015	▶1	U	R.DNIQGITK.P	
♂1064	▶32	468.2076	934.4006	934.4032	-2.84	0	43	0.002	▶1	U	R.DAVTYTEH.A	
♂1195	▶77	495.2912	988.5679	988.5706	-2.68	0	54	0.00061	▶1	U	K.VFLENVIH.D	
♂1197	▶1	499.7450	997.4754	997.4791	-3.64	0	21	0.079	▶1	U	K.TVTAMDVVY.A	
♂1276	▶74	507.7434	1013.4722	1013.4740	-1.76	0	51	0.0011	▶1	U	K.TVTAMDVVY.A + Oxidation (M)	
♂1479	▶1	548.8443	1095.6741	1095.6764	-2.13	0	40	0.0047	▶1	U	R.IQGITKPAIR.R	
♂1502	▶1	563.7943	1125.5740	1125.5740	-0.028	1	25	0.042	▶1	U	R.KTVTAMDVVY.A	
♂1541	▶44	378.8506	1133.5299	1133.5353	-4.74	0	27	0.45	▶2	U	R.DAVTYTEHAK.R	
♂1601	▶117	567.7730	1133.5314	1133.5353	-3.44	0	56	0.00056	▶1	U	R.DAVTYTEHAK.R	
♂1753	▶77	571.7906	1141.5667	1141.5689	-1.97	1	48	0.0031	▶1	U	R.KTVTAMDVVY.A + Oxidation (M)	
♂1849	▶80	590.8124	1179.6102	1179.6135	-2.82	0	85	7.7e-007	▶1	U	R.ISGLIYEETR.G	
♂1910	▶1	605.8659	1209.7172	1209.7193	-1.73	0	34	0.029	▶1	U	D.NIQGITKPAIR.R	
♂2146	▶93	442.5884	1324.7434	1324.7463	-2.17	0	35	0.0081	▶1	U	R.DNIQGITKPAIR.R	
♂2157	▶88	663.3791	1324.7437	1324.7463	-1.92	0	72	8.5e-006	▶1	U	R.DNIQGITKPAIR.R	
♂2222	▶1	663.8491	1325.6837	1325.6901	-4.82	0	42	0.0028	▶1	U	K.TVTAMDVVYAALK.R + Oxidation (M)	
♂2224	▶1	442.9149	1325.7288	1325.7303	-1.15	0	23	0.93	▶1	U	R.DNIQGITKPAIR.R + Deamidated (R)	
♂2225		663.8739	1325.7332	1325.7303	2.23	0	33	0.017	▶1	U	R.DNIQGITKPAIR.R + Deamidated (R)	
♂2236	▶1	668.8641	1335.7137	1335.7034	7.67	1	36	0.03	▶1	U	R.ISGLIYEETRGV.L	
♂2237	▶1	668.8641	1335.7137	1335.7146	-0.71	1	45	0.0079	▶1	U	K.RISGLIYEETR.G	
♂2404		371.2178	1480.8419	1480.8474	-3.67	1	28	0.023	▶1	U	L.RDNIQGITKPAIR.R	
♂2460	▶3	494.6224	1480.8454	1480.8474	-1.33	1	61	7.1e-005	▶1	U	L.RDNIQGITKPAIR.R	
♂2463	▶58	741.4302	1480.8459	1480.8474	-0.98	1	69	1.1e-005	▶1	U	L.RDNIQGITKPAIR.R	
♂2467	▶1	494.9510	1481.8313	1481.8314	-0.084	1	40	0.012	▶1	U	L.RDNIQGITKPAIR.R + Deamidated (R)	S2162
♂2470	▶1	741.9241	1481.8337	1481.8314	1.55	1	47	0.0029	▶1	U	L.RDNIQGITKPAIR.R + Deamidated (R)	S2157
♂2471		741.9268	1481.8390	1481.8314	5.15	1	22	0.12	▶1	U	L.RDNIQGITKPAIR.R + Deamidated (R)	S2097

		Score	Mass	Matches	Sequences	emPAI	
☑ 2.1	♂H31T_HUMAN	636	15613	327 (69)	8 (3)	0.80	Histone H3.1t OS=Homo sapiens GN=HIST3H3 PE=1 SV=3

▼*414 peptide matches (14 non-duplicate, 400 duplicate)*

☑ Auto-fit to window

Query	Dupes	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	1	2	Peptide
♂398		330.6938	659.3731	659.3755	-3.62	0	28	0.13	▶1	U	■		K.LPFQR.I
♂415	▶1	344.7056	687.3966	687.3989	-3.35	0	24	0.95	▶1	U	■		R.VTIMPK.D
♂454	▶36	352.7025	703.3905	703.3938	-4.70	0	29	0.17	▶1	U	■		R.VTIMPK.D + Oxidation (M)
♂519	▶24	358.2069	714.3992	714.4024	-4.47	0	25	0.67	▶6	U	■		K.DIQLAR.R
♂596	▶27	394.7409	787.4672	787.4704	-4.09	1	32	0.066	▶1	U	■		R.KLPFQR.I
♂603	▶1	395.2316	788.4486	788.4545	-7.48	1	28	0.32	▶1	U	■		R.KLPFQR.I + Deamidated (R)
♂695	▶49	416.2480	830.4815	830.4861	-5.61	0	57	0.0004	▶1	U	■		K.STELLIR.K
♂820	▶1	422.7555	843.4964	843.5000	-4.29	1	29	0.27	▶3	U	■		K.RVTIMPK.D
♂835	▶47	425.7167	849.4188	849.4266	-9.17	1	41	0.019	▶1	U		■	A.LEARDMK.F + Oxidation (M)
♂897	▶88	425.7171	849.4196	849.4232	-4.29	0	48	0.0035	▶1	U	■		R.ELAQDFK.T
♂1000	▶73	430.7536	859.4926	859.4969	-2.73	1	34	0.04	▶1	U	■		K.RVTIMPK.D + Oxidation (M)
♂1291		344.8686	1031.5840	1031.5876	-3.51	0	21	0.97	▶1	U	■		R.YRPGTVALR.E
♂1293	▶1	516.7995	1031.5844	1031.5876	-3.14	0	28	0.15	▶1	U	■		R.YRPGTVALR.E
♂1370	▶38	523.2838	1044.5530	1044.5561	-3.23	2	10	1.9	▶6	U		■	R.EKRLQDVGT.F

Figure 1.41. List of peptides identified from Histone H4 (above) and from Histone H3.1t (below); deamidated peptides are enclosed. Individual ions scores > 35 indicate identity or extensive homology (p<0.05)

These peptides ought to contain the putative citrulline at the C-terminus, thus assuming a trypsin cleavage, but in both cases the identifications were below the threshold score that meant that they were not completely reliable. Since even for the scan 2097 of the peptide RDNIQGITKPAIR there has been an underscored identification of a deiminated peptide it was important to check the quality of the MS/MS spectra of these two peptides, like it has been done with peptide RDNIQGITKPAIR. Moreover, given that the presence of the peptides with a mass increase of 1 Da compared to the unmodified forms has been highlighted, it was also very interesting to figure out which other modification might be the responsible of such increase, if not the deimination.

Mass ranges were extracted around the value of the unmodified peptide DNIQGITKPAIR (M^{2+}=663.373) and the peptide bearing the increment (M^{2+}=663.873). The chromatographic behavior showed the latter eluting in tail of the unmodified and a partial overlap was evident between their isotopic patterns (Figure 1.42).

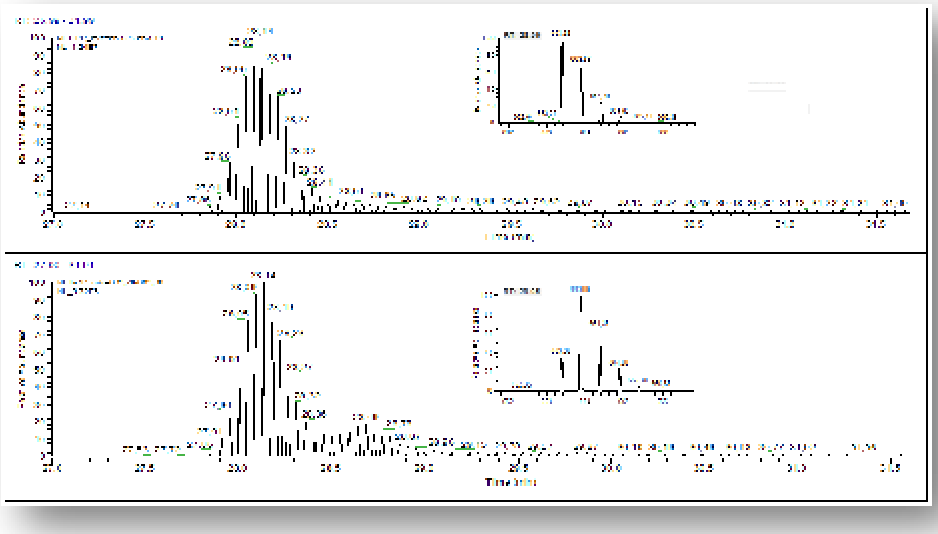


Figure 1.42. XICs of the peptide DNIQGITKPAIR and its modified version, including the corresponding integrated mass scans

Peptide DNIQGITKPAIR is included in the sequence 25-36 of histone H4 as result of trypsin activity with no missed cleavage and it is a great part of the region covered by the peptide RDNIQGITKPAIR. MS/MS spectra of the couple of peptides DNIQGITKPAIR highlighted the absence of a great part of the mono charged ions of the b series making not possible the detection of eventual amino acidic differences within this series (Figure 1.43).

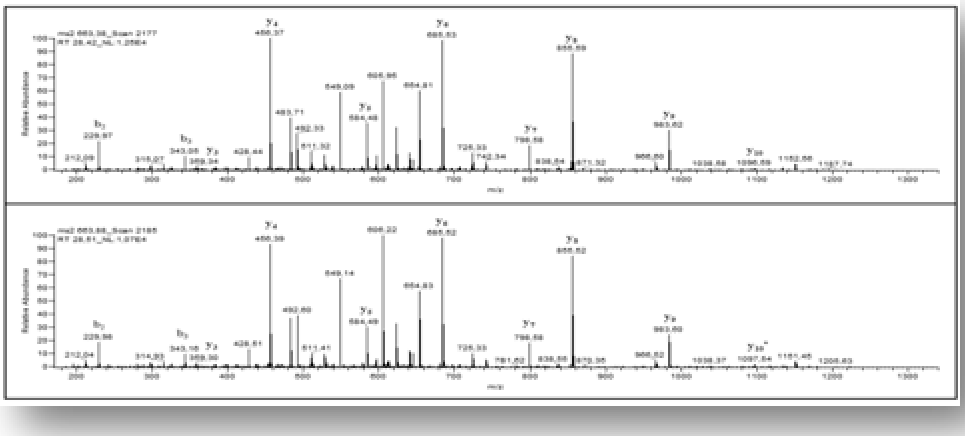


Figure 1.43. MS/MS spectra of DNIQGITKPAIR peptide (above) and its modified form (below)

Instead the y series presented several signals shared with the unmodified peptide whose presence had masked signals rising from the modified peptide. Only the signal y_{10} contained an increment of 0.984 Da but the intensity of the signal was too low to be trustable.

Therefore, even in this case MS/MS spectrum did not provide a final answer concerning the exact modification site within the peptide of interest. The incomplete chromatographic separation of unmodified and modified peptides heavily impaired the interpretation of the MS/MS spectrum of the modified peptide.

Similar consideration rose from the other peptide KLPFQR (M^{2+}=394.74) whose modified version with the mass increment of 1 Da (M^{2+}=395.23) was evident from the MS scan extracted by the XIC (Figure 1.44).

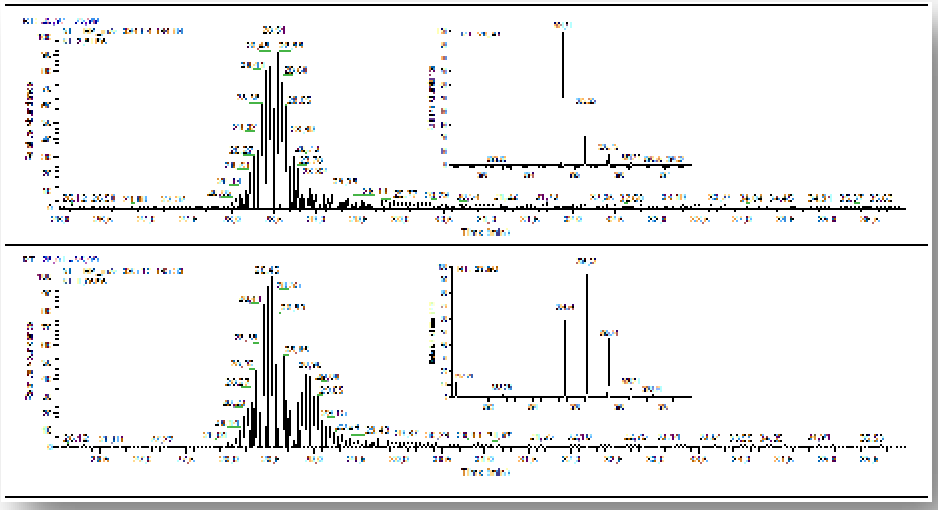


Figure 1.44. XICs of the peptide KLPFQR and its modified version, including the corresponding integrated mass scans

In order to check the presence of the modification at amino acidic level MS/MS spectra of unmodified and modified peptides were compared (Figure 1.45).

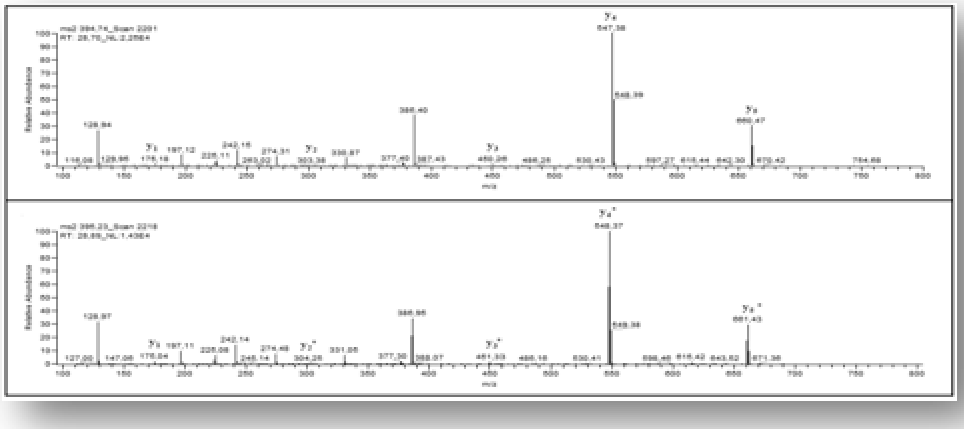


Figure 1.45. MS/MS spectra of KLPFQR peptide (above) and its modified form (below)

In this case looking at the y series it was found the presence of the increment starting from ion y_2 that was compatible with a glutamine deamidation, which might have been the responsible for the 0.984 Da mass increase observed in the precursor.

Considering the retention times of the peptides with sequence RDNIQGITKPAIR, DNIQGITKPAIR and KLPFQR it was clear that they all eluted in the same temporal window crowding the full MS scan with a lot of precursors. In particular within the retention time of the citrullinated peptide RDNIQGITKPAIR (M^{2+}=741.93) other precursor were present such as DNIQGITKPAIR (M^{2+}=663.33) and the majority of them showed signals of higher intensity (Figure 1.46).

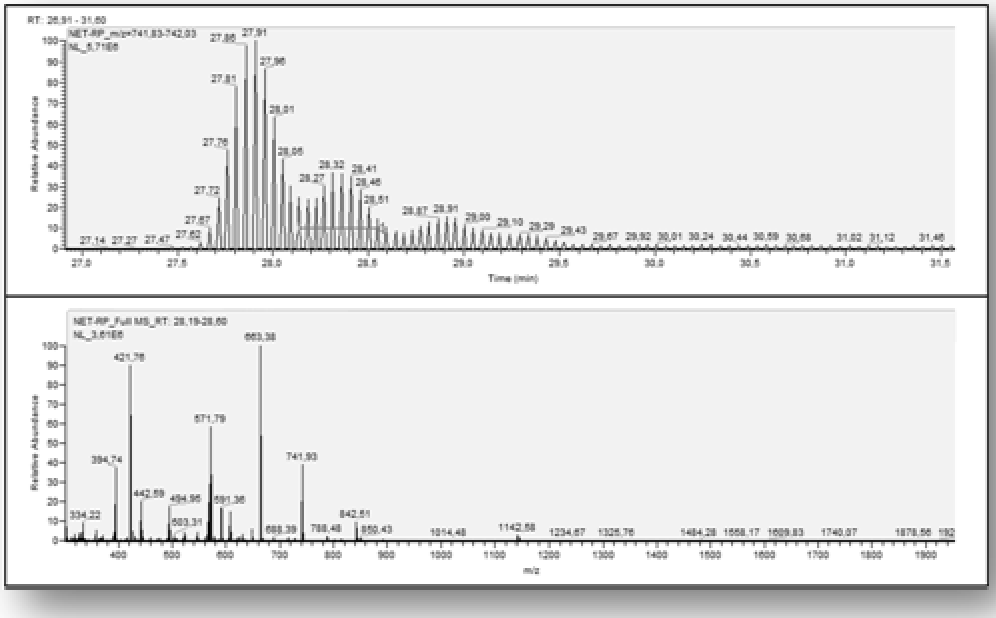


Figure 1.46. XIC of the citrullinated peptide (above) and full MS scan rising from the integration of the chromatographic peak of the citrullinated peptide (below)

Since the MS method is set to send to the collision cell only the three most intense precursor ions and being the citrullinated peptide RDNIQGITKPAIR the less intense among them it was sampled few times as evident from the only three MS/MS spectra of double-charged precursors selected for the fragmentation. This situation explained well which is the essential difficulty in the detection of low abundant peptides.

In conclusion these LC-MS/MS analyses performed on the NETs digest using a RP stationary phase provided very useful information concerning the peptide that had been found to be citrullinated with the derivatization technique:

- MS/MS spectra of the citrullinated peptide revealed that the amino acidic sequence was not the one expected (DNIQGITKPAIRR) but it was slightly different (RDNIQGITKPAIR) because of a combined trypsin-chymotrypsin activity that produced a semi-tryptic peptide. This discovery moved the attention from residues 36 and 37 to residues 24 and 36 as possible deimination sites.
- The contaminant peptide with MH^+=1480.072, as shown by MALDI spectrum, was proved to be related to the citrullinated and the unmodified peptides because sharing the same amino acidic sequence and probably bearing an amidated aspartic acid.
- The three peptides eluted in a narrow temporal window revealing that an uncompleted chromatographic separation had realized; moreover several peptides eluted together with the citrullinated one thus reducing the probability for this to be selected for the fragmentation.
- Only three useful MS/MS spectra could be extrapolated for the citrullinated peptide RDNIQGITKPAIR and a spectral interpretation was approached. However the very low intensity signals combined with the contamination from the unmodified peptide hindered an undoubted assignment of the deimination site. The MS/MS spectra seemed more compatible with the presence of a deiminated Arg24 instead of the Arg36 but deepen analyses were required.

Interesting results came from the processing of the same analyses in order to highlight the presence of other citrullinated peptide:

- The database search confirmed the identification and the assignment of the citrullinated peptide RDNIQGITKPAIR, indicating Arg24 as the deimination site; for one scan the results were unclear demonstrating that the identification of this PTM is very difficult even when performed by bioinformatic tools.
- Two further peptides (DNIQGITKPAIR from Histone H4 and KLPFQR from Histone H3.1t) were found to be possibly deiminated but the underscored identifications necessitated a manual analysis of the spectra. Once again the partial co-elution of the unmodified peptide from the modified one was the biggest hinder for a correct assignment. For peptide KLPFQR partial indications

suggested that the modification was not the arginine deimination but the isobaric glutamine deamidation.

An aspect that clearly emerges from these analyses is the real difficulty in the characterization of low expressed components such as peptides bearing a modification, spread inside a complex sample that provides interfering molecules. LC-MS/MS analyses had identified few scans for each of the modified peptides (deiminated or deamidated). On one hand this proved the high sensitivity of the employed system but on the other hand it highlighted the importance to take care of this small number of scans and carefully analyzing the spectra to catch all the information they bring.

1.3.5.3 HILIC analysis of NETs digest

The rationale behind the use of HILIC standard phase for the separation of citrullinated peptides from the unmodified forms had been already described.

In RP the modified peptide RDNIQGITKPAIR eluted after the unmodified thanks to its slightly higher hydrophobicity that resulted in a longer retention by the phase, even if not enough to allow a complete chromatographic separation.

Since the citrullinated and the unmodified peptides share similar polar properties a polar phase such as the ZIC-HILIC® ought to be more incisive than an hydrophobic phase in their separation. Moreover the orthogonality of the two configurations ensure an opposite order of elution of the peptides so that, with HILIC, citrullinated peptides elute before the unmodified. This aspect might be advantageous because the very first part of the chromatographic peak of the citrullinated peptide should not suffer the contamination from the unmodified expected to elute later. So if few citrullinated precursors can be isolated for the fragmentation there is the possibility to have "pure" MS/MS spectra available for a final certain interpretation of the fragment ions pattern.

The sample to submit to these analyses needed to be properly prepared in an organic medium being water the elution buffer. The tryptic digest contained a saline solution of ammonium bicarbonate therefore it was dried and redissolved in 90% MeCN.

The same instrumental set-up adopted for the standard peptides was used.

XICs for the three forms of the peptide RDNIQGITKPAIR were extrapolated in order to check their chromatographic behavior. The XIC of the peptide containing a putative amidated aspartate represented the reference chromatogram because it did not contain any contribute coming from the other (Figure 1.47, panel A). From its shape it was clear a feature already seen for the standard peptide that was the peak tailing even if in this case a general large peak width was observed.

This was a really unwanted aspect which even other peptides were supposed to suffer.

XIC of the unmodified peptide (Figure 1.47, panel B) showed a massive overlapping with the previous one but they were not the analytes target of the analyses therefore it was not a great problem. XIC (Figure 1.47, panel C) of the citrullinated peptide, instead, showed an extra peak before them, as expected.

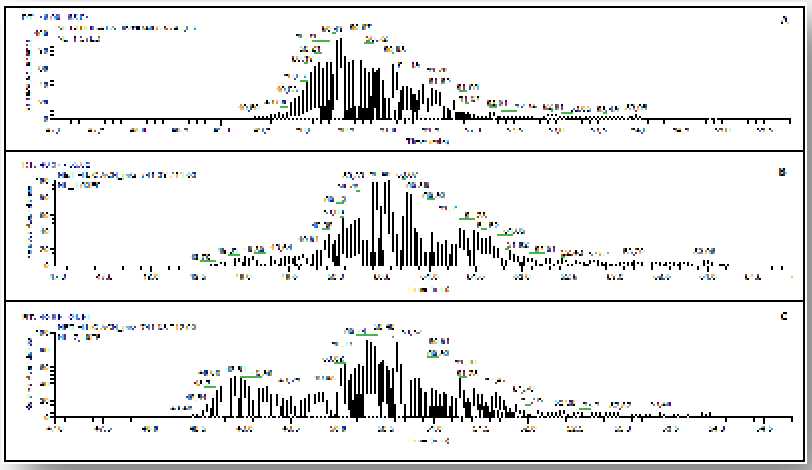


Figure 1.47. XICs of the contaminant peptide with M^{2+}=740.94 (panel A), the unmodified with M^{2+}=741.42 (panel B) and the citrullinated with M^{2+}=741.92 (panel C)

At a first glance it clearly emerged that the desired separation between the citrullinated peptide and the unmodified version substantially did not take place mainly because of the mentioned marked band broadening responsible of an excessive chromatographic dispersion of the analytes. To retrieve more detailed information, full MS scans were extracted sampling various retention times in correspondence of the peak of the citrullinated peptide (Figure 1.48).

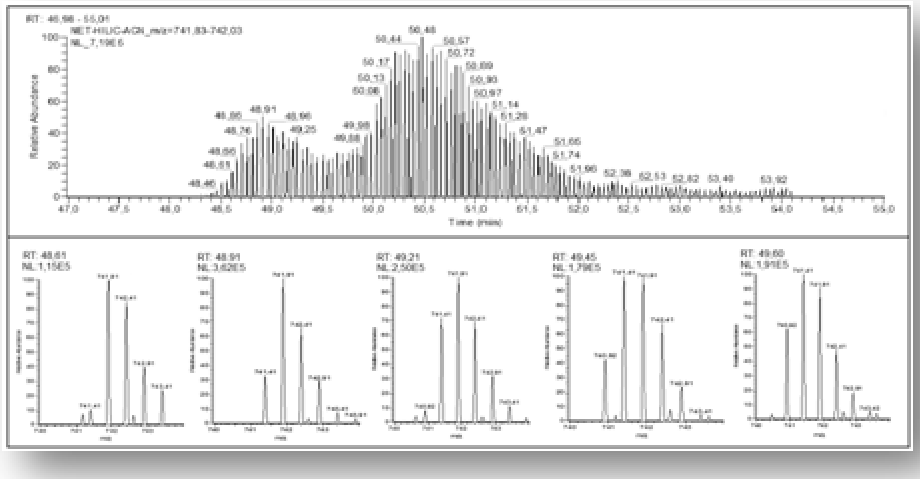


Figure 1.48. Full MS scan at different retention times, within the peak of the citrullinated peptide

Looking at the extracted scans it was clear that the citrullinated peptide did not ever eluted alone but there was always a minimum contribution from the other two species.
At retention time of 48.61 min the citrullinated peptide was flanked by a small amount of the unmodified whose intensity increased so that at 49.45 min reached the intensity of the citrullinated and at 49.60 passed it. In the mean time even the peptide with suspected amidated aspartate grew in intensity from 49.21 min when it was barely visible until 49,60 when its amount become relevant. As consequence of this repeated incomplete chromatographic separation, the corresponding MS/MS spectra of the citrullinated peptide were supposed to be greatly contaminated by the signals of the unmodified and contaminant peptides.
Three MS/MS spectra of double and triple-charged precursors at 741.9 m/z could be extracted from the chromatogram and an anomalous situation was observed.
The interpretation of the extracted MS/MS spectra pointed out that an isobaric peptide of the citrullinated RDNIQGITKPAIR was also present, namely the peptide with sequence TVTAMDVVYALKR that in presence of an oxidized methionine acquired a mass of 1482.798 Da. Since they shared the same mass and eluted in the same moment both were selected for the fragmentation and the subsequent MS/MS obviously contained contributions from both of them. Moreover the precursor selected was a triple-charged one so it was not useful.
Only one MS/MS spectrum (scan 5328) did not contain signals from the peptide TVTAMDVVYALKR but, unfortunately, it corresponded to a retention time of 49.54 min which was a critical chromatographic region because of the elution of both the unmodified and the contaminant peptide. Accordingly, there were very few signals with the increment of 0.984 Da so that it was not possible to extrapolate the exact deimination position (Figure 1.49).

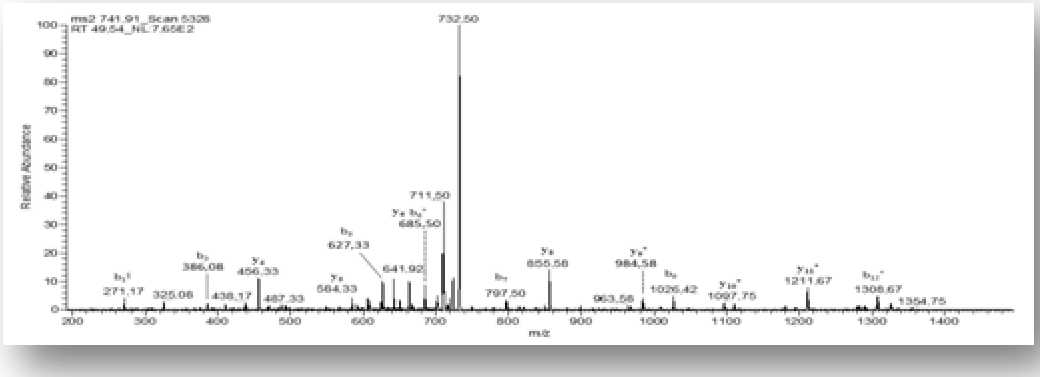


Figure 1.49. MS/MS spectrum of the precursor at 741.91 Da of the RDNIQGITKPAIR sequence; §=b1 ion from the peptide at mass 740.9 Da

In this case, more than in the previous RP-LC analysis, contaminations from the other similar peptides were consistent and masked the incremented signals of the citrullinated peptide thus hindering its interpretation. So the same problems pointed out with RP-LC-MS/MS analysis, once again, emerged. This time there was a further problem created by the peptide TVTAM*DVVYALKR co-eluting with the isobaric citrullinated RDNIQGITKPAIR peptide.
Corresponding MS/MS spectra resulted heavily contaminated.
It should also be noticed that the sample preparation required for HILIC analysis had imposed steps of drying and re-suspension that could have affected peptide recovery. In order to check if, after such preparation, the citrullinated peptide was still present in the sample in a sufficient amount to be detected, further runs was performed changing the MS method, in particular, the scan events cycle. LTQ analyzer was set to work in the top 2 configuration that contemplated a data-dependent isolation of the precursors to be fragmented in the first two selections and a targeted isolation of the citrullinated peptide in the third selection. This kind of setting augments the sensitivity toward the precursor of interest and it is particularly useful in presence of low abundant peptides that otherwise would be rarely selected in a data-dependent mode. With this analysis a large number of MS/MS spectra of precursors of mass 741.9 Da corresponding to the citrullinated peptide were extrapolated, thus indicating that the peptide was present and with such MS method that had imposed its selection it was possible to highlight its presence. However, extracting the XIC of both the citrullinated and the unmodified peptides became evident that no separation between the two forms had realized while in the previous analysis a minimum separation was observed.
Full MS scan showed that all the three peptides RDNIQGITKPAIR were eluting at the same retention time without any sign of retention by the HILIC stationary phases and also the base peak chromatogram was poor of peaks, altogether suggesting that a chromatographic trouble had happened.
This was confirmed by comparing the base peak chromatogram of a standard BSA digest that was run twice in the same sample sequence, one at the beginning and one at the end, in order to provide information regarding the reproducibility of the method. A robust and reliable chromatographic method should provide reproducible results inter- and intra-analysis unless some unexpected events took place instead those chromatograms showed that something had gone wrong. Indeed the same sample showed different retention times not only compared to the previous analysis but also within the same sample list (Figure 1.50) indicating that the system was not robust in the operative mode it was working.

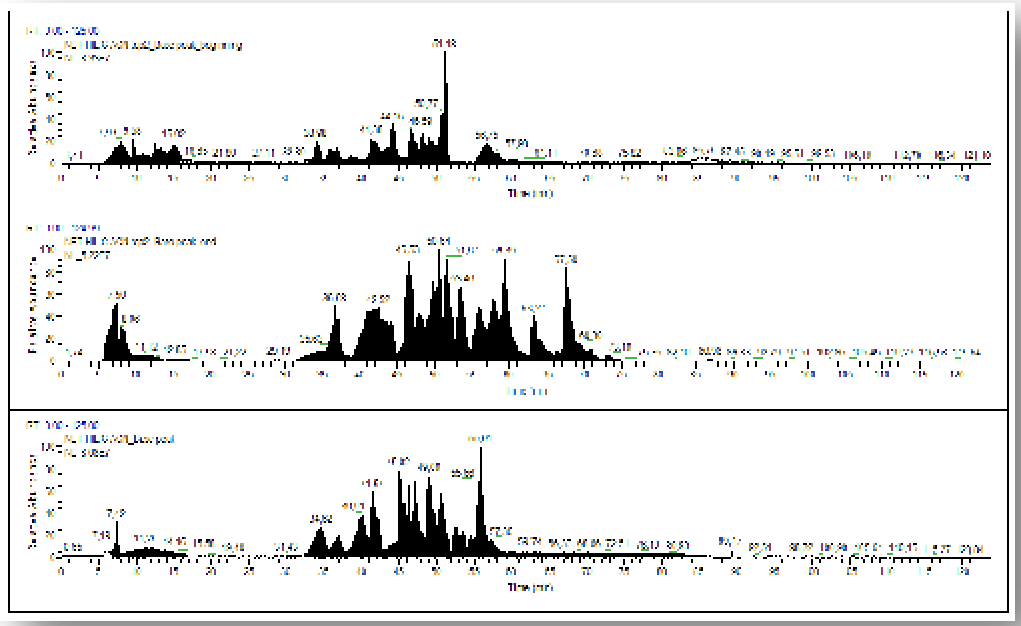


Figure 1.50. Base peak chromatogram of the standard BSA digest at the beginning of the sample list (above) and at the end (in the middle) and from the previous analysis (below)

It is known that a chromatographic system based on HILIC stationary phase is a sensible system due to the extremely long time of reconditioning of the column required to ensure the mechanisms of retention to correctly take place. From the comparison of the BSA analysis run at the beginning and at the end of the sample list it was evident that in the first case a poor separation of the peptides had realized while at the end it was greatly enhanced.

The missed separation regarded also the other two peptide of interest, namely DNIQGITKPAIR and KLPFQR, whose XICs revealed a similar situation.

In conclusion all the HILIC analysis performed on the NETs digest turned not to be equal to the expectation due to the combination of the failed chromatographic separation and of collateral issues which prevented the prefigured purposes to be achieved.

The aims of making a direct detection of citrullines by means of LC-MS/MS analysis has faced consistent problems, the most important of which was the interfering from unmodified peptide. The analysis on the standard peptides demonstrated to be successful when HILIC phase was used while, once employed on the real samples, it failed. Further evaluations concerning the real applicability of this technology for our purposes will be necessary.

1.4 CONCLUSIONS

RA sera contain a variety of autoantibodies but those specific for citrullinated antigens are of utmost interest, being highly specific for the disease, probably endowed with pathogenic potential and possibly related to the basic mechanisms of the disease.

Several lines of evidence indicate that histone deimination is the result of PAD4 activity. It has been shown that PAD4 is localized in nuclei[73] and can convert multiple arginine sites of H3 and H4 into citrulline.[74] PAD4 can in fact deiminate Arg2, Arg8, Arg17 and Arg26 on histone H3[75] and Arg3 on the N-terminal sequence shared by H2a and H4.[76] PAD4 activation and histone deimination is triggered in HL60 cells and blood granulocytes by a variety of stimuli, such as Ca^{2+} ionophore, LPS, TNF, chemoattractant peptides and H_2O_2.[77] Under these stimuli, neutrophils, eosinophils and mast cells extrude NETs, undergoing a novel form of cell death that requires reactive oxygen species produced by NADPH-oxidase. In neutrophils stimulated with Ca^{2+}-ionophore, the qualitative and quantitative NETs composition has been analyzed by a proteomic approach.[78] Core histones are the most abundant proteins, accounting for approximately 70% of all NET-associated proteins. H3 and H4 are present in lower amounts than H2B and H2A and their molecular weight is reduced when compared with nuclear histones, suggesting post-translational modifications.

In this project citrullinated H4 has been identified as a novel substrate for ACPAs, thus showing that NETosis can be the source of this antigen, and establishing a novel connection between innate and adaptive immunity in RA. Initially, immunoblotting experiments allowed us to detect a reactivity of RA sera with a band whose identity with H4 was suggested by specific antibody recognition and demonstrated by MALDI-TOF analysis. Citrullination of H4 was shown by reactivity with anti-citrulline antibodies and proved by MALDI-TOF analysis after peptide derivatization. Subsequently, RA sera reactivity with isolated NETs proteins was observed; again, the band was identified as deiminated H4 using specific antibodies and, most importantly by MALDI-TOF analysis of tryptic peptides from the digested band before and after derivatization. As confirmation of the specific reactivity of RA sera with H4, further validation results were obtained through studies performed by the Clinical Immunology and Allergy Unit of the University of Pisa (Prof. Paola Migliorini). ELISA assays using *in vitro* deiminated human purified H4 as antigen and RA sera showed that the binding of patients' sera was significantly higher than the binding of normal blood donors. Altogether these observations proved that RA sera react with deiminated H4.

The presence of deimination was demonstrated on the tryptic digests deriving from activated neutrophils using a chemical modification with antipyrine and 2,3-butanedione, which are the same reagents employed in the Anti-modified Citrulline

detection kit, a very frequently used tool to detect deiminated proteins by immunoassays.[79] The reaction includes a two steps addition of the reagents to the ureido group of citrullines gaining a final mass increment of 238.11 Da in the citrullinated peptide whose MALDI spectrum also shows a characteristic isotopic pattern for the presence of minus 1 and 2 Da signals.[30] The experimental conditions were optimized on several citrullinated and not citrullinated peptides pairs and on mixtures of low concentrated citrullinated peptide spiked into a BSA tryptic digest, thus miming a real biological sample.

When these derivatization experiments were performed on the nuclear and NETs samples, a signal at m/z 1720.8, which was absent before, become visible in the MALDI spectra of the derivatizated proteolytic digests. This signal was compatible with the mass expected for the derivatizated histone H4 peptide DNIQGITKPAI(cit)R with MH^+=1482.8. The low intensity of the derivatized peptide's signal prevented the precursor to be isolated and fragmented through tandem mass spectrometry and thus the deimination was not confirmed by a corresponding MS/MS spectra.

Moreover the derivatization technique did not provide information regarding the exact deimination position since two arginines were available to modification and in the literature there are no formal evidences that allow to exclude citrullination of the C-terminal ones.[29]

The second part of the research was a deeper investigation conducted on NETs trypsin digest to better delineate the deimination pattern of H4 starting from the peptide DNIQGITKPAIRR whose modification status has been demonstrated through derivatization. In addition we performed a search for possible further citrullinated H4 peptides, exploiting a different strategy.

The challenges in the detection of *in vivo* citrullinated proteins through proteomic methods are clearly presented in a recent review from which important considerations emerge.[29] The majority of the approaches described in the literature to highlight the presence of citrulline residues have been tested either on standard peptides, available in large amounts, or on standard proteins deiminated *in vitro*. Only few of these studies face the fundamental problem of working with proteins citrullinated *in vivo* that is the very low amount of citrullinated peptides which dare the analytical sensitivity of the mass spectrometers.

Another important consideration is that the combination of more than one approach is necessary in order to provide complementary information, difficult to be achieved by a single method.

In the context of this project, considering the issues left unresolved by the derivatization approach initially implemented, a direct detection of citrullines residues by means of an

LC-MS/MS analysis appeared as the most appropriate, since the direct interpretation of citrullinated peptides' MS/MS spectra could provide the requested sequence.
This approach is much more troublesome, if compared to the modification strategies, and in fact only few articles report its successful implementation.
We decided to test two different stationary phases since we were aware that the chromatographic separation played the most important role in this analysis. The HILIC phase was found to be more efficient than the reverse phase in the separation of a pair of standard citrullinated and not citrullinated peptides, allowing us to assume that it could be successfully applied to the NETs tryptic digest. Both the analyses clearly demonstrated that the peptide at MH^+=1482.8 did not correspond to the sequence DNIQGITKPAIRR but to the isobaric semitryptic peptide RDNIQGITKPAIR, thus changing the possible deimination site of H4. However several difficulties were met that prevented the establishment of the exact deimination position on that peptide and did also not allow to asses which modification was responsible for the 0.984 Da mass increase found in two other histone peptides, DNIQGITKPAIR and KLPFQR.
Regarding the peptide RDNIQGITKPAIR the presence of the higher concentrated unmodified form and a form bearing a presumable aspartate amidation severely affected the analysis by eluting in close proximity to the citrullinated, thus hampering a correct peptide sequencing to be achieved. Moreover the presence of an isobaric peptide of the citrullinated contributed to make more complicated the interpretation of the results by reducing its sampling by the mass spectrometer.
Taken together, these observations confirm that a direct detection of citrullinated peptides in real samples by interpretation of their MS/MS spectra is a quite challenging task because of both their low amount, combined to the presence of unmodified peptides that act as interfering agents. Taking into account the complexity of the sample, we do believe that to move forward this project it will be necessary to optimize the chromatographic method on a tryptic digest of *in vitro* deiminated histone H4, in order to reproduce the conditions of the real sample, but at higher concentrations. Moreover, we concluded that the use of HILIC phase, which proved to be more efficacious than the reverse phase when applied to standard peptides, requires a still longer phase of conditioning in order to create the optimal conditions for the separation mechanisms characteristic of these columns.

In conclusion the present research shows two different interesting aspects.
On one hand the description of autoantibodies reactive with citrullinated histones raises important questions on the mechanisms leading to their production. The prevailing idea is that neutrophils by NETosis expose autoantigens to immune cells in the presence of danger signals, but whether this is sufficient for the induction of autoantibodies in a normal or autoimmune background is still to be settled. As neutrophils are the most

abundant cells in synovial fluid, NETosis may create targets for ACPAs in rheumatoid synovia, where these autoantibodies are produced. The analysis of NETs composition in healthy subjects and in RA patients and the study of their immunogenic properties may help to unravel the relationship between NETosis, citrullinated histones and induction of ACPAs in RA patients.

On the other hand the need to clearly highlight the presence of deiminated residues in the protein target by means of proteomic technologies represents a methodological challenge that has been faced both using an already described derivatization strategy and trying a different approach, which has been previously barely described. A direct detection of citrulline residues still requires considerable efforts, as confirmed by the few successful examples reported in the literature. Accordingly, it is worthy to proceed with the idea of using different stationary phases that may offer a solution to the still unresolved problem of the separation of citrullinated peptides from the corresponding unmodified forms.

1.5 EXPERIMENTAL PART

Neutrophils isolation and stimulation

Neutrophils were obtained from buffy coat of healthy blood donors and isolated using discontinuous gradient centrifugation according to *English et al.*[56]

Purified neutrophils were resuspended at 2x106 cell/ml in Locke's solution (NaCl 150 mM, Hepes-HCl 10 mM pH 7.3, $CaCl_2$ 2 mM, Glucose 0.1%) and incubated at 37°C for 15 min with or without A23187 4 µM (Sigma-Aldrich, Saint Louis, MO, USA), followed by a further incubation of one hour in serum-free RPMI at 37°C.

NETs preparation

Neutrophils were seeded in 6 wells plates in RPMI at 2x106 cells/w and activated with 100 nM phorbol myristate acetate (PMA) for 4 hours at 37°C.[80] After removing the medium, the wells were washed 2x10 min with D-PBS and incubated for 20 min at 37°C with 10U/ml DNase I (Sigma) in RPMI. DNase activity was stopped by adding EDTA 5 mM (final concentration). The samples were then centrifuged at 3000g to remove intact cells and intact nuclei; the supernatants containing NETs proteins were processed as described below.

Histone extraction

Stimulated and unstimulated neutrophils and NETs were incubated overnight in H_2SO_4 0.2 M at 4°C with agitation. Acid extracted proteins were then precipitated with 33% trichloroacetic acid (TCA) for 2 hours at 4°C, washed twice with acetone and suspended in ddH_2O. Protein concentrations were determined using BCA Protein Assay (Pierce, Rockford, IL, USA).

SDS-PAGE and immunoblotting

Acid extracted proteins from stimulated and unstimulated neutrophils and from NETs were resolved in a 16.5% Tris-Tricine-SDS PAGE (Bio-Rad, Hercules, CA, USA) under non reducing conditions and blotted onto PVDF (Millipore, Billerica, MA, USA). The membrane strips were saturated for 30 min at room temperature in TBS containing 5% BSA and 0.05% Tween-20 and incubated overnight at 4°C with human sera diluted 1:200, purified antibodies 50 µg/ml , anti Histone H4 (Upstate, Millipore), and anti histone H4 (citrulline 3) (Upstate, Millipore) rabbit antisera diluited 1:500.

Anti Citrulline (modified) detection kit (Upstate, Millipore) was used to detect deiminated proteins, following the manufacturer's instructions.

Peroxidase activity was visualized by means of enhanced chemiluminescence using LuminataTM Western HRP Substrate (Millipore). Images were acquired and analyzed using the VersaDoc Imaging System and QuantityOne analysis software (Bio-Rad).

In-gel digestion

The gel bands of interest were chopped into small pieces, washed with water and shrinked with acetonitrile. Reduction and alkylation were performed in a solution of 10 mM DTT (56°C, 45 min) and iodoacetamide 55 mM (room temperature, 30 min, in the dark), respectively. Shrinking and rehydratation were carried out with acetonitrile and 100 mM ammonium bicarbonate in order to completely remove Coomassie staining. Proteins were digested by adding a solution of 12 ng/µl of trypsin (type IX-S, from porcine pancreas, Sigma Aldrich) in 10 mM ammonium bicarbonate (37°C, overnight). Digestions were stopped with 10% TFA and the supernatants were recovered for subsequent mass spectrometry analyses.

Chemical modification

Derivatization of citrulline residues was performed as described previously.[30,31] Briefly 10 µl of proteolytic digest were added to 10 µl of 50 mM antipyrine in presence of 20 µl of 10% TFA; 10 µl of freshly prepared 2,3-butanedione 50 mM were, finally, added to start the reaction. After 2h of incubation at 37°C, in the dark the samples were SpeedVac-concentrated up to 20 µl and desalted by Zip-Tip C18 pipette tips (Millipore).

MALDI-TOF analysis

MALDI-MS measurements were performed using an Ultraflex III TOF-TOF instrument (Bruker Daltonics, Bremen, Germany), operating in reflectron positive ion mode between a mass range of 860-4000 m/z. Ions were formed by a pulsed UV laser (λ = 337 nm) beam. The instrumental conditions were: UIS1 = 25 kV; UIS2 = 21.65 kV; reflectron potential: 26.3 kV; delay time = 0 nsec.

External mass calibration was done using the Peptide Calibration Standard, basing on the monoisotopic values of $[M+H]^+$ of Angiotensin II, Angiotensin I, Substance P, Bombesin, ACTH clip (1-17), ACTH clip (18-39), Somatostatin 28 at m/z 1046.542, 1296.685, 1347.736, 1619.823, 2093.087, 2465.199 and 3147.471, respectively.

For the analyses of the proteolytic digests the dried droplet deposition method was applied using a solution of α-cyano-4-hydroxycinnamic acid spotted in an AnchorChip (Bruker Daltonics) target plate. Peptide mass fingerprint data were obtained for each sample and MS-MS analyses were performed on selected precursors. MALDI-TOF-TOF experiments were carried out using the LIFT device. The instrumental parameters were: UIS1=8 kV; UIS2=7.2 kV; ULIFT1= 19 kV.

Mascot searches were done against the Swiss-Prot database (release 2012_10) calculating 2 maximum missed cleavage for trypsin, carbamidomethylation of cysteine as fixed modification, oxidation of methionine as variable modification, 30 ppm and 0.6 Da as precursors and fragments tolerances, respectively.
For the analyses of the derivatized digests the same instrumental conditions were applied.

LC-MS/MS analysis

Peptide analysis were performed on a *nano*LC-*nano*-ESI-LTQ-Orbitrap system (Thermo Scientific, San Jose, CA, USA).
The *nano*LC system used was an UltiMate™ 3000 apparatus (Dionex) which comprised a LPG-3600 dual-gradient pump, a FLM-3000-Flow manager module and a WPS-3000 thermostatted well plate autosampler (set at 5°C).
Reversed-phase separations were performed on a Acclaim® PEPMAP100 C18 *nano*column (Dionex) (3 µm particle size, 100 Å; column dimension: 75 µm id x 5 cm) using the four different eluents. For the loading sample mobile phase A was made of 0,1% TFA while phase B of acetonitrile in presence of 0,1% TFA; for the gradient mobile phase A was made of water:acetonitrile 97:3 with 0,1% of formic acid while phase B of water:acetonitrile 3:97 with 0,1% of formic acid.
1 µl of the standard peptide or 5 µl of tryptic digest were loaded through µlpickup injection onto a Acclaim® PEPMAP100 C18 trapping column (Dionex) (5 µm particle size, 100 Å; column dimension: 100 µm id x 5 mm) at a flow rate of 20 µl/min. The 60 min lasting gradient proceeded at 300 nl/min and included the following steps: for the first 10 min the phase B increased from 4 to 40 %, from 10 to 30 min it increased until the 65%, from 30 to 35 min it remained at 65%, in a minute rapidly reached the 90% were it lasted for 4 min until 40 min, in a minute phase B decreased to 4% where it remained until 60 min.
HILIC separations were performed on a ZIC® HILIC columns (Sequant, Umèa, Sweden) (5µm particle size, 200 Å; column dimension: 100 µm id × 10 cm) using eluents with the same composition of the previously described.
Since no ZIC® HILIC trapping columns were equipped on the LC system, a direct on-column loading was performed by injecting 1 µl of sample through a full-loop injection. Peptides were separated using a mobile phase gradient of 125 min, at a 300 nl/min including the following steps: 10 min of 97% phase B to allow the sample to reach the column after the injection, from 10 to 60 min the phase B decreased until the 50% where it remained until 70 min, next 5 min it came back to the 97% of B where it lasted until 125 min.

The LTQ-Orbitrap mass spectrometer was equipped with a *nano*ESI source through a positive ion electrospray ionization with a spray voltage of 3.1 kV being applied to the needle.
The other spectrometer's tuning parameters were: capillary voltage at 45 V, tube lens voltage at 230 V, capillary temperature at 220°C.
Two different MS method were applied. In the data-dependent acquisition (DDA) the cycle consisted in a survay MS scan with 15000 resolving power within 300-2000 Th, followed by MS/MS fragmentation of the three most intense precursor ions by collision-induced dissociation using a normalized collision energy of 30% in the linear trap. In top 2 each full mass spectrometric (MS) scan (15000 resolving power) was followed by two MS/MS scans performed in DDA modality and one MS/MS scan performed on the target m/z value of 741.9 Th ; the collision energy was of 30%.
Single charge precursors were automatically excluded from MS/MS acquisition and m/z of precursors already fragmented was dynamically excluded (repeat count 2, repeat duration 15sec, exclusion duration 30sec).
Calibration of the Orbitrap analyzer was performed by infusion of a Thermo standard calibration mixture (caffeine, tetrapeptide MRFA, and Ultramark).
LC-MS/MS data was controlled by the Chromeleon software integrated with Xcalibur 2.0.7 (Thermo Fisher Scientific) to manage both the chromatographic and mass spectrometric aspects. Each run was converted to .mgf file using RAW2msm utility using default parameters (intensity-weighing the parent ion m/z over the LC elution profile and keeping top 6 most intense fragment ions per 100 m/z units).

Data Analysis

Peak lists were searched on a local server using the Mascot 2.2 algorithm (Matrix Science Ltd., London, UK), selecting semitrypsin as the proteolytic enzyme and allowing two missed cleavages, against SwissProt databank (release: 2012_05). Oxidation (Met) and deimination (Arg) were selected as variable modifications while carbamidomethylation (Cys) was selected as fixed modification; mass tolerance for precursor ions was set at 10 ppm and for fragment ions at 0.6 Da. At least 2 significant sequences for each protein were selected as criterion filter to accept identified proteins. For MS/MS searches individual ions scores >35 were reported as indicative of identity or extensive homology ($p<0.05$).

CHAPTER 2

pH regulated formation of side products in the methylation and ethylation approach for differential labeling of peptides in relative quantitative experiments

2.1 INTRODUCTION

2.1.1 Quantitative proteomics

Biological and biomedical research deals with the study of processes and molecular mechanisms that take place inside cells at different levels (e.g. genomic, proteomic, metabolic) and that support all the cellular functions (e.g. proliferation, communication, surviving, development). These cellular processes and functions are influenced and regulated by various stimuli and the recognition of the biological changes correlated to them is a crucial task.[81,82] Toward this aim it is evident the need to set up analytical tools capable to measure and quantify these changes; accordingly, proteomics is playing an increasingly important role enabling to measure the expression level of proteins, as tangible and measurable response to perturbation-induced changes within cells. In fact in the biomedical field the possibility to correlate anomalous cellular responses, such as increased/decreased expression of specific proteins, with the establishment of a pathological condition represents a desirable scenario.

Proteomic technologies can provide a significant contribution comparing samples from healthy and diseased subjects and mapping possible quantitative differences associated with specific conditions. The fast development of mass spectrometry by both a methodological and an instrumental point of view, together with a contemporary improvement in efficiency of HPLC techniques and materials, has provided powerful tools. Various approaches have been developed using 2D gels or MS techniques, the latter possibly based on chemical derivatization (Figure 2.1).[83]

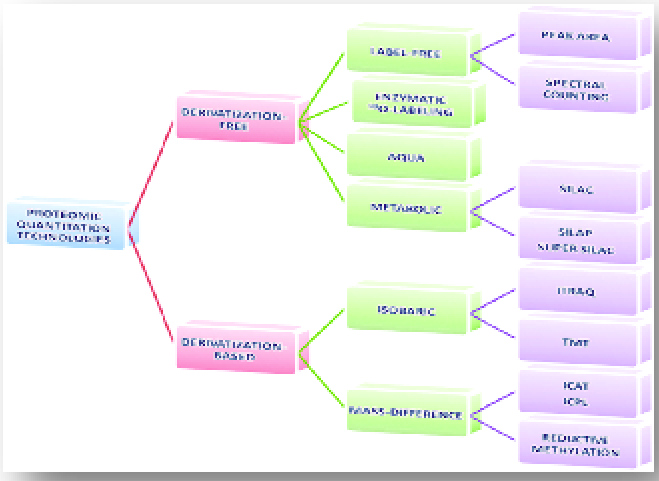


Figure 2.1. Quantitative proteomics toolbox[83]

Derivatization-free technologies include different strategies which have in common the easy sample preparation and the lowest possible limit of quantification.[83] However they provide less accurate measurements, reason for which efforts are required to keep a very good reproducibility at each experimental step. On the other hand development of chemical derivatization-based methods complies the need of increasing the throughput screening of samples by means of a single experiment. Indeed differential labeling allows a simultaneous analysis of different conditions (namely "multiplexed quantitation"), thus minimizing the systematic variation.

A basic concept of all MS-based methods is that MS itself is not a quantitative technology because mass spectrometric response, besides analyte amount, is also dependent on various other factors such as sample preparation, instrument conditions and analyte physicochemical properties. Nevertheless quantitative proteomics is asked to compare results coming from different analyses, situation that clearly collides with this problem. For derivatization-free methods this standstill has been overtaken by using standardized procedures based on highly performing LC-MS instruments, several experimental replicates for each condition and statistical analyses in order to ensure reproducibility and statistical significance. At variance, in the case of derivatization-based methods, a single analysis of the two mixed samples is sufficient to overcome that problem, since each couple of native and stable isotope-labeled peptides share the same chromatographic and mass spectrometric properties.[82]

In the last two decades a great effort has been done to develop chemical labeling techniques for relative quantitation experiments by employing chemical reagents able to specifically react with functional groups of proteins/peptides (e.g. sulfhydryl, amine and carboxyl groups) and modify them with labeled tags (Figure 2.2).

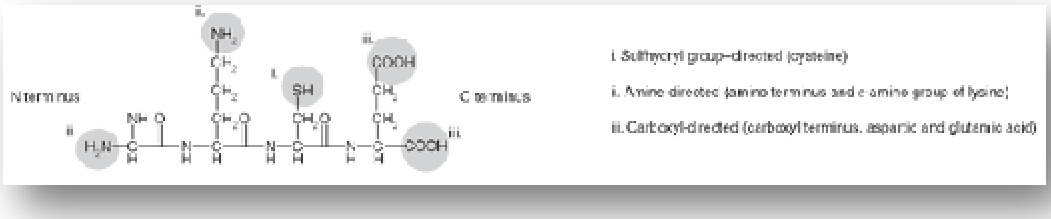


Figure 2.2. *Stable isotope-bearing chemical reagents target reactive sites on a protein or peptide*[84]

The rational of employing MS strategies based on derivatization methods for quantitative analysis relies on the resolution of mass spectrometers, namely their power in separating isomers with different stable isotope composition (isotopomers).[84] As a general evaluation, an ideal and reliable isotopic labeling reaction must produce the smallest possible increase of sample complexity, thus it should be very selective, quantitative, possibly fast, and give no side-products. Labeling reactions with different

stable isotope-bearing chemicals lead to the formation of "heavy" and "light" isotopomeric peptides coeluting in LC runs, or showing an almost negligible chromatographic isotopic effects (Figure 2.3 A).

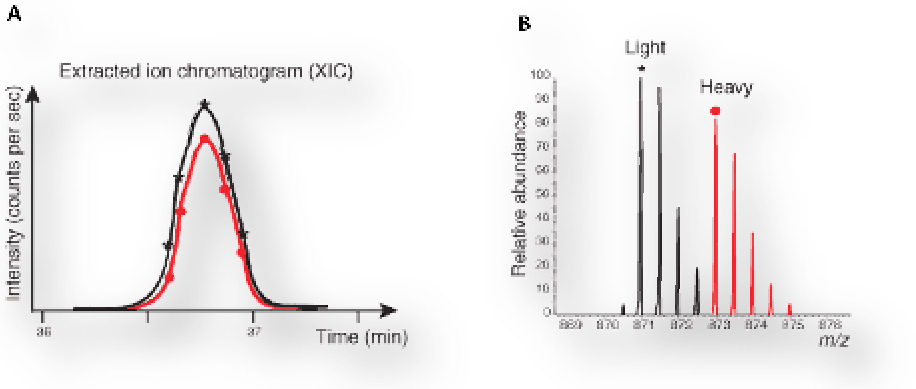


Figure 2.3. Co-elution of differentially labeled peptides (A) and corresponding MS spectrum showing the isotopic cluster for each of them (B)

A good separation of the signals of the isotopomers is also mandatory because overlapping between peaks corresponding to light and heavy labeled substrates would result in distorted ratios and misrepresenting quantitations (Figure 2.3 B). A mass difference of 4 Da is sufficient in the mass range below 1900 Da, but for heavier molecules a significant overlap between signals occurs and a wider spacing, in terms of mass units, is necessary.

The most commonly employed commercial reagents are designed to fulfill the above mentioned requirements and to offer different approaches according to researchers' needs and instrumental equipment, as they all have their advantages and disadvantages. Through these kits the quantitation step can be carried out at MS level by using isotopic tags as in the case of ICAT™ (isotope-coded affinity tag) and ICPL™ (isotope-coded protein labeling) or at MS/MS level by using isobaric tags as iTRAQ™ (Isobaric tags for relative and absolute quantitation) and TMT™ (Tandem Mass Tag). The ICAT™ reagent, for example, selectively labels cysteine residues at the peptide level and quantitation is performed in MS spectra after an avidin-affinity chromatographic step.[85] This approach has the great advantage of reducing the complexity of the sample but sometimes the low abundance of cysteine residues can be a crucial drawback as information about cysteine lacking peptides are lost. The more recently developed ICPL™ kits ensure a wider coverage of labeled peptides as multiplex quantification can be performed at the MS level, with a selective reaction at the primary amino groups on proteins, prior to digestion.[86] Isobaric reagents, TMT™ and iTRAQ™ produce no increase to the sample complexity nevertheless a slight instability

of reagents and the quantitation performed on a single MS2 spectrum represent a weakness of this strategy.[87,88]

The difficult synthesis of these reagents and their good efficiency always reflects into high costs, which can be sometimes a considerable problem. Cheaper reagents exist and can be also employed but they sometimes suffer of poor selectivity, give unstable products, or are selective for low abundant residues.

In this scenario stable-isotope labeling strategies based on reductive amination of peptides with formaldehyde or acetaldehyde as alkylating agents represent interesting alternatives. The reaction is simple, fast, specific and undoubtedly inexpensive if compared with commercially available kits, an ensemble of features that support its employment in quantitation experiments in proteomics.

2.1.2 Reductive amination

Chemical tagging by reductive amination involves any primary amino groups present in a peptide or a protein, namely all the N-terminal amino groups and all the ε-amino groups of lysine residues. Reaction with aldehydes generates an imine that is reduced to secondary amine with $NaCNBH_3$; then a further sequence of addition and reduction leads to a tertiary amine as the final product. By working in a reducing environment with an excess of formaldehyde or acetaldehyde, dimethylation or diethylation of amino groups is respectively obtained. According with the presence or the absence of a C-terminal lysine, mono or di-derivatized products are expected upon labeling peptides generated by digestion with trypsin.

For a relative quantitative analysis two proteins from samples of different conditions are digested by protease then labeled with d(0)-formaldehyde and d(2)-formaldehyde respectively, as example, and finally mixed together to be analyzed by LC-MS. The differentially labeled peptides are co-eluted in the chromatographic step and a pair of signals is detected at the MS level with a mass shift of 28 Da and 32 Da respectively, for each derivatized reactive site (Figure 2.4).

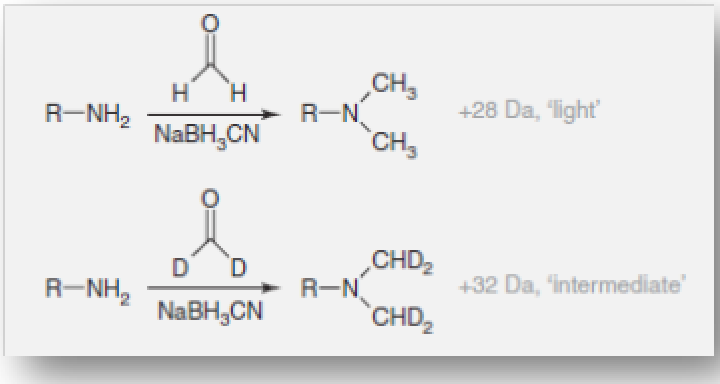


Figure 2.4. Reductive amination using d(0)-formaldehyde and d(2)-formaldehyde in presence of $NaCNBH_3$

With d(0)-acetaldehyde and d(4)-acetaldehyde, instead, we can obtain a mass increase of 56 and 64 Da respectively for each derivatized site, thus a mass separation between the differentially labeled peptides that grows up to 16 Da for doubly tagged substrates.

The application of dimethylation in stable isotope labeling procedures for quantitative purposes has been firstly reported by *Hsu et al* demonstrating the reaction to be fast (5 minutes) and quantitative, working at pH 5-6.[89] In other papers modifications have been made to the protocol, basically changing the reaction buffer and increasing the incubation time up to one hour.[90,91] Multiplex experiments have been recently reported employing also d(2)-^{13}C-formaldehyde and $NaCNBD_3$ without a sensible increase in terms of cost.[92] Although the mild reducing agent, $NaCNBH_3$, is known to be more selective and efficient towards protonated imines, dimethylation has also been performed with neutral and slightly basic conditions and little effect of pH change was observed in a systematic investigation with formaldehyde, in the range between 3.0 and 8.2.[93]

Moreover it has been described that the conversion of amino groups from primary to tertiary increases the proton affinity of the derivatized sites, positively affecting peptides coverage in the identification process and improving *de novo* sequencing.[94] Concerning this, a further useful aspect is the enhancement of the a_1 ion in MS/MS spectra that has also been reported to allow quantitation at the protein level.[95]

On the other hand the use of diethylation labeling for peptide quantification has found less confirmations as evident from a poor literature concerning the use of acetaldehyde as labeling agent, except for small molecules.[96] Only recently a diethylating labeling based on different acetaldehydes isotopomers and 2-picoline borane instead of $NaCNBH_3$ has been successfully applied on the proteome analysis of an unicellular algae.[97]

The chemistry of both formaldehyde and acetaldehyde in biological systems features interesting aspects for proteomics investigations but it has been deeply investigated only in few papers. Formaldehyde has been widely used to fix tissues and cells and it is known as a cross-linking agent involving a wide set of amino acid residues.[98] Besides reacting with the N-terminal amino groups and with lysine it is also able to attack tryptophan, histidine, arginine and cysteine residues to form a metylol adduct that, only in the case of primary amino groups, undergoes rearrangement to imine. A further reaction can involve the generated Schiff-base and other residues, like glutamine, asparagines, tryptophan, histidine, arginine, cysteine, and tyrosine, in intra or intermolecular cross-links. *Metz et al* reported a systematic investigation on a series of synthetic peptides, each one featuring a different residue likely reactive towards formaldehyde, in the presence of glycine or a reducing agent.[99,100]
Considerations rising from these studies are important to delineate the reactivity of formaldehyde toward chemical groups exposed by the proteins and, in a proteomic perspective, invite not to underestimate its ease of use as derivatization agent for quantitation experiments.

2.1.3 Quantitation using the Mascot Distiller software

Whatever the labeling strategy used for quantitative proteomic experiments, proper algorithms of analysis are asked to extrapolate the ultimate light/heavy (L/H) ratio. Moreover the analysis of proteomic samples requires the employment of LC-MS/MS analysis through which the co-eluting isotopomeric couples of peptides could be isolated for subsequent fragmentation. Corresponding peaks list containing all the precursors and the products related information (mass, charge, intensity) could be submitted to data bank search and then processed to get quantitation results.
Concerning this workflow it is important to note that the quantitation step follows the identification step therefore the number of proteins available for quantitation is limited to the proteins that are recognized in all the samples that are tested (Figure 2.5). Protein abundance and sample complexity are significant factors that affect the availability of proteins for mass spectrometric quantitation.

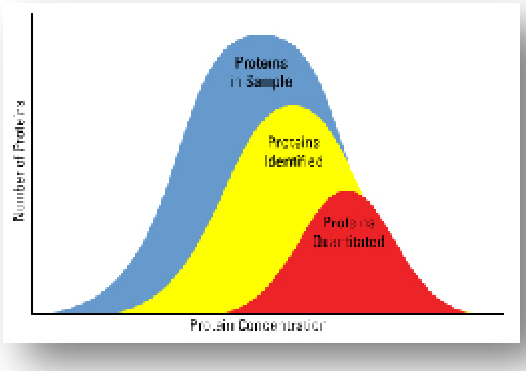


Figure 2.5. Protein availability for quantitative proteomic analysis is limited

Different approaches can be followed to achieve protein quantitation using mass spectrometry data.[101] Measuring the relative intensities of fragment peaks at fixed m/z values within an MS/MS spectrum (reporter method) or of sequence ion fragment peaks within an MS/MS spectrum (multiplex method) or considering the relative intensities of extracted ion chromatograms for precursors within a single data set (precursor method) represent few examples.

The Mascot Distiller software is able to perform quantitation analysis applying some of the above mentioned protocols. In particular for reductive amination it makes use of the precursor method, since any chemistry that creates a precursor mass shift adopts this method.

The precursor protocol requires information from the raw data file that are not present in the peak list, so the quantitation report is generated by Mascot Distiller, which has access to both the Mascot search results and the raw data. It processes the raw data and submits the search to Mascot Server; when the search is over results are returned to Distiller that can generate the quantitation results reporting the final L/H protein ratios (Figure 2.6).

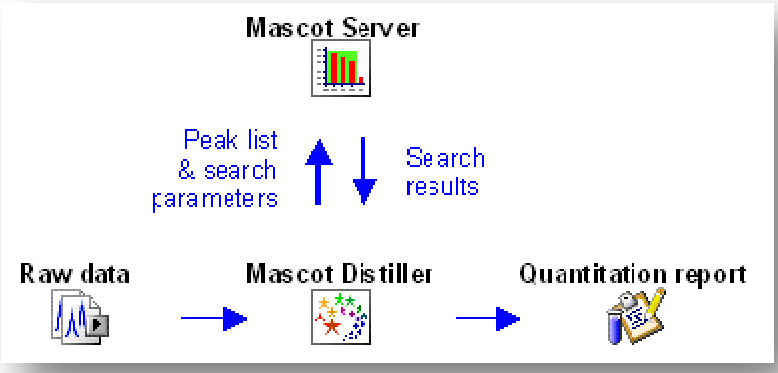


Figure 2.6. Mascot distiller workflow

Quantitation is performed at peptide level but reported at the protein level; the final protein ratio will result improved as more peptide measurement data points are provided. To accept a peptide ratio Distiller uses three parameters that measure the quality of the experimental data:

- Fraction refers to the fraction of the peak area in the precursor region accounted for by the labeled peptide (called component); this is to control the presence of interfering peaks in the same area.
- Correlation coefficient is a measure of the fitting between the predicted and the observed precursor isotope distribution. This second parameters tests the shape of the distribution to avoid interference from other peptides.
- Standard error reflects the variability around the reported ratio that is the result of a least squares fit to the heavy and light components from the scans in the XIC peak. The standard error is a measure of the reliability of the ratio.

The final protein ratio can be calculated from the set of peptides as:

- Median if it is the median of the peptides ratio
- Average if it is the geometric mean of the peptides ratios
- Weighted if it is calculated from a weighted average of the peptide ratios. The weighting factors are the summed intensities of the components.

In conclusion Mascot Distiller implements quantitation measuring the relative intensities of extracted ion chromatograms for precursors in survey scans and uses a quantitation method that encapsulates all the settings for searching, processing, and reporting of quantitation data.

2.2 OBJECTIVES

One of the most crucial issues and growing area of modern proteomics is the relative quantification of the differential expression of proteins in two or more samples representing different conditions of biological systems.

Among the most common stable-isotope labeling strategies, the reaction of aldehydes with peptides in the presence of $NaCNBH_3$ features many attractive aspects that enable its employment in quantitation proteomics experiments.

The reactivity of aldehydes toward proteins has been deeply investigated only in few papers, reporting how the Schiff-base generated in the reaction with primary amino groups is sensitive to any nucleophile site of the protein, easily undergoing cyclizations and cross-linking. The almost simultaneous addition of a reducing agent should repress any other reaction but the dialkylation of primary amines and indeed, only slight problems with side-reactions have been reported, however easily overcome by reducing reaction time.

Considering the known predisposition of aldehydes to promote cross-reactions with proteins, we estimated appropriate to study more in detail the impact of the reductive amination with either formaldehyde and acetaldehyde in proteomic quantitative experiments.

Moreover, considering the limited available literature and the variability in terms of pH and reaction times in the reported protocols, we wanted to fix general rules for their employment in stable isotope peptides labeling, particularly for the diethylation approach. Indeed acetaldehyde and its heavy isotopomers can be used as diethylating agents for providing a wider mass separation between the differentially labeled peptides, attractive condition when working with multiply charged peaks on low resolution mass spectrometers.

A parallel investigation was performed on synthetic peptides changing reaction conditions, in order to evaluate in particular the role of pH, since this was the most divergent parameter among the described protocols. Based on preliminary observations derived from both the methylation and the ethylation of model peptides, a consistent series of tests were performed trying to highlight the nature of unexpected collateral signals.

Before undertaking experiments on complex biological samples, the occurrence of unwanted side-products was monitored on a single protein digest and a mixture of proteins digest.

2.2.1 Scheme of the work

The research was developed as follows:

- Two commercially available peptides, TDP and ACTH, were labelled with both formaldehyde and acetaldehyde, testing buffers at different pH values and different reagents concentration. Reactions were monitored by MALDI-TOF MS.
- Standard tryptic peptides were employed to deeply investigate the nature of a signal at minus 2 Da than the expected and whose incidence was found to be pH related and more intense for acetylated peptides. ESI-MS acquisitions, LC separations combined to peptide sequencing experiments and guanidination reactions allowed the situation to be clarified.
- Synthetic peptides were expressly produced with amino acids substituted in critical positions in order to evaluate their role in the occurrence of a signal at minus 4 Da than the expected, in MALDI spectra. Their specific behavior in reductive amination and guanidination reactions was studied to delineate its characteristics and to limit its formation.
- More complex samples including a single digested protein and a mixture of digested proteins were differentially labeled and LC-MS/MS analyzed; a Mascot Distiller processing was applied and evaluations concerning the impact of the side-products were extrapolated.

2.3 RESULTS AND DISCUSSION

2.3.1 Dimethylation and diethylation labeling of standard peptides

Preliminary tests were performed on two commercially available peptides, DRVYIHPFHLVIHN (renin substrate tetradecapeptide human, TDP, EM: 1758.9) and RPVKVYPNGAEDESAEAFPLEF (adrenocorticothropic hormone fragment 18-39, ACTH, EM: 2464.2), featuring respectively one and two reactive sites. Although both substrates were not highly representative of a typical tryptic digest, as neither arginine or lysine were present at the C-termini, they were thought to be suitable for some initial experiments. The reactions were performed with both non-deuterated formaldehyde and acetaldehyde employing buffers lacking primary amino groups thus no TRIS or ammonium salts were introduced in the reaction environment.

The behavior of the reaction at different pH value was the first important task to check therefore four buffers were prepared: 0.1 M triethylammonium bicarbonate (pH = 8.8), 0.1 M triethylammonium bicarbonate (pH = 8.0), 0.1 M tetraethylammonium acetate (pH = 6.9) and 0.1 M sodium acetate (pH = 5.3).

The reaction mixtures were analyzed by MALDI-TOF MS to verify the presence of the desired products and the complete absence of the starting material.

For TDP it was observed that increasing the pH a signal at 1785.9 m/z became more evident, featuring two mass units less than the expected dimethylated product of TDP ($[M+H]^+$: 1787.9) with a mass increase of 28 Da (Figure 2.7).

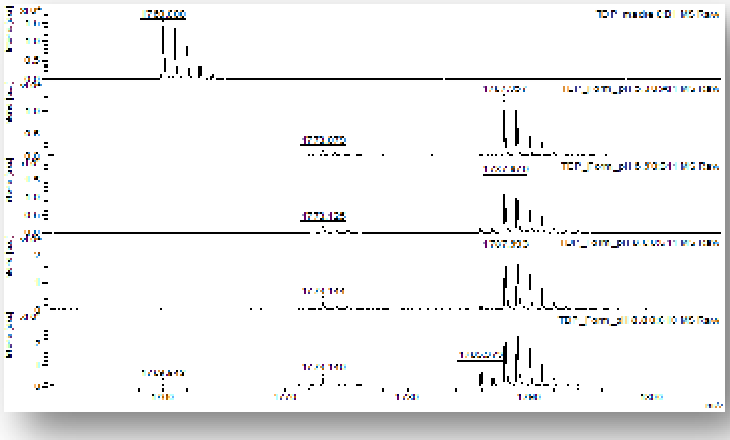


Figure 2.7. MALDI spectra of unlabeled TDP at 1759.9 m/z, dimethylated product at 1787.9 m/z and side-product at 1785.9 m/z after 10 min reaction with formaldehyde, at increasing pH from 5.3 to 8.8

Instead the reaction with acetaldehyde was much slower respect to with formaldehyde. This is not surprising due to a reduced electrophilicity of the secondary iminic carbon.
No side product was observed in acid condition, as well as with formaldehyde, but, on increasing pH, a +54 Da product ($[M+H]^+$: 1814.0), became evident beside the expected +56 Da diethylated compound ($[M+H]^+$: 1815.9), being even the most abundant observed when triethylammonium bicarbonate at pH = 8.8 was employed as reaction buffer (Figure 2.8).

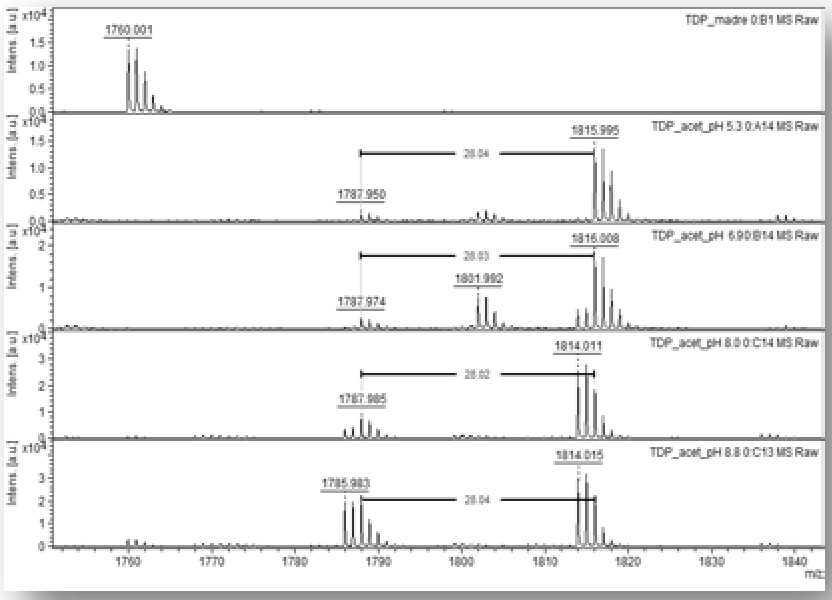


Figure 2.8. MALDI spectra of unlabeled TDP at 1759.9 m/z, diethylated product at 1815.9 m/z and side-product at 1814.0 m/z after 6 h reaction with acetaldehyde, at increasing pH from 5.3 to 8.8

Moreover the ethylation did not reach completeness after 6 h of incubation at 30°C, as evident from the presence of the monoethylated product at $[M+H]^+$: 1787.9.
The presence of signals at +26 or +54 Da for methylation and ethylation, was not reported among the dimethyl labeling approaches described in literature for proteomics quantitative experiments. Only *Li et al.* mentioned the occurrence of a side product with a mass increase 2 Da lower than expected, in reactions carried on for 15 minutes instead of 5.[102]
Metz et al., investigating formaldehyde-derived intramolecular cross-links, reported the presence of a small amount of peptide adduct with mass increase of 26 Da after the addition of the aldehyde and $NaCNBH_3$ to peptides.[99]
The resulting ring structure was found to be an imidazolidinone (Figure 2.9). Moreover formaldehyde and acetaldehyde are known to form stable methylene bridge as results

of intramolecular rearrangement that leads to the formation of an imidazolidinone derivative.[103,104,105]

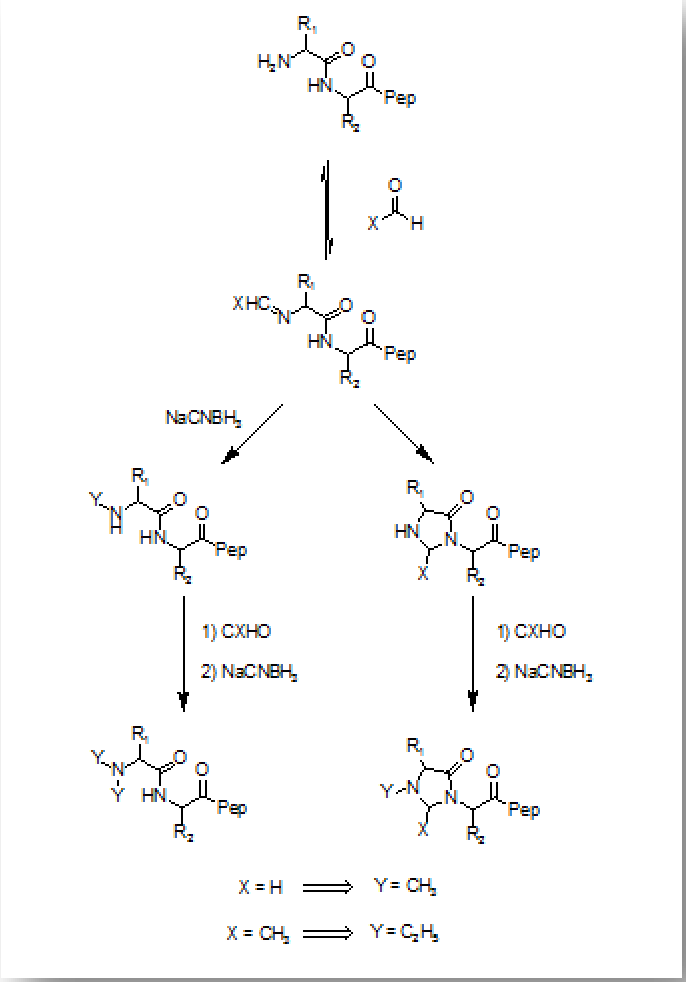


Figure 2.9. The Schiff-base generated by the reaction between aldehydes and the terminal amino group of peptides can be reduced by hydride addiction (workflow on the left) or undergo cyclization by nucleophilic addition of the contiguous peptidic nitrogen (workflow on the right)

The same experimental conditions used for the peptide TDP were applied to the peptide ACTH and, again, unexpected signals were found. In particular the signal of a secondary product with a 52 mass increase ($[M+H]^+$: 2517.3) was observed, beside the signal of the tetramethylated peptide ($[M+H]^+$: 2521.3) in the reaction with formaldehyde at the very basic pH. No side-products were observed performing the derivatization in acidic conditions.

Interestingly there was no evidence with acetaldehyde of a +108 Da product, corresponding to the +52 Da observed with formaldehyde (Figure 2.10).

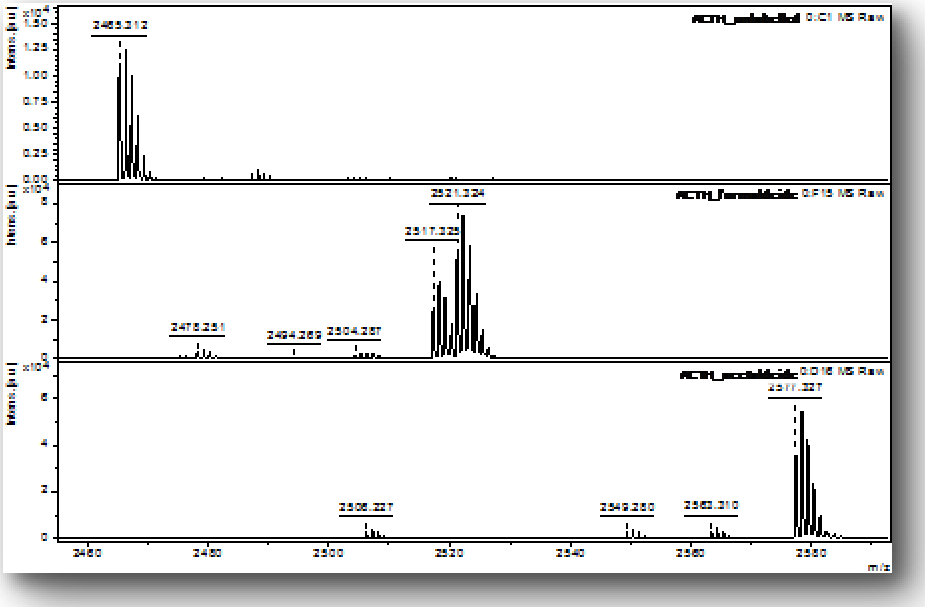


Figure 2.10. ACTH unlabelled; the reaction with formaldehyde and $NaCNBH_3$ yields a considerable amount of a side product with 4 mass units less than the expected tetramethylated product. No side product in the reaction with acetaldehyde

In the same work by *Metz et al* it was, also, described the presence of an adduct with a mass increase of 24 Da (instead of 28) in formaldehyde-treated peptides, in presence of $NaCNBH_3$.[99]

It was demonstrated that two methylene bridge occurred between the side chains of two closed residues, lysine and arginine (Figure 2.11).

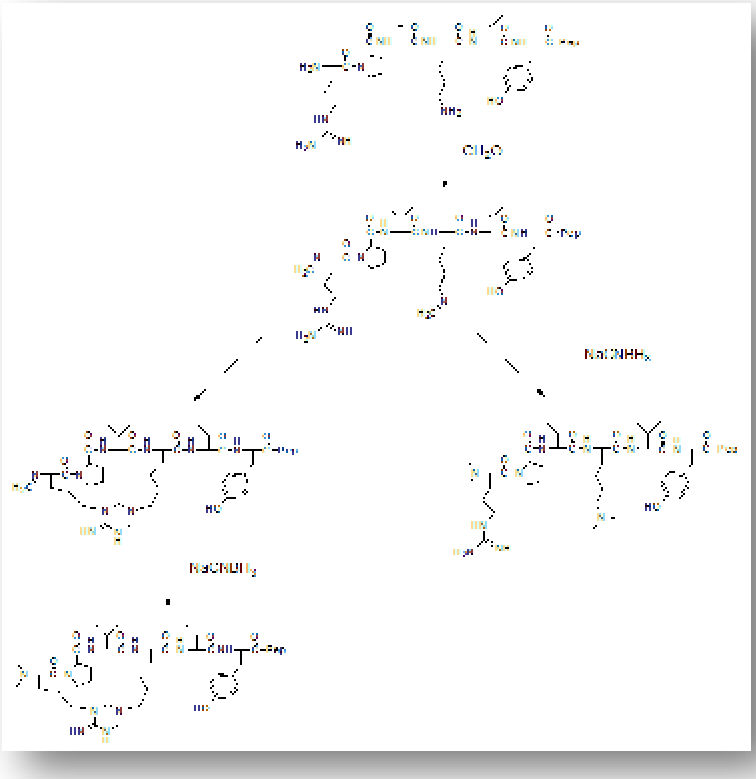


Figure 2.11. Reactive side of ACTH. The Schiff-base at the lysine residue can undergo cross-link with arginine promoting the formation of the peptide adduct with 4 Da minus than the correctly tetramethylated (workflow on the left)

The earlier tests performed on two different peptides showed the presence of two overlapping signals in the MALDI spectra, corresponding to the expected labeled peptides, and to much less abundant products featuring 2 or 4 mass units less.

The 2 Da minus signal observed for TDP peptide was consistent with a rearrangement involving the Schiff-base to generate an imidazolidinone derivative at the N-terminus with the formation of a methylene bond that implies the loss of two hydrogen atoms. The reactions performed with formaldehyde or acetaldehyde revealed a similar trend, with such signals showing a higher intensity, the higher was the pH.

At variance, for the 4 Da minus signals observed with ACTH peptide, the apparent loss of four hydrogens could be ascribed to the formation of two methylene bridges in a rearrangement that probably involves the lysine in position 4 and the N-terminal arginine in an intramolecular cross-link.

Considerations rising from the work of *Metz et al* provided interesting cues, but evidences that similar rearrangements and structures were the real responsible for the signals observed in our experiments should be more consistently obtained. Accordingly, an extensive and systematic investigation was conducted firstly to confirm the presence of the mentioned signals using other peptides, secondly to check if they

were ionization-derived artifacts or real chemical species and finally to explain their real chemical nature.

2.3.2 Insights into the observed side products

Four synthetic peptides were employed to better delineate the nature of the side-chains modifications observed with TDP and ACTH. These new peptides were more representative of a tryptic digest and more closely modeled a real proteomic sample: GTFTASQNYLR (**1**, EM: 1256.6) and EITFTVLASR (**2**, EM: 1135.6), featuring a single reactive site at the N-terminus, and SIHVDIYSFPK (**3**, EM: 1304.7) and SLEVTFTPVIEDIGK (**4**, EM: 1646.9), containing a C-terminal lysine that provides a further derivatizable amino group. Dimethylation and diethylation reactions were performed changing different experimental conditions such as the concentration of $NaCNBH_3$ and aldehydes that was varied in the range 0.2-2.0 M and 0.5-3.0 M respectively. In these conditions where a large excess of reagents were present no substantially differences were found. Moreover even changing the matrix and the solvent for the crystallization during the sample spotting in the MALDI target plate no significant differences were observed.

Instead, similarly to previous results, when reductive methylation of peptides was performed at the four different pH values (5.3; 6.9; 8.0 and 8.8) some differences were observed. No traces of the +26 or +54 Da products were observed at the lower pH (Figure 2.12) while slightly visible signals at the more basic pH were present.

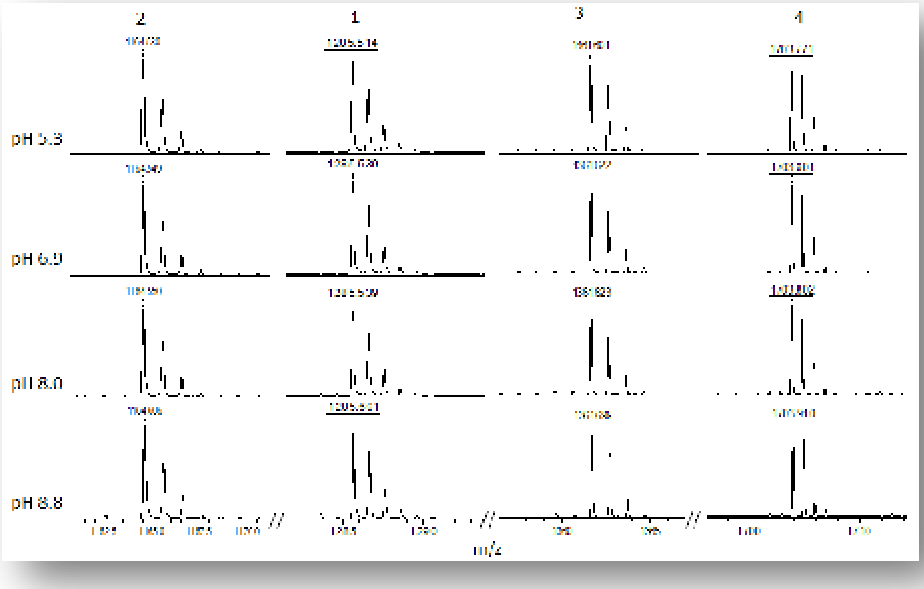


Figure 2.12. MALDI spectra of methylated peptides 1-4 at different pH values

Such results downsized previous observations made on methylated TDP peptide where the presence of the signal was clearly evident at pH 8.8.
On the contrary when the four peptides were subjected to ethylation the pH related differences in the presence of the minus 2 Da signal became evident, similarly to that observed with acetylated TDP. At acid and neutral pH, spectra were almost clear, while at the basic pH of 8.0 and 8.8, in the case of peptides **2**, **3** and **4**, the minus 2 Da signals were even higher than the corresponding ethylated products (Figure 2.13).

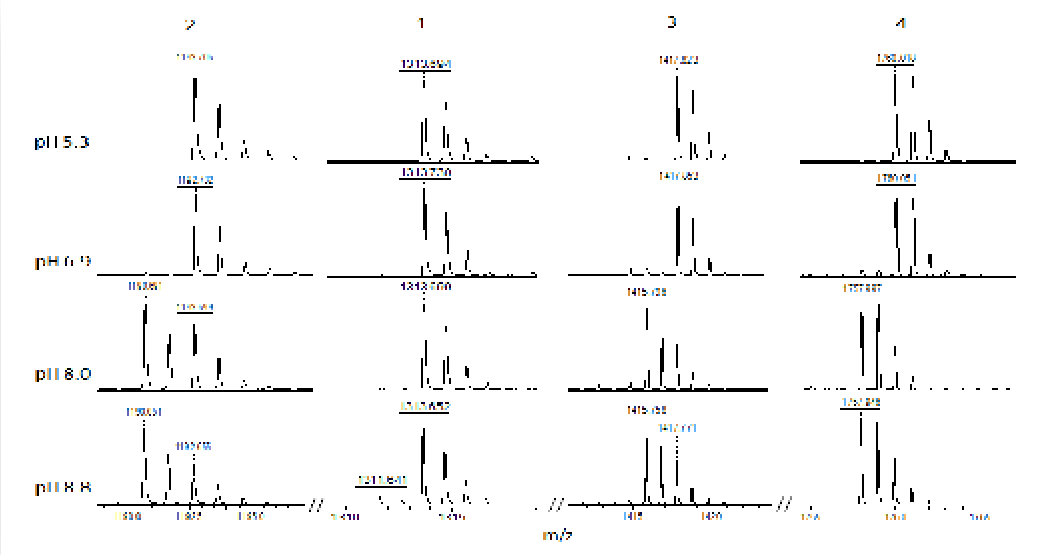


Figure 2.13. MALDI spectra of ethylated peptides 1-4 at different pH values

It was also interesting to note that peptide **1** did not show the presence of this signal even at the more basic pH. It was clear that the chemical environment of the primary amino groups played a crucial role in regulating the competition between the intermolecular addition of hydride and other intramolecular or intermolecular nucleophilic attacks to the generated imine.
All the previous reactions, monitored through MALDI-TOF analyses, confirmed both the presence of the minus 2 Da signals and the general trend observed in the early experiments with TDP.
In order to highlight the nature of the minus 2 Da signals several tests were performed. Initially, the tetramethylated and tetraethylated peptide **3** formed at basic pH was directly infused on an ESI-qTOF Ultima® mass spectrometer to check the presence of the minus 2 Da signals using an ESI source. In both cases the double-charged precursors at 681.48 and 709.33 m/z were preceded by an intense minus 2 Da signal at 680.52 and 708.34 m/z, respectively.

If the minus 2 Da signal was an artifact deriving from the ionization process, both MALDI and ESI were able to produce it and it was not exclusively correlated to the MALDI desorption process (Figure 2.14).

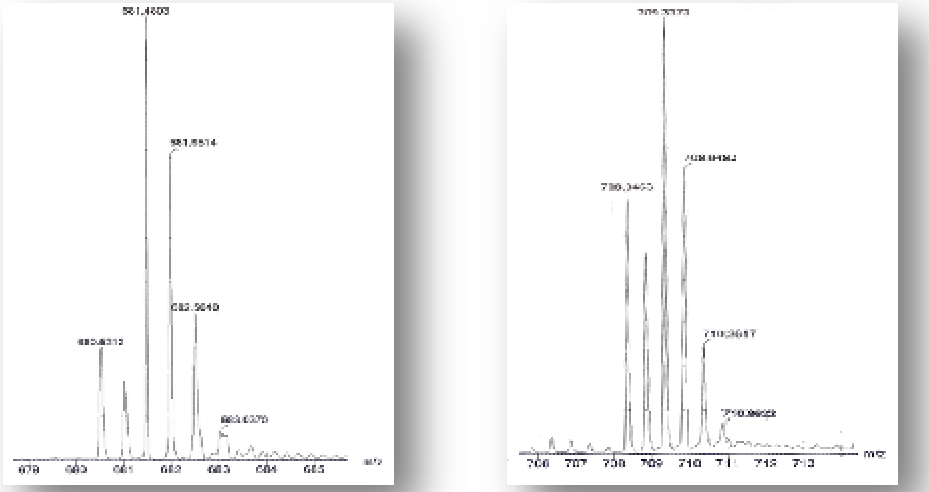


Figure 2.14. ESI spectra of tetramethylated double-charged peptide 3 (681.48 Da) (on the left) and tetraethylated double-charged peptide 3 (709.33 Da) (on the right), at pH 8.8. Evidences of high intense minus 2 Da signals (680.52 Da and 708.34 Da)

To more deeply investigate the nature of the signals a LC separation was performed on the tetramethylated peptide **3**; products were manually collected and the corresponding MS/MS spectra were analyzed. The chromatogram profile showed the presence of two peaks, the first eluting corresponding to the tetramethylated peptide and the second one to the side product (Figure 2.15).

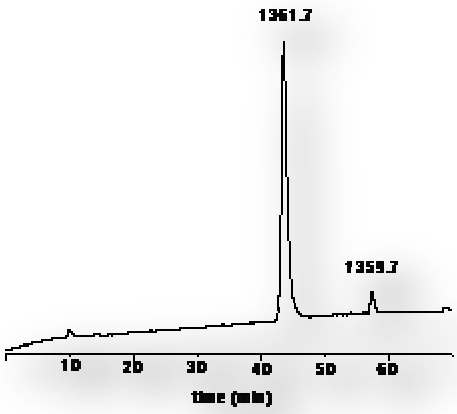
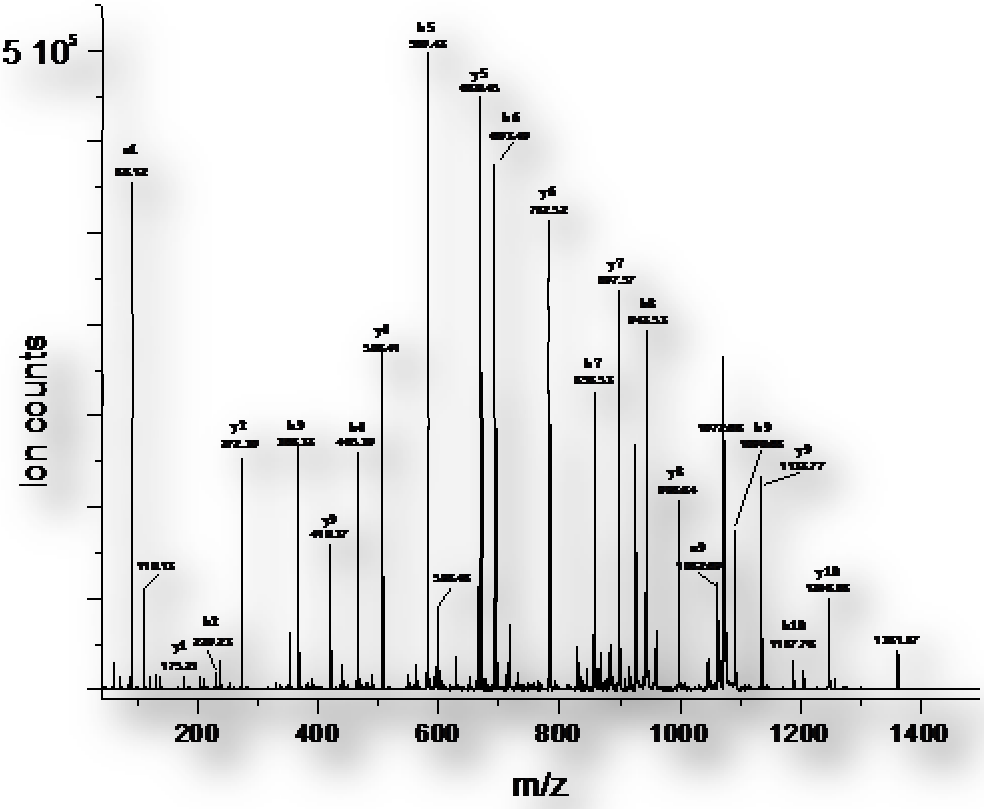


Figure 2.15. The side product of the reaction of peptide 3 with formaldehyde elutes after the expected tetramethylated compound in LC run

These data confirmed that the signal observed in MALDI and ESI owned to a real chemical species rose as side product after the derivatization, and not to an artifact produced during the ionization process.

Moreover fragmentation spectra, holding peptide sequencing information, helped the localization of the mass decrement at the amino acid level. In the MS/MS spectrum of the first eluted product (1361.7 m/z) the expected enhancement of the a_1 ion was observed, the y series was completely detected while in the b series the b_1 ion missed. The y sequence was exactly the same in the spectrum of the second eluted compound (1359.7 m/z) whereas all the b ions were shifted by 2 mass units and the a_1 ion was almost suppressed, providing evidences that the modification between the two products was localized at the N-terminus (Figure 2.16).



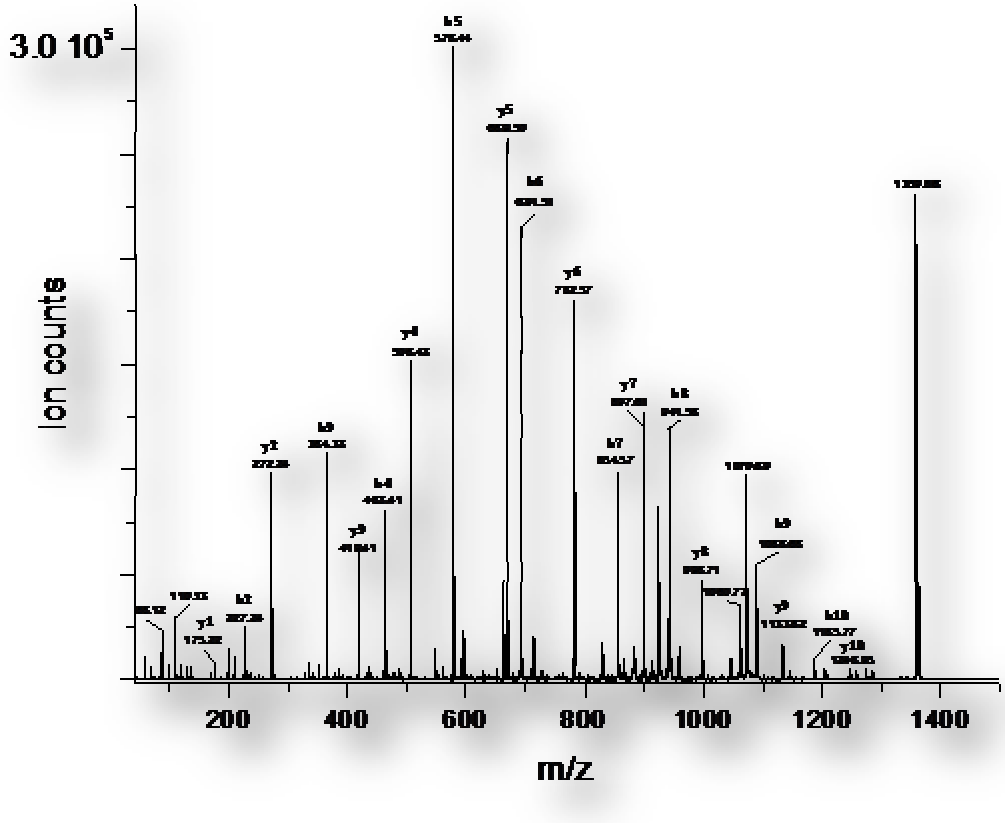


Figure 2.16. MALDI-CID spectrum of the first eluted compound (above) and MALDI-CID spectrum of the second eluted compound (below)

Further evidences supporting the imidazolidinone structure proposed by *Metz et al* came from 2MEGA (N-terminal dimethylation after lysine guanidination) and 2ETGA (N-terminal ethylation after lysine guanidination) experiments. Basically lysine residues were guanidinated with O-methylisourea so that the following dimethylation or diethylation could involve only the N-termini of the peptide.

The 2MEGA experiment at pH 8.8 showed a slightly visible +26 signal in presence of a higher abundant expected peptide at 1375.67 m/z; instead with 2ETGA reactions the +26 signal at 1401.7 m/z was more intense than the expected at 1403.7 m/z (Figure 2.17).

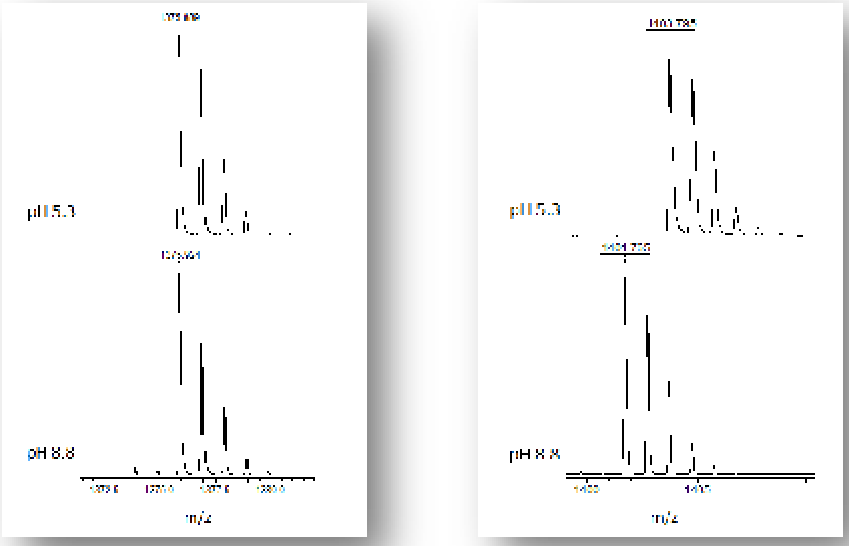


Figure 2.17. 2MEGA (on the left) and 2ETGA (on the right) reactions on peptide 3, at pH 5.3 (spectra above) and 8.8 (spectra below)

These further reactions highlighted the presence of the signal at minus 2 Da in a peptide bearing a blocked lysine residue, confirming that the N-terminal was the site responsible for the rearrangement, being the only reactive site of this peptide.

The peculiar modifications observed with ACTH were studied more in detail by using *ad hoc* synthesized decapeptides, designed to reproduce the presumed reactive region of the peptide. RPVKVYPNGA (**5**, EM: 1099.6) was synthesized to replicate the same sequence of ACTH from the N-terminus to the Ala10 and APVKVYPNGA (**6**, EM: 1014.6) to prevent the cross-link by replacing Arg1 with Ala. The corresponding acetylated peptides AcRPVKVYPNGA (**7**, EM: 1141.6) and AcAPVKVYPNGA (**8**, EM: 1056.6) featured a protection at the N-terminus with the aim at avoiding its involvement in the rearrangement. Finally AcAPVKVAPNGA (**9**, EM: 964.5) featured the replacement of Tyr6 with Ala, eliminating a further possible residue that could be involved in a cross-link with the dimethylated Lys4.
These peptides were derivatized with both formaldehyde and acetaldehyde to explore the characteristics of the rearrangement, and analyzed by MALDI-TOF MS.
Peptide **5** and **7** behaved exactly as ACTH in the reaction with formaldehyde, thus the expected +56 Da product ([M+H]+ 1156.6) was observed at pH 5.3 with no evidence of side-products; instead in the reaction at pH 8.8 a secondary product with a 52 Da mass increment was visible in the MALDI spectrum (Figure 2.18).

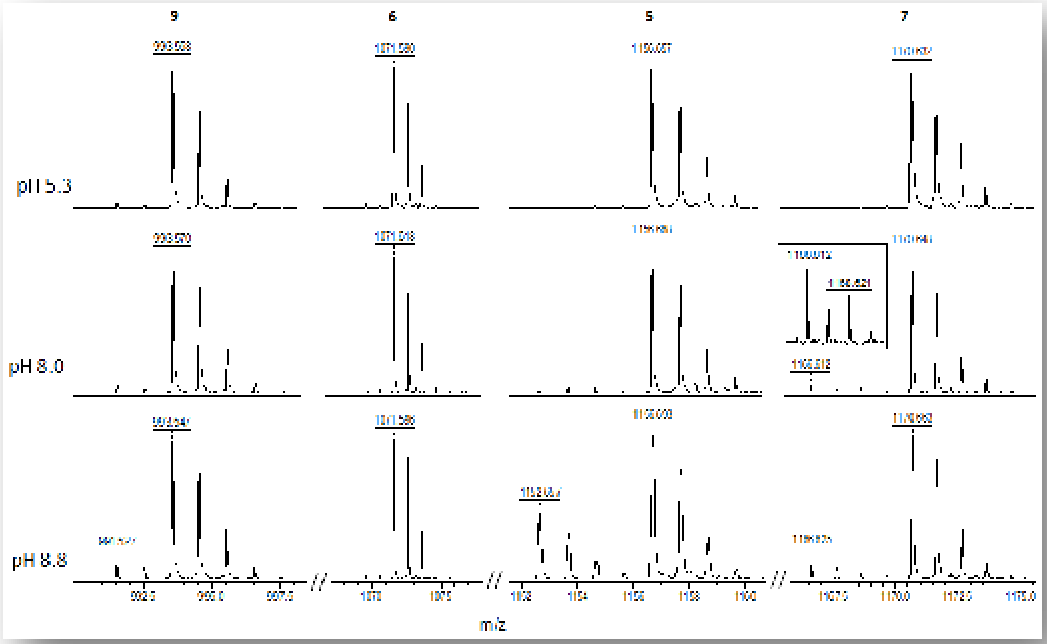


Figure 2.18. Formaldehyde-derivatized peptide 5, 6, 7 and 9

Peptides **6** and **9,** bearing substituted arginine and tyrosine, respectively, lacked that signal thus supporting the hypothesis that a cross-link between the side chain of arginine and lysine was the responsible for the +52 Da signal. This was confirmed by 2-MEGA reactions where, lysine was transformed in homoarginine and the expected increment of 70 Da was not accompanied by the minus 4 Da signal.

Interestingly the formation of the 52 Da signal was found to be almost completely suppressed when formaldehyde was added after $NaCNBH_3$, namely when the imine is formed in a preexistent reducing environment (Figure 2.19). Otherwise, the short time elapsed between the addition of formaldehyde and the subsequent introduction of the reducing agent is enough to let the cross-link to take place.

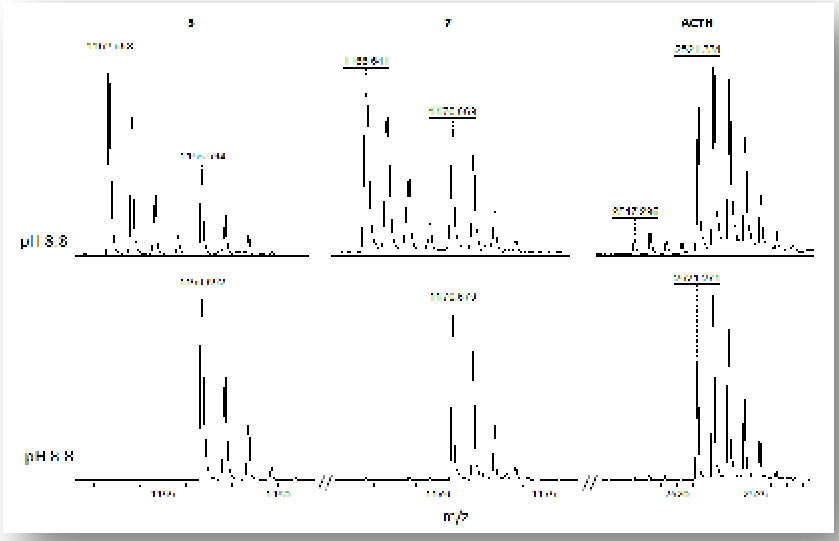


Figure 2.19. Peptide 5,7 and ACTH derivatized with formaldehyde and $NaCNBH_3$ (above); adding formaldehyde after $NaCNBH_3$ greatly reduces the amount of the rearrangement (below)

These results highlighted another aspect that was the presence of a small amount of a secondary product featuring 2 mass units less than the expected derivatized peptide, in all the MALDI spectra. Similar results derived also from peptide **8** where no side products were expected, as arginine was replaced by alanine and the N-terminus was protected by acetylation. Instead, once dimethylated, its MALDI spectrum showed a +26 signal but no evidence of a corresponding product was found in the ESI-MS spectrum (Figure 2.20) or in LC-MS runs.

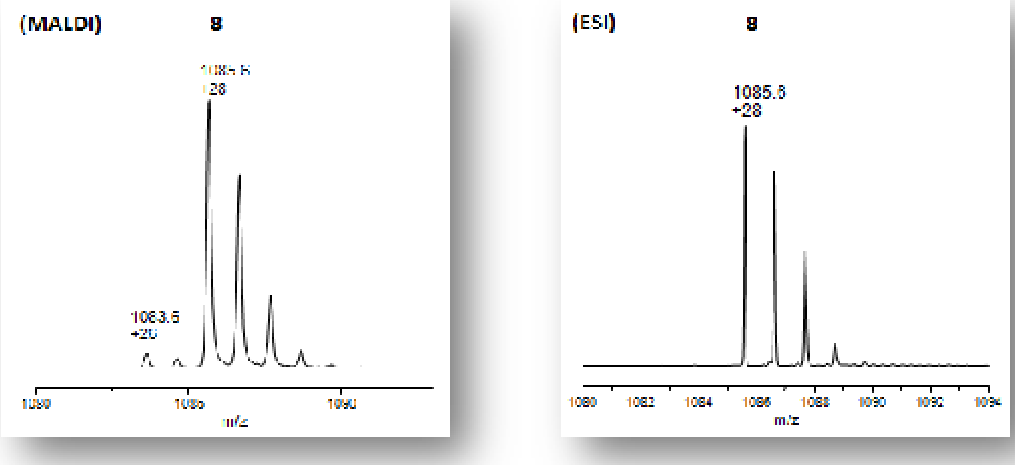


Figure 2.20. MALDI spectrum (on the right) and ESI spectrum (on the left) of peptide 8

These evidences supported the hypothesis that probably the observed signal did not correspond to the formation of a secondary product in the reaction, but rather to a

rearrangement of the tertiary amine in the respective immonium ion during the matrix assisted desorption event. This phenomenon seems to be related to a rearrangement that takes place in the MALDI source.[106] Laser desorption is known to be "harder" than electrospray ionization, so what we observe in this case is probably the product of a rearrangement of the tertiary amine that leads to the loss of two hydrogen atoms generating of an immonium ion that gives a $[M-H]^+$ signal (Figure 2.21).

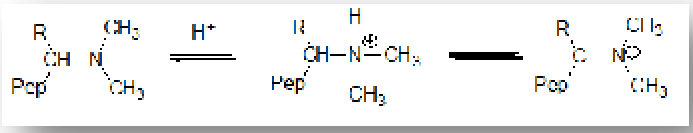


Figure 2.21. Possible mechanism for the presence of $[M-H]^+$ ion formation for tertiary amines

It is difficult to attribute the minus 2 Da signal to the imidazolidinone or the immonium ion, that are real chemical species or a MALDI artifact, respectively. However these hypotheses are supported by the observation that formation of these signals is forbidden at low pH and in ESI source, respectively.

In the reaction with acetaldehyde on peptides **5**-**9** the formation of a cross-link product was never observed, even if the reducing agent was added three hours after the introduction of acetaldehyde at pH 8.5. Steric hindrance is probably the main reason for this difference with dimethylation, together with an evident overall reduced reactivity.

Taking into account the results obtained from dimethylation and diethylation of the first employed peptides, TDP and ACTH, of the tryptic peptides **1**-**4** and the ACTH-derived synthetic peptides **5**-**9**, some considerations raised.

The behavior of ACTH or its peptide analogues **5**-**9** toward formaldehyde is undoubtedly peculiar and interesting but, fortunately, such rearrangements should be very rare when a tryptic digest is subjected to reductive amination. An intramolecular cross-link between arginine and lysine can take place only if a missed cleavage occurs and both residues are present in the same peptide at a proper distance from each other.

On the other hand, the imidazolidinone rearrangement at the N-terminus showed a higher frequency of occurrence, especially with acetaldehyde above pH 7.0, and could negatively affect quantitation experiments increasing the complexity of the sample.

In order to evaluate the real impact of the imidazolidinone formation in dimethyl and diethyl labeling procedures for quantitative experiments, more complex samples were prepared and a proper quantitation software was employed to extrapolate light/heavy ratios.

2.3.3 Evaluation of stable-isotope labeling procedures in proteomic model samples

Mixtures of heavy and light peptides labeled with both formaldehyde and acetaldehyde were analyzed by MALDI-TOF MS to check the progress of the reactions and to monitor the impact of the imidazolidinone on the quantitation. Methylated peptides confirmed to suffer a moderate presence of this side product while ethylated peptides were substantially covered by the imidazolidinone at the basic pH.
As example are reported the signals of peptides **2** and **4**, bearing one and two reactive sites, respectively. At pH 8.8 the signal of the heavy-labeled peptide **2**, because of the presence of the imidazolidinone and its isotopic pattern, partially overlapped with the signal of the corresponding light labelled, covering the spacing of 8 Da between them (Figure 2.22).

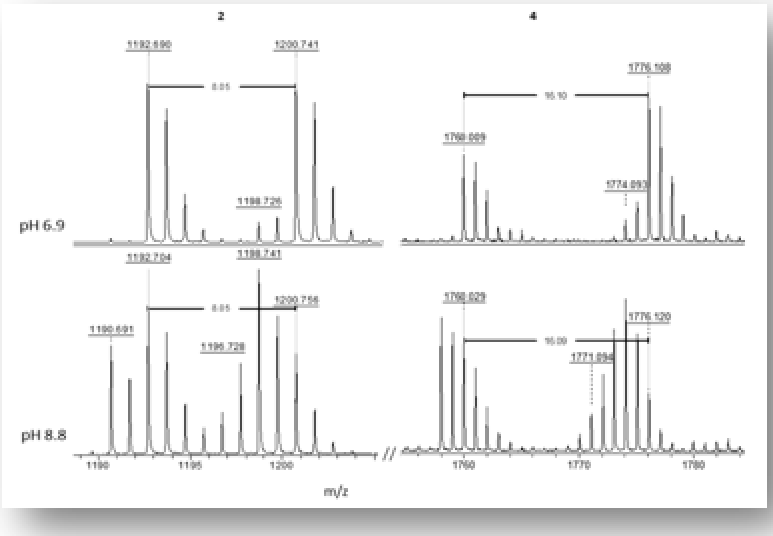


Figure 2.22. Peptide 2 and 4 ethylated at pH 6.9 (above) and at pH 8.8 (below)

For peptide **4**, instead, the spacing of 16 Da was sufficient to separate the two groups of signals. The comparison with the corresponding reactions performed at neutral pH, where a low amount of imidazolidinone was present, clearly highlighted how the presence of this rearrangement enhances the complexity of the spectra hampering a correct quantitation to be done.
In addition, it should be mentioned that reactions with both isotopomeric acetaldehydes were confirmed to be significantly slower than reactions with formaldehyde, requiring subsequent additions of the reagents to be done in order to achieve a complete ethylation. This aspect is crucial because a labeling reaction must provide a complete conversion of the peptides to be considered reliable for quantitative purposes.

For a more comprehensive evaluation of these aspects a differential labeling was performed on a BSA digest with formaldehyde and acetaldehyde at both neutral and basic pH. Mixtures of differentially methylated or ethylated digest were prepared with a theoretic light/heavy ratio of 0.7, 2.5 and 0.5, analyzed through LC-MS/MS runs and the data processed using Mascot Distiller software.

According to the labeling agent the amino acid modification included in the databank search were the dimethylation or the diethylation of lysine and the N-term and the quantitation methods used for the processing were Dimethylation [MD] and Diethylation [MD], both present in the Mascot Server.

As an example we report the results from the methylated and ethylated samples at pH 6 with protein light/heavy ratio of 0.7; results obtained from other methylated samples at the two pH values at different protein ratio are also discussed. In the quantitation report each experimental light/heavy ratio is reported as the geometric mean of the peptide ratios with the corresponding geometrical standard deviation. Moreover, for each peptide, in addition to the protein ratio the three mentioned quality parameters are included so that peptides bearing improper values are automatically excluded.

The reaction of methylation performed on the BSA digest proved to be reliable since the experimental ratio turned to be very close to the theoretic, 0.76 versus 0.70; there was also sufficient homogeneity among the 15 peptides used for the quantification (Figure 2.23). Outlier peptides, exceeding the standard error threshold or the fraction threshold were excluded from the analysis.

Hit	Accession	Score	Mass	L/H	SD(geo)	#
1	ALBU_BOVIN	944	73168	**0.7622**	**1.104**	**15**
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4						

	z	Sequence	Incl.	L/H	Std.Err.	Fraction	Correlation	Intensity
1	2	ATEEQLK	X	0.6840	0.01303	0.6331	0.9975	39.40
2	2	CCTESLVNR	X	0.7508	0.03507	0.8941	0.9947	263.6
3	2	KQTALVELLK		0.4122	0.1432	0.1690	0.9967	71.98
4	3	DTHKSEIAHR		0.7962	0.1003	0.7276	0.9954	17.66
5	3	FKDLGEEHFK	X	0.9060	0.006235	0.7566	0.9926	348.5
6	2	FKDLGEEHFK	X	0.7704	0.03439	0.4559	0.9987	45.75
7	3	SLHTLFGDELCK		0.8792	0.05113	0.7526	0.9922	590.1
8	2	SLHTLFGDELCK	X	0.7070	0.02245	0.7933	0.9976	168.4
9	3	TCVADESHAGCEK	X	0.8568	0.02205	0.8076	0.9967	553.0
10	2	TCVADESHAGCEK		0.7183	0.05669	0.6273	0.9985	31.82
11	2	ETYGDMADCCEK	X	0.7923	0.02156	0.8351	0.9990	247.1
12	2	DAFLGSFLYEYSR	X	0.8775	0.01130	0.9695	0.9993	1004
13	3	DAFLGSFLYEYSR	X	0.5320	0.05113	0.6061	0.9559	55.29
14	2	MPCTEDYLSLILNR	X	0.7769	0.02065	0.9590	0.9976	290.6
15	3	MPCTEDYLSLILNR	X	0.6583	0.01536	0.6970	0.9940	85.77
16	3	ECCHGDLLECADDR	X	0.6599	0.02145	0.9263	0.9994	206.3
17	3	LFTFHADICTLPDTEK	X	0.7028	0.05946	0.7779	0.9987	887.8
18	3	CCAADDKEACFAVEGPK	X	0.8144	0.06762	0.5670	0.9941	285.9
19	3	RHPYFYAPELLYYANK	X	0.7570	0.05478	0.7391	0.9903	223.2
20	3	GLVLIAFSQYLQQCPFDEHVK	X	0.7711	0.01321	0.8755	0.9949	146.1

Figure 2.23. Quantitation report of methylated samples at pH 6 (above) where the theoretic light/heavy ratio was 0.7; table of identified sequences (below)

The XICs of the two components of one BSA peptide, namely SLHTLFGDELCK, are reported as example showing the co-elution of the two differentially labeled peptides (Figure 2.24).

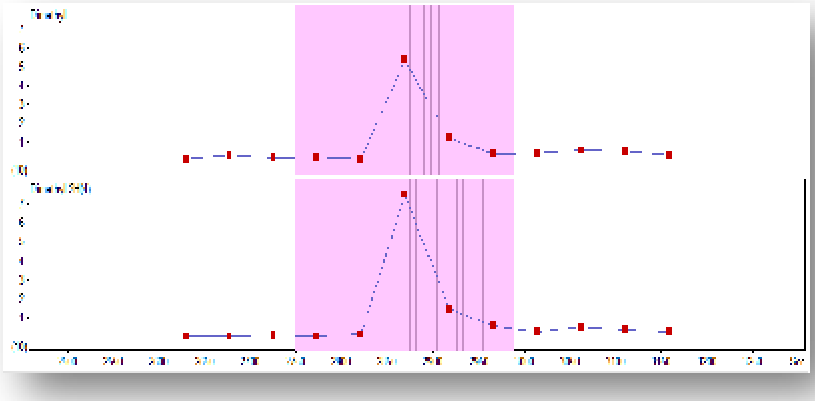


Figure 2.24. XICs of the two components of peptide SLHTLFGDELCK

The scan markers on the XICs are for survey scans but only the scan included in the shady window are integrated to determine the ratios while the darker bands are the scans for which have been found Mascot matches. Database matches for both the components in a ratio are not required for each pair of peptides because as long as there is a match for one component, the "partner" is inferred from its mass and co-elution.

The spectrum of the two components allowed to make a visual judgment about their relative intensities showing that the intensity of the double-charged light peptide at mass 738.373 Da was about 2/3 lower than the heavy peptide at 742.294 Da (Figure 2.25).

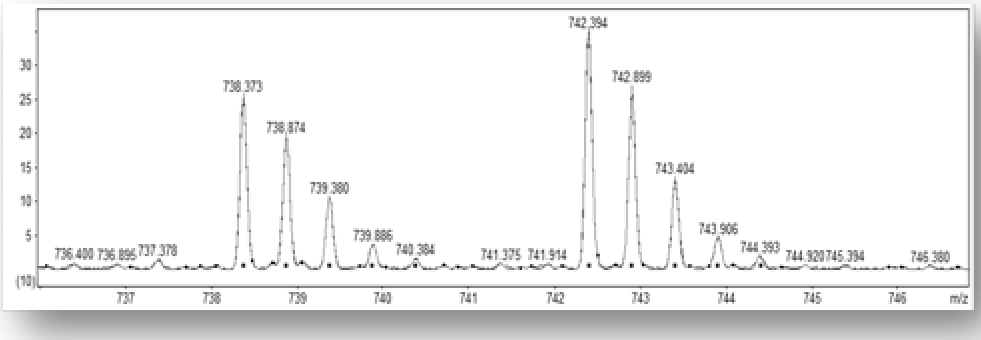


Figure 2.25. Observed mass spectrum of the two components of the peptide SLHTLFGDELCK

When data concerning ethylated BSA digest were processed, the protein was recognized based on 22 labeled peptides that were correctly identified by means of

their fragmentation spectra. Unfortunately the calculated protein ratio was higher than the expected with a reported value of 3.3908 (against the theoretic value of 0.7) based on highly variable singular peptide ratios.

As it was previously noted with the mixture of standard peptides the ethylation reaction was much more troublesome than the corresponding methylation one. In order to achieve a satisfactory labeling yield several additions of acetaldehydes and $NaCNBH_3$ were required and a constant monitoring of the progress of the reaction had been performed. Considering that, it is not unexpected that the ethylation of the BSA digest had faced similar problems hampering the reactions to reach the completeness and thus provoking a wrong protein light/heavy ratio. Similar results were obtained in repeated experiments of ethylation of complex samples indicating that an overall optimization of the labeling, including a constant monitoring of the progression of the reaction, was essential to achieve an efficient yield of conversion.

At variance, the positive results obtained in the previous reactions with formaldehyde supported its use as reliable labeling agent. In order to provide further evidences the BSA digest differently derivatized at pH 6 was prepared in a different light/heavy ratio, namely 2.5, and analyzed by means of LC-MS/MS runs. Even in this case, Mascot Distiller processing correctly calculated a protein ratio of 2.421 with a deviation of 1.184 and once again all the quantified peptides exhibited homogeneous ratios.

Interesting results were obtained with the differential methylation of the protein digest performed at pH 8.0 which was a critical value for the formation of the imidazolidinone derivative, as observed with standard peptides. Data processed considering the light and heavy methylation returned a ratio of 0.5267, very close to the theoretic which was 0.5. When the same data were analyzed searching for peptide rearranged to form the imidazolidinone derivative no corresponding matches were found, indicating that the impact of such side product in the methylation approach was limited. It would have been very interesting to evaluate the impact of the imidazolidinone also on the differentially ethylated digest since the results obtained with the peptides had proved a significant incidence of that rearrangement at basic pH. Because of previous considerations, it was not possible to evaluate this aspect and quantify the impact on proteomic samples.

Instead, given the results obtained with formaldehyde on the single protein digest other experiments were conducted on more complex samples in order to further validate the strategy and to insist monitoring its behavior at both the "safe" and "hazardous" pH.

A sample containing five proteins was prepared including BSA, myoglobin, alcohol dehydrogenase, cytochrome C and carbonic anhydrase 2. The proteins were mixed in different amounts so to include a wide range of concentrations. Moreover different light/heavy ratios were thought for each protein so as to have available a sample that could shape as much as possible a proteomic sample. The two mixture of proteins

were separately subjected to light and heavy methylation at pH of 6.0 and 8.0, and then mixed to be submitted to LC-MS/MS analysis.

Quantitation processing of reactions led at pH 6.0 allowed proteins to be measured with ratios that resulted very close to the theoretic ones (Figure 2.26), confirming the results obtained with the single protein experiment.

Protein	Theoretic L/H
Cytochrome C	1
Alcohol Dehydrogenase 1	2.46
Serum Albumin	0.5
Myoglobin	1
Carbonic Anhydrase	0.25

Quantitation Table

	Accession	Score	Mass	L/H	SD(geo)	#	Description
1	CYC_HORSE	358	12472	0.9738	1.1005	11	Cytochrome c OS=Equus caballus GN=CYCS PE=1 SV=2
2	ADH1_YEAST	337	37929	**2.4951**	**1.1607**	**14**	Alcohol dehydrogenase 1 OS=Saccharomyces cerevisiae GN=ADH1 PE=1 SV=4
3	ALBU_BOVIN	336	72927	0.6563		15	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
4	MYG_EQUBU	324	17605	1.0295		11	Myoglobin OS=Equus burchelli GN=MB PE=1 SV=2
5	CAH2_BOVIN	252	29673	**0.2900**	**1.4328**	**11**	Carbonic anhydrase 2 OS=Bos taurus GN=CA2 PE=1 SV=3

Figure 2.26. Theoretic light/heavy ratio (above) and quantitation report (below) of the methylated mixture of five proteins, at pH of 6.0

It was also really encouraging to find out that, once again, the peptides pairs belonging to the same protein exhibited a light/heavy ratio in a wide agreement with the final protein ratio. Moreover single proteins were identified by means of several labelled peptides, indicating an high yield of derivatization also in a more complex sample.

Interestingly, similar results came from the processing of the methylation reaction performed at pH 8.0. Experimental ratios followed the theoretical ones (Figure 2.27) and the research for imidazolidinone derivatives did not retrieve any significant match.

Quantitation Table

	Accession	Score	Mass	L/H	SD(geo)	#	Description
1	CYC_HORSE	425	12472	1.0334	1.1049	12	Cytochrome c OS=Equus caballus GN=CYCS PE=1 SV=2
2	CAH2_BOVIN	360	29673	**0.3156**	**1.3647**	**11**	Carbonic anhydrase 2 OS=Bos taurus GN=CA2 PE=1 SV=3
4	ADH1_YEAST	322	37929	**2.4846**	**1.1671**	**10**	Alcohol dehydrogenase 1 OS=Saccharomyces cerevisiae GN=ADH1 PE=1 SV=4
5	MYG_EQUBU	263	17605	**1.1947**	**1.1921**	**11**	Myoglobin OS=Equus burchelli GN=MB PE=1 SV=2
6	ALBU_BOVIN	210	73168	**0.5022**	**1.1075**	**9**	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Figure 2.27. Quantitation report of the methylated mixture of five proteins, at pH of 8.0

The results obtained with the methylation of increasing complex samples starting from mixture of few peptides to mixture of tryptic peptides from different proteins confirmed the reliability of the labeling using formaldehyde. Moreover the impact of the imidazolidinone product did not appear to be significant, as already observed with preliminary experiments, since research algorithm did not find it out.

Instead with the ethylation labeling the incidence of the imidazolidinone had been proved to be significant even at the neutral pH, showing a potential risk of misleading quantitative interpretations. However a sound evaluation of the impact in more complex samples, like it has been done with methylation, was not achieved for the occurrence of other problems. Indeed acetaldehyde labeling strategy was not successful just starting from simple peptides mixtures revealing that the same experimental conditions used for methylation were not able to reproduce similar satisfactory results with this reaction, mainly in terms of derivatization yield.

2.4 CONCLUSIONS

Quantitative proteomics is a powerful approach for understanding global protein dynamics in a cell, tissue or organism. To achieve this goal an accurate quantitation of protein expression is necessary and the development of suitable methods is one of the most challenging and rapidly evolving areas of proteomics. Among the mass spectrometry-based technologies the use of chemical tags allowing comparative quantitation of proteins in different samples has been achieved by means of different stable isotope labeling reactions.

In this work of thesis a reductive amination reaction has been employed using aldehydes that, in presence of $NaCNBH_3$, targets all the free amino groups of peptides, namely N-termini and lysine residues, leading to their alkylation. A differential isotope labeling can be achieved by employing d(0)-formaldehyde and d(2)-formaldehyde or d(0)-acetaldehyde and d(4)-acetaldehyde that produce isotopomeric peptides featuring a mass difference of 4 or 8 Da for each dimethylated or diethylated derivatization site, respectively.

The chemistry of the reaction between formaldehyde and an amino group is well understood despite the lack of a comprehensive scenario regarding the chemistry of aldehyde modifications within proteins. However the description of cross-linking with some amino acids in formaldehyde-treated protein suggests to watch out for collateral unwanted products that may rise from the derivatization of complex proteomic samples. In fact the absence of side products is one of the basic requirements for a reaction to become a reliable method of stable isotopic labeling of peptides, in order to avoid significant error in the quantitative analysis through the mass spectra.

Given these premises a targeted evaluation on the reductive amination with both formaldehyde and acetaldehyde has been initially conducted using standard peptides, in order to monitor the progress of the reactions.

Our experiments have shown that reductive amination on tryptic peptides leads to the desired dimethylation, or diethylations, at N-termini and at lysine residues only when pH ensures a rapid action of the hydride and reduces the nucleophilicity of other amino groups.

On the contrary, when using basic pH, some rearrangements can take place involving peptide N-terminal in an imidazolidinone derivative or involving the arginine side chain in a methylene bridge with dimethylated lysine. Several experiments were led to prove the nature of such adducts including ESI source mediated acquisitions, guanidination reactions, LC separations, MS/MS spectra interpretation and *ad hoc* designed synthetic peptides.

We observed that the incidence of the rearrangement was widely different among the tested peptides suggesting that the chemical environment of the reactive site greatly influences its formation. The adduct between the dimethylated lysine and the arginine by means of two methylene bonds can be also suppressed if a reducing environment is created, for instance by the addition of the hydride before the introduction of the aldehyde.
Furthermore it was possible to highlight the presence of an immonium ion rising after MALDI desorption featuring the same minus 2 Da signal as imidazolidinone but this apparent side product could be a problem in a LC-MALDI experiment only.
The impact of the imidazolidinone in quantitative experiments was monitored using mixture of peptides and more complex samples of tryptic digestion of a single or various proteins. While the reactions on peptides were MALDI analyzed, these last were analyzed by LC-MS/MS and results processed using the Mascot Distiller package. In agreement with the situation observed with standard peptides, it was found that for methylation the occurrence of the rearrangement was mostly negligible. Moreover the quantitation results were in accordance with the theoretic, confirming its reliability as isotope labeling reaction.
The use of acetaldehyde, instead, revealed more problematic aspects; an higher incidence of the imidazolidinone was observed starting from neutral pH and became the predominant signal at the most basic pH. Moreover experiments conducted on standard peptide mixtures proved a slow kinetic of the reaction requiring subsequent additions of the reagents to be performed to reach completeness. In the case of more complex samples, such as a digested protein, these further additions were not probably sufficient to overcome this problem. Indeed a strong disagreement was retrieved between the calculated and the theoretic light/heavy ratio.
Further experiments will be necessary to monitor the imidazolidinone adduct in biological complex proteomic samples and a revision of the experimental conditions employed in the ethylation strategy will be also fundamental to evaluate its applicability in peptide labeling.

In conclusion this project has partially highlighted how diverse can be the scenarios when we undertake a derivatization task on a complex ensemble of substrates.
By a strictly proteomic point of view we can definitely fix, as a general rule, that reductive amination for differential isotope labeling must be performed at pH 5-6 to be absolutely sure that no undesired side reactions can take place. In these conditions reductive amination actually fits with the requirements of a suitable technique for isotope coding in relative quantitation of proteins, including a considerable reduction of costs. The application of higher pH values, as already reported in the literature, should not bring to any significant drawback with formaldehyde, at least as we work below pH

8.0. Conversely, when acetaldehyde is used, the incidence of collateral reactions recommends not to work above pH 6.0 in order to limit the formation of side products.
By a chemical point of view we observed side products in reductive amination either with formaldehyde and acetaldehyde. Apart from these evidences based on our experiments, we cannot describe a comprehensive picture, or make an assessment of those residues that are the most likely promoter of a rearrangement at the N-terminus, or individuate the lysine residues more prone to be involved in cross-links. Thus a further more extensive investigation could be interesting to widen the knowledge on aldehyde-induced modification on peptides and proteins.

2.5 EXPERIMENTAL PART

Instruments and methods

MALDI experiments were performed on an UltraFlex III MALDI-TOF/TOF mass spectrometer (Bruker-Daltonics, Bremen, Germany). Data were acquired in positive reflectron mode. Acquisitions were performed in the mass range 200-3000 Thomson (m/z) with voltages of 25 and 21.7 kV for the first and second ion extraction stages, 9 kV for the lens, 26.3 and 13.8 kV for reflector 1 and 2 respectively. TOF calibration was performed using the Peptide Calibration Standard, basing on the monoisotopic mass of bradikinin (clip 1-7) $[M+H]^+$=757.3992, angiotensin II $[M+H]^+$=1046.542, angiotensin I $[M+H]^+$=1296.685, Substance P $[M+H]^+$=1347.7354, bombesin $[M+H]^+$=1619.823, ACTH (clip 1-17) $[M+H]^+$=2093.087, ACTH (clip 18-39) $[M+H]^+$=2465.199, somatostatin $[M+H]^+$=3147.471. MS and MS/MS data were processed and analyzed by the Bruker FlexAnalysis 3.0 software.

Chromatographic separation of peptide **3** from the corresponding imidazolidinone was performed on a Dionex Ultimate capillary-HPLC system equipped with Famos autosampler and Switchos microcolumn switching device (LC Packings, Amsterdam, The Netherlands).
5 µL of the reaction mixture were loaded onto a RP in-line precolumn (Dionex/LC Packings) and washed with 0.1% formic acid at a flow rate of 0.030 ml/min, for 3min. Separation was performed on a C18 RP capillary column (Dionex/LC Packings) whose temperature was set at 30°C. A linear gradient elution was applied at a flow rate of 3 µL/min (eluent A, 0.1% formic acid in 95% water and 5% acetonitrile; eluent B: 0.1% formic acid in 95% acetonitrile and 5% water). from 5 to 40% of B within 25 min.
The Ultimate system was fully controlled by the Dionex™ Chromeleon® software.

LC-MS/MS experiment for quantitation were performed on a Waters *nano*Aquity UPLC system interfaced with a Waters Q-TOF Premier mass spectrometer equipped with a *nano*ESI source.
Peptides, 2 µL full loop injection, were trapped on a 5 µm Symmetry C18 column (180 µm × 620 mm) and washed for 3 min at 0.3 µL/min with 0.1% TFA. Peptides were then eluted and separated using a 60 min RP gradient at 250 nL/min (35% ACN over 45 min) on a 1.7 mm BEH 130 C18 NanoEase™ (75 µm × 625 mm) *nano*scale LC column. Eluent A was 0.1% formic acid and eluent B 0.1% formic acid in ACN.

The column temperature was set at 50 °C. Lock mass ([Glu1]-fibrinopeptide B, 250 fmol/μL) was constantly infused by the NanoAcquity auxiliary pump at a constant flow rate of 250 nL/min.
The Q-Tof Premier™ mass spectrometer was operated in positive mode.
The Q-TOF was set to perform automated data-dependent acquisition switching from MS to MS/MS mode and returning to the MS mode using data dependent criteria. Multiply charged precursors were selected and fragmented automatically with the collision energy in MS/MS varied according to the m/z and the charge state of the precursor ion. MS survey scans were acquired within a mass range of 300–1600 m/z, with a 0.8 s duration and an interscan delay of 0.1 s. A dynamic exclusion window was set to 30 s.
All the LC-MS/MS system was under the control of the Water software MassLynx®.

Reductive amination

As a general procedure for standard peptides methylation, 50 μL of reaction buffer (0,1 M) were added to 10 μL of peptide solution (1 mg/mL) in a 0.2 mL eppendorf tube. 2.5 μL of $NaCNBH_3$ (1 M) were added and the solution vortexed for 30 s, then 2.5 μL of formaldehyde solution (0.5 M) were added and the mixture vortexed again. After 1 h incubation at 25° C, the reaction was quenched with 2 μL of 10% ammonia solution to consume the excess of formaldehyde.
The ethylation labeling procedure followed the same criteria except for the incubation time that was prolonged until 6 h at 30°C; after the first 4 h a supplemental addiction of both acetaldehydes and $NaCNBH_3$ were performed and allow to react for other 2 h.

Guanidination

10 μL o-methyl isourea solution (0.5 M) were added to 10 μL of a solution of peptide (1 mg/mL) in a 0.2 mL eppendorf tube at 60°C and pH was adjusted to 11 with 1 μL of 2 M NaOH. The reaction was kept at 60° C for 1 h and quenched with 1 μL of TFA 10%.

Sample preparation

Before MALDI analysis all the samples have been desalted by ZipTipμ-C18 treatment and spotted with a solution of α-cyano-4-hydroxycinnamic acid (5 mg/mL in 50% acetonitrile, 0,1% TFA).

Peptides synthesis

Peptides from **5** to **9** were synthesized on an automatic batch synthesizer (APEX 396, Advanced ChemTech) equipped with a 40-well reaction block, following the fluorenylmethoxycarbonyl (Fmoc)/tert-butyl (tBu) solid-phase peptide synthesis (SPPS)

strategy. Fmoc amino acids and resins were purchased from Calbiochem–Novabiochem.

Fmoc deprotections were performed in 30 min with 20% piperidine in dimethylformamide (DMF). Coupling reactions (repeated twice) were performed for 45 min by using the Fmoc-protected amino acids dissolved in N,N Dimethylformamide (DMF), in presence of a solution of 2M N,N-diisopropylethylamine (DIPEA) in N-Methylpyrrolidone (NMP) and a 0.5 M solution of 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) in DMF.

Acetylation of peptides from **7** to **9** was performed with a solution of acetic anhydride (20 eq) in N-Methylmorpholine (NMM) and dichloromethane (DCM).

Peptide cleavage from the resin and deprotection of the amino acid side chains were carried out in 3 h with TFA/TIS/H_2O (95:2.5:2:5) (vol:vol:vol:). The resin was filtered off, and the solution was concentrated. The crude products were precipitated with cold Et_2O, centrifuged, and lyophilized.

Products were analyzed by analytical RP-HPLC (Alliance, model 2695 equipped with a diode array detector, Waters) using a Jupiter C18 (5 mm, 250x4.6 mm) column (Phenomenex) at 1 ml/min and interfaced to an ESI-MS source (Micromass ZQ).

The solvent systems used were A (0.1% TFA in H_2O) and B (0.1% TFA in acetonitrile).

Products were purified by preparative RP-HPLC (model 600, Waters) on a Jupiter C18 column (10 mm, 25 cmx10 mm) at 4 ml/min by using the same solvent systems reported above. Characterization of the products was performed with the LCQ Advantage liquid chromatography electrospray ionization mass spectrometer (ThermoFinnigan).

In solution digestion

15 µL of a solution of the protein mixture in TEAB 0.1 M was prepared in order to contain 50 µg of proteins. 1.3 µL of 0.1 M DTT, 1.5 µL of 0.2 M IAA and again 0.25 µL of 0.1 M DTT were subsequently added, leaving the mixture for 1h at 37° C, 1h at r.t. in dark, and 20m at 37° C respectively. The sample was digested over night at 37° C by adding 2 µL of trypsin (0.5 µg/µL) and was stopped with 2 µL of 1% TFA.

Data bank search and quantitation

LC-MS/MS data were processed using the package Mascot Distiller whose features has been previously described. Raw data were processed using the default processing options described. Databank searches were performed within the SwissProt database (released version: Sept2012; 537505 sequences, 190795142 residues) without any restriction except for the BSA samples search that was limited to "other mammalia" taxonomy. Precursor tolerance of 20 ppm and fragment tolerance of 0.1 Da were set. Cysteine carbamidomethylation was set as fixed modification and methionine oxidation

as variable; according to the sample the specific modifications induced by the labeling strategy were considered as variable. Dimethyl (K), dimethyl (N-term), dimethyl:2H(4) (K), dimethyl:2H(4) (N-term), diethyl (K), diethyl (N-term), diethyl:2H(8) (K); diethyl:2H(8) (N-term) where selected among the Mascot Server panel. Four more modifications regarding the imidazolidinone derivatives where edited using the configuration editor of the Mascot Server: imidazol-methyl L, imidazol-methyl H, imidazol-ethyl L and imidazol-ethyl H.

Relative quantitative analysis has been made by measuring the ratio between the signals of the light and heavy labeled peptides (L/H) in the mass spectra, using the Quantitation Tool of Mascot Distiller. Dimethylation [MD] and diethylation [MD] were pre-existing quantitation methods while one further method was edited to perform the quantitation of the imidazolidinone in methylation reactions.

CHAPTER 3

Alpha actinin is specifically recognized by multiple sclerosis autoantibodies isolated using an N-glucosylated peptide epitope

3.1 INTRODUCTION

3.1.1 Multiple Sclerosis

Multiple Sclerosis (MuS) is an autoimmune neurodegenerative disease characterized by a progressive loss of the myelin sheath of nerves and a consequent gradual reduction of motor, cognitive and sensorial skills. Neurons are able to communicate by sending electrical signals (action potentials) that travel along the myelin-isolated nervous fibers. Whenever these are subjected to anatomic damages severe backlashes are recorded in the organism.

The degenerative process can potentially target any cerebral region and the neurological symptoms can vary among patients as well as the progression of the disease. Common features shared with other autoimmune diseases are the alternating remitting and relapsing phases which show unpredictable onset, duration and disappearance. In the late phase of the disease patients enter a sort of irreversible stadium characterized by a progressive worsening of the systemic conditions.

An evident clinical symptom of MuS concerns the occurrence of neuroglial hypertrophic plaques and focuses of demyelination among the central nervous system (CNS), which account for the name of the disease. Demyelination has been proved to be largely mediated by immunocompetent factors thus suggesting a strong autoimmune component.[107]

MuS shows a multifactorial pathogenesis characterized by a significant activation of the immune system and a strong inclination to develop in genetically predisposed subjects exposed to environmental factors such as bacterial/viral infections or toxic agents.

Several hypothesis have been proposed to explain the trigger of the autoimmune response; both autoreactive T cells and demyelinating antibodies risen outside the CNS are thought to enter the hematoencephalic barrier. Adhesion molecules over-expressed by endothelial cells would allow T lymphocytes' entry thus promoting the activation of other immunocompetent cells (i.e. antigen presenting cells, APCs) after the release of pro-inflammatory cytokines. Activation of this choral immune response might result in the development of a local inflammatory reaction with a progressive myelinic and oligodendrocyte impairment (Figure 3.1).

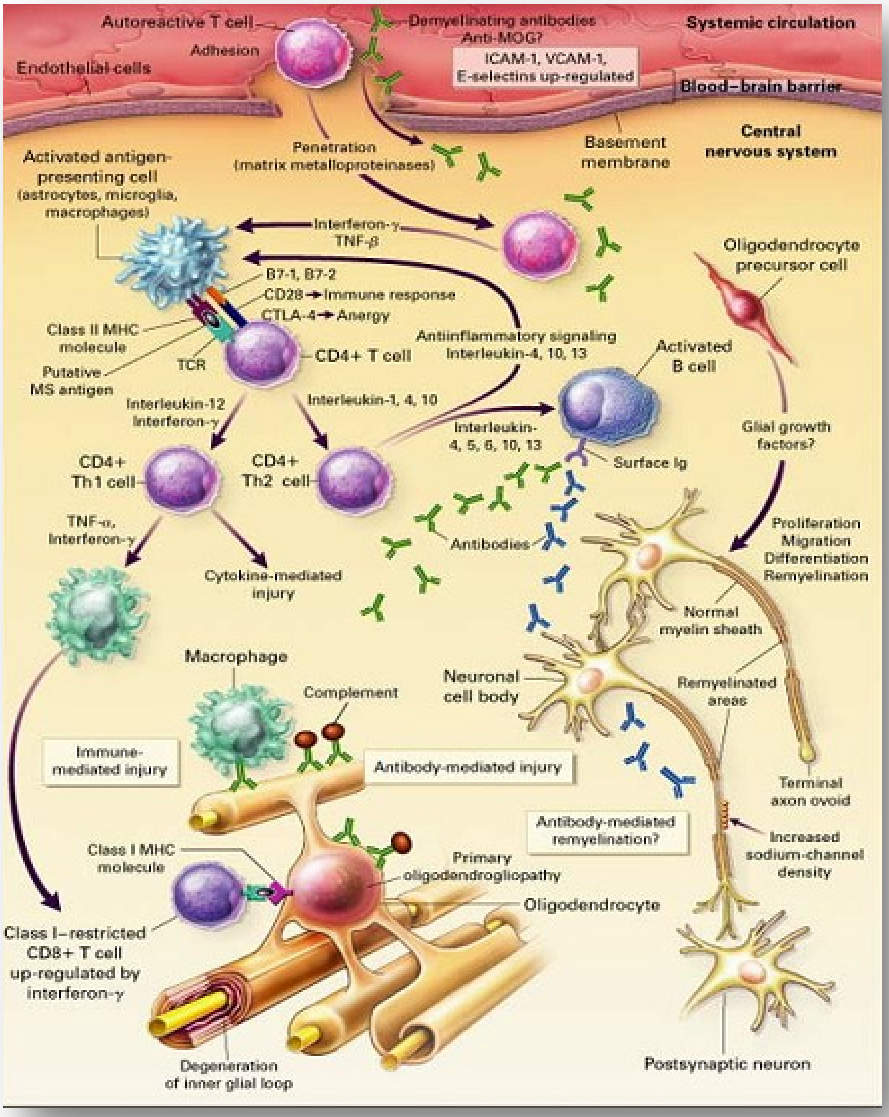


Figure 3.1 Hypothetical mechanism of neuronal damage in MuS

Four different patterns of MuS pathology have been proposed:[108]

- Cell-mediated demyelinization;
- Antibody-mediated demyelination;
- Active myelin destruction;
- Oligodendrogliopathy or oligodendrocyte dystrophy.

Rather than a single pathology MuS is, thus, considered as a set of diseases exhibiting different onsets but sharing a strong inflammatory reaction and a progressive disability in afflicted patients. Four internationally recognized clinical forms of MuS have been described: relapsing remitting (RR), secondary progressive (SP), primary progressive (PP) and progressive relapsing (RP).

Like other autoimmune diseases, as already described in the case of rheumatoid arthritis, the possibility to perform an early diagnosis is of mandatory importance to develop appropriate therapies. Moreover the disease has a large social impact, with spiraling costs increasing with the progression of disability.[109] If progression can be delayed, the quality of life and independence of MuS patients will improve and the cost to health care system and society will decrease accordingly.[110]

Up to now, the drugs used to treat MuS are able to manage only the symptoms by reducing the severity and the frequency of exacerbation. Corticosteroid, interferons (Betaseron and Avonex) and Copaxone are the therapeutic approaches most frequently employed in the management of the disease.

On the contrary, targeted treatments are not still available since the pathogenetic mechanisms involved in the establishment of MuS have not been unraveled, as well as the CNS myelin antigens responsible of the triggering of the autoimmune response. Most of the putative antigens belong to the myelin family (myelin basic protein, proteolipid lipoprotein, and myelin oligodendrocyte glycoprotein).

However disappointing results have been obtained in extensive studies attempting to develop immunological assays employing these antigens. These assays showed only a limited clinical value, due to low sensitivity that compromises their ability to discriminate between various inflammatory CNS diseases.[111] In fact, the recently revised criteria for the diagnosis of MuS (issued by the International Panel on Diagnosis of MuS) are essentially based on neurologist's clinical observation, supported by magnetic resonance imaging (MRI), and do not include any confirmation derived from immunological assay.[112]

Recently, a number of non-myelin antigens have been reported as putative biomarkers of MuS. These findings highlighted the complexity of the disease and were instrumental in the advancement of novel interpretations of this pathology. For example, by applying a proteomic approach, *Mathey et al* identified neurofascin and contactin-2 as candidate antigens for MuS, demonstrating that antibodies against each of these proteins cause axonal injury and demyelinating lesions in the cortex.[113,114] Since high neurofascin expression occurs predominantly in the brain of MuS patients characterized by chronic progressive disease, it appears that these antigens might be implicated mainly in the degenerative phase of the pathology. Similarly, an exhaustive search for antigens recognized by IgG autoantibodies derived either from sera or cerebrospinal fluid of MuS patients, revealed the presence of both oligodendroglia (transketolase, CNPase) and cytoskeletal proteins (radixin, actin interactin protein 1) in human brain.[115]

Due to the ambiguous identification of the antigens implicated in the disease the development of serum biomarkers to employ for the diagnosis, monitoring, and prognosis of MuS remains a challenging issue.

3.1.2 An N-glucosylated peptide detecting disease-specific autoantibodies, biomarkers of multiple sclerosis

The heterogeneous clinical spectrum of MuS has suggested that a single diagnostic probe is unlikely to serve as a general prognostic tool. Therefore, the development of a number of biomarkers, each one specific for different pathophysiological mechanisms, will be important for better understanding disease pathogenesis and for future drug development in MuS.[116] The role of autoantibodies as potential biomarkers of autoimmune diseases arise from several studies such as those regarding the role of IgM autoantibodies against myelin oligodendrocyte glycoprotein (MOG), reported to be elevated during the first demyelinating event.[117,118]

Taking into account these considerations, at the Interdepartmental Laboratory of Peptide and Protein Chemistry and Biology of the University of Florence an approach has been pursued to develop a specific antigenic probe able to identify a population of MuS patients in which the presence of autoantibodies, recognizing the antigenic probe, was correlated with disease activity.[119]

A "chemical reverse approach" based on the use of synthetic peptides as antigenic probes was employed for the identification of autoantibodies as biomarkers of autoimmune diseases. It was called "reverse" because the autoantibodies circulating in biological fluids of patients guided the selection of the probes. "Chemical" because autoantibodies recognition drived selection and optimization of the best "chemical" structure among defined peptide libraries. Thus, antibodies provided the model for the chemical structure of peptides, that mimicked native epitopes of the antigens involved in the disease pathogenesis. At variance, conventional methods are based on preselected myelin components, which displayed variable efficiency in the recognition properties for autoantibodies.

This strategy is particularly promising when post-translational modifications of proteins are involved in antibody recognition. In fact, the use of synthetic peptides allows the introduction of specific chemical modifications to reproduce the structures that trigger the autoimmune response, in contrast to extracted or recombinant proteins.

Post-translational modifications can play a crucial role in the autoimmunity mechanisms by creating new antigens no longer recognized by the immune system; in particular, aberrant glycosylation can affect the immune response and profoundly influence immune tolerance.[120]

Using this approach the efficient detection of autoantibodies in a large group of MuS patients guided the selection of an optimized synthetic glucopeptide antigen called CSF114(Glc).

Synthetic libraries of peptides bearing unmodified residues or glycosylated residues were analyzed through ELISA assays using MuS patients' sera to check their capacity

to recognize autoantibodies. Peptides were designed to bear small variations of amino acid sequence or sugar moiety in order to rationally achieve optimal conformation of the epitope recognized by antibodies.

The screening of this peptide collection enabled the selection of the sequence of CSF114(Glc), characterized by a N-linked glucose moiety, as the best ligand for antibody binding thus proving that the key feature for antibodies recognition was the glucosylated asparagine residue assembled on the top of a β–hairpin structure. Results were validated on a relevant number of MuS patients confirming that CSG114(Glc) detects serum autoantibodies in a subset of approximately 30% of patients and exhibits an immunoreactive response correlated with the relapsing remitting trend of the disease.[121] Subsequently, the value of the glucopeptide was proved at cellular level by means of immunohistochemistry experiments that revealed CSF114(Glc) specific antibodies to be able to recognize myelin and oligodendrocyte autoantigens in brain tissue.

Therefore CSF114(Glc) acts as a simple, reliable, and an efficient antigenic probe able to detect, isolate, and characterize antibodies as biomarkers of MuS.

3.2 OBJECTIVES

Sophisticated approaches have recently led to the identification of novel autoantigens associated with MuS, e.g. neurofascin, contactin, CNPase and other T cell receptor membrane anchored proteins. These putative antigens, although differing from the conventional myelin derivatives, are conceptually based on an animal model of experimental autoimmune encephalomyelitis. A novel approach, which has shown promising results in terms of predictive value for the MuS pathology, is based on the documented post-translational modifications of native antigens present in the CNS. Pathophysiological post-translational glycosylation of native antigens generates neoantigens that trigger the immune system to generate autoantibodies which escape conventional diagnostic procedures. Previous studies has shown that glycosylated molecules can act as neo antigens surrogates that serve as probes for the detection of disease-related autoantibodies present in the serum of MuS patients.[122,123]

The glucopeptide CSF114(Glc), developed by our group, is capable of identifying and measuring MuS-related autoantibodies whose levels are correlated with clinical assessment of MuS activity and MRI profile of brain lesions. Furthermore, CSF114(Glc) affinity-purified autoantibodies from MuS serum specifically stains myelin and oligodendrocytes antigens in human brain histological specimens.

The CSF114(Glc) peptide represents an unconventional approach since its structure is completely unrelated to myelin oligodendrocyte glycoprotein or any other myelin derivative and is not linked to any particular pathogenetic hypothesis. The main characteristic of CSF114(Glc) is its conformational propensity to form a β-turn that exposes the sugar moiety, perhaps the key element, for recognition by the MuS-relevant autoantibodies.[124,125]

Given this background we wondered whether CSF114(Glc) affinity-purified antibodies from MuS patients' sera, could be useful in back-tracking CNS antigens implicated in the MuS-related autoimmune response. An immunoproteomic approach was followed to identify putative antigens from rat brain tissues using serum MuS autoantibodies purified through CSF114(Glc) affinity columns, hereafter termed "CSF114(Glc) antibodies". MALDI mass spectrometry analysis of the isolated antigens were performed to achieve the protein identifications and further validation experiments were conducted to confirm the recognition.

3.2.1 Scheme of the work

- CSF114(Glc) was covalently immobilized on sepharose columns using a cyanogen-bromide activated resin and used to isolate reactive antibodies from MuS patients' sera and from healthy subjects as negative control.
- Rat brain proteins were extracted using a protocol based on alternating extraction phases and centrifugation steps; the yield of the extraction was checked through SDS-PAGE.
- Protein extract was run on a monodimensional SDS-PAGE to be subsequently immunoblotted with anti-CSF114(Glc)-antibodies; reactive bands from corresponding gel were excised and submitted to trypsin digestion for the MALDI-MS analysis.
- Commercial standard proteins corresponding to the identified ones where run and immunodetected by the same isolated antibodies to confirm the previous recognitions.

3.3 RESULTS AND DISCUSSION

3.3.1 Detection of immunoreactive proteins in rat brain homogenate using anti-CSF114(Glc) antibodies

As the primary aim of this study was to investigate whether antibodies detected by CSF114(Glc) glycopeptide might reveal specific antigens present in brain tissue, the first step of the research concerned the purification of the antibodies. To achieve this goal an affinity chromatography system was prepared immobilizing CSF114(Glc) to a solid support and then sera of MuS patients positive to CSF114(Glc) were pushed through the column to specifically isolate reactive antibodies.

The chromatographic bed was assembled using CNBr activated sepharose beads able to covalently block the peptide. The coupling reaction was performed in presence of an amine-free basic buffer so that only the primary amines of the peptide could react with both the cyanate esters and imidocarbonates groups exhibited by the resin. After an overnight incubation the net result was a covalent coupling of the ligand to the sepharose matrix. Substituted imidocarbonate derivatives compared to charged isourea derivatives are the preferred resultant structures (Figure 3.2).

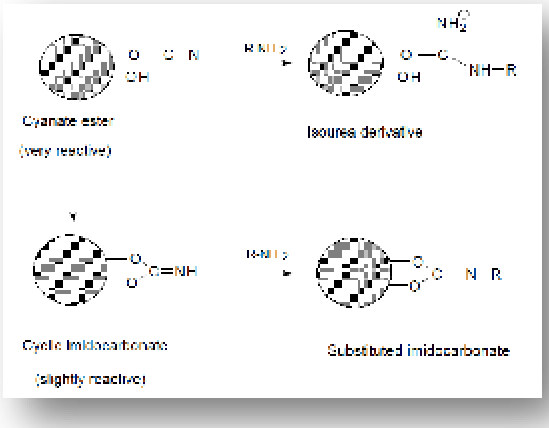


Figure 3.2. Reaction scheme of the peptide coupling to the CNBr-activated Sepharose beads

Autoantibodies responding to CSF114(Glc) were isolated from MuS patients' sera and from normal blood donors' sera, the latter used as negative control. In particular six MuS sera were taken among the 250 previously analyzed for autoantibodies by ELISA with the glycopeptide and belonged to patients in the relapsing-remitting disease stage.[119]

Isolation and purification of antibodies were performed loading each filtered serum onto the column and incubating it for one hour at room temperature; non-specific antibodies were removed with several washes while specific antibodies retained by the column were eluted with an acidic buffer, neutralized with a basic medium and finally concentrated in a small volume. The purified antibody titer of each serum was evaluated by means of ELISA test on CSF114(Glc) coated plates. Results indicated a wide range of antibody levels, from 4- to 20-fold that present in the pooled control serum, taken from blood donors of matching age (Figure 3.3).

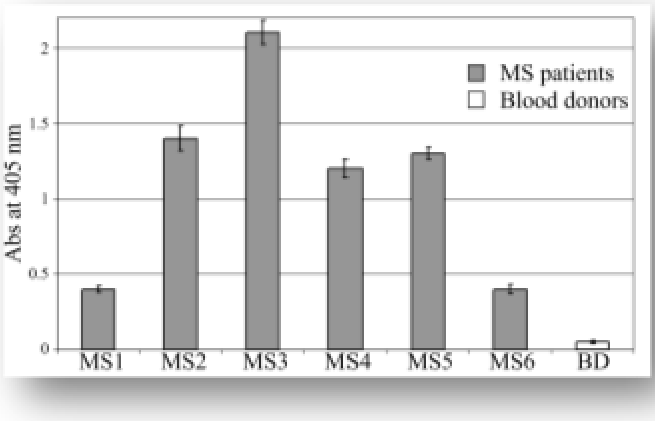


Figure 3.3. ELISA absorbance at 405 nm of autoantibodies purified by affinity column functionalized with CSF114(Glc) glycopeptide (MuS patients and pool of normal blood donor's sera)

The second step of the research concerned the extraction of proteins from brain tissues. As a source of putative antigens the brain was chosen from adult rats, a species in which the above mentioned proteins are conserved as compared to human, and which for its availability, avoids problems of protein degradation. In addition, brain homogenization might disclose proteins (putative autoantigens), which are hidden by the integrity of brain tissue. Rat brain homogenate was obtained after sonication of the tissue, previously minced into small pieces; after a slow centrifugation cellular debris were removed and supernatant re-centrifugated for a longer time. The final pellet was re-suspended in a specific buffer and incubated for one hour; soluble proteins were recovered, the concentration measured by spectrophotometry and the protein profile was examined on SDS-PAGE.

Affinity purified anti-CSF114(Glc)-IgGs from each MuS patient were challenged through western blot against the separated rat brain proteins and three major bands were revealed (Figure 3.4).

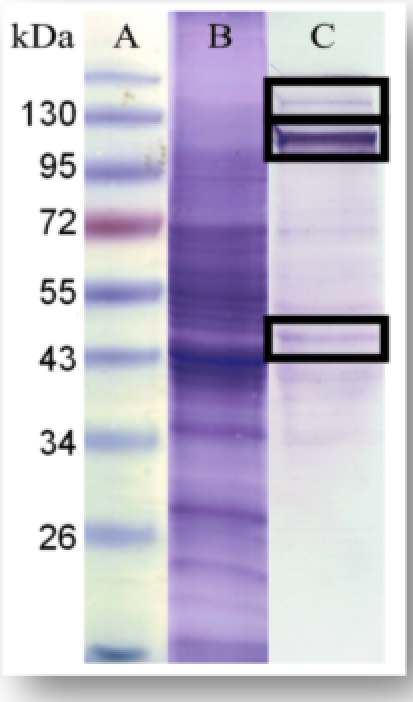


Figure 3.4. Rat brain proteins are selectively recognized by CSF114(Glc) specific IgGs from MuS patients

The bands were characterized by different *Mr*: 130 kDa, 98 kDa and two close bands at 47kDa. The 98 kDa band was recognized by antibodies purified from all patients examined (6/6), while recognition of the other bands occurred only in 3 patients sera.
The finding of these reactivities was in agreement with the mentioned immunoistochemistry results performed on neurological tissues with CSF114(Glc) purified antibodies that had highlighted a diffuse reactivity throughout the white matter.

3.3.2 Identification of antigens through proteomic approach

A thorough proteomic approach was applied to the immunoreactive bands above described to identify the nature of proteins present in rat brain homogenate detected by CSF114(Glc) antibodies.
After excision and trypsinization of individual bands, MALDI analyses were performed obtaining PMFs data for each protein (Figure 3.5).

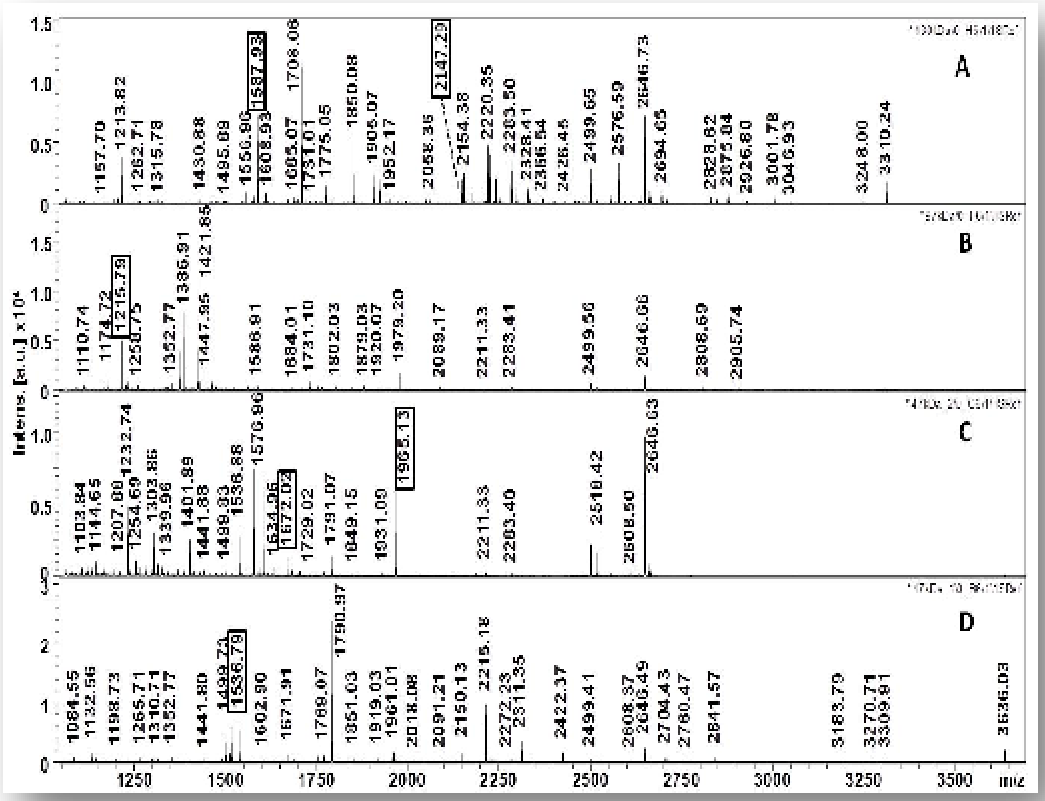


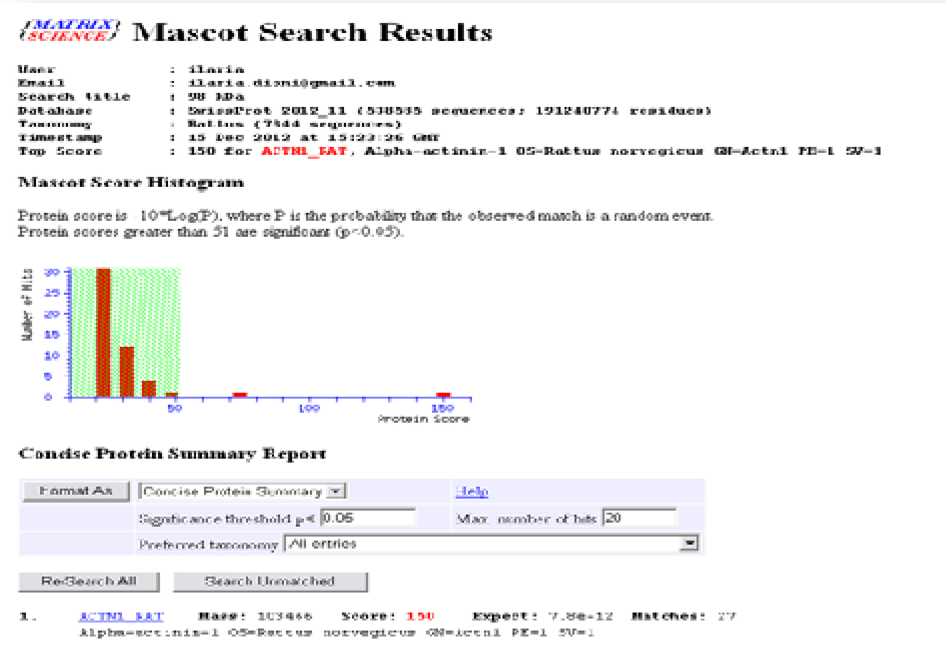
Figure 3.5. MALDI-TOF MS analyses of the four trypsin-digested bands excised from the SDS-PAGE

The search for these fingerprint signatures in the SwissProt rattus database, enabled the identification of the rat brain proteins sensitive to recognition by human antibodies:

- Alpha fodrin corresponding to the 130kDa band,
- Alpha actinin 1 corresponding to the 98 kDa band;
- 2',3'-cyclic-nucleotide 3'-phosphodiesterase and creatine kinase for the two 47 kDa bands.

Results of databank searches follow (Figure 3.6).


Mascot Search Results
User : ilaria
Search title : 130 kDa
Mascot Score Histogram
Protein scores greater than 51 are significant (p<0.05).
Concise Protein Summary Report
Format As
Help
Re-Search All


Mascot Search Results
User : ilaria
Email : ilaria.dinni@gmail.com
Search title : 98 kDa
Database : SwissProt 2012_11 (538585 sequences; 191240774 residues)
Taxonomy : Rattus (7844 sequences)
Timestamp : 15 Dec 2012 at 15:23:26 GMT
Top Score : 150 for ACTN1_RAT, Alpha-actinin-1 OS=Rattus norvegicus GN=Actn1 PE=1 SV=1
Mascot Score Histogram
Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 51 are significant (p<0.05).
Protein Score
Concise Protein Summary Report
Format As
Concise Protein Summary
Help
Significance threshold p< 0.05
Max. number of hits 20
Preferred taxonomy All entries
Re-Search All
Search Unmatched
1. ACTN1_RAT Mass: 103466 Score: 150 Expect: 7.8e-12 Matches: 27
Alpha-actinin-1 OS=Rattus norvegicus GN=Actn1 PE=1 SV=1

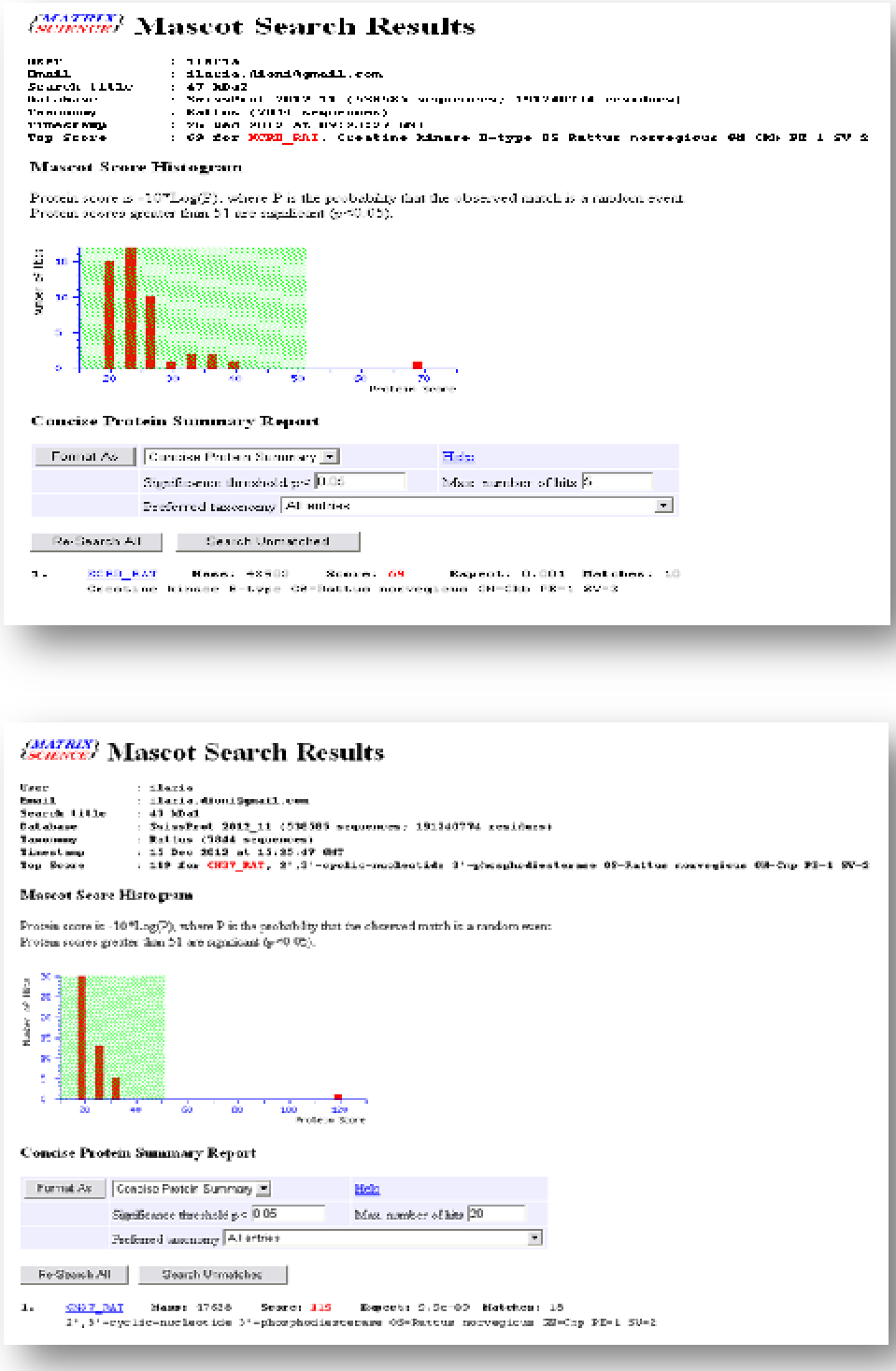


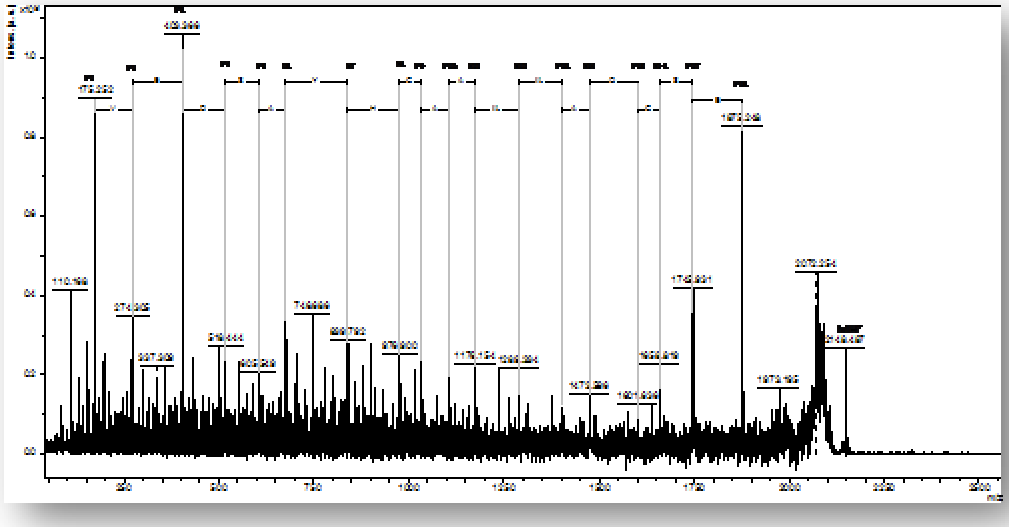
Figure 3.6. Databank searches of the four digested bands starting from the one at 130kDa, 98 kDa and the two at 47 kDa

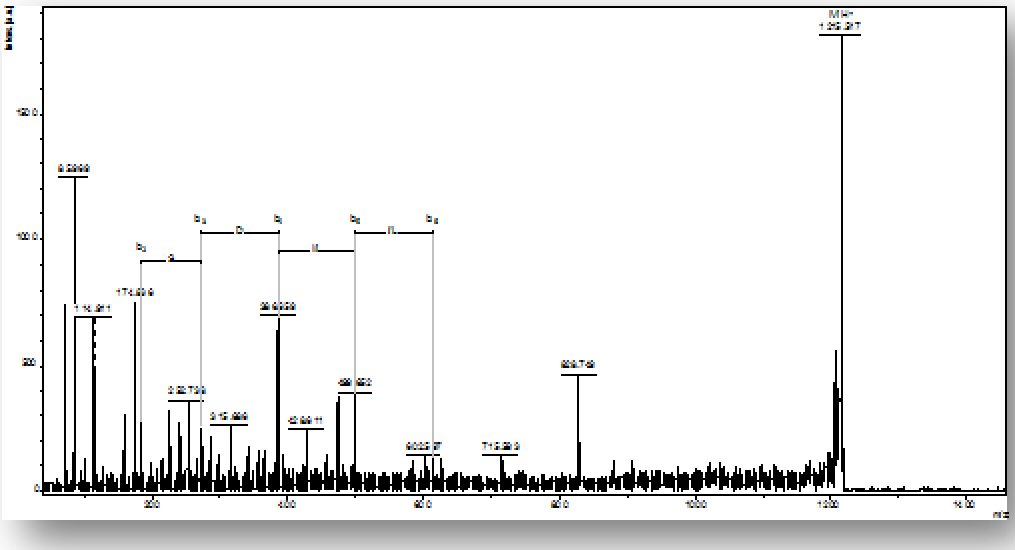
Identifications were confirmed by MS/MS analyses of selected peptides for each protein (Table 3-I).

SDS band	Identified protein	Accession number	Mascot search Score	Matched / Searched	Sequence coverage	MSMS Sequence confirmation	Mass	Position
130 kDa	alpha fodrin	P16086	258	51 / 87	25.9%	SADESGQALLAAGHYASDEVR REELITNWEQIR	2147 1587	419-439 336-347
98 kDa	alpha actinin	Q6T487 rat	150	29 / 101	38.5%	LASDLLEWIR	1215	282-291
47 kDa [2]	creatine kinase	P07335	69	13 / 58	35.2%	TFLVWINEEDHLR GTGGVDTAAVGGVFDVSNADR	1672 1965	224-236 321-341
47 kDa [1]	2',3' CNPase	P13233	119	16 / 96	35.2%	AIGALLYAQGDVVR	1536	261-274

Table 3-I. Identification of putative autoantigens following peptide mass fingerprint search on Mascot engine and MS/MS analyses of selected signals for each protein

One MS/MS spectra for each identified protein is reported consecutively (Figure 3.7).





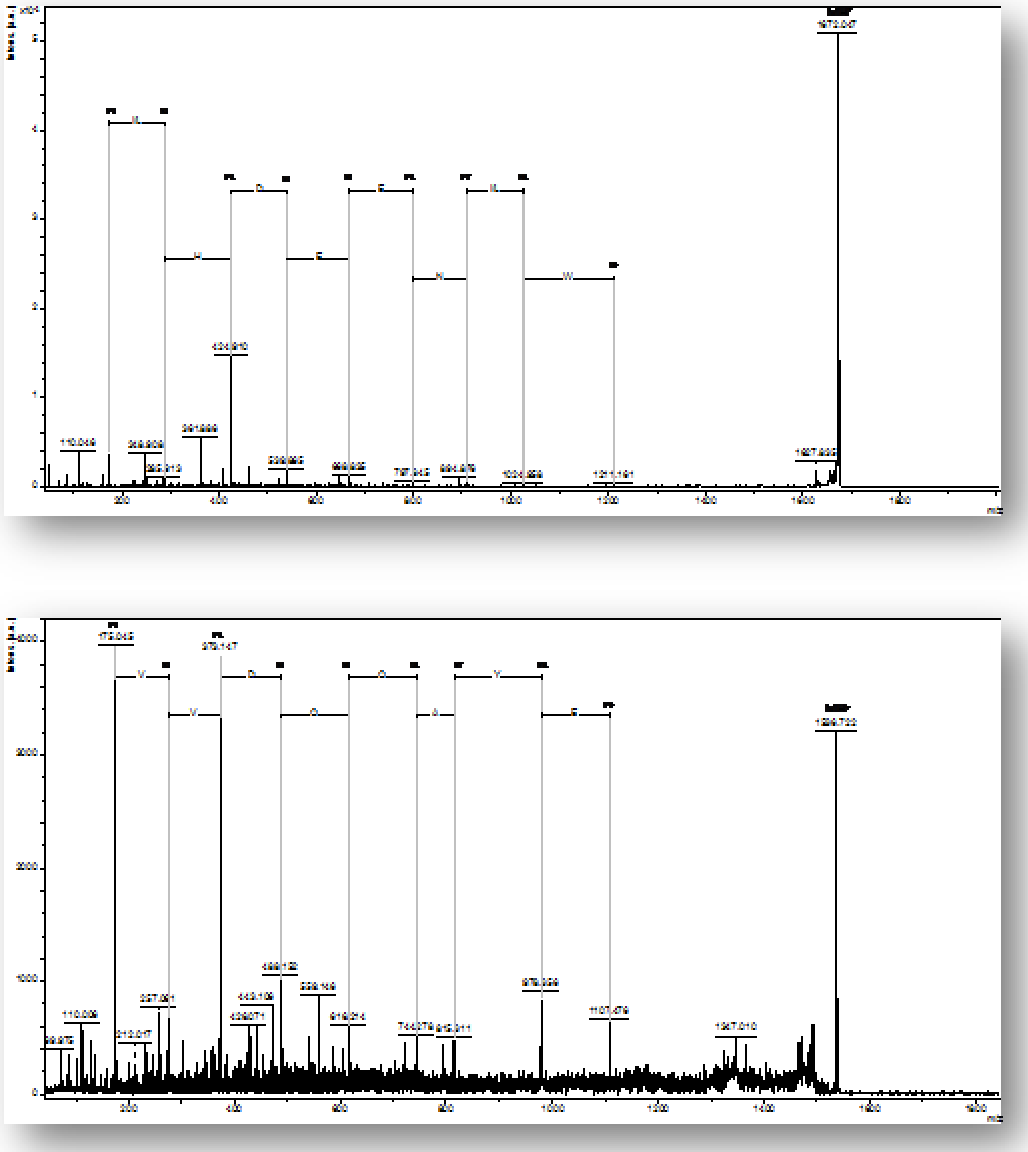


Figure 3.7. MS/MS spectra of precursor at 2147 m/z from alpha fodrin, precursor at 1215 m/z from alpha actinin, precursor at 1672 m/z from cretin kinase and precursor at 1536 m/z from CNPase

Mass spectrometric analyses had revealed two proteins, namely alpha fodrin and alpha actinin, that belong to the spectrin family and are associated with the cytoskeleton. The other two proteins, CNPase and creatine kinase, are, instead, enzymes involved in cellular metabolic processes.

3.3.3 Validation of the identified rat brain antigens

To rule out the possibility that the mass spectrometry analysis might have identified proteins of similar Mr co-localizing in the same band (false positive) the proteins

identified required a further validation. With the use of monodimensional electrophoresis as a protein separation technique, indeed, protein co-localization is unavoidable. At the same time the use of more powerful separation techniques, such as 2D electrophoresis, was hampered by the limited quantity of CSF114(Glc) antibodies available, due to the low yield of the purification procedure from MuS patients' sera.

For this reason the four identified bands, (i.e. alpha fodrin, alpha actinin 1, CNPase and creatine kinase) and their standard counterparts were compared side by side against both IgGs monoclonal antibodies and CSF114(Glc) antibodies. A sequence homology analysis indicated that each standard protein exhibits a significant level of sequence overlapping with the corresponding rat sequence (86.6% – 95.4%), and in the case of alpha actinin, with the corresponding human form (96.9%).

Monoclonal antibodies recognized all standard proteins except alpha fodrin, possibly due to a faulty protein, demonstrating the immunogenicity of the commercial proteins. Similarly monoclonal antibodies also recognized all extracted proteins in similar fashion as CSF114(Glc) antibodies previously did, confirming the presence of the MS-identified proteins in the brain extract. The most interesting results came from the standard proteins challenged by CSF114(Glc) antibodies because only alpha actinin provided a positive band, while all other proteins failed to respond to the antibodies (Figure 3.8).

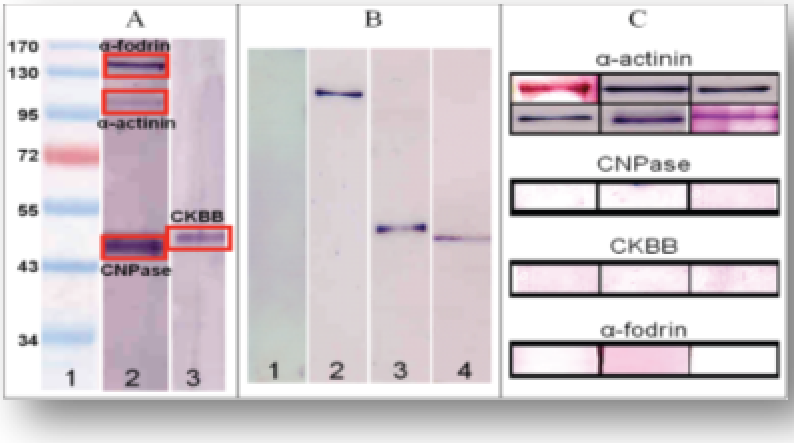


Figure 3.8. Western blot of rat brain proteins with CSF114(Glc) specific IgGs from MuS patients (panel A); authentic proteins challenged with commercially available monoclonal antibodies (panel B) and with anti-CSF114(Glc) antibodies (panel C)

The positive response obtained with alpha actinin was consistent with initial results, showing the presence of specific autoantibodies against this protein in all patient sera.

3.4 CONCLUSIONS

CSF114(Glc) is a glycopeptide developed by the Laboratory of Peptide and Protein Chemistry and Biology (PeptLab) as specific probe for detection of autoantibodies of relapsing remitting subtype of MuS. It was selected by means of a "chemical reverse approach" that allowed autoantibodies response to guide the design and the optimization of its chemical structure. This peptide, bearing a N-glucosylated moiety, revealed to be the most efficient probe in recognizing specific autoantibodies in approximately the 30% of MuS patients' population.

The usefulness of CSF114(Glc) not only resides in the employment as antigenic probe for the detection of antibodies possibly biomarkers of MuS but also as investigative tool to help in delineating the pathogenetic mechanisms of the disease. Indeed serum autoantibodies ought to reflect the presence of corresponding antigens and peptides recognized by those autoantibodies might mimic structural features of neo-epitopes. Therefore an autoantibodies population proven to be relevant for the MuS pathology might be used to look for the proteins that originally could have primed its release, namely the native autoantigens. Since immunohistochemistry experiments had clearly proved the CNS specificity of anti-CSF114(Glc) antibodies, possible reactive proteins were searched among the brain tissue of rats.

Experiments were designed to include four sequential stages: (a) purification of autoantibodies from MuS sera; (b) detection of antigens displaying immunoreactivity toward the autoantibodies; (c) MALDI identification of putative antigens; (d) assessment of immunoreactivity of putative antigens and their authentic protein counterparts against either specific monoclonal antibodies or the purified autoantibodies.

The CSF114(Glc) antibodies, purified and enriched through affinity chromatography, were able to detect three distinct protein bands in the rat brain, which upon MALDI-PMF and MS/MS analyses led to the identification of four proteins, i.e. alpha fodrin, alpha actinin 1, CNPase and creatine kinase. However, with the exception of alpha actinin 1, CNPase, alpha fodrin and creatin kinase, in their authentic form, failed to be recognized by CSF114(Glc) antibodies. They consistently provided a negative immunoreactive response when challenged with the MuS patient derived autoantibodies against the glycosylated probe. This failure might be attributed to post-translational modifications of antigens, well documented particularly for CNPase.[126]

The salient finding of this work was the identification of alpha actinin 1 as the only molecule, which provided immunoreactive responses to the CSF114(Glc) antibodies for both the extracted rat brain protein and the reference standard forms.

Alpha actinin 1 is an ubiquitous cytoskeletal protein, which in its isoforms 1 and 4, both termed non-muscle, has been reported to be involved in autoimmune diseases such as systemic lupus erythematosus and autoimmune hepatitis.[127] To our knowledge, this is the first time that alpha actinin 1 has been identified as one of the targets recognized by specific autoantibodies associated with MuS.

Alpha actinin 1 appears to be implicated in the organization of the cytoskeleton and through its proximity to adherens junctions, might be involved in T-cells migration. As component of adherence and tight junctions, it also modulates blood brain barrier functions.

Anti alpha actinin 1 autoantibodies have been found in lupus nephritis and in autoimmune hepatitis (AIH), and their detection in the latter pathology is an important diagnostic tool. Because these diseases are characterized by inflammatory and degenerative components, prevailing features also of MuS, the identification of alpha actinin 1 as a major autoantibodies in MuS suggests that the underlying mechanisms producing these derangements are similar for a number of autoimmune pathologies. Its detection by autoantibodies from MuS patients, besides signaling its involvement in the pathology, implies a derangement of the barrier shield, leading to exposure of the neurovascular niche to further degenerative injuries.

It is of interest that the use of CSF114(Glc) antibodies approach, which is not linked to any particular pathogenetic mechanism, discloses antigens undetected by conventional approaches, e.g. alpha actinin 1. It appears that the "a priori" designation of a specific pathogenetic autoimmune model for MuS, poses limitations as it reveals antigens confined to that specific mechanism. Recent reports have described an approach for the identification of diagnostically useful antibodies, which bears similarities with our approach that is not based on predefined knowledge of antigens.[128,129] Both studies successfully employ peptide libraries, either synthetic or phage displayed, for capturing autoantibodies involved in immune mediated disorders of the CNS.

In conclusion the distinct advantage of this study lies in the use of purified autoantibodies specific for the N-glucosylated peptide CSF114(Glc), which recognize a limited repertoire of relevant antigens. In contrast, the extensive use in other studies of unselected Ig fractions had accounted for the detection of antigens having remote relevance for the disease.

In our view the observed cross-reactivity between the N-glucosylated peptide and alpha actinin 1 fits the well-known hypothesis that common pathogens are associated with the onset of MuS, suggesting that molecular mimicry may cooperate with the adaptive immune system leading to recognition of brain antigens.[130] In addition the finding of the double reactivity of specific MuS autoantibodies toward a synthetic

peptide, CSF114(Glc), and a specific protein, alpha actinin 1, constitute a further validation of CSF114(Glc) as antigenic probe able to detect antibodies as biomarkers of MuS.

Our findings support, also, the hypothesis that alpha actinin 1, a cytoskeleton protein implicated in inflammatory/ degenerative autoimmune diseases (lupus nephritis and autoimmune hepatitis), might be regarded as a novel MuS autoantigen, perhaps a prototypic biomarker for the inflammatory/degenerative process typical of the disease. Although it is difficult to assess, at this point, what is the role of alpha actinin 1 in the pathogenesis of MuS, its detection in several autoimmune diseases, which underscores its involvement in inflammatory/degenerative pathologies, warrants further investigations.

3.4 EXPERIMENTAL PART

CSF114(Glc)-Sepharose Coupling

CSF114(Glc)-Sepharose column was prepared by using Cyanogen-Bromide Activated resin (CNBr-Sepharose® 4B, Sigma). CNBr-Sepharose beads (60 mg, coupling capacity 30-40 mg of ligand per ml) were washed thoroughly with 1 mM HCl, equilibrated with coupling buffer (0.1 M $NaHCO_3$ containing 0.5 M NaCl, pH 8.3) and coupled to 1 mg of glycopeptide CSF114(Glc). Unreacted ligand was washed and unreacted sites were blocked with 0.2 M glycine (pH 8). Column was washed thoroughly and equilibrated with PBS.[119]

Affinity Purification

Multiple Sclerosis patients' serum from six individual patients (MuS1 – MuS6) was diluted in PBS (1:10) and filtered through 0.22 µm filter (MILLEX®, GS). The study was approved by the institutional ethic committee. All patient included corresponded to relapsing-remitting MuS and had no comorbidities. Filtered serum was loaded onto the column and incubated for one h at room temperature. Column was washed with PBS buffer to remove non-specific antibodies, after washing specific antibodies were eluted with 0.1 M glycine buffer, pH 2.6. Eluted fractions were neutralized with 0.1 M $NaHCO_3$ pH 8.3 and concentrated on Amicon Ultra-0.5 ml, 50K (Millipore). Immunoreactivity of anti-CSF114(Glc) IgGs was examined by ELISA on CSF114(Glc) coated plates.[121]

Extraction of Brain Proteins from Rat

120 mg of frozen rat brain was excised and minced into small pieces. Sample was dissolved in 1 ml PBS and sonicated (Brandelin SONOPLUS Probe KE-76, Sigma) for 2 min in ice. Homogenate was centrifuged at 10,000 g for 15 min and debris were removed. Supernatant was re-centrifuged at 31,870 x g for 90 min at 4°C. Pellet was solubilized in 0.01% Triton X-100 buffer containing a cocktail of protease inhibitors (Complete™ Mini tablets, Roche Applied Science) and NaN_3 (0.05%) at 37°C for one h. Total protein concentration was 17.9 mg/ml, measured by spectrophotometry (NanoDrop 1000, Thermo Scientifics). 10 µl of sample were loaded and examined on SDS-PAGE 12%.

SDS-PAGE and Western Blot Analysis

Protein samples were subjected to SDS-PAGE 12% under reducing condition (2% v/v β-mercaptoethanol). The following standard proteins were obtained from commercial sources: recombinant human alpha fodrin (kindly provided by Aesku Diagnostics

GmbH, Germany), alpha actinin isoform 1 (brain), extracted from chicken gizzard (Sigma, Italy), recombinant human CNPase expressed in E. coli (Abcam, USA), recombinant human creatine kinase expressed in Pichia pastoris (Sigma, Italy). Gels were either stained with Coomassie blue or used to transfer onto nitro cellulose membrane (0.45 µm; BIO-RAD) at 220 mA for two h at room temperature. Membrane was blocked with milk powder 2% in TBST (0.1%) and incubated with affinity purified anti-CSF114(Glc)-IgGs (MuS patients and normal blood donors) for 1 h at room temperature. Membrane was washed thoroughly with TBST (0.1%) and incubated with alkaline phosphatase conjugated anti-IgGs (1:500; Sigma) for 1 h at room temperature. Membrane was washed as above with TSBT (0.1%) and blots were developed using AP conjugated substrate kit (BIO-RAD).

Excision and digestion

The samples were obtained by excision and digestion of the bands from SDS-PAGE 12% of rat brain homogenate, after comparison with the Western blot results to identify the bands of interest. The bands were excised and minced into small pieces, washed with NH_4HCO_3 50 mM and shrunk with acetonitrile. Reduction and alkylation were performed with βME 20 mM (56°C, 30 min) followed by 2-iodoacetamide 55 mM (20°C, 20 min, in the dark). Proteins were digested in a buffer containing 25 mM NH_4HCO_3, 5 mM $CaCl_2$ and ~20 ng/µl of trypsin (37°C, overnight). Peptides were extracted from gel in two steps by adding 25 mM NH_4HCO_3 and 5% formic acid; each step was followed by addition of acetonitrile to shrink the gel and maximize the peptide recovery.

MALDI Analysis

MALDI-MS measurements were performed using an Ultraflex III TOF-TOF instrument (Bruker Daltonics, Bremen, Germany), operating in reflectron positive ion mode. Ions were formed by a pulsed UV laser (λ = 337 nm) beam. The instrumental conditions were: UIS1 = 25 kV; UIS2 = 21.65 kV; reflectron potential: 26.3 kV; delay time = 0 nsec. The samples were dried, re-suspended in acetonitrile and desalted by Zip-Tip C18 pipette tips (Millipore). Zip-tipped samples were spotted on the stainless steel sample plate with α-cyano-4-hydroxycinnamic acid matrix (5 mg in H_2O/Acetonitrile/0.1% TFA). 1µL of the peptide eluted solution was deposited on the stainless steel sample holder, and allowed to dry before introduction into the mass spectrometer. External mass calibration was done using the Peptide Calibration Standard, basing on the monoisotopic values of $[M+H]^+$ of Angiotensin II, Angiotensin I, Substance P, Bombesin, ACTH clip (1-17), ACTH clip (18-39), Somatostatin 28 at m/z 1046.542, 1296.685, 1347.736, 1619.823, 2093.087, 2465.199 and 3147.471, respectively. MALDI-TOF-TOF experiments were carried out using the LIFT device. The instrumental parameters were: UIS1=8 kV; UIS2=7.2 kV; ULIFT1= 19 kV.

The resulting data were processed using the FlexAnalysis 2.4 software (Bruker Daltonics) and optimized for databank search. Searches were performed using the Mascot Server 2.3 search engine against the Swiss-Prot database (release 2012_8, subset Rattus, 7800 entries) considering up to one missed tryptic cleavage, monoisotopic peptide mass tolerance of 500 ppm, and fragment ion mass tolerance of 0.3 Da. Carbamidomethyl modification of cysteine and oxidation of methionine were considered as appropriate.

REFERENCES

[1] G.S. Firestein. Evolving concepts of rheumatoid arthritis. Nat, **2003**, *423*, 356-361.

[2] R. Yamada, A. Suzuki, X. Chang, K. Yamamoto. Citrullinated proteins in rheumatoid arthritis. Front. Biosci., **2005**, *10*, 54-64.

[3] N. Wegner, K. Lundberg, A. Kinloch, B. Fisher, V. Malmström, M. Feldmann, P.J. Venables. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. Immunol. Rev., **2010**, *233*, 34-54.

[4] P. Emery, F.C. Breedveld, M. Dougados, J.R. Kalden, M.H. Schiff, J.S. Smolen. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. Ann. Rheum. Dis., **2002**, *61*, 290-297.

[5] R.L. Nienhuis, E. Mandema. A new serum factor in patients with rheumatoid arthritis; the antiperinuclear factor. Ann. Rheum. Dis., **1964**, *23*, 302–305.

[6] B.J. Young, R.K. Mallya, R.D. Leslie, C.J. Clark, T.J. Hamblin. Anti-keratin antibodies in rheumatoid arthritis. Br. Med. J., **1979**, *2*, 97–99.

[7] M. Simon, E. Girbal, M. Sebbag, C. Vincent, C. Masson-Bessiere, J.J. Durieux. The antiperinuclear factor and the so called antikeratin antibodies are the same rheumatoid arthritis specific autoantibodies; J.Clin. Invest., **1995**, *95*, 2672-2679.

[8] E. Girbal-Neuhauser, J. J. Durieux, M. A. P. Dalbon, M. Sebbag, C. Vincent, M. Simon, T. Senshu, C. Masson-Bessière, C. Jolivet-Reynaud, M. Jolivet, G. Serre. The epitopes targeted by the rheumatoid arthritis-associated antifilaggrin autoantibodies are post translationally generated on various sites of (pro)filaggrin by deimination of arginine residues. J. Immunol., **1999**, *162*, 585–594.

[9] G.A. Schellekens, B.A. de Jong, F.H. van den Hoogen, L.B. van de Putte, W.J. van Venrooij. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J. Clin. Invest., **1998**, *101*, 273–281.

[10] H.S. El-Gabalawy, J.A. Wilkins. Anti-Sa antibodies: prognostic and pathogenetic significance to rheumatoid arthritis. Arthritis Res Ther., **2004**, *6*, 86-89.

[11] E.R. Vossenaar, N. Deprés, E. Lapointe, A. van der Heijden, M. Lora, T Senshu, W.J. Venrooij, H.A. Ménard. Rheumatoid arthritis specific anti-Sa antibodies target citrullinated vimentin. Arthritis Res. Ther., **2004**, *6*, R142–R150.

[12] G. Steiner, J. Smolen. Autoantibodies in rheumatoid arthritis and their clinical significance. Arthritis Res. **2002**, *4*, S1-S5.

[13] S. Dubucquoi, E. Solau-Gervais, D. Lefranc, L. Marguerie, J. Sibilia , J. Goetz, V. Dutoit, A.L. Fauchais, E. Hachulla, R.M. Flipo, L. Prin. Evaluation of anti-citrullinated filaggrin antibodies as hallmarks for the diagnosis of rheumatic diseases. Ann. Rheum. Dis., **2004**, *63*, 415–419.

[14] M.M. Nielen, D. van Schaardenburg, H.W. Reesink, R.J. van de Stadt, I.E. van der Horst-Bruinsma, M.H. de Koning, M.R. Habibuw, J.P. Vandenbroucke, B.A. Dijkmans. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum., **2004**, *50*, 380–386.

[15] B. Combe at al., EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann. Rheum. Dis., **2007**, *66*, 34-45.

[16] W.J. van Venrooij, J.J.B.C. Joyce van Beers, G.J.M. Pruijn. Anti-CCP Antibody, a Marker for the Early Detection of Rheumatoid Arthritis. Ann. N.Y. Acad. Sci., **2008**, *1143*, 268–285.

[17] A.H.M. van der Helm-van Mil, K.N. Verpoort, F.C. Breedveld, R.E.M. Toes, T.W.J Huizinga. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. Arthritis Res. Ther., **2005,** *7*, R949–R958.

[18] J. Vencovský, S. Macháček, L. Šedová, J. Kafková, J. Gatterová, V. Pešáková, S. Růžičková, Autoantibodies can be prognostic markers of an erosive disease in early rheumatoid arthritis. Ann. Rheum. Dis., **2003**, *62*, 427–430

[19] N. Wegner, K. Lundberg, A. Kinloch, B. Fisher, V. Malmström, M. Feldmann, P.J. Venables, Autoimmunity to specific citrullinated proteins gives the first clues to the ethiology of rheumatoid arthritis. Immun. Rev., **2010**, *233*, 34-54.

[20] K. Tilleman, A. Union, T. Cantaert, S. De Keyser, A. Daniels, D. Elewaut, F. De Keyser, D. Deforce. In pursuit of B-cell synovial autoantigens in rheumatoid arthritis: Confirmation of citrullinated fibrinogen, detection of vimentin, and introducing carbonic anhydrase as a possible new synovial autoantigen. Proteomics Clin. Appl., **2007**, *1*, 32-46.

[21] P.K. Gregersen, J. Silver, R.J. Winchester. The shared epitope hypothesis: an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum., **1987**, *30*, 1205-1213.

[22] X. Chang, J. Han, L. Pang, Y. Zhao, Y. Yang, Z. Shen. Increased PADI4 expression in blood and tissues of patients with malignant tumors. BMC Cancer., **2009**, *9*, 40-51.

[23] P. Li, H. Yao, Z. Zhang, M. Li, Y. Luo, P.R. Thompson, D.S. Gilmour, Y. Wang. Regulation of p53 target gene expression by peptidylarginine deiminase 4. Mol. Cell Biol., **2008**, *28*, 4745-4758.

[24] M.A. Moscarello, F.G. Mastronardi, D.D. Wood. The role of citrullinated proteins suggests a novel mechanism in the pathogenesis of multiple sclerosis. Neurochem Res., **2007,** *32*, 251-256.

[25] H. Uysal, K.S. Nandakumar, C. Kessel, S. Haag, S. Carlsen, H. Burkhardt, R. Holmdahl. Antibodies to citrullinated proteins: molecular interactions and arthritogenicity. Immunol Rev., **2010**, *233*, 9-33.

[26] E.R. Vossenaar, A.J. Zendman, W.J. van Venrooij,G.J. Pruijn. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. Bioessays, **2003**, *25*, 1106-1118.

[27] A. Suzuki *et al.*, Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat. Genet., **2003**, *34*, 395-402.

[28] T. Senshu, T. Sato,T. Inoue, K. Akiyama et al., Detection of citrulline residues in deiminated proteins on polyvinylidene difluoride membrane. Anal. Biochem., **1992**, *203*, 94–100.

[29] M. De Ceuleneer, K. Van Steendam, M. Dhaenens, D. Deforce. In vivo relevance of citrullinated proteins and the challenges in their detection. Proteomics, **2012**, *12*, 752–760.

[30] A. Holm, F. Rise N.Sessler, L.M. Sollid, K. Undheim, B. Fleckenstein. Specific modification of peptide-bound citrulline residues. Anal. Biochem., **2006**, *352,* 68–76.

[31] M. Stansland, A. Holm, A. Kiehne, B. Fleckenstein. Targeted analysis of protein citrullination using chemical modification and tandem mass spectrometry. Rapid Commun. Mass Spectrom., **2009**, *23,* 2754–2762.

[32] M. De Ceuleneer, De Wit V, K. Van Steendam, F. Van Nieuwerburgh, K. Tilleman, D. Deforce. Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. Rapid Commun. Mass Spectrom., **2011**, *25*, 1536-1542.

[33] A.E. Tutturen, A.Holm, M. Jørgensen, P. Stadtmüller, F. Rise, B. Fleckenstein. A technique for the specific enrichment of citrulline-containing peptides. Anal. Biochem., **2010**, 403, 43-51.

[34] K.L. Bicker, V. Subramanian, A.A. Chumanevich, L.J. Hofseth, P.R. Thompson. Seeing citrulline: development of a phenylglyoxal-based probe to visualize protein citrullination. J. Am. Chem. Soc., **2012**, *134*, 17015-17018.

[35] K. Takahashi. The reaction of phenylglyoxal with arginine residues in proteins. J. Biol. Chem., **1968**, *243*, 6171-6179.

[36] K. Kubota, T. Yoneyama-Takazawa, K.Ichikawa. Determination of sites citrullinated by peptidylarginine deiminase using ^{18}O stable isotope labeling and mass spectrometry. Rapid Commun. Mass Spectrom., **2005**, *19*, 683-688.

[37] M. Hermansson, K. Artemenko, E. Ossipova, H. Eriksson, J. Lengqvist, D.Makrygiannakis, A.L.Catrina, A.P. Nicholas, L.Klareskog, M. Savitski, R.A.Zubarev, P.J.Jakobsson. MS analysis of rheumatoid arthritic synovial tissue identifies specific citrullination sites on fibrinogen. Proteomics Clin. Appl., **2010**, *4*, 511-518.

[38] G.Hao, D. Wang, J. Gu, Q. Shen, S.S. Gross, Y. Wang. Neutral loss of isocyanic acid in peptide CID spectra: a novel diagnostic marker for mass spectrometric identification of protein citrullination. J. Am. Soc. Mass Spectrom., **2009**, *20*, 723–727.

[39] A.J. Creese, M.M. Grant, I.L.C. Chapple, H.J. Cooper. On-line liquid chromatography neutral loss-triggered electron transfer dissociation mass spectrometry for the targeted analysis of citrullinated peptides. Anal. Methods, **2011**, *3*, 259–266.

[40] M. De Ceuleneer, K. Van Steendam, M. Dhaenens, D. Elewaut, D. Deforce. Quantification of citrullination by means of skewed isotope distribution pattern. J. Proteome Res., **2012**, 11, 5245-5251

[41] S. Tsvetkov, E. Ivanova, L. Djondjurov. The pool of histones in the nucleosol and cytosol of proliferating Friend cells is small, uneven and chasable. Biochem J. **1989**, *264*, 785-791.

[42] L. Chang, S.S. Loranger, C. Mizzen, S.G. Ernst, C.D. Allis, A.T. Annunziato. Histones in transit: cytosolic histone complexes and diacetylation of H4 during nucleosome assembly in human cells. Biochemistry. **1997**, *36*, 469-80.

[43] J. Rosenbluh, E. Hariton-Gazal, A. Dagan, S. Rottem, A. Graessmann, A. Loyter. Translocation of histone proteins across lipid bilayers and Mycoplasma membranes. J Mol Biol. **2005**, *345*, 387-400.

[44] E. Hariton-Gazal, J. Rosenbluh, A. Graessmann, C. Gilon, A Loyter. Direct translocation of histone molecules across cell membranes. J Cell Sci. **2003**, *116*, 4577-4586.

[45] C.B. Park, M.S. Kim, S.C. Kim. A novel antimicrobial peptide from Bufo bufo gargarizans. Biochem. Biophys. Res. Commun. **1996**, *218*, 408–413.

[46] I.Y. Park, C.B. Park, M.S Kim, S.C. Kim. Parasin I, an antimicrobial peptide derived from histone H2A in the catfish, Parasilurus asotus. FEBS Lett. **1998**, 437, 258–262.

[47] G.A.Birkemo, T. Luders, O. Andersen, I.F. Nes, J. NissenMeyer. Hipposin, a histone-derived antimicrobial peptide in Atlantic halibut (Hippoglossus hippoglossus L.). Biochim. Biophys. Acta, **2003**, *1646*, 207–215.

[48] S.A. Patat, R.B. Carnegie, C. Kingsbury, P.S. Gross, R. Chapman, K.L. Schey. Antimicrobial activity of histones from hemocytes of the Pacific white shrimp. Eur J Biochem. **2004**, *271*, 4825-4833.

[49] V. Brinkmann, U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D.S. Weiss, Y. Weinrauch, and A. Zychlinsky. Neutrophil extracellular traps kill bacteria. Science. **2004**, *303*,1532–1535.

[50] Y. Wang, M. Li, S. Stadler, S. Correll, P. Li, D. Wang, R. Hayama, L. Leonelli, H. Han, S.A. Grigoryev, C.D. Allis, S.A. Coonrod. Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation. J Cell Biol. **2009**, *184*, 205-213.

[51] S. Sidoli, L. Cheng, O.N.Jensen. Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. J Proteomics. **2012**, *75*, 3419-3433.

[52] M. Von Köckritz-Blickwede, V. Nizet. Innate immunity turned inside-out: antimicrobial defense by phagocyte extracellular traps. J Mol Med. **2009**, *87*, 775-783.

[53] T.A. Fuchs, U. Abed, C. Goosmann, R. Hurwitz, I. Schulze, V. Wahn, Y. Weinrauch, V. Brinkmann, A. Zychlinsky. Novel cell death program leads to neutrophil extracellular traps. J Cell Biol. **2007**, *176*, 231-41.

[54] K. Kessenbrock, M. Krumbholz, U. Schönermarck, W. Back, W.L. Gross, Z. Werb, H.J. Gröne, V. Brinkmann, D.E. Jenne. Netting neutrophils in autoimmune small-vessel vasculitis. Nat Med. **2009**, *15*, 623-625.

[55] A. Hakkim, B.G. Fürnrohr, K. Amann, B. Laube, U.A. Abed, V. Brinkmann, M. Herrmann, R.E. Voll, A. Zychlinsky. Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. Proc Natl Acad Sci U S A., **2010**, *107*, 9813-9818.

[56] D. English, B.R. Andersen. Single-step separation of red blood cells. Granulocytes and mononuclear leukocytes on discontinuous density gradient of ficoll-hypaque. J. Immunol. Methods, **1974**, *5*, 249-252.

[57] I. Striz, I. Trebichavsky. Calprotectin – a pleiotropic molecule in acute and chronic inflammation. Physiol. Res., **2004**, *53*, 245-253.

[58] C. Perera, H.P. McNeil, C.L. Geczy. S100 Calgranulins in inflammatory arthritis. Immunol. Cell Biol., **2010**, *88*, 41-49.

[59] Webinar: Nano LC Accuracy for Proteomics Applications. Thermo Fischer scientific & Eksigent. http://www.labmanager.com/?articles.view/articleNo/4814/article/Webcast--Nano-LC-Accuracy-for-Proteomics-Applications

[60] R.T. Gallagher, M.P. Balogh, P. Davey, M.R. Jackson, I. Sinclair, L.J. Southern. Combined electrospray ionization-atmospheric pressure chemical ionization source for use in high-throughput LC-MS applications. Anal. Chem., **2003**, *75*, 973-977.

[61] M. Karas, U. Bahr, T. Dülcks. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J. Anal. Chem., **2000**, *366*, 669-676.

[62] R. Juraschek, T.Dülcks, M. Karas. Nanoelectrospray-More Than Just a Minimized-Flow Electrospray Ionization Source. J. Am. Soc. Mass. Spectrom., **1999**, *10*, 300-308.

[63] A. Schmidt, M. Karas. Effect of Different Solution Flow Rates on Analyte Ion Signals in Nano-ESI MS,or: When Does ESI Turn into Nano-ESI? J. Am. Soc. Mass Spectrom., **2003**, *14*, 492-500.

[64] GE Healthcare Handbook. "Hydrophobic Interaction and Reversed Phase Chromatography. Principles and Methods".

[65] W. Jiang, G. Fischer, Y. Girmay, K. Irgum. Zwitterionic stationary phase with covalently bonded phosphorylcholine type polymer grafts and its applicability to separation of peptides in the hydrophilic interaction liquid chromatography mode. J. Chromatogr. A., **2006**, *1127*, 82-91.

[66] G. Hao, D.C. Wang, J. Gu, Q Y. Shen, S.S. Gross and Y. M. Wang. Neutral loss of isocyanic acid in peptide CID spectra: a novel diagnostic marker for mass spectrometric identification of protein citrullination. J. Am. Soc. Mass Spectrom., **2009**, *20*, 723–727.

[67] R. Raijmakers, J.J. Van Beers, M El-Azzouny, N.F. Visser, B. Božič, G.J. Pruijn, A.J. Heck. Elevated levels of fibrinogen-derived endogenous citrullinated peptides in synovial fluid of rheumatoid arthritis patients. Arthritis Res. Ther., **2012**, *14*, R114-R123.

[68] K. Sergeant, C. Pinheiro, J.F. Hausman, C. P. Ricardo, J. Renaut. Taking Advantage of Nonspecific Trypsin Cleavages for the identification of Seed Storage Proteins in Cereals. J. Prot. Res., **2009**, *8*, 3182-3190.

[69] P. Rehulka, J. Chmelík, G. Allmaier. The influence of nonspecific cleavage sites on identification of low molecular mass proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry with seamless post-source decay fragment ion analysis. Rapid Commun. Mass Spectrom., **2005**, *19*, 79-82.

[70] http://web.expasy.org/findpept/

[71] A. Gattiker, W.V. Bienvenut, A. Bairoch, E. Gasteiger. FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification. Proteomics., **2002**, *2*, 1435-1444.

[72] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, D.F. Hochstrasser.Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol., **1999,** *112,* 531-52.

[73] K. Nakashima, T. Hagiwara, M. Yamada. Nuclear localization of peptidylarginine deiminase V and histone deimination in granulocytes. J. Biol. Chem., **2002**, *277*, 49562-49568.

[74] Y. Wang, J. Wysocka, J. Sayegh, Y.H. Lee, J.R. Perlin, L. Leonelli, L.S. Sonbuchner, C.H. McDonald, R.G. Cook, Y. Dou, R.G. Roeder, S. Clarke, M.R. Stallcup, C.D. Allis, S.A. Coonrod. Human PAD4 regulates histone arginine methylation levels via demethylimination. Science, **2004**, *306*, 279-283.

[75] G.L. Cuthbert, S. Daujat, A.W. Snowden, H. Erdjument-Bromage, T. Hagiwara, M. Yamada, R. Schneider, P.D. Gregory, P. Tempst, A.J. Bannister, T. Kouzarides. Histones deimination antagonizes arginine methylation. Cell, **2004**, *118*, 545-553.

[76] T. Hagiwara, Y. Hidaka, M. Yamada. Deimination of histone H2A and H4 at arginine 3 in HL-60 granulocytes. Biochemistry, **2005**, *44*, 5827-5834.

[77] I. Neeli, S.N. Khan, M. Radic. Histone deimination as a response to inflammatory stimuli in neutrophils. J. Immunol., **2008**, *180*, 1895-1902.

[78] C.F. Urban, D. Ermert, M. Schmid, U. Abu-Abed, C. Goosmann, W. Nacken *et al.* Neutrophils extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against Candida albicans. PLoS Pathogens, **2009**, *5*(10):e1000639.

[79] A. Kinloch, K. Lundberg, R. Wait, N. Wegner, N.H. Lim, A.J.W. Zendman, T. Saxne, V. Malmstrom, P.J. Venables. Synovial Fluid Is a Site of Citrullination of Autoantigens in Inflammatory Arthritis. Arthritis Rheum., **2008**, *58*, 2287-2295.

[80] Q. Remijsen, T. Vanden Berghe, E. Wirawan, B. Asselbergh, E. Parthoens, R. De Rycke, S. Noppen, M. Delforge, J. Willems, P. Vandenabeele. Neutrophils extracellular trap cell death requires both autophagy and superoxide generation. Cell Res., **2011**, *21*, 290-304.

[81] Julka, F. Reigner. Quantification in proteomics through stable isotope coding: a review. J. Proteome Res., **2004**, *3*, 350-363.

[82] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster. Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem., **2007**, *389*, 1017-1031.

[83] X. Yao. Derivatization or not: a choice in quantitative proteomics. Anal. Chem., **2011**, *83*, 4427-4439.

[84] R. Westermeier, T. Naven, H.-R. Höpker. Proteomics in practice. A guide to successful experimental design 2nd Ed.Wiley-vch, **2008**.

[85] S.P. Gygi, B. Rist. S.A. Gerber, F. Turecek, M. Gelb, R. Aebersold. Quantitative analysis of complex mixtures using isotope-coded affinity tags. Nat. Biotechnol., **1999**, *17*, 994-999.

[86] A. Schmidt, J. Kellermann, F. Lottspeich. A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics, **2005**, *5*, 4-15.

[87] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem., **2003**, *75*, 1895-1904.

[88] P.L. Ross, Y.N. Huang, J. Marchese. Multiplexed protein quantitationin saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics, **2004**, *3*, 1154-1169.

[89] J.L. Hsu, S.Y. Huang, N.H. Chow, S.H. Chen. Stable-isotope dimethyl labeling for quantitative proteomics. Anal. Chem., **2003**, *75*, 6843-6852.

[90] C. Ji, L. Li. Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS. J. Proteome Res., **2005**, *4*, 734-742.

[91] P.J. Boersema, T.T. Aye, T.A.B. Van Veen, A.J.R. Heck, S. Mohammed. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics, **2008**, *8*, 4624-4632.

[92] P.J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomix. Nat. Protoc., **2009**, *4*, 484-494.

[93] Q. Fu, L. Li. De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. Anal. Chem., **2005**, *77*, 7783-7795.

[94] J.L. Hsu,S.-Y. Huang, J.-T. Shiea, W.-Y. Huang, S.-H. Chen. Beyond quantitative proteomics: signal enhancement of the a1 ion as a mass tag for peptide sequencing using dimethyl labeling. *J. Proteome Res.*, **2005**, *4*, 101-108.

[95] J.L. Hsu, S.-H. Chen, D.-T. Li, F.-K. Shi. Enhanced a1 fragmentation for dimethylated proteins and its application for terminal identification and comparative protein quantitation. J. Proteome Res., **2007**, *6*, 2376-2383.

[96] C. Ji, W. Li, X.D. Ren, A.F. El-Kattan, R. Kozak, S. Fountain, C. Lepsy. Diethylation labeling combined with UPLC/MS/MS for simultaneous determination of a panel of monoamine neurotransmitters in rat prefrontal cortex microdialysates. Anal. Chem., **2008**, *80*, 9195-9203.

[97] M.E. Barrios-Llerena, J.C. Pritchard, L.E. Kerr, T. Le Bihan. The use of a novel quantitation strategy based on Reductive Isotopic Di-Ethylation (RIDE) to evaluate the effect of glufosinate on the unicellular algae *Ostreococcus tauri*. J. Proteomics, **2011**, *74*, 2798-2809.

[98] C.H. Fox, F.B. Johnson, J. Whiting, P.P. Roller. Formaldehyde fixation. J. Histochem. Cytochem., **1985**, *33*, 845-853.

[99] B. Metz, G.F.A. Kersten, P. Hoogerhout, H.F. Brugghe, H.A.M. Timmermans, A. de Jong, H. Meiring, J. ten Hove, W.E. Hennink, D.J.A. Crommelin, W. Jiskoot. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. J. Biol. Chem., **2004**, *279*, 6235-6243.

[100] B. Metz, G.F.A. Kersten, G.J. Baart, A. de Jong, H. Meiring, J. ten Hove, M.J. van Steenbergen, W.E. Hennink, D.J. Crommelin, W. Jiskoot. Identification of formaldehyde-induced modifications in proteins: reactions with insulin. Bioconjug. Chem., **2006**, *17*, 815-822.

[101] S.E. Ong, M. Mann. Mass spectrometry-based proteomics turns quantitative. Nat. Chem. Biol., **2005**, 1, 252-262.

[102] S.S. DeKeyser, L. Li. Matrix-assisted laser desorption/ionization Fourier transform mass spectrometry quantitation *via* in cell combination. Analyst, **2006**, *131*, 281-290.

[103] K.P. Braun, R.B. Cody, D.R. Jones, C.M. Peterson. A structural assignment for a stable acetaldehyde-lysine adduct. J. Biol. Chem., **1995**, *270*, 11263-11266.

[104] L.F. Fowles, E. Beck, S. Worrall, B. Shanley, J. de Jersey. The formation and stability of imidazolidinone adducts from acetaldehyde and model peptides. Biochem. Pharmacol., **1996**, *51*, 1259-1267.

[105] A.J. Heck, P.J. Bonnici, E. Breukink, D. Morris, M. Wills. Modification and inhibition of vancomycin group antibiotics by formaldehyde and acetaldehyde. Chem. Eur. J., **2001**, *7*, 910-916.

[106] X. Lou, A. J. H. Spiering, B. F. M. de Waal, J. L. J. van Dongen, J. A. J. M. Vekemans, E. W. Meijer. Dehydrogenation of tertiary amines in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, J. Mass Spectrom., **2008**, *43*, 1110-1122.

[107] K.Bashir, J.N. Whitaker. Handbook of multiple sclerosis. Lippincott Williams & Wilkins, **2003**.

[108] C. Lucchinetti, W. Brück, J. Parisi, B. Scheithauer, M. Rodriguez, H. Lassmann. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. Ann. Neurol., **2000**, *47*, 707-717.

[109] K. Whetten-Goldstein, F.A. Sloan, L.B. Goldstein, E.D. Kulas. A comprehensive assessment of the cost of multiple sclerosis in the United States. Mult. Scler., **1998**, *4*, 419-425.

[110] C. Miltenburger, G. Kobelt. Quality of life and cost of multiple sclerosis. Clin. Neurol. Neurosurg., **2002**, *104*, 272-275.

[111] P. H. Lalive, M. G. Hausler, H. Maurey, Y. Mikaeloff, M. Tardieu, H. Wiendl, M. Schroeter, H. P. Hartung, B. C. Kieseier, T. Menge. Highly reactive anti-myelin oligodendrocyte glycoprotein antibodies differentiate demyelinating diseases from viral encephalitis in children. Mult. Scler., **2011**, *3*, 297-302.

[112] C. H. Polman, S. C. Reingold, B. Banwell, M. Clanet, J. A. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos, F. D. Lublin, X. Montalban, P. O'Connor, M. Sandberg-Wollheim, A. J. Thompson, E. Waubant, B. Weinshenker, J. S. Wolinsky. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol., **2011**, *69*, 292-302.

[113] E. K. Mathey, T. Derfuss, M. K. Storch, K. R. Williams, K. Hales, D. R. Woolley, A. Al-Hayani, S. N. Davies, M. N. Rasband, T. Olsson, A. Moldenhauer, S. Velhin, R. Hohlfeld, E. Meinl, C. Linington. Neurofascin as a novel target for autoantibody-mediated axonal injury. J. Exp. Med., **2007**, *204*, 2363-2372.

[114] T. Derfuss, K. Parikh, S. Velhin, M. Braun, E. Mathey, M. Krumbholz, T. Kumpfel, A. Moldenhauer, C. Rader, P. Sonderegger, W. Pollmann, C. Tiefenthaller, J. Bauer, H. Lassmann, H. Wekerle, D. Karagogeos, R. Hohlfeld, C. Linington, E. Meinl. Contactin-2/TAG-1-directed autoimmunity is identified in multiple sclerosis patients and mediates gray matter pathology in animals. Proc. Natl. Acad. Sci. U. S. A., **2009**, *106*, 8302-8307.

[115] L. Lovato, R. Cianti, B. Gini, S. Marconi, L. Bianchi, A. Armini, E. Anghileri, F. Locatelli, F. Paoletti, D. Franciotta, L. Bini, B. Bonetti. Transketolase and 2',3'-cyclic-nucleotide 3'-phosphodiesterase type I isoforms are specifically recognized by IgG autoantibodies in multiple sclerosis patients. Mol. Cell. Proteomics., **2008**, *7*, 2337-2349.

[116] B. Bielekova, R. Martin. Development of biomarkers in multiple sclerosis. Brain., **2004**, *127*, 1463-1478.

[117] T. Berger, P. Rubner, F. Schautzer, R. Egg, H. Ulmer, I. Mayringer, E. Dilitz, F. Deisenhammer, M Reindl. Antimyelin antibodies as a predictor of clinically definite multiple sclerosis after a first demyelinating event. N Engl J Med., **2003**, *349*, 139-145.

[118] S. Gaertner, K.L. de Graaf, B. Greve, R. Weissert. Antibodies against glycosylated native MOG are elevated in patients with multiple sclerosis. Neurology, **2004**, *63*, 2381-2383.

[119] F. Lolli, B. Mulinacci, A. Carotenuto, B. Bonetti, G. Sabatino, B. Mazzanti, A. M. D'Ursi, E. Novellino, M. Pazzagli, L. Lovato, M. C. Alcaro, E. Peroni, M. C Pozo-Carrero, F. Nuti, L. Battistini, G. Borsellino, M. Chelli, P. Rovero, A. M. Papini. An N-glucosylated peptide detecting disease-specific autoantibodies, biomarkers of multiple sclerosis. Proc. Natl. Acad. Sci. U. S. A., **2005**, *102*, 10273-10278.

[120] H.A. Doyle, M.J. Mamula. Post-translational protein modifications in antigen recognition and autoimmunity. Trends Immunol., **2001**, *22*, 443-449.

[121] F. Lolli, B. Mazzanti, M. Pazzagli, E. Peroni, M. C. Alcaro, G. Sabatino, R. Lanzillo, V. Brescia Morra, L. Santoro, C. Gasperini, S. Galgani, M. M D'Elios., V. Zipoli, S. Sotgiu, M. Pugliatti, P. Rovero, M. Chelli, A. M. Papini. The glycopeptide CSF114(Glc) detects serum antibodies in multiple sclerosis. J. Neuroimmunol., **2005**, *167*, 131-137.

[122] J. Brettschneider, T. D. Jaskowski, H. Tumani, S. Abdul, D. Husebye, H. Seraj, H. R. Hill, E. Fire, L. Spector, J. Yarden, N. Dotan, J. W. Rose. Serum anti-GAGA4 IgM antibodies differentiate relapsing remitting and secondary progressive multiple sclerosis from primary progressive multiple sclerosis and other neurological diseases. J. Neuroimmunol., **2009**, *217*, 95-101.

[123] M. Schwarz, L. Spector, M. Gortler, O. Weisshaus, L. Glass-Marmor, A. Karni, N. Dotan, A. Miller. Serum anti-Glc(alpha1,4)Glc(alpha) antibodies as a biomarker for relapsing-remitting multiple sclerosis. J. Neurol. Sci., **2006**, *244*, 59-68.

[124] A. Carotenuto, A. M. D'Ursi, B. Mulinacci, I. Paolini, F. Lolli, A. M. Papini, E. Novellino, P. Rovero. Conformation-activity relationship of designed glycopeptides as synthetic probes for the

detection of autoantibodies, biomarkers of multiple sclerosis. J. Med. Chem., **2006**, *49*, 5072-5079.

[125] A. Carotenuto, M. C. Alcaro, M. R. Saviello, E. Peroni, F. Nuti, A. M. Papini, E. Novellino, P. Rovero. Designed glycopeptides with different beta-turn types as synthetic probes for the detection of autoantibodies as biomarkers of multiple sclerosis. J. Med. Chem., **2008**, *51*, 5304-5309.

[126] M. J. Walsh, J. M. Murray. Dual implication of 2',3'-cyclic nucleotide 3' phosphodiesterase as major autoantigen and C3 complement-binding protein in the pathogenesis of multiple sclerosis. J. Clin. Invest., **1998**, *101*, 1923-1931.

[127] K. G. Oikonomou, K. Zachou., G. N. Dalekos. Alpha-actinin: a multidisciplinary protein with important role in B-cell driven autoimmunity. Autoimmun. Rev., **2011**, *10*, 389-396.

[128] M. M. Reddy, R. Wilson, J. Wilson, S. Connell, A. Gocke, L. Hynan, D. German, T. Kodadek. Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. Cell, **2011**, *144*, 132–142.

[129]H. B. Larman, Z. Zhao, U. Laserson, M. Z. Li, A. Ciccia, M. A. Gakidis, G. M. Church, S. Kesari, E. M. Leproust, N. L. Solimini, S. J. Elledge. Autoantigen discovery with a synthetic human peptidome. Nature Biotech.,, **2011**, *29*, 535–541.

[130] J. E. Libbey, L. L. McCoy, R. S. Fujinami. Molecular mimicry in multiple sclerosis. Int. Rev. Neurobiol., **2007**, *79*, 127-147.

RINGRAZIAMENTI

Nel corso di questo percorso durato tre anni sono tante le persone che ho incontrato e tanti i contributi che ho ricevuto a livello umano e professionale, ciascuno dei quali è stato importante per rendere il mio dottorato un'esperienza significativa.

Vorrei, quindi, ringraziare tutto il laboratorio PeptLab di Firenze ed il Prof. Paolo Rovero, per avermi dato l'opportunità di svolgere questo dottorato e grazie alle mie insostituibili colleghe Giulia Pacini e Giada Rossi, per esserci state in ogni momento.

Un grande e caloroso ringraziamento va a tutto quanto il Laboratorio di Proteomica e Metabonomica di Roma diretto dal Prof. Andrea Urbani, per avermi gentilmente ospitato per tutto il mio primo anno di dottorato facendomi sentire parte del gruppo. In particolare ringrazio il Dr. Stefano Levi Mortera per essere stato una guida sapiente e paziente, le Dr.sse Valeria Marzano e Simona D'Aguanno per tutti gli insegnamenti che ho ricevuto e le Dr.sse, ed assolutamente amiche, Cristina Neri e Viviana Greco per i consigli e l'appoggio che non mi hanno mai fatto mancare.

Ringrazio, inoltre, la Prof.ssa Paola Migliorini ed il Dr. Federico Pratesi (Unità di Immuno-allergologia Clinica; Università di Pisa) per la stretta collaborazione nel progetto artrite reumatoide. Ringrazio anche il Centro di servizi di Spettrometria di Massa (CISM) dell'Università di Firenze, in particolare le Dr.sse Francesca Boscaro ed Elena Michelucci ed il Dr. Giuseppe Pieraccini, per aver fornito un importante supporto ad una parte della mia ricerca.

Tuttavia i ringraziamenti più sentiti vanno ai miei genitori, a mio fratello, ai miei nonni per avermi accompagnata con costante e forte sostegno in questa mia esperienza ed a Daniele, per avermi spronato ad esser fedele alle mie convinzioni, in ogni circostanza.