Biomarker search in rheumatic diseases through conventional and emerging proteomic technologies

Lucía González Rodríguez

Memoria presentada para optar al grado de doctor internacional

Tesis doctoral UDC 2019

Directores: Dr. Francisco Javier Blanco García y Dra. Cristina Ruiz Romero

Tutor: Dra. María del Carmen Veiga Barbazán

Programa de doctorado en Biotecnología Avanzada

Grupo de Investigación de Reumatología. Unidad de Proteómica. Instituto de Investigación Biomédica de A Coruña

The directors of this doctoral thesis, Dr. Francisco Javier Blanco García and Dr. Cristina Ruiz Romero,

CERTIFY THAT:

This work, entitled “Biomarker search in rheumatic diseases through conventional and emerging proteomic technologies”, was carried out by Lucía González Rodríguez, under our supervision at the Institute of Biomedical Research of A Coruña, and encompasses the requiring originality and scientific rigor conditions to qualify for international PhD in Advanced Biotechnology.

A Coruña, September 2nd, 2019

Dr. Francisco J. Blanco García

Dra. Cristina Ruiz Romero

UNIVERSIDADE DA CORUÑA

The tutor of this doctoral thesis, Dr. María del Carmen Veiga Barbazán,

CERTIFIES THAT:

This work, entitled “Biomarker search in rheumatic diseases through conventional and emerging proteomic technologies”, was carried out by Lucía González Rodríguez at the Institute of Biomedical Research of A Coruña, and encompasses the requiring originality and scientific rigor conditions to qualify for international PhD in Advanced Biotechnology.

A Coruña, September 2nd, 2019

Dra. María del Carmen Veiga Barbazán

Part of the research presented in this thesis has been carried out during three different national and international mobility stages under the supervision of:

- Anabel Marina Ramírez, Head of the Proteomics Facility at CBM-Severo Ochoa in Madrid, during a 1-month predoctoral research stay in this center in February 2017.

- Francesc Canals, Head of the Proteomics Facility at VHIO in Barcelona, during a 1-month predoctoral research stay in this center in November 2017.

- Christoph Borchers and David Schibli, Director and Associate Director of the UVic Genome BC Proteomics Centre in Victoria BC (Canada), during a 3-month international predoctoral research stay in this center from September until November 2018.

A mi familia,

AGRADECIMIENTOS

Después de estos cuatro años de investigación predoctoral, quiero agradecer de manera general a todas aquellas personas que, de una manera u otra, han participado y formado parte en esta trayectoria y desarrollo de tesis doctoral.

En primer lugar, me gustaría comenzar agradeciendo a mis directores de tesis, el Dr. Francisco Blanco y la Dra. Cristina Ruiz, por haberme dado la oportunidad y haber confiado en mí para realizar todo este trabajo, así como por haberme sabido guiar con su experiencia durante estos cuatro años. A mi tutora de tesis, Maria del Carmen Veiga, por su amabilidad y su gran dedicación a la hora de resolver todas las dudas y trámites durante mi etapa predoctoral, así como por haber aceptado ser mi tutora.

En especial, me gustaría agradecer a mis compañeras de grupo de proteómica por todo el apoyo brindado desde el primer momento. A Valentina, por haberme enseñado todo lo que sabe y ser un ejemplo a seguir, por tu implicación en el grupo y tu manera de ver y hacer las cosas más fáciles. A Patri, gracias por todos los momentos que hemos pasado juntas, que no han sido pocos, de los más tensos en la QTRAP a los más graciosos en Cádiz (jeje). A Rocío, gracias por tu alegría y tu ayuda incondicional, eres una gran compañera y ejemplo de trabajo en equipo, espero que todo te vaya bien en esta etapa predoctoral. A Lucía, gracias por acogerme desde el primer momento y apoyarme, eres un gran ejemplo de perseverancia y una magnífica compañera, por ello sé que conseguirás siempre lo que te propongas. A Bea, gracias por tu alegría, compañerismo, templanza, en definitiva, gracias por ser como eres y mostrarnos tu positividad ante los problemas. Ha sido un placer trabajar con todas vosotras en esta etapa y os doy las gracias por haberme aguantado también en mis malos momentos.

A todos los compañeros del INIBIC, en especial a los del grupo "Coffee", y al grupo de Biología Molecular de la UDC, gracias por haberme acogido y ayudado siempre en todo lo que he necesitado.

A Anabel Marina, Carlos García, Francesc Canals, Nuria Colomé, Angela Jackson, David Schibli, Bjorn Frohlich y Christoph Borchers por haberme aceptado, guiado y ayudado durante mis estancias de movilidad en Madrid, Barcelona y Victoria (Canadá), las cuales me han ayudado a crecer como investigadora. Sois gente maravillosa tanto profesionalmente como personalmente, lo cual hace que todo haya resultado ser más fácil con vuestra ayuda y amplia experiencia.

Por último, pero no menos importante, quiero dar las gracias a mi familia, mi hermana Alba, Yaya, mis abuelos, ... por haber estado siempre ahí y apoyarme en todo. No olvidando a los pequeños de la familia: a Luna, mi gran compañera incondicional, de la que tantas cosas he aprendido, sigues presente en cada momento, siendo este uno de ellos; a Chase, por hacerme reír cada día y animarme en los peores momentos. A Pablo, gracias por aguantarme y apoyarme durante todo este tiempo juntos, por cuidarme día a día, intentar sorprenderme siempre y conseguir que todos estemos unidos, en familia. Pero en especial, a mis padres, porque siempre me han ayudado en todo lo que han podido y sé que cada mérito alcanzado es un orgullo para ellos, pero también para mí, ya que sin su apoyo y esfuerzo hubiese sido más difícil poder realizar todo aquello que me he propuesto. Y por ello, es a ellos a quienes le dedico especialmente esta tesis, GRACIAS.

ABSTRACT / RESUMEN / RESUMO

ABSTRACT

There are more than 200 rheumatic diseases, which are the second leading cause of incapacitation worldwide and affect about 2 billion people, mainly women. This thesis focuses especially on two of the most common: osteoarthritis (OA) and rheumatoid arthritis (RA). Despite the differences between them, they all produce similar symptoms (joint pain, stiffness and even loss of functionality), although their current diagnosis and treatment are limited. The standard diagnosis includes the use of imaging techniques, which can be harmful to the patient, and the analysis of non-specific markers. OA treatment is aimed at alleviating pain and improving the functionality of the joints, although joint replacement may be even necessary at the last stages of the disease; while RA treatment includes various therapeutic strategies through trial-error. Due to these limitations in diagnosis and treatment, there is a great need to find biomarkers that facilitate the understanding of the pathological processes that take place, in order to better specify the diagnosis and effective treatment for the management of these diseases.

Particularly, in this thesis, different proteomic studies based on mass spectrometry (MS) and immunoassays have been carried out, with the aim of finding proteins with potential biomarker value in RA and OA, which could facilitate classification, diagnosis and subsequent selection of effective therapies.

In the first study, the search for a biomarker protein panel has been carried out for monitoring disease activity in RA patients, one of the major limitations for the corresponding selection of effective and specific therapies that could help to remit the characteristic acute episodes of this pathology. Initially, a panel of 11 candidate proteins has been identified by MS and relative quantification with iTRAQ labeling in depleted plasma pools of RA patients with extreme disease activities. In order to confirm the trend of these 11 proteins, 80 individual plasma samples of RA patients with different disease activities have

been analyzed by targeted MS, with the aid of standard labeled peptides, of which 5 have been verified as potential biomarkers. Finally, 4 of them (SAA1, AACT, HPT and A1AG1) have been validated by ELISA in 420 plasma samples from patients with RA, healthy donors and controls of other analogous diseases. Hence, this validated panel could be useful for the diagnosis of autoimmune inflammatory rheumatic diseases such as RA and disease activity monitoring, crucial for proper follow-up and treatment.

In the second and third studies, two different types of more emerging proteomic techniques have been carried out in order to find biomarkers in OA. On the one hand, two independent peptide studies have been carried out by means of MS in cartilage secretomes and synovial fluid and serum samples from OA patients and healthy donors. Peptidomics can be of great interest in OA, since the deterioration of cartilage is linked to various processes, among which, the action of proteases that produce neopeptides with possible biomarker role for OA diagnosis or monitoring stands out. Thus, 8 PRELP, CLUS, CILP1, COMP and MGP neopeptides were significantly altered in cartilage secretomes of OA patients, whose release seems to be also dependent on the affected joint. In addition, 6 neopeptides belonging to APOA4, ITIH4, CO3 and KNG1 have been altered in serum pools of OA patients and healthy donors, although no confirmatory statistical analysis has been performed yet. On the other hand, two hybrid immunoaffinity-based techniques coupled to mass spectrometry (IA-MS) have been developed for the detection of two specific cartilage proteins (CILP1 and PRG4) in serum samples. An immuno-MALDI (iMALDI) assay for the detection of PRG4 and another of SISCAPA-MRM for the analysis of CILP1 have been developed. Although the tendency of these proteins is to be increased in OA patients versus healthy controls, this alteration has not become significant yet in the analyzed small cohort of serum samples.

RESUMEN

Existen más de 200 enfermedades reumáticas, las cuales son la segunda causa de incapacitación a nivel mundial y afectan a cerca de 2 billones de personas, fundamentalmente a mujeres. Esta tesis se centra especialmente en dos de las más comunes: artrosis (OA) y artritis reumatoide (RA). A pesar de las diferencias que existen entre ellas, todas producen síntomas similares (dolor, rigidez y pérdida de funcionalidad articular), aunque sus actuales diagnósticos y tratamientos son limitados. El diagnóstico habitual incluye el uso de técnicas de imagen, las cuales pueden resultar dañinas para el paciente, y el análisis de marcadores no-específicos. El tratamiento de la OA está dirigido a paliar el dolor y mejorar la funcionalidad de las articulaciones, llegando incluso a ser necesario el reemplazo articular; mientras que el de la RA incluye diversas estrategias terapéuticas mediante ensayo-error. Debido a estas limitaciones, hay una gran necesidad de encontrar biomarcadores que faciliten el entendimiento de los procesos patológicos que tienen lugar con el fin de definir mejor el diagnóstico y tratamiento efectivos para el manejo de estas enfermedades.

Particularmente, en esta tesis, se han realizado diferentes estudios proteómicos basados en espectrometría de masas (MS) e inmunoensayos, con el fin de encontrar proteínas con potencial papel biomarcador en RA y OA y así facilitar su clasificación, diagnóstico y posterior selección de terapias efectivas.

En el primer estudio, se ha llevado a cabo la búsqueda de un panel de proteínas biomarcadoras para monitorizar la actividad de la enfermedad en pacientes con RA, una de las mayores limitaciones para la adecuada selección de terapias efectivas y específicas que ayuden a remitir los episodios agudos característicos de esta patología. Inicialmente, se ha identificado mediante MS y cuantificación relativa con marcaje iTRAQ, un panel de 11 proteínas candidatas en muestras de plasma agrupadas y deplecionadas de pacientes RA con actividades de enfermedad extremas. Con objeto de confirmar la tendencia de

estas 11 proteínas, 80 muestras de plasma individuales de pacientes RA con diferentes actividades de enfermedad han sido analizadas mediante MS dirigida, con ayuda de péptidos estándar marcados, de las cuales 5 han sido verificadas como potenciales biomarcadores. Finalmente, 4 de ellas (SAA1, AACT, HPT y A1AG1) han sido validadas mediante ELISA en 420 muestras de plasma de pacientes con RA, donantes sanos y controles de otras enfermedades análogas. Así, este panel validado podría ser útil para el diagnóstico de enfermedades reumáticas inflamatorias autoinmunes como la RA y la monitorización de actividad de enfermedad, crucial para el adecuado seguimiento y tratamiento.

En el segundo y tercer estudios, se han llevado a cabo dos tipos diferentes de técnicas proteómicas más emergentes con el fin de encontrar biomarcadores en OA. Por un lado, se han realizado dos estudios peptidómicos independientes mediante MS en secretomas de cartílago y en muestras de líquido sinovial y suero de pacientes OA y donantes sanos. La peptidómica puede ser de gran interés en la OA, ya que el deterioro del cartílago está ligado a diversos procesos, entre los que destaca la acción de las proteasas que producen neopéptidos con posible papel biomarcador de diagnóstico o monitorización de la OA. Así, 8 neopéptidos de PRELP, CLUS, CILP1, COMP y MGP se vieron significativamente alterados en secretomas de cartílago de pacientes OA, cuya liberación parece ser dependiente de la articulación afectada. Además, 6 neopéptidos pertenecientes a APOA4, ITIH4, CO3 y KNG1 se han visto alterados en muestras agrupadas de suero de pacientes OA y donantes sanos, aunque en este caso no se han realizado análisis estadísticos confirmatorios. Por otro lado, se han puesto a punto dos técnicas híbridas de inmunoafinidad acopladas a espectrometría de masas (IA-MS) para la detección de dos proteínas específicas de cartílago (CILP1 y PRG4) en muestras de suero. Se ha desarrollado un ensayo iMALDI para la detección de PRG4 y otro de SISCAPA-MRM para el análisis de CILP1. Aunque la tendencia de estas proteínas se ve aumentada en pacientes OA frente a controles sanos, esta alteración no llega a ser significativa en la pequeña cohorte de muestras séricas analizadas.

RESUMO

Hai máis de 200 enfermidades reumáticas, que son a segunda causa de incapacitación a nivel mundial e afectan a preto de 2.000 millóns de persoas, principalmente mulleres. Esta tese céntrase especialmente en dúas das máis comúns: a artrose (OA) e a artrite reumatoide (RA). A pesar das diferenzas entre elas, todas producen síntomas similares (dor, rixidez e incluso perda de funcionalidade articular), aínda que os actuais diagnósticos e tratamentos son limitados. O diagnóstico estándar inclúe o uso de técnicas de imaxe, que poden ser prexudiciais para o paciente, e a análise de marcadores non específicos. O tratamento da OA está dirixido a aliviar a dor e mellorar a funcionalidade das articulacións, podendo incluso ser necesaria a substitución articular; mentres que o da RA inclúe varias estratexias terapéuticas mediante proba-erro. Debido a estas limitacións no diagnóstico e tratamento, hai unha gran necesidade de atopar biomarcadores que faciliten a comprensión dos procesos patolóxicos que teñen lugar co fin de especificar mellor o diagnóstico e o tratamento eficaz para o manexo destas enfermidades.

Particularmente, nesta tese, realizáronse diferentes estudos proteómicos baseados na espectrometría de masas (MS) e inmunoensaios, co fin de atopar proteínas con potencial papel biomarcador en RA e OA e facilitar así a súa clasificación, diagnóstico e posterior selección de terapias eficaces.

No primeiro estudo, levouse a cabo a busca dun panel de proteínas biomarcadoras para controlar a actividade da enfermidade en pacientes con RA, unha das principais limitacións para a correspondente selección de terapias eficaces e específicas que axudan a remitir os característicos episodios agudos desta patoloxía. Inicialmente, un grupo de 11 proteínas candidatas foi identificado por MS e cuantificación relativa co etiquetado de iTRAQ, en mostras de plasma agrupadas e deplecionadas de pacientes con RA con actividades de enfermidade extrema. Para confirmar a tendencia destas 11 proteínas, analizáronse 80 mostras

plasmáticas individuais de pacientes con RA con diferentes actividades de enfermidade mediante MS dirixida, coa axuda de péptidos marcados estándar, das que 5 verificáronse como biomarcadores potenciais. Finalmente, 4 delas (SAA1, AACT, HPT e A1AG1) foron validadas por ELISA en 420 mostras de plasma de pacientes con RA, doantes sans e controis doutras enfermidades análogas. Así, este panel validado podería ser útil para o diagnóstico de enfermidades reumáticas inflamatorias autoinmunes como a RA e o seguimento da actividade da enfermidade, crucial para un seguimento e tratamento adecuados.

No segundo e terceiro estudos, leváronse a cabo dous tipos diferentes de técnicas proteómicas máis emerxentes para atopar biomarcadores na OA. Por unha banda, leváronse a cabo dous estudos peptidómicos independentes mediante MS en secretomas de cartilaxe e en mostras de líquido sinovial e soro de pacientes con OA e doantes sans. A peptidómica pode ser de grande interese na OA, xa que a deterioración da cartilaxe está ligada a diversos procesos, entre os que destaca a acción das proteasas que producen neopéptidos con posible papel biomarcador para o diagnóstico ou o seguimento da OA. Así, 8 neopéptidos de PRELP, CLUS, CILP1, COMP e MGP víronse significativamente alterados nos secretomas de cartilaxe de pacientes con OA, cuxa liberación parece depender tamén da articulación afectada. Ademais, 6 neopéptidos pertencentes a APOA4, ITIH4, CO3 e KNG1 mostráronse alterados en mostras séricas agrupadas de pacientes con OA e doantes sans, aínda que neste caso non se realizaron análises estatísticas confirmativas. Doutra banda, desenvolvéronse dúas técnicas híbridas de inmunoafinidade acopladas a espectrometría de masas (IA-MS) para a detección de dúas proteínas de cartilaxe específicas (CILP1 e PRG4) en mostras de soro. Elaborouse un ensaio inmuno-MALDI (iMALDI) para a detección de PRG4 e outro de SISCAPA-MRM para a análise de CILP1. Aínda que a tendencia destas proteínas vese aumentada en pacientes con OA fronte a controis saudables, esta alteración non se fai significativa na pequena cohorte de mostras de soro analizadas.

OUTLINE

Abbreviations

ACN: acetonitrile.

CILP: Cartilage Intermediate Layer Protein.

COMP: Cartilage Oligomeric Matrix Protein.

ECM: Extracellular Matrix.

ESI: Electrospray Ionization.

FA: Formic Acid.

HSA: Human Serum Albumin.

IA: Immunoaffinity.

iMALDI: immuno- Matrix-Assisted Laser Desorption/Ionization.

IMIDs: Immune-Mediated Inflammatory Diseases.

iTRAQ: Isobaric Tags for Relative and Absolute Quantitation.

MALDI: Matrix-Assisted Laser Desorption/Ionization.

MGP: Matrix Gla Protein.

MMPs: Matrix Metalloproteinases.

MRM: Multiple Reaction Monitoring.

MS: Mass Spectrometry.

OA: Osteoarthritis.

PRELP: Prolargin.

PRG4: proteoglycan-4, superficial zone protein or lubricin.

PsA: Psoriatic Arthritis.

RA: Rheumatoid Arthritis.

RMDs: Rheumatic and Musculoskeletal Diseases.

SISCAPA: Stable Isotope Standards and Capture by Anti-Peptide Antibodies.

SF: Synovial Fluid.

SLE: Systemic Lupus Erythematosus.

TCA: Trichloroacetic Acid

TFA: Trifluoroacetic Acid.

UF: Ultrafiltration.

Figure index.

Table index.

1. Introduction

1. The human joint.

The adult human body is composed of 206 bones. Each bone is connected to at least another bone, with the exception of the hyoid bone (neck). The location where bones come together is called the joint, which allows for movement between the bones. Joints are classified both structurally (based on the nature of the material that separates the bones) and functionally (in function of their movement). Thus, fibrous joints (poor movement), cartilaginous joints (semi-mobile) and synovial joints (high mobility) can be distinguished.

Synovial joints are the most common type of joint in the body, characterized by the presence of a joint cavity. This cavity is surrounded by an external articular capsule of fibrous connective tissue, which protects the joint. Lining the inner articular capsule surface is a thin synovial membrane, whose cells secrete synovial fluid (SF). SF is a thick, slimy fluid that provides lubrication (which reduces bone friction) and nourishes the articular cartilage as well. The articular cartilage is one of the most important joint components as it covers the bone therefore cushioning pressure forces on the joints and preventing direct bone friction (**Figure 1**). Outside of their articulating surfaces, the bones are also connected by ligaments, muscles and tendons, which strengthen and support the joint (by anchoring the bones together and preventing their separation), allow the sliding of joint surfaces and help stabilize movement (1).

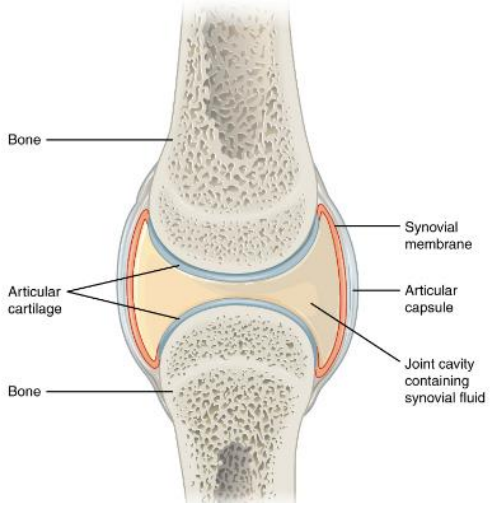


Figure 1. Synovial joint components (1).

1.1. Articular cartilage composition.

Articular cartilage is a 2-4 mm thin layer of specialized connective tissue with singular viscoelastic properties, which provides a smooth, lubricated surface yielding low-friction articulation and facilitates the loads transmission to the subjacent subchondral bone (SB). It is composed of hyaline cartilage, the most abundant cartilage type in the body responsible for bone formation and usually found also in trachea and costal cartilage. Moreover, the articular cartilage is mainly constituted by a dense, porous and permeable extracellular matrix (ECM) and chondrocytes (around 2% of the total cartilage volume), although it is absent of blood vessels, nerves or lymphatics, unlike other tissues. Chondrocytes are highly specialized, metabolically active cells, responsible for cartilage ECM development, maintenance and repair by producing enzymes, growth factors, and inflammatory mediators. The ECM is primarily composed of water, collagens (around 70% dry weight, mostly type II collagen) and proteoglycans or PRGs (20-30% dry weight), such as aggrecan, decorin, biglycan, fibromodulin, lumican, perlecan and superficial zone protein (also called lubricin or PRG4). Although other structural and/or regulatory non-collagenous proteins and glycoproteins are present in lesser amount. Fibronectin (FINC), Cartilage Intermediate Layer Protein (CILP) or Cartilage Oligomeric Matrix Protein (COMP) are structural non-collagenous proteins, while gp-39/YKL-40, Matrix Gla Protein (MGP), Chondromodulins (I and II) and Transforming Growth Factors α and β are considered to have metabolic regulatory roles (2, 3).

Despite its simple appearance, cartilage is a high heterogeneous tissue, which exhibits different compositions depending on its depth (**Figure 2**). Thus, four zones are differentiated in the articular cartilage: the superficial zone (SZ), the middle zone (MZ), the deep zone (DZ) and the calcified zone (CZ). The outer protective SZ makes up approximately 10–20% of the cartilage thickness, which is in direct contact with synovial fluid. Immediately deep, the MZ represents the

next 40–60%, the bottom 30–40% is the DZ, and finally the CZ, in direct contact with the SB. Chondrocyte distribution and morphology differ in the different zones, as well as the collagen fibers organization. These fibers originate from the CZ perpendicularly to the joint surface, and then change their orientation in the MZ to become parallel to the articular surface in the SZ. This particular disposition, provides the tissue with unique biomechanical characteristics, enabling the cartilage to resist the sheer, tensile, and compressive articulation forces. Moreover, the secretion and prevalence of the different cartilage proteins also differ among zones, likely due to their zone-specific functionality (2-4).

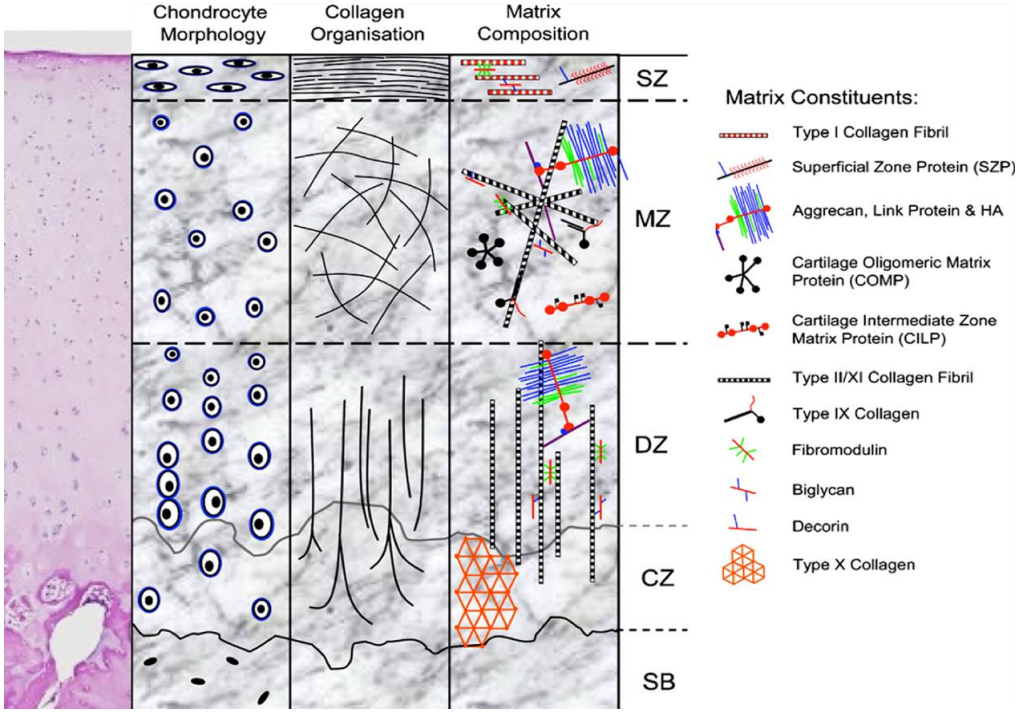


Figure 2. Schematic representation of cartilage morphology, structure and composition (4).

2. Rheumatic Diseases.

"Rheumatism" refers to various painful medical conditions which mainly affect joints, bones, cartilage, tendons, ligaments and muscles. Rheumatic and musculoskeletal diseases (RMDs) are diverse and complex affections of non-traumatic origin studied in the rheumatology medical field. RMDs are generally characterized by persistent joint pain, inflammation (joint swelling, stiffness, redness, and/or warmth) and a consequent reduction of range motion, flexibility and function in one or more areas of the musculoskeletal system. However, other symptoms such as tenderness, joint deformity, extreme fatigue, lack of energy and weakness may also arise.

RMDs can also affect other systems or internal organs such as heart, lungs, eyes, skin, blood vessels and other connective tissues (**Figure 3**), hence its diversity and complexity (5). Thus, more than 200 different RMDs are reported, although Osteoarthritis (OA), Rheumatoid arthritis (RA), Systemic Lupus Erythematosus (SLE), Ankylosing Spondylitis (AS), Psoriatic Arthritis (PsA), Systemic Sclerosis (SSc), Sjogren's Syndrome, and Gout are among the most common ones. Despite all the differences that exist between them, symptoms as pain, joint stiffness and functional capacity loss are prevalent in all of them. For instance, RA is characterized by joint inflammation due to autoimmune processes, whereas OA leads to wear and joint degeneration without need for inflammation to occur. In contrast, osteoporosis is characterized

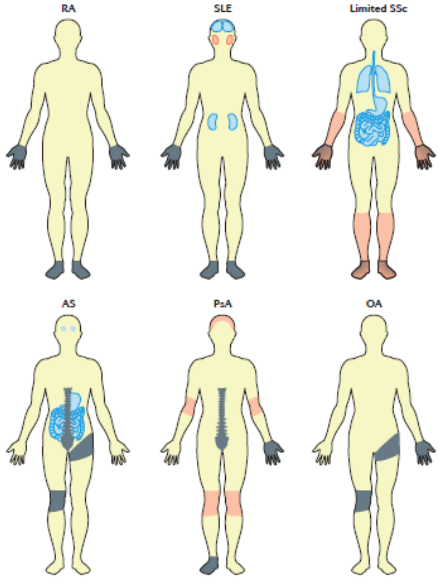


Figure 3. Joint (grey), skin (red) and organ (blue) regions affected in different RMDs (5).

by bone mass density loss and porosity increment, ultimately increasing the risk of fracture.

RMDs were reported as the second global leading cause of disability with around 2 billion people affected worldwide (6). Moreover, most of these affections are long term or chronic and worsen over time. In severe cases, they can even result in significant or complete disability. RMDs have a considerable impact on life quality and expectancy and impose huge financial costs. In Europe alone, RMDs are associated with an economic burden of over 200 billion euros per year, becoming the most expensive diseases for European health care systems. Besides, RMDs affect all ages and both genders, although women are more frequently affected than men. The risk of developing a rheumatic disease is increased due to different factors as smoking, obesity, genetics, age, occupation and socioeconomic background (7).

Clinical RMDs diagnosis are typically assessed by blood tests (e.g. erythrocyte sedimentation rate or ESR, C-reactive protein or CRP, anti-nuclear antibodies or ANAs, rheumatoid factor or RF) and diverse imaging approaches. However, a preferred RMDs diagnosis method was not reported to date and the clinical tools used are dependent on the primary care providers' own knowledge. Moreover, most of the available effective treatments (mainly disease modifying drugs and/or biological therapies) only help to manage pain and control arthritis symptoms, and the patient response may become rather variable and particular.

2.1. Osteoarthritis.

Osteoarthritis (OA) is the clinical and pathological result of a variety of disorders that ultimately produces structural and functional failure of synovial joints. OA has been initially considered a disease of the articular cartilage, but recent studies have reported the involvement of the entire joint organ (8).

Although the articular cartilage loss seems to be the primary change, a combination of cellular and biomechanical stresses causes several secondary changes, including SB remodeling, osteophytes formation, bone marrow lesions development, change in the synovium, joint capsule, ligaments, periarticular muscles, and meniscal tears and extrusion (9).

2.1.1. Pathophysiology.

OA is a complex multifactorial process in which mechanical circumstances have a central role, although inflammatory, and metabolic components are also remarkably involved (**Figure 4**). During the OA process, cartilage suffers composition changes consequently losing its integrity and increasing its susceptibility to disruption by physical forces. Initially, erosions are only at the SZ, then deeper fissures are expanded to the CZ. Bone turnover and vascularity are increased in the SB, likely associated with the development of SB marrow lesions. The osteophytes are generated at the joint margins by endochondral ossification, probably due to inflammatory biological factors, overload and/or abnormal joint kinematics. Proliferating synoviocytes release proinflammatory molecules, increasing tissue hypertrophy and vascularity. Thus, the usual cartilage ECM turnover is mediated by hypertrophic chondrocytes which, in an attempt at repair, increase their synthetic activity and release proinflammatory and ECM degradative enzymes. Then, any injury or degenerative process which could disturb chondrocyte activity, may lead to failure on maintaining the homeostatic balance between ECM synthesis and degradation. Hence, OA is an active and dynamic alteration caused by an imbalance between the joint repair and destruction, not a passive degenerative disorder or so-called and commonly described wear-and-tear disease (9-11).

The primary pro-inflammatory proteinases involved in ECM degradation include: matrix metalloproteinases or MMPs (such as collagenases, gelatinases

and stromelysins), a disintegrin and metalloproteinase with thrombospondin motifs or ADAMTS, elastases, calpains, caspases and cathepsins (such as cathepsin B and D). Particularly, collagenases MMP-13, MMP-1 and MMP-14; gelatinases MMP-2 and MMP-9, stromelysin MMP-3; and aggrecanases ADAMTS-4 and ADAMTS-5 were found to degrade cartilage ECM collagens and/or proteoglycans (2, 3, 12-14). Moreover, the role of inflammation mediators is also noteworthy and different cytokines, such as interleukins (ILs) and/or tumor necrosis factor alpha (TNF-α), stimulate the production of the previous matrix proteases and induce proinflammatory and catabolic pathways.

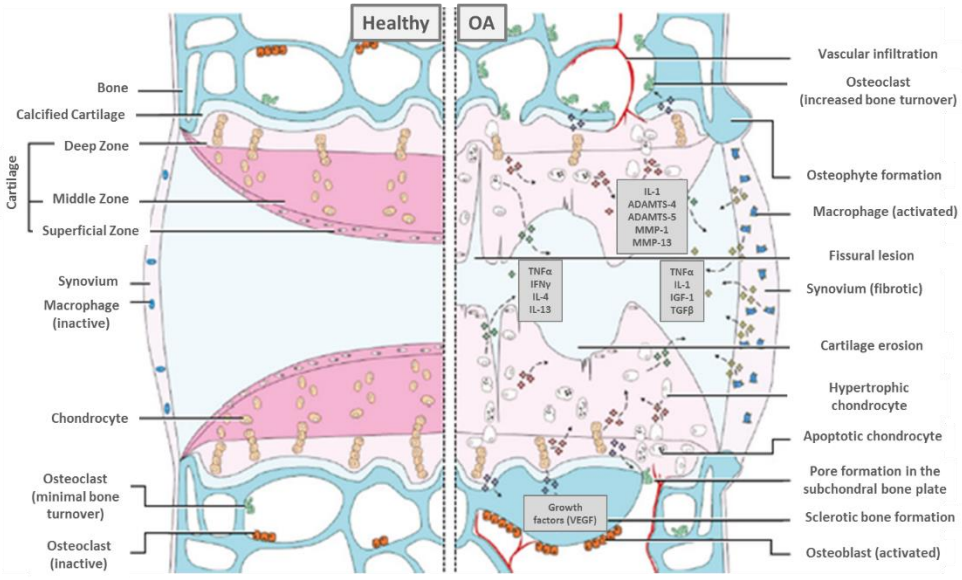


Figure 4. Major factors involved in OA pathogenesis compared to healthy joint (10).

2.1.2. Prevalence, burden and risk factors.

OA is the most common chronic RMD which affects around 250 million people, 15% of the world population (9, 10). Forecasts anticipate that around 29% of the adult population will suffer OA by 2032 and that OA will be a major cause of morbidity and physical limitation among people over 40 years old (10, 11). Besides, individuals over age 65 have higher probabilities of suffering OA.

Clinically, knee is the most common OA location (**Figure 5**), implying the 85% OA burden worldwide, followed by hand and hip (10). Moreover, women are more likely to suffer OA than men, as well as a more severe condition (15).

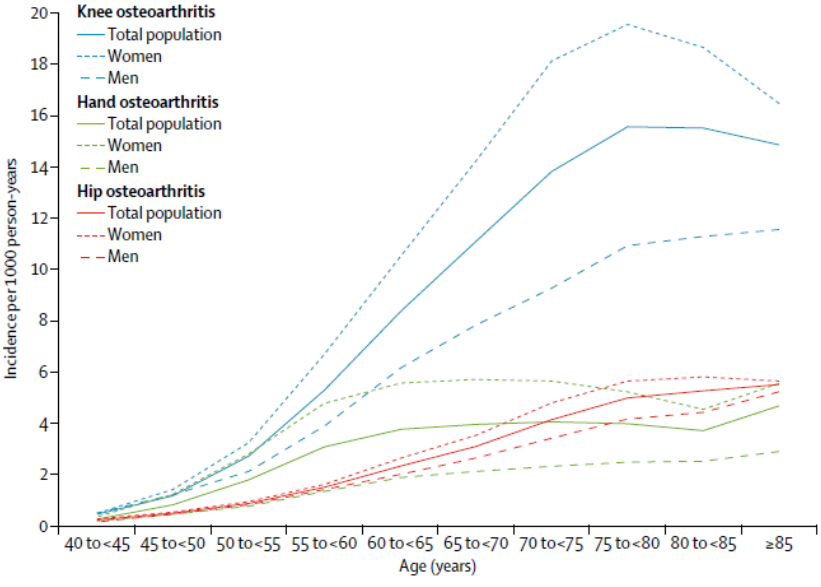


Figure 5. OA incidence depending on the OA location and affected gender (10).

The OA burden entails direct and indirect costs, as well as other intangible costs to the individual such as pain, activity limitation, decreased quality of life, fatigue or reduced social participation. On the one hand, the direct OA medical cost in various high-income countries has been estimated to account for 1-2.5% of their gross domestic product (GDP). Particularly, joint replacements and the consequent hospital stays represent the major proportion of these health-care costs. On the other hand, indirect costs are challengingly estimated to be up to eightfold greater than the direct costs, therefore underestimating the real OA burden. The major indirect cost is mainly attributed to productivity loss and specifically due to a reduced employment rate (16).

Several systemic and local risk factors contribute to OA etiology. Their influence may vary attending to the affected joint(s) and the disease stage,

development, progression and diagnosis. Thus, age, gender, race, genetics, congenital/developmental conditions, bone density, smoking and diet are among the most representative systemic factors. Particularly, age is considered one of the strongest and most common factors, females are more likely to suffer OA, genetics contribute around 40-80% and low selenium and vitamins C, D and K intake are associated with irregular and decreased bone and cartilage development and integrity. Local risk factors include obesity/overweight, injury/surgery, occupation, high-impact physical activities/sports, mechanical factors, joint alignment/shape and laxity. Specially, these local factors are likely related with joint hit, joint overload, uneven articulation load, joint overuse, muscle weakness and/or atrophy. Moreover, each of these OA risk factors might instigate different pathogenic OA pathways, so distinct stratified OA components have been proposed on this basis: inflammatory, mechanical overload, metabolic and cell senescence (10, 11, 15, 17).

2.1.3. Diagnosis and treatment.

The current OA diagnosis is usually based on the American College of Rheumatology (ACR) established criteria, which involves the assessment of different clinical manifestations, biological parameters and imaging evidences (18). The main OA clinical manifestations include pain, stiffness, reduced movement, swelling, crepitus and increased age (unusual before age 40) in the absence of systemic features (such as fever). ESR and CRP levels are usually analyzed as markers of inflammation although they are not specific for OA diagnosis. In advanced OA stages, joint space narrowing (JSN), osteophytes and sometimes SB changes can be seen in plain radiographs or by magnetic resonance imaging (MRI) (8, 19). Thus, the Kellgren and Lawrence system has been commonly used for radiologically assessing the OA severity using five grades (0, 1, 2, 3, 4), where zero represents the absence of radiographic OA features and 4 corresponds to the presence of large and numerous osteophytes, marked JSN,

severe sclerosis and definite bone deformity (20). However, although radiography is the most common used technique for OA diagnosis, its frequent use can be harmful to the patient. Moreover, it is not a very sensitive technique for detecting incipient changes and the visualization of soft tissues such as cartilage is not possible, making it difficult to evaluate OA progression and cartilage destruction. In contrast, the MRI technique is more sensitive and specific so minimal cartilage structural changes can be detected, although it is very expensive and difficult to apply in the clinic (21, 22).

OA treatment choices can be divided into four main categories: nonpharmacologic, pharmacologic, complementary and alternative, and surgical. In general, treatment should begin with the safest and least invasive remedies before proceeding to more invasive and expensive therapies (**Figure 6**). However, apart from pain reduction management, joint function improvement and surgical intervention at the end OA stage, there are not current cures or effective therapeutic treatments for OA (11, 19).

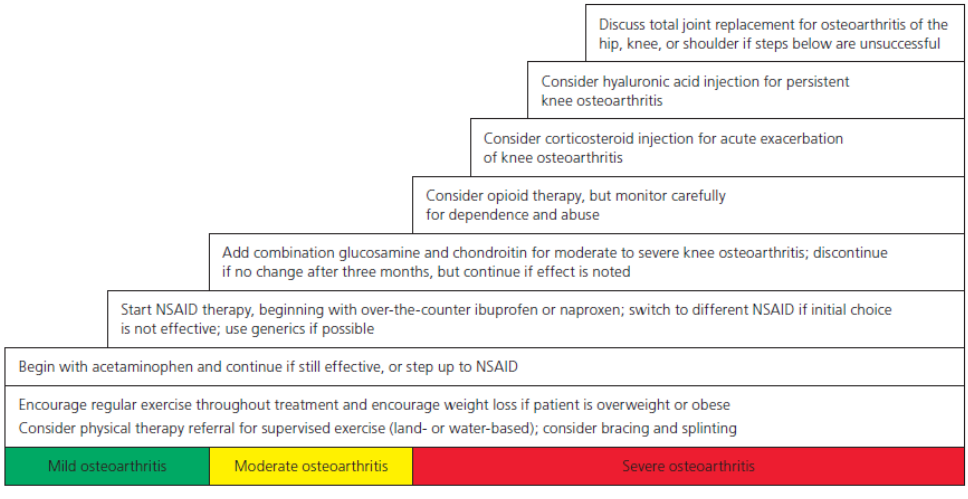


Figure 6. Recommended stepped-care approach for OA treatment (19).

Hence, there is an unmet clinical need for more precise and accessible OA monitoring and diagnosis tools which could help to better understand the OA

etiology and discover more effective OA treatments. In general terms, the main OA diagnosis limitation is likely due to the unknown, clinically silent duration of the disease, where metabolic alterations have already occurred at the tissue level. Consequently, patients are normally diagnosed in advanced OA stages, when it is already difficult to recover the functional capacity of the affected individual. All limitations and needs considered, there is an increased interest for finding biological markers that reflect articulation remodeling changes that allow us to make an early diagnosis of the disease, evaluate its prognosis and monitor alternative and personalized therapies for each patient (23).

2.2. Rheumatoid Arthritis.

Rheumatoid arthritis (RA) is a long-term chronic inflammatory autoimmune disorder, primarily characterized by persistent joint inflammation, ultimately leading to cartilage and bone damage, deformity and disability, as well as other possible systemic complications (24, 25). Thus, RA can also affect extra-articular organs and systems, such as skin, lung, eyes, gastrointestinal tract and cardiovascular system (24, 26).

2.2.1. Pathophysiology.

The exact cause of RA is still unknown, although several autoimmune and inflammatory processes, involving numerous cellular components, soluble mediators, adhesion molecules and autoantibodies (such as RF and ACPAs, markers of autoimmune dysfunction and RA erosion respectively), contribute to synovitis, swelling, inflammation and unfavorable structural changes in joints and internal organs. The leading hypothesis is that RA pathogenesis results from a complex interaction between genetic and environmental factors, which induce an aberrant activation of the innate and adaptive immune system, finally leading to RA symptoms and establishment of the disease (**Figure 7**) (27, 28).

Figure 7. RA development phases and the corresponding factors/features involved in each stage (28).

The process mainly involves immune tolerance failure, inflammatory cytokine and ECM protease production (MMP), macrophage differentiation into osteoclasts (which later resorb and destroy bone) and the adverse autoantigen appearance with antigen-specific T and B cells activation (**Figure 8**). T and B cells appertain to the adaptive immune system, while cells from the innate immune system (such as macrophages, monocytes, dendritic cells, neutrophils, natural killer cells and/or mast cells among others) activate the phagocytosis process which produces the release of specific inflammatory cytokines. The cascade of these events ultimately leads to hyperplastic inflammation of the synovial lining, as well as cartilage and SB destruction (24, 29-31).

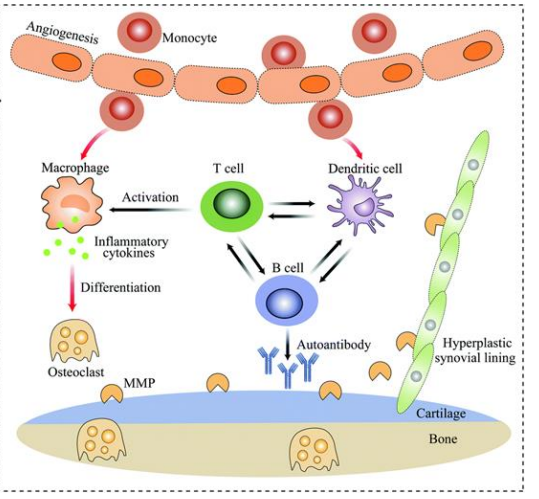


Figure 8. Multifactorial RA pathogenesis (31).

2.2.2. Prevalence, burden and risk factors.

RA is also one of the most common RMD, which affects around 0.5-1% of the population worldwide. Approximately 75% of individuals who develop RA are

women. Although RA can occur in patients at any age, incidence was found to rise with age, especially in patients around 60. RA entails a considerable individual and socioeconomic burden. The individual burden mainly arises from musculoskeletal alterations, which parallelly cause decreased physical function and life quality, and could also result in enhanced comorbid and mortal risks. The socioeconomic burden includes the leading direct medical costs, although indirect costs (primarily derived from functional and work disability, and/or lowered social participation) also account for an extensive load (30, 32, 33).

The main RA risk factors include gender, genetics and environmental factors. On the one hand, genetics contribute to 40-60% of the total RA risk, which is increased 2-to-10-fold in first-degree relatives. Thus, more than one hundred genetic loci have been associated with RA risk, most of which implicate immune mechanisms, while some of them also contribute to other chronic inflammatory diseases. Particularly, alleles of human leukocyte antigens (HLA) represent around 40% of the genetic RA risk. These alleles contain similar amino acid sequences, the so-called shared-epitope (SE), which are known to highly contribute to RA pathogenesis. Cigarette smoking, obesity and diet, microbiota and infections, occupation and atmospheric agents (such as air pollution and/or exposure to pesticides, herbicides and insecticides) are among the most representative environmental factors. The strongest contribution to the environmental RA risk is the exposure to smoke, which accounts for 20–30% of the risk. Hence, female sex, exposure to tobacco smoke, occupational dust (silica), air pollution, sodium, red meat and iron consumption, obesity and low vitamin D intake and levels, increase the risk of developing RA; whereas fish and omega 3 fatty acid consumption, moderate alcohol intake and healthy diet, decrease it (24-26, 29, 34).

2.2.3. Diagnosis and treatment.

RA patients generally present tender and swollen joints, morning joint stiffness, and elevated CRP, ESR, RF and ACPA levels. RA swelling is generally "soft" due to synovitis and effusion, while it is harder and bony in OA. However, these manifestations are not strictly specific from RA and can be also present in other types of arthritis or autoimmune conditions (35). RA is typically classified into two subtypes: seropositive and seronegative RA (30). Seropositive RA implies high serum levels of RF and ACPA autoantibodies and is associated with more severe joint damage, symptoms and increased mortality. Additionally, leukocyte and neutrophil cell count in SF obtained by arthrocentesis (joint aspiration), plain radiographs, ultrasonography and MRI are among the general diagnostic tools for RA. Moreover, the most recent RA classification criteria were established by a cooperative committee of the ACR and the European League Against Rheumatism (EULAR) in 2010 (36). The classification criteria resulted in a score from 0 to 10, which grade different parameters such as joint involvement, serology (RF, ACPA), acute-phase reactants (CRP, ESR) and symptoms duration. A patient with a score of 6 or more points can then be classified as suffering from RA. However, the classification goal is to maximize homogeneous populations for study purposes, whereas the aim of diagnosis is to correctly identify individual patients. Hence, further research is required in order to validate the previous classification criteria for diagnosis (30).

After RA classification, evaluation of physical function and assessment of disease activity are crucial for patient follow-up. Thus, physical function is typically evaluated with the Health Assessment Questionnaire Disability Index (HAQ-DI) at every clinical visit. Additionally, combined measures defining remission and low, moderate or high disease activity have been developed such as the disease activity score using 28 joints (DAS28), the simplified disease activity index (SDAI) and/or the clinical disease activity index (CDAI) among others. The

cited scores generally include specific variables such as tender and swollen joint counts, acute phase reactant(s) levels (CRP or ESR) and/or patient/physician global assessment(s). These measures are generally required when using a treat-to-target (TTT) strategy in pharmacologic therapies, where treatments are adapted until remission (primarily for early RA) or low disease activity (likely in established RA) conditions are achieved (**Figure 9**). ACR and EULAR recently conjointly developed new remission criteria, based on a Boolean approach or on a strategy using the SDAI or CDAI criteria (37), since remission according to DAS28-ESR/CRP results in high frequency of false-positive responses, especially when drugs affecting the acute-phase response are used. There is an almost linear relationship between disease activity and physical function loss or damage progression, although there is a large variation in the resulting disease activity classification according to the various indices. Furthermore, about one third of the RA patients in remission have also shown radiographic evidence of joint damage progression and/or synovitis (30, 38-41).

Importantly, disease-modifying treatments should be initiated as soon as possible after RA diagnosis, in order to prevent irreversible structural damage and chronic disability. Although RA is incurable, the current managing therapeutic strategies, including individual and/or combined nonpharmacological and pharmacological treatments, allow excellent disease control. Nonpharmacologic management includes educating patients about RA. Pharmacological therapies involve the use of anti-inflammatory agents, such as nonsteroidal anti-inflammatory drugs (NSAIDs) and glucocorticoids for fighting against pain and inflammation; as well as synthetic and biologic disease-modifying antirheumatic drugs (DMARDs), which work as immunomodulatory agents that target cellular and/or molecular RA pathogenesis pathways. Generally, if low disease activity or remission is not attained at 6 months after a particular therapy is applied, other treatments should be re-evaluated (**Figure 9**). Clinical remission, meaning absence of significant inflammation signs with/without treatment, occurs in 20%

or less of the RA patients, whereas remission or low disease activity with continuing treatment may happen in up to 75% of RA patients. (30, 38-40, 42).

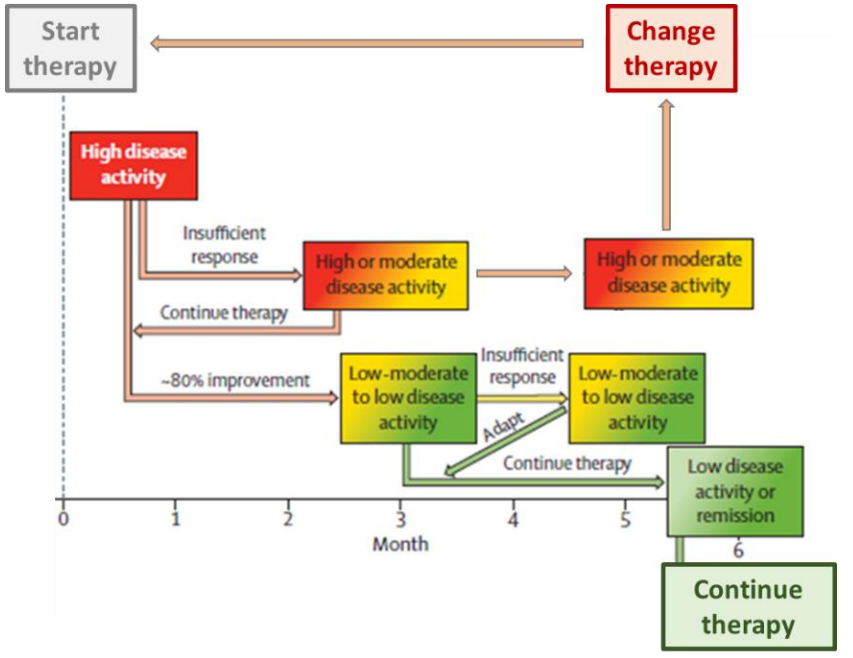


Figure 9. Treat to target strategy by disease activity monitoring (30).

However, there is still considerable unmet needs in RA: 1) based on the limitations of the current disease activity measurements, there is a great interest in developing novel multi-biomarker tests for reliable RA monitoring and subsequent therapy selection and optimization; 2) full or stringent remission is not typical or sustained without continuing treatment, with the consequences and side effects that chronic treatment may entail, such as higher risk of infections and comorbidities or economical costs (30, 38, 39, 41).

2.2.4. Rheumatoid Arthritis as a Rheumatic Immune-Mediated Inflammatory Disease.

Immune-Mediated Inflammatory Diseases (IMIDs) are chronic and highly disabling disorders that involve an inappropriate or excessive immune response caused by dysregulation of the normal cytokine balance, ultimately causing inflammatory injury in different organs/systems. Depending on where, when, and in what microenvironments the dysregulation occurs, over- or underproduction of certain cytokines may have diverse inflammatory effects. Thus, interleukins 1, 2 and 12, interferon γ, and tumor necrosis factor α are pro-inflammatory, whereas interleukins 4 and 10, and transforming growth factor β are anti-inflammatory (43).

RA, as well as other mentioned rheumatic diseases such as AS, PsA, SSc and SLE, are considered IMIDs which primarily affect joints. However, there are many other IMIDs that affect other organs/systems, such as psoriasis (skin), Crohn's disease (digestive tract), ulcerative colitis (intestine) and autoimmune hepatitis (liver) among others. Despite shared pathophysiology and therapy, autoimmune conditions are not typically seen as interrelated. Thus, in spite of the existence of the "Endocrine, Metabolic, and Immunity Disorders" Major Diagnostic Category (MDC), most IMIDs fall into MDCs specific to the organs/systems they affect. Around 7% of the population suffers from an IMID and their burden involves different linked comorbidities (especially cardiovascular disease), healthcare use and life quality impairment, among other factors. The majority of the IMID patients (90%) present only one IMID, although around 8% of the patients may conjointly present two different IMIDs. The prevalence is generally higher in women and in patients of 40-80 years old. Although in RA and PsA patients the prevalence raises with age, in other IMIDs the age distribution is variable. As seen in RA, genetic and environmental factors play an important role in the development of these disorders (43-47).

Earlier and more specific diagnosis as well as more efficient therapeutics help to improve IMID patients' management. On the one hand, it is necessary and appropriate to include additional interdisciplinary medical fields for correct IMID diagnosis, since more than one organ/system is generally involved. However, finding the correct diagnosis is still usually challenging. Around 4-32% of all rheumatic IMID patients are not adequately diagnosed when only applying the current classification criteria, since one of the prerequisites for positive classification is typically the exclusion of suffering similar conditions. Nevertheless, an overlap between similar diseases is feasible, converting the diagnosis into a more complex process. Therefore, there is a great and challenging need for specific biomarkers that could distinguish among related IMIDs in terms of diagnosis and disease activity follow-up (46, 48). On the other hand, IMIDs share the same therapeutic aims: rapid inflammation control, tissue damage prevention, quality of life improvement and long-term disease remission achievement. Anti-cytokine therapies have lately yielded therapeutic progresses in extremely different conditions. However, since adjusting cytokine balance may reduce or unintendedly boost the inflammatory effects, the current challenge is to identify exactly to who and when to apply these therapies in order to save time and avoid adverse reactions (49, 50).

3. Proteomics.

The term proteome was firstly designated by Marc Wikins in 1996, who linked the terms PROTein and genOME, for defining the complete set of proteins that a cell or organism could express (51). Biochemically, proteins are macromolecules composed of one or more chains of specifically ordered amino acids with unique functions. Biologically, proteins are essential components of the body's tissues, organs and systems, required for their structure, function, and regulation. Enzymes, hormones and antibodies are among the major protein classes. The word proteomics was later introduced in 1997 (52) in order to describe the discipline compelled to study the different proteomes, as well as the proteins' structure and function(s).

The general protein synthesis was firstly stated by Crick in 1958 (53) under the concept of The Central Dogma. Briefly, the DNA (genome) preserves the genetic information, which is self-replicated and transcribed into messenger ARN (mARN), conforming the transcriptome, and finally translated to proteins (proteome) by means of the genetic code. However, while the genome is the same for all cells, the proteome is characteristic of each cell type. Then, the proteome, unlike the genome, is highly complex and dynamic, since its components may vary in each cell as a consequence of several diverse processes (**Figure 10**). Gene activation or deletion, alternative splicing of the mRNA and/or more than 200 post-translational modifications (PTMs) are some of the processes that affect the complexity of the proteome, which are influenced as well by other circumstances such as physiological and pathological conditions, environmental changes, stress situations, drug exposure, ... Thus, due to the fact that the same gene is involved in the formation of various protein variants, the total number of proteins is highly increased from the number of genes/transcripts that generate them. Nevertheless, the dynamism of the proteome has an advantage over the genome,

since the proteome displays a real image of what happens in the organism, crucial to comprehend the gene function (54-56).

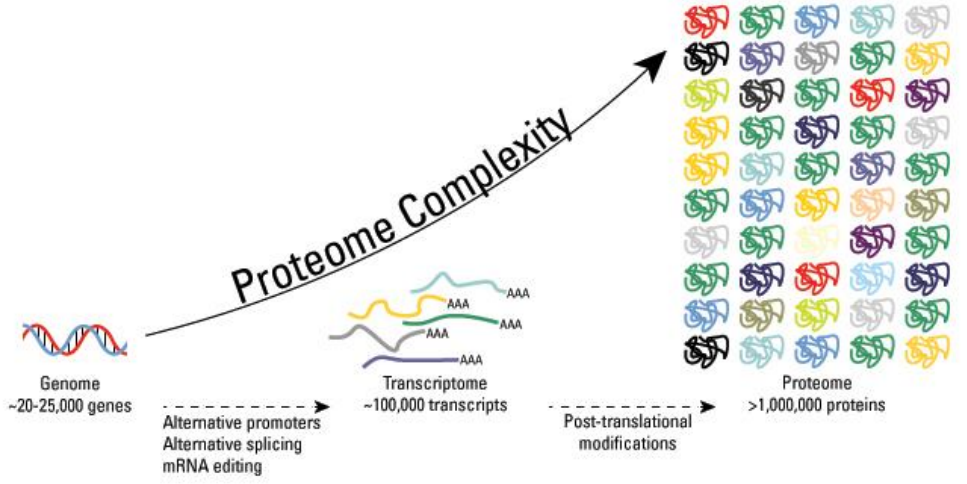


Figure 10. Proteome, transcriptome and genome comparison and the influencing factors affecting the proteome complexity. Figure from: www.creative-proteomics.com

In view of the success of the Human Genome Project in 2001 (57), a gene-centric approach, the Human Proteome Project, has been proposed in 2010 to generate a complementary human proteome map about the proteins expressed from each gene locus with the final aim of facilitating in-depth studies to understand human biology and diseases (58).

3.1. Main proteomic applications: Biomarker Discovery.

Proteomics can technically be classified in three types: protein expression, structural and functional proteomics, although all of them are interrelated and complementary (**Figure 11**) (56, 59). Structural proteomics is the determination of the 3D atomic protein structures, which aims to identify all the proteins within a protein complex or organelle, determine their location, and characterize all protein-protein interactions. Functional proteomics is focused on the elucidation of the biological function of the unknown proteins and the definition of diverse

cellular mechanisms at the molecular level such as protein signaling, disease mechanisms or protein-drug interactions. Expression proteomics is the quantitative study of protein expression for proteome profiling and proteome comparison between different samples/conditions, also termed differential proteomics. Differential expression proteomics is widely used and crucial in biomedical sciences, since it allows to determine which proteins are significantly altered in specific pathological processes (biomarker discovery), therefore applied for the development of this thesis, and also aids to characterize particular molecular and/or biological mechanisms or also to identify pharmacological targets.

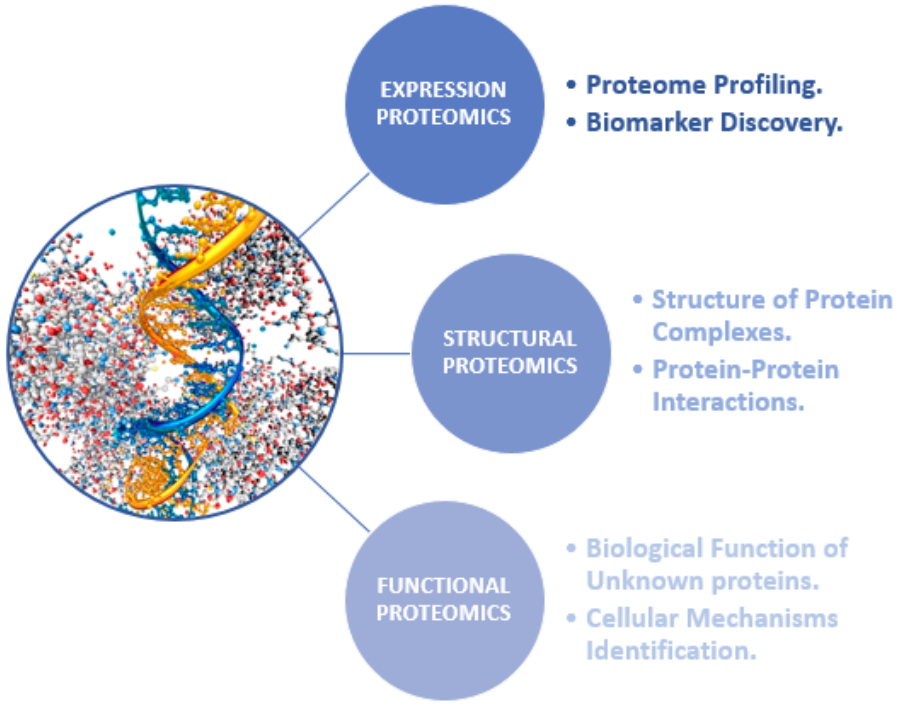


Figure 11. Proteomics classification and main purposes.

A biomarker is a molecule, feature or property which can be objectively measured and evaluated as an indicator of specific physiological or pathological processes or responses to particular exposures and/or pharmacologic therapies. Biomarkers should offer high sensitivity and specificity values in order to be

confirmed as valid indicators for clinical use. Particularly, sensitivity is proportional to the number of true positives that are correctly identified by the biomarker (most of the people with the target disease are detected), while specificity is significant if the proportion of true negatives that are correctly identified is elevated (most of the people without the disorder are excluded). Different types of biomarkers are used in clinics depending on the purpose, such as diagnostic (for disease detection and stratification), monitoring (for disease follow-up), prognostic (for predicting disease evolution) or predictive (for treatment selection) biomarkers among others (60, 61). This thesis is mainly focused on the first two biomarker types for disease diagnosis and monitoring.

Efficient novel biomarker development is suggested to be achieved by a phased approach that progressively shifts from an unbiased discovery phase to a targeted biomarker validation before final clinical evaluation, which could take many years since the biomarker is preclinically described. Generally, different proteomic strategies, such as mass spectrometry (MS) methods and/or immunoassays, are used in this biomarker pipeline including discovery, qualification, verification and validation phases (**Figure 12**). These and other proteomic techniques will be later reviewed in further detail, as most of them are implemented in this thesis. Briefly, a non-targeted MS approach (shotgun proteomics) is firstly performed in the discovery phase for relative quantitation of 1000s of protein analytes in a small number of samples. Significantly and differentially expressed proteins found between the studied conditions undergo the following qualification and verification phases in order to confirm the differential expression of these potential biomarkers on additional samples by higher-specificity MS techniques or immunoassays. Finally, a smaller number of the verified promising biomarkers is validated on 100s of samples prior to clinical evaluation on 1000s (62-64).

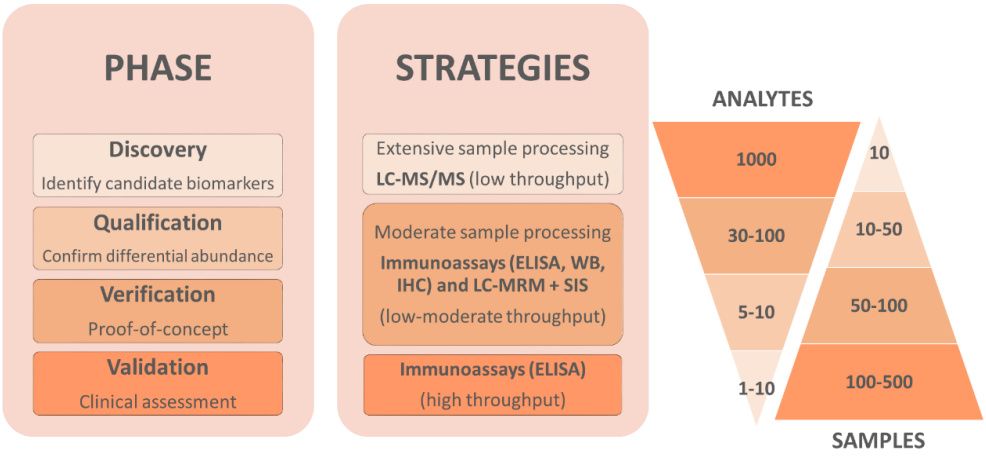


Figure 12. Pipeline of the development of novel biomarker candidates using proteomic technologies (63).

Different biological specimens have been widely studied in order to find specific OA and RA protein biomarkers, although not many of these biomarkers have reached the target clinical phase. On the one hand, proteomic studies have been focused in highly related joint specimens such as direct joint components (cartilage tissue, synovial membrane, SB), analogous cells (chondrocytes, synoviocytes) and/or their corresponding conditioned media or secretomes. On the other hand, joint-proximal SF, blood (plasma and/or serum) and other non-invasive body fluids (urine, saliva, tears) are expected to be excellent sources of OA/RA protein biomarkers since they come in contact with the previous joint specimens and pick up their secreted proteins and potential biomarkers. However, these released proteins are diluted in the cited body fluids to a degree that usually makes them nearly undetectable by the current available proteomic techniques (65). Thus, as seen in **Figure 13**, the majority of the plasma proteins that are released from tissues are of low abundance with concentrations ranging from ng/mL to pg/mL, making their detection difficult by cause of the presence of higher abundance and non-relevant proteins for biomarker discovery (66). Therefore, due to the complexity of the human plasma/serum proteome, there is a great need for developing new proteomic technologies that could beat the

challenging discovery of new OA/RA biomarkers, especially in these samples. In this thesis, a wide range of specimens has been studied for the search of OA and RA biomarkers including cartilage tissue, cartilage secretome, SF, serum and/or plasma, using diverse conventional and/or emerging proteomic strategies.

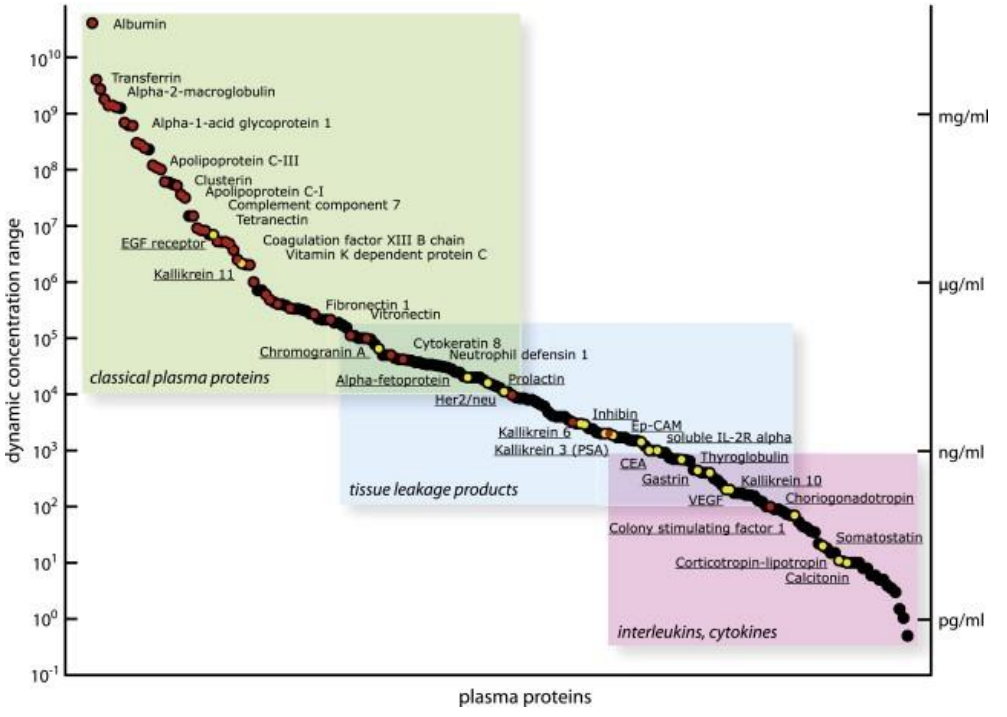


Figure 13. Protein categories included in the dynamic plasma concentration range: classical plasma proteins, tissue leakage products, signaling compounds (interleukins, cytokines). Red dots indicate proteins identified by the HUPO plasma proteome initiative, yellow dots represent currently utilized biomarkers (66).

3.2. Conventional and emerging proteomic strategies for biomarker search.

Proteomic strategies typically comprise three main steps: sample collection from biological or clinical specimens and storage; protein isolation and purification by precipitation, fractionation, enrichment and depletion methods; and protein analysis using immunoassays or MS (**Figure 14**). Sample collection and protein isolation and purification make up the sample preparation procedure,

which principal aim is to obtain high-quality, proteome-representative protein samples for maximizing protein identification and quantitation in the subsequent protein analysis step. It is the most fundamental step, especially in differential proteomics, where achieving reliable results is essential for dealing with minor variances among the studied conditions. None, one or combined protein isolation and purification strategies may be performed depending on the sample's complexity and the applied protein analysis technique (67).

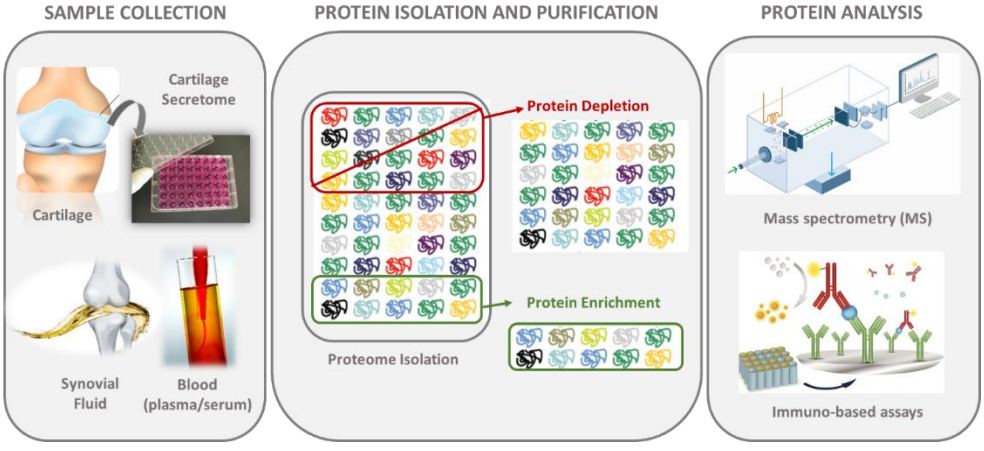


Figure 14. General steps in proteomic strategies for OA/RA biomarker discovery.

3.2.1. Sample preparation for proteomic analysis.

As seen in **Figure 12**, biomarker discovery, qualification and verification phases need moderate-to-extensive sample processing in order to satisfactorily identify and quantify the maximum number of protein biomarker candidates. As described in the previous section, blood samples are the most complex and dynamic among the samples used for biomarker discovery, therefore the need for a more extensive sample preparation in these particular specimens. By reason of this challenging complexity and since most of this thesis is focused on the search of biomarkers in serum and plasma, sample preparation strategies will be centered in these specimens, although they may be also applied in other OA/RA-related samples. Thus, the following protein isolation and purification methods

may be applied to detect tissue-leaking proteins in blood samples, in order to gain resolution and sensitivity for subsequent protein identification and quantitation.

3.2.1.1. Protein depletion methods.

The top twelve most abundant plasma proteins account for approximately 95% of the total protein content (albumin representing the 50-60%), while the rest of the plasma proteins is present in a wide dynamic range, covering more than 10 orders of magnitude in terms of concentration (**Figure 15**).

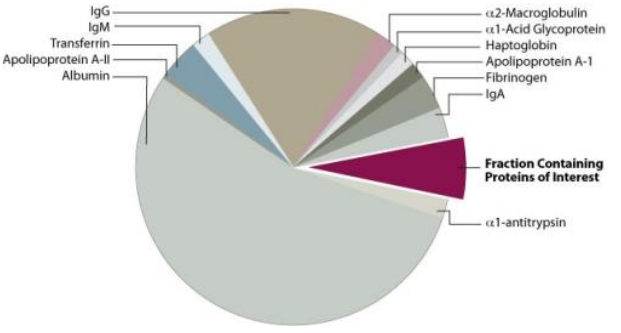


Figure 15. Most abundant proteins in serum/plasma (R&D Systems).

Therefore, different protein depletion strategies have been developed for reducing or removing one or various high abundance proteins, with the main aim of enhancing detection sensitivity and identification of lower abundance proteins or peptides in blood samples or other dynamic body fluids. The most common and efficient depletion strategy is based on the use of specific antibodies against these major proteins on chromatographic columns or solid phase extraction cartridges (68-70), although more affordable techniques as centrifugal ultrafiltration using different molecular weight cutoff (MWCO) membranes (71, 72), chemical protein precipitation (73) or commercial kits (74) have also been quite useful for fast and straightforward depletion of major proteins. However, while the number of analyzed and identified proteins is likely to increase after performing the adequate depletion, other proteins and/or peptides might be lost during the depletion process due to nonspecific binding to the depleted larger

carrier proteins and/or interacting surfaces. Therefore, it is crucial to broadly optimize the applied depletion procedures and analyze the bound and unbound fractions after depletion to verify that target proteins are not inadvertently lost (67, 75, 76).

Particularly, centrifugal ultrafiltration separates proteins into two compartments by centrifugation against diverse semipermeable MWCO membranes. Depending on the molecular weight (MW) cut-off used, the low MW proteins are forced to pass through the membrane, while the high MW proteins are retained in the sample chamber. The cut-off of 30 kDa is generally used for albumin and immunoglobulins depletion since their MWs are over 50 kDa, and represent more than 75% total protein content.

Protein precipitation (defined as the isolation of proteins from a solution as a solid, after altering its solubility by the addition of a specific reagent) is probably the simplest, most economic and accessible technique to isolate proteins and remove contaminants (detergents, unwanted cellular products) that could interfere in the posterior chromatography (baseline lift, noise, column deterioration) or MS (deteriorated ionization, ion source blockage) steps. It is typically performed by adding water-miscible organic solvents (acetone, acetonitrile, methanol, ethyl acetate), inorganic acids (trichloroacetic acid (TCA), perchloric acid) or salts (ammonium/zinc sulfate) to the sample in a 3:1 or 4:1 ratio (v/v) and mixing. Thus, insoluble proteins aggregate and form a pellet that can be easily removed from the remaining liquid by centrifugation or filtration. However, since proteins precipitate under different conditions, some valuable proteins may be lost during the procedure, influencing posterior protein identification and quantitation (77-79).

3.2.1.2. Protein enrichment methods.

Enrichment strategies aim to enhance the concentration of characteristic proteins and reduce sample complexity for subsequent identification and quantitation, especially when the target proteins represent a minor fraction or have been post-translationally modified. Enrichment strategies include electrophoretic and chromatographic fractionation techniques, as well as targeted immunoprecipitation methodologies. They basically separate different protein types and enrich the target protein(s) in smaller fractions depending on specific characteristics such as size, MW, charge, hydrophobicity or immunoaffinity (IA) with the aim of gaining resolution and sensitivity in the posterior protein analysis (67).

Immunoprecipitation (IP) is defined as the small-scale affinity purification of a particular antigen using a specific antibody that is immobilized to a solid support, commonly to Protein A and/or G magnetic particles (**Figure 16**). Proteins A and G show specificity for the heavy chains on the Fc (fragment crystallizable) antibody region, orienting the immobilized antibodies with antigen-binding sites facing outward. IP was adapted from traditional affinity chromatography, although instead of using a packed column, IP uses a low number of beads per experiment and incubation steps are performed in a batch-wise manner. Beads are pelleted to the bottom of the tube

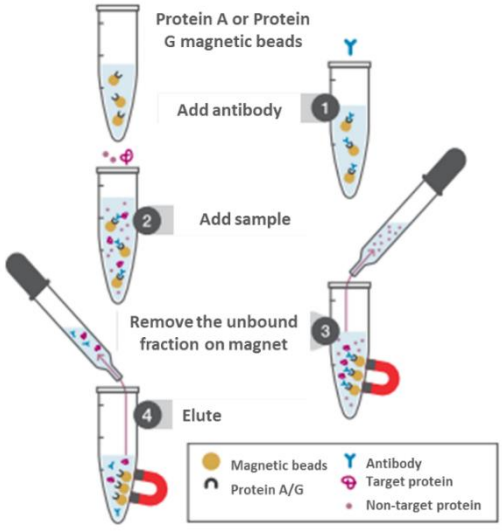


Figure 16. General immunoprecipitation protocol for antigen purification with magnetic particles (Adapted from ThermoFisher webpage).

by centrifugation or separated with the aid of a magnet so that the solution containing the unbound proteins or antibody excess can be removed. Unlike affinity chromatography, the IP goal is to isolate just enough protein to be able to measure it by western blotting or MS methods for combining the selectivity of the immunoaffinity (IA) extraction with the specificity of the MS detection (80, 81). The different combined IA-MS strategies are later reviewed in the emerging strategies section, since they were applied in this thesis for protein biomarker search.

Gel electrophoresis (GE) is used for separating and enriching proteins based on their MW and/or isoelectric point (pH at which the proteins net charge is neutral). Polyacrylamide and agarose porous matrices are commonly used in GE, since they act as a molecular sieve. Generally, protein separation by GE is performed within one dimension (1D) or two dimensions (2D) and the resulting separated proteins can then be visualized with various staining methods such as Coomassie, silver, zinc, fluorescent and functional group-specific stains (**Figure 17**). 1D or Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is the most common GE technique. It basically separates proteins based on their MW, since proteins are denatured and uniformly and negatively charged by heating the samples in the presence of SDS so, when a current is applied, all SDS-bound proteins within a sample migrate towards the positively charged electrode, uniquely based on their mass. Moreover, 2D-PAGE is used for separation of more complex protein mixtures, in the first dimension by the isoelectric point (using isoelectric focusing in a pH gradient) and in the posterior second dimension by protein MW, providing higher resolution than SDS-PAGE alone. However, these GE techniques are not reproducible and sensitive enough for detecting low MW proteins (67, 82).

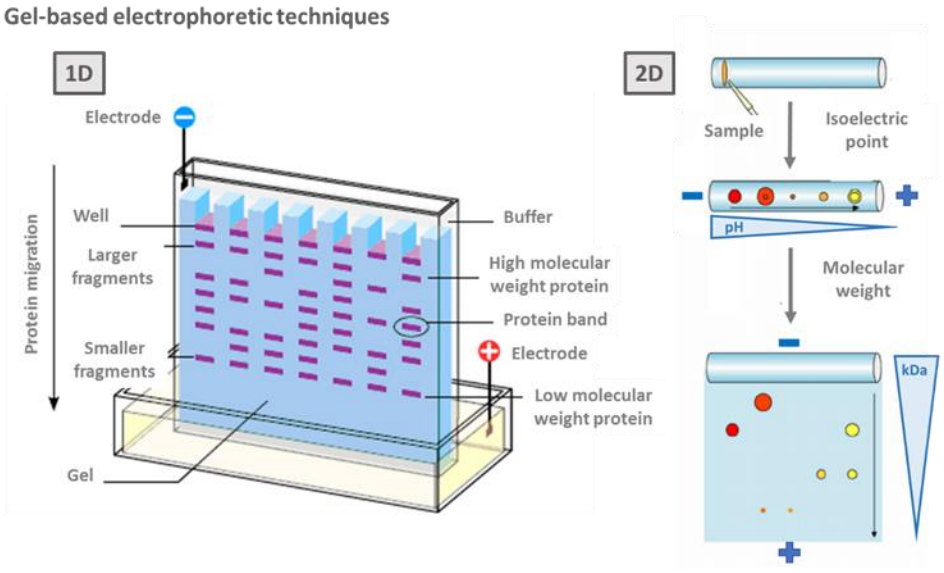


Figure 17. Gel electrophoretic techniques for protein fractionation and enrichment.

Column liquid chromatography (LC) techniques are generally used for protein and/or peptide fractionation, based on interactions between proteins/peptides in the solution or mobile phase (MP) and the solid support or stationary phase (SP) within a column. Depending on the interaction between the target protein/peptide and the solid support, different types of LC may be performed in proteomics such as reversed-phase (RPC), hydrophobic-interaction (HIC), ion-exchange (IEC), size-exclusion (SEC) or affinity (AC) chromatography (**Figure 18**). Both RPC and HIC are based on hydrophobic interactions between the proteins in a polar MP and a nonpolar SP. However, in the case of RPC, the nonpolar SP (hydrophilic silica beads covered with hydrophobic carbon chains or hydrophobic polymer matrix) is more hydrophobic than the HIC SP (polystyrene/divinyl benzene, agarose). This leads to stronger hydrophobic interactions that must be reversed using increasing amounts of non-polar, organic solvents (acetonitrile, methanol), for successful gradual elution of the more polar to the less polar proteins in RPC. Although in HIC the proteins are also progressively eluted in the same order, the mechanism is different since the hydrophobic interaction is firstly promoted by an increase of salt content and later

impeded due to the effect of lowering the salt content (ammonium/sodium sulfate, sodium/potassium chloride). Furthermore, RPC is also used for desalting and concentrating the hydrophobic components of a sample (ionic salts pass through the RPC column while hydrophobic components are concentrated on the column until elution) converting it in an ideal technique to couple with MS, especially if combined with other chromatography techniques (multidimensional LC), allowing for high-resolution, fast and accurate protein identification and characterization in proteomics. Thus, although RPC is the most frequently used LC method for protein analysis, IEC, SEC or AC are also employed for protein fractionation according to protein charge, molecular size or biospecific recognition, respectively. Nevertheless, the production of fractions in LC techniques results in increasing numbers of the analyte per specimen, causing the need to focus on a single or several fraction(s) which contain the target proteins (67, 79, 83).

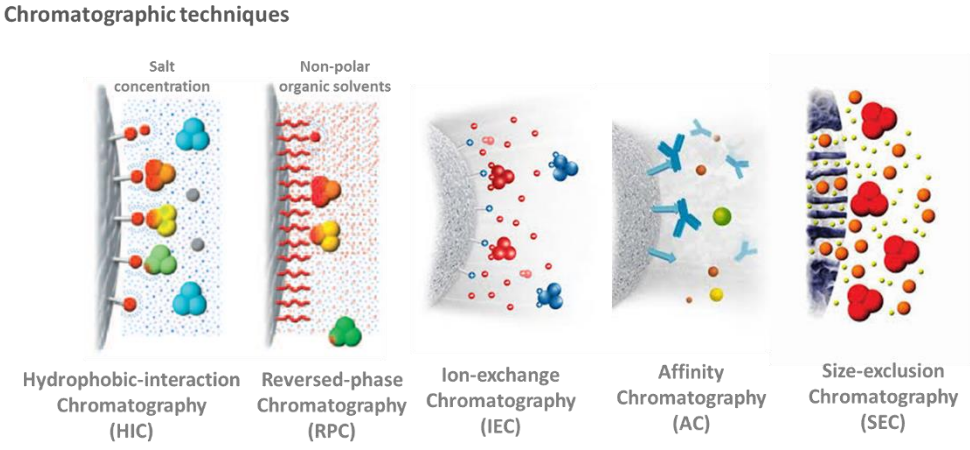


Figure 18. Chromatographic techniques for protein fractionation and enrichment.

3.2.2. Mass spectrometry techniques for biomarker discovery and verification.

Mass spectrometry (MS) is an analytical method for sensitive detection, identification and quantitation of molecules in simple and complex mixtures

based on the mass-to-charge ratios (m/z) of the corresponding resulting ions. In order to perform MS analysis, three different steps are carried out in a mass spectrometer: 1) ionization of the sample; 2) separation of the resulting ions according to their m/z; 3) specific ion count. **Figure 19** shows the components of any mass spectrometer, which require high vacuum conditions in order to maintain all components at low pressure and avoid ion-gas molecule collision. First, the inlet system introduces the sample into the ion source, where the components are converted into gaseous ions. Then, the resulting gaseous ion flow is accelerated in the mass analyzer, and ions are sorted according to their m/z ratio. Finally, ions reach the detector, which converts the ion beam into an electrical signal that can be processed, stored and interpreted by a data system with the aim of obtaining the mass spectrum (83, 84).

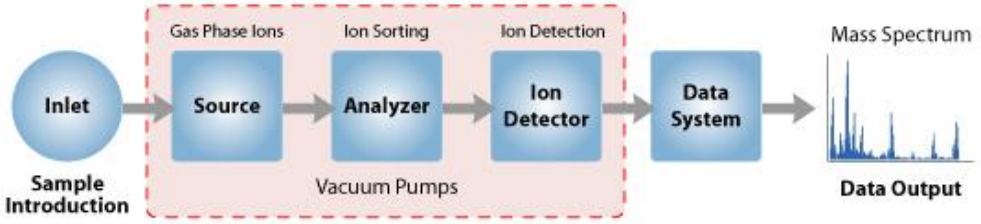


Figure 19. Components of a mass spectrometer. Figure from premierbiosoft.

Two types of MS information are typically obtained from the proteins/peptides analyzed: protein/peptide masses (MS) and their fragmentation patterns or product ions (tandem mass spectrometry or MS/MS). MS/MS uses two consecutive stages of mass analysis: 1) isolation of a particular protein or peptide precursor ion of interest by its m/z (MS); 2) m/z analysis of the product ions formed by spontaneous or induced fragmentation of the selected precursor ion (MS/MS) by different techniques as CID, SID or IRMPD (**Figure 20**). Thus, MS/MS spectra contain precursor fragments that are related to the corresponding peptide sequence, which, in conjunction with the MS spectra, help to elucidate the protein/peptide sequence identification.

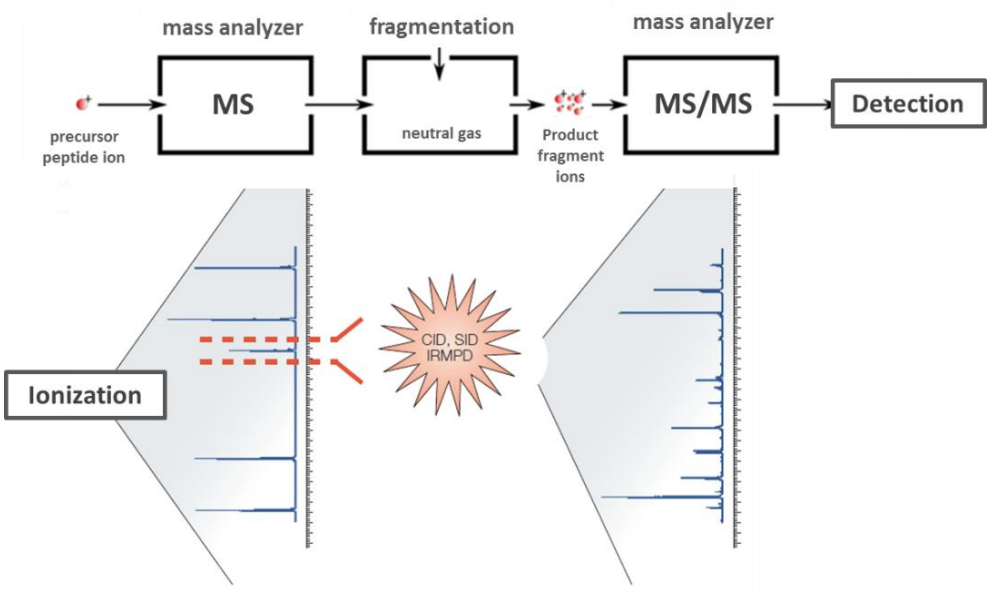


Figure 20. Peptide MS and MS/MS analysis. CID (Collision-induced dissociation), SID (Surface-induced dissociation), IRMPD (Infrared multiple photon dissociation) are different fragmentation techniques.

The mass spectrometer instrument designation is usually dependent on the kind of ion source and mass analyzer(s) used. The two most applied ion source techniques for MS peptide/protein analysis are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). However, mass analyzers are the most essential components of the mass spectrometer since they define the instrument's performance in terms of mass accuracy, resolution, sensitivity, and MS/MS capability. The most used mass analyzers for peptide/protein analysis are time of flight (TOF), quadrupole (Q), ion trap (IT), Fourier transform ion cyclotron resonance (FTICR or FT) and orbitrap, which can be categorized as well as scanning and ion-beam (TOF and Q) or trapping (IT, FT and orbitrap) mass analyzers. They can be used alone or, more frequently, in combination for boosting the instrument's versatility and performance (**Figure 21**).

Instrument	IT/LIT	QQQ	QTOF	IT-TOF	TOF-TOF	LIT-FT/ Orbitrap
Mass accuracy	+[b]	+	++	++	++	+++
Resolution	+	+	++	++	++	+++
Sensitivity	++/+++	++	++	++	++	+++
Speed	++/+++	+/+++[c]	++	++	+++	++
Dynamic range	+/++	++	++	++	++	+++
Additional Features	MS^n	MRM, precursor scan		MS^n	high energy CID	MS^n

[a] Abbreviations: CID, collision-induced dissociation; FT, Fourier transform; IT, ion trap; LIT, linear ion trap; MRM, multiple reaction monitoring; MS^n, multi-stage MS/MS; QQQ, triple quadrupole; TOF, time of flight.
[b] +, ++ and +++ stand for low, good, and excellent, respectively.
[c] The speed for full scan is low; the speed for MRM is very high.

Figure 21. Comparison of single and/or combined mass analyzers (83).

3.2.2.1. Ion sources.

On the one hand, in MALDI, ions are obtained from mixing and crystalizing the sample with a molar excess of an organic matrix on a metallic plate, after applying a particular laser energy (**Figure 22**). This matrix serves to reduce analyte intramolecular interactions and rapidly absorb large amounts of energy. It is crucial to perform an adequate matrix preparation and homogenous crystallization since this directly influence sample ionization, spectra quality, resolution, reproducibility and correlation between the signal and the analyte's concentration. Briefly, when the crystalized sample is bombarded by a laser operating in the ultraviolet (UV), the matrix absorbs UV light, promoting matrix and analytes sublimation and ionization in the spectrometer, therefore resulting in protonated gas molecules that are subsequently directed towards the mass analyzer using an electric field. The choice of a matrix depends on the wavelength irradiation and the type of sample analyzed. The most used MALDI matrices are the α-cyano-4-hydroxycinnamic acid (α-cyano or CHCA), the 3,5-dimetoxy-4-hydroxycinnamic acid (sinapinic acid or SA), and the 2,5-dihydroxybenzoic acid (DHB), which absorb energy at the UV wavelength of 337 nm, since all present aromatic groups. CHCA is the most sensitive and commonly used matrix for

peptides and small proteins, which induces higher fragmentation, whereas SA is mainly used for peptides and large proteins, and DHB for peptides, small proteins and their glycosylated variants. MALDI predominately generates singly charged molecular peptide/protein ions, so the sample spectral interpretation is greatly straightforward. Moreover, due to its femtomole sensitivity and buffer-tolerance, MALDI has become an excellent option for the analysis of protein digests.

On the other hand, in ESI, ions are formed by spraying a liquid sample containing the corresponding peptides and proteins from the tip of a fine capillary, at atmospheric pressure and high voltage conditions (**Figure 22**). Briefly, in this spray process, often assisted by pneumatic nebulization, sample particles are dispersed as charged microdroplets (electrospray) formed in a very high electric field by applying a 2.5-6 kV voltage. These droplets become highly charged and solvent droplet evaporation is reasonably rapid to produce drops of smaller size that, when increasing the load density and exceeding the surface tension, cause the molecules to separate, via coulombic repulsion, into several analytes with single or multiple charges (after picking up one, two, or more protons from the solvent). These new drops are finally transferred to the mass analyzer by action of the electric field. The number of charges acquired by a molecule is practically the same as the number of possible protonation sites, which also correlate with the length of the peptide (usually one proton per 1000 Da), then the most common ion charges are two or three for tryptic peptide fragments. Reduced or no fragmentation is generally observed in ESI for peptides, and the most popular ESI operating mode is in combination with HPLC. However, since the signal intensity produced in an ESI mass spectrometer depends on the analyte's concentration rather than its absolute amount, it is desirable to use nano-LC instruments to achieve better sensitivity, as they typically operate at low flow rates (100 to 300 nl/min) and are compatible with small column diameters (75 μm).

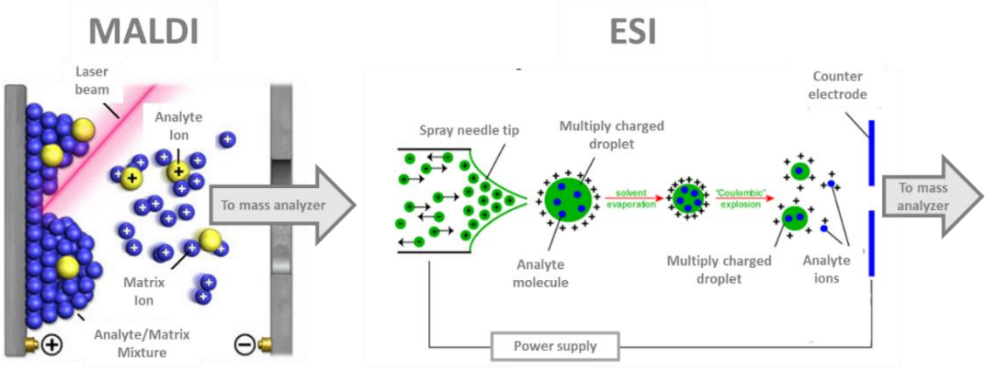


Figure 22. Scheme of MALDI and ESI ion sources.

3.2.2.2. Mass analyzers.

In TOF, ions of the same charge are accelerated to the same kinetic energy (E_k) before directed into a flight tube in which the ions of different velocities (flight times) are separated before detection (**Figure 23**). Thus, since E_k is constant for all the ions, their velocities are inversely proportional to their molecular masses in accordance to the E_k equation ($E_k=(1/2)mv^2$). As a result, ions of greater mass fly more slowly than lighter ones and reach the detector later. TOF analyzers are generally coupled to MALDI ion sources, however, MALDI-TOF instruments present low resolution, partially due to little distribution differences in the initial ion E_k, which is typically corrected by an energy-focusing device (ion mirror or reflectron). TOF can be coupled to other analyzers such as Q and IT, for combining the TOF resolution with the Q/IT MS/MS capability, as well as to another TOF (separated by a collision cell), where the first TOF is used to select ions, which are fragmented in a high-energy collision in the collision cell, and the resulting fragment ions are measured by the second TOF.

In Q, ions of a particular m/z travel through two pairs of parallel rods, by continuously adjusting radiofrequency (RF) and direct-current (DC) voltages, until they reach the detector, while the other ions collide with the rods (**Figure 23**). Q

presents slow scan speeds and limited mass accuracy and resolution. Q is frequently combined with other analyzers as in triple quadrupole (QQQ), QTRAP and QTOF instruments. In QQQ, three Qs are disposed in series, while in QTRAP, the third Q (Q3) can also function as a linear IT with the aim of enhancing the MS/MS scan speed and sensitivity. Four different types of scans can be typically performed on QQQ for different analytical goals, where Q2 is normally used as a collision cell for fragmenting the precursor ions. Thus, product ion scanning (for obtaining the full MS/MS spectra of the selected peptides), neutral loss scanning (for detecting a particular functional group), precursor ion scanning (for detecting specific fragment ions) and multiple reaction monitoring (MRM) can be performed on a QQQ. MRM is frequently conducted for quick peptide quantitation by using known precursor-fragment pairs (transitions) for a particular peptide.

In ITs, ions are trapped and accumulate over time by the application of electrical fields before they are specifically ejected for MS, MS/MS or MS^n detection (**Figure 23**). They are usually coupled to ESI ion sources and present excellent MS/MS capabilities since ions can be also fragmented in the same trap as if it were a MS/MS system. ITs are sensitive, fast, robust, and relatively inexpensive although they are limited in terms of mass accuracy and resolution. Therefore, linear ion trap (LIT) and linear trap quadrupole (LTQ) analyzers have been developed, which present higher trapping capacity and better performance, resulting in ITs replacement.

In FTICR, ions are trapped on a cyclotronic m/z-dependent movement caused by a powerful magnetic field. When a RF voltage is applied, ions (whose cyclotronic frequency coincides with the transmitted voltage) lengthen their motion orbits and their frequencies are detected in function of time (**Figure 23**). The applied energy can also be adjusted to produce fragmentation or to expel the ions from the trap. This frequency-time spectrum is then converted using Fourier

transform methods to a mass-intensity spectrum. FTICR present very high mass accuracy, resolution, dynamic range, and sensitivity, although they also pose low peptide fragmentation efficiency and high operational expenses and complexity, which are normally addressed by using a hybrid LIT-FTICR instrument to enhance usability and MS/MS capability.

In orbitraps, ions are trapped and orbit around a central electrode while oscillating in axial direction by static electrostatic fields (**Figure 23**). The oscillation frequency of the oscillations is dependent on the m/z of the ion, which is similarly measured as in FTICR. Orbitrap displays high mass accuracy and resolution, with lower operational expense since it does not need a powerful magnetic field. Hybrid LIT-Orbitrap is typically used for time efficient analysis of complex peptide mixtures, since both analyzers can work in parallel as they have their own detectors. Thus, the LIT first send ions to the orbitrap for high-resolution MS analysis, then, while the orbitrap is working on the MS scan, the LIT conducts high-speed MS/MS.

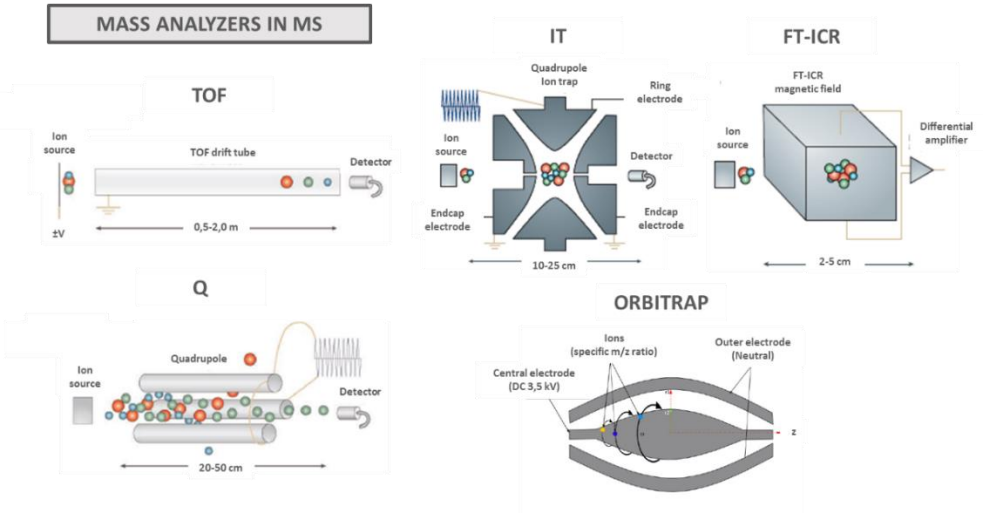


Figure 23. Scheme of TOF, Q, IT, FT-ICR and orbitrap mass analyzers.

3.3.2.3. Protein and peptide analysis and quantitation.

There are two main strategies for protein MS analysis: bottom-up and top-down. In the bottom-up approach, enzymatically-digested peptides from proteins are detected, whereas in the top-down approach, intact proteins are directly analyzed by ESI-MS. The top-down strategy provides complete compositional protein characterization, although its application is limited due to low sensitivity and difficulty for ionizing many proteins (especially larger than 30 kDa) and isolating target proteins from complex mixtures, while maintaining their solubility in ESI-MS-compatible buffers. Therefore, the bottom-up strategy is much more commonly used, since sample solubilization and fractionation prior to MS is easier for peptides than for proteins and the obtained digested peptides are readily analyzed by MS. However, this approach entails an additional sample digestion step before MS analysis and protein information is partially lost, as only a small fraction of the peptides is usually detected. Briefly, protein digestion is the process of breaking proteins by their peptide bonds, resulting in the generation of shorter protein fragments or peptides. In order to conduct the protein digestion, proteins need to be: 1) denatured (by heat or chemical agents such as SDS or urea); 2) reduced (by breaking the disulfide bridges with beta-mercaptoethanol and dithiothreitol (DTT)); 3) sulfhydryl-alkylated with iodoacetamide (IAA) (to avoid the formation of the disulfide bridges again); and finally 4) digested by peptidase enzymes which catalyze the peptide bond cleavage, often between two particular amino acids in order to generate predictable and reproducible protein fragments. The trypsin enzyme is typically used for protein digestion, mainly due to its specificity and the ability to generate tryptic peptides of proper length for MS analysis, with a lysine or arginine on the carboxyl end, which facilitates peptide ionization, fragmentation and sequence information. Protein digestion can be performed in gel, after protein GE separation of size-specific protein(s), or in solution, prior to LC fractionation (83).

In order to quantitate proteins/peptides by MS, the intensity of the corresponding analyte signal is normally used, since it is usually dependent on its abundance. However, this relationship is limited by several factors such as peptide ionization and desorption efficiencies, matrix effects, MS detector fluctuations and/or sample preparation variability, which are typically overcome by using labeled versions of the endogenous analytes. Thus, there are two major quantitation methods: the label-free and the labeling-based approaches. The label-free methods are not really accurate, although they are simpler and less expensive techniques that provide higher analytical depth in a large number of samples. In contrast, the labeling-based approaches analyze different tagged protein/peptide analogs, which yield similar ionization efficiencies and MS response signals in comparison to the corresponding non-labeled proteins/peptides. If the labeled-standard amount is known, absolute quantitation can be performed in several samples, although the proteins/peptides can also be differentially and simultaneously labeled in a constrained number of samples for relative quantitation. Along this line, two MS proteomic strategies are generally used for biomarker discovery and verification (**Figure 12**): shotgun proteomics using relative-quantitation techniques for initial unbiased, large-scale screening of potential differentially-expressed candidate biomarkers and targeted proteomics using absolute-quantitation strategies for reproducible, sensitive and accurate verification of the previously selected candidate biomarker targets (85).

On the one hand, different label-free or labeling approaches can be performed in order to determine relative protein abundances by comparing signal intensity or peak area of the corresponding peptides or reporter ions. Thus, Stable Isotope Labeling by/with Amino acids in Cell culture (SILAC), Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) and Tandem Mass Tags (TMT) are among the most common labelling strategies for relative quantitation. Thus, the iTRAQ approach was used in this thesis with the aim of screening potential

biomarker candidates across eight different samples (iTRAQ 8-plex) (**Figure 24**). Shortly, iTRAQ reagents are composed of three different regions: 1) nicotinoyloxy-succinimide esters, which covalently bind peptide primary amines (by the lysine and the amino termini peptide groups); 2) carbonyl balancer groups of different lengths (so all labeled peptide versions present the same mass, producing only one MS signal); 3) up to eight different reporter groups, which are released during peptide fragmentation and relatively quantified based on the corresponding peak intensities in MS/MS since they present different m/z (113.1; 114.1; 115.1; 116.1;17.1; 118.1; 119,1 and 121,1 in the 8-plex version).

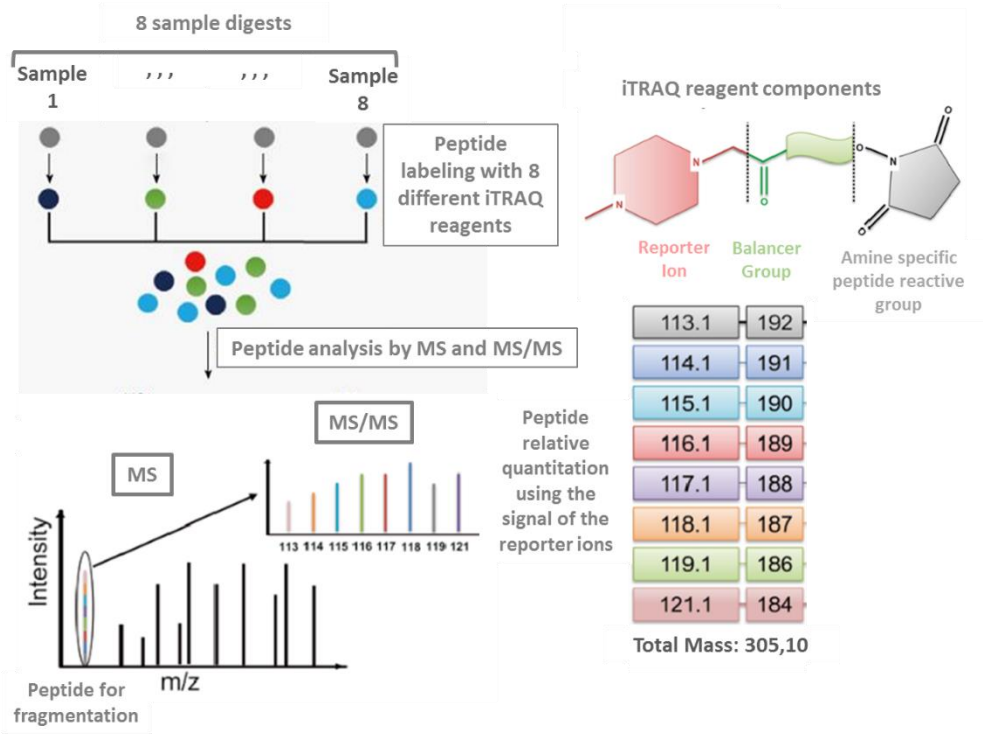


Figure 24. iTRAQ 8-plex workflow.

On the other hand, targeted proteomic MS strategies, such as MRM or parallel reaction monitoring (PRM), are typically used for multiplex protein absolute quantification of particular candidate biomarkers with the aid of SIS (Stable Isotope-labeled Standard) peptides (**Figure 25**). SIS peptides basically contain stable isotope labeled amino acids (commonly ^{13}C, ^{15}N or ^{2}H) which make

them distinguishable from their non-labeled analogs since they present different masses (therefore different m/z), while they share similar physiochemical properties and chemical reactivity (important for avoiding LC-MS variability). Briefly, known concentrations of several SIS peptides (also termed heavy peptides) from one or several target protein(s) are spiked into the sample digests and analyzed by the cited targeted MS strategies. As a result, peak intensities/areas of both the light endogenous peptides (of the sample digest) and the corresponding heavy peptides are compared and quantified by yielding light/heavy ratios (L/H), which are later used for calculating the absolute unknown light peptide concentrations, since the concentration of the spiked SIS peptide is already known.

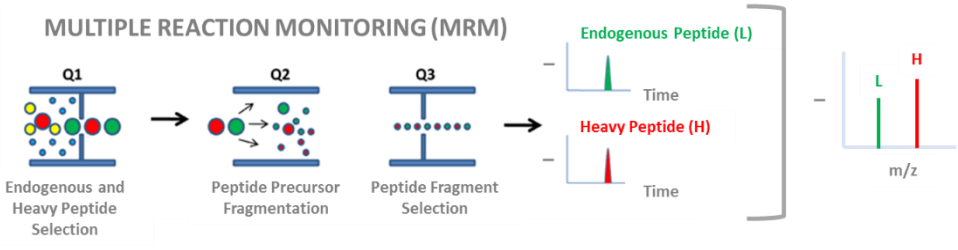


Figure 25. MRM workflow for simultaneous endogenous (L, green) and heavy (H, red) peptide analysis and quantitation using the L/H ratio.

3.2.3. Immuno-based techniques for biomarker validation.

Although targeted LC-MS/MS strategies (MRM, PRM) could also be used for multiplex biomarker validation in many samples, Enzyme Linked Immunosorbent Assays (ELISAs) are best suited for analyzing independent biomarkers on a large number of samples in well microtiter plates. ELISA is a highly sensitive immuno-based technique, widely used for detecting and quantifying the abundance of a particular antigen, mainly for diagnostic purposes. An ELISA could be conducted in three different formats, although the sandwich ELISA construct is the most common, which entails the following steps (**Figure 26**): 1) a capture antibody is immobilized on a well's solid surface; 2) the diluted sample is added and the target protein is specifically bound to the capture antibody; 3) the

unbound material is washed, followed by the addition of an enzyme-conjugated detection antibody (specific for a different antigen epitope), which binds to the target antigen; 4) any free detection antibody is removed by washing and a specific substrate is added, producing a chromogenic reaction, where the antibody-conjugated enzyme acts on the substrate producing a colored product. The color intensity can be measured by a spectrophotometer or an ELISA reader in order to compare the detected signal with a standard curve and determine the exact amount of antigen in the sample.

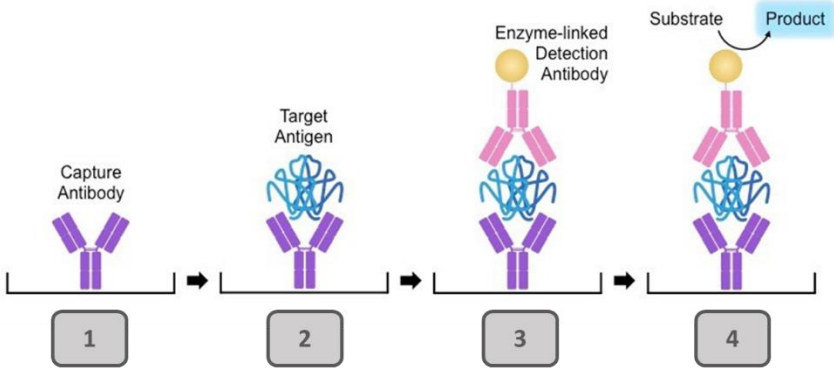


Figure 26. Sandwich ELISA workflow. Image modified from ib.bioninja.com

3.2.4. Emerging proteomic strategies for biomarker search.

3.2.4.1. Peptidomics.

Peptidomics is one of the proteomics branched field, which is presently given higher consideration as a proteomic biomarker discovery tool since it targets endogenously produced protein fragments that may indicate the presence and/or status of particular pathological conditions as endogenous peptides may be generated by active synthesis or aberrant proteolytic processing of larger precursor proteins. **Figure 27** shows the differences between both proteomic and peptidomic workflows for protein and endogenous peptide analysis respectively.

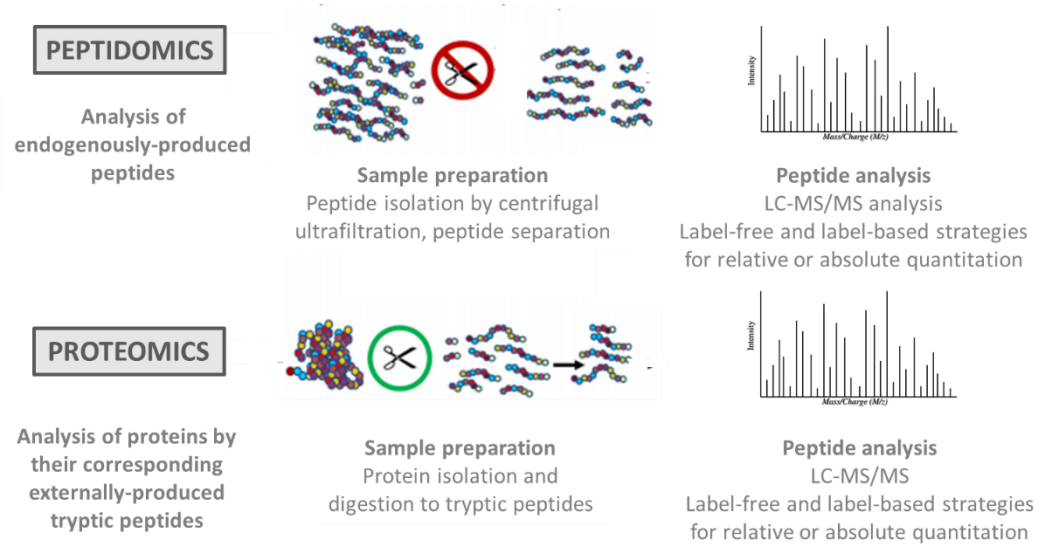


Figure 27. Peptidomics and proteomics workflow comparison.

However, peptidomic analysis is really challenging due to the low abundance of the endogenous peptides, which are constantly exposed to post-translational processing by site-specific proteases and proteasome degradation, making their detection difficult. Currently, several proteomic strategies are applied for peptide isolation from tissue extracts and body fluids, although centrifugal ultrafiltration using MWCO membranes with cutoffs of less than 10 kDa is the most common technique for simple and rapid peptidome extraction, where the lower molecular weight fraction is kept, in contrast to their use for protein depletion. However, this technique also leads to partial peptide losses and undesired contamination so, in some cases, more effective peptide isolation techniques are needed for being able to detect and quantify peptidomic traces. Additionally, peptides could also be first fractionated with different electrophoretic and chromatographic techniques in order to facilitate the peptidome analysis, which is typically done by different LC-MS/MS strategies using different label-free or labeling methods for relative and/or absolute quantitation. After peptide identification and quantification, different studies can be performed in order to find the relationship with the considered pathology such as: differential peptide comparison between conditions, peptide mapping inside

the protein sequence and enzyme recognition for biological enzymatic understanding and/or peptide structural and functional analysis. (86, 87).

3.2.4.2. Hybrid immunoaffinity-MS methods.

The most promising protein biomarkers are likely to be detected in particular disease-related specimens since they are much more likely present in higher/lower abundance in the corresponding affected organs and/or systems. However, in order to ultimately implement these tissue-specific biomarkers for clinical routine disease diagnosis and/or monitoring, they should be finally detected in more accessible specimens, such as plasma or serum. On the one hand, LC-MS/MS strategies lack the sensitivity to detect low abundant proteins in these convenient biofluids, even after extensive fractionation. On the other hand, the number of actual protein biomarker candidates identified has been exponentially raising, and sensitive immunoassays such as ELISA do not hold the sufficient multiplex ability for detecting all of them in a single analysis. Therefore, hybrid immunoaffinity-MS methods (IA-MS) have emerged as promising techniques in proteomics since they couple the high multiplexing capability and specificity of the different MS strategies with the vast sensitivity of the IA techniques (88-90). Two IA-MS techniques have been developed in this thesis for the detection of two cartilage-related proteins: an immuno-MALDI (iMALDI) assay and a Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA) technique. **Figure 28** shows the workflows for both IA-MS strategies. Both techniques have the advantage of using only one antibody instead of two, as in a conventional sandwich ELISA, where coupling issues and higher costs may be encountered during the ELISA development process. Besides, SIS peptides are initially spiked in the sample digests for performing precise absolute quantitation with the two cited strategies. iMALDI and SISCAPA workflows are nearly the same and comprise three main phases: 1) immunoprecipitation of the light and heavy peptide targets with the aid of anti-protein or anti-peptide antibodies coupled to

magnetic protein A or protein G beads; 2) acidic elution of both targeted peptide variants in tube (SISCAPA) or on plate (iMALDI); 3) MS analysis by shotgun MALDI-MS or targeted LC-ESI-MRM, respectively.

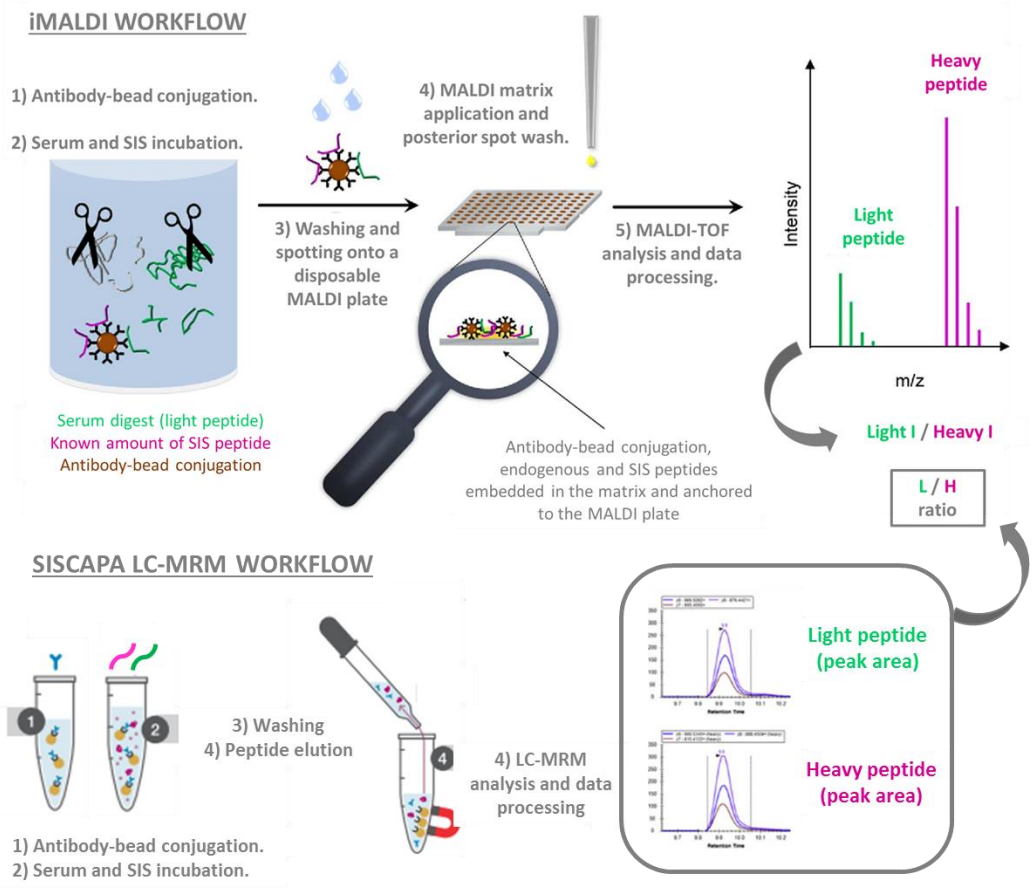


Figure 28. iMALDI and SICAPA LC-MRM workflows.

11. Objectives

OBJECTIVES

The general aim of this thesis is the search of protein biomarkers to improve the diagnosis and/or monitoring of patients with rheumatic diseases by proteomic strategies, using MS techniques and/or immuno-based assays. In order to conduct this principal objective, the following particular aims were projected:

1. To identify a panel of protein biomarker candidates for monitoring the disease activity in plasma from RA patients using the conventional proteomic pipeline for biomarker discovery.

2. To perform a peptidomic study on cartilage secretomes, synovial fluid and serum samples for the detection of a panel of endogenous peptides as possible OA biomarkers.

3. To develop two immunoaffinity-mass spectrometry (IA-MS) techniques for the detection and absolute quantification of two cartilage-related proteins with biomarker value for OA in serum samples.

III. Chapters

1. Chapter I: Detection of a panel of protein biomarker candidates for monitoring RA disease activity in plasma using the conventional pipeline for biomarker discovery.

RA monitoring after diagnosis is extremely important for a favorable disease surveillance, although more sensible and selective monitoring tools are needed in clinic in order to attain this aim. Currently, proteomics has proved to be a powerful tool for biomarker discovery in many clinical fields with the aim of disease diagnosis and monitoring, as well as treatment assignation and reconduction if needed. Detecting multi-panels of protein markers for RA activity monitoring has become very interesting for researchers (91), and companies such as Vectra, who are already selling blood profiling kits for RA patients.

In this chapter, a three-stage approach has been used to characterize the plasma proteomic profile associated with disease activity in a cohort of plasma samples obtained from the IMID (Immune-Mediated Inflammatory Diseases) Consortium. This cohort is composed of plasma samples from patients suffering different prevalent IMIDs including Rheumatoid Arthritis (RA), Psoriasis (Ps), Psoriatic Arthritis (PsA), Systemic Lupus Erythematosus (SLE), Crohn's Disease (CD), and Ulcerative Colitis (UC). After conducting the first stage (discovery stage) by shotgun mass spectrometry (MS), a panel of candidate biomarkers for diagnosis and disease activity monitoring was identified using a random selection of 80 patients with different extreme RA activities from the IMIDs cohort (n = 40 with high RA activity and n = 40 with low RA activity). In the second stage (verification stage), a targeted proteomic strategy was employed to verify the most significant candidate biomarkers from the discovery phase panel, using the same 80 samples of this first stage. After this verification step, a third phase (validation stage) was conducted, broadening the number of RA samples from the IMID cohort, in order to determine the uniqueness of the panel for the diagnosis and monitoring of RA activity in plasma. For this last purpose, two similarly sized

subgroups of patients showing extreme disease activity (i.e., very high and very low disease activity) were selected as disease controls within the PsA and SLE groups. Additionally, plasma samples from an independent cohort of healthy donors (HD) were also measured. A summarized scheme of the number of samples used in the study can be seen in **Table 1. 1.**

Table 1. 1. Plasma samples used in each phase of the study.

Disease	Disease Activity Score	Discovery cohort		Verification cohort		Validation cohort	
		LA	HA	LA	HA	LA	HA
RA	Disease Activity Score 28	40	40	40	40	79	91
PsA	Disease Activity Score 28	-	-	-	-	40	40
SLE	Selena-Sledai and BILAG	-	-	-	-	40	40
HD	-	-	-	-	-	90	

1.1. Materials and Methods.

1.1.1. Discovery Phase: Selection of biomarker candidates for RA disease activity monitoring using the relative quantitation iTRAQ labeling strategy and shotgun mass spectrometry.

Foremost, in order to identify a first panel proteomic biomarkers that could distinguish among patients with extreme RA activities, a first screening stage was conducted. For the determination of the disease activity level of these patients, the DAS28 was used, defining patients with low RA activity (DAS28: 2.6-3.2) and high RA activity (DAS28: >5.1). DAS clinically stands for Disease Activity Score, which examines the 28 joints normally affected in RA, employed for quantifying the RA activity. For this purpose, 80 plasma samples from patients with different RA activities were randomly selected from the IMIDs cohort (40 patients with high RA activity and 40 patients with low RA activity). Relative analysis was performed by carrying out an iTRAQ labelling strategy, with the aim of attaining an initial panel of possible RA activity distinguishing biomarkers, which would later be ascertained in the following verification phase.

1.1.1.1. Sample preparation.

Initially, the 80 plasma samples selected were pooled in 8 groups attending to the RA activity degree, minimizing the possible patient inter-variability. To later conduct an 8-plex analysis with iTRAQ labels, pools were firstly designed. Thereby, in order to achieve a relative RA activity comparison, four pools represented the high RA activity condition (HA-RA) and the other four, the low RA activity condition (LA-RA). Each pool was prepared with 10 µl of each individual plasma sample of the corresponding RA activity, and each pool was made up of 10 plasma samples, giving a final pool volume of 100 µL.

Knowing the high dynamic range of the plasma and the subsequent problems in the analysis that could arise due to this feature (mostly sensitivity limitations), a simple, fast and reproducible albumin-depletion method was applied to these plasma pools. It is widely known that without performing the removal of the most predominant proteins in plasma, the analysis of the less abundant proteins could be hampered and limited. Due to the fact that albumin is the most prevalent protein in plasma (60% of the total protein content in plasma is albumin), an albumin-chemical-depletion method using ethanol and trichloroacetic acid (TCA) was tuned up to be implemented in this study. The main concern about choosing this depletion method, was the maintenance of the less prevalent proteins and the loss of further interesting information they could develop (ie. immunoglobulins could have a paper on the RA activity changes as it is an autoimmune disorder disease).

Due to the preceding factors, a study of a chemical depletion method using a lab-tuned ethanol and TCA protocol was previously performed (92, 93), with the aim of assessing its application on the actual study. To accomplish this objective, 3 serum samples (around 20µg of total protein each one) were albumin depleted, trypsin digested and analyzed by nanoLC-MALDI-TOF/TOF using the ProteinPilot software for protein identification. In order to compare the data of this depletion method with a commercial depletion technique, an HSA Depletion Kit was purchased from Thermo Scientific. The same samples were albumin depleted as stated in the kit's protocol and subsequently treated the same way as the last samples were. The cited samples were additionally processed and analyzed as well without performing any depletion technique, in order to use them as a control, for later comparing the proteomic and peptidomic profiles obtained in each condition. The results of this experimental study can be seen in **Annex I**.

After verifying the capacity of the method, the 8 plasma pools depletions

were individually carried out. For this aim, 20 µL of each pool were added to 200 µL of ethanol 1% TCA. The mixtures were shacked for 3 minutes and centrifuged for 5 minutes at 1500 g and 4^{o}C. The pellets containing the albumin-depleted plasma were formed and kept, while the supernatants which mostly included TCA-albumin complex were discarded. Afterwards, the pellets were washed once with 200 µL of methanol by manual shaking and centrifuging for 5 minutes at 2000 g and 4^{o}C. At last, the supernatants were again dismissed and the remaining pellets were resuspended in 20 µL of urea 6M. Depleted and non-depleted plasma pool profiles were evaluated by SDS-PAGE and silver nitrate staining, using a similar amount of total protein content in each of the wells provided.

After plasma depletion, 10 µL of each of the reconstituted pellets (around 60-80 µg of protein content) were individually digested. For this aim, 30 µL of water and 20 µL of Dissolution Buffer were added to 10 µL of each depleted plasma pool. 30 µL of each resulting pool solution (around 30-40 µg of protein content) were trypsin-digested following the iTRAQ kit procedure, using the cited kit's reagents. Firstly, 2 µL of the Reducing Agent were added to each pool solution and the mixtures were incubated for 1 hour at 60^{o}C with continuous shaking. Secondly, the solutions were incubated with 1 µL of Cysteine Blocking for 10 minutes at room temperature. At last, samples were digested with Trypsin Promega at a rate 1:30–1:40 (µg trypsin : µg total protein) at 37^{o}C overnight. After verifying the digestions were successfully carried out by trypsin using MALDI-TOF analysis, the iTRAQ labelling procedure was immediately performed.

1.1.1.2. iTRAQ labelling.

iTRAQ labelling reagents were prepared as specified in the kit procedure. Each of the 8 labels were added to each of the 8 digested pools. Labels 113, 114, 115, 116 correspond to 4 of the pools made out of 40 HA-RA plasma samples; whereas labels 117, 118, 119 and 121 represent 4 of the pools depicting the other

40 LA-RA plasma samples. After adjusting the pH to 7.5 with Dissolution Buffer, incubating them individually for 3 hours at room temperature and lowering the organic content of each labelled pool with 100 µL of water (in order to stop the labelling procedure), the 8 labelled pools were combined. The mixture was then separated in two aliquots (with around 140 µg protein content each) and then dried in a speedvac. Each half was resuspended in 150 µL TFA 0.5% and desalted with 3 homemade StageTips, each one made up from 5 C18 Empore discs, following an internal tuned up protocol. Once the samples were pooled, albumin-depleted, trypsin-digested, iTRAQ labelled, combined and desalted, two different fractionation and LC-MS/MS strategies were followed for later comparison. A scheme of the pathway followed in this phase can be seen in **Figure 1. 1**.

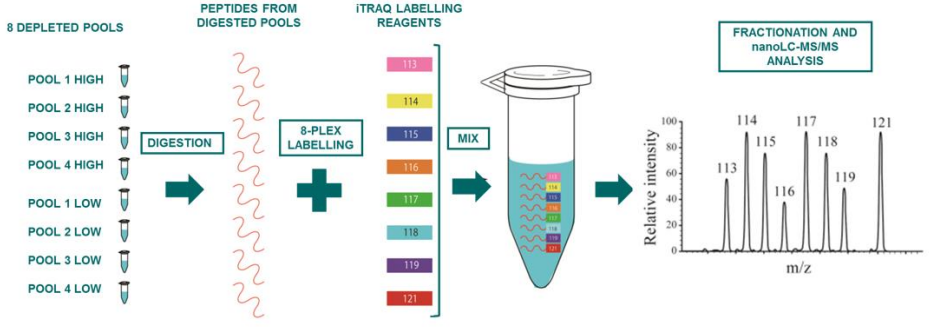


Figure 1. 1. Sample preparation and analysis scheme followed in the discovery phase.

1.1.1.3. Fractionation and LC-MS/MS analysis.

On the one side, a primary off-line reversed phase chromatography at basic pH (pH 10) step was conducted using one of the cleaned aliquots (approximately 140 µg protein content). Dried labeled peptides were resuspended in 140 µL of buffer A (10 mM ammonium hydroxide, 5% acetonitrile) and 130 µL were injected into an HP 1200 system (Agilent). The separation was performed on a C18 reversed-phase column (Zorbax Extend C18, 100 × 2.1 mm id, 3.5 µm, 300Å; Agilent), employing a chromatographic gradient of 100 minutes at

a flow of 200 µL/min (**Table 1. 2**). Sixty fractions were obtained (1 fraction: 1.5 minutes, 300 µL) from minute 9, monitoring the protein content at λ= 214 nm, which measures peptide absorbance. The chromatogram was obtained using a UV detector at 214 nm. Sixty fractions were collected every 90 s using a Gilson FC203B fraction collector (Gilson, Middleton, WI, USA). The cited 60 fractions were combined post-collection based on the peak intensity of the UV trace (protein content) and attending to the % of ACN needed for their elution in the following LC-MS/MS experiment. Thirty new fractions were obtained, for lowering the complexity of the initial sample, individually dried in a vacuum concentrator and resuspended in 10 µL of mobile phase A (2% ACN, 0.1% TFA) and/or stored at −20 °C for the next step of analysis.

Table 1. 2. Chromatographic gradient using HPLC with Zorbax column.

Gradient (minutes)	Buffer A (5% ACN, 10mM NH_4OH)	Buffer B (90% ACN, 10mM NH_4OH)
0	100	0
10	100	0
70	40	60
72	0	100
74	0	100
78	100	0
100	100	0

The final 30 peptide fractions were secondary separated offline using reversed phase chromatography in a nanoLC system (Tempo, Eksigent, Dublin, CA, USA). Peptides were desalted for 10 min and loaded onto a C18 column (Integrafit C18, Proteopep II, 75 µm id, 10.2 cm, 5 µm, 300 Å; New Objective). Peptides were eluted at a flow rate of 0.35 µL/min during a 1 hour-linear gradient using mobile phase B with different ACN % in 0.1% TFA. The reversed-phase fractions were collected and mixed with matrix α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich) at 4 mg/mL 70% ACN in 0.1% TFA at a flow rate of 1.2 µL/min, using the Sun Collect MALDI Spotter/Micro Collector (SunChrom Wissenschaftliche Geräte GmbH, Germany). Fractions collected every 15 s were then spotted onto a matrix-assisted

laser desorption/ionization (MALDI) plate for MS analysis. Chromatograms corresponding to each of the 30 fractions were composed by 240 spots.

MS data acquisition was accomplished for each chromatogram in a positive ion mode in a MALDI-TOF/TOF instrument (4800 ABSciex, Framingham, (MA), USA) using the 4000 Series Explorer software version 3.5.1 (ABSciex). MS spectra from m/z 800−4000 were acquired for each fraction using 1500 laser shots processed with internal calibration. We used 3 fmol/spot of angiotensin, diluted in the matrix, as an internal standard (m/z = 1046.50), and a laser intensity of 3800. After screening of all LC-MALDI sample positions in MS positive reflector mode, the fragmentation of automatically selected precursors was performed at a collision energy of 1 kV with CID gas (air). Up to 12 of the most intense ion signals per spot position with signal/noise ratio (S/N) above 80 were selected as precursors for MS/MS acquisition, excluding common trypsin autolysis peaks and matrix ion signals. The number of shots was 1800 for MS/MS, and the laser intensity was set to 4700. A second MS/MS spectra were acquired excluding the precursors selected in the previous MS/MS run. Precursors with S/N > 30 were selected to identify low abundant proteins that were not identified in the previous run, using a fixed laser intensity of 3800 kV and 1500 shots/spectrum.

On the other side, the first fractionation was made using around 100 µg protein content of the second cleaned aliquot. In this case, the HPLC Dionex Ultimate 300 with the column Waters Xbridge C18 5 µm (4.6 x 150 mm) was used, employing a chromatographic run of 95 minutes at a 150µL/min flow rate (**Table 1. 3)**. Thirty fractions were collected (1 fraction: 2.5 min, 375 µL) from minute 10 and combined in 5 final interleaved fractions (i.e. final fraction 1 was made out of fractions 1, 6, 11, 16, 21 and 26; final fraction 2 was composed of fractions 2, 7, 12, 17, 22 and 27; …). Half of the content from each individual fraction was used for getting 5 final fractions, which were were individually dried in a speedvac.

Table 1. 3. Chromatographic gradient using HPLC Dionex Ultimate 300 and Waters Xbridge column.

Gradient (minutes)	Solvent A (10mM NH_4OH, pH 9.3)	Solvent B (80% MeOH, 10mM NH_4OH, pH 9.3)
0	98	2
10	95	5
20	75	25
65	30	70
70	0	100
75	0	100
77	98	2
95	98	2

A 2-µg aliquot of each fraction was subjected to 1D-nano LC ESI-MSMS analysis using a nano liquid chromatography system (Eksigent Technologies nanoLC Ultra 1D plus, SCIEX, Foster City, CA) coupled to a high-speed Triple TOF 5600 mass spectrometer (SCIEX, Foster City, CA) with a Nanospray III source. The analytical column used was a silica-based reversed phase Acquity UPLC M-Class Peptide BEH C18 Column, 75 µm × 150 mm, 1.7 µm particle size and 130 Å pore size (Waters). The trap column was a C18 Acclaim PepMapTM 100 (Thermo Scientific), 100 µm × 2 cm, 5 µm particle diameter, 100 Å pore size, switched on-line with the analytical column. The loading pump delivered a solution of 0.1% formic acid in water at 2 µl/min. The nano-pump provided a flow-rate of 250 nl/min and was operated under gradient elution conditions. Peptides were separated using a 150 minutes gradient ranging from 2% to 90% mobile phase B (mobile phase A: 2% acetonitrile, 0.1% formic acid; mobile phase B: 100% acetonitrile, 0.1% formic acid). Injection volume was 5 µl.

Data acquisition was performed with a TripleTOF 5600 System (SCIEX, Foster City, CA). Data was acquired using an ion spray voltage floating (ISVF) 2300 V, curtain gas (CUR) 35, interface heater temperature (IHT) 150, ion source gas 1 (GS1) 25, declustering potential (DP) 100 V. All data was acquired using information-dependent acquisition (IDA) mode with Analyst TF 1.7 software

(SCIEX, USA). For IDA parameters, 0.25 s MS survey scans (mass range of 350–1250 Da) were followed by 35 MS/MS scans of 100 ms (mass range of 100–1800, total cycle time was 4 s). Switching criteria were: ion m/z greater than 350 and smaller than 1250, with charge state of 2–5 and an abundance threshold of more than 90 counts (cps). Former target ions were excluded for 15s. IDA rolling collision energy (CE) parameters script was used for automatically controlling the CE. The mass spectrometry data obtained were processed using PeakView® 2.2 Software (SCIEX, Foster City, CA) and exported as mgf files.

1.1.1.4. Mass Spectrometry Data Analysis

The iTRAQ 8-plex data collected from both LC-MS/MS analysis was interpreted using the ProteinPilot 4.0 software (ABSciex). The Paragon algorithm in ProteinPilot software served as the default search program with trypsin as the digestion agent and MMTS as a fixed modification of cysteine. Biological modifications were programmed in the algorithm. Each MS/MS spectrum was searched in the Uniprot/Swissprot database (downloaded in 2014) for Homo sapiens. Only proteins identified with at least 95% confidence, or a Prot Score (protein confidence measure) of at least 1.3 were reported. The ProteinPilot software also calculated a confidence percentage, the Unused score, which reflects the probability of a hit being a “false positive,” at a 95% confidence. While this software automatically accepts all peptides with an identification level >1%, only proteins having at least one peptide above the 95% confidence level were initially recorded. Searches against a concatenated database containing both forward and reversed sequences allowed the false discovery rate to be kept at 1%. The results obtained from ProteinPilot were exported to Microsoft Excel for further analysis. Statistical analysis using t-test was performed by using the MultiExperimet Viewer 4.9 software. Proteins differentially detected and relatively quantified between the HA-RA labelled pools and the LA-RA labelled pools, with a difference in their ratios greater than 0.3 and with high significance

($p < 0.05$), were selected as possible candidates to be confirmed in the following verification phase.

1.1.2. Verification Phase: Confirmation of protein biomarker candidates differential expression by multiplex targeted mass spectrometry (MRM) using SIS peptides for semi-absolute quantitation.

Having selected the protein candidates from the previous discovery phase, which could have an outstanding paper for the procurement of a panel of possible RA activity biomarkers, an individual plasma sample analysis was performed using targeted proteomics. For this purpose, tryptic unique peptides from the chosen candidate proteins were selected based on the signals and intensities measured by MRM on a QTRAP 5500. Stable Isotope-labeled Standards (SIS) were produced in order to semi-quantitatively prove the previous results on individual plasma samples. The samples selected for this stage were the same as the ones used in the discovery phase: 80 plasma samples (40 from patients with HA-RA and 40 from patients with LA-RA).

1.1.2.1. MRM method configuration and refinement.

From the proposed panel of protein candidates, peptide databases from the GPMdb and peptides reported on the discovery phase data were consulted in order to generate a set of tryptic peptides suitable for the proteomic analysis by MRM. All the peptides were chosen on a basis of being unique from the proteins they proceed, tryptic, with an acceptable signal and MRM peak shape and preferable best candidates for MRM as defined in the GPMdb. Not only were the tryptic peptides studied but also their most suitable charges and transitions for their monitoring by MRM. These measurements were performed on a QTRAP 5500 by injecting 2 µL of plasma digest (around 1 µg of protein content) under the conditions explained later for MRM analysis**.** After electing the most suitable

tryptic fragments, SIS peptides (crude purity) were ordered and purchased from JPG (Germany).

1.1.2.2. Isotopically Labelled Synthetic Peptides.

The SIS peptides contained different stable isotope labeled amino acids, depending on the tryptic end of each SIS. The modifications carry ^{13}C and ^{15}N isotopes, leading to different mass shifts on dependence of the labeled amino acid. For Lysine, K, ($^{13}C_6$,$^{15}N_2$-Lys), a +8 Da mass shift was added and for Arginine, R, ($^{13}C_6$,$^{15}N_4$-Arg), a +10 Da mass shift. With the aim of verifying the signal of the SIS, an equimolar amount of each SIS at 100 fmol/µL was combined, in order to make an initial stock mixture of SIS peptides and analyze it by targeted proteomics (MRM). Even though all the SIS had the same concentration, the signals and intensities observed differed for each SIS peptide. Furthermore, due to the different concentrations and signals of the endogenous (END) targeted peptides in the measured plasma samples. SIS peptides were individually adjusted by concentration in a final stock mixture, correlating their signals with those of the endogenous peptides. Balancing the END/SIS ratios helps to reduce the analytical variability in MRM measurements. Equilibration of each SIS concentration was based on the relative END/SIS responses obtained from the LC−MRM-MS analysis of different SIS mixtures spiked in a plasma tryptic digest. Aliquots of the final SIS stock mixture were prepared and stored at −80°C until their use, minimizing the number of freeze−thaw cycles to which the peptides are subjected, in order to reduce the prompt SIS peptide degradation.

1.1.2.3. Sample preparation.

At first, 2 µL of each of the 80 plasma samples with the extreme RA activities were diluted to a final volume of 20 µL with urea 6M thiourea 2M. The solution was shaken at room temperature for 1 hour, leaving enough time to

accomplish protein denaturation. Then, 2 µL of this solution (around 10 µg of protein content) was pulled apart in order to conduct in solution trypsin digestions of the 80 plasma samples. For this intention, 3 µL of Ammonium Bicarbonate 25 mM and 1.25 µL of DTT 50 mM were affixed to the previous 2 µL of diluted plasma. The mixture was incubated for 1 hour at 37^{o}C, aiding the reduction of the disulfide bonds between proteins. Subsequently, 6.25 µL of IA 100 mM were added to the mixture and incubated for 45 minutes at room temperature in the absence of any light, blocking free sulfhydryl groups present in cysteines. At last, pH was adjusted to 7 with 10 µL of AmBic 25 mM and trypsin (Promega) was added at a ratio (1 µg trypsin : 20 µg protein content). Plasma digestion was attained overnight at 37^{o}C. After this period, trypsin activity was defused by acidifying the solution with 10 µL TFA 2%. In order to quantify the elected peptide levels in the plasma digests, 10 µL of the final SIS stock mixture were added to each sample and a desalting procedure using homemade StageTips (3M Empore SPE-C18 disk, 47 mm, Sigma Aldrich) was conducted. END and SIS peptides were eluted and evaporated until dryness in a speedvac (Thermo, USA) for 6 hours. Each sample was then resuspended in 20 µL of mobile phase (98% H_2O, 2% ACN, 0.1% FA) for later MRM method validation and MRM analysis.

1.1.2.4. MRM method validation.

Reconstituted SIS peptides and plasma tryptic digests were analyzed by LC-MS/MS in a nanoLC system (TEMPO) coupled to a 5500-QTRAP instrument (Sciex). For this aim, these digest solutions were first injected at a 3 µL/min flow during 10 minutes and desalted using a C18 column (5 µm, 300 A, 100 µm * 2 cm, Acclaim PepMap, Thermo Scientific, USA). Subsequently, the tryptic digests were separated on C18 nanocolumns (75 µm inner diameter, 15 cm, Acclaim PepMap 100, Thermo Scientific, USA) at a flow rate of 300 nL/min. The gradient showed in **Table 1. 4** was selected for MRM analysis after testing different gradients. The 5500-QTRAP mass spectrometer was interfaced with a nano-spray source

equipped with an uncoated fused silica emitter tip (20 µm inner diameter, 10 µm tip; New Objective, Woburn, MA) and operated in the positive ion mode. MS source parameters were set as follows: IS 2,600 V, IHT 150°C, GS2= 0, CUR= 20, GS1= 25 psi and high CAD. MS compound parameters were set to 10 for the EP and to 15 for CXP. Skyline software was used to predict and optimize collision energies for each peptide, both END and SIS. Q1 and Q3 were set to 0.7 Da unit/unit resolution and the pause between mass ranges was set to 3 ms. In order to confirm the identity of the peptides, an MRM Information Dependent Acquisition (IDA) experiment was performed for each peptide. The mass spectrometer was instructed to switch from MRM to EPI scanning mode when an individual MRM signal exceeded 1,000 counts. Each precursor was fragmented a maximum of twice before being excluded for 10s and the masses were scanned from 250 to 1,000 Da. The rolling collision energy (CE) option was employed to automatically ramp up the CE value in the collision cell as the m/z value increased.

Table 1. 4. Chromatographic gradient selected for nanoLC-ESI-QTRAP analysis.

Gradient (minutes)	% Solvent A (5% ACN in 0.1% FA)	% Solvent B (95% ACN in 0.1% FA)
0	95	5
5	85	15
45	65	35
46	5	95
56	5	95
57	95	5
70	95	5

An interference screening was carried out based on the fragment-ion ratios of the transitions and using the exported responses from the SIS and END peptides. The final SIS stock solution alone in buffer was also analyzed to confirm that there was no END peptide signal due to contamination from the SIS peptide synthesis. Once the END/SIS responses were evaluated, the MRM method was validated by producing multiplexed standard curves for each peptide by triplicate. 13 points curves were produced by maintaining the digested pool amount at 1µg

of protein content (constant END) and varying the SIS concentration (from 0 to 1200 fmol/µL). The linear regression, sensitivity, dynamic range, LLOQ and ULOQ of each calibration curve were calculated using the online software "QUALIS-SIS: Automatic Standard Curve Generation for Multiplexed MRM Data Analysis". Additionally, the LLOD was calculated for each peptide using the following formula: LLOD= 3/10 LLOQ.

1.1.2.5. MRM analysis.

A targeted proteomic analysis of the previously chosen peptides was conducted in the 80 plasma samples. For this aim, 2 µL of each plasma digest (around 1 µg of protein content), from the last suspension in 20 µL of buffer A (5% ACN in 0.1% FA), were injected and analyzed by nanoLC-QTRAP using the conditions detailed in the previous MRM method validation section. Two subsequent injections were acquired for each sample followed by two different blanks. A first blank of ACN and second consecutive blank of buffer A were run between each sample injection to avoid carry over in the nanoLC system. Data analysis was accomplished using Skyline software for method refinement, optimization and peak integration. Raw files were imported to Skyline and integration was manually inspected to ensure correct peak detection and accurate integration. Coelution of all the transitions of the target or END peptide with the SIS peptide was considered as accurate MRM signal. The Savitzky-Golay method was applied to smooth the data points. NAT/SIS peak area ratios were used to calculate the relative peak area ratio of each peptide. Peak area ratio reports were exported from Skyline to give further information of each measured peptide replicate as: retention time (RT), mean, standard deviation, coefficient of variation (% CV), fold change and p-values comparing HA-RA and LA-RA samples. Quantitation was conducted by single point measurements. Additionally, two-tailed Mann-Whitney U-test and box plot graphs were performed using GraphPad Prism 5.01 software (La Jolla, CA, USA). A p-value <0.05 was considered

statistically significant. Conversion of END/SIS ratios to µg of protein/µL plasma was calculated for each significant peptide using the next equations:

fmol END/µL plasma= (END/SIS ratio) * ($[SIS]_{injected}$) * (2µL injected) * (10: dilution factor)^2 / (2µL plasma)

µg protein/µL plasma= (END/SIS ratio) * ([SIS] final SIS stock solution) * (CF:50) * (Protein MW: Da) / (10^9: fg->µg)

Where:

$[SIS]_{injected}$ = ($[SIS]_{final\ SIS\ stock\ solution}$) * (10µL final SIS stock solution added) / (20µL buffer A resuspension)

Conversion factor (CF): (10µL final SIS stock solution added) / (20µL buffer A resuspension) * (2µL injected) * (10: dilution factor)^2 / (2µL plasma) = 50

1.1.3. Validation Phase: proof of the verified candidate proteins as potential biomarkers for clinical use by individual immuno-based strategies (ELISA).

In order to validate and confirm the diagnostic paper of the selected protein candidates from the previous verification stage, a larger number of plasma samples from patients with extreme RA activities, and patients with analogous disabilities were selected from the same IMIDs cohort as disease controls. Furthermore, plasma samples from healthy donors were also selected.

1.1.3.1. ELISA set up and analysis

ELISA development kits were purchased from Bio-Techne (DuoSet, Minneapolis, USA) and Sino Biological (Matched ELISA Pair Set, Beijing, China).

Full-matching the immunogen sequence of the antibodies in the development kits with the significant peptide fragments analyzed by MRM was a requirement for kit selection. Additionally, TMB Substrate Solution was acquired from ThermoFisher (Massachusetts, USA). H_2SO_4 1N was used as Stop Solution. Optical density measurements were assessed at 450nm, using the Infinite M200 Nanoquant plate reader (Tecan, Switzerland). Wavelength correction for optical imperfections in the plate was set to 550nm, as readings made directly at 450 nm without correction may be higher and less accurate.

Sample dilution tests were performed as detailed in each specific ELISA kit, using samples with the lowest and highest concentration measured by MRM. Two samples of each concentration range were chosen in order to cover the dynamic range of each protein in plasma samples, and obtain the optimal sample dilution for the following analysis. Once the sample dilution and the standard curve range were set up for the analysis, a previous verification test was conducted, with the aim of corroborating the ELISA (immunoaffinity based) measurements with the MRM ones (MS based). Twenty samples, previously measured by MRM, were randomly selected and analyzed by ELISA. Protein concentration trends, grouped by disease activity, were compared between MRM and ELISA analysis to support the usage of the developed ELISA assays in the validation phase.

420 plasma samples were analyzed using each developed ELISAs, among them: 170 samples from patients with RA; 160 samples from patients with SLE (80) and PsA (80), both with extreme disease activities as well; and 90 samples from healthy donors without notice of suffering any of the related disorders. Sample concentrations were calculated based on each plate's calibration curve, employing the most adequate regression for each curve with GraphPad Prism software. Moreover, a RA plasma pool, constituted by 10 arbitrary plasma samples of patients with RA were measured in three replicate(s) per plate. Intra-

plate and inter-plate variabilities were determined using the calculated pool concentrations in each plate, accepting CVs lower than 20%. Furthermore, inter-plate sample concentrations were corrected by using a normalization factor (n). This factor was calculated for every plate analyzed by dividing the protein concentration's median of all plates against the protein concentration median of each single plate. Thence, each plate's protein concentration was corrected by multiplying each plate's normalization factor by each plate's protein concentration value.

1.2. Results.

1.2.1. Discovery phase.

After completing the analysis with the two different fractionations and LC-MS/MS strategies, a comparison of the identified and modulated proteins and peptides was accomplished. Identifications were done at 95% confidence level using the ProteinPilot software, after stablishing different analysis conditions as the sample description and the data processing. For instance, MMTS alkylation, trypsin digestion, urea denaturation and Homo Sapiens species assumptions were firstly assigned for sample description. Furthermore, bias and background correction, biological modifications ID focus, uniport_sprot2014 database with a through ID search effort and a detected protein threshold (ProtScore) higher than 1.3 (which corresponds to a 95% confidence level) were also launched for data processing. Tag normalizations were done against the 113 labeled condition in order to obtain the different relative ratios between the 8 labeled pools for relative quantitation. As it can be seen in **Table 1. 5,** 184 proteins were detected in the analysis of the 30 fractions by MALDI-TOF/TOF; whereas only 164 proteins were detected when analyzing less fractions (5) by ESI-tTOF. The same happens with the proteins before grouping (270 vs 213), these include all detected proteins and their corresponding isoforms. Nevertheless, the number of distinct peptides and spectra identified were higher when working with ESI-tTOF, thus achieving a higher coverage of the proteins identified with this methodology (37097 vs 10316). Contrarily, the percentage of total spectra used for the protein identification was higher with MALDI-TOF/TOF than with ESI-tTOF (64.3% vs 22.2%).

Table 1. 5. Comparison of the two LC-MS/MS methodologies followed for protein and peptide identification.

Fractions obtained / analyzed	LC-MS/MS	Proteins Detected	Proteins Before Grouping	Distinct Peptides	Spectra Identified	% Total Spectra
60 / 30	nanoLC (offline) MALDI-TOF/TOF	184	270	5699	10316	64.3
30 / 5	nanoLC (online) ESI-tTOF	164	213	11352	37097	22.2

Additionally, a first comparison of the number of common proteins identified was executed using the Venny 2.1 online software. As it can be seen in the Venn diagram (**Figure 1. 2**), 149 proteins were prevalent in both approaches, 15 proteins were exclusively identified by ESI-tTOF and 35 by MALDI-TOF/TOF. In total, taking into account both methods, 199 proteins were identified. These three protein lists are presented in **Annex I**. Most of the proteins determined just by ESI-tTOF are immunoglobulins (9 out of 15), whereas the proteins determined merely by MALDI-TOF/TOF (35) were more diverse.

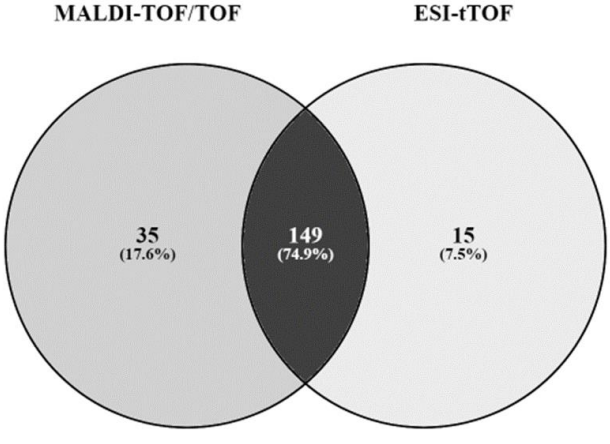


Figure 1. 2. Number of proteins commonly and exclusively identified using the two LC-MS/MS methodologies tested.

In order to carry out the differential study of the iTRAQ-labeled pools, the MultiExperimentViewer software was used. Statistical analysis was performed in order to discover which of the 199 proteins were significantly modulated. Relative quantitation was achieved by considering the iTRAQ ratios of the 8 labeled digested pools. As 4 of the labeled conditions appertained to the HA-RA group and the other 4 to the LA-RA group, average iTRAQ ratios for both groups were calculated for each group and protein. A final list of 11 proteins were found to be significantly altered between the two groups (p-value < 0.05) with a cutoff difference of 0.3 minimum among them and a suitable molecular function (**Table 1. 8**). This protein panel was defined from comparing the two modulated protein lists generated after evaluating the two LC-MS/MS outcomes (**Table 1. 6** and **Table 1. 7**). All these proteins were conveniently related with the RA process and its effects (acute phase response, inflammatory response, immune response, bone mineralization, cytokine secretion, osteoblast differentiation, anti-inflammatory properties, …). Since this fact, they became biomarker candidates for RA activity monitoring which needed to be further assessed. Thus, the following verification phase was carried out.

Table 1. 6. Modulated proteins found by nanoLC-MALDI-TOF/TOF.

LC-MALDI-TOF/TOF (6 proteins)					
Protein	**Number of Peptides**	**HA-RA**	**LA-RA**	**p-value**	**HA-LA difference**
Haptoglobin	**73**	**1,18**	**0,69**	**0,01**	**0,49**
Kininogen-1	**31**	**1,07**	**1,57**	**0,02**	**-0,50**
Afamin	**9**	**0,93**	**1,36**	**0,02**	**-0,43**
Alpha-2-HS-glycoprotein	**20**	**1,20**	**1,80**	**0,02**	**-0,60**
Selenoprotein P	*2*	*0,94*	*1,06*	*0,03*	*-0,12*
Histidine-rich glycoprotein	**23**	**0,97**	**1,29**	**0,03**	**-0,31**
Ig lambda chain V-II region BUR	*1*	*0,88*	*1,23*	*0,04*	*-0,36*
Cholesteryl ester transfer protein	*2*	*0,98*	*1,13*	*0,04*	*-0,16*
Serum amyloid A-1 protein	**4**	**0,82**	**0,50**	**0,05**	**0,32**
Ig kappa chain V-I region HK101 (Fragment)	*10*	*0,97*	*0,78*	*0,05*	*0,20*

Table 1. 7. Modulated proteins found by nanoLC-ESI-tTOF.

LC-ESI-tTOF (7 proteins)					
Protein	**Number of Peptides**	**HA-RA**	**LA-RA**	**p-value**	**HA-LA difference**
Insulin-like growth factor-binding protein 3	**2**	**0,99**	**1,32**	**0,01**	-0,33
Alpha-1-acid glycoprotein 1	**14**	**0,98**	**0,37**	**0,02**	0,61
Hyaluronan-binding protein 2	**1**	**0,97**	**0,65**	**0,02**	0,32
Alpha-2-antiplasmin	*9*	*0,84*	*1,17*	*0,03*	*-0,33*
Alpha-1-antichymotrypsin	**24**	**0,83**	**0,45**	**0,03**	0,38
Alpha-1-acid glycoprotein 2	**8**	**1,15**	**0,71**	**0,04**	0,44
Ig kappa chain V-I region BAN	*2*	*0,97*	*0,78*	*0,04*	*0,18*
Alpha-2-HS-glycoprotein	**14**	**1,16**	**1,56**	**0,05**	-0,39
Complement C4-A	*81*	*0,94*	*1,13*	*0,05*	*-0,19*
Serum amyloid A-1 protein	**1**	**0,81**	**0,30**	**0,05**	0,50

Table 1. 8. Modulated proteins found after comparing the two LC-MS/MS strategies.

Protein	**HA-RA**	**LA-RA**	**p-value**	**HA-LA**
Afamin	0,93	1,36	0,02	-0,43
Alpha-1-acid glycoprotein 1	0,98	0,37	0,02	0,61
Alpha-1-acid glycoprotein 2	1,15	0,71	0,04	0,44
Alpha-1-antichymotrypsin	0,83	0,45	0,03	0,38
Alpha-2-HS-glycoprotein	1,20	1,80	0,02	-0,60
Haptoglobin	1,18	0,69	0,01	0,49
Histidine-rich glycoprotein	0,97	1,29	0,03	-0,31
Hyaluronan-binding protein 2	0,97	0,65	0,02	0,32
Insulin-like growth factor-binding protein 3	0,99	1,32	0,01	-0,33
Kininogen-1	1,07	1,57	0,02	-0,50
Serum amyloid A-1 protein	0,81	0,30	0,05	0,50

1.2.2. Verification phase.

In order to obtain a final MRM method for monitoring the 11 proteins (chosen from the previous phase), a large study for the selection of their corresponding unique peptides was conducted. Two MRM methods were firstly tested by MRM on plasma digests as explained in the "MRM method configuration and refinement" section. Each method contained 68 tryptic peptides, while each peptide was tracked with different charges (normally charge 2 and 3, but also with charge 1 if needed). Furthermore, at least 5 transitions were scanned per precursor. After several injections to confirm the peptides responses, a final MRM method with 11 proteins, 26 peptides, 52 precursors and 152 transitions was elected. Most of the proteins were sufficiently covered and/or defined by at least 2 peptides, except for SAA1. In this specific case, there was only one peptide that could fulfill the MRM requirements of peak intensity and/or shape definition. Some proteins needed to be defined by three peptides when one of the peptides partially satisfied the named demands. A1AG1 and A1AG2 are proteins of the same family, hence they share a large part of their sequence and there are only a couple of tryptic peptides that can distinguish among both proteins. Due to this matter, 3 peptides were also selected for monitoring and discriminating these two proteins. Twenty-six analogous SIS peptides were purchased with the objective of verifying the tendency of the proteins observed in the preceding phase.

The signals of the commercial SIS peptides were tested by MRM and their concentrations were balanced all together to secure a good END/SIS output. This was done by spiking different SIS mixtures in tryptic digests and measuring them using MRM. Once settled, the concentration of each SIS peptide added to the final stock mixture varied among 25, 50, 100 and 250 fmol/µL, as indicated in **Table 1. 9**. Calibration curves using the 13 points data from the triplicates (except for the blank) where analyzed by the Qualis-SIS software as presented in **Table 1. 9**.

Table 1. 9. SIS peptide concentration for later analysis of each peptide measured conjointly by MRM and calibration curve regression, LOD, LOQ, curve points and dynamic range data using the QUALIS-SIS software.

PROTEIN GENE	PEPTIDE SEQUENCE	[SIS] (fmol/µL)	QUALIS-SIS						
			Regression	R^2	LLOD	LLOQ	ULOQ	N points	Dynamic range (orders)
SAA1	SFFSFLGEAFDGAR	100	y=0.0827x-1.69	0.984 (<20%)	7,5	25	1200	6	>48 (2)
AFM	ESLLNHFLYEVAR	100	y=0.112x-0.806	0.983 (<20%)	3	10	1200	7	>120 (2)
	DADPDTFFAK	25	y=0.169x-0.0062	0.988 (<20%)	1,5	5	1200	8	>240 (3)
	FTFEYSR	25	y=0.153x-0.317	0.991 (<20%)	1,5	5	1200	8	>240 (3)
AHSG	EHAVEGDCDFQLLK	250	y=0.00472x+0.0915	0.969 (<20%)	3	10	1200	7	>120 (2)
	FSVVYAK	250	y=0.0186x-0.0025	0.990 (<20%)	3	10	1200	7	>120 (2)
HRG	GGEGTGYFVDFSVR	250	y=0.0313x-0.182	0.984 (<20%)	3	10	1200	7	>120 (2)
	ADLFYDVEALDLESPK	100	y=0.0668x-0.348	0.986 (<20%)	3	10	1200	7	>120 (2)
	DGYLFQLLR	50	y=0.0803x-0.191	0.983 (<20%)	1,5	5	1200	8	>240 (3)
HABP2	FTCACPDQFK	250	-	-	-	-	-	-	-
	VVLGDQDLK	25	-	-	-	-	-	-	-
IGFBP3	YGQPLPGYTTK	25	-	-	-	-	-	-	-
	FLNVLSPR	25	-	-	-	-	-	-	-
KNG1	YNSQNQSNNQFVLYR	250	y=0.00228x-0.00735	0,97 (<20%)	15	50	1200	5	>24 (2)
	TVGSDTFYSFK	100	-	-	-	-	-	-	-
	YFIDFVAR	100	y=0.0371x-0.103	0.979 (<20%)	1,5	5	1200	8	>240 (3)
A1AG1	EQLGEFYEALDCLR	250	y=0.0052x-0.0255	0.985 (<20%)	3	10	1200	7	> 120 (2)
	SDVVYTDWK	250	y=0.0023x+0.015	0.839 (25%)	0,3	1	600	8	600 (2)
	YVGGQEHFAHLLILR	250	y=0.0072x-0.0686	0.994 (<20%)	7,5	25	1200	6	> 48 (2)
A1AG2	EHVAHLLFLR	250	y=0.013x-0.226	0.997 (<20%)	1,5	5	1200	8	> 240 (3)
	EQLGEFYEALDCLCIPR	250	y=0.0030x-0.0308	0.991 (<20%)	7,5	25	1200	6	> 48 (2)
	SDVMYTDWK	250	y=0.0072x+0.0087	0.989 (<20%)	1,5	5	1200	8	> 240 (3)
AACT	AVLDVFEEGTEASAATAVK	250	y=0.0017-0.0023	0.976 (<20%)	7,5	25	1200	6	> 48 (2)
	EIGELYLPK	250	y=0.0255+0.0357	0.993 (<20%)	3	10	1200	7	> 120 (2)
HPT	VGYVSGWGR	250	y=0.0038x+0.0007	0.995 (<20%)	1,5	5	1200	8	> 240 (3)
	VTSIQDWVQK	250	y=0.0038-0.0027	0.993 (<20%)	3	10	1200	7	> 120 (2)

Most of the peptides measured comprised LLOQs between 5 and 25 fmol/µL, aside from one of the KNG1 peptides with a LLOQ of 50 fmol/µL. All ULOQs were 1200 fmol/µL, satisfying the linear regression conditions (R^2> 0.97), except for one of the A1AG1 peptides, whose ULOQ needed to be lowered down to 600 fmol/µL, still without conforming to the guidelines (CV<20%, R^2>0.95). The number of points suited in the most favorable curves were among 6 and 8, except for one of the KNG1 peptides, where only 5 points suited the curve specifications. All measured peptides operated between 2 and/or 3 orders of magnitude, along a dynamic range from 24 to 240 (at least). Nevertheless, there are 5 peptides which were not capable of fulfilling the software requirements (3 consecutive linear points and point triplicates with CVs less than 20% in accuracy and precision). HABP2 and IGFBP3 proteins were not qualified for MRM quantification, as all four measured peptides associated with them didn't behave suitably, after

examining the demands of the corresponding standard curves. The same situation takes place with one of the KNG1 peptides, although in this case, there are two more peptides operating well.

After accomplishing the MRM method validation for the 21 measured peptides, the 80 RA plasma samples were individually digested and analyzed in duplicates using the defined LC-MS/MS strategy. Only some punctual duplicates of the different peptide L/H ratios showed CVs higher than 20%, albeit in none of the cases CVs were higher than 30. Using the Skyline software, statistical analysis determined a significant difference between the HA-RA and the LA-RA groups in some of the peptides, as illustrated in **Table 1. 10**. In this table, the p-values of each peptide can be seen, as well as the fold change obtained by comparing the L/H ratios of the HA-RA group to the LA-RA one at 95% of confidence level and 78 degrees of freedom. Although the 5 significant peptides (p-values < 0.05) are highlighted, additional 4 peptides with a p-value close to significance (p-value: 0.05-0.1) were also appended. These last peptides corresponded to the same proteins that yield a p-value < 0.05, reinforcing their possible disease activity biomarker feature as a whole protein. SFFS peptide from SAA1 achieved the highest fold change (2.135) at the lowest p-value; while HPT attained a significant p-value with the VGYV peptide with the second highest fold change (1.489) in contrast to the other non-significant peptide with the third greatest fold change (1.386). Moreover, AACT, A1AG1 and A1AG2 showed significant and non-significant peptides with lower fold changes and roughly 1, meaning they display a slighter difference when performing disease activity diagnosis. **Figure 1. 3** shows the tendency of the cited peptides when comparing both RA disease activity status. Most of the samples are within the average subgroup range when comparing both RA activity categories for most of the significant peptides. Reversely, SFFS peptide from SAA1 exhibits a heterogeneous dispersion of the data, particularly when looking at the high activity subcategory.

Table 1. 10. Statistically significant peptides measured by MRM after the analysis of 80 plasma samples from patients with extreme RA activities.

Protein	Peptide	Fold Change (HA/LA)	p value
SAA1	**SFFSFLGEAFDGAR**	**2,135**	**0,009**
AACT	**AVLDVFEEGTEASAATAVK**	**1,239**	**0,011**
A1AG2	**EQLGEFYEALDCLCIPR**	**1,174**	**0,020**
A1AG1	**SDVVYTDWK**	**1,217**	**0,030**
HPT	**VGYVSGWGR**	**1,489**	**0,042**
AACT	EIGELYLPK	1,182	0,061
A1AG1	EQLGEFYEALDCLR	1,163	0,066
A1AG1	YVGGQEHFAHLLILR	1,207	0,073
HPT	VTSIQDWVQK	1,386	0,080

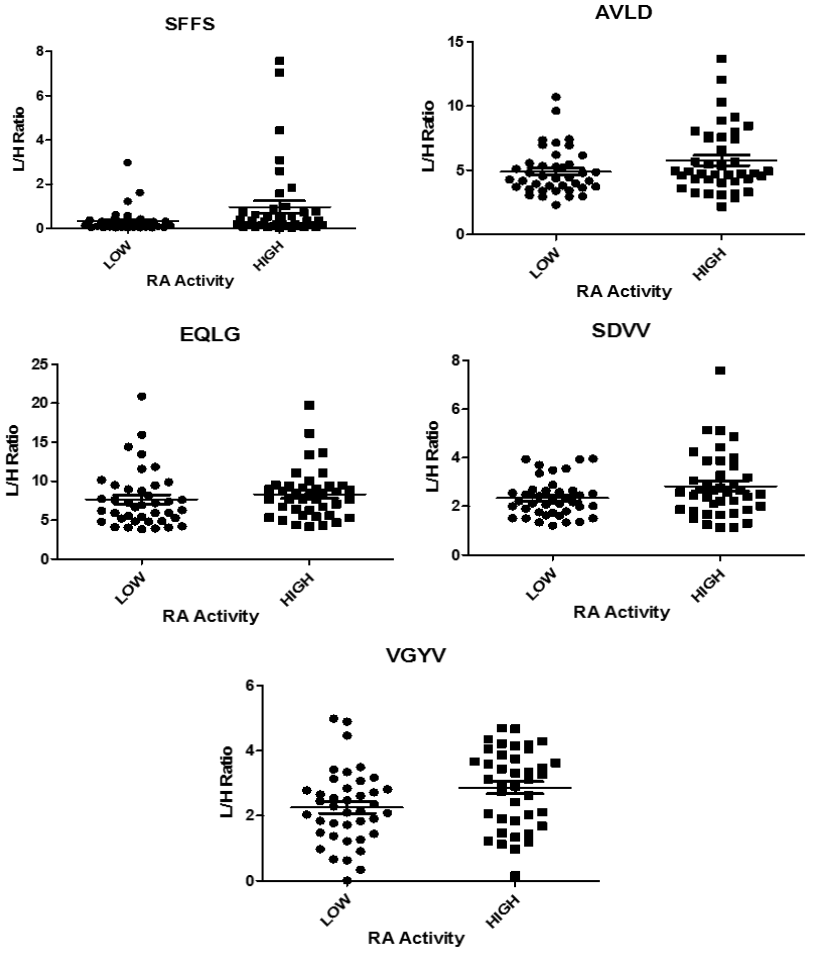


Figure 1. 3. MRM peptide modulation of the statistically significant peptides after analyzing the 80 plasma samples of patients with extreme RA activities.

Once the significant peptides were examined, protein concentrations were calculated using the L/H ratios by single point measurement with the prior formula. If looking deeper to the average protein concentration of the different peptides and conditions in **Table 1. 11** and **Figure 1. 4**, it can be noticed that most of the peptides which correspond to the same protein don't correlate in terms of determining the concentration of their matching protein. For instance, AACT SIS peptides were both added at the same concentration (250 fmol/µL), therefore they should measure a similar amount of the protein they appertain. In contrast, even if they prove to result in a similar fold change (AVLD FC: 1,18; EIGE FC: 1,19), the protein average concentrations for both of the activity subgroups are quite disparate among them (AACT concentration is around 21 times higher when measuring AVLD than EIGE). A similar trend can be foreseen when examining the A1AG1 peptide concentrations, where SDVV and EQLG demonstrate at least twice higher A1AG1 concentrations than YVGG. On the other hand, both HPT peptides manifest a quite coordinated protein concentration assessment when comparing the two activity categories.

Table 1. 11. Quantitation of the peptides measured by MRM with a p-value <0.1 and their correlation in terms of their corresponding absolute protein concentration and/or relative quantitation (FC: Fold Change) when discriminating RA patients with HA and LA.

Protein_Peptide	[SIS added] (fmol/µL)	L/H Ratio		Protein Average Concentration (ng/µL)		
		LA-RA	HA-RA	LA-RA	HA-RA	FC: HA/LA
SAA1_SFFS	100	0,33	0,97	22,17	65,43	2,95
AACT_AVLD	250	4,89	5,77	2914,18	3439,73	1,18
AACT_EIGE	250	0,23	0,28	139,73	166,65	1,19
A1AG2_EQLG	250	7,66	8,32	2260,93	2453,36	1,09
A1AG1_SDVV	250	2,33	2,83	685,79	831,69	1,21
A1AG1_EQLG	250	1,97	2,23	580,28	654,68	1,13
A1AG1_YVGG	250	0,96	1,07	283,14	313,99	1,11
HPT_VGYV	250	2,26	2,86	1274,88	1616,92	1,27
HPT_VTSI	250	2,35	3,02	1326,77	1704,13	1,28

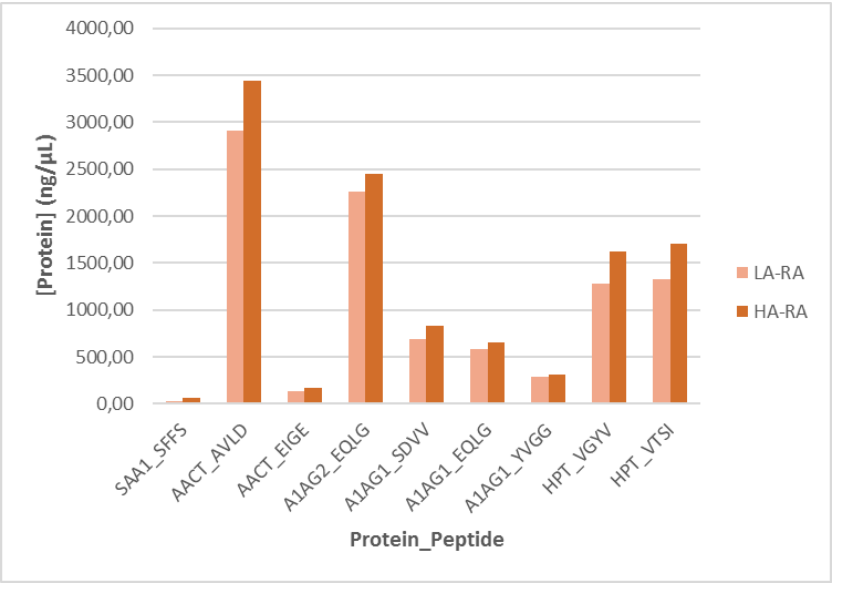


Figure 1. 4. Absolute concentration of each modulated peptide (p-value < 0.1) in the two different disease activity groups (LA-RA and HA-RA).

1.2.3. Validation phase.

After verifying the MRM peptidic results of the previous phase, 5 proteins were selected for conducting a more extensive study using a larger number of plasma samples. At this point, the already cited plasma samples from patients with analogous diseases (rheumatic IMIDs) were included, as well as healthy controls. In order to measure the concentration of each protein, DuoSet kits were purchased attending to the previously described requirements and set up for carrying out the immuno-analysis. Only 4 kits were selected (SAA1, AACT, HPT and A1AG) for measuring the previous 5 significant proteins. This decision was taken after considering the only availability of A1AG Duoset and the difficulty of distinguishing both A1AG1 and A1AG2 proteins using an immunoassay, as a result of their similar sequence. Around 89% of the protein sequence is common for both proteins as shown in **Figure 1. 5**, after performing a BLAST protein sequence alignment among them. Actually, there are only a few amino acids (21) and areas where A1AG1 (Query) and A1AG2 (Sbjct) differ.

Score	Expect	Method	Identities	Positives	Gaps
343 bits(880)	2e-127	Compositional matrix adjust.	162/183(89%)	172/183(93%)	0/183(0%)

```
Query  1    QIPLCANLVPVPITNATLDQITGKWFYIASAFRNEEYNKSVQEIQATFFYFTPNKTEDTI  60
            QIPLCANLVPVPITNATLD+ITGKWFYIASAFRNEEYNKSVQEIQATFFYFTPNKTEDTI
Sbjct  1    QIPLCANLVPVPITNATLDRITGKWFYIASAFRNEEYNKSVQEIQATFFYFTPNKTEDTI  60

Query  61   FLREYQTRQDQCIYNTTYLNVQRENGTISRYVGGQEHFAHLLILRDTKTYMLAFDVNDEK  120
            FLREYQTRQ+QC YN++YLNVQRENGT+SRY GG+EH AHLL LRDTKT M    ++DEK
Sbjct  61   FLREYQTRQNQCFYNSSYLNVQRENGTVSRYEGGREHVAHLLFLRDTKTLMFGSYLDDEK  120

Query  121  NWGLSVYADKPETTKEQLGEFYEALDCLRIPKSDVVYTDWKKDKCEPLEKQHEKERKQEE  180
            NWGLS YADKPETTKEQLGEFYEALDCL IP+SDV+YTDWKKDKCEPLEKQHEKERKQEE
Sbjct  121  NWGLSFYADKPETTKEQLGEFYEALDCLCIPRSDVMYTDWKKDKCEPLEKQHEKERKQEE  180

Query  181  GES  183
            GES
Sbjct  181  GES  183
```

Figure 1. 5. A1AG1(Query) and A1AG2 (Sbjct) sequence comparison showing an 89% identity.

Dilution tests were performed assessing different protein concentrations by comparing them with the theoretical advised concentrations of the analogous ELISA kits. After doing so, it was concluded that plasma samples for measuring SAA1 needed a 1/20 dilution (using a 10 point-standard curve ranging from 0 to

900 ng/mL), while the analysis of A1AG and AACT needed a 1/1000000 dilution (using 8 point-standard curves varying from 0 to 4000 pg/mL and from 0 to 8000 pg/mL respectively). In order to detect the HPT levels in the plasma samples, a 1/250000 dilution and an 8 point-standard curve extended from 0 to 8000 pg/mL were required. Furthermore, a technical verification strategy was conducted in order to confirm the validity of the ELISA measurements to those of the previous MRM analysis. As seen in **Table 1. 12**, plasma concentrations and disease activity trend of each peptide and or protein were compared when using both methodologies. On the one hand, SAA1 plasma levels and HA/LA ratios followed the same tendency when comparing the LA-RA and HA-RA subgroups, although they were higher when measured by MRM. The same inclination is followed when looking at the AVLD peptide of AAACT and the AACT ELISA measurements, while the other AACT peptide (EIGE) shows lower MRM concentrations but higher fold change. A1AG1 levels of the MRM peptides are quite heterogenous though they all display a higher but similar fold change when comparing the disease activity subgroups. On the other hand, when comparing the HPT concentrations calculated using the measured MRM peptides and ELISA protein levels, the HPT levels calculated by MRM were higher but the tendency was different in both methodologies. In this particular case, MRM data shows a slight modulation between both subgroups, notwithstanding ELISA results exhibit no alteration at all (fold change is close to 1). Besides, correlation between the MRM and ELISA samples analyzed ranged from low-moderate (HPT; R^2: 0.24; r: 0.49) to moderate (A1AG1, AACT; R^2: 0.46, 0.59; r: 0.68, 0.77) or high (SAA1; R^2: 0.78; r: 0.88) as shown in **Figure 1. 6**. Interestingly, no MRM-ELISA correlation is seen when the A1AG2 significant peptide (EQLG) is compared. Based on this previous data, it can be concluded that even knowing the difference between the protein levels calculated by MRM and/or ELISA, the tendency when comparing the two disease activity subgroups remains the same, assuring the ELISA technology as a good candidate for continuing the validation study.

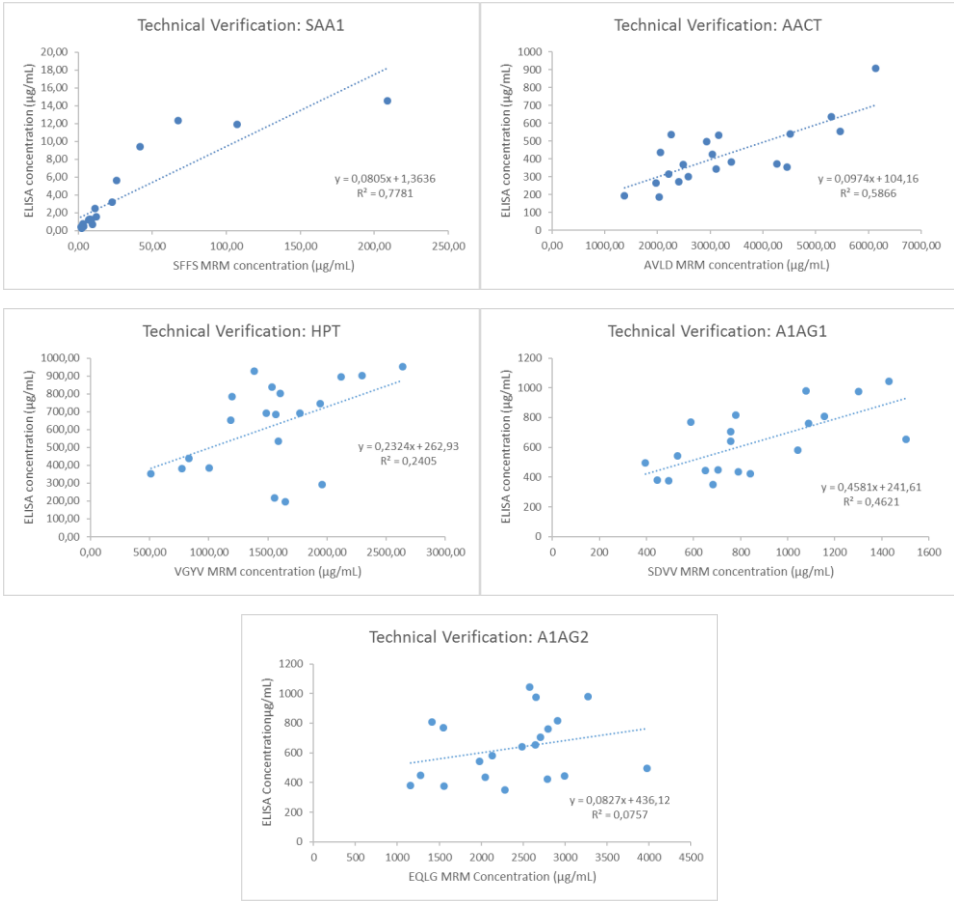


Figure 1. 6. Correlation of the absolute protein concentration when analyzing the statistically significant analogous peptides measured by MRM and the ELISA protein data for the same 20 plasma samples.

Table 1. 12. MRM and ELISA absolute and relative protein concentration comparison attending to the disease activity grade.

Technical Verification	[SAA1] (µg/mL)		[AACT] (µg/mL)			[HPT] (µg/mL)			[A1AG1] (µg/mL)			
	MRM_SFF	ELISA	MRM_AVL	MRM_EIG	ELISA	MRM_VGY	MRM_VTS	ELISA	MRM_YVG	MRM_EQL	MRM_SDV	ELISA
LA-RA (10 samples)	29,53	1,97	2598,40	126,95	371,06	1292,51	1321,16	625,08	276,41	559,95	762,25	584,47
HA-RA (10 samples)	93,36	5,68	3914,57	194,64	471,60	1766,97	1868,19	611,83	357,36	707,88	939,70	678,40
Fold Change (HA/LA)	**3,16**	**2,88**	**1,51**	**1,53**	**1,27**	**1,37**	**1,41**	**0,98**	**1,29**	**1,26**	**1,23**	**1,16**

After justifying the use of the ELISA technique in this validation phase, the cited 420 samples were analyzed in 5 plates per protein. The sample concentration in each plate was weighted and counterbalanced with the corresponding standard curve fit. Moreover, plate variability was assessed by estimating the intraplate and interplate CVs of the 3 RA plasma pool replicates

(**Annex I**). For SAA1 and AACT, intraplate CVs were lower than 15% in all 5 plates, whereas interplate CVs remained under 20%. Contrarily, HPT and A1AG variabilities show up to be discrepant in some of the intraplate (i.e. HPT: Plate 1 and Plate 4; A1AG: Plate 1, Plate 2 and Plate 5) and interplate CVs, with values over 20% for the 3 pool replicates. Plate 4 of HPT and Plate 2 of A1AG CVs could be corrected when excluding one of the replicates, although no amendment could be done for the rest of the fluctuating replicates.

Moreover, in order to normalize all this data, every corresponding protein value was corrected by using the already explained normalization factor (n) formula. Each protein statistics were compared with and without data normalization in order to examine the significance loss or gain by cause of this applied correction (**Table 1. 13** and **Table 1. 14**). Beginning with the disease group comparison, after performing the corresponding Mann Whitney test for each disease group match, it can be noticed that, in general, the significance is either maintained or increased when the median normalization is performed in any of the cases. Attending to the disease activity state of each of the IMID groups and the HD group, the tendency is hardly the same, although there are some little exceptions. SAA1 shows a higher significance when median normalization of the data is applied, while the significance between the LA-PsA and the HA-PsA is lost in the normalization. In this particular case, 3 potential outliers have been detected, each of them appertaining to 3 different groups (HD, HA-PsA and LA-PsA) as shown in **Figure 1. 7**. After their removal, the significance between the LA-PsA and the HA-PsA was again recovered, while the rest of the group comparisons significances were preserved and/or ameliorated. Contrary to this trends, different statistic HPT outcomes were achieved when studying the different disease activity stages of patients with IMIDs and/or HD. According to the present data, there is a loss of significance when HA-SLE and LA-SLE patients or HA-SLE and HD are contrasted versus a gain of significance when comparing RA and SLE patients suffering HA episodes. Interestingly, statistics notably improves when

normalization is applied to A1AG values, for the disease activity studies. Altogether, normalization of the previous protein data generally shows to maintain and/or increase the differences and significance among the studied groups. Thus, normalized data with outlier removal, in the case of SAA1, was used for the rest of the analysis.

Table 1. 13. Disease group comparison of the obtained ELISA data without and with median normalization and/or outlier removal if needed (as in SAA1). Taking into account the following significance criteria: p< 0.05 (*), p< 0.01 (**), p< 0.001 (***), p< 0.0001 (****).

Group Comparison (Mann Whitney)	SAA1			AACT	
	No Normalization	Median Normalization	Median Normalization without outliers	No Normalization	Median Normalization
HD vs RA	****	****	****	****	****
HD vs SLE	****	****	****	****	****
HD vs PsA	****	****	****	****	****
RA vs SLE	****	****	****	**	**
RA vs PsA	****	****	****	**	***
SLE vs PsA	ns	ns	ns	ns	ns

Group Comparison (Mann Whitney)	HPT		A1AG	
	No Normalization	Median Normalization	No Normalization	Median Normalization
HD vs RA	****	****	***	****
HD vs SLE	*	*	***	****
HD vs PsA	***	***	*	***
RA vs SLE	**	**	ns	ns
RA vs PsA	ns	ns	ns	ns
SLE vs PsA	ns	ns	ns	ns

Table 1. 14. Disease activity group comparison of the obtained ELISA data without and with median normalization and/or outlier removal if needed (as in SAA1). Taking into account the following significance criteria: p< 0.05 (*), p< 0.01 (**), p< 0.001 (***), p< 0.0001 (****).

Group Comparison (Mann Whitney)	SAA1			AACT	
	No Normalization	Median Normalization	Median Normalization without outliers	No Normalization	Median Normalization
HD vs HA-RA	****	****	****	****	****
HD vs LA-RA	****	****	****	****	****
HD vs HA-SLE	****	****	****	****	****
HD vs LA-SLE	****	****	****	*	*
HD vs HA-PsA	****	****	****	****	****
HD vs LA-PsA	**	***	***	**	**
HA-RA vs LA-RA	***	****	****	****	****
HA-SLE vs LA-SLE	ns	ns	ns	*	**
HA-PsA vs LA-PsA	*	ns	*	*	**
HA-RA vs HA-SLE	***	***	***	*	*
HA-RA vs HA-PsA	**	***	***	**	**

Group Comparison (Mann Whitney)	HPT		A1AG	
	No Normalization	Median Normalization	No Normalization	Median Normalization
HD vs HA-RA	****	****	****	****
HD vs LA-RA	*	*	ns	**
HD vs HA-SLE	***	ns	*	***
HD vs LA-SLE	ns	ns	***	***
HD vs HA-PsA	****	****	*	**
HD vs LA-PsA	ns	ns	ns	*
HA-RA vs LA-RA	***	****	*	**
HA-SLE vs LA-SLE	**	ns	ns	ns
HA-PsA vs LA-PsA	****	**	ns	ns
HA-RA vs HA-SLE	ns	***	ns	ns
HA-RA vs HA-PsA	ns	ns	ns	ns

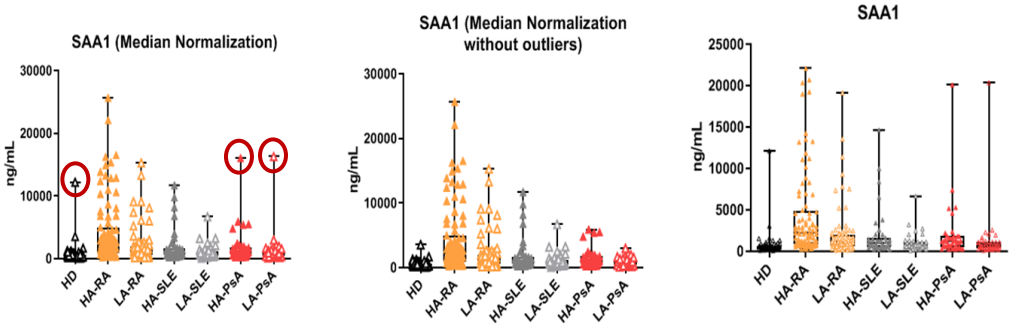


Figure 1. 7. SAA1 comparison of normalized data with or without outliers and data without normalization or outlier removal measured by ELISA.

A deeper analysis of this previous data can be explored for each protein in **Annex I**. After checking these quality control factors, statistical analysis was performed on the 420 plasma samples, distinguishing 2 comparisons: disease identification and disease activity relation. On the one hand, **Figure 1. 8** and **Table 1. 15** illustrate the disease comparison of the 3 IMIDs patients and the HD for each of the measured proteins.

Table 1. 15. Final protein statistics for differentiating patients with different IMIDs and/or HD.

Group Comparison	SAA1 (Normalized Outliers Removal)			AACT (Normalized)		
	Kruskal-Wallis: p< 0,0001		Mann Whitney	Kruskal-Wallis: < 0.0001		Mann Whitney
HD vs RA	-154.7	****	**<0.0001**	-132.6	****	**<0.0001**
HD vs SLE	-94.41	****	**<0.0001**	-85.79	****	**<0.0001**
HD vs PsA	-87.14	****	**<0.0001**	-83.58	****	**<0.0001**
RA vs SLE	60.26	***	**<0.0001**	46.84	*	**0.0022**
RA vs PsA	67.53	***	**<0.0001**	49.05	*	**0.0007**
SLE vs PsA	7.269	ns	0.7417	2.219	ns	0.7979
Group Comparison	**HPT (Normalized)**			**A1AG (Normalized)**		
	Kruskal-Wallis: p< 0,0001		Mann Whitney	Kruskal-Wallis: p< 0,0001		Mann Whitney
HD vs RA	-87.91	****	**<0.0001**	-75.16	****	**<0.0001**
HD vs SLE	-38.23	ns	**0.0241**	-85.63	****	**<0.0001**
HD vs PsA	-71.83	***	**0.0002**	-55.64	*	**0.0009**
RA vs SLE	49.68	*	**0.0019**	-10.47	ns	0.499
RA vs PsA	16.08	ns	0.3764	19.53	ns	0.1921
SLE vs PsA	-33.61	ns	0.0635	30	ns	0.092

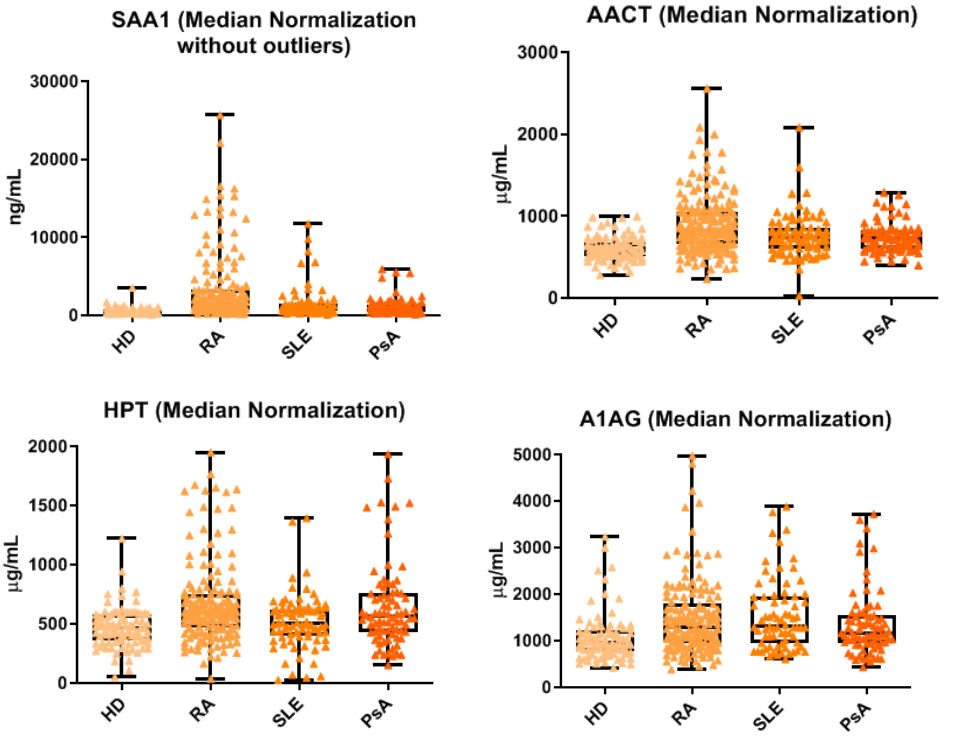


Figure 1. 8. SAA1, AACT, HPT and A1AG absolute concentrations measured by ELISA on 420 plasma samples from patients with different IMIDs (RA, SLE, PsA) and healthy donors (HD).

The average tendency of every protein is to be elevated in RA patients (except for A1AG, which exhibits a slightly greater concentration in SLE patients), while HD present lower protein concentrations (**Annex I**). All the 4 proteins seem to differentiate patients with any of the investigated IMIDs from the healthy donors (Mann Whitney test: $p < 0.05$). Additionally, every protein distinguishes RA patients from HD with a strong relevance ($p < 0.0001$), although SAA1 shows a greater difference between any of these groups. Moreover, only SAA1 and AACT discriminate RA patients from patients with other related IMIDs as SLE and PsA. None of the proteins can categorize SLE and PsA patients when examined together, HPT and A1AG almost do so as they are close to significance ($p= 0.0635$ and $p= 0.092$ respectively). In conclusion, SAA1 exhibits a superior biomarker role when comparing the 4 studied disease profiles, specifically when classifying RA patients from other patients with related IMIDs and/or HD, followed by AACT.

On the other hand, **Figure 1. 9** and **Table 1. 16** illustrate the disease activity modulation in IMID patients and healthy donors. HA-RA patients register the greatest concentrations in any of the measured protein (except for A1AG, where it was greater in HA-SLE patients), while HD involve the lowest protein concentrations for any of them. Furthermore, all protein concentrations are more elevated on the high activity condition (HA) than the corresponding low activity condition (LA) for any of the mentioned IMIDs (**Annex I**). Besides, SAA1, A1AG and AACT can differ all IMID patients with different activities from HD, whereas only AACT distinguishes the HA and LA conditions for every IMID tested. As stated in the previous discovery and verification phases, all proteins show a significant difference between HA-RA and LA-RA patients, as well as between HA-RA or LA-RA patients and HD. Beyond, SAA1 and HPT could classify PsA patients with HA and/or LA. SLE and PsA patients with high activity could also be recognized from HA-RA patients when looking at the SAA1 and/or AACT levels. In view of this data, all 4 proteins are attractive biomarkers for RA activity monitoring, specially SAA1

and AACT which could also discriminate between patients with related IMIDs when they undergo high disease activity stages.

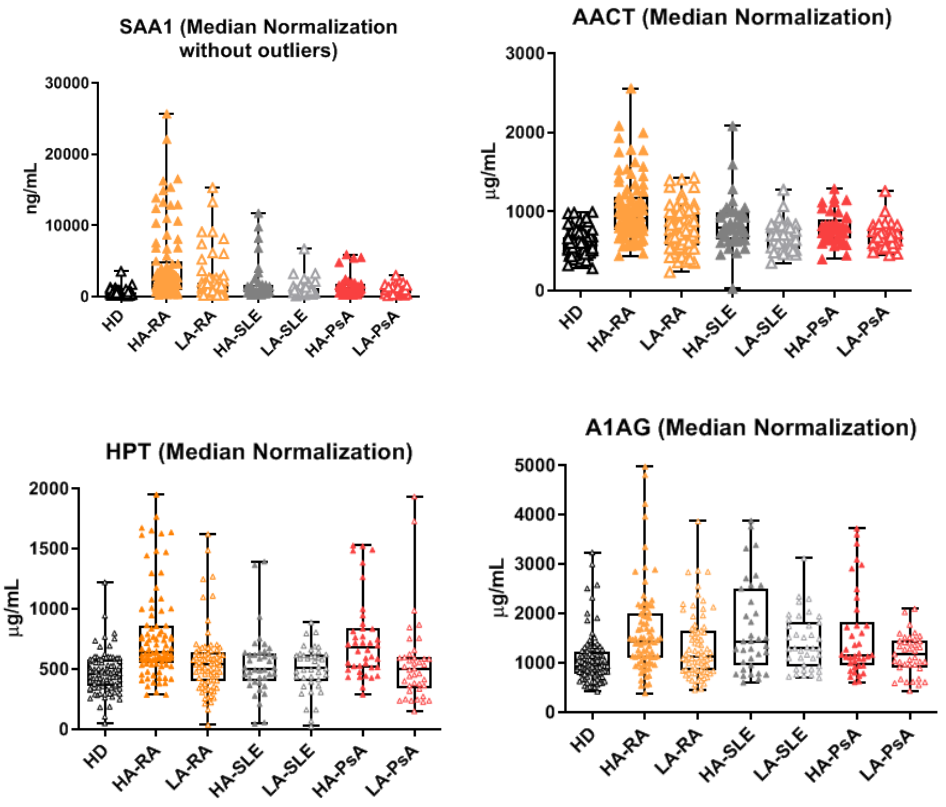


Figure 1. 9. SAA1, AACT, HPT and A1AG absolute concentrations measured by ELISA on 420 plasma samples from IMID patients with extreme activities and HD.

Table 1. 16. Final protein statistics for differentiating IMID patients with extreme activities and/or HD.

Group Comparison	SAA1 (Normalized Outliers Removal)			AACT		
	Kruskal-Wallis: p< 0,0001		Mann Whitney	Kruskal-Wallis: < 0.0001		Mann Whitney
HD vs. HA-RA	-184,9	****	**<0.0001**	-173.5	****	**<0.0001**
HD vs. LA-RA	-121	****	**<0.0001**	-85.55	***	**<0.0001**
HD vs. HA-SLE	-108,7	****	**<0.0001**	-122.5	****	**<0.0001**
HD vs. LA-SLE	-80,46	**	**<0.0001**	-49.11	ns	**0.0147**
HD vs. HA-PsA	-111,1	****	**<0.0001**	-115.1	****	**<0.0001**
HD vs. LA-PsA	-63,82	ns	**0.0006**	-52.03	ns	**0.0021**
HA-RA vs. LA-RA	63,86	**	**<0.0001**	87.96	****	**<0.0001**
HA-SLE vs. LA-SLE	28,26	ns	**0.2538**	73.38	ns	**0.0042**
HA-PsA vs. LA-PsA	47,26	ns	**0.046**	63.09	ns	**0.0051**
HA-RA vs. HA-SLE	76,16	*	**0.0002**	51.02	ns	**0.012**
HA-RA vs. HA-PsA	73,8	*	**0.0002**	58.39	ns	**0.0015**

Group Comparison	HPT Normalized			A1AG Normalized		
	Kruskal-Wallis: p< 0,0001		Mann Whitney	Kruskal-Wallis: p< 0,0001		Mann Whitney
HD vs. HA-RA	-125.4	***	< 0.0001	-97.75	****	<0.0001
HD vs. LA-RA	-44.09	ns	0.0171	-50.17	ns	0.0063
HD vs. HA-SLE	-40.45	ns	0.0631	-94.34	***	0.0001
HD vs. LA-SLE	-36.01	ns	0.0758	-77.82	**	0.0002
HD vs. HA-PsA	-120.5	***	< 0.0001	-74.44	*	0.001
HD vs. LA-PsA	-23.2	ns	0.5039	-37.31	ns	0.0377
HA-RA vs. LA-RA	81.31	***	< 0.0001	47.58	ns	0.0096
HA-SLE vs. LA-SLE	4438	ns	0.8663	16.52	ns	0.4363
HA-PsA vs. LA-PsA	97.28	**	0.0013	37.13	ns	0.1683
HA-RA vs. HA-SLE	84.95	**	0.0002	3.411	ns	0.9253
HA-RA vs. HA-PsA	4928	ns	0.8106	23.31	ns	0.3203

1.3. Discussion.

In this study, a 3-stage proteomic approach has been followed for the discovery, verification and validation of putative plasma biomarkers that could be able to discriminate between patients with different IMIDs going across extreme disease activity stages and healthy individuals. As stated by many authors (62, 64), the biomarker pipeline is based on a reverse interrelationship among the number of samples used for the analysis and the number of proteins that arise as possible biomarkers. In this study, while 80 pooled plasma samples were used at the beginning in the discovery phase, identifying 199 proteins, 4 of these proteins and 420 other plasma samples were analyzed at the end in the validation phase. In the middle, a verification of 11 of the 199 proteins was assessed in the same individual 80 plasma samples before proceeding to the final stage, following the previous biomarker discovery concept. Deeper in each phase, some remarks must be detailed taking into account each of the procedures followed and the corresponding outcomes obtained.

Even though plasma samples are preferred over serum due to better reproducibility (94), preanalytical variability factors as sample selection, handling and preparation may play an important role in this initial critical stage. It is necessary to take into account that the plasma samples used in the discovery phase were not only pooled, but also chemically depleted. Making pools of the plasma samples and/or depleting them are considered to be double-edged swords. Patients are wider represented when samples are pooled by groups, although at the same time possible low abundant biomarkers are diluted with the risk of missing possible patient-specific characteristics. Depletion of high-level proteins, such us albumin, is required when the identification of as many proteins as possible is needed, as in the case of a wide discovery. Notwithstanding, many low-abundant proteins might be unintentionally removed during these primary steps, missing a remarkable information.

The tuned TCA/EtOH protocol followed herein has demonstrated to be quite reproducible and effective on the HSA removal, without endangering the identification of the rest of the proteins compared. These results have been contrasted with a commercially available HSA depletion kit (Pierce™ Albumin Depletion Kit, Thermo Scientific), giving the TCA/EtOH tuned protocol significant advantages due to its simplicity, reproducibility, accessibility and competitive cost. This depletion kit uses an immobilized Cibacron Blue dye agarose resin, which seems to interact with albumin by its hydrophobic fatty acid-anion and bilirubin-binding sites (95) and depends on the dye loading, pH, HSA initial concentration, ionic strength and/or temperature (96, 97). In contrast, the mechanism that takes place in the depletion of HSA using the TAC/EtOH protocol is presumably based on the formation of a TCA-albumin complex (using TCA as a protein precipitating agent) that is soluble in different organic solvents as EtOH (93). It is thought that the protein acquisition of the hydrophobic trichloromethyl group might increase the solubility of the TCA-protein complex in an organic solvent.

Generally, biomarker discovery is conducted with non-targeted/shotgun relative quantitation techniques, achieving results as up-or-down regulation of proteins and/or fold changes. In this case, the iTRAQ isotopic labeling technique was used for the “bottom-up” relative quantitation of a protein, as it is calculated based on the intensities of peptides that appertain to the analogous protein. As seen in the screening results, 199 proteins were identified in total using the two different fractionation and LC-MS/MS approaches. 149 proteins were commonly identified using both methods, while 15 were exclusively identified by ESI-tTOF and 35 by MALDI-TOF/TOF. This difference in the number of proteins identified with each method is likely to be directly related with the number of fractions analyzed, which were higher by MALDI-TOF/TOF (30) than by ESI-tTOF (5), as well as with the type of fractionation performed. Focusing on the MALDI-TOF/TOF fractionation, the first 60 portions were combined into 30, taking into account the hydrophobicity and total peptidic content of the final fractions. Each fraction was

again partitioned by nanoLC after adjusting the needed organic content for the chromatography separation of the corresponding peptidic fraction. In the case of the ESI-tTOF fractionation, the initial 30 fractions collected were joined in 5 final interleaved fractions, picking and mixing fractions with dispersed hydrophobic grades. The resulting 5 inserted fractions were separated by nanoLC using the same organic content. The first fractionation presents less peptides per final fraction and a more specific second fractionation attending to the required organic content, so the number of peptides separated and identified is higher than in the second LC-MS/MS analysis tested. Nonetheless, the ESI-tTOF approach offers a higher number of peptides and spectra identified, meaning a higher coverage of the proteins detected with this method but a lower % of total spectra used for identification. Therefore, in terms of protein identification, the whole MALDI-TOF/TOF approach is offered as the best option for this discovery phase even taking into consideration that the MS equipment is not so advanced as the tTOF.

From this screening phase, 11 proteins were selected as possible biomarker candidates for RA activity monitoring, attending to their statistical significance and relevance on the studied disease. Most of these proteins, as A1AG, AACT, HPT and/or SAA, are inflammatory markers and also known as acute phase reactants (APRs). APRs function as mediators and/or inhibitors of inflammation, immune response regulators, transport proteins for products generated during the inflammatory process, and/or play an active role in tissue repair and remodeling (98). The other remaining elected proteins are also interesting as they are involved in other processes like the binding of Vitamin E (which exhibits anti-inflammatory properties) as AFM (99), the osteoblast differentiation as IGFBP3 (100) and the bone mineralization as AHSG (54).

After reviewing this protein panel, the following verification phase was conducted by MRM, named as “Method of the Year for 2012” by Nature Methods

(101). The MRM ability of rapidly analyzing several numbers of biomarkers in only one LC/MRM-MS analysis (high multiplexing) makes it well-suited for the verification of large panels of biomarkers. Concretely, MRM is a tandem mass spectrometric technique (MS/MS) performed on triple-quadrupole mass spectrometers that allows the quantitation of hundreds of peptides and, by extrapolation, quantitation of their parent proteins. In this study, 11 proteins and 26 proteotypic peptides were simultaneously monitored in a single MRM method, later including their SIS peptide analogs. The use of SIS peptides helps to control the reproducibility and variability of MRM assays when both are applied together, and it is considered as the "gold standard" MS-based quantitation method (102). For each peptide, the SIS concentrations needed to be balanced in order to obtain a good NAT/SIS signal. This SIS-concentration balance reduced the analytical variation between injections, ameliorated accuracy of the quantitation and broadened the linear range of the calibration curves. Actually, the majority of the CVs were all lower than 20% when examining the 80 plasma samples and/or the points of the final calibration curves, all measured with SIS peptides in duplicates and triplicates, respectively. Moreover, the linear range of 21 of the 26 selected peptides fulfilled the standard curve requirements stated in the FDA guidelines, already accounted with the Qualis-SIS software (103). These conditions include 20% deviation at LLOQ and 15% for the rest of the qualified levels, in terms of precision and accuracy criteria. 2 proteins (HABP2 and IGFBP3) could not be measured by MRM, as the 4 analogous peptides didn't accomplish the linear range requirements. Particularly, both peptides showed irregular peak shapes in Skyline and the coelution of their transitions was quite unstable as well, failing to give a good quantitation of the corresponding proteins analyzed. In order to solve this issue and being able to measure these proteins, additional HABP2 and IGFBP3 peptides would need to be researched. The same happens with one of the KNG1 peptides, but in this case, there are other 2 peptides that worked perfectly for the KNG1 MRM detection and/or quantification.

Plasma samples were not depleted this time, as the MRM methodology permits the analysis and quantitation of proteins whose concentrations fall within 5 or more orders of magnitude when using undepleted plasma (104). Whence, the 80 plasma samples used in the previous screening phase were individually analyzed by MRM with the aid of concentration-balanced SIS peptides. Different concentrations of SIS peptides for each of the protein corresponding peptides were used, as it happens with the AFM, HRG and/or KNG1 peptides. This fact is related with the nature and ionization of the peptide and not with the concentration of the peptide in the sample. If a peptide ionizes worse than others, the concentration of this peptide needs to be increased for ameliorating its signal. Opposingly, when a peptide ionizes too well, its concentration needs to be lowered down in order to avoid MRM signal saturation problems and/or light peptide interference. This light peptide contamination can come from the SIS peptide synthesis, which is not seen when lower SIS peptide concentrations are used.

The nanoLC-MRM analysis yielded significant differences between the HA-RA and LA-RA groups in 5 of the 9 remaining proteins analyzed, although the initial up/down-regulation trend from the discovery phase was followed as well by all the proteins/peptides. This tendency corroborates the relative-quantitation results obtained in the previous stage, confirming both methodologies, iTRAQ labelling and MRM analysis, as possible MS-based techniques for RA activity monitoring. It is important to take into account that the first one is based on a relative-quantitation while the second one is absolute. iTRAQ quantification gives fold changes while MRM quantitation calculates NAT/SIS ratios, based on the NAT and SIS peak areas, which could be converted in absolute protein concentration values whether the cited formula is applied. Furthermore, two different quantification strategies could be followed for MRM quantification: single point measurement (SPM) and/or using the calibration curve regression of each protein. In this study, the absolute quantification was achieved via SPM, although it could

have also been done using the curve equations as they all offer a satisfying regression ($R^2 > 0.97$). Peak areas from every transition of each peptide were used for quantitation. Generally, the highest transition is the one used for quantification (quantifier), while all the transitions are used for the peptide identification (qualifiers) (105-107).

Another important factor to mention is related with the different significance and concentrations of the measured peptides, which determine the protein biomarker role and/or levels in plasma (108). One example of the same protein peptides showing different p values is HPT. Whereas the VGYV peptide significantly differs between patients with extreme activities (p= 0.042), the other VTSI peptide from HPT doesn't do it significantly, although it is somehow close (p= 0.08). The same happens with A1AG1, A1AG2 and AACT peptides. Conversely, when looking at the absolute concentration of the peptides in plasma, both HPT peptides display comparable levels in either HA-RA or LA-RA groups; while AACT peptide concentrations differ considerably (around 21 times higher with the AVLD peptide). Normally, the peptide that yields the highest concentration is reported since, theoretically, it has been digested and measured with the greatest efficiency (109). Consequently, it is important to note that, apparently, the peptide of each parent protein with the highest concentration is normally also the one that offers the highest significance (AVLD from AACT and SDVV from A1AG1). Interestingly, even if the peptides differ in absolute protein concentration when measuring one particular protein, the HA-RA/LA-RA peptide fold changes remain quite resembling in every protein case. All peptide fold changes match when relative protein quantitation is determined by MRM analysis as a consequence of using SIS peptides for between sample analysis normalization.

Five (SAA1, HPT, AACT, A1AG1 and A1AG2) of the nine proteins measured by MRM were successfully verified and could go through the following validation phase. MRM analysis is a good technique when many proteins are measured at

the same time. Nevertheless, when the number of samples to be examined increases, as in this last validation phase, a technique that analyzes many samples at the same time is needed. Enzyme linked immunosorbent assays (ELISAs) are best suited for this situation where only a few biomarkers need to be verified or validated in a large number of samples. In this case, the 5 verified proteins were validated by ELISA on 420 plasma samples of patients with different IMIDs and/or healthy donors. However, ELISAs require antibodies against each targeted protein or peptide, and high-quality ELISA assays are often unavailable for some proteins (110), as here in the case of A1AG1 and A1AG2 proteins. Due to the high homology of both proteins, the ability of the specific A1AG1 and/or A1AG2 antibodies of binding the corresponding protein is limited, as they could bind to the 89% shared protein sequence. This would lead to failures in quantifying the correct protein of the ELISA, which does not occur when a MS approach is followed, as they quantify proteins based on their corresponding proteotypic peptides. However, in order to choose the most convenient ELISA kit, the immunogen sequence was asked to the R&D systems development team, who confirmed the A1AG1 peptides measured by MRM were included in this sequence. As more A1AG1 peptides were close to significance than A1AG2, the commercial R&D systems ELISA kit was used under the limitation of a presumable detection of both A1AG1 and A1AG2 proteins.

Other restraints of using ELISA kits is their limited multiplexing capabilities and cross-reactivity issues. About this last matter, it should be noted that false-positive and/or negative interferences could cause over and/or underestimation of the ELISA protein concentrations. Rheumatoid factor (RF) is widely mentioned as an important cause of false-positive interference in immunoassays. RF is a kind of autoantibody against the fragment c portion of IgG, IgA, IgM, IgE and/or IgD. In sandwich immunoassays, RF can link the capture and the detection antibodies together and falsely increase the measured ELISA signal (111). Plasma RF levels are elevated in about 2% of healthy people and 20% of people over 60 years old.

Particularly, high levels of plasma RF are present in about 80% of RA patients, and they are also elevated in patients with other IMIDs (112). This is the reason why RF interference should be accounted and blocked/or removed when performing ELISA analysis in samples from patients with any IMID. Nevertheless, this non-specific binding is prompt to decrease as a function of serum/plasma dilution (113). In this project, 1/20, 1/250000 and 1/1000000 plasma dilutions were assessed after performing the dilution tests for the 4 purchased ELISA kits. Except for the 1/20 plasma dilution needed for the SAA1 analysis, the rest are quite large dilutions and should not provoke cross-reactivity problems due to non-specific binding. Therefore, only the original blocking agents of each particular ELISA protocol were used to avoid this interference issue.

Another important fact taken into account in this study was the correlation between the MRM and ELISA analysis. For this aim, a prior small technical verification was conducted. Most of the MRM and ELISA HA-RA/LA-RA fold changes follow a similar trend for SAA1, AACT and A1AG, with the exception of HPT. When measuring 20 of the 80 plasma samples by ELISA, the FC is altered from 1.37-1.41 when HPT is analyzed by MRM to 0.98 in the case of ELISA measurements. Moreover, SAA1, AACT and A1AG show high to moderate MRM-ELISA correlations (R^2: 0.78, 0.59 and 0.46 respectively), while HPT demonstrates low correlation (R^2: 0.24). The flipped FC regulation and the low MRM-ELISA correlation seen in the HPT analysis is presumably due to the different protein sequences measured by each technique. Thus, the HPT peptides measured by MRM are located on the beta chain (**Annex I**), while the HPT ELISA analysis is done based on the whole protein sequence, including both the alpha and beta chains. Finally, the MRM concentrations attending to the significant peptides of SAA1, AACT, HPT and A1AG1 were higher than those determined by ELISA. As a result, MRM may have likely measured the total protein present in the sample, while ELISA might have determined only the most accessible free protein, possibly due

to deficient simultaneous capture and detection antibody binding in the sandwich system.

Finally, the ELISA validation results showed a significant increase of these four proteins in RA patients with high activity, as well as a differential abundance in patients with any rheumatic IMID and HD, therefore corroborating their value as potential biomarkers for RA activity monitoring and rheumatic IMID diagnosis and classification. Moreover, AACT also distinguishes between SLE patients with extreme disease activities, while HPT and SAA1 significantly differ among PsA patients with low and high disease activity.

In summary, four proteins (SAA1, HPT, AACT and A1AG1) have been found significantly altered in RA patients with extreme disease activities and patients with similar rheumatic IMIDs and/or HD. Interestingly, these proteins have been extensively defined as acute phase and/or disease activity markers in several inflammatory conditions including RA (114-121), although they have not been established together yet as a multi-biomarker panel for clinical utility. However, although a more extensive statistical analysis needs to be further performed for validating this panel in clinics, it seems to provide satisfactory values for detecting similar rheumatic IMID patients with different disease activities. Particularly, SAA1 has displayed better outcomes for patient classification and monitoring. In fact, this protein has been already considered as a possible indicator of response to treatment (122, 123). Altogether, this protein panel could serve as a profiling tool for rheumatic IMID classification; RA, SLE and PsA disease activity monitoring; as well as for treatment selection or therapy re-assignment in clinics, which is usually performed following a trial-error procedure in RA patients.

2. Chapter II: Peptidomics study on OA cartilage secretome, synovial fluid and serum samples for the detection of a panel of OA-related endogenous peptides as possible OA biomarkers.

OA is clinically silent in most patients in their early stages; thus, the deterioration of cartilage is already extensive at the time of diagnosis. Therefore, the development of strategies for early diagnosis and accurate monitoring of disease progression is among the major research goals in OA. OA is characterized by the loss of structural constituents from the articular cartilage ECM (124). The ECM maintains and supports chondrocytes within their natural physicochemical micro-environment, and the degradation and release of cartilage proteins can vary according to the stage of the disease process (125). Degradation of cartilage ECM proteins by specific proteinases is one of the main factors involved in OA pathology which contribute to disease progression. Several proteases have been extensively described as responsible for degradation of cartilage ECM proteins in OA such as MMPs, ADAMTSs, cathepsins, calpains, caspases, among others (13). Therefore, the presence of cartilage-characteristic proteins and their degradation products in proximal or peripheral body fluids (such as synovial fluid, blood or urine) has been extensively evaluated to assess their biomarker usefulness. As examples confirming this hypothesis, the increase of the type II collagen fragment CTXII in urine has demonstrated a predictive value for disease progression (126, 127), and elevated levels of COMP in serum are correlated with the presence of OA and disease severity (128). Altogether, the ability to detect biomarkers of cartilage degradation and/or inflammation in biological samples may be helpful to improve OA diagnosis, predict its progression and/or develop effective therapeutic strategies. In this area, proteomics has demonstrated to be a powerful tool for biomarker discovery in OA research (65, 129).

The term “peptidomics” was introduced as a branch derived from proteomics to define the quantitative and qualitative analysis of endogenous

peptides (also named neopeptides) in biological samples, primarily by LC or biochip platforms coupled to various forms of MS (87). A specific neopeptide can be released from a protein due to the existence or progression of a specific disease. Therefore, peptidomics has been appealing for biomarker studies since the knowledge that is generated may present a dynamic view of health status: peptides are created by a complex and fluid interaction of proteases, activators, inhibitors and protein substrates (130). Due to many difficulties, biomarker discovery of endogenous peptides in complex samples is challenging and require systematic peptide extraction to achieve successful analysis (71). In this work, the principal aim was to characterize the profile of neopeptides present in conditioned media (secretomes) from human articular cartilage and quantitatively compare these profiles between healthy and osteoarthritic tissues from different joints. The resulting peptides and additional neopeptides from a subsequent study were measured as well in synovial fluid (SF) and serum samples in order to confirm their presence in more clinical-accessible samples. This would allow not only to identify potential neopeptide biomarker candidates, but also to foster the understanding of specific protease pathways that may be relevant for cartilage ECM destruction, which is one of the most characteristic features of OA.

CHAPTER II

Part of this chapter has been recently accepted for publication in Molecular & Cellular Proteomics as:

- Patricia Fernandez-Puente*, **Lucía González-Rodríguez***, Valentina Calamia, Florencia Picchi, Lucía Lourido, María Camacho-Encina, Natividad Oreiro, Beatriz Rocha, Rocío Paz, Anabel Marina, Carlos García, Francisco J. Blanco and Cristina Ruiz-Romero. *Analysis of endogenous peptides released from osteoarthritic cartilage unravels novel pathogenic markers*. Molecular & Cellular Proteomics July 27, 2019, mcp.RA119.001554. https://doi.org/10.1074/mcp.RA119.001554.
  *These authors contributed equally to this work.

2.1. Materials and Methods.

2.1.1. Identification and targeted MS analysis of neopeptides secreted from articular cartilage.

2.1.1.1. Human articular cartilage specimens.

Articular cartilages for the proteomic analysis were obtained either from femoral heads or condyles of patients with OA undergoing hip or knee replacement, and donors with no history of joint disease (N). All tissue samples were provided by the Tissue Bank and the Autopsy Service at Hospital Universitario de A Coruña. The study was approved by the local Ethics Committee (Galicia, Spain). OA patients were diagnosed following the criteria determined by the American College of Rheumatology (131). The experimental workflow followed for the neopeptidomic analysis of articular cartilage and the identification of OA markers is later summarized in **Figure 2. 8**. It comprises 3 different stages: discovery/screening stage, MRM method development and final validation of the resulting neopeptides using the established MRM method. Particularly, cartilage samples from 5 patients were used for the screening analysis (3 OA and 2 N), from additional 21 patients were employed for MRM development (13 OA and 8 N), and from further 40 patients in the validation studies (22 OA and 18 N). The demographic characteristics of the donors are detailed in **Table 2. 1**.

Table 2. 1 Characteristics of the articular cartilage explants employed in this work. Two different explants were obtained per OA tissue: one from the unwounded and another from the wounded zone. Thus, the total number of samples analyzed is duplicated for OA cartilage. Cartilage tissue conditions are expressed as: N: healthy cartilage; UZ: Unwounded Zone of OA cartilage; WZ: Wounded Zone of OA cartilage.

Screening					
Dx		n	% Female	Age (mean±SD)	Mankin (mean)
N		2	33.3	77.33±4.16	1.5
OA		3	0	66±11.31	2.5 (UZ) 7.6 (WZ)
Total number of samples: 8 (2N, 3UZ, 3WZ)					
MRM Development					
Dx		n	% Female	Age (mean±SD)	Mankin (mean)
Hip	N	6	33.3	77.67±8.16	1.5
	OA	5	100	82.2±6.02	3.6 (UZ) 6.2 (WZ)
Knee	N	2	0	56±2.83	1.5
	OA	8	62.5	82.5±9.26	3.2 (UZ) 9 (WZ)
Total number of samples: 34 (8N, 13UZ, 13WZ)					
Validation					
Dx		n	% Female	Age (mean±SD)	Mankin (mean)
Hip	N	13	38.46	76.38±12.24	1.7
	OA	10	70	77.8±9.02	3.3 (UZ) 9.3 (WZ)
Knee	N	5	40	70.6±13.6	2.6
	OA	12	41.67	73.93±6.97	5 (UZ) 9.8 (WZ)
Total number of samples: 62 (18N, 22UZ, 22WZ)					

2.1.1.2. Histological-histochemical grading of cartilage.

A modified Mankin scoring was employed for the histopathological classification of the severity of lesions on all the cartilage samples employed in this work (132). Briefly, tissue sections (4 µm) were stained with hematoxylin and eosin to evaluate cellular architecture, and with toluidine blue and safranin O/fast green to visualize the matrix proteoglycan content. Three different aspects of the score were determined and summed up: cartilage structure (0-7 points), cellular abnormalities (0-2 points) and matrix staining (0-4 points), leading to a scale that ranges between 0 and 13. The Mankin score between 0–2 represents normal cartilage, 3–5 superficial fibrillation, 6–7 moderate cartilage destruction, 8–10 severe damage of cartilage, and over 10 complete loss of cartilage.

2.1.1.3. Cartilage explants culture and obtention of secretomes.

Tissue explants were obtained from the dissection of N and OA hip and knee cartilages as previously described (133). Among the OA cartilage samples, the wounded zones (WZ) were differentiated from those corresponding to the area adjacent to the lesion, or unwounded zones (UZ). Three 6-mm explants were cut from each zone/condition using a sterile biopsy punch. After extensive washes with PBS, the discs were placed into 96-well plates (one disc/well), containing 200 µL of serum-free DMEM supplemented with 100 units/mL penicillin and 100 µg/mL streptomycin to avoid contamination. Plates were incubated overnight at 37°C, 5% CO_2. The collection time line of conditioned media (secretomes) was optimized based on previous experience (133) and after appraising representative proteomic and peptidomic profiles along 7 days. Secretomes from day 1 were discarded and replaced with fresh medium in order to remove serum contamination from surgery. Then, secretome samples were collected at days 2 and 5 from each explant culture, achieving a total volume of 600 µL per condition and day (3 explants of 200 µL each) and finally frozen at -80^{O} C until processing.

2.1.1.4. Secretome Processing.

In order to isolate the endogenous peptides, present in the conditioned secretome media, different combinations of ultrafiltration (UF) and solid phase extraction (SPE) were explored. This led to the following final protocol described, which showed the highest number of unique peptides and the lowest serum contamination in the conditioned media. Thus, secretomes from the same donor and condition (WZ, UZ or N) collected at days 2 and 5 were mixed together in a total volume of 1200µL. The endogenous peptides were concentrated by ultrafiltration using Amicon Ultra-4 devices (10 kDa MWCO, Merck Millipore, Bedford, MA). The resulting eluted volumes (fractions comprising peptides of < 10 kDa), were dried in a vacuum concentrator. The samples were cleaned twice

prior to LC-MS/MS analysis, first by homemade Stage Tips containing six C18 Solid Phase Extraction Disks (Empore), and then using commercial NuTip C18 (Glygen).

2.1.1.5. Isotopically Labelled Synthetic Peptides.

Synthetic isotope-labelled standard peptides (SIS peptides, crude purity) were purchased from Thermo Scientific, (USA). These peptides incorporated fully atom labelled ^{13}C and ^{15}N isotopes at the different amino acids (labelled position; mass shift) as Alanine ($^{13}C_3$,^{15}N-Ala; +4 Da) (A), Proline ($^{13}C_5$,^{15}N-Pro; +6 Da) (P), Valine ($^{13}C_5$,^{15}N-Val; +6 Da) (V), Leucine ($^{13}C_6$,^{15}N-Leu; +7 Da) (L), Lysine ($^{13}C_6$,$^{15}N_2$-Lys; +8 Da) (K), or Arginine ($^{13}C_6$,$^{15}N_4$-Arg; +10 Da) (R). Individual stocks of each peptide ranging from 2.25-19.5 µg/µL were made. Then, equal volumes of each peptide were combined to make the standard mixture solution. Finally, a 1/5000 dilution of this mixture was made as the SIS stock solution in a concentration range of 1.78-17.6 pmol/µL of each peptide. Aliquots were kept at -20^{o}C until LC-MS/MS analysis. The processed cartilage secretome samples used to develop the targeted MRM method were reconstituted in 7 µL of buffer A (5% acetonitrile in 0.1% formic acid) or 7µL of SIS peptide stock solution, whereas the set of samples used for the validation was uniquely reconstituted in 7 µL of the SIS peptide stock solution.

2.1.1.6. Discovery phase analysis by shotgun LC/MS-MS.

Eight dried secretome desalted samples (n=8, 2N, 3UZ, 3WZ) were resuspended in 10 µL of 0.1% formic acid (FA) and analyzed by LC-MS/MS in an Easy-nanoLC II system coupled to LTQ-Orbitrap-Velos-Pro mass spectrometer (Thermo Scientific). The peptides were concentrated by reverse phase chromatography using a 0.1mm × 20 mm C18 RP precolumn (Proxeon), and then separated using a 0.075mm x 100 mm C18 RP column (Proxeon) operating at 0.3 µL/min. Peptides were eluted using a 90-min gradient from 5 to 40% solvent B

(Solvent A: 0,1% FA in water, solvent B: 0,1% FA, 80% acetonitrile in water). ESI ionization was performed using a Nano-bore emitters Stainless Steel ID 30 µm (Proxeon) interface. The Orbitrap resolution was set at 30.000. Peptides were detected in survey scans from 400 to 1600 amu (1 µscan), followed by ten data dependent MS/MS scans (Top 10), using an isolation width of 2 m/z units (in mass-to-charge ratio units), normalized collision energy of 35%, and dynamic exclusion applied during 30 seconds periods.

2.1.1.7. Design and development of the Multiple Reaction Monitoring (MRM) method.

Firstly, in order to design the MRM method, target peptides were chosen from the identified peptides detected on the previous discovery phase. The peptide selection criteria were the following: 1) peptides with the highest Xscore (>3) using the Proteome Discoverer 1.3 software, 2) peptides present in at least 4 of the 6 secretome samples analyzed in the discovery phase and 3) peptides belonging to cartilage ECM proteins. The five most intense transitions for each suitable precursor were elected based on data deposited in the MS/MS library using the Skyline software (134). Thus, peptide precursors and fragment ion masses were selected on this basis and assayed for MRM analysis.

Therefore, 34 processed and desalted secretome samples (8N, 13UZ, 13WZ) reconstituted in 7 µL of 5% ACN 0.1% FA or SIS stock solution, were analyzed in total for achieving the development of the final MRM method. In the first stages of the MRM method development, secretome samples were reconstituted in 5% ACN 0.1% FA in order to test the detection of the selected peptides by a different LC-MS/MS strategy from the discovery phase. Differently, at the end of the method development, SIS peptides were used for confirming the identification of the selected peptides as well as for electing the most adequate precursor charges and/or transitions.

Thus, endogenous and SIS peptides were analyzed by LC-MS/MS using a nanoLC system (TEMPO, Eksigent) coupled to a 5500-QTRAP instrument (Sciex). The mass spectrometer was interfaced with nanospray sources equipped with uncoated fused silica emitter tips (20 µm inner diameter, 10 µm tip, NewObjective, Woburn, MA) and was operated in the positive ion mode. The protocol and settings for desalting and separation in the MRM method was performed essentially as previously described by our group (135). Skyline was used to predict and optimize collision energies (CE) and declustering potential (DP) for each peptide. Q1 and Q3 were set to unit/unit resolution (0.7 Da) and the pause between mass ranges was set to 3 ms. MRM analysis was conducted with up to 152 transitions per run (dwell time, 15 ms; cycle time 3s).

2.1.1.8. Validation phase analysis by targeted LC/MS-MS using the designed MRM method.

In order to carry out the validation analysis, 23 peptides were selected and included in the final MRM method developed based on the following criteria: good signal in the MRM method, co-elution of at least 3 transitions and detection using the MIDAS workflow (136). With this aim, the best MRM transitions for these peptides were pooled in one scheduled-MRM method with a 45-min gradient. In this validation stage, 62 processed and desalted secretome samples (18N, 22UZ, 22WZ) were reconstituted in 7 µL of SIS stock solution and analyzed under the previous LC-MS/MS conditions using the nanoLC system (TEMPO, Eksigent) coupled to 5500-QTRAP (Sciex). **Table 2. 2** shows the final list of peptides quantified in this work, whereas all the monitored peptide transitions and the corresponding settings for their analysis are enumerated in **Annex II**.

2.1.1.9. Data analysis.

Peptide identification from raw data from the LTQ-Orbitrap was carried out using the SEQUEST algorithm (Proteome Discoverer 1.3, Thermo Scientific). The following constraints were used for the searches: no enzyme and tolerances of 10 ppm for precursor ions and 0.8 Da for MS/MS fragment ions, search against decoy database (integrated decoy approach) using false discovery rate (FDR) < 0.01. Data from the 5500 QTRAP were analyzed with ProteinPilot 4.0 (Sciex), using the Paragon algorithm as default search program using no enzyme and modifications criteria. Raw files from the MRM analysis were imported to Skyline and integration was manually inspected to ensure correct peak detection and accurate integration. The Protease Specificity Prediction Server (PROSPER) tool (137) was employed to search enzymes putatively involved in the cleavage of the endogenous peptides that had been identified in this work.

2.1.1.10. Statistical analysis.

A $p < 0.05$ was considered statistically significant and all statistical tests were two-sided. GraphPad Prism 5.0 (Graphpad, San Diego, CA, USA) was used to compare medians among the three different conditions of patients and healthy controls (WZ-UZ-Control), and a Kruskal–Wallis test's multiple comparison was conducted. Mann–Whitney U tests were performed to evaluate the significance of discrimination between the disease classes and the control cohort. Receiver operator characteristic (ROC) analysis was performed to quantify the overall ability of a peptide to classify the tissue as OA or healthy. The ROC curves were smoothed, compared and threshold computed using the R package pROC 2018 (138).

2.1.2. Neopeptide detection in synovial fluid and serum samples for OA monitoring.

2.1.2.1. Neopeptide extraction method comparison in serum.

In order to elect the endogenous peptide separation protocol for analyzing neopeptides in serum, different strategies were examined. Firstly, in order to deplete the most abundant/large proteins from serum and simplify the posterior peptidomic extraction, 3 different depletion techniques were followed: a chemical depletion technique using acetonitrile (ACN), the EtOH/TCA chemical depletion method already used in the previous chapter and a physical depletion method using 30kDa-size ultra-centrifugal filters. Secondly, an ultrafiltration (UF) using Amicon Ultra-4 devices (10 kDa MWCO, Merck Millipore, Bedford, MA) was performed for concentrating the endogenous peptides. Lastly, the resulting neopeptide-enriched solutions were desalted using homemade StageTips with different C18 Empore discs, depending on the processed serum quantity. Moreover, 50µL and 100µL of the same serum were used per test in order to study the influence of the initial serum load and the different neopeptide enrichment protocols.

An slightly modified ACN depletion protocol for peptidome enrichment was tested (139), which likely releases many carrier-bound molecular species in a high reproducible manner. It consisted in diluting 50/100µL serum 1:2 with ACN, shaking during 30’ and spinning for another 30’ at 13000rpm and 4°C. The supernatants were collected while the resulting pellets were discarded since they contained the precipitated most abundant proteins. Finally, the supernatants were diluted in 10 times more of the initial serum volume with 2% ACN 0.1% FA for conducting the subsequent 10kDa UF.

The EtOH/TCA strategy was carried out as in the previous chapter. Thus, 50/100µL serum were diluted 1/10 with ethanol 1% TCA and the resulting mixtures were shacked for 3 minutes and centrifuged for 5 minutes at 1500g and 4ºC. The pellets containing the albumin-depleted plasma were formed and kept, while the supernatants which mostly included TCA-albumin complex were discarded. Afterwards, the pellets were washed once with 400/700µL of methanol by manual shaking and centrifuging for 5 minutes at 2000g and 4ºC. At last, the supernatants were again dismissed and the remaining pellets were resuspended in 10 times more of the initial serum volume with 2% ACN 0.1% FA for performing the posterior 10kDa UF.

In order to perform the 30kDa UF procedure, 50/100µL serum were diluted 1/10 with 2% ACN 0.1% FA. Amicon devices were firstly rinsed with 200µL 2% ACN 0.1% FA in order to wet, clean and condition the filter. Firstly, 3/5 of the diluted serum were processed for 10' at 4500rpm and 4°C. Subsequently, the other 2/5 were added and gently mixed with the remaining unfiltered serum in order to free the filter and prevent sample clogging. Again, serum was centrifuged for another 10' at 4500rpm and 4°C. The extracted <30kDa fractions of the serum (approximately being the same initial dilution volume) were placed apart for later separating the neopeptide content by 10kDa UF.

Finally, once most of the proteomic content was depleted, the remaining low abundant and low molecular weight proteomic and peptidomic content was concentrated by 10kDa UF. The same process as in the previous 30kDa protocol was followed with only one slight change. Thus, Amicon devices were firstly rinsed with 200µL 2% ACN 0.1% FA and 3/5 of the resulting depleted serum solutions were centrifuged for 10' at 4500rpm and 4°C. Subsequently, the other 2/5 were added, gently mixed with the remaining unfiltered solutions and centrifuged for 20' at 4500rpm and 4°C. The <10kDa fractions of the serum were lastly dried in a vacuum concentrator, cleaned using homemade StageTips C18 (with 3 discs when

50µL serum were processed and 6 when 100µL serum were tested), dried again and stored at -20°C until LC-MS/MS analysis.

2.1.2.2. Shotgun LC-MS/MS analysis on serum and synovial fluid samples.

Firstly, in order to select the best neopeptide extraction method, the previously processed sera were analysed using a label-free strategy by Enhanced Mass and Enhanced Resolution (EM ER). Samples were reconstituted in 12µL of 5% ACN 0.1% FA and 5µL were injected onto the nanoLC system (TEMPO) coupled to a 5500-QTRAP instrument (AB Sciex) to detect the enriched neopeptides. After precolumn desalting using a C18 column (5 µm, 300A, 100 µm∗2 cm, Acclaim PepMap, Thermo Scientific, USA) at a flow of 3 µL/min during 10 min, neopeptides were separated on C18 nanocolumns (75 µmid, 15 cm, AcclaimPepMap 100, Thermo Scientific, USA) at a flow rate of 300 nL/min using a linear gradient of 120 min. The mass spectrometer was interfaced with a nanospray source equipped with an uncoated fused silica emitter tip (20 µm inner diameter, 10 µm tip, NewObjective, Woburn,MA) andwas operated in the positive ionmode. MS Source parameters were as follows: ion spray voltage (IS) 2600 V, interface heater temperature (IHT) 150 °C, ion source gas 2 (GS2) was 0, curtain gas (CUR) was 20 and ion source gas 1 (GS1) was 25 psi, and collision gas (CAD) high. MS compound parameters were set to 10 for the entrance potential (EP) and to 15 for the Collision cell exit potential (CXP). Each MS/MS fragmentation from the 5500 QTRAP was used to search for protein candidates using ProteinPilot software (AB Sciex, version 4.0). Due to the lack of accurate fragments generated in the QTRAP system, the search was conducted with low and high confidence settings (20% and 95% respectively). The enrichment protocol that identified more neopeptides was used for their analysis on serum and synovial fluid samples.

Once the neopeptide enrichment protocol was assessed, 9 serum samples (3 from hip OA patients, 3 from knee OA patients and 3 from healthy

donors) and 2 synovial fluids (SF) were prepared using the selected method. The processed samples were analyzed by shotgun LC-MS/MS in an Easy-nLC II system coupled to LTQ-Orbitrap-Velos-Pro mass spectrometer (Thermo Scientific) in order to identify a larger number of cartilage-related neopeptides in these fluids. The peptides were concentrated by reverse phase chromatography using a 0.1mm × 20 mm C18 RP precolumn (Proxeon), and then separated using a 0.075mm x 100 mm C18 RP column (Proxeon) operating at 0.3 µL/min. Peptides were eluted using a 90-min gradient from 5 to 40% solvent B (Solvent A: 0,1% FA in water, solvent B: 0,1% FA, 80% acetonitrile in water). ESI ionization was performed using a Nano-bore emitters Stainless Steel ID 30 µm (Proxeon) interface. The Orbitrap resolution was set at 30.000. Peptides were detected in survey scans from 400 to 1600 amu (1 µscan), followed by ten data dependent MS/MS scans (Top 10), using an isolation width of 2 m/z units (in mass-to-charge ratio units), normalized collision energy of 35%, and dynamic exclusion applied during 30 seconds periods.

2.1.2.3. Targeted LC-MS/MS analysis on serum samples and synovial fluids.

As for the secretome neopeptide validation, an MRM method was developed by selecting the peptides identified in the previous shotgun LC-MS/MS analysis of the 9 serum and 2 SF samples. The criteria for choosing them was the following: 1) peptides appertaining to ECM or OA-related proteins, 2) peptides corresponding to proteins commonly identified in cartilage secretome from the discovery phase, SF and serum specimens. The four most intense transitions for each suitable precursor were elected based on data deposited in the MS/MS library using the Skyline software. Several MRM methods were firstly designed, with less than 200 transitions examined per method. The MRM method development and LC-MS/MS analysis of serum/SF were performed exactly as for the 23 cartilage secretome neopeptide's one, also with the aid of the corresponding SIS peptides. Additional 9 serum and 4 SF samples were used for the development of an ultimate MRM method with 21 serum/SF neopeptides.

Lastly, two serum pools, one from 3 OA patients and another from 3 healthy donors, were analyzed using both cartilage secretome and serum/SF MRM methods in order to detect and quantify the 23 initially found cartilage-related neopeptides and the 21 new possible OA-related endogenous peptides. Thus, 100µL of each serum pool were depleted by the ACN method and subsequently enriched using by a 10kDa UF technique as previously described. After the processed serum pools were desalted and dried, 18µL of a mixture of the 21 SIS peptides (previously combined and diluted 1/5000) were used for reconstituting the samples and 4µL were later analyzed using the serum/SF MRM method.

In order to calculate the serum concentration of the detected neopeptides different calculations were performed in Excel. First, the SIS peptide quantity injected (µg) was calculated by multiplying the initial SIS concentration (mg/mL) by the 1/5000 SIS dilution factor (for sample reconstitution) and by 4µL of injection volume. Subsequently, NAT/SIS ratios were multiplied by the µg SIS peptide injected and by a factor of 18/4 in order to calculate the neopeptide (NAT) content in the initial 100µL serum. Neopeptide concentration in serum was expressed in ng/mL and/or pg/mL.

2.1.2.4. Data analysis.

Peptide identification was carried out from LTQ-Orbitrap raw data using the SEQUEST algorithm (Proteome Discoverer 1.3, Thermo Scientific) and PEAKS software. The following constraints were used for the searches: no enzyme and tolerances of 10 ppm for precursor ions and 0.8 Da for MS/MS fragment ions, search against decoy database (integrated decoy approach) using false discovery rate (FDR) < 0.01 and peptide/protein de novo sequencing. Data from the 5500 QTRAP were analyzed with ProteinPilot 4.0 (Sciex), using the Paragon algorithm as default search program using no enzyme and modifications criteria. Raw files

from the MRM analysis were imported to Skyline and integration was manually inspected to ensure correct peak detection and accurate integration. NAT/SIS ratios were measured and exported from Skyline as well. The Protease Specificity Prediction Server (PROSPER) tool (137) was employed to search enzymes putatively involved in the cleavage of the endogenous peptides that had been identified in this work.

2.2. Results.

2.2.1. Identification and targeted MS analysis of neopeptides secreted from articular cartilage.

2.2.1.1. Identification of endogenous peptides released from articular cartilage by shotgun LC-MS/MS.

Firstly, the histological characteristics of the cartilage explants were assessed by Mankin scoring, as previously stated, in order to define the differences between the tested explant samples (UZ, WZ and N) in accordance to the state of the cartilage. On the one hand, the healthy cartilages analyzed in this work had a Mankin score of 1.76±0.48, therefore defining them as normal tissue. On the other hand, in the OA tissue, explants were obtained both from the macroscopically normal zone (UZ) with an average Mankin score of 3.52±0.92 and the lesion (WZ) with a Mankin score of 8.38±1.47. Thus, taking into account the Mankin scoring classification, the explants from the defined UZ correspond to a cartilage with superficial fibrillation, whereas the ones from the WZ represent cartilages with severe damage.

The screening step resulted in the identification of 1175 different peptides corresponding to 101 unique proteins that were released from 8 hip or knee articular cartilages to the conditioned media. The complete list of the number of neopeptides that were identified per examined secretome, and their correspondent parent proteins, is shown in **Annex II**. A higher number of peptides was found in OA compared to normal tissue, although the result was not statistically significant (p= 0.17). The parent proteins identified with the highest score and highest number of peptides were ECM structural constituents, such as COMP, PRELP or Fibronectin (FINC). Several of them were specifically characteristic of the articular cartilage ECM, such as COMP, Cartilage Intermediate

Layer Protein 1 (CILP1) or Proteoglycan 4 (PRG4).

2.2.1.2. Development of targeted methods for the quantitative analysis of endogenous peptides released from articular cartilage.

From the identified 1175 endogenous peptides in the previous discovery phase, 54 peptides (belonging to 17 proteins) were explored for the development of the MRM method, which was carried out using secretome samples from 11 hip and 10 knee cartilages (**Table 2. 1**). These peptides were chosen under the already cited criteria: highest identification score (>3) in the screening phase, identified in the majority of samples and belonged to proteins expressed in articular cartilage. Then, the final MRM method was designed with the aid of SIS peptides for the detection and quantification of 23 endogenous peptides, which showed the best performance in the analysis. To confirm and normalize the resulting data, a scheduled MRM method including light/endogenous and heavy/SIS peptides was finally developed.

The parent proteins of these 23 peptides are Matrix Gla Protein (MGP), COMP, CILP1, PRELP, Dermcidin (DCD), FINC, Clusterin (CLUS), Glia Derived Nexin (GDN) and Collagen Alpha-1 (II) Chain (CO2A1). The list of endogenous peptides included in this targeted analysis is shown in **Table 2. 2**, as well as their corresponding molecular/biological functions, collected from UniProt resource (140).

Table 2. 2. MRM method for quantifying the 21 selected neopeptides found in articular cartilage secretomes. Bold letters indicate the stable isotope-labeled amino acid in each peptide. Protein functions were procured from the UniProt resource (Universal Protein Repository).

Protein Name	Function	Peptide Sequence
Matrix Gla protein	Associates with the organic matrix of bone and cartilage.	NANTFISPQQ**R**
	Inhibitor of bone formation.	NTFISPQQ**R**
Cartilage oligomeric matrix protein	Plays a role in the structural integrity of cartilage via its interaction with other ECM such as collagens and fibronectin and in the OA pathogenesis.	AEPGIQ**L**KAV
		AVAEPGIQ**L**K
		V**L**NQGREIVQT
Cartilage intermediate layer protein 1	Plays a role in cartilage scaffolding.	DEGDTFPL**R**
	Overexpression may lead to impair chondrocyte growth and matrix repair in aging and OA cartilage.	NLEPRTGF**L**SN
		STATAAQTD**L**NFIN
Prolargin	May anchor basement membranes to the underlying connective tissue.	DSNKIETI**P**N
		SDGVFK**P**DT
		SSD**L**ENVPH
		DLENVPHL**R**
Fibronectin	Involved in osteoblast compaction, essential for osteoblast mineralization.	SSGSGPFTDVRA**A**
		TSSGSGPFTDVRA**A**
Dermcidin	Displays antimicrobial activity limiting skin infection.	DAVEDLESVG**K**
	Exhibits proteolytic activity on the C-terminal of Arg and Lys.	ENAGEDPGLA**R**
Clusterin	Prevents stress-induced aggregation of blood plasma proteins. Inhibits formation of amyloid fibrils.	ASHTSDSDVPSGVTE**V**
		ASHTSDSDVPSGVTEV**V**
		GEDQYYLRVTT**V**
Glia-derived nexin	Serine protease inhibitor with activity toward thrombin, trypsin, and urokinase. Promotes neurite extension by inhibiting thrombin. Binds heparin.	SEDGTKASAATTAI**L**
		AVAQTDLKEP**L**KV
Collagen alpha-1 (II) chain	Specific for cartilaginous tissues, essential for embryonic development of the skeleton, linear growth and the ability of cartilage to resist compressive forces	AGPPGPVG**P**AGGP
		AGPSGPRGPPG**P**VGP

Moreover, the area under the curve for the endogenous peptides (obtained from Skyline) was plotted for each peptide in cartilage samples from the UZ and WZ of OA and healthy tissues. Certain peptides belonging to CILP1 (DEGDTFPLR) and PRELP (DSNKIETIPN, DLENVPHLR) were found to be mostly increased in the WZ of OA cartilages when compared to UZ and/or healthy donors. A representative example of the results obtained with this analysis can be seen for the peptide DSNKIETIPN in **Figure 2. 1**.

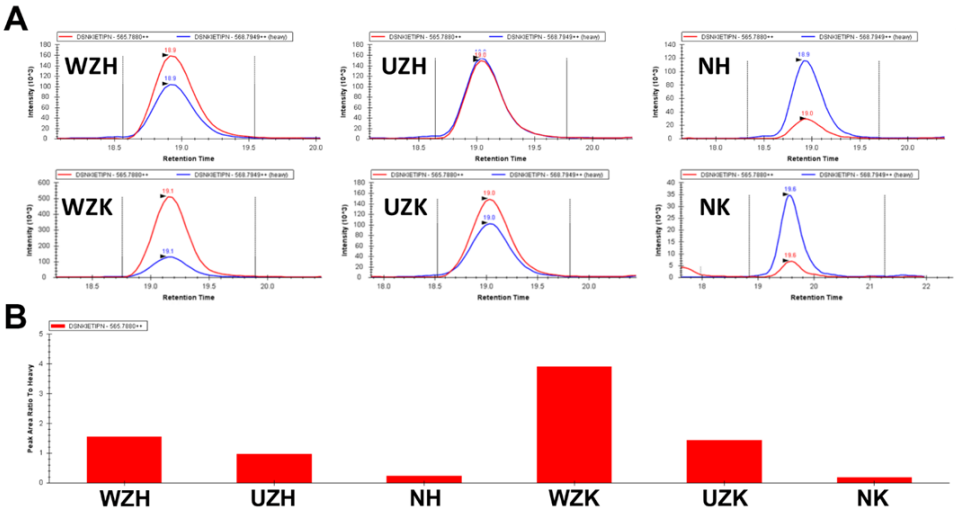


Figure 2. 1. Multiple Reaction Monitoring (MRM) MS quantification of endogenous peptides. **A)** Representative chromatograms of the endogenous peptide PRELP_DSNKIETIPN in a pool of secretome samples (n=3) from hip (upper row) and knee (lower row). The endogenous peptides (light) are represented in red, whereas the SIS peptides (heavy) are displayed in blue. The spiked SIS amount into each sample was kept constant. **B)** Chart plot representing the peak area ratio normalized to the heavy peptide standard for each type of sample. WZH, wounded zone from OA hip; UZH, unwounded zone from OA hip; NH, healthy hip; WZK, wounded zone from OA knee; UZK, unwounded zone from OA knee; NK, healthy knee.

2.2.1.3. Quantification of endogenous peptides in OA and healthy cartilage secretomes.

The validation study was carried out using the scheduled MRM method with the corresponding SIS peptides on 62 secretome samples obtained from hip (n=33) and knee (n=29) cartilages. All the quantification data (expressed as peak area ratios of light/heavy peptides) from the peptides in the secretome of different zones of OA cartilage (UZ and WZ) and healthy donors in the different joints are shown in **Annex II**. After statistical analysis of the results, four endogenous peptides were found to be differentially released from OA cartilage compared to healthy tissue with a significant p-value. Among these, two peptides from PRELP (DSNKIETIPN and DLENVPHLR) and one from MGP (NTFISPQQR) were

differentially released independently of the OA cartilage zones (**Figure 2. 2A**). Furthermore, the same tendency was detected in the OA WZ compared to control donors for these peptides and the peptide DEGDTFPLR from CILP1. Thus, all of them were increased in the OA WZ vs healthy cartilage secretomes (**Figure 2. 2B**). Finally, the peptide DSNKIETIPN (PRELP) was differentially released in the UZ compared to normal cartilage, and also between the two OA cartilage zones.

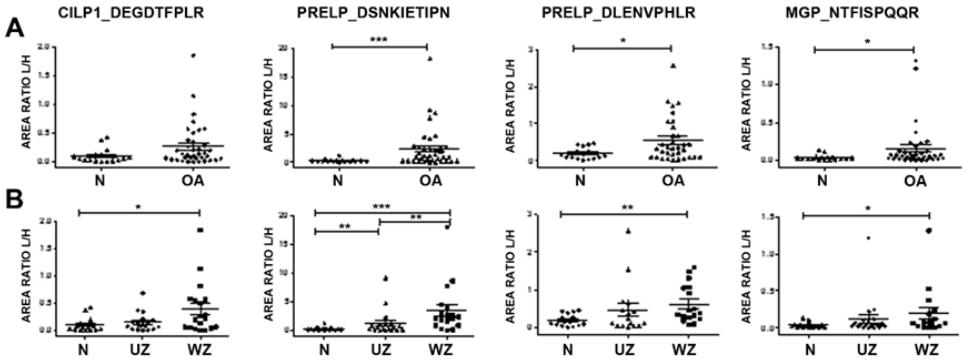


Figure 2. 2. Differential endogenous peptides released from osteoarthritic articular cartilage. **A)** Comparison between OA (n=44) and normal tissue (n=18). **B)** OA samples were classified into those from the unwounded zone of the tissue (UZ, n=22) and from the wounded (WZ, n=22). The results are expressed as area ratios (light/heavy, L/H). Data were analyzed using Mann-Whitney test and plotted as means ± SEM for each condition. p*<0.05, p**<0.005 p***<0.0005.

2.2.1.4. Differential release of endogenous peptides from knee and hip articular cartilages.

The precedent targeted peptide quantification evidenced a differential release of certain neopeptides depending on the joint that was studied (p<0.05), which are shown in **Table 2. 3** and **Figure 2. 3**.

Table 2. 3. Fold changes of endogenous peptides differentially released from knee and hip articular cartilage with a significant p-value (<0.05). Data obtained using the MS stats tool from Skyline software.

Knee vs Hip cartilage			
Protein	**Peptide Sequence**	**Fold Change**	**Mann Whitney test (p-value)**
sp|P08493|MGP_HUMAN	**NANTFISPQQR**	2.29 (95% CI:1.05 to 5)	0.0174
sp|P08493|MGP_HUMAN	**NTFISPQQR**	3.03 (95% CI:1.43 to 6.45)	0.0006
sp|P49747|COMP_HUMAN	**AEPGIQLKAV**	5.11 (95% CI:1.98 to 13.18)	0.0039
sp|P51888|PRELP_HUMAN	**DSNKIETIPN**	3.88 (95% CI:1.73 to 8.73)	0.0035
OA Knee vs OA Hip cartilage			
Protein	**Peptide**	**Fold Change Result**	**P-Value Mann Whitney**
sp|P10909|CLUS_HUMAN	**GEDQYYLRVTTV**	3.47 (95% CI:1.15 to 10.45)	0.0296
sp|P10909|CLUS_HUMAN	**ASHTSDSDVPSGVTEV**	4.43 (95% CI:1.29 to 15.24)	0.0383
sp|P49747|COMP_HUMAN	**AEPGIQLKAV**	8.09 (95% CI:2.55 to 25.64)	0.0251
sp|P51888|PRELP_HUMAN	**DSNKIETIPN**	3.47 (95% CI:1.17 to 10.29)	0.0319
Healthy Knee vs Healthy Hip cartilage			
Protein	**Peptide**	**Fold Change Result**	**P-Value Mann Whitney**
sp|P08493|MGP_HUMAN	**NTFISPQQR**	3.54 (95% CI:1.3 to 9.67)	0.0177

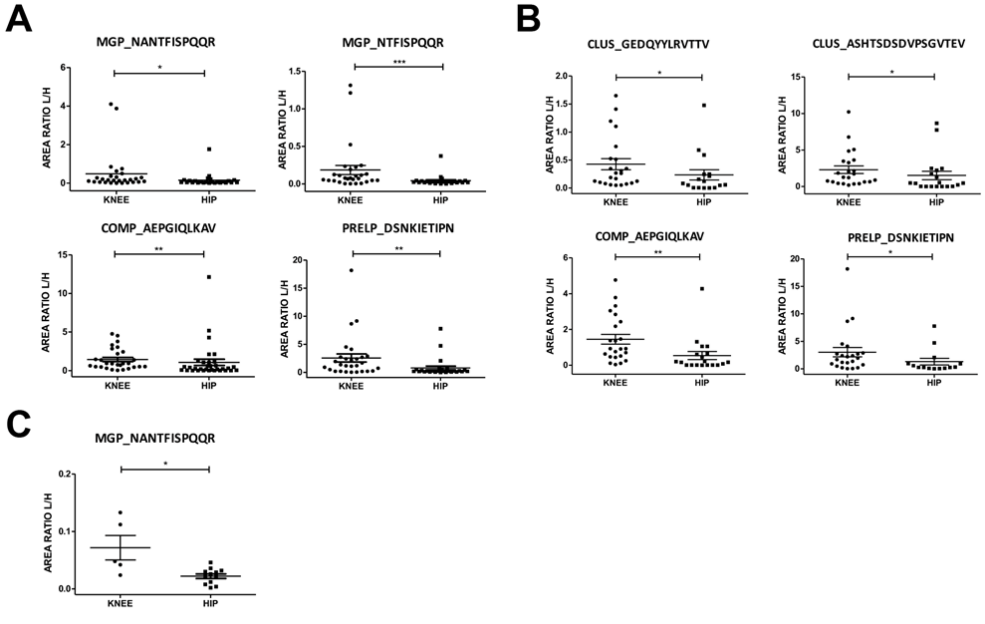


Figure 2. 3. Differential release of endogenous peptides from hip and knee articular cartilages. Scattering plots showing the distribution of the Area light/heavy (L/H) ratios of representative endogenous peptides. The data were analyzed using Mann-Whitney test and plotted as means ± SEM for each condition. **A)** Knee (n=29) vs hip (n=33), **B)** OA knee (n=23, 12 WZ and 11 UZ) vs OA hip (n=20, 10 WZ and 10 UZ), and **C)** Healthy knee (n=5) vs healthy hip (n=13). p* < 0.05, p**<0.005 p***<0.0005.

In all cases, the release was higher from the knee tissue. Comparison of the conditioned media of all knee (n=29) and hip (n=33) cartilage samples demonstrated the increased release from knee of endogenous peptides corresponding to the MGP (NANTFISPQQR and NTFISPQQR), COMP (AEPGIQLKAV) and PRELP (DSNKIETIPN), with fold changes ranging from 2.29 to 5.11 (**Figure 2. 3A**). In OA cartilage, the peptide AEPGIQLKAV (COMP) has a remarkable 8-fold change ratio higher in knee vs hip, while DSNKIETIPN from PRELP and GEDQYYLRVTTV and ASHTSDSDVPSGVTEV from CLUS also showed significant differences (**Figure 2. 3B).** Considering only the healthy tissues (knee n=5 and hip n=13), one peptide was increased in the knee samples (NTFISPQQR, from MGP) with a fold ratio of 3.54 (**Figure 2. 3C**). Given these joint-characteristic profiles, the differences in the release of neopeptides were examined independently in hip and knee.

In hip samples, two peptides from CLUS were increased in the conditioned media of healthy cartilage compared to OA tissue: ASHTSDSDVPSGVTEVV and GEDQYYLRVTTV (**Figure 2. 4A**). When the different zones in the diseased cartilage were considered (**Figure 2. 4B**), these two peptides showed a significant lower release from the wounded zone of the tissue (WZ) in comparison to the healthy tissue. The same happens with another peptide from CLUS, ASHTSDSDVPSGVTEV, and the AEPGIQLKAV peptide from COMP.

In knee samples, two endogenous peptides from PRELP were significantly increased in the conditioned media of OA tissue from N: DSNKIETIPN and DLENVPHLR (**Figure 2. 5A**). Considering the two zones of OA tissue separately, these two peptides showed an enhanced release in the WZ from N (**Figure 2. 5B**). Besides, the peptide DEGDTFPLR from CILP1 displayed a similar tendency. Interestingly, the peptide DSNKIETIPN exhibited the most significant differences, which were also detectable in the macroscopically normal zone of OA tissue (UZ) vs WZ.

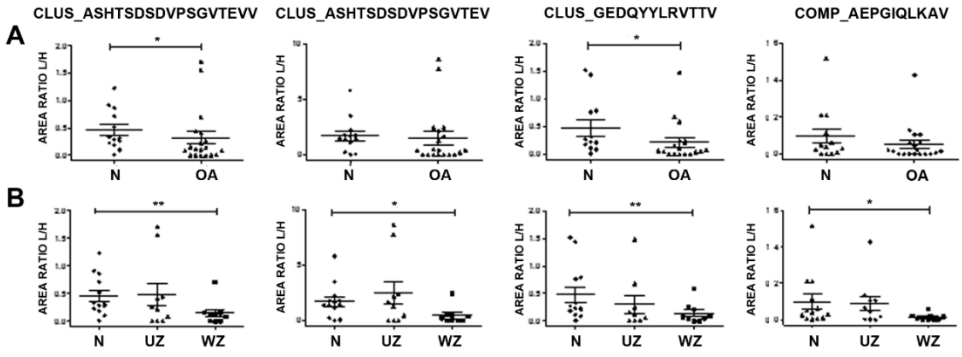


Figure 2. 4. Differential endogenous peptides released from hip articular cartilage. Scattering plots show peptide abundances in hip cartilage secretomes. **A)** Comparison between OA (n=20) and normal tissue (n=13). **B)** OA samples were classified into those from unwounded zones (UZ, n=10) or wounded zones (WZ, n=10). The results are expressed as area ratios (light/heavy, L/H). Data were analyzed using Mann-Whitney test and plotted as means ± SEM for each condition. P*<0.05 and p**<0.005.

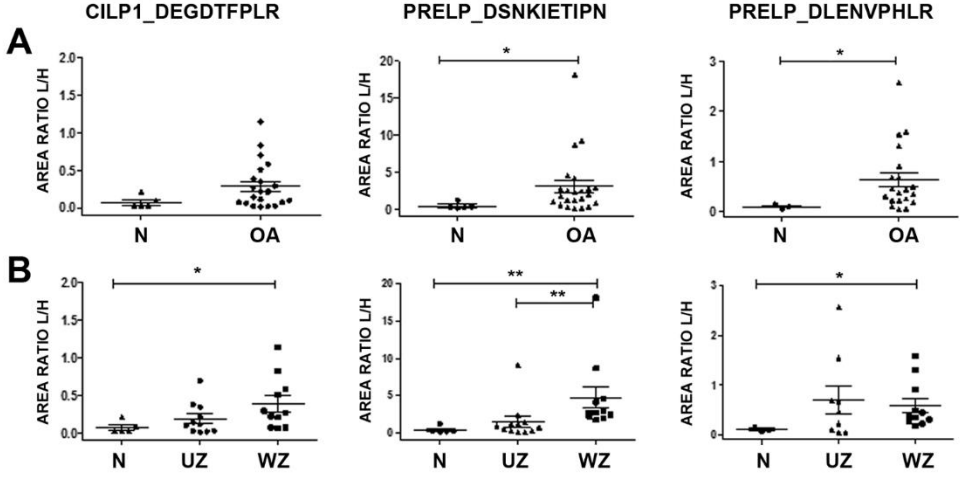
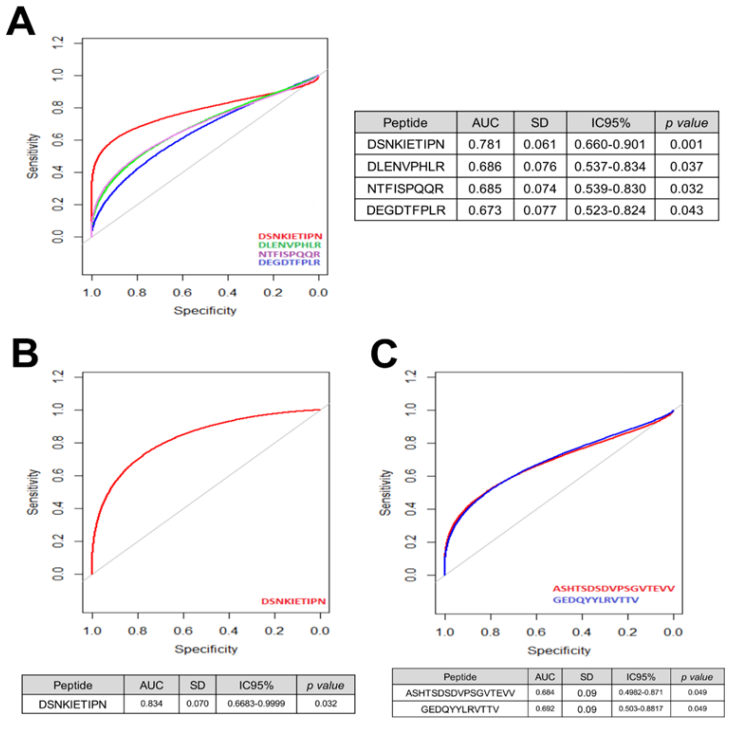


Figure 2. 5. Differential endogenous peptides released from knee articular cartilage. Scattering plots showing the abundance of each peptide in knee cartilage secretomes. **A)** Comparison between OA (n=24) and normal tissue (n=5). **B)** OA samples were classified into those from unwounded zones (UZ, n=12) or wounded zones (WZ, n=12). The results are expressed as area ratios (light/heavy, L/H). Data were analyzed using Mann-Whitney test and plotted as means ± SEM for each condition. P*<0.05 and p**<0.005.

2.2.1.5. Value of the identified peptides as biomarkers of articular cartilage degradation.

To evaluate the putative biomarker value of the endogenous peptides that have been identified, an analysis by receiver operator characteristic (ROC) curves was performed. As illustrated in **Figure 2. 6A**, the peptide DSNKIETIPN showed an area under the curve (AUC) of 0.781 [IC 95%: (0.660-0.901), p= 0.001], being the best candidate to discriminate healthy vs OA tissue independently of the target joint. Considering only the knee, the AUC of this peptide increased up to 0.834 (**Figure 2. 6B**). On the other hand, two peptides from CLUS (ASHTSDSDVPSGVTEVV and GEDQYYLRVTTV) displayed significant AUCs when analyzing the hip tissue exclusively (**Figure 2. 6C**).



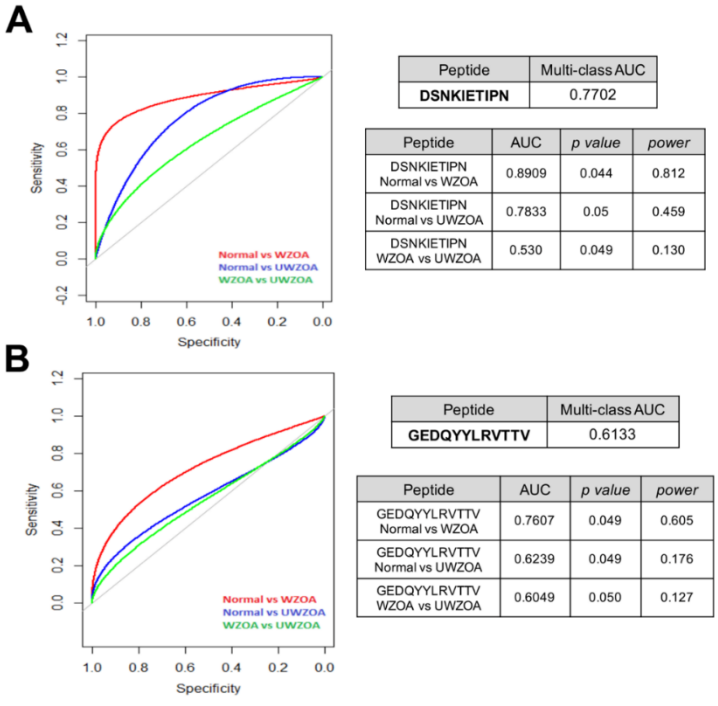
Peptide	AUC	SD	IC95%	*p value*
DSNKIETIPN	0.781	0.061	0.660-0.901	0.001
DLENVPHLR	0.686	0.076	0.537-0.834	0.037
NTFISPQQR	0.685	0.074	0.539-0.830	0.032
DEGDTFPLR	0.673	0.077	0.523-0.824	0.043

Peptide	AUC	SD	IC95%	*p value*
DSNKIETIPN	0.834	0.070	0.6683-0.9999	0.032

Peptide	AUC	SD	IC95%	*p value*
ASHTSDSDVPSGVTEVV	0.684	0.09	0.4982-0.871	0.049
GEDQYYLRVTTV	0.692	0.09	0.503-0.8817	0.049

Figure 2. 6. Receiver operator characteristic (ROC) curves of the biomarker peptides identified in this work. **A)** The release of four peptides discriminates OA vs healthy articular cartilage with significant p value ($p<0.05$), **B)** The peptide DSNKIETIPN from prolargin differentiates knee OA from healthy tissue, and **C)** Two peptides from clusterin discriminate hip OA from healthy tissue.

Finally, this analysis was also performed splitting the OA tissue in zones (**Figure 2. 7**). In this case, again the best results were obtained for the peptide DSNKIETIPN in knee, showing a good biomarker value (AUC= 0.783) in OA but macroscopically normal cartilage. Comparing healthy knee tissue with the damaged zones of knee OA, this AUC increased up to 0.891. In hip, the performance of GEDQYYLRVTTV was worse, but still highly significant (AUC normal vs WZOA= 0.761).



Peptide	Multi-class AUC
DSNKIETIPN	0.7702

Peptide	AUC	*p value*	*power*
DSNKIETIPN Normal vs WZOA	0.8909	0.044	0.812
DSNKIETIPN Normal vs UWZOA	0.7833	0.05	0.459
DSNKIETIPN WZOA vs UWZOA	0.530	0.049	0.130

Peptide	Multi-class AUC
GEDQYYLRVTTV	0.6133

Peptide	AUC	*p value*	*power*
GEDQYYLRVTTV Normal vs WZOA	0.7607	0.049	0.605
GEDQYYLRVTTV Normal vs UWZOA	0.6239	0.049	0.176
GEDQYYLRVTTV WZOA vs UWZOA	0.6049	0.050	0.127

Figure 2. 7. Receiver operator characteristic (ROC) curves of the best biomarker peptides differentiating disease and zone in knee **(A)** or hip **(B)** articular cartilage. The inset tables show the metrics obtained for each peptide in normal (healthy), unwounded (UW) or wounded (W) zones of OA tissue from each joint.

An experimental workflow of the steps followed in this study can be seen in **Figure 2. 8**. At the end of the validation phase, 8 neopeptides from PRELP, MGP, CILP1, CLUS and COMP were found to be candidate markers of the cartilage turnover in OA. On the one hand, two peptides from PRELP (DSNK and DLEN), one from MGP (NTFI) and another from CILP1 (DEGD) significantly differentiate articular cartilage from patients with OA and healthy donors, independently from

the examined joint. On the other hand, the same peptides from PRELP and CILP1 can differ between OA damaged and healthy knee tissues, whereas three peptides from CLUS (ASHT…V, ASHT…VV and GEDQ) and one peptide from COMP (AEPG) discern hip cartilages from OA patients and healthy donors. Altogether, these 8 neopeptides (coming from the degradation of these 5 proteins) could be useful to assess the cartilage turnover for monitoring the OA diagnosis (N vs OA) and progression on the early stages of the disease (N vs UZ) as well as on the last stages (N vs WZ and UZ vs WZ).

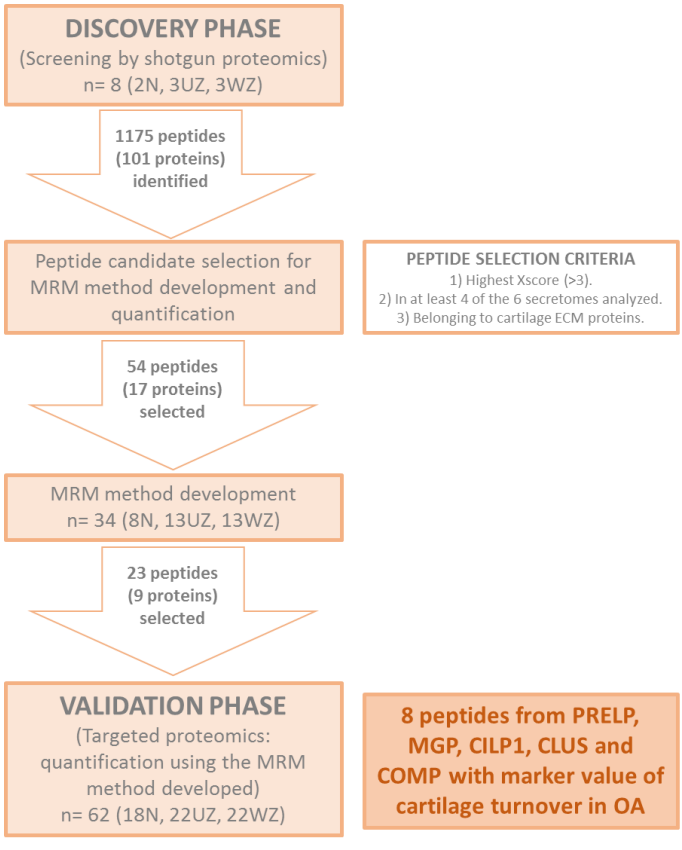


Figure 2. 8. Experimental workflow for the analysis of neopeptides of articular cartilage.

2.2.1.5. Neopeptide enzymatic likely-cleavage.

To sum up, 8 neopeptides were significantly found to be differentiated between OA and N tissue. After searching by PROSPER tool the likely-acting endogenous enzymes, different proteases were found to be acting over the 8 described neopeptides (**Table 2. 4**).

Table 2. 4. Enzymes likely acting over the corresponding neopeptide. Asterisks show the enzyme cleaving the starting or the ending aminoacid of the peptide.

Protein_Neopeptides	Metalloprotease	Serine protease
CILP1_DEGDTFPLR	MMP-9	Glutamyl peptidase I
PRELP_DSNKIETIPN		Cathepsin G, Glutamyl peptidase I, elastase-2
PRELP_DLENVPHLR		
MGP_NTFISPQQR		Chymotrypsin A (cattle-type)
COMP_AEPGIQLKAV	MMP-9*	Elastase-2
CLUS_ASHTSDSDVPSGVTEV	MMP-9	Elastase-2, Glutamyl peptidase I
CLUS_ASHTSDSDVPSGVTEVV	MMP-9	Elastase-2, Glutamyl peptidase I
CLUS_GEDQYYLRVTTV		Signalase (animal) 21 kDa component, Glutamyl peptidase I

2.2.2. Neopeptide detection in synovial fluid and serum samples for OA monitoring.

2.2.2.1. Neopeptide identification in synovial fluid and serum samples by shotgun LC-MS/MS.

First of all, the same serum sample was processed using 3 different methodologies in order to enrich the neopeptide content and select the one that displays the most satisfactory performance and identifies the highest number of neopeptides. As stated, precedent physical and chemical depletions (ACN, EtOH/TCA and 30kDa UF) were used for removing the most abundant/large serum proteins, while a posterior 10kDa UF was used for enriching the neopeptide serum load. **Table 2. 5** shows the number of neopeptides and their corresponding proteins identified with the ProteinPilot software at 20% and 95% confidence levels after performing the EM ER analysis of the differently processed serum. In total, 6 conditions were compared when taking into account the 3 depletion methods and the 2 initial serum loads tested. The ACN depletion method of 100µL serum turned to be the most satisfactory strategy for detecting a larger number of neopeptides when compared to the other cases (265 at 95% confidence and 605 at 20% confidence). Interestingly, the depletion of a larger serum load (100µL) leads to an increase in the number of peptides identified in both chemical depletions (ACN and EtOH/TCA), whereas this number decreases when the physical UF depletion is performed (from 555 peptides to 478 at 20% confidence and from 179 to 121 at 95% confidence).

Table 2. 5. Number of proteins and the corresponding neopeptides identified in the EM ER analysis of differently processed serum loads. Identifications were attained with the ProteinPilot software at different confidences (20% and 95%).

Depletion method	Serum Load (µL)	Number (20% Conf)		Number (95% Conf)	
		Proteins	Peptides	Proteins	Peptides
30kDa UF	50	71	555	6	179
	100	54	478	4	121
ACN	50	64	557	1	63
	100	70	605	7	265
EtOH/TCA	50	69	445	2	46
	100	55	513	5	122

Additionally, the precedent processed samples were analyzed using the previous cartilage secretome MRM method although none of the 23 neopeptides were identified. Thus, a new discovery LC-MS/MS strategy was followed for the identification of cartilage/OA-related neopeptides in serum and SF samples. For this aim, 100µL of 9 sera and 2 SF samples, from OA patients and healthy donors, were depleted using the ACN method followed by a 10kDa UF neopeptide enrichment. The resulting cleaned SF and serum neopeptides were analyzed by shotgun MS using an LTQ-Orbitrap.

2.2.2.2. Development of MRM methods for the quantitative analysis of neopeptides in synovial fluid and serum.

More than 1000 neopeptides were conjointly identified in the examined SF and serum samples from the previous shotgun strategy. In order to develop another MRM method for monitoring these SF and serum neopeptides, 253 peptides appertaining to different 45 proteins were selected based on the already established criteria (**Annex II**). Therefore, 6 MRM methods were designed and a final scheduled MRM method including 16 proteins, 21 peptides, 42 precursors and 152 transitions was accomplished (**Table 2. 6**) for the detection of neopeptides in serum and SF samples. Hence, 9 serum and 4 SF samples were used for carrying out the previous MRM development with the aid of the

corresponding SIS peptides. Moreover, these samples were also analyzed using the cartilage secretome MRM method (23 neopeptides), although again, none of the 23 neopeptides were detected in these samples.

Table 2. 6. MRM method for quantifying the selected 23 endogenous peptides identified in serum and synovial fluid samples. Parenthesis indicate the stable isotope-labelled amino acid in each peptide.

Protein name	Protein Accesion Number (Gene)	Peptide Sequence
Actin, cytoplasmic 1	P60709 (ACTB_HUMAN)	**Q(K)DSYVGDEAQS Y(V)GDEAQSKRG**
Alpha-2-HS-glycoprotein	P02765 (FETUA_HUMAN)	**SLGSPSGEVSHP(R)**
Alpha-2-macroglobulin	P01023 (A2MG_HUMAN)	**HGPEG(L) FYESD(V)M**
Apolipoprotein A-I	P02647 (APOA1_HUMAN)	**EEYTKK(L)N**
Apolipoprotein A-IV	P06727 (APOA4_HUMAN)	**QVNTQAEQ(L) SLAELGGHLDQQVEE(F)**
Apolipoprotein E	P02649 (APOE_HUMAN)	**TVGS(L)AG**
Complement C3	P01024 (CO3_HUMAN)	**IHWESASLL(R)**
Dermcidin	P81605 (DCD_HUMAN)	**PGLARQ(A)P**
Inter-alpha-trypsin inhibitor heavy chain H1	P19827 (ITIH1_HUMAN)	**DRVTG(V)DTD**
Inter-alpha-trypsin inhibitor heavy chain H4	Q14624 (ITIH4_HUMAN)	**MNFRPGV(L)S GLPGPPDVPDHAAYHP(F) NVHSGSTF(F)**
Kininogen-1	P01042 (KNG1_HUMAN)	**RPPGFSP(F)**
Malate dehydrogenase, mitochondrial	P40926 (MDHM_HUMAN)	**LGIGKVSS(F)E**
Matrix Gla protein	P08493 (MGP_HUMAN)	**NTFIS(P)QQ**
Sulfhydryl oxidase 1	O00391 (QSOX1_HUMAN)	**AAPGQEP(P)EHMAE**
Transcription factor Maf	O75444 (MAF_HUMAN)	**AGGAGGAGGGG(P)AS**
Transthyretin	P02766 (TTHY_HUMAN)	**YSTTAVVTN(P)KE**

2.2.2.3. Quantification of endogenous peptides in sera from OA patients and healthy donors.

Finally, 2 serum pools, one from OA patients (OA, 3) and another from healthy donors (HD, 3) were analyzed with both the previous cartilage secretome MRM method (monitoring 23 neopeptides) and the new serum/SF MRM method (assessing 21 neopeptides). Repeatedly, none of the 23 neopeptides from the cartilage secretome MRM method was detected while 6 peptides from KNG1 (RPPGFSPF), APOA4 (SLAELGGHLDQQVEEF), CO3 (IHWESASLLR) and ITIH4 (NVHSGSTFF, GLPGPPDVPDHAAYHPF and MNFRPGVLS) were successfully

identified with the aid of the corresponding 6 peptides. The MRM analysis of these peptides in the OA and HD pools is shown in **Table 2. 7**, where NAT/SIS ratios were obtained by measuring the peak areas of both the neopeptide (NAT) and the corresponding SIS peptide peaks using the Skyline software. Neopeptide concentrations in serum were calculated as explained using the NAT/SIS ratios and expressed in ng/mL and/or pg/mL serum.

Table 2. 7. MRM analysis of the 6 detected neopeptides from KNG1, APOA4, CO3 and ITIH4 in OA and N serum pools. The initial SIS peptide concentration is located on the right of each peptide and expressed in mg/mL. NAT/SIS ratios were exported from Skyline after peak processing.

MRM analysis	KNG1_RPPGFSPF	4 mg/mL	APOA4_SLAELGGHLDQQVEEF	4,8 mg/mL
	µg SIS injected (4µL)	0,0032	µg SIS injected (4µL)	0,00384
	NAT/SIS	ng/mL serum	NAT/SIS	ng/mL serum
OA pool	0,3075	44,28	0,02	3,46
HD pool	0,1598	23,01	0,0561	9,69
MRM analysis	CO3_IHWESASLLR	3,3 mg/mL	ITIH4_NVHSGSTFF	3,3 mg/mL
	µg SIS injected (4µL)	0,00264	µg SIS injected (4µL)	0,00264
	NAT/SIS	pg/mL serum	NAT/SIS	ng/mL serum
OA pool	0,0067	795,96	0,0185	2,20
HD pool	0,0074	879,12	0,0162	1,92
MRM analysis	ITIH4_GLPGPPDVPDHAAYHPF	2,8 mg/mL	ITIH4_MNFRPGVLS	3,8 mg/mL
	µg SIS injected (4µL)	0,00224	µg SIS injected (4µL)	0,00304
	NAT/SIS	ng/mL serum	NAT/SIS	pg/mL serum
OA pool	0,4959	49,99	0,0004	54,72
HD pool	0,2783	28,05	0,0005	68,40

As seen in **Figure 2. 9**, neopeptides one peptide from KNG1 (RPPGFSPF) and two peptides from ITIH4 (GLPGPPDVPDHAAYHPF and NVHSGSTFF) were increased in the OA serum pool; whereas the other detected neopeptide from ITIH4 (MNFRPGVLS), a peptide from APOA4 (SLAELGGHLDQQVEEF) and a peptide from CO3 (IHWESASLLR) showed the opposite trend and were elevated in the HD serum pool. Moreover, 4 of these neopeptides were in the ng/mL range while the other 2 stayed around the lower pg/mL level.

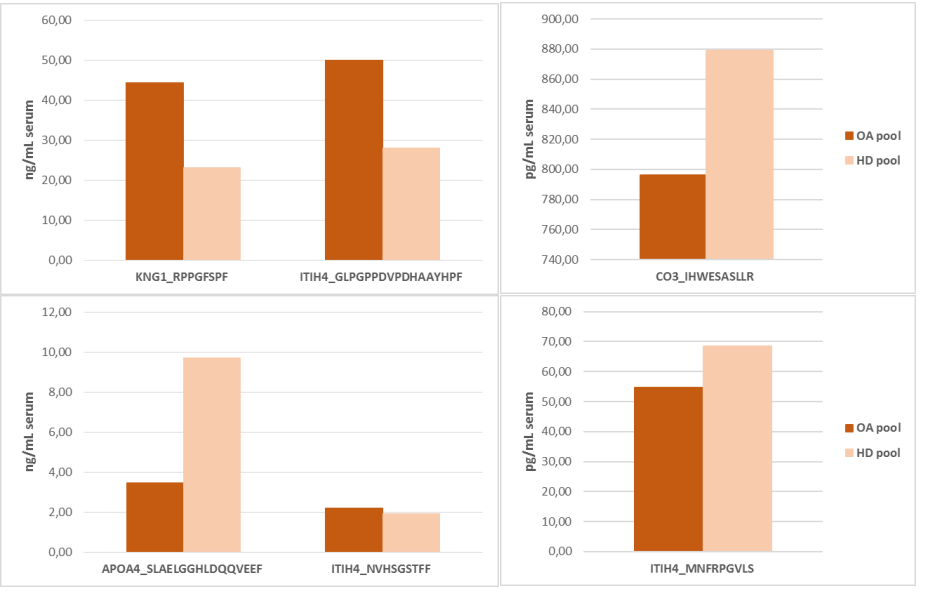


Figure 2. 9. Graphs showing the concentrations of the 6 MRM-detected neopeptides in OA/HD serum pools.

2.2.2.4. Neopeptide enzymatic likely-cleavage.

After searching the specific enzymes likely-acting on these natural protein digestions, cysteine and serine proteases and metalloproteases were found to be related with some of the detected neopeptides using the PROSPER tool (**Table 2. 8**). Most of these proteases are presumably related with the OA progression and cartilage turnover as later discussed.

Table 2. 8. Enzymes likely acting over the corresponding neopeptide. Asterisks show the enzyme cleaving the starting or the ending amino acid of the peptide.

Protein_Neopeptides	Cysteine protease	Metalloprotease	Serine protease
APOA4_SLAELGGHLDQQVEEF	Cathepsin K		Chymotrypsin A (cattle-type)*
KNG1_RPPGFSPF		MMP-3	
CO3_IHWESASLLR	Cathepsin K	MMP-9	Cathepsin G*, Elastase-2
ITIH4_GLPGPPDVPDHAAYHPF			Cathepsin G, Thylakoidal processing peptidase
ITIH4_NVHSGSTFF			Elastase-2
ITIH4_MNFRPGVLS		MMP-9*	Cathepsin G

2.3. Discussion.

Peptides are constantly generated in vivo either by active synthesis and proteolytic processing of larger precursor proteins, often yielding protein fragments that mediate a variety of physiological or pathological functions. Given that abnormal proteolysis is a hallmark of various diseases, many studies turned to focus on the peptidome as a source of biomarkers (86). The investigation of peptides in a system-wide manner could facilitate the identification of potential biomarkers, the identification of protease-substrate relationships and the profiling of pathological degradation processes. Considering that the process of articular cartilage ECM degradation is a hallmark for OA, a neopeptidomic profiling of this pathological situation was performed, for the first time, without the use of any exogenous stimulus that may provide artifact results. Previous studies on endogenous peptides in OA have all employed models using either well known OA-related proteinases (141) or inducers of cartilage degradation such as mechanical damage or proinflammatory cytokines (142, 143).

On the one hand, the peptidomic analysis started with a first discovery phase on conditioned media from cartilage explants, which is likely to be the deepest characterization of cartilage neopeptides. Interestingly, more peptides and with higher signals were detected in secretomes from knee samples than from hip, which depicts the differences between these two joints and also indicates a higher turnover in the knee that was not described in previous proteomic analyses performed directly on the tissue (144, 145). Data mining showed that most of the identified proteins were cartilage ECM proteins or proteins with well-established matrix functions, such as collagens and proteoglycans. Although some of the parental proteins of many of these neopeptides have been reported for the first time in cartilage-derived samples (such as salivary acidic proline-rich phosphoprotein 1/2) many of them had been previously associated with OA. Notably, this neopeptide list includes the detection

of previously known OA biomarkers, such as CTXII (peptides GPDPLQYMRA, DPLQYMRA and SAFAGLGPRE, from the C-telopeptide fragment of type II collagen). Altogether, these facts further evidence the usefulness of secretome analysis as a source of cartilage-characteristic biomarkers (143, 146).

Next, in the validation step, a targeted MS-based method was developed for the quantification of a selected panel of these neopeptides in secretomes. MRM enables the multiplexed monitoring of neopeptides, and also of modified variants of specific proteins (147), without the need of developing complex targeted immunoassays (148) and providing a higher specificity than antibody-based tests. The present method was then applied for an exhaustive analysis on 62 secretomes from articular cartilage, which allowed to obtain statistically significant results of the differences. Eight endogenous peptides from PRELP, MGP, CILP1, CLUS and COMP were found to be differentially released from OA compared to healthy tissue.

Remarkably, decreased amounts of three neopeptides from CLUS and one from COMP were found in hip OA samples (**Figure 2. 4**). This is in accordance with the disease-related significant decrease of these two proteins in articular cartilage that has been recently described (145). CLUS, also known as Apolipoprotein J, is a secreted protein that regulates apoptosis and inflammation. A few studies have observed elevated CLUS in cartilage and synovial fluid in early OA (149, 150). Furthermore, increased CLUS levels in SF and serum showed statistically significant associations with joint space narrowing after adjustment for age and sex (151). However, IL-1α-stimulated cartilage explants have shown to produce decreased levels of CLU compared to untreated cartilage (145, 152). An analogous discrepancy happens with COMP: although this protein is decreased in knee and hip OA articular cartilage ($p=0.007$) (145), it is well known that its elevated levels in serum are associated with OA severity (127, 153). An explanation for this might be that these higher levels of CLUS and COMP in OA SF and serum/plasma could

represent the activation of a compensatory, but ultimately ineffective, protective pathway. The cartilage COMP degradation with OA progression and the appearance of elevated COMP levels in serum could also be due to a protein leak, from the cartilage compartment to the blood torrent, as COMP degrades. In knee, the disease-related increased release of one neopeptide from CILP1 and two from PRELP was observed. This increment was significant from the WZ zones of the tissue in all cases, although in the case of the peptide DSNKIETIPN from PRELP it was also detectable in the macroscopically normal zone in comparison to the wounded zone.

Furthermore, the ROC analysis showed the best results for this peptide (**Figure 2. 6**), with and AUC of 0.834 for the classification of the tissue as OA or healthy. Interestingly, an slightly longer DSNKIETIPN neopeptide from PRELP was identified in a previous study as the relatively most abundant peptide from an in vitro digestion of cartilage with ADAMTS4 (141). The contribution of the aggrecanases ADAMTS4 and ADAMTS5 to cartilage destruction in OA has been widely established (14, 154), although it has not been completely resolved. PRELP is a small leucine-rich proteoglycan highly abundant in cartilage (155, 156) that binds the basement membrane heparan sulfate proteoglycan perlecan through its N-terminal region, and collagens (type I and II) through its 12 leucine-rich repeat (LRR) domains. An increase in DSNKIETIPN, localized in the 7th LRR domain of the protein, denotes PRELP breakage with a loss of half its LRR domains for collagen binding. Thus, the statistically significant increase of this neopeptide in OA cartilage that is demonstrated in the present work depicts the role of PRELP as mediator of ADAMTS4 catabolic effects in articular cartilage.

On the other hand, different strategies were compared in order to concentrate the neopeptide content in serum and SF samples. The development of sensitive analytical techniques for effective, efficient and reproducible peptidome isolation and analysis still remains a challenge. Nevertheless, every

peptidomic extraction technique displays several advantages and disadvantages, which need to be tested and identified for achieving the most adequate performance (157). It is important to take into consideration that different peptidomic isolation methods produce unique sets of peptide ion masses (158). Thus, complementary combinations of tandem peptide extraction methods are likely required in order to effectively cover the serum peptidome. In this study, a preliminary ACN precipitacion for the depletion of the most abundant serum proteins and a subsequent 10kDa centrifugal ultrafiltration for peptidome enrichment were found to be the most satisfactory techniques for neopeptide isolation. The highest number of neopeptides was found when this methodology was applied. Moreover, when the serum load was increased the number of neopeptides identified was also raised except when the 30kDa UF was conducted, likely due to the tendency for clogging boosted due to the serum matrix increment (157).

Furthermore, there are not many MRM peptidomic studies with SIS peptides on serum or SF samples since most of them are based on shotgun approaches. In this work, shotgun and MRM MS analysis were performed for the identification of cartilage and OA-related naturally produced peptides in serum and synovial fluid samples. After the MRM measurement of 44 neopeptides (from the 2 designed MRM methods), 6 peptides corresponding to KNG1, CO3, APOA4 and ITIH4 were differentially detected on serum pools from OA patients and HD. However, none of the neopeptides appertaining to the first cartilage secretome MRM method (23 peptides) were successfully detected in SF or serum. Interestingly, other peptidomic studies have also detected the same and/or similar peptide fragments from ITIH4, CO3, KNG and APOA4 in serum or plasma (159, 160). Moreover, all these peptides were examined in a few SF samples due to the low availability of these specimens. However, most of the identified proteins and/or neopeptide fragments have been also identified in SF samples by other authors (161) as the KNG1 and CO3 neopeptides (162).

Interestingly, the enzymatic digestion of the 6 peptides from KNG1, CO3, APOA4 and ITIH4 and the 8 peptides from PRELP, MGP, CILP1, CLUS and COMP was presumably due to three different enzyme types: cysteine proteases (cathepsin K), serine proteases (as cathepsin G and elastase-2) and metalloproteases (MMP-3 and MMP9). These proteases are related with the cartilage turnover, joint inflammation and/or OA progression to all appearances (12, 13) and the majority have been commonly reported to act over the detected cartilage and serum neopeptides of the study. Specifically, the increase of cathepsins has extensively been associated with OA severity and joint inflammation (163, 164). On the one hand, cathepsin G seems to have special roles in the development of inflammation by promoting the migration of neutrophils, monocytes and antigen presenting cells (APCs) and activating different protease-activated receptors (PARs) (165, 166). Interestingly, cathepsin G activates other known OA-related cytokines as interleukins and TNF-α (167-169) and degrades collagen and proteoglycans (170, 171). On the other hand, cathepsin K may digest cartilage and bone matrix components therefore contributing to the development of OA lesions as widely studied (172-174). Many strategies have been broadly examined for inhibiting the cathepsin K action over the cartilage and bone, transforming it into an attractive therapeutic target (175-177). Moreover, matrix metalloproteases (MMPs) are genuinely involved in the cartilage extracellular matrix degradation and therefore cartilage and bone damage are intensely described (178, 179). Particularly, increased levels of the MMP-9 gelatinase and the MMP-3 stromelysin are likely associated with the OA pathogenesis as widely described (180-184). Besides, elastase-2, also known as neutrophil elastase, may also induce joint inflammation and cartilage destruction (185, 186) and has been converted into a potential OA treatment target (187).

In summary, a peptidomic analysis for the discovery and validation of novel neopeptides associated with the degradation of human articular cartilage

in osteoarthritis has been performed. This has enabled not only to obtain an exhaustive neopeptidomic profile of this tissue, but also the validation of a panel of eight endogenous peptides that are characteristically released during the pathogenic process. The peptide DSNKIETIPN, from PRELP, showed the best metrics as a biomarker of OA cartilage, proving to be the most promising candidate for the development of assays aimed at its detection and quantification in biological fluids. Furthermore, another discovery and targeted analysis of neopeptides in synovial fluid and serum samples was performed. Six additional peptides were detected in serum samples from OA patients and healthy donors, although a more exhaustive analysis must be explored for confirming their OA monitoring role. Altogether, these neopeptides seem to be cleaved by different described proteases, presumably related with the OA progression, cartilage turnover and joint inflammation as MMP-3, MMP-9, cathepsin K, cathepsin K and elastase-2.

3. Chapter III: Development of two immunoaffinity-mass spectrometry (IA-MS) techniques for the detection and absolute quantification of two cartilage-related proteins in serum samples from OA patients.

At present, the major limitation in diagnostic and monitoring tools for Osteoarthritis (OA) has caused a considerable interest in finding specific biomarkers that could reflect both qualitative and quantitative variations of the remodeling of joint tissue. Cartilage Intermediate Layer Protein 1 (CILP1) and Proteoglycan 4 (PRG4), also known as Lubricin or Superficial Zone Protein (SZP), have been described as putative biomarkers of cartilage degradation in OA. These proteins are also extremely interesting in cartilage bioengineering approaches, since they are markers of specific layers of this highly structured tissue (CILP1 of the intermediate layer and PRG4 of the superficial layer) as seen in **Figure 2**. More in depth, CILP1 probably plays a role in cartilage scaffolding and may inhibit TGFB1-mediated induction of cartilage matrix genes via its interaction with TGFB1. Besides, it has the ability to suppress IGF1-induced proliferation and sulfated proteoglycan synthesis, and inhibits ligand-induced IGF1R autophosphorylation. CILP1 overexpression may lead to impair chondrocyte growth and matrix repair and indirectly promote inorganic pyrophosphate (PPi) supersaturation in aging and osteoarthritis cartilage. PRG4, on his side, plays a role in boundary lubrication within articulating joints and prevents protein deposition onto cartilage from synovial fluid by controlling adhesion-dependent synovial growth and inhibiting the adhesion of synovial cells to the cartilage surface.

The aim of this chapter is to develop, validate and compare two different sensitive and multiplexed immunoaffinity mass spectrometry (IAMS) methods for the evaluation of these two cartilage-related protein biomarkers useful for OA diagnosis and/or monitoring. The outcome of this study is enclosed in a multicenter project called OA-BioMark from the CIBER-BBN Intramural Program. Four groups are involved in the procurement of the project: Nanobiotechnology

for Diagnostics (Nb4D) group, Nanoparticle and Peptide Chemical Group (NPCG), Biomaterials Centre, Universidad Politécnica de Valencia (CBIT-UPV) and Tissue Bioengineering and Cell Therapy Group from the Universitary Hospital of A Coruña (GBTTC-CHUAC) as the coordinator research group. Moreover, the specific objectives were:

1. Selection of two proteotypic peptides from each target protein: CILP1 and PRG4.

2. Generation of polyclonal antibodies specific against these proteotypic peptides.

3. Development of two multiplexed IAMS methods:

- iMALDI.
- SISCAPA-nano-LC-MRM.

4. Analytical validation of the previous methods for the detection and quantification of the two cartilage-related markers in synovial fluids and sera from OA patients and controls (healthy donors).

3.1. Materials and Methods.

3.1.1. Selection of CILP1 and PRG4 proteotypic peptides.

3.1.1.1. Sample preparation.

In order to conduct the selection of CILP1 and PRG4 proteotypic peptides, human cartilage as well as non-depleted and depleted serum samples from OA patients and/or HD were analyzed by MRM.

On the one hand, human knee articular cartilage samples were obtained from adult donors undergoing joint surgery. Full thickness cartilage or cartilage slices were snap frozen in liquid nitrogen and posteriorly stored at −80°C. For protein extraction, cartilages were pulverized in a freezer mill. The resulting powder was transferred to Eppendorf tubes and dissolved in 6M Urea and 2% SDS. Then, samples were vortexed and sonicated three times for approximately 1s, and later agitated overnight at 4°C. The extracted material was then centrifuged at 4°C for 20min at 14,000rpm. Supernatants, containing cartilage proteins, were transferred to new tubes and precipitated overnight with ice cold acetone at −20°C. Protein content was collected by centrifugation, forming a protein pellet which was washed once with ice cold acetone and subsequently air-dried. Finally, the pellets were dissolved in 25mM ammonium bicarbonate (AmBic).

On the other hand, crude serum samples were immunodepleted from the 20 most abundant proteins (albumin, IgG, transferrin, fibrinogen, IgA, α2-Marcroglobulin, IgM, α1- Antitrypsin, complement C3, haptoglobulin, apolipoprotein A1, A3 and B; α1- Acid Glycoprotein, ceruloplasmin, complement C4, C1q; IgD, prealbumin, and plasminogen) in order to boost the detection of less-abundant proteins. For this aim, 150µL serum were filtered through 0.2µm cellulose acetate membrane tube filters (CLS8160, Costar Spin-X centrifuge,

Corning Life Sciences, Amsterdam, The Netherlands) at 14,000g for 45min. The filtered serum (90µL) was then injected into a high-performance liquid chromatography system (1200 Series, Agilent Technologies, Palo Alto, CA), equipped with a Proteo-Prep20 Plasma Immunodepletion LC Column (Sigma-Aldrich, St. Louis, MO). Buffer A (ProteoPrep equilibration buffer, (PBS), Sigma) was used in the equilibration and washing steps. ProteoPrep elution buffer (0.1M Glycine-HCl, pH 2.5 with Octyl β-D-glucopyranoside, Sigma) was used for elution of the bound and highly abundant proteins from the column. The chromatograms of the pools were monitored using a UV detector at 280nm at room temperature according to the protocol supplied by the manufacturer. Finally, the flow-through fraction containing low-abundance proteins was collected and stored at -20^{o}C.

The cartilage protein concentrations were determined by the Bradford assay, while depleted and non-depleted serum samples were quantified using the nanoDrop instrument at 580nm. For in-solution digestion, 10µg of proteins from cartilage and serum samples were dissolved in denaturing and reducing buffer (6M Urea/2M Thiourea, 25mM AmBic, 10mM Dithiothreitol (DTT)) for 1h at 37°C, and cysteins were alkylated with 50mM iodoacetamide (IA) for 45 min in the dark. Samples were diluted with 25mM AmBic to a final concentration of 1M Urea for conducting a posterior protein trypsin-digestion. Thus, Promega Grade Trypsin (Promega) was added at a 1:25 ratio (enzyme:protein) for 16h at 37°C. Samples were acidified with TFA and subsequently desalted using in-house made stage tips (3M Empore SPE-C18 disk, 47mm, Sigma Aldrich). Samples were finally dried using a speed-vacuum (Thermo, USA) and stored at -20^{o}C for later nanoLC-MS/MS analysis.

3.1.1.2. Selection of proteotypic peptides by multiple reaction monitoring (MRM).

The Skyline software v1.3 (134) was used for creating an empirical data set of the target CILP1 and PRG4 peptides. Different proteotypic peptides with the highest spectral counts, only fully tryptic peptides, with no missed cleavages, with a length between 8 and 30 amino acids and devoid of methionine and cysteine residues were chosen. In addition, sequences that may cause incomplete digestion, such as continuous sequences of arginine (R) or lysine (K) and a proline (P) at the C-terminal side of R or K, were also excluded since partial tryptic hydrolysis at the peptide bond is often observed in MS/MS. Furthermore, if possible, peptides were selected from different areas of the protein sequence, in order to attain the maximum protein coverage. The top transitions were selected for method development on the basis of the presence of abundant y ions at m/z greater than that of the precursor. In the absence of high-m/z y ions, the most abundant fragment b ions were selected. Additionally, the Global Proteome Machine database was used to select peptides from target proteins that were frequently detected (multiple experiments). The resulting CILP1 and PRG4 peptides, coming from the approaches described above, were pooled in a final optimized MRM method.

MRM analysis was performed injecting 1µg of digested samples onto the nanoLC-5500 QTRAP using a linear gradient of 70 min. Thus, protein digests were analyzed by LC-MS/MS using a nanoLC system (TEMPO) coupled to a 5500-QTRAP instrument (AB Sciex). After precolumn desalting using a C18 column (5µm, 300A, 100µm $*$ 2cm, Acclaim PepMap, Thermo Scientific, USA) at a flow of 3 µL/min during 10min, tryptic digests were separated on C18 nanocolumns (75µm id, 15cm, Acclaim PepMap 100, Thermo Scientific, USA) at a flow rate of 300 nL/min. The mass spectrometer was interfaced with a nanospray source equipped with an uncoated fused silica emitter tip (20µm inner diameter, 10µm tip, NewObjective,

Woburn, MA) and was operated in the positive ion mode. MS Source parameters were as follows: ion spray voltage (IS) 2600V, interface heater temperature (IHT) 150°C, ion source gas 2 (GS2) was 0, curtain gas (CUR) was 20 and ion source gas 1 (GS1) was 25psi, and collision gas (CAD) high. MS compound parameters were set to 10 for the entrance potential (EP) and to 15 for the Collision cell exit potential (CXP). Skyline was used to predict and optimize collision energies for each peptide. Q1 and Q3 were set to unit/unit resolution (0.7Da) and the pause between mass ranges was set to 3ms. In order to confirm the identity of the peptides, a MRM Information Dependent Acquisition (IDA) experiment was performed for each peptide. The mass spectrometer was instructed to switch from MRM to enhanced product ion (EPI) scanning mode when an individual MRM signal exceeded 1000 counts. Each precursor was fragmented a maximum of twice before being excluded for 10s and the mass were scanned from 250 to 1000 Da.

Data analysis of the proteotypic peptides for the target proteins was performed using Skyline. Raw files were imported to Skyline and integration was manually inspected to ensure correct peak detection, absence of interferences, and accurate integration. MRM signal was defined as the detection of all the transitions from the endogenous peptide exactly coeluting. Moreover, the MS/MS data generated were analyzed with the ProteinPilot software. The Paragon algorithm in ProteinPilot software served as the default search program for protein identification, with trypsin as the digestion agent and IAA as a fixed modification of cysteine. Biological modifications were programmed in the algorithm. The searches for peptide mass fingerprints and tandem MS spectra were performed using the SwissProt database.

Finally, two proteotypic peptides from each protein, showing the best peak signals by MRM and most confident identifications in cartilage samples, were chosen for the following generation of antibodies. The reason for measuring

2 peptides per protein at first, is because one of them might not later fulfill all the antibody development and/or IAMS analysis conditions. Thus, the protein quantitation could be ensured at least with one of the proteotypic peptides. Importantly, the selected peptides should not be detected in non-depleted and/or depleted sera for reasonably conducting IAMS analysis.

3.1.2. Generation of polyclonal antibodies specific against the selected proteotypic peptides.

3.1.2.1. Peptide synthesis, sequence confirmation and peptide conjugation (NPCG).

Two different strategies were followed for peptide synthesis. On the one hand, PRG4 selected proteotypic peptides were synthesized on a Liberty Lite microwave peptide synthesizer (CEM corporation) using a Rink amide resin with the first amino previously manually incorporated. Peptides were elongated using a Fmoc/tBu synthesis strategy. The Fmoc elimination was performed with 10% of piperazine in NMP-EtOH mixture and diisopropylcarbodiimide (DIC) and Oxyma Pure were used as coupling reagents and DMF as solvent. Once the peptide elongation was finished the N-terminal amino was acetylated manually following standard protocols. This step helps to neutralize the peptide charge, increasing their ability to enter cells and reducing the effect of charged N-termini during the posterior ELISA binding (188). Moreover, the acetylation may also increase the metabolic stability of peptides as well as their ability to resist enzymatic degradation by aminopetidases, exopeptidases, and synthetases. On the other hand, CILP1 proteotypic peptides were synthesized manually, using the 2-Chloro-Trityl Chloride resin following an Fmoc/tBu synthesis strategy as well. The elongation was carried out using DIC and hydroxybenzotriazole (HOBt) as coupling reagents.

Peptides were cleaved from the resin by acydolisis treatment with TFA-H_2O-TIS (95:2.5:2.5), characterized by HPLC and HPLC-MS and finally purified by semiprep HPLC-MS to reach a minimum of 95% purity (determined by HPLC-PDA at 220 nm).

3.1.2.2. Production of the specific antibodies for the selected biomarkers (Nb4D).

Peptides or small molecular weight biomarkers do not elicit an immunogenic response. Therefore, they need to be conjugated to biomacromolecules to render them immunogenic. In order to generate the specific polyclonal antibodies, the preparation of the immunogens and competitors was firstly carried out in order to immunize the rabbits and finally obtain the corresponding antibodies. Thus, succinimidyl iodoacetate (SIA) bioconjugates were prepared by mixing, under stirring overnight at 4°C, 2.8µmol of each corresponding peptide and 3mg of a purified SIA-activated protein solution. This solution consisted on 3.9mg SIA dissolved in 500 µL anhydrous dimethylformamide (DMF) added dropwise to 4.5mL of a protein solution (horseshoe crab hemocyanin (HCH) or bovine serum albumin (BSA), 15 mg each) in borate buffer and mixed under stirring for 4h at RT. The resulting activated protein solution was ultimately purified by size-exclusion chromatography using a HiTrap desalting Sephadex G-25 Superfine column (Amersham Biosciences) and borate buffer as eluent. Finally, the resulting peptide conjugates were purified by dialysis, lyophilized and stored freeze-dried at -20°C.

The rabbit immunizations were carried out in the animal facility of the Research and Development Center (CID) of the Spanish Research Council (CSIC), minimizing suffering of the animals. The antisera (As) was obtained by immunizing female white New Zealand rabbits with the generated peptide conjugates. Along the immunization process, the evolution of the antibody titer was assessed by non-competitive indirect ELISA. Thus, the binding of serial dilutions of the

different As to microtiter plates coated with a fixed concentration of the peptide conjugate at 1 mg/mL was assessed and measured. After 6 immunizations, the animals were exsanguinated, and the blood was collected in vacutainer tubes provided with a serum separation gel. As were obtained by centrifugation at 4°C for 10min at 10,000 rpm, and stored at 4°C for subsequent use or if not at −80°C in the presence of 0.02% NaN_3.

At the end, the polyclonal antibodies contained in the As were precipitated by ammonium sulfate and lastly resuspended in 10mM PBS. Subsequently, the precipitated antibodies were then injected into the chromatographic system Äkta Prime plus in order to separate the corresponding anti-peptide antibodies. This process was carried out using an immuno-affinity column, with the previously obtained peptide conjugates immobilized on it. Thus, the specifically generated anti-peptide antibodies were bound to the immobilized analogous peptide conjugates and later eluted with 100mM Glycine-HCl (pH 2.7). Furthermore, the unbound fraction obtained was assessed, in order to monitor the immunoaffinity chromatography performance, by non-competitive indirect ELISA, equally to when the As antibody titer was monitored. After purifying the specific anti-peptide antibodies, pH was adjusted with 1.5M Tris to 7-8 for optimal antibody performance and filtered by 0.2µm Millex polyvinylidene difluoride (PVDF) devices. Antibody concentration was finally measured at 280nm using a plate reader.

3.1.3. Development and analytical validation of iMALDI for the detection of CILP1 and PRG4 in sera from OA patients and controls (healthy donors).

3.1.3.1. Sample preparation.

Serum samples from patients suffering OA and healthy donors (HD) were trypsin digested using the following in-plate-batched-protocol. Firstly, a

denaturation and reducing mix was prepared with 2.7g Urea, 140mg Trizma pre-set crystals (pH 8.1), 500µL BondBreaker TCEP (0.5M bottle conc) and made up to 5mL with water (Water, Optima™ LC/MS Grade, Thermo Fisher Scientific). Further 879µL of water were added to the previous solution and 20µL of this final solution was aliquoted per well in a 1.1mL 96-well plate and then dried overnight at 37°C in a dry incubator. Later, 12µl of each sample was added per dried mix well and incubated for 30' at RT with shaking (1000rpm) in order to conduct the protein denaturation and disulfuric Cys bond reduction of the samples. Subsequently, 10µL of 0.1M iodoacetamide were added to each well and incubated for other 30' at RT in the dark with shaking (1000rpm) with the aim of blocking the free -SH Cys groups. Afterward, 160µL of 0.2M Trizma pH 8.1, 0.015% CHAPS was added per well in order to lower down the Urea concentration, before the tryptic digestion was performed overnight, at 37°C, using 15µL of 2.33µg/µL Worthington-TPCK treated trypsin in 10mM HCl at 1:20 ratio (µg trypsin:µg total protein content). Finally, 5µL of 1.14µg/µL TLCK in 1mM HCl were added in each well and shacked at 1000 rpm for 5' at RT in order to arrest trypsin digestion and maintain basic the pH of the sample. The final total volume of each well-digested sample is approximately of 200µL, as the lyophilized denaturation and reducing mix would also add extra volume when resuspended.

Stable-isotope-labeled standard (SIS) peptide analogues of the four selected proteotypic peptides were purchased to JPT. The SIS peptides contained different stable isotope labeled amino acids, depending on the tryptic end of each SIS. The modifications carry ^{13}C and ^{15}N isotopes, leading to different mass shifts on dependence of the labeled amino acid. For Lysine, K, ($^{13}C_6$,$^{15}N_2$-Lys), a +8 Da mass shift was added and for Arginine, R, ($^{13}C_6$,$^{15}N_4$-Arg), a +10 Da mass shift. In this case, SpikeTides TQL peptides were ordered to achieve a maximum absolute peptide quantitation. Peptides are attached to a JPT's QTag, a small chemical tag which can be released by trypsin and does not interfere with posterior measurements. QTag provides robust and reproducible quantitation via HPLC-UV

or HPLC-MS and UV, converting it in a very accurate method (CV= 5%). 1 nmol vial of each SIS peptide was resuspended in 100µL 20% ACN 80% 0.1M Ammonium Bicarbonate (AmBic) obtaining a SIS stock solution of 10 pmol/µL. 20µL of each SIS stock solution peptide (200pmol) were individually digested, during 5 hours at 37°C, with 60µL Trypsin Promega in 25mM AmBic at 0.16µg/µL , in order to maintain the basic pH and lower down the ACN % in solution to 5% before digestion. Lastly, the solution was acidified with 120µL 0.1% FA until a pH of 3.5 in order to stop the trypsin digestion, obtaining a final SIS peptide concentration of 1pmol/µL.

3.1.3.2. Antibody-bead conjugation.

In order to couple the antibodies to the beads, different tests were performed based on these two reagents. Three different commercial beads were tested in order to select the ones that give better outcomes for the IAMS procedure and/or analysis. They include Dynabeads Protein A and Dynabeads Protein G 30 µg/µL (Invitrogen, Thermo Fisher Scientific), as well as MagReSyn Protein A microspheres (ReSyn Biosciences). Moreover, in order to covalently couple the antibody to the beads, the disuccinimidyl suberate (DSS) crosslinker was also tested. This crosslinking step is normally utilized when only the bounded molecules want to be eluted from the antibody, conserving intact the antibody-bead conjugation, which could be then reused after extreme acidic peptide elution. Other parameters as antibody-bead saturation and/or antibody/bead interferences (negative controls) were also evaluated, as explained in the actual and following sections.

The protocol for conjugating the different beads to the antibodies previously generated was hardly the same all the time, with little changes when one or the other bead type was used. In general, the antibody-bead coupling was performed in big quantities for lowering the variability of the assay. Previously,

MagReSyn Protein A beads were disaggregated using an ultrasound probe in order to facilitate their handling as they tend to aggregate and stick together quite hard. Firstly, beads were washed and conditioned to the optimal incubation solution. The objective of prior washing the magnetic beads was exchanging the Tween20 detergent of the bead slurry solution with CHAPS. CHAPS is a less noisy detergent in terms of mass spectrometry analysis as it only shows one peak on the spectrum, while Tween20 shows several polymer peaks which could hamper the detection of low-abundant peptides. For this aim, 20µL of bead slurry were placed on a magnet for 30'' while the supernatant was removed. Subsequently, ten total washes were done allowing 10'' for the wash and 30'' for pelleting the beads on the magnet and a posterior supernatant withdraw. Thus, seven washes were done with 600µL of 25:75 ACN : PBSC (PBS buffer with 0.015 % CHAPS) and three with 600µL of PBSC. Ultimately, beads were resuspended in the initial 20µL volume with PBSC. Secondly, each antibody was added to the washed bead suspension at the rate of 0.2 µg antibody/µL bead slurry, as recommended by the manufacturer. The bead-antibody mixture was incubated for at least 1 h at room temperature while shaking at 1000rpm. The unbound antibody was washed three times with 600µL PBSC on a magnet and the antibody-bead conjugation was resuspended in 200 µL PBSC and kept at 4°C until its use. The antibody is at a final concentration of 0.02 µg/µL, whereas the beads are at 3 µg/µL in the 200µL PBSC solution. Besides, antibody-bead saturation was evaluated by coupling increasing antibody amounts on the same bead quantity and detecting a constant SIS load.

Furthermore, the DSS crosslinking protocol was once tested with the aim of eliminating a posteriorly detected peptide interference, likely coming from the immunoaffinity purification of the generated antibodies. For this purpose, beads were washed and antibodies were coupled to the washed beads as stated in the previous paragraph. Antibody conjugated beads were washed twice on a magnet with 80 µL PBS. After the last wash, all liquid was withdrawn and 100 µL of 10mM DSS in dimethyl sulfoxide (DMSO) was added and incubated at RT for 30' while

rotating. The crosslinking reaction was then quenched with 10 µL of Tris HCl 1M (pH 7.5-8) and incubated at RT for 15' with rotation. Thereafter, the crosslinked antibody-beads were washed three times on a magnet with 600 µL PBSC. After the last wash, all supernatant was discarded and the crosslinked antibody-beads were again resuspended in 200 µL PBSC, in order to compare their performance with the original non-crosslinked antibody-bead conjugates. 50µL of the crosslinked antibody-beads in PBSC (5µL of the initial bead slurry) were washed on a magnet 3 times with 150µL 50mM Glycine-HCl (pH 3) in order to elute the interfering peptide and then other 3 times with 150µL PBSC. Washed crosslinked antibody-beads were lastly resuspended in the initial 50µL of PBSC in order to compare the performance of the washed and non-washed crosslinked antibody-beads or non-crosslinked antibody beads.

Additionally, the peptide interference was also intended to be eliminated by performing a preliminary acidic ultrafiltration (UF) of the generated antibodies previous to the antibody-bead coupling. In this case, the interfering peptides (<1kDa) are eluted and separated from the heavy antibodies (~150kDa) using 0.5mL 10kDa Amicon Ultra Centrifugal Filters (Merck). For this purpose, 500µL of each immunoaffinity purified antibody at a known concentration were acidified with 1µL FA in order to obtain a final 0.2% FA solution (pH: 2.5). Subsequently, the acidified antibodies were ultra-filtrated with the 10kDa cited centrifugal filters during 20 minutes at 4°C and 4000g. In this step the interference peptides were eluted from the antibodies due to the acidic reaction. While the antibodies remained on the filter device, the peptides passed through them, separating both reagents. Afterwards, 200µL 0.1% FA were added to the upper antibody solution and mixed by gently pipetting. Once more, the preceding ultrafiltration steps were repeated 3 more times and the final antibody concentrate (around 120µL) was diluted with PBS to the initial antibody solution concentration. Finally, the pH solution was confirmed to be around 6-6.5 for optimal antibody behavior and environment.

3.1.3.3. Sample incubation and MALDI plate spotting.

In order to proceed with the sample incubation, tests were performed in duplicates and/or triplicates depending on the newness of the experiment. Besides the sample type or test conducted, 10µL of washed and crosslinked / only crosslinked / no crosslinked antibody-bead conjugations were generally used per individual experiment (1µL of the original bead slurry). Additionally, 10µL of a known and MALDI-TOF detectable SIS concentration and different amounts and/or volumes of sample digests diluted in PBSC were added for each experiment.

Firstly, different parameters attending to the sample incubation were evaluated as, for example, the incubation platform (plates and/or tubes), time (hours or overnight), temperature (at RT or 4°C) and/or technique (shacking or rotating). Thus, serum digests were tested for detecting and studying the natural/endogenous peptide signals. The final sample incubation protocol for serum samples, after examining some of the previously cited parameters, remained as follows. First, 10µL of the final antibody-bead conjugation in PBSC (1µL of the original bead slurry), 10µL of a known SIS peptide amount (fmol) and 200µL of sample digest were mixed in 1.1mL 96-well plates. This mixture was incubated during 3 hours at room temperature (RT) by shaking at 1000rpm. Beads were washed three times with the aid of a multichannel pipette and a magnet, allowing a mixing time of 30'' and a pelleting time of 15''. The first wash was performed with 80µL 15% ACN 85% PBSC and the resuspended beads were switched to clean wells in order to remove matrix interferences as lipids or other unspecific molecular traces. Then, the second wash was done with 100µL 15% ACN 85 % 5mM AmBic and the third wash with 100µL 5mM AmBic. Lastly, the remaining beads were resuspended with 5µL 5mM AmBic and spotted onto a disposable 96-spot µFocus Microflex MALDI plate (Hudson Surface Technology) and let until dryness (**Figure 3. 1**).

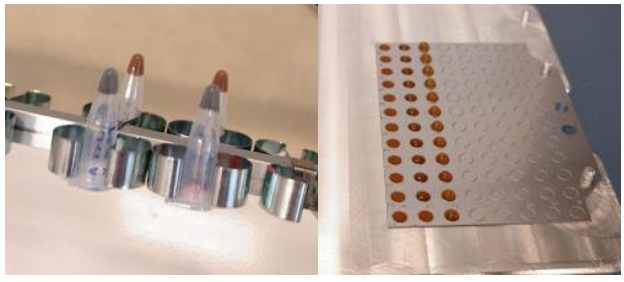


Figure 3. 1. Beads rotating during the incubation step and bead spotting onto the disposable MALDI plate.

Finally, in order to perform the matrix spot and wash steps, 3mg/mL HCCA (Alpha-Cyano-4-hydroxycinnamic acid) matrix in 70% ACN 0.1% TFA in 7mM Ammonium Citrate and 7 mM Ammonium Citrate (AmCit) solutions were freshly prepared. Then, 1.5µL 3mg/mL HCCA matrix solution were spotted onto each target spot and let dry. Thus, when applying the HCCA matrix, the peptides that were bounded to the antibodies were then eluted (due to the acidic pH of the matrix) and anchored in the same target spot. Each spot was later washed for three times with 5µL 7 mM AmCit allowing 5'' incubation time and let again dry for posterior MALDI-TOF analysis. The aim of this spot wash is to remove salts and traces that could interfere with the subsequent MS analysis.

Besides, some method controls were tested within each particular protocol for monitoring the different parameters examined. On the one hand, an already validated iMALDI assay for the FLNVLSPR peptide detection (corresponding to Insulin-like growth factor-binding protein 3) was used as the positive control. Specifically, 0.5 µg of antipeptide-antibody were added per µL of the initial bead slurry for the antibody-bead conjugation, while 500 fmol of commercial light and heavy peptides were incubated, spotted and analyzed by MALDI-TOF for controlling the different assay conditions. On the other hand, two different negative controls were also evaluated. A first negative control that monitors the unspecific bound of the measured peptides with the beads. In this case, the beads undergo the exact antibody-bead conjugation procedure but

without adding the antibody, as well as the rest of the distinct steps of the sample incubation, elution, spotting and analysis procedures. In order to check the absence of unspecific bound of the measured peptides with other antibodies that are not the corresponding antipeptide-antibodies, a second negative control was also implemented when any of the parameters were changed. For this aim, 0.5 µg of antipeptide-antibody were added per µL of the initial bead slurry for the antibody-bead conjugation. Then, the same sample digest and SIS peptide amounts tested were incubated, spotted and analyzed by MALDI-TOF.

3.1.3.4. MALDI-TOF analysis and data interpretation.

After conducting the immunoaffinity capture of the corresponding peptides (NAT and SIS), these are anchored to the spots on a MALDI target, ready for the MS analysis. MALDI-TOF analysis was conducted on a Bruker Microflex LFR mass spectrometer. The MS analysis was performed under the following parameters: medium mass range between 800-4000 Da, up to 800 Da suppression in deflection mode and 60Hz laser frequency among others. One mass spectrum per spot was acquired by summing a total of 1000 shots in two different positive ion modes: linear positive (LP) or reflector positive (RP). The spectrum was acquired at a fixed laser intensity and a random walk of 15 shots at each raster spot on the complete sample. LP and RP AutoXecute methods were developed in the Bruker FlexControl 3.3 software in order to automatically acquire LP and/or RP data from each sample target. After the MS data was acquired, FlexAnalysis 3.4 software was used for automatic smoothing, baseline subtraction and mass list find on LP and RP data. Mass lists were generated in the same software using the different Centroid or Snap algorithms. Peak intensities, areas and/or S/N of the light and heavy peptides were recorded and analyzed in order to find the optimal signal response for the data analysis of the assay. L/H ratios were then calculated based on the selected signal method and NAT absolute concentrations

in the samples could be also determined by multiplying the L/H ratios by the amount of spiked SIS peptides.

3.1.3.5. iMALDI assay validation.

After the whole antibody-bead conjugation, sample incubation, spotting and analysis protocols were evaluated and finally determined, the iMALDI validation of the succeeding peptide(s) was accomplished. This validation was done following the Assay Characterization Guidance Document (version 1.0) from the CPTAC Assay Development Working Group (189).

First, in accordance to these guidelines, Experiment 1. Response Curve was conducted with the aim of determining the lower limit of quantification (LLOQ), the limit of detection (LOD) and the linear range. An 8-point calibration curve was designed based on the particular analyte's sample concentration, which should fall in the center of the curve. Each point of the curve was analyzed in four replicates, each replicate being an independent experiment. For the totality of the experiment, the same batch of antibody-bead conjugation and plasma sample digest was used for the sample incubation, in order to minimize the variability coming from the sample preparation. Both reagents were consistently added to every experiment as in the cited protocols, 10 µL (1 µL bead slurry) and approximately 200 µL pooled serum digest. Contrarily, serial dilutions of the SIS peptides (freshly digested, aliquoted and frozen to -80°C) were added to the sample digest in the same volume of 10 µL as stated in the previous sample incubation section. All curves and their corresponding points were prepared and analyzed at the same time and under the same conditions for maximum homogeneity. After recovering the data of the 4 calibration curves (L/H or NAT/SIS ratios), the curve that showed the most variable results was discarded, taking into account the 8 %CV of the corresponding points of the curve. The NAT and SIS peak intensities of the three remaining 8-point curves were used for building the

response curve with the aid of the already cited QUALIS-SIS software (103). The software determined most of the curve's parameters as linear regression, LLOQ and dynamic range with a precision and accuracy lower than 20%. Additionally, the LLOD was calculated for each peptide using the following formula: LLOD= 3/10 LLOQ.

Second, after confirming the linear range and LLOQ of the assay, Experiment 2. Mini-validation of repeatability was performed with the aid of assessing the intra-assay and inter-assay variability, as well as the accuracy and precision of the method. Therefore, 6 replicates of three different concentrations of SIS spiked peptides in a constant pooled serum digest were measured in 5 different days. The three SIS concentrations were included in the previously tested linear range of the assay and were determined based on the already calculated LLOQ. Thus, a low concentration was defined as 2.5 times the LLOQ of the assay, a medium concentration as 10 times the LLOQ, while 50 times the LLOQ corresponded to a high concentration. All reagents and bead-antibody conjugations were freshly made each of the 5 days in order to determine the intra-day and inter-day variabilities of the three low-medium-high concentrations. Hence, the assay precision for each low/medium/high concentration was determined by calculating the average %CVs for 5 of the 6 concentration replicates measured for one of the days (n=5 for each concentration level). Likewise, assay accuracy for each low/medium/high concentration was calculated using the selected 5 replicates of the 5 days for each concentration level. The assay accuracy was interpreted as the % nominal, which is calculated with the following formula: average of the experimental concentration value/average of the theoretical concentration value*100 (n=5 replicates*5 days=25 for each concentration level). The average intra-day variability was calculated as the average of the 5-individual day %CVs average of the 5 replicates; while the average inter-day variability was the average of the 5-individual replicate %CVs average for the 5 days. Finally, the

total assay variability was determined by the square root of the squared average intra-day variability plus the squared average inter-day variability.

Ultimately, after confirming the assay response and variability accomplished the cited requirements, 38 sera from patients with OA (22) and HD (16) were analyzed in triplicates with the validated assay. In addition to these samples, the same previous 8-point calibration curve was also simultaneously processed and analyzed in triplicates as previously stated. The absolute concentration of each serum sample was calculated and compared using two different methodologies: Single Point Measurement (SPM) and Standard Curve Regression Measurement (SCRM). SPMs were determined by using the average L/H ratios for each sample while SCRMs were calculated using the QUALIS-SIS software. It is important to note that, when using the QUALIS-SIS software for this purpose, there are some samples, the ones that fulfill the confidence requirements, which are determined by SCRM, while the others that do not accomplish these requisites are calculated by SPM. Nevertheless, the software itself automatically detects which measurements should be done using one or the other methodology. Statistical analysis was performed using the GraphPad Prism software. More specifically, Mann Whitney tests were performed in order to see differences between OA patients and HD using the SPM and/or SCRM measurements.

3.1.3.6. iMALDI technical validation by ELISA.

In order to assess the technical validity of the developed iMALDI for measuring PRG4 in serum samples, the same 38 sera (22 OA and 16 HD) were analyzed by a commercial PRG4 ELISA from Aviva Systems Biology (San Diego, USA). The antibody epitopes of this ELISA contain the PRG4 surrogate peptide, precedingly used for iMALDI assay development. Additionally, 34 serum samples (17 OA and 17 HD) of the same cohort were measured conjointly in order to

confirm the detected PRG4 abundance in these specimens. Serum dilution was first checked for performing the final ELISA analysis.

3.1.4. Development and analytical validation of SISCAPA-nanoLC-MRM for the detection of CILP1 in sera from OA patients and controls (healthy donors).

3.1.4.1. Sample preparation.

Serum samples from patients suffering OA and healthy donors (HD) were trypsin digested using the following protocol. Firstly, 0.8mg urea were added per µL serum in order to obtain a final urea concentration of approximately 8M for denaturing the serum protein samples. Then, pH was checked and corrected to 8 with 40mM AmBic for optimal protein digestion. Subsequently, 0.015µL 700mM DTT were added per µL of denatured sample and the mixture was incubated for 1h at RT in order to conduct the disulfuric Cys bond protein reduction. Afterwards, 0.045µL of 700mM iodoacetamide were added per µL of the digestion solution and immediately incubated for 30' at RT in the dark with the aim of blocking the free -SH Cys groups. In order to blockade the excess of IAA, 0.075µL of 500mM N-Acetyl-L-cysteine (NAC) were added per µL of the total solution and incubated during 15' at RT. Next, a specific volume of 50mM AmBic was added to dilute the urea to 1M, before the tryptic digestion was performed overnight, at 37°C with 1µg/µL Worthington-TPCK treated trypsin in at 1:20 ratio (µg trypsin : µg total protein content). Finally, a volume of 10mM TLCK in 1mM HCl, corresponding to a 2-fold-molar excess from trypsin, was added and the mixture was incubated for 30' at RT in order to arrest trypsin digestion and maintain basic the pH of the sample.

Different approaches were followed when conducting the serum digestion, depending on the SISCAPA step. On the one hand, a first large pool of sera was used for setting the parameters of all the SISCAPA method procedure.

Considering the digests needed to be later diluted 7.5 times in PBSC, as explained in the next section, the pool digest volume was too high for SISCAPA performance in 1.5mL tubes. Thus, the pool digest was passed through SepPack OASIS HLB C18 cartridges (Waters) in order to concentrate the digest in a lower volume. On the other hand, when the SISCAPA method was completely refined, serum samples were individually digested for carrying out the validation of the SISCAPA procedure. Lower volumes of the digests were obtained for avoiding the explained SepPack concentration procedure of every individual sample. All final protein digests were kept at 4°C if used the same day or frozen at -20°C for later analysis.

The same Stable-Isotope-labeled Standard (SIS) peptide analogues purchased to JPT and used in the iMALDI procedure were employed for conducting the SISCAPA assays. As these peptides have a trypsin-digestible JPT's QTag attached, SIS peptides were trypsin digested individually. Thus, 1 nmol vial of each SIS peptide was resuspended in 100 µL 20% ACN 80% 0.1M Ammonium Bicarbonate (AmBic) obtaining a SIS stock solution of 10 pmol/µL and intermediate SIS peptide solutions at 2 pmol/µL were prepared. Then, 5 µL of these intermediate solutions (10 pmol) were diluted with 34 µL 25 mM AmBic in order to lower down the ACN concentration to less than 10% for optimal digestion performance. Subsequently, 10 pmol of each SIS peptide were incubated, during 6 hours at 37°C, with 1 µL Trypsin Gold (Promega) at 0.2 µg/µL. Lastly, the solution was acidified with 60 µL 0.1% FA, or until pH was around 3.5, in order to stop the trypsin digestion. Final SIS peptide solutions at 100 fmol/µL were obtained, aliquoted and stored at -20°C for further SISCAPA analysis.

3.1.4.2. Antibody-bead conjugation, serum digest and SIS peptide incubation and peptide elution.

Protein A Dynabeads (Thermofisher Scientific) were used for immunoprecipitation of CILP1-IVG and CILP1-TFL peptides. First of all, the magnetic beads were washed for exchanging the Tween20 detergent with CHAPS. Thus, 500 µL bead slurry were washed three times with 500 µL 0.03% (p/v) CHAPS in PBS (PBSC) with the aid of a magnet. After the last wash, beads were finally resuspended in 500 µL PBSC. Then, in order to perform the antibody-bead conjugation, 15 µL of washed Protein A Dynabeads (0.45mg), and 3.6 µg of antibody were incubated in a final volume of 115 µL with PBSC during 1.5 h at RT on a Multi-Rotator PTR-35 (Grant-bio). Importantly, the total antibody amount permitted per mg of bead was 8 µg, due to the bead capacity. So, when both CILP1-IVG and CILP1-TFL SISCAPA assays wanted to be combined for multiplex analysis, the antibody amount was halved and 1.8 µg of each antibody were linked to 15 µL beads.

After all the antibodies were linked to the beads, a known SIS peptide amount and serum digest (together in a constant experimental volume) were incubated overnight at 4°C while rotating. This fixed volume was calculated by adding 7.5 times more PBSC of the total summed serum digest and SIS peptide volumes for conducting a more loosely and comfortable sample incubation process. After the CILP1-IVG and CILP1-TFL NAT and/or SIS peptides were immunoprecipitated, the supernatant was discarded with the aid of a magnet. Subsequently, 4 washes were done on the magnet for removing the non-bounded molecules and/or peptides. The first three washes were done with 50µL PBSC and a last wash with 50µL ACN for removing the harder stacked interfering and undesirable molecules. Finally, the immunoaffinity-bounded NAT and SIS peptides were eluted with 40µL 0.1% FA, the solution was separated from the beads on a magnet and placed on a vial for subsequent nanoLC-MS/MS analysis by MRM.

In addition to every SISCAPA experiment tested, a solely-incubated PBSC control (in the absence of any serum digest and/or SIS peptide) was included in every analysis, in order to check interfering signals for optimal MRM performance. Accordingly, as stated in the previous iMALDI section, a peptide contamination coming from the affinity-purified antibodies was discovered. Thus, CILP1-IVG and CILP1-TFL antibodies were cleaned by acidic ultrafiltration (UF) as explained in the iMALDI protocol. However, MRM analysis of the cited PBSC controls were also performed in order to check and confirm this interfering peptide presence. Additionally, cross-reactivity tests were also performed in order to examine the affinity of the antibodies for their corresponding peptides. For this purpose, both CILP1-IVG and CILP1-TFL SIS peptides were simultaneously incubated with each of the generated CILP1 anti-peptide antibodies. The absence of an unintended peptide peak is crucial for later multiplex purposes, as the peptide analytes could be under- or over-estimated provoking further quantitation issues.

3.1.4.3. nanoLC-MS/MS analysis and data interpretation.

The immuno-enriched peptide mixtures were analyzed using a linear ion trap Velos-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Instrument control was performed using Xcalibur software package, version 2.2.0 (Thermo Fisher Scientific, Bremen, Germany). Peptide mixtures were fractionated by on-line nanoflow liquid chromatography using an EASY-nLC 1000 system (Proxeon Biosystems, Thermo Fisher Scientific) with a two-linear-column system. 10µL of the final elution were loaded onto a trapping guard column (Acclaim PepMap 100 nanoviper, 2 cm long, ID 75 µm and packed with C18, 3 µm particle size from Thermo Fisher Scientific) at 4uL/min. Then, samples were eluted from the analytical column (25 cm long, ID 75 µm and packed with Reprosil Pur C18-AQ, 3 µm particle size from Dr Maisch). Elution was achieved by using a mobile phase from 0.1% FA in water (Buffer A) and acetonitrile with 0.1% FA (Buffer B) and applying a linear gradient from 0 to 35% of buffer B for 40 min

at a flow rate of 300 nL/min. Ions were generated applying a voltage of 1.9 kV to a stainless steel nano-bore emitter (Proxeon, Thermo Fisher Scientific), connected to the end of the analytical column, on a Proxeon nano-spray flex ion source.

Target peptide sequences were loaded into Skyline to generate a precursor ion mass list in order to build the mass spectrometer acquisition method. The LTQ Orbitrap Velos mass spectrometer was operated in parallel reaction monitoring (PRM) mode. PRM was used to acquire full MS/MS spectra from target peptides from the generated mass list. Collision-induced dissociation fragmentation was performed in the linear ion trap. Accumulation of ions for MSMS was performed with an AGC target value of 5000. The maximum ion accumulation time was 50 ms. The normalized collision energy was set to 38% and one microscan was acquired per spectrum. For all precursor masses an isolation width of 2 Da was defined. Finally, raw data was imported to Skyline software in order to analyze the results. Chromatographic ion extraction of the five transitions giving the highest intensities were used to quantify each peptide. Skyline was used to calculate ratios between the unlabeled peptide and the labeled internal standard (L/H) that were used for quantification or SPM.

3.1.4.4. SISCAPA assay validation.

Once more, in order to validate the SISCAPA assay for multiplexed CILP1-IVG and CILP1-TFL quantitation, a response curve was firstly performed. For this aim, the antibody bead conjugation was done with 15µL Protein A Dynabeads, 1.8µg CILP1-IVG (UF) antibody and 1.8µg CILP1-TFL (UF) antibody. Furthermore, a constant serum digest and different SIS peptide amounts were incubated per experiment or point of the curve. In order to create a 6-point calibration curve, SIS peptide amounts ranging from 0 to 500 fmol (CILP1-IVG) and from 0 to 1000 fmol (CILP1-TFL) were analyzed in duplicates. For this experiment, SIS peptides were digested separately from the serum pool and different amounts of each

peptide were added to each assay depending on the point of the curve tested. The incubated serum digest volume per experiment was 178µL and contained around 600µg protein content. All the duplicates were done simultaneously, using the same batch-reagents and analyzed in continuous for avoiding further variability issues.

Besides, 12 serum samples were individually digested and SISCAPA enriched using the previous multiplexed CILP1-IVG and CILP1-TFL conditions for antibody-bead conjugation and sample incubation. For this purpose, 178µL of individually serum digests, at a concentration of 3.52µg total protein/µL serum, were incubated with 200fmol of each CILP1 SIS peptides. Sample concentrations were calculated by SPM.

3.2. Results.

3.2.1. Selection of CILP1 and PRG4 proteotypic peptides.

After performing MRM analysis of the chosen CILP1 and PRG4 peptide candidates on cartilage and serum digests, four peptides (two per target protein) were selected for conducting IAMS assays. The peptides chosen for CILP1 were IVGPLEVNVR and TFLVGNLEIR, from now on CILP1-IVG and CILP1-TFL respectively; whereas the PRG4 peptides selected were GGSIQQYIYK and GFGGLTGQIVAALSTAK, from now on PRG4-GGS and PRG4-GFG, correspondingly. **Figure 3. 2** and **Figure 3. 3** show the location of the selected CILP1 and PRG4 proteotypic peptides for IAMS performance respectively. Notably, both CILP1 peptides were located in the chain 2 of the protein, whereas PRG4 peptides appeared at the end of the protein sequence.

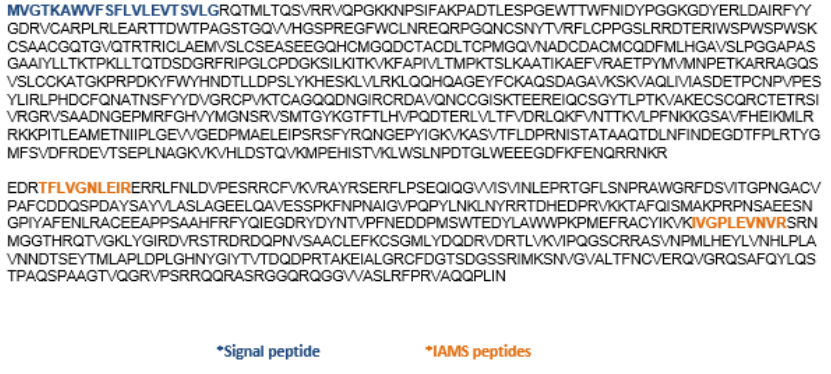


Figure 3. 2. CILP1 sequence (divided in the two corresponding chains) and selected proteotypic peptides for IAMS (orange).

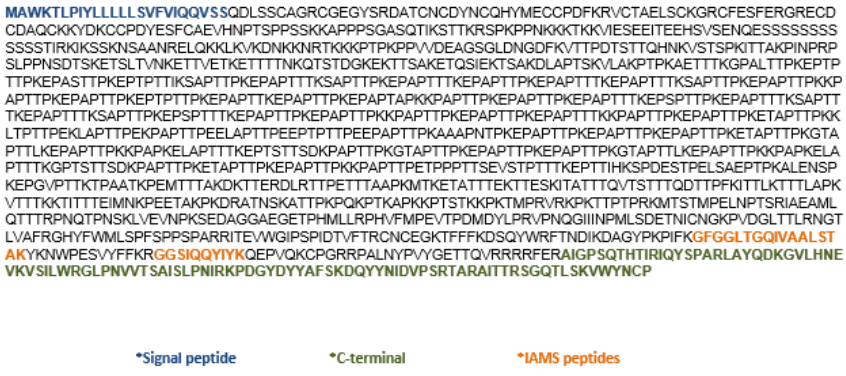


Figure 3. 3. PRG4 sequence and selected proteotypic peptides for IAMS (orange).

On the one hand, **Annex III** show the four selected CILP1 and PRG4 proteotypic peptide MRM chromatograms of cartilage and depleted and/or non-depleted serum samples. On the other hand, MS/MS data was analyzed and searched for protein identification using the ProteinPilot software in order to corroborate the presence of the peptide and the analogous protein. All cartilage digests showed an adequate MRM peak detection (measured transitions coeluted and their ranks matched with the corresponding database MS/MS spectrum). Peptide presence was additionally corroborated by ProteinPilot search for protein identification as well. Furthermore, the four peptides were not detected (or hardly detected as in the case of PRG4-GGS, although transition patterns were different than in cartilage) in non-depleted serum samples. In addition, CILP1 and/or PRG4 peptide and/or protein identifications were not achieved by ProteinPilot search. Finally, CILP1 peptides were not MRM encountered when sera were depleted, whereas PRG4 peptides were considerably detected by MRM, although none of the peptides/proteins were identified by protein database search.

After confirming the presence of the examined CILP1 and PRG4 peptides in cartilage and their absence in non-depleted serum, the complementary SIS peptides were ordered to JPT as previously stated. Moreover, the analogous

peptide and immunogen synthesis were attained and the corresponding anti-peptide antibodies were generated and purified in order to perform the two different IAMS assays. **Table 3. 1** shows the overall characteristics (sequence length, hydrophobicity, molecular weight (MW) and isoelectric point (pI) of the selected CILP1 and PRG4 peptides. Furthermore, **Table 3. 2** indicates the MW of the NAT and SIS peptides (labeled in red), as well as the difference between the MW of the posterior and precedent peptides, arranged by MW (useful for defining the iMALDI suitability since they should not overlap).

Table 3. 1. CILP1 and PRG4 peptide characteristics.

Protein	Peptides	Sequence length (aa)	Hydrophobicity	MW monoisotopic (Da)	Theoretical pI
CILP1	TFLVGNLEIR	10	34,61	1160,6554	7
	IVGPLEVNVR	10	24,99	1094,6448	7
PRG4	GFGGLTGQIVAALSTAK	17	32,53	1589,8778	10,1
	GGSIQQYIYK	10	22,22	1155,5925	9,6

Table 3. 2. Molecular weight (MW) information of the selected IAMS light and heavy labeled (red) peptides.

Protein	Peptides	MW average (Da)	ΔMW (Da)
CILP-1	IVGPLEVNVR	1095,3193	
CILP-1	IVGPLEVNVR (+10)	1105,3193	10
PRG4	GGSIQQYIYK	1156,3153	51
CILP-1	TFLVGNLEIR	1161,3783	5
PRG4	GGSIQQYIYK (+8)	1164,3153	3
CILP-1	TFLVGNLEIR (+10)	1171,3783	7
PRG4	GFGGLTGQIVAALSTAK	1590,8533	419
PRG4	GFGGLTGQIVAALSTAK (+8)	1598,8533	8

3.2.2. Development and analytical validation of iMALDI for the detection of CILP1 and PRG4 in sera from OA patients and controls (healthy donors).

As described in the antibody-bead conjugation section, different beads were tested. MagReSyn Protein A beads were discarded due to their difficult handling during the whole protocol as explained in the cited subdivision. When comparing Protein A and Protein G Dynabeads (**Figure 3. 4**), it can be seen that Protein G beads demonstrate better performance for both CILP1-IVG and/or CILP1-TFL iMALDI assays. Specifically, higher SIS intensity signals for the same spiked peptide amount (1000 fmol) were determined. Furthermore, Protein A beads did not behave as properly as Protein G, considering their handling in some steps of the procedure. From this first assay, it can be noticed that IVG and TFL light peptides were detected while no sample digest was added in the incubation step. Since this is a peptide interference when measuring any of the samples suggested, there is a high need for its removal.

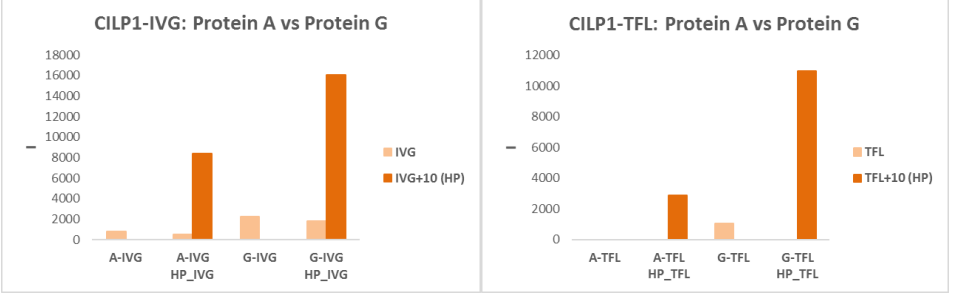


Figure 3. 4. Comparison of the use of Dynabeads Protein A (A) or Protein G (G) for the CILP1-IVG and CILP1-TFL iMALDI assays. Both assays were performed with 2µg of anti-peptide antibody per experiment. Experiments were also conducted with and without the incubation of the corresponding HP.

Two different techniques were used for eliminating this peptide background: the use of the DSS crosslinker in the antibody-bead conjugation procedure and the preliminary acidic ultrafiltration of the antibodies. On the one hand, the use of the DSS crosslinker did not function as expected (**Figure 3. 5**) and

most of the interfering peptide remained on the antibodies after the acidic peptide wash with glycine (G-IVG-XW and G-IVG-XW-HP experiments). Furthermore, the immunoprecipitation of the same SIS peptide amount (1000 fmol) was diminished when the antibodies were crosslinked to the beads. Since these facts, the DSS crosslinking procedure was not elected for peptide interference removal. On the other hand, a preliminary acidic UF of the antibodies was tested as previously described. The UF antibodies were linked to the beads and 1000 fmol of each SIS peptide were added to the corresponding experiments. With this methodology, the intensity signal of the endogenous peptide was effectively silenced, as seen in the CILP1-IVG iMALDI experiments from **Figure 3. 6** and **Figure 3. 7**. Discordantly, when taking into account the CILP1-TFL assay, the peptide interference was not present this time as in **Figure 3. 4** (G-TFL). Despite this controversy, the performance of the assay seems to be boosted, since the same SIS peptide amount was tested (1000 fmol) and the intensity signal increased when the UF antibodies were conjugated for both assays. PRG4-GGS and PRG4-GFG assays did not present this interference so data is not shown for these experiments.

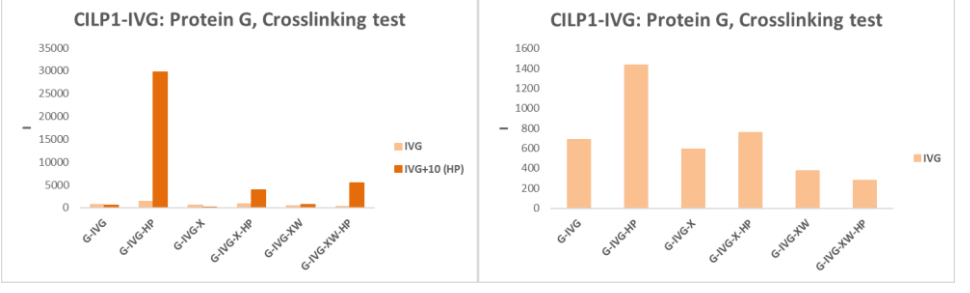


Figure 3. 5. DSS crosslinking test for the removal of the peptide interference in the CILP1-IVG iMALDI assay. G-IVG experiments were tested without the crosslinker, while G-IVG-X and G-IVGXW experiments were performed with the DSS crosslinking procedure, without and with the acidic glycine wash respectively.

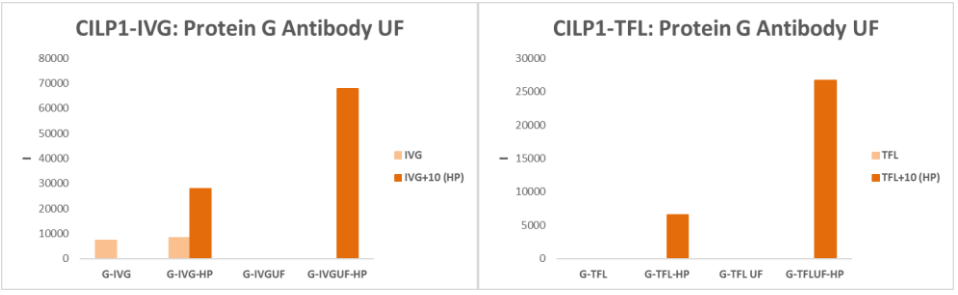


Figure 3. 6. Acidic ultrafiltration (UF) of the antibodies for eliminating the peptide interference in CILP1-IVG and CILP1-TFL MALDI assays. Both assays were performed with 2µg of anti-peptide antibody per experiment. Experiments were also conducted with and without the incubation of the corresponding HP.

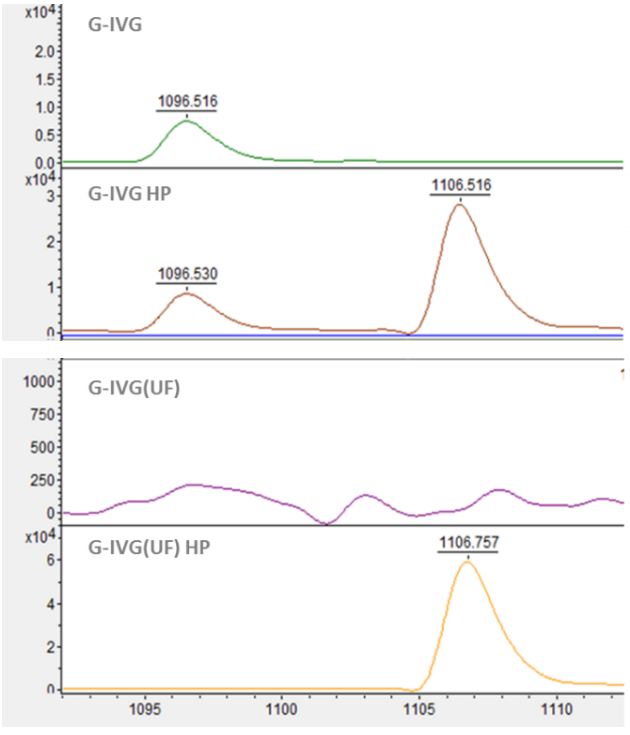


Figure 3. 7. MALDI spectra comparing the use of antibodies without or with acidic UF and CILP1-IVG iMALDI performance when analyzing the corresponding CILP1-IVG HP (around 1106 m/z). G-IVG and G-IVG HP spectra show the peptide interference at m/z 1096.516 or 1096.530 when using the original antibodies, while this peptide background is eliminated when the preliminary acidic UF of the antibodies is performed in both G-IVG (UF) experiments.

Once the bead type (Dynabeads Protein G) and the peptide background removal technique (prior acidic ultrafiltration of the antibodies) were determined, the bead saturation was checked for selecting the appropriate antibody-bead ratio (**Figure 3. 8**). In the case of CILP1-IVG and CILP1-TFL, the antibodies used were firstly cleaned from the peptide interference using the acidic ultrafiltration procedure, while PRG4-GGS and PRG4-GFG assays used the originally purified antibodies without any posterior modification. Different amount of antibody (from 0.2 to 2µg) were conjugated to the same bead amount (1µL bead slurry) for detecting the same SIS peptide amount (1000 fmol). After checking the collected results, the following antibody µg per µL of bead slurry were selected for proper antibody-bead conjugation: 2µg/µL (CILP1-IVG), 1.5µg/µL (CILP1-TFL), 1µg/µL (PRG4-GGS) and 1µg/µL (PRG4-GFG). In most of the experiments, the saturation level selection was quite clear, although in the case of PRG4-GFG the antibody-bead ratio was not so obvious. Particularly, the experiment with the highest SIS signal level and antibody amount was chosen as the appropriate antibody-bead ratio in order to ensure the optimal assay performance.

Figure 3. 8. Antibody-bead saturation tests for CILP1-IVG, CILP1-TFL, PRG4-GGS and PRG4-GFG iMALDI assays. At the x-axis: amount of antibody incubated onto 1µL bead slurry. At the y-axis: intensity signal of the spiked SIS peptide at a constant amount of 1000 fmol.

After settling the antibody-bead conjugation parameters (bead type and antibody-bead ratio), pooled serum digests were incubated with the corresponding determined antibody-bead couplings. Routinely, positive and negative controls were analyzed in order to check the validity of every tested iMALDI experiment. On the one side, the FLNVLSPR iMALDI assay, used as the positive control, has given favorable results as both the light and heavy peptide peaks can be clearly detected by iMALDI in any of the sample incubation conditions tested. One spectrum example of this positive control is shown in **Figure 3. 9**. On the other side, the spectra for the negative control (using only the beads and a pool of sera) reveals some issues in terms of sample background interference (**Figure 3. 10**). In this spectra, a couple of intrusive background peaks are present in the m/z area where most of the CILP1-IVG, CILP1-TFL and PRG4-GGS peptides are detected. To the contrary, the other negative control which takes into account the specificity of the antibodies to their corresponding peptides was not giving any misbehaving outcomes so data is not shown due to the absence of the tested SIS peptide peaks.

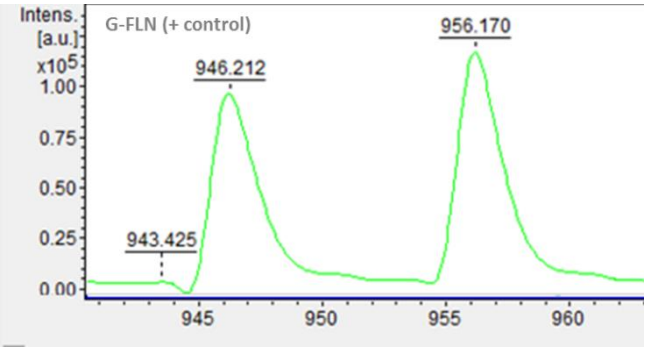


Figure 3. 9. Positive iMALDI control spectrum using the FLNVLSPR assay. On the left the 946.212 peak corresponds to 500 fmol of light peptide while 956.170 corresponds to the same amount of the analogous heavy labeled peptide.

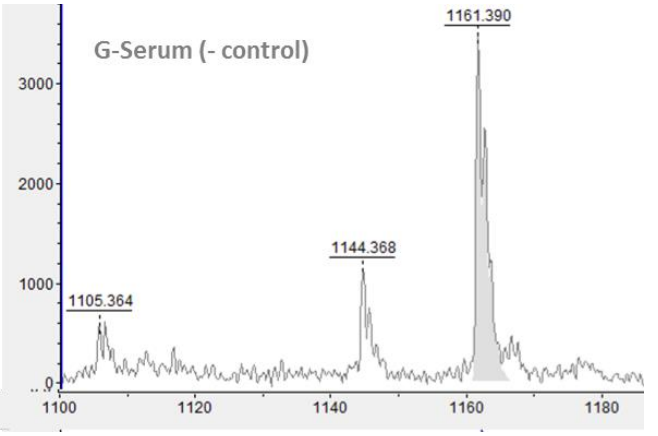


Figure 3. 10. Negative iMALDI control spectrum. Protein G Dynabeads were washed and coupled to 12µL of a pooled serum digest for checking unspecific bounding of the sample to the beads.

In order to detect the corresponding CILP1 and/or PRG4 peptides in serum samples, 12µL of a pooled serum digest and 500-1000 fmol of SIS peptides were incubated with the corresponding antibody-bead conjugation. **Figure 3. 11** shows the 4 CILP1 and PRG4 iMALDI assays performed as described before. Specifically, CILP1-IVG and PRG4-GGS endogenous peptide peaks (1096.530 and 1156.125 m/z, respectively) are not markedly detected, even when the serum digest quantity was increased (data not illustrated). In contrast, CILP1-TFL and PRG4-GFG endogenous peptide peaks (1162.396 and 1592.144 m/z respectively) can be quite clearly observed. As stated in the last paragraph, some sample background peaks can be found in some of the experiments. Specifically, there is one peak arounf 1161-1162 m/z that can be seen in both the negative control (**Figure 3. 10**) and in the PRG4-GGS iMALDI assay (**Figure 3. 11**). Moreover, this interference peak presents the same m/z of the CILP-TFL endogenous peak (1162.396 m/z), so it is likely that the endogenous TFL is simultaneously measured with this interference. Since this could lead to overestimation measurements of CILP1-TFL, this iMALDI assay was not prompt for undergoing the following method validation. The same decision was taken for both CILP1-IVG and PRG4-GGS iMALDI assays due to the low detectable endogenous concentration and/or peak

intensities observed in serum samples. In conclusion, only the PRG4-GFG iMALDI assay offers an adequate performance in terms of serum sample detection, absence of interfering sample background as well as acceptable peak shapes and intensities among other detection parameters.

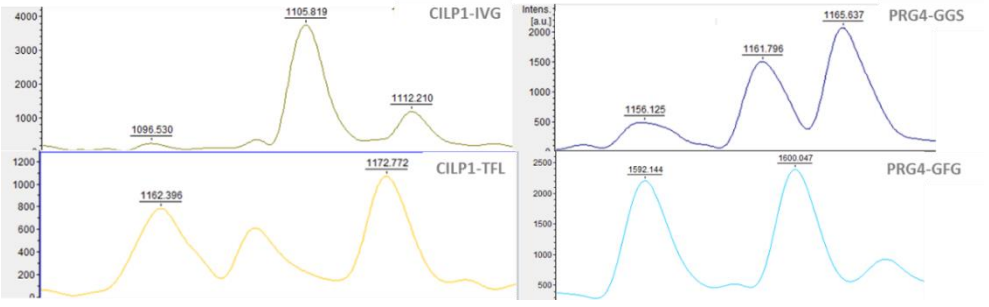
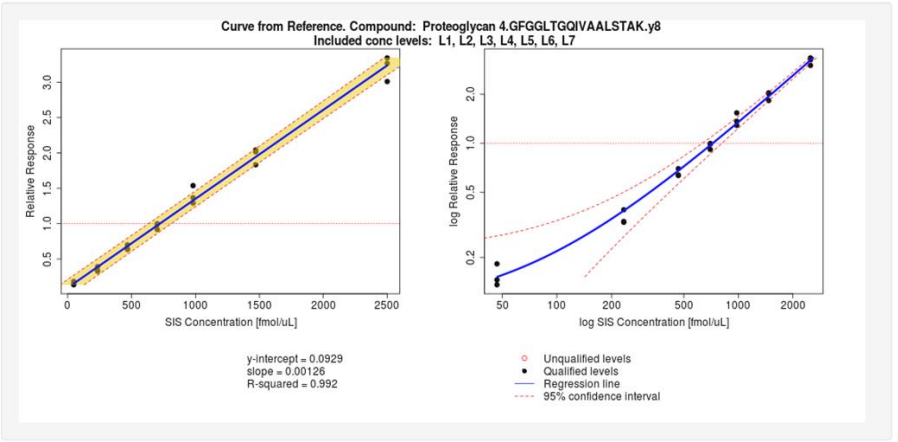


Figure 3. 11. Serum digest (12µL) and SIS peptide incubation spectra for the different CILP1 and PRG4 iMALDI assays.

Finally, in sight of the previous conclusions, the PRG4-GFG iMALDI method validation was carried out. In the first place, the iMALDI response curve was performed and analyzed in four replicates. When looking at the curve (**Figure 3. 12**), an exceptional regression (R^2: 0.992) can be observed when taking into account three of the four curve replicates measured. The curve was calculated using the QUALIS-SIS software, which determines the H/L ratios (relative responses) for every replicate point tested and plots them against the corresponding SIS peptide amount added. This software does all the calculations and interpretations based on a 10% of accuracy and/or precision. Thus, the LLOQ for the sample measurement is stated as 46.685 fmol, and the amount of NAT (endogenous peptide) in the 12µL pooled serum is around 664.306 fmol, based on the software calculations. The LLOD (3/10*LLOQ) would be then around 14 fmol as well. **Figure 3. 13** shows the spectra for the different 8 points of one of the four replicates of the calibration curve tested. Particularly, the peak intensity for the SIS peptide (around 1600 m/z) is increased as the SIS peptide amount is incremented as well.



	Criteria	Qualification Requirements
1	Precision Excluding LLOQ	av. <20% CV per level
2	Precision at LLOQ	av. <20% CV
3	Accuracy Excluding LLOQ	20%
4	Accuracy at LLOQ	20%
5	Min. No. of Levels	3 consecutive
6	No. of Acceptable Replicates/Level	per qualified level

	Metric	Value
1	LLOQ (ng/mL)	6924.090
2	LLOQ (fmol/uL)	46.685
3	ULOQ (ng/mL)	370785.000
4	ULOQ (fmol/uL)	2500.000
5	Dynamic Range	53.550
6	Conc NAT in Reference (ng/mL)	98526.110
7	Conc NAT in Reference (fmol/uL)	664.308

Figure 3. 12. PRG4-GFG iMALDI calibration curve regression and calculation of LLOQ and sample concentration using the QUALIS-SIS software.

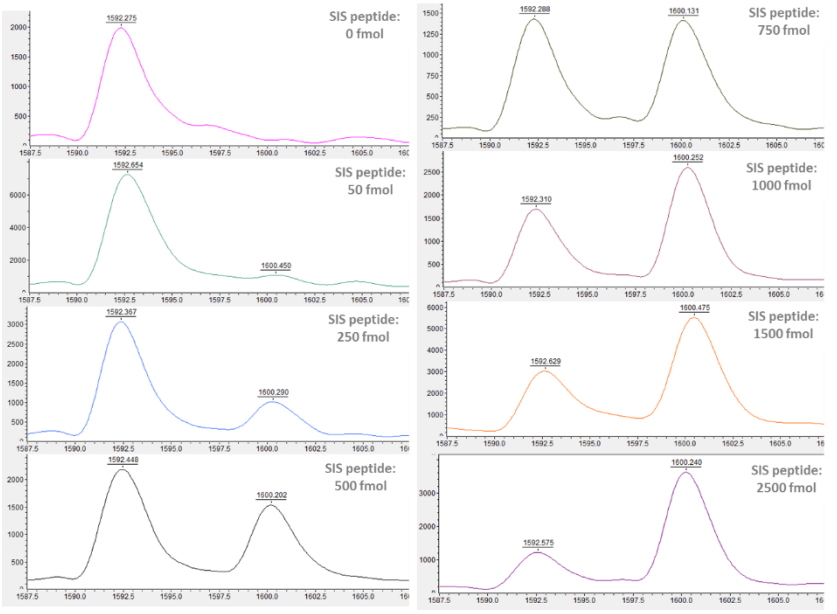


Figure 3. 13. PRG4-GFG iMALDI calibration curve spectra using 12µL pooled serum digest and increasing amounts of SIS peptide (from 0 to 2500 fmol).

In the second experiment of the method validation, once the linear range of the assay and the LLOQ were determined, the variability of the assay was also verified. Low, medium and high spiked SIS quantities (117.5, 470 and 2350 fmol respectively) were calculated based on the LLOQ (47 fmol). These three concentration levels were analyzed during 5 days using 5 replicates per day and level. For this aim, 12µL of pooled serum digest and the corresponding SIS peptide amount were incubated in each experiment. In terms of assay precision, different % CV were obtained depending on the concentration level and the day of the analysis. Thus, the precision of the low concentration level varied from 4.30 to 14.03, whereas the medium stayed between 2.85 and 6.95 and the high among 1.95 and 12.88 % CV (**Annex III**). Moreover, the assay accuracy was also determined in terms of % nominal (**Annex III**). The % nominal for the low concentration level was 153.31%, while for the medium and high levels the % nominal ameliorated since they were closer to 100%, being 112.72% and 90.71% respectively. Furthermore, intraday, interday and total variability data was determined (**Table 3. 3**). Thus, average intraday CVs (%) remained under 10% in every concentration level tested, while only the medium and high concentration stages remained below this percentage when examining the interday CVs. Likewise, focusing on the total variability, the medium concentration level solely presented CV (%) beneath 10%, although the low and high concentration CVs (%) stayed under 15%.

Table 3. 3. PRG4-GFG iMALDI assay variability (total, intraday and interday).

Variability (% CV)	LOW (117,5 fmol)	MED (470 fmol)	HIGH (2350 fmol)
Average Intraday	6,79	4,38	7,19
Average Interday	12,77	7,84	8,14
Total Variability	14,47	8,98	10,86

After validating and confirming the linear response, good regression (R^2: 0.992) and the low variability (< 15%) of the PRG4-GFG iMALDI assay, serum samples from 22 patients suffering OA and 16 healthy donors were analyzed in

triplicates. For this aim, serum samples were individually and simultaneously digested and the resulting digests were incubated along with 750fmol PRAG4-GFG SIS peptide. This SIS peptide amount was chosen based on the NAT/SIS signal obtained in the previous calibration curve (**Figure 3. 13**) were both peaks were balanced and showed hardly the same intensities. Once again, the identical calibration curve was prepared in triplicates with the rest of the serum samples, so their concentration could be determined by standard curve measurement (SCM) using the QUALIS-SIS software (**Figure 3. 14**). Single point measurement (SPM) was also employed by the QUALIS-SIS software when the qualifying parameters for using SCM were not achieved. Both measuring methods were later compared for quantifying PRG4-GFG in serum samples. The regression of the actual response curve was as adequate as the first curve performed (R^2= 0.990) and the LLOQ was approximately the same as well (46.685 fmol), confirming the previous curve results (**Annex III**).

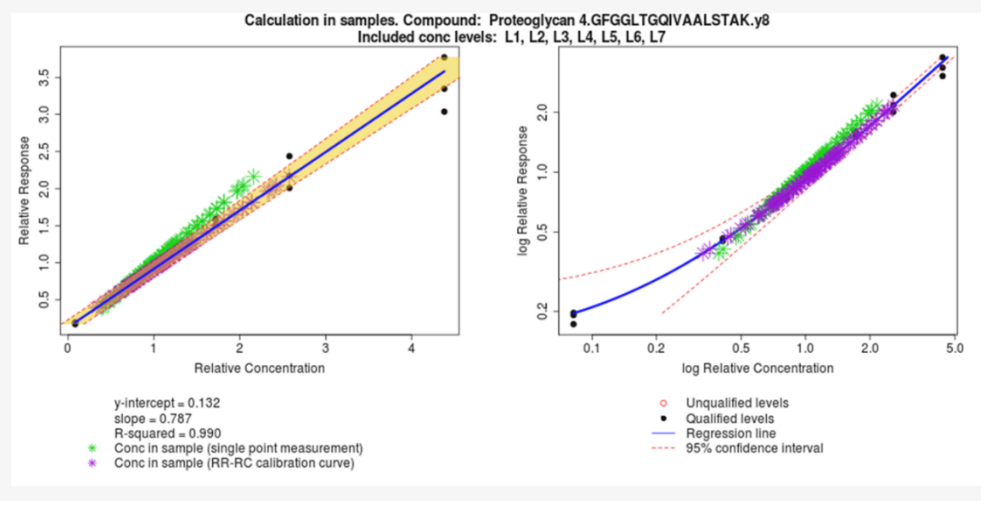


Figure 3. 14. PRG4-GFG iMALDI analysis of 38 sera by Standard Curve Measurement using QUALIS-SIS.

Average PRG4-GFG concentrations in fmol/µL serum and µg/mL serum were calculated for the OA patients and HD using both SCM and SPM methods (**Annex III**). When comparing both groups, either using the SCM or the SPM, there

is not a significant difference so OA patients could not be discriminated from HD using the PRG4-GFG iMALDI assay (**Table 3. 4** and **Figure 3. 15**). The PRG4-GFG serum concentration of both subgroups was settled around 66.83-67.27 fmol/µL and 9.60-9.67 µg/mL (SPM) or 61.98-62.73 fmol/µL and 8.91-9.01 µg/mL (SCM). Differences between the SPMs and SCMs (%) were lower than 20% in all the values except for two samples which differences were around 23-24% (**Annex III**) (109).

Table 3. 4. PRG4-GFG average concentration (fmol/µL and µg/mL) in serum samples from OA patients and HD calculated with SPM and SCM.

PRG4-GFG (Average)	(fmol/µL serum)		(µg/mL serum)	
	SPM	SCM	SPM	SCM
OA	66,83	61,98	9,60	8,91
HD	67,27	62,73	9,67	9,01

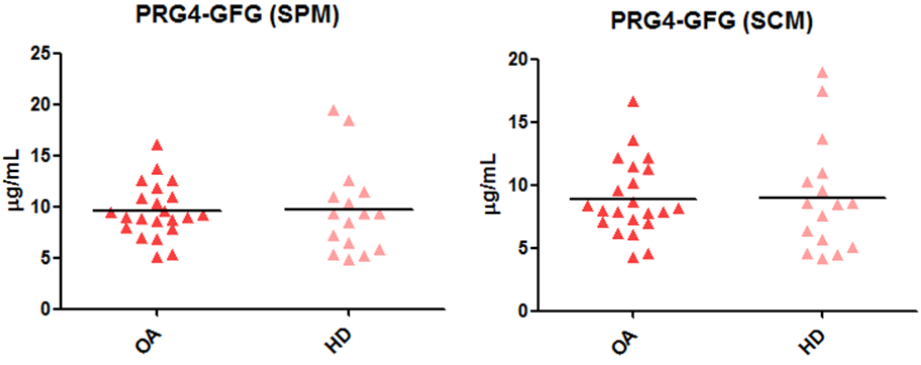


Figure 3. 15. SPM and SCM graphs detecting PRG4-GFG in serum from OA patients and HD by iMALDI.

Finally, the same serum samples were analyzed by ELISA in order to technically validate the developed PRG4 iMALDI assay. After performing dilution tests, samples were chosen to be diluted 1/100 with sample diluent for PRG4 ELISA detection using one replicate per diluted sample. Moreover, a PRG4 calibration curve ranging from 100 to 0 ng/mL was performed by using 2/3 serial dilutions of the PRG4 standard with sample diluent, obtaining a total of 8 points

including the blank. Calibration curves were analyzed in duplicates and a final calibration curve using the average of each point was applied for quantifying the samples (**Figure 3. 16**).

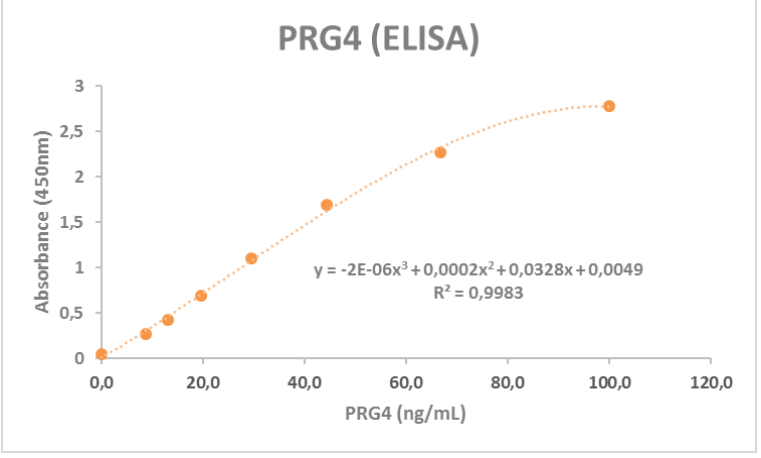
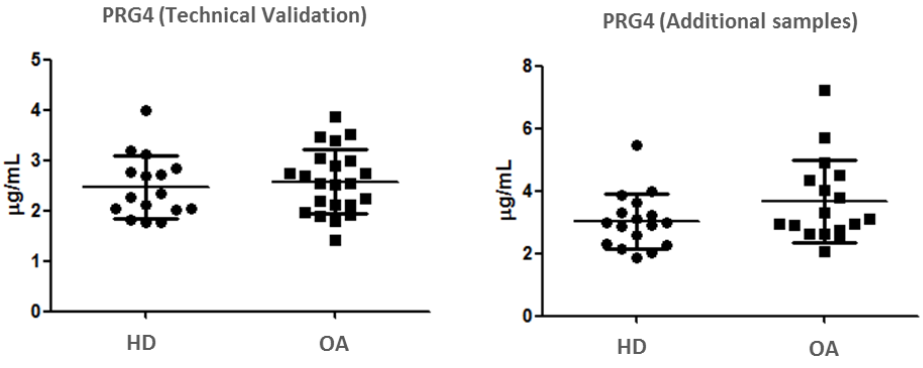


Figure 3. 16. PRG4 ELISA calibration curve used for sample quantitation (averaged duplicate points).

After performing the ELISA, samples were quantified using the previous calculated standard curve. As seen in **Figure 3. 17**, PRG4 levels of the samples analyzed by the developed PRG4 iMALDI did not show significant difference between OA and HD (p-value: 0.605), differing 0.1 as in the iMALDI assay, although this time OA PRG4 levels were slightly higher than in HD samples. Interestingly, the same upregulated trend in OA was better seen when additional OA and HD samples were analyzed, displaying a non-significant (p-value: 0.191) but greater difference between samples (3.67 vs 3.03 µg/mL, respectively).



ELISA	PRG4 (µg/mL)	
	Technical validation	Additional samples
OA	2,57 (22)	3,67 (17)
HD	2,47 (16)	3,03 (17)
p- value	0,605	0,191

Figure 3. 17. Serum sample PRG4 quantitation using the iMALDI samples for accomplishing technical iMALDI validation (22 OA and 16 HD) and additional serum samples (17 OA and 17 HD).

Nevertheless, although no significant and similar differences were found between OA patients and HD, when analyzing the same serum samples by the developed PRG4 iMALDI SPM or SCM or the commercial PRG4 ELISA, the correlation of these measurements by both assays was quite poor, where r=0.37 (R^2: 0.14) when comparing iMALDI (SPM) and ELISA measurements and r=0.36 (R^2: 0.13) for iMALDI (SCM)-ELISA comparison (**Figure 3. 18**).

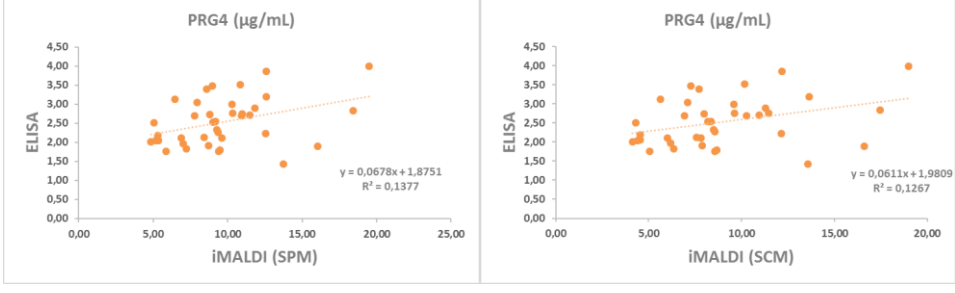


Figure 3. 18. iMALDI and ELISA correlation when measuring PRG4 in serum samples (22 OA and 16 HD).

3.2.3. Development and analytical validation of SISCAPA-nanoLC-MRM for the detection of CILP1 in sera from OA patients and controls (healthy donors).

First of all, the MRM method was designed for both CILP1-IVG and CILP1-TFL endogenous/light (NAT/L) and heavy (SIS/H) peptides. Thus, in order to verify the retention times and peak transitions for both CILP1 peptides, 200fmol of each digested SIS peptide were jointly analyzed by MRM. **Figure 3. 19** shows both chromatograms where CILP1-IVG and CILP1-TFL SIS peptide peaks can be observed at 25.7' and 30.5' respectively. Moreover, the figure also represents the selected optimal transitions for both peptides (located on the right side of their corresponding precursor peaks), which display a satisfactory co-elution for peptide qualification and identification.

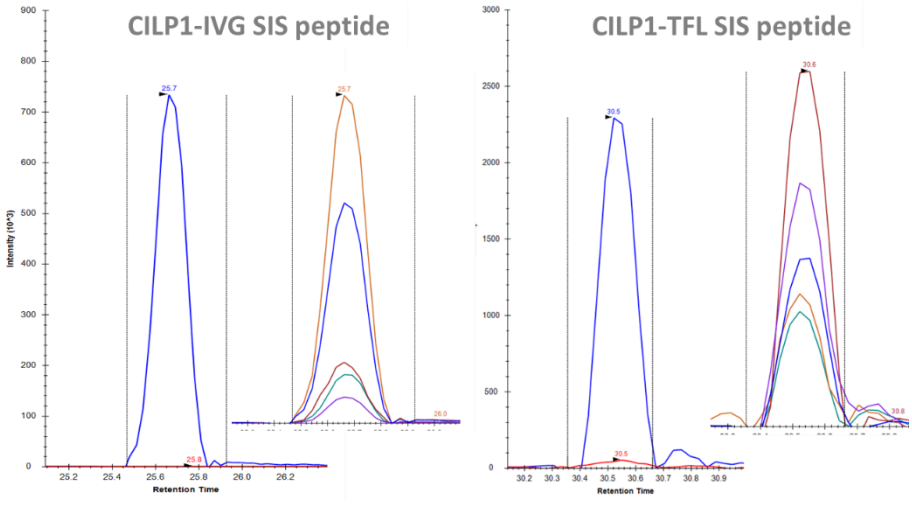


Figure 3. 19. CILP1-IVG and CILP1-TFL SIS peptide MRM chromatograms.

Posteriorly, a study of the cited contaminant peptides, likely coming from the affinity-purified antibodies, and SISCAPA performance tests were conducted in order to define both CILP1-IVG and CILP1-TFL assays. For this aim, three experimental controls were used. Firstly, in order to test the presence of the interfering peptides, CILP1-IVG and CILP1-TFL antibody-bead conjugations were incubated only with PBSC. As seen in experiments "A-CILP1-IVG PBSC" and "A-

CILP1-TFL" from **Figure 3. 20**, intense interfering NAT/L peaks (red) can be seen at 25.6' and 30.5' respectively. Therefore, the presence of the already cited contaminant peptides for both anti-peptide antibodies was also confirmed by SISCAPA-MRM analysis, since any peak should arise on these tests. Secondly, CILP1-IVG and CILP1-TFL antibody-bead conjugations were incubated with both CILP1-IVG and CILP1-TFL SIS peptides in order to test the performance of the assay and the cross-reactivity of the antibodies. On the one hand, experiments "A-CILP1-IVG SIS" and "A-CILP1-TFL SIS" show the proper SISCAPA enrichment of the corresponding SIS peptides to their antibodies. CILP1-IVG and CILP1-TFL SIS peptide peaks (blue) are observed at 25.8' and 30.5' respectively. Furthermore, the corresponding NAT peptide peaks (red) are also present, corroborating again the existence of the contaminant peptides. Additionally, the simultaneous analysis with the analogous SIS peptides could then indicate the amount of interfering peptide present. Thus, peptide interferences (NAT) were calculated by SPM, being around 276fmol for CILP1-IVG and 97fmol for CILP1-TFL. On the other hand, experiments "A-CILP1-IVG SIS-TFL" and "A-CILP1-TFL SIS-IVG" examine the cross-reactivity of CILP1-IVG and CILP1-TFL for the contrary corresponding SIS peptides. In both chromatograms, the nonappearance of the matching SIS peptide peaks (in "A-CILP1-IVG SIS-TFL" around 30.5' and in "A-CILP1-TFL SIS-IVG" around 25.7') confirms there is not cross-reactivity of the antibodies for the inverse-matched peptides. This fact facilitates the posterior multiplex capability of both assays, as each of them only measures their analogous analyte.

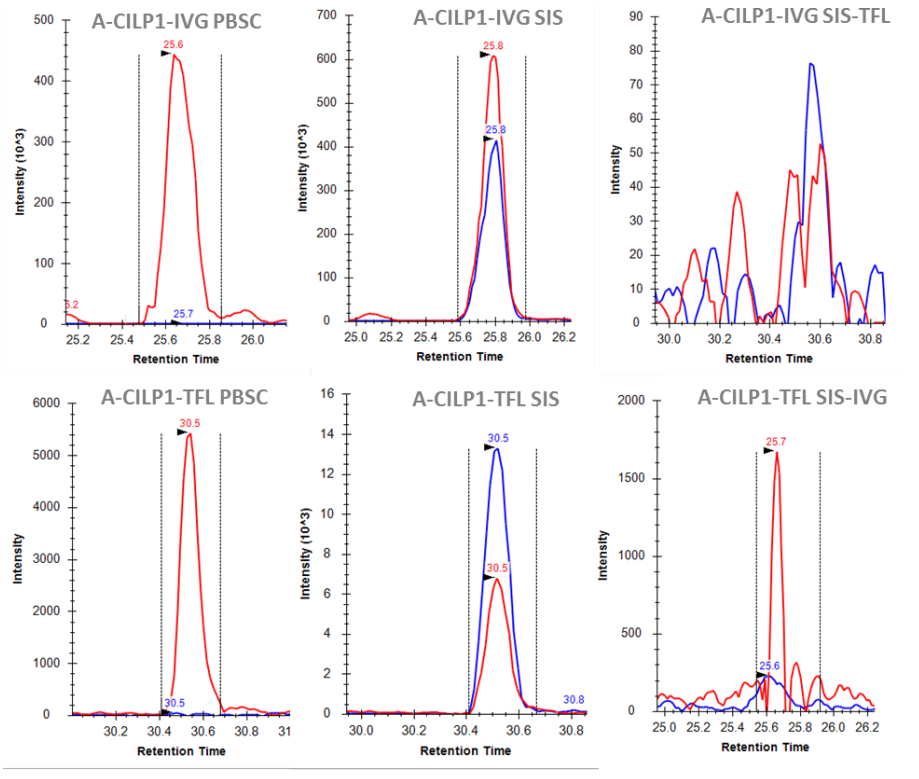


Figure 3. 20. CILP1-IVG and CILP1-TFL assays for checking peptide interference and SISCAPA performance. Protein A and CILP1-IVG/CILP1-TFL antibody couplings are represented as A-CILP1-IVG and A-CILP1-TFL, respectively. NAT peptides are represented in red, whereas SIS peptides are shown in blue. The first chromatogram of each assay displays the incubation with PBSC, while the second shows the enrichment of the corresponding SIS peptide and the last illustrates the absence of cross-reactivity when CILP1-TFL SIS peptide is enriched by CILP1-IVG SISCAPA and vice versa.

Within sight of the previous facts, the acidic ultrafiltration of the anti-peptide antibodies was carried out in order to remove the peptide contamination. Then, the UF antibodies were coupled to Protein A Dynabeads and 200fmol of each corresponding SIS peptides were incubated. As a control, the same procedure was simultaneously followed, although non-UF anti-peptide antibodies were used this time. **Figure 3. 21** shows the chromatograms of the cited SISCAPA-MRM experiments. For both CILP1-IVG and CILP1-TFL assays, the corresponding peptide interference (NAT peak in red at 25.7' and 30.5' respectively) was

successfully removed as seen in the "A-CILP1-IVG(UF) SIS" and "A-CILP1-TFL SIS" chromatograms, where only the SIS peptide peaks (blue) are observed. Moreover, a graph on the right side of each CILP1 assay, showing the peak area for both NAT and SIS peptides, is illustrated. According to these charts, the disappearance of the peptide interference (red) can be seen, as well as the increase of each SIS peptide signal, when the UF of both CILP1 antibodies is performed. Furthermore, SPMs were done for quantifying the peptide contamination. Both CILP1-IVG and CILP1-TFL NAT interfering peptide amounts were reduced from 242 to 0.54fmol and from 97 to 0.56 fmol respectively. Thus, the anti-peptide antibodies, used for the rest of the experiments conducted, were previously processed with the defined acidic UF protocol.

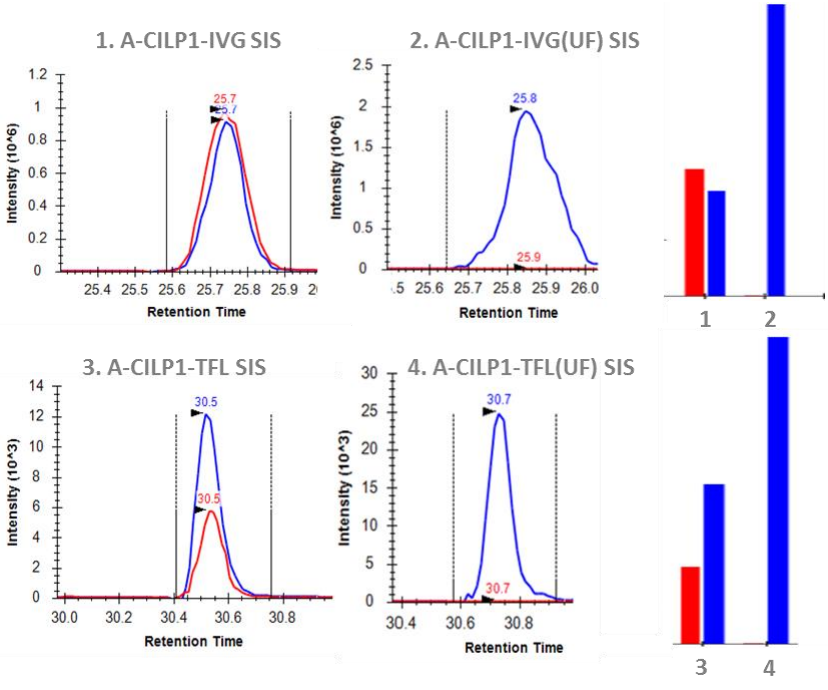


Figure 3. 21. CILP1-IVG and CILP1-TFL SISCAPA tests for removing the corresponding peptide interferences. A-CILP1-IVG SIS and A-CILP1-TFL SIS experiments were carried out with the original affinity-purified anti-peptide antibodies, whereas A-CILP1-IVG(UF) SIS and A-CILP1-TFL(UF) SIS experiments were conducted with the same antibodies after performing a previous acidic UF. 200fmol of the corresponding SIS peptide were added per experiment. On the right, the NAT (red) and SIS (blue) peptide peak areas were graphed for both CILP1-IVG and CILP1-TFL assays.

After settling the antibody-bead coupling, 12 serum samples from OA patients and HD were pooled and digested in order to perform CILP1 multiplexed SISCAPA analysis by combining both CILP1-IVG and CILP1-TFL SISCAPA assays. For this purpose, 1.8µg of each CILP1 anti-peptide antibody were coupled to 15µL Protein A Dynabeads in each experiment. Then, 200 fmol of each SIS peptide and 18µL of the pooled serum digest were immediately incubated overnight and posteriorly analyzed by MRM. **Table 3. 5** displays the multiplex analysis for both CILP1-IVG and CILP1-TFL assays, where NAT NAT amounts per µL serum were 1.95 fmol/µL and 6.38 fmol/µL respectively. Interestingly, CVs were lower than 20% in both CILP1-IVG and CILP1-TFL analysis (0.32% and 17.99% specifically). Hence, multiplex analysis were used for the following SISCAPA-MRM method validation.

Table 3. 5. CILP1-IVG and CILP1-TFL SISCAPA assays when measured conjointly (multiplex).

Multiplex SISCAPA experiment	NAT/SIS ratio	fmol NAT /µL serum	Average fmol NAT /µL serum	CV (%)
A-CILP1-IVG Multiplex	0,18	1,95	**1,95**	0,32
	0,18	1,96		
A-CILP1-TFL Multiplex	0,51	5,57	**6,38**	17,99
	0,65	7,20		

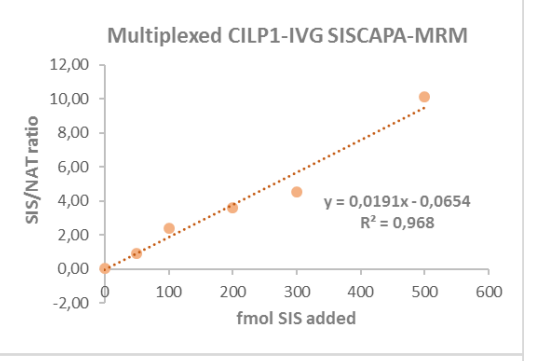
After the multiplex capability for both SISCAPA CILP1 assays was confirmed, the validation of the multiplex CILP1-IVG and CILP1-TFL SISCAPA-MRM method was conducted. For this aim, a 6-point response curve was designed and carried out in duplicates. Again, 1.8µg of each anti-peptide antibody were coupled to 15µL Dynabeads. Subsequently, 18µL of a pool of the 12 individually digested sera were incubated with the corresponding SIS CILP1-IVG and CILP1-TFL amounts (fmol) per experimental point of the curve (**Table 3. 6**). **Annex III** shows the average SIS/NAT ratios for every point of the curve examined as well as the CV (%) for each of them. Focusing on the CILP1-IVG curve, CVs for all point were lower than 20% except for the point were no SIS peptide was added (29.82%). This CV is reasonably typical since the addition of SIS peptides normalizes NAT signals and

when they are not added, the variability could be greater. In contrast, 3 points of the CILP1-TFL curve displayed CVs above 20%: 1000fmol SIS (41.46%), 750fmol SIS (21.70%) and 250fmol SIS (23.72%). Nevertheless, the regression for both multiplex CILP1-IVG and CILP1-TFL SISCAPA assays were quite satisfactory (R^2=0.968 and R^2=0.9947, respectively) as displayed in **Figure 3. 22**.

Table 3. 6. Multiplex CILP1-IVG and CILP1-TFL response curve, showing the 6 points designed for conducting the validation of the combined SISCAPA assays.

Multiplex CILP1-IVG and CILP1-TFL Response Curve						
fmol SIS CILP1- IVG	500	300	200	100	50	0
fmol SIS CILP1-TFL	1000	750	500	250	125	0

fmol SIS added	SIS/NAT ratio
500	10,11
300	4,51
200	3,61
100	2,41
50	0,90
0	0,02



fmol SIS added	SIS/NAT ratio
1000	4,91
750	3,66
500	2,56
250	1,62
125	0,82
0	0,04

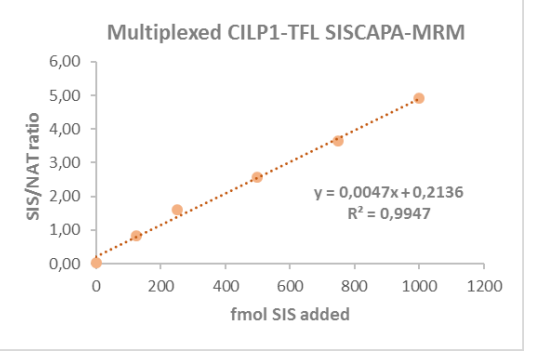


Figure 3. 22. CILP1-IVG and CILP1-TFL calibration curves. On the x-axis the SIS amount added in each multiplexed CILP1 SISCAPA assay and on the y-axis the SIS/NAT ratio obtained by dividing the corresponding peak area of the spiked SIS peptide against the peak area of a constant NAT peptide in serum.

Moreover, 12 serum samples, 6 from OA patients and 6 HD, were examined with the validated multiplexed CILP1 SISCAPA assay. Likely, the same

antibody-bead ratio and the serum volume were used for SISCAPA performance and enrichment. Additionally, 200fmol of each SIS peptide were added per experiment as well in order to later perform SPM. As seen in **Table 3. 7**, the average concentration of CILP1-IVG in HD was 1.28µg/mL serum while in OA patients it was slightly higher, 1.38µg/mL serum. There was not a great difference among the two groups since the fold change (OA/HD) was 1.08. Equitably, the CILP1-TFL average concentration in OA patients was greater than in HD (8.76µg/mL serum vs 6.19µg/mL serum respectively). Nonetheless, fold change (OA/HD) was fairly elevated to 1.42 in contrast with CILP1-IVG. However, none of the differences have reached the statistically significance for being acclaimed as cartilage biomarkers for monitoring OA yet (**Figure 3. 23**).

Table 3. 7. Average CILP1-IVG and CILP1-TFL concentrations in serum expressed as fmol/µL and µg/mL serum.

CILP1 Peptide	Group	Average NAT (fmol/µL serum)	Average NAT (µg/mL serum)
IVGPLEVNVR	HD	9,80	1,28
	OA	10,60	1,38
TFLVGNLEIR	HD	47,49	6,19
	OA	67,25	8,76

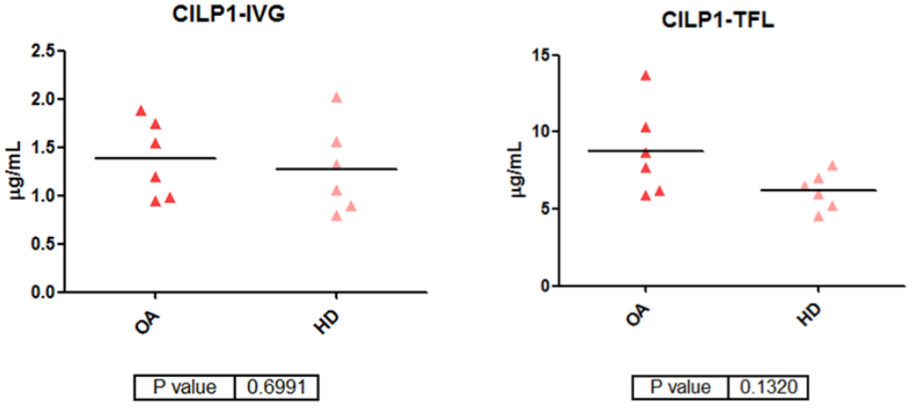


Figure 3. 23. CILP1-IVG and CILP1-TFL graphs representing the 12 sera concentrations (µg/mL) from OA and HD. Under each representation, the p-values obtained after performing Mann Whitney test comparing both group concentrations.

4.3.3. Discussion.

IA-MS assays provide numerous advantages over MS methods and ligand binding assays (LBAs) in terms of specificity, selectivity, sensitivity, capacity for multiplexing or method developing time, although these statements depend on the nature of the analyte(s) involved (90). Therefore, IAMS methods are generally used when the specificity of the LBAs is not enough or need further testing, or when the target analyte needs to be enriched for increasing MS sensitivity (88). This is the case for analyzing CILP1 and PRG4 in sera, since no conventional ELISA was found with satisfactory performances and targeted MS analysis could not detect them properly in serum, without the need of any other preliminary depletion step. Thus, different objectives have been accomplished in this project for performing different IAMS analyses of CILP1 and PRG4 cartilage-related proteins in sera from OA patients and HD.

First of all, two proteotypic peptides for each protein were selected after performing MRM analysis on cartilage and serum digests, attending to the established requirements. On the one hand, the presence of these proteotypic peptides was identified in cartilage digests, since they appertain to proteins conferred in this tissue. On the other hand, no detection or identification of these proteins was accomplished when the selected proteotypic peptides were analyzed in serum digests, likely on account of being medium to low-abundant proteins. Thus, in order to detect these minor peptides/proteins in sera, different depletion methods might be applied. Depletion of the most abundant proteins in serum is generally the most applied technique for being capable of measuring and quantitate low-level proteins. However, it is widely known that this procedure might cause the removal of non-specifically-bound proteins and therefore the loss of possible relevant information, as already explained in the introduction and chapter I (68). Anyhow, depletion of the 20 most abundant serum proteins was carried out in serum samples and the four chosen CILP1 and PRG4 peptides were

analyzed by MRM in the depleted serum digests. The results confirm the absence of CILP1 peptides also in the depleted serum while PRG4 peptides were detected this time when serum was depleted. This was possibly due to a higher PRG4 concentration in serum or by reason of the picked peptide nature and behavior on MRM analysis. Nevertheless, other depletion strategies as affinity-enrichment are more suitable when the verification or validation of defined possible biomarkers are performed. This is due to the fact that the protein of interest (POI) is specifically isolated from the complex serum matrix instead of manipulating the whole matrix in order to simplify the analysis of the POI, misleading valuable data as a result of a not-so-specific depletion. Due to the incapability of detecting and/or identifying CILP1 in and PRG4 in non-depleted serum, IAMS assays were developed for the analysis of these cartilage-related proteins in this fluid.

On the one hand, immunoMALDI (iMALDI) assays were optimized for analyzing CILP1 and PRG4 proteins in serum samples. Although the four CILP1 and PRG4 iMALDI assays operated appropriately, as demonstrated with the binding and analysis of the corresponding SIS peptides, only one PRG4 peptide (PRG4-GFG) could be conveniently measured by iMALDI. On the minus side, the other three peptides from CILP1 and PRG4 did not fulfilled the procedure requirements, since different facts emerged when testing the sera using the set up iMALDI protocol. For instance, interfering peaks appeared on the m/z range spectrum of the corresponding peptides, hampering the adequate iMALDI analysis due to overlapping issues. Additionally, the peak signal was not enough for measuring CILP1-IVG peptide in serum (NAT). In order to overcome these problems, deeper protocol adjustments should have been taken in order to tackle the different issues (washing steps, digestion procedure, bead handling, etc…). Nevertheless, this iMALDI set up was done during a 3-months international research stay in Victoria (BC, Canada), so time was the limiting circumstance for not proceeding with further optimization of the non-succeeding CILP1 and PRG4 assays. Besides, it should be considered that, at first, the election of the target peptides was not

done acknowledging their use for iMALDI and more attention was given to their performance in nanoLC-ESI-MRM. In line with the results, an initial iMALDI interfering test, using the stated negative control, should have been performed in order to avoid undesirable overlapping problems. This experiment would help the selection of the iMALDI peptides, as they could be chosen not only by their sensitivity but also by their selectivity within the m/z spectrum. Withal, the four SIS peptide signals displayed satisfactory selectivities and sensitivities even when iMALDI was performed.

Thus, PRG4-GFG iMALDI assay was validated in terms of assessing the linear range and variability of the method. Regression for the two curves performed in triplicates was outstanding (R^2=0.99) and, in both cases, the points of the curve displayed CVs (%) below the universally established 20% in terms of accuracy and precision. Besides, method variability data proved to be quite satisfactory whether they are compared using the Assay Development Guidelines from the CPTAC Assay Development Working Group (189) or the FDA Guidance for Industry: Immunogenicity Testing of Therapeutic Protein Products — Developing and Validating Assays for Anti-Drug Antibody Detection (190). Both guidelines accept variability scores (including precision, intraday, interday and total variability) up to 20% (CV %), requirements that were fulfilled when the PRG4-GFG iMALDI validation was performed. Specifically, the medium level concentration (approximate serum's level concentration) displayed variability values under 10% (CVs %) in terms of precision (2.85-6.95%), intraday, interday and total variabilities (4.38, 7.84 and 8.98% respectively). The accuracy of the method was also quite suitable as % nominal was 112.72%, close enough to the 100% exact value. Despite, extreme level concentrations (low and high) still displayed admissible variability values (CVs below 15%) in terms of precision (1.95-14.03%), intraday (6.79-7.19%), interday (8.17-12.77%) and total (10.86-14.47%) variabilities; while accuracy displayed % nominal (low and high) of 153.31% and 90.71% respectively.

Hence, PRG4-GFG was measured in 38 sera from OA patients and HD with the validated singleplex iMALDI assay. The levels of PRG4-GFG in serum were around 67 fmol/µL or 10 µg/mL for both subgroups. However, no statistical significance was reached among OA and HD groups, meaning PRG4 is not likely to indicate the molecular diagnosis of OA when serum samples are accounted in the analysis. Furthermore, the same samples were measured by a commercial PRG4 ELISA, and again, no significant differences were found between OA and HD groups since the PRG4 serum levels for both conditions were around 2.5 µg/mL, about 4 times lower measurement values than in the developed iMALDI. Thus, it is likely that the developed iMALDI assay may have measured nearly the total PRG4 present in serum, while ELISA might have only detected the most accessible PRG4 protein, or either due to the fact that both antibodies in ELISA need to perform well simultaneously, while in iMALDI only the capture antibody is requested to yield satisfactory binding. Moreover, iMALDI and ELISA measurements were poorly correlated (r: 0.36-0.37). Some explanation for this difference could be the performance of the different antibodies used in each test, which could yield differential protein captures in the corresponding assays, as well as the use of a single antibody in iMALDI and two antibodies in the sandwich ELISA. Nevertheless, 34 additional serum samples (17 OA and 17 HD) were analyzed by ELISA, and the difference among OA and HD conditions was increased this time, although it did not involve significant outcomes yet (62, 191). This indicates the need to increase sample size to obtain established conclusions.

On the other hand, CILP1 peptides were analyzed in sera using a multiplex SISCAPA strategy. The examined NAT (serum) and SIS peptides from CILP1 could be satisfactory detected by singleplex analysis in this biological fluid, in the way that the corresponding transitions co-eluted quite adequately and orderly. Then, both CILP1 SISCAPA assays (CILP1-IVG and CILP1-TFL) were combined in order to obtain a final multiplex CILP1 SISCAPA trial. This set up was conducted during a month visit in Vall d'Hebron Hospital (Barcelona, Spain). Due to time constrictions,

it was not possible to perform further protocol optimization, variability testing and/or a broaden number of sera measurements. Nonetheless, both multiplex CILP1-IVG and CILP1-TFL SISCAPA assays displayed competent regression values (R^2=0.968 and R^2=0.995 respectively) when linear responses were examined. Due to the timing factor, each curve point could only be performed in duplicates. Consequently, some of the point CVs (%) were a bit higher than the generally accepted 20% for CILP1-TFL, as curve points are usually done in triplicates. In consonance with the time parameter, instrumentation is also a restricting issue when conducting SISCAPA experiments, as chromatography takes up more than 30 minutes per sample injected, while by iMALDI the individual measurements are in the range of seconds. In addition, chromatography comes with the need for including blanks between samples in order to avoid carry over from precedent samples, lengthening as well the duration of SISCAPA analysis.

CILP1 was measured in 12 sera from OA patients and HD using the validated multiplex SISCAPA assay. Remarkably, CILP1 levels were differentially measured when analyzing one or the other peptide. The average CILP1-IVG concentration in serum was around 10 fmol/µL while the average CILP1-TFL was approximately 57 fmol/µL serum. Furthermore, average CILP1-IVG concentration in serum from OA patients was higher than the one in HD (1.38 µg/mL vs 1.28 µg/mL serum respectively). Likely, average CILP1-TFL levels in serum were as well increased in OA vs HD (8.76 vs 6.19 µg/mL). Nonetheless, none of the differences were statistically significant, only CILP1-TFL (with a p-value= 0.132) showed a tendency to significance that could be demonstrated by incrementing the number of samples analyzed.

Even though the immuno-enrichment protocols are similar and performed with the same anti-peptide antibodies, the iMALDI showed to be not sensitive enough for detecting CILP1-IVG, while it could be measured by SISCAPA with an approximate 10 fmol/µL serum concentration. SISCAPA methods include

a chromatographic step which increases the selectivity and sensitivity since the separation of the analytes is carried out before the MS analysis, while by iMALDI no fractionation is conducted and low concentration measurements could be hampered by higher levels of other analytes. Notwithstanding, both iMALDI and SISCAPA methods could detect CILP1-TFL. SISCAPA measurements for CILP1-TFL displayed values around 57 fmol/µL serum, almost 6 times more than CILP1-IVG. This may be the reason why CILP1-TFL, at a higher concentration than CILP1-IVG, could be measured not only by SISCAPA but also by iMALDI. Other parameters to be accounted include the particular nature and MS behavior of the corresponding peptides when analyzed in different MS instruments with diverse ionization sources. However, due to the interfering MS peak at the same m/z, the quantification of CILP1-TFL by iMALDI could not be done, while by SISCAPA, peptides are chromatographically separated and fragmented for attaining a more sensitive and selective peptide identification.

Furthermore, immunogen contamination was observed and confirmed in both CILP1 IAMS assays. It was likely that traces of CILP1 peptides, used for the affinity-purification, came off from the column when the corresponding anti-peptide antibody was eluted and could then be detected by IAMS (192). Fortunately, this interference could be successfully removed by performing a size-ultrafiltration (UF), aided by a preliminary acidic elution of the peptide from the antibody, since both molecular sizes are extreme (1kDa vs 150kDa, approximately). Interestingly, once the contaminant peptides were removed from the antibodies their binding capacity was also boosted, as reflected in the SIS peptide measurement by both IAMS methodologies (**Figure 3. 6** and **Figure 3. 21**). Therefore, the generated polyclonal antibodies seem to be quite robust since they put up with several extreme pH adjustments during their handling. These radical pH transitions could have provoked conformational changes in the antibody, destroyed the corresponding antigen complementarity and finally led to dysfunctionality (193), although in this case pH was not a major effecting factor.

Neither CILP1 nor PRG4 seemed to statistically discriminate between patients with OA and healthy donors, although their tendency was to be increased in serum from OA patients. On the one side, PRG4 expression was significantly elevated in synovial fluid (SF) from OA patients (150.81 g/mL) compared with HD (34.82 g/mL), as well as in the previous ELISA outcomes in serum (194). Likewise, PRG4 concentrations were significantly elevated in SF from horses naturally suffering OA (152µg/mL) vs healthy horses (68 µg/mL) (195). Taken together, these results suggest that PRG4 may be upregulated as part of the endogenous repair response to protect joints from OA affected cartilages. In contrast to this trend, PRG4 SF levels were found to be around 287.1 µg/mL in HD and 146.5 in OA patients (196), approximately 15-29 times more concentrated than in serum (10 µg/mL). This opposite tendency in SF could indicate a decrease in cartilage boundary–lubricating ability when OA is present, due to different possible mechanisms as lowered expression/synthesis of PRG4, increased degradation of PRG4, or increased loss of PRG4 from the joint capsule through an inflamed synovium (197). The principal mechanism proposed for the reduction of PRG4 synthesis is the degradation activity of neutrophil-derived enzymes and inflammatory mediators present in post-traumatic SF. On the other side, CILP1 expression was also found to be increased in OA cartilages vs normal cartilages with fold changes around 2-4 depending on the OA severity (198), possibly as an attempt to repair/replace the extracellular matrix (199).

Until the date, CILP1 and PRG4 levels in serum have not been measured together. Although they could not be validated for OA monitoring yet, both IAMS assays are ready to be used for this or other aims. The developed methods could be used for performing further studies in a larger and more defined cohort of samples (in order to qualify them for an specific clinical use) or either for monitoring other rheumatic diseases such as rheumatoid arthritis. Moreover,

they could also serve as alternative methods for analyzing CILP1 and PRG4 in SF and cartilage samples in a more accurate and sensible way.

IV. General discussion

The search of new molecular biomarkers for disease diagnosis, prognosis, monitoring and prediction of treatment response is currently increasing and has become a priority in clinical research. This is mainly due to the fact that several pathologies cannot be diagnosed in time, provoking irreparable damage in patients (as in OA and RA) and difficulties at evaluating novel effective treatments. Within the field of biomarkers, proteomics has demonstrated to be a powerful discipline due to its ability to identify proteins with potential disease biomarker value. Therefore, numerous proteomic strategies have been developed with the final aim of characterizing proteins as potential disease biomarkers for diagnosis, patient stratification and development of personalized therapies. This thesis has been focused on the search for potential OA and RA biomarkers by combining different proteomic strategies centered on mass spectrometry and immuno-based techniques.

On the one hand, RA is a complex rheumatic disease since its pathogenesis is highly related with an aberrant autoimmunity and inflammatory process, therefore its consideration as a rheumatic IMID. Particularly, RA patients suffer many flares or periods of increased disease activity during which the symptoms are more severe. One of the main needs for monitoring these disease after diagnosis is to find sensitive and specific biomarkers for monitoring the disease activity at the molecular level, in order to try to maintain the patient in remission conditions using personalized therapies (TTT). Therefore, many researchers and companies are putting efforts into finding biomarker panels for effectively measuring the activity of the disease. Hence, in this thesis, the previously described conventional biomarker discovery pipeline has been carried out in order to find a panel of potential disease activity biomarkers in plasma from RA patients, which could help to monitor RA patients and select the most suitable therapies in clinics for mitigating the severe symptoms during the mentioned activity flares.

The conventional biomarker discovery pipeline is mainly composed of 4 stages (discovery, qualification, verification and validation), although the qualification and verification phases could be performed simultaneously since they both proof the results obtained in the previous discovery. In order to find biomarkers of disease activity in plasma from RA patients, the 3-stage approach was followed, including discovery, verification and validation phases. In the initial discovery phase, 186 proteins were identified by nanoLC-MS/MS and 11 of them were found to be significantly altered between extreme RA disease activity conditions using the iTRAQ strategy. Then, these 11 biomarker candidates were verified by nanoLC-MRM with the aid of SIS peptides for absolute quantitation and 5 peptides from SAA1, AACT, HPT, A1AG1 and A1AG2 were found to be significantly increased in RA patients with high disease activity. Finally, SAA1, AACT, HPT and A1AG1 proteins were validated by individual ELISA in a larger cohort of samples from RA patients, HD and other rheumatic IMID patients (SLE and PsA). The results showed a significant increase of these four proteins in RA patients with high activity, as well as a differential abundance in patients with any rheumatic IMID and HD, therefore corroborating their value as potential biomarkers for RA activity monitoring and rheumatic IMID diagnosis and classification. Moreover, AACT also distinguishes between SLE patients with extreme disease activities, while HPT and SAA1 significantly differ among PsA patients with low and high disease activity.

Altogether, the biomarker search outline followed comprised: 1) an initial discovery of biomarker candidates by relative quantitation using the iTRAQ labeling strategy and a shotgun proteomic approach, which allow to measure a low number of samples but many analytes at the same time while offering differential protein abundance results; 2) a subsequent verification of the differential abundance of the previously discovered biomarker candidates by targeted proteomics using a developed MRM method and SIS peptides for absolute and simultaneous quantitation of these candidates, which tendency

found in the initial discovery phase was confirmed in this case in more samples; 3) a final validation of the verified biomarker candidates by singleplex ELISA, the gold standard for clinical biomarker validation, which allows to measure many samples at the same time but does not have the sufficient multiplexing capability for measuring all at the same time. This phase also included a prior technical validation for confirming the suitability of the immunoassay for the analysis of the verified tryptic peptides and therefore proteins that were differentially detected by MS, important for an effective analysis of the same verified biomarkers when the technical format used is changed. In this case, the correlation between the protein measurements by the targeted MRM method (MS) and the corresponding ELISA were all moderate-to-high, except for one of the proteins (A1AG2) since the ELISA immunogen sequences of the antibodies did not present the tryptic peptide that yielded significant outcomes by MRM.

In conclusion, a panel of 4 proteins (SAA1, AACT, HPT, A1AG1) were found as potential biomarkers for RA activity monitoring and rheumatic IMID diagnosis and classification in a cohort of 420 plasma samples, using a 3-stage biomarker discovery approach which included discovery, verification and validation phases. Besides, AACT was as well validated as a promising biomarker for SLE activity monitoring, whereas HPT and SAA1 were confirmed as markers of disease activity monitoring in PsA patients.

On the other hand, OA is the most prevalent rheumatic disease in society and it is clinically silent in its initial stages. Therefore, when the patient comes to the consultation with any symptoms, the disease is already quite advanced causing the impossible recovering of the regular individual functionality. Besides, there is no effective therapy against the disease and most of them are based on pain and inflammation relief. Since OA is mainly characterized by articular cartilage degradation, the finding of specific markers involved in these processes would help to develop disease diagnostic and monitoring methods. Particularly,

in this thesis, two different approaches have been applied for OA biomarker discovery: an extensive peptidomic study on cartilage secretomes, synovial fluid and serum samples and the development of two immuno-affinity mass spectrometry methods for the detection of two cartilage-specific proteins in serum.

Thus, a peptidomic approach has been performed in order to better understand the complex molecular interactions that lead to cartilage ECM deterioration. Specific neopeptides may be generated under the action of particular molecular processes and proteases, likely causing the existence or progression of a certain disease. In this study, 8 neopeptides from PRELP, MGP, CILP1, CLUS and COMP were differentially detected in healthy and OA cartilage secretomes. Differences between the neopeptide content were found to be related with the target joint, with a higher neopeptide release in the secretomes obtained from knee tissue samples. Particularly, CLUS and COMP neopeptides were found to be significantly altered in hip secretomes, while CILP1 and PRELP neopeptide abundance varied in knee secretomes, suggesting a joint-dependent neopeptide release in terms of OA progression. Notably, one of the PRELP neopeptides (DSNKIETIPN) has demonstrated to hold an excellent biomarker value (AUC of 0.891 in ROC curve statistics) when comparing healthy tissue secretomes with the ones from the wounded and unwounded OA tissue zones. Although these neopeptides were satisfactorily detected by MRM in cartilage secretome samples, they could not be detected in synovial fluid or serum samples, after testing several peptidome isolation strategies and using the cited MRM method.

Therefore, a new discovery phase was performed for OA-related neopeptide detection in these body fluid specimens. The combination of an initial protein precipitation with acetonitrile and a subsequent centrifugal ultrafiltration was found to be the most effective method for neopeptide isolation in serum.

After using this method for the detection of neopeptides in serum samples, 6 peptides from KNG1, APOA4, CO3 and ITIH4 were found to be altered in pooled sera from OA patients and HD. Altogether, the cited 14 neopeptide sequences (8 neopeptides detected in cartilage secretomes and 6 neopeptides found in sera) were searched by the PROSPER tool, which identifies the likely-acting endogenous enzymes. OA-related proteases such as MMP-9, MMP-3, elastase-2 and cathepsin G were found to be possibly acting over most of the mentioned neopeptides, therefore boosting their value as candidate markers for OA diagnosis and monitoring. However, more extensive analysis should be performed in order to confirm their biomarker role.

The 8 neopeptides found in cartilage secretomes - mostly coming from cartilage-associated proteins such as MGP, CILP1 or COMP- could not be detected in sera or synovial fluid after extensive peptidome isolation and analysis by conventional LC-MS/MS. Therefore, these neopeptides need to be validated in more clinically-accessible samples such as synovial fluid, serum or plasma, for instance using IA-MS strategies. These hybrid techniques aid to enhance sensitivity and specificity conditions for detecting low abundant molecules such as these particular neopeptides in body fluids, which are typically in less than ng/mL concentration ranges. Furthermore, the trend found in the 6 neopeptides detected in sera needs to be further examined in independent and more numerous serum samples, in order to perform extensive statistical analysis for confirming the found neopeptide tendency and use them in clinics. Although they do not appertain to cartilage-related proteins, their release could still be linked to the OA progression since the proteases involved in the OA process could be acting over these common proteins presented in the surrounding cells and systems.

Finally, and following an alternative approach for the measurement of cartilage-related proteins, two hybrid IA-MS strategies were developed for the detection of low abundant tissue-specific proteins in complex mixtures such as

serum. In this case, CILP1 and PRG4 proteins, which are present in intermediate and superficial cartilage zones respectively, were detected in serum by iMALDI or SISCPA-MRM IA-MS techniques with the aid of SIS peptides and anti-peptide antibodies. On the one hand, the iMALDI was developed for quantifying only PRG4 while multiplexed SISCAPA measurements were conducted for the quantification of the intermediate and less accessible CILP1 protein. It is important to notice that within these two IA-MS methods, iMALDI is likely to be the most applicable one for clinical validation since it does not require any chromatographic stage as in SISCAPA (200). Chromatography, and specially nanoLC, is normally one of the most tedious obstacles for clinical validation due to specialized personal and equipment requirements. However, in this particular case, iMALDI could not be performed as a multiplex method since many detection issues aroused in the analysis of the selected CILP1 peptides, considering that they were initially designated for SISCAPA-MRM analysis. Seeing that iMALDI does not involve any previous fractionation, it is crucial that the antibodies present the required specificity for avoiding the bind of untargeted molecules since those could hamper the MS analysis making it difficult to differentiate the target peptide peaks in the MS spectrum. Particularly, many interfering peaks were visualized when trying to measure the selected CILP1 light or heavy peptides, therefore precluding to perform a satisfactory quantification of the target analytes. This situation was solved when SISCAPA-MRM analysis were performed due to the fact that nanoLC fractionation involves the separation the target peptides during the chromatographic run and the targeted MRM analysis facilitates the detection and identification of specific peptides, since MS/MS measurements are also performed.

Additionally, PRG4 levels were also measured by a commercial ELISA for confirming the validity of the developed PRG4 iMALDI assay. From this comparison, it is important to note that higher values for the same samples measured were obtained when using the iMALDI method (around 4 times higher).

This could be related with many factors such as the Hook effect that normally affects sandwich immunoassays, where higher antigen concentrations could saturate the antibodies causing underestimation of the antigen levels. Another factor for the higher PRG4 iMALDI values could be that nearly the total quantitation of the antibody-captured PRG4 peptide was conducted by iMALDI, whereas the total PRG4 protein may not be as accessible for ELISA capture/detection antibodies. In line with this, two antibodies are used in ELISA and both are required to simultaneously bind to the target protein, while in iMALDI, only one antibody is needed for efficient protein detection.

Although no statistical differences were found when CILP1 and PRG4 were analyzed in serum samples from OA patients and HD, an increased abundance trend of both cartilage-related proteins was found in serum from OA patients (p-value < 0.200). This suggests the loss of cartilage ECM proteins in cartilage of OA patients, hallmark of the disease process, which are therefore increased in the more or less proximal biological fluids. However, this tendency should be further examined in a greater number of samples in order to confirm the protein biomarker value and method validity for clinical diagnosis or monitoring.

Altogether, the work performed in this Thesis has resulted in obtaining a collection of methodologies for the analysis of proteins related with the molecular processes in different rheumatic diseases, and identifying biomarker candidates to facilitate the RA and OA management. Although in some cases the small sample size employed to validate these techniques has hindered the obtention of conclusive results, the developed methods will be undoubtedly useful for performing further large-scale analyses on clinical cohorts and therefore enabling their qualification for further specific clinical uses.

V. Conclusions

According to the chapters presented and discussed in this thesis, the following final conclusions were established per chapter:

1. A three-stage proteomic approach has been performed, following the conventional biomarker pipeline, and a panel of four proteins was finally validated as potential plasma biomarkers for activity monitoring of three rheumatic IMID (RA, SLE and PsA).

1.1. The initial discovery phase analysis lead to the identification of 199 proteins, out of which 11 were found to be significantly altered between RA patients with extreme disease activities.

1.2. In the subsequent verification phase, an MRM method was developed for the absolute quantitation of the 11 biomarker candidates for RA disease activity monitoring, detecting 26 proteotypic peptides. By these means, five peptides from SAA1, HPT, AACT, A1AG1 and A1AG2 were found to be significantly increased in RA patients with high disease activity.

1.3. In the last validation phase, SAA1, HPT, AACT and A1AG1 were measured by ELISAs in independent plasma samples from RA patients and healthy donors, as well as patients with other rheumatic IMIDs such as SLE and PsA. The levels of these four proteins were significantly higher in RA patients with high activity, suggesting the potential value of these proteins as biomarker candidates for RA disease activity monitoring, as well as for rheumatic IMID classification.

2. A peptidomic study was performed on human cartilage secretomes, synovial fluid and serum samples, using shotgun and targeted MS strategies for the discovery and absolute quantitation of neopeptides associated with the proteases involved in the degradation of human articular cartilage in OA.

2.1. In the initial discovery phase, performed on healthy and OA cartilage secretomes by shotgun MS, a total of 1,175 neopeptides from 101 proteins were identified.

2.2. An MRM method was developed for absolute neopeptide quantitation. This targeted method measured 23 neopeptides from 9 proteins, and detected eight neopeptides from PRELP, MGP, CILP1, CLUS and COMP as potential marker candidates for cartilage turnover monitoring in OA.

2.3. In a parallel discovery phase on synovial fluid and serum samples, the combination of protein precipitation with ACN and subsequent ultrafiltration using 10 kDA MWCO membrane filters was found to be the most effective technique for neopeptide isolation.

2.4. Using the selected neopeptide isolation method, more than 1000 serum/synovial fluid neopeptides were identified by LC-MS/MS. From them, 21 neopeptides corresponding to 16 proteins were finally assembled in MRM method, out of which six (from APOA4, ITIH4, CO3 and KNG1) showed differential abundances in pooled sera from OA patients and HD. However, a higher number of samples should be analyzed to prove their biomarker value for OA monitoring in serum.

3. Two IA-MS strategies (iMALDI and SISCAPA) were developed for the detection and absolute quantitation of two cartilage-related proteins (CILP1 and PRG4) in synovial fluid and serum samples, using specific anti-peptide antibodies and SIS peptides.

3.1. An iMALDI protocol was set up for the analysis of PRG4 in sera with good response curve regression (R^2: 0.992) and total variability values lower than 15%.

3.2. The developed iMALDI assay showed no significant differences between OA and HD samples.

3.3. A multiplexed SISCAPA assay was developed for the absolute quantitation of two tryptic CILP1 peptides. CILP1-TFL levels were slightly increased in OA vs HD (8.76 vs 6.19 µg/mL, p-value: 0.132).

VI. Bibliography

1. OpenStax. Anatomy and Physiology: OpenStax; 2016.
2. Sophia Fox AJ, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. Sports Health. 2009;1(6):461-8.
3. Carballo CB, Nakagawa Y, Sekiya I, Rodeo SA. Basic Science of Articular Cartilage. Clin Sports Med. 2017;36(3):413-25.
4. Di Bella C, Fosang A, Donati DM, Wallace GG, Choong PF. 3D Bioprinting of Cartilage for Orthopedic Surgeons: Reading between the Lines. Front Surg. 2015;2:39.
5. Ospelt C, Frank-Bertoncelj M. Why location matters - site-specific factors in rheumatic diseases. Nat Rev Rheumatol. 2017;13(7):433-42.
6. Al Maini M, Adelowo F, Al Saleh J, Al Weshahi Y, Burmester GR, Cutolo M, et al. The global challenges and opportunities in the practice of rheumatology: white paper by the World Forum on Rheumatic and Musculoskeletal Diseases. Clin Rheumatol. 2015;34(5):819-29.
7. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: A systematic analysis of the World Health Organization Burden of Diseases Database2019. annrheumdis-2019 p.
8. Hunter DJ, Felson DT. Osteoarthritis. BMJ. 2006;332(7542):639-42.
9. Man GS, Mologhianu G. Osteoarthritis pathogenesis - a complex process that involves the entire joint. J Med Life. 2014;7(1):37-41.
10. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet. 2019;393(10182):1745-59.
11. Chen D, Shen J, Zhao W, Wang T, Han L, Hamilton JL, et al. Osteoarthritis: toward a comprehensive understanding of pathological mechanism. Bone Res. 2017;5:16044.
12. Judith Farley VCDaJSM. Proteases and Cartilage Degradation in Osteoarthritis. Principles of Osteoarthritis - Its Definition, Character, Derivation and Modality-Related Recognition: IntechOpen; 2012.
13. Troeberg L, Nagase H. Proteases involved in cartilage matrix degradation in osteoarthritis. Biochim Biophys Acta. 2012;1824(1):133-45.
14. Verma P, Dalal K. ADAMTS-4 and ADAMTS-5: key enzymes in osteoarthritis. J Cell Biochem. 2011;112(12):3507-14.
15. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med. 2010;26(3):355-69.
16. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol. 2014;10(7):437-41.
17. Allen KD, Golightly YM. State of the evidence. Curr Opin Rheumatol. 2015;27(3):276-83.
18. Altman RD, Gold GE. Atlas of individual radiographic features in

osteoarthritis, revised. Osteoarthritis Cartilage. 2007;15 Suppl A:A1-56.
19. Sinusas K. Osteoarthritis: diagnosis and treatment. Am Fam Physician. 2012;85(1):49-56.
20. KELLGREN JH, LAWRENCE JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494-502.
21. Guermazi A, Hayashi D, Roemer FW, Felson DT. Osteoarthritis: a review of strengths and weaknesses of different imaging options. Rheum Dis Clin North Am. 2013;39(3):567-91.
22. Ding C, Zhang Y, Hunter D. Use of imaging techniques to predict progression in osteoarthritis. Curr Opin Rheumatol. 2013;25(1):127-35.
23. Rousseau JC, Delmas PD. Biological markers in osteoarthritis. Nat Clin Pract Rheumatol. 2007;3(6):346-56.
24. Calabresi E, Petrelli F, Bonifacio AF, Puxeddu I, Alunno A. One year in review 2018: pathogenesis of rheumatoid arthritis. Clin Exp Rheumatol. 2018;36(2):175-84.
25. Karami J, Aslani S, Jamshidi A, Garshasbi M, Mahmoudi M. Genetic implications in the pathogenesis of rheumatoid arthritis; an updated review. Gene. 2019;702:8-16.
26. Berman S, Bucher J, Koyfman A, Long BJ. Emergent Complications of Rheumatoid Arthritis. J Emerg Med. 2018;55(5):647-58.
27. Gerlag DM, Raza K, van Baarsen LGM, Brouwer E, Buckley CD, Burmester GR, et al. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. Annals of the Rheumatic Diseases. 2012;71(5):638.
28. Mankia K, Emery P. Preclinical Rheumatoid Arthritis: Progress Toward Prevention. Arthritis Rheumatol. 2016;68(4):779-88.
29. Croia C, Bursi R, Sutera D, Petrelli F, Alunno A, Puxeddu I. One year in review 2019: pathogenesis of rheumatoid arthritis. Clin Exp Rheumatol. 2019;37(3):347-57.
30. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet. 2016;388(10055):2023-38.
31. Wang Q, Sun X. Recent advances in nanomedicines for the treatment of rheumatoid arthritis. Biomater Sci. 2017;5(8):1407-20.
32. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the Global Burden of Disease 2010 study. Annals of the Rheumatic Diseases. 2014;73(7):1316.
33. Sokka T, Kautiainen H, Pincus T, Verstappen SM, Aggarwal A, Alten R, et al. Work disability remains a major problem in rheumatoid arthritis in the 2000s: data from 32 countries in the QUEST-RA study. Arthritis Res Ther. 2010;12(2):R42.

34. Deane KD, Demoruelle MK, Kelmenson LB, Kuhn KA, Norris JM, Holers VM. Genetic and environmental risk factors for rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2017;31(1):3-18.

35. Falkenburg WJJ, van Schaardenburg D. Evolution of autoantibody responses in individuals at risk of rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2017;31(1):42-52.

36. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2010;69(9):1580-8.

37. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Ann Rheum Dis. 2011;70(3):404-13.

38. Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis: A Review. JAMA. 2018;320(13):1360-72.

39. Gibofsky A. Epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis: A Synopsis. Am J Manag Care. 2014;20(7 Suppl):S128-35.

40. Gibofsky A. Overview of epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis. Am J Manag Care. 2012;18(13 Suppl):S295-302.

41. Oderda GM, Lawless GD, Wright GC, Nussbaum SR, Elder R, Kim K, et al. The potential impact of monitoring disease activity biomarkers on rheumatoid arthritis outcomes and costs. Per Med. 2018;15(4):291-301.

42. Littlejohn EA, Monrad SU. Early Diagnosis and Treatment of Rheumatoid Arthritis. Prim Care. 2018;45(2):237-55.

43. Williams JP, Meyers JA. Immune-mediated inflammatory disorders (I.M.I.D.s): the economic and clinical costs. Am J Manag Care. 2002;8(21 Suppl):S664-81; quiz S82-5.

44. Puig L, Ruiz de Morales JG, Dauden E, Andreu JL, Cervera R, Adán A, et al. [Prevalence of ten Immune-mediated inflammatory diseases (IMID) in Spain]. Rev Esp Salud Publica. 2019;93.

45. El-Gabalawy H, Guenther LC, Bernstein CN. Epidemiology of immune-mediated inflammatory diseases: incidence, prevalence, natural history, and comorbidities. J Rheumatol Suppl. 2010;85:2-10.

46. Santos M, Cordeiro A, M. Gil V. Immune-Mediated Inflammatory Rheumatic Diseases2015. 113-32 p.

47. Rahman P, Inman RD, El-Gabalawy H, Krause DO. Pathophysiology and pathogenesis of immune-mediated inflammatory diseases: commonalities and differences. J Rheumatol Suppl. 2010;85:11-26.

48. Schirmer M, Duftner C, Dejaco C. Challenges in the diagnosis of chronic immune-mediated rheumatic diseases. Discov Med. 2013;15(82):160-5.
49. Kuek A, Hazleman BL, Ostör AJ. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. Postgrad Med J. 2007;83(978):251-60.
50. Baker KF, Isaacs JD. Novel therapies for immune-mediated inflammatory diseases: What can we learn from their use in rheumatoid arthritis, spondyloarthritis, systemic lupus erythematosus, psoriasis, Crohn's disease and ulcerative colitis? Ann Rheum Dis. 2018;77(2):175-87.
51. R. Wilkins M, Sanchez J-C, Gooley A, Appel R, Humphery-Smith I, F. Hochstrasser D, et al. Progress with Proteome Projects: Why All Proteins Expressed by a Genome Should Be Identified and How To Do It1996. 19-50 p.
52. James P. Protein identification in the post-genome era: the rapid rise of proteomics. Q Rev Biophys. 1997;30(4):279-331.
53. CRICK FH. On protein synthesis. Symp Soc Exp Biol. 1958;12:138-63.
54. Lee CC, Bowman BH, Yang FM. Human alpha 2-HS-glycoprotein: the A and B chains with a connecting sequence are encoded by a single mRNA transcript. Proc Natl Acad Sci U S A. 1987;84(13):4403-7.
55. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, et al. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis. 1995;16(7):1090-4.
56. Graves P, Haystead T. Molecular Biologist's Guide to Proteomics2002. 39-63; table of contents p.
57. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
58. A gene-centric human proteome project: HUPO--the Human Proteome organization. Mol Cell Proteomics. 2010;9(2):427-9.
59. Lau AT, He QY, Chiu JF. Proteomic technology and its biomedical applications. Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai). 2003;35(11):965-75.
60. Jain KK. The handbook of biomarkers. Totowa, N.J. ; London: Humana; 2010.
61. Bauer DC, Hunter DJ, Abramson SB, Attur M, Corr M, Felson D, et al. Classification of osteoarthritis biomarkers: a proposed approach. Osteoarthritis Cartilage. 2006;14(8):723-7.
62. Parker CE, Borchers CH. Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. Mol Oncol. 2014;8(4):840-58.
63. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation:

the long and uncertain path to clinical utility. Nat Biotechnol. 2006;24(8):971-83.
64. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R. On the development of plasma protein biomarkers. J Proteome Res. 2011;10(1):5-16.
65. Ruiz-Romero C, Blanco FJ. Proteomics role in the search for improved diagnosis, prognosis and treatment of osteoarthritis. Osteoarthritis Cartilage. 2010;18(4):500-9.
66. Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol. 2009;3(1):33-44.
67. Bose U, Wijffels G, Howitt CA, Colgrave ML. Proteomics: Tools of the Trade. Adv Exp Med Biol. 2019;1073:1-22.
68. Bellei E, Bergamini S, Monari E, Fantoni LI, Cuoghi A, Ozben T, et al. High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins. Amino Acids. 2011;40(1):145-56.
69. Jaros J, Guest P, Bahn S, Martins-de-Souza D. Affinity Depletion of Plasma and Serum for Mass Spectrometry-Based Proteome Analysis2013.
70. Millioni R, Tolin S, Puricelli L, Sbrignadello S, Fadini GP, Tessari P, et al. High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. PLoS One. 2011;6(5):e19603.
71. Greening DW, Simpson RJ. Characterization of the Low-Molecular-Weight Human Plasma Peptidome. Methods Mol Biol. 2017;1619:63-79.
72. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics. 2003;2(10):1096-103.
73. Colantonio DA, Dunkinson C, Bovenkamp DE, Van Eyk JE. Effective removal of albumin from serum. Proteomics. 2005;5(15):3831-5.
74. Polaskova V, Kapur A, Khan A, Molloy MP, Baker MS. High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. Electrophoresis. 2010;31(3):471-82.
75. Kovàcs A, Guttman A. Medicinal chemistry meets proteomics: fractionation of the human plasma proteome. Curr Med Chem. 2013;20(4):483-90.
76. Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. Proteomics. 2005;5(13):3292-303.
77. Feist P, Hummon AB. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. Int J Mol Sci. 2015;16(2):3537-63.

78. Wells DA. Bioanalytical Applications: Solid-Phase Extraction. 2013.
79. Chutipongtanate S, Chatchen S, Svasti J. Plasma prefractionation methods for proteomic analysis and perspectives in clinical applications. Proteomics Clin Appl. 2017;11(7-8).
80. Kaboord B, Perr M. Isolation of proteins and protein complexes by immunoprecipitation. Methods Mol Biol. 2008;424:349-64.
81. Uljon SN, Mazzarelli L, Chait BT, Wang R. Analysis of proteins and peptides directly from biological fluids by immunoprecipitation/mass spectrometry. Methods Mol Biol. 2000;146:439-52.
82. Valcu C-M, Valcu M. Reproducibility of Two-Dimensional Gel Electrophoresis at Different Replication Levels. Journal of Proteome Research. 2007;6(12):4677-83.
83. Zhang G, Annan RS, Carr SA, Neubert TA. Overview of peptide and protein analysis by mass spectrometry. Curr Protoc Protein Sci. 2010;Chapter 16:Unit16.1.
84. Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. Nat Rev Drug Discov. 2003;2(2):140-50.
85. Wang H, Shi T, Qian WJ, Liu T, Kagan J, Srivastava S, et al. The clinical impact of recent advances in LC-MS for cancer biomarker discovery and verification. Expert Rev Proteomics. 2016;13(1):99-114.
86. Lai ZW, Petrera A, Schilling O. The emerging role of the peptidome in biomarker discovery and degradome profiling. Biol Chem. 2015;396(3):185-92.
87. Dallas DC, Guerrero A, Parker EA, Robinson RC, Gan J, German JB, et al. Current peptidomics: applications, purification, identification, quantification, and functional analysis. Proteomics. 2015;15(5-6):1026-38.
88. Ackermann BL. Understanding the role of immunoaffinity-based mass spectrometry methods for clinical applications. Clin Chem. 2012;58(12):1620-2.
89. Ackermann BL. Hybrid immunoaffinity--mass spectrometric methods for efficient protein biomarker verification in pharmaceutical development. Bioanalysis. 2009;1(2):265-8.
90. Ackermann BL, Berna MJ. Coupling immunoaffinity techniques with MS for quantitative analysis of low-abundance protein biomarkers. Expert Rev Proteomics. 2007;4(2):175-86.
91. Centola M, Cavet G, Shen Y, Ramanujan S, Knowlton N, Swan KA, et al. Development of a multi-biomarker disease activity test for rheumatoid arthritis. PLoS One. 2013;8(4):e60635.
92. Fu Q, Bovenkamp DE, Van Eyk JE. A rapid, economical, and reproducible method for human serum delipidation and albumin and IgG removal for proteomic analysis. Methods Mol Biol. 2007;357:365-71.
93. Chen YY, Lin SY, Yeh YY, Hsiao HH, Wu CY, Chen ST, et al. A modified protein

precipitation procedure for efficient removal of albumin from serum. Electrophoresis. 2005;26(11):2117-27.

94. Omenn GS. THE HUPO Human Plasma Proteome Project. Proteomics Clin Appl. 2007;1(8):769-79.

95. Leatherbarrow RJ, Dean PD. Studies on the mechanism of binding of serum albumins to immobilized cibacron blue F3G A. Biochem J. 1980;189(1):27-34.

96. Kassab A, Yavuz H, Odabaşi M, Denizli A. Human serum albumin chromatography by Cibacron Blue F3GA-derived microporous polyamide hollow-fiber affinity membranes. J Chromatogr B Biomed Sci Appl. 2000;746(2):123-32.

97. Altintaş EB, Denizli A. Efficient removal of albumin from human serum by monosize dye-affinity beads. J Chromatogr B Analyt Technol Biomed Life Sci. 2006;832(2):216-23.

98. Gianazza E, Miller I, Palazzolo L, Parravicini C, Eberini I. With or without you - Proteomics with or without major plasma/serum proteins. J Proteomics. 2016;140:62-80.

99. Jerkovic L, Voegele AF, Chwatal S, Kronenberg F, Radcliffe CM, Wormald MR, et al. Afamin is a novel human vitamin E-binding glycoprotein characterization and in vitro expression. J Proteome Res. 2005;4(3):889-99.

100. Li J, Jin D, Fu S, Mei G, Zhou J, Lei L, et al. Insulin-like growth factor binding protein-3 modulates osteoblast differentiation via interaction with vitamin D receptor. Biochem Biophys Res Commun. 2013;436(4):632-7.

101. Method of the Year 2012. Nature Methods. 2012;10:1.

102. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol. 2005;1(5):252-62.

103. Mohammed Y, Percy AJ, Chambers AG, Borchers CH. Qualis-SIS: automated standard curve generation and quality assessment for multiplexed targeted quantitative proteomic experiments with labeled standards. J Proteome Res. 2015;14(2):1137-46.

104. Percy AJ, Chambers AG, Yang J, Borchers CH. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. Proteomics. 2013;13(14):2202-15.

105. Percy AJ, Yang J, Hardie DB, Chambers AG, Tamura-Wells J, Borchers CH. Precise quantitation of 136 urinary proteins by LC/MRM-MS using stable isotope labeled peptides as internal standards for biomarker discovery and/or verification studies. Methods. 2015;81:24-33.

106. LeBlanc A, Michaud SA, Percy AJ, Hardie DB, Yang J, Sinclair NJ, et al. Multiplexed MRM-Based Protein Quantitation Using Two Different Stable Isotope-Labeled Peptide Isotopologues for Calibration. J Proteome Res. 2017;16(7):2527-

36.
107. Rezeli M, Sjödin K, Lindberg H, Gidlöf O, Lindahl B, Jernberg T, et al. Quantitation of 87 Proteins by nLC-MRM/MS in Human Plasma: Workflow for Large-Scale Analysis of Biobank Samples. J Proteome Res. 2017;16(9):3242-54.
108. Percy AJ, Chambers AG, Yang J, Hardie DB, Borchers CH. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. Biochim Biophys Acta. 2014;1844(5):917-26.
109. Percy AJ, Mohammed Y, Yang J, Borchers CH. A standardized kit for automated quantitative assessment of candidate protein biomarkers in human plasma. Bioanalysis. 2015;7(23):2991-3004.
110. Haab BB, Paulovich AG, Anderson NL, Clark AM, Downing GJ, Hermjakob H, et al. A reagent resource to identify proteins and peptides of interest for the cancer community: a workshop report. Mol Cell Proteomics. 2006;5(10):1996-2007.
111. Tate J, Ward G. Interferences in immunoassay. Clin Biochem Rev. 2004;25(2):105-20.
112. Xu L, Yu Z, Fan W, Wang X, Xie M, Xu Y, et al. Negative interference in serum HBsAg ELISA from rheumatoid factors. PLoS One. 2013;8(11):e80620.
113. Güven E, Duus K, Lydolph MC, Jørgensen CS, Laursen I, Houen G. Non-specific binding in solid phase immunoassays for autoantibodies correlates with inflammation markers. J Immunol Methods. 2014;403(1-2):26-36.
114. Calvin J, Neale G, Fotherby KJ, Price CP. The relative merits of acute phase proteins in the recognition of inflammatory conditions. Ann Clin Biochem. 1988;25 (Pt 1):60-6.
115. Chard MD, Calvin J, Price CP, Cawston TE, Hazleman BL. Serum alpha 1 antichymotrypsin concentration as a marker of disease activity in rheumatoid arthritis. Ann Rheum Dis. 1988;47(8):665-71.
116. Yildirim K, Karatay S, Melikoglu MA, Gureser G, Ugur M, Senel K. Associations between acute phase reactant levels and disease activity score (DAS28) in patients with rheumatoid arthritis. Ann Clin Lab Sci. 2004;34(4):423-6.
117. Kang MJ, Park YJ, You S, Yoo SA, Choi S, Kim DH, et al. Urinary proteome profile predictive of disease activity in rheumatoid arthritis. J Proteome Res. 2014;13(11):5206-17.
118. Park YJ, Yoo SA, Hwang D, Cho CS, Kim WU. Identification of novel urinary biomarkers for assessing disease activity and prognosis of rheumatoid arthritis. Exp Mol Med. 2016;48:e211.
119. Targońska-Stępniak B, Majdan M. Serum amyloid A as a marker of persistent inflammation and an indicator of cardiovascular and renal involvement in patients with rheumatoid arthritis. Mediators Inflamm. 2014;2014:793628.

120. Cunnane G, Grehan S, Geoghegan S, McCormack C, Shields D, Whitehead AS, et al. Serum amyloid A in the assessment of early inflammatory arthritis. J Rheumatol. 2000;27(1):58-63.
121. Shen C, Sun XG, Liu N, Mu Y, Hong CC, Wei W, et al. Increased serum amyloid A and its association with autoantibodies, acute phase reactants and disease activity in patients with rheumatoid arthritis. Mol Med Rep. 2015;11(2):1528-34.
122. Hwang YG, Balasubramani GK, Metes ID, Levesque MC, Bridges SL, Moreland LW. Differential response of serum amyloid A to different therapies in early rheumatoid arthritis and its potential value as a disease activity biomarker. Arthritis Res Ther. 2016;18(1):108.
123. Connolly M, Mullan RH, McCormick J, Matthews C, Sullivan O, Kennedy A, et al. Acute-phase serum amyloid A regulates tumor necrosis factor α and matrix turnover and predicts disease progression in patients with inflammatory arthritis before and after biologic therapy. Arthritis Rheum. 2012;64(4):1035-45.
124. Guilak F, Nims RJ, Dicks A, Wu CL, Meulenbelt I. Osteoarthritis as a disease of the cartilage pericellular matrix. Matrix Biol. 2018;71-72:40-50.
125. Bay-Jensen AC, Hoegh-Madsen S, Dam E, Henriksen K, Sondergaard BC, Pastoureau P, et al. Which elements are involved in reversible and irreversible cartilage degradation in osteoarthritis? Rheumatol Int. 2010;30(4):435-42.
126. Huebner JL, Bay-Jensen AC, Huffman KM, He Y, Leeming DJ, McDaniel GE, et al. Alpha C-telopeptide of type I collagen is associated with subchondral bone turnover and predicts progression of joint space narrowing and osteophytes in osteoarthritis. Arthritis Rheumatol. 2014;66(9):2440-9.
127. Kraus VB, Collins JE, Hargrove D, Losina E, Nevitt M, Katz JN, et al. Predictive validity of biochemical biomarkers in knee osteoarthritis: data from the FNIH OA Biomarkers Consortium. Ann Rheum Dis. 2017;76(1):186-95.
128. Clark AG, Jordan JM, Vilim V, Renner JB, Dragomir AD, Luta G, et al. Serum cartilage oligomeric matrix protein reflects osteoarthritis presence and severity: the Johnston County Osteoarthritis Project. Arthritis Rheum. 1999;42(11):2356-64.
129. Ruiz-Romero C, Fernández-Puente P, Calamia V, Blanco FJ. Lessons from the proteomic study of osteoarthritis. Expert Rev Proteomics. 2015;12(4):433-43.
130. Mahendru S, Roy K, Kukreti S. Peptide Biomarkers: Exploring the Diagnostic Aspect. Curr Protein Pept Sci. 2017;18(9):914-9.
131. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum.

1986;29(8):1039-49.
132. Pascual Garrido C, Hakimiyan AA, Rappoport L, Oegema TR, Wimmer MA, Chubinskaya S. Anti-apoptotic treatments prevent cartilage degradation after acute trauma to human ankle cartilage. Osteoarthritis Cartilage. 2009;17(9):1244-51.
133. Lourido L, Calamia V, Mateos J, Fernández-Puente P, Fernández-Tajes J, Blanco FJ, et al. Quantitative proteomic profiling of human articular cartilage degradation in osteoarthritis. J Proteome Res. 2014;13(12):6096-106.
134. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966-8.
135. Fernández-Puente P, Calamia V, González-Rodríguez L, Lourido L, Camacho-Encina M, Oreiro N, et al. Multiplexed mass spectrometry monitoring of biomarker candidates for osteoarthritis. J Proteomics. 2017;152:216-25.
136. Unwin RD, Griffiths JR, Whetton AD. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). Nat Protoc. 2009;4(6):870-7.
137. Song J, Tan H, Perry AJ, Akutsu T, Webb GI, Whisstock JC, et al. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PLoS One. 2012;7(11):e50300.
138. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
139. Merrell K, Southwick K, Graves SW, Esplin MS, Lewis NE, Thulin CD. Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. J Biomol Tech. 2004;15(4):238-48.
140. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158-D69.
141. Zhen EY, Brittain IJ, Laska DA, Mitchell PG, Sumer EU, Karsdal MA, et al. Characterization of metalloprotease cleavage products of human articular cartilage. Arthritis Rheum. 2008;58(8):2420-31.
142. Peffers MJ, Thornton DJ, Clegg PD. Characterization of neopeptides in equine articular cartilage degradation. J Orthop Res. 2016;34(1):106-20.
143. Wang Y, Li Y, Khabut A, Chubinskaya S, Grodzinsky AJ, Önnerfjord P. Quantitative proteomics analysis of cartilage response to mechanical injury and cytokine treatment. Matrix Biol. 2017;63:11-22.
144. Önnerfjord P, Khabut A, Reinholt FP, Svensson O, Heinegård D. Quantitative proteomic analysis of eight cartilaginous tissues reveals

characteristic differences as well as similarities between subgroups. J Biol Chem. 2012;287(23):18913-24.

145. Hsueh MF, Khabut A, Kjellström S, Önnerfjord P, Kraus VB. Elucidating the Molecular Composition of Cartilage by Proteomics. J Proteome Res. 2016;15(2):374-88.

146. Clutterbuck AL, Smith JR, Allaway D, Harris P, Liddell S, Mobasheri A. High throughput proteomic analysis of the secretome in an explant model of articular cartilage inflammation. J Proteomics. 2011;74(5):704-15.

147. Miyamoto S, Stroble CD, Taylor S, Hong Q, Lebrilla CB, Leiserowitz GS, et al. Multiple Reaction Monitoring for the Quantitation of Serum Protein Glycosylation Profiles: Application to Ovarian Cancer. J Proteome Res. 2018;17(1):222-33.

148. Lorenzo P, Aspberg A, Saxne T, Önnerfjord P. Quantification of cartilage oligomeric matrix protein (COMP) and a COMP neoepitope in synovial fluid of patients with different joint disorders by novel automated assays. Osteoarthritis Cartilage. 2017;25(9):1436-42.

149. Fandridis E, Apergis G, Korres DS, Nikolopoulos K, Zoubos AB, Papassideri I, et al. Increased expression levels of apolipoprotein J/clusterin during primary osteoarthritis. In Vivo. 2011;25(5):745-9.

150. Ritter SY, Subbaiah R, Bebek G, Crish J, Scanzello CR, Krastins B, et al. Proteomic analysis of synovial fluid from the osteoarthritic knee: comparison with transcriptome analyses of joint tissues. Arthritis Rheum. 2013;65(4):981-92.

151. Ritter SY, Collins J, Krastins B, Sarracino D, Lopez M, Losina E, et al. Mass spectrometry assays of plasma biomarkers to predict radiographic progression of knee osteoarthritis. Arthritis Res Ther. 2014;16(5):456.

152. Wilson R, Belluoccio D, Little CB, Fosang AJ, Bateman JF. Proteomic characterization of mouse cartilage degradation in vitro. Arthritis Rheum. 2008;58(10):3120-31.

153. Neidhart M, Hauser N, Paulsson M, DiCesare PE, Michel BA, Häuselmann HJ. Small fragments of cartilage oligomeric matrix protein in synovial fluid and serum as markers for cartilage degradation. Br J Rheumatol. 1997;36(11):1151-60.

154. Yang CY, Chanalaris A, Troeberg L. ADAMTS and ADAM metalloproteinases in osteoarthritis - looking beyond the 'usual suspects'. Osteoarthritis Cartilage. 2017;25(7):1000-9.

155. Wilson R, Diseberg AF, Gordon L, Zivkovic S, Tatarczuch L, Mackie EJ, et al. Comprehensive profiling of cartilage extracellular matrix formation and maturation using sequential extraction and label-free quantitative proteomics. Mol Cell Proteomics. 2010;9(6):1296-313.

156. Cillero-Pastor B, Eijkel GB, Kiss A, Blanco FJ, Heeren RM. Matrix-assisted

laser desorption ionization-imaging mass spectrometry: a new methodology to study human osteoarthritic cartilage. Arthritis Rheum. 2013;65(3):710-20.
157. Mahboob S, Mohamedali A, Ahn SB, Schulz-Knappe P, Nice E, Baker MS. Is isolation of comprehensive human plasma peptidomes an achievable quest? J Proteomics. 2015;127(Pt B):300-9.
158. Aristoteli LP, Molloy MP, Baker MS. Evaluation of endogenous plasma peptide extraction methods for mass spectrometric biomarker discovery. J Proteome Res. 2007;6(2):571-81.
159. Zheng X, Baker H, Hancock WS. Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer. J Chromatogr A. 2006;1120(1-2):173-84.
160. Koomen JM, Li D, Xiao LC, Liu TC, Coombes KR, Abbruzzese J, et al. Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. J Proteome Res. 2005;4(3):972-81.
161. Kamphorst JJ, van der Heijden R, DeGroot J, Lafeber FP, Reijmers TH, van El B, et al. Profiling of endogenous peptides in human synovial fluid by NanoLC-MS: method validation and peptide identification. J Proteome Res. 2007;6(11):4388-96.
162. Kamphorst J. Systems biology of osteoarthritis2019.
163. Ben-Aderet L, Merquiol E, Fahham D, Kumar A, Reich E, Ben-Nun Y, et al. Detecting cathepsin activity in human osteoarthritis via activity-based probes. Arthritis Res Ther. 2015;17:69.
164. Yasuda Y, Kaleta J, Brömme D. The role of cathepsins in osteoporosis and arthritis: rationale for the design of new therapeutics. Adv Drug Deliv Rev. 2005;57(7):973-93.
165. Sambrano GR, Huang W, Faruqi T, Mahrus S, Craik C, Coughlin SR. Cathepsin G activates protease-activated receptor-4 in human platelets. J Biol Chem. 2000;275(10):6819-23.
166. Gao S, Zhu H, Zuo X, Luo H. Cathepsin G and Its Role in Inflammation and Autoimmune Diseases. Arch Rheumatol. 2018;33(4):498-504.
167. Wiedow O, Meyer-Hoffert U. Neutrophil serine proteases: potential key regulators of cell signalling during inflammation. J Intern Med. 2005;257(4):319-28.
168. Padrines M, Wolf M, Walz A, Baggiolini M. Interleukin-8 processing by neutrophil elastase, cathepsin G and proteinase-3. FEBS Lett. 1994;352(2):231-5.
169. Takahashi A, de Andrés MC, Hashimoto K, Itoi E, Oreffo RO. Epigenetic regulation of interleukin-8, an inflammatory chemokine, in osteoarthritis. Osteoarthritis Cartilage. 2015;23(11):1946-54.
170. McGowan SE. Mechanisms of extracellular matrix proteoglycan

degradation by human neutrophils. Am J Respir Cell Mol Biol. 1990;2(3):271-9.
171. Capodici C, Berg RA. Cathepsin G degrades denatured collagen. Inflammation. 1989;13(2):137-45.
172. Morko JP, Söderström M, Säämänen AM, Salminen HJ, Vuorio EI. Up regulation of cathepsin K expression in articular chondrocytes in a transgenic mouse model for osteoarthritis. Ann Rheum Dis. 2004;63(6):649-55.
173. Petrey AC, Flanagan-Steet H, Johnson S, Fan X, De la Rosa M, Haskins ME, et al. Excessive activity of cathepsin K is associated with cartilage defects in a zebrafish model of mucolipidosis II. Dis Model Mech. 2012;5(2):177-90.
174. Kozawa E, Cheng XW, Urakawa H, Arai E, Yamada Y, Kitamura S, et al. Increased expression and activation of cathepsin K in human osteoarthritic cartilage and synovial tissues. J Orthop Res. 2016;34(1):127-34.
175. Yamashita T, Hagino H, Hayashi I, Hayashibara M, Tanida A, Nagira K, et al. Effect of a cathepsin K inhibitor on arthritis and bone mineral density in ovariectomized rats with collagen-induced arthritis. Bone Rep. 2018;9:1-10.
176. Nwosu LN, Gowler PRW, Burston JJ, Rizoska B, Tunblad K, Lindström E, et al. Analgesic effects of the cathepsin K inhibitor L-006235 in the monosodium iodoacetate model of osteoarthritis pain. Pain Rep. 2018;3(6):e685.
177. Hayami T, Zhuo Y, Wesolowski GA, Pickarski M, Duong LT. Inhibition of cathepsin K reduces cartilage degeneration in the anterior cruciate ligament transection rabbit and murine models of osteoarthritis. Bone. 2012;50(6):1250-9.
178. Rose BJ, Kooyman DL. A Tale of Two Joints: The Role of Matrix Metalloproteases in Cartilage Biology. Dis Markers. 2016;2016:4895050.
179. Murphy G, Lee MH. What are the roles of metalloproteinases in cartilage and bone damage? Ann Rheum Dis. 2005;64 Suppl 4:iv44-7.
180. Zeng GQ, Chen AB, Li W, Song JH, Gao CY. High MMP-1, MMP-2, and MMP-9 protein levels in osteoarthritis. Genet Mol Res. 2015;14(4):14811-22.
181. Galasso O, Familiari F, De Gori M, Gasparini G. Recent findings on the role of gelatinases (matrix metalloproteinase-2 and -9) in osteoarthritis. Adv Orthop. 2012;2012:834208.
182. Lipari L, Gerbino A. Expression of gelatinases (MMP-2, MMP-9) in human articular cartilage. Int J Immunopathol Pharmacol. 2013;26(3):817-23.
183. Chen JJ, Huang JF, Du WX, Tong PJ. Expression and significance of MMP3 in synovium of knee joint at different stage in osteoarthritis patients. Asian Pac J Trop Med. 2014;7(4):297-300.
184. Georgiev T, Ivanova M, Kopchev A, Velikova T, Miloshov A, Kurteva E, et al. Cartilage oligomeric protein, matrix metalloproteinase-3, and Coll2-1 as serum biomarkers in knee osteoarthritis: a cross-sectional study. Rheumatol Int. 2018;38(5):821-30.

185. Muley MM, Reid AR, Botz B, Bölcskei K, Helyes Z, McDougall JJ. Neutrophil elastase induces inflammation and pain in mouse knee joints via activation of proteinase-activated receptor-2. Br J Pharmacol. 2016;173(4):766-77.
186. Hilbert N, Schiller J, Arnhold J, Arnold K. Cartilage degradation by stimulated human neutrophils: elastase is mainly responsible for cartilage damage. Bioorg Chem. 2002;30(2):119-32.
187. Muley MM, Krustev E, Reid AR, McDougall JJ. Prophylactic inhibition of neutrophil elastase prevents the development of chronic neuropathic pain in osteoarthritic mice. J Neuroinflammation. 2017;14(1):168.
188. Arispe N, Diaz JC, Flora M. Efficiency of histidine-associating compounds for blocking the alzheimer's Abeta channel activity and cytotoxicity. Biophys J. 2008;95(10):4879-89.
189. CPTAC Assay Characterization Guidance Document. Version 1.0. [Available from: https://proteomics.cancer.gov/sites/default/files/assay-characterization-guidance-document.pdf.
190. FDA Immunogenicity Testing of Therapeutic Protein Products: Developing and Validating Assays for Anti-Drug Antibody Detection. [Available from: https://www.fda.gov/media/119788/download.
191. Lin D, Alborn WE, Slebos RJ, Liebler DC. Comparison of protein immunoprecipitation-multiple reaction monitoring with ELISA for assay of biomarker candidates in plasma. J Proteome Res. 2013;12(12):5996-6003.
192. Shah B, Reid JD, Kuzyk MA, Parker CE, Borchers CH. Developing an iMALDI method. Methods Mol Biol. 2013;1023:97-120.
193. Reverberi R, Reverberi L. Factors affecting the antigen-antibody reaction. Blood Transfus. 2007;5(4):227-40.
194. Neu CP, Reddi AH, Komvopoulos K, Schmid TM, Di Cesare PE. Increased friction coefficient and superficial zone protein expression in patients with advanced osteoarthritis. Arthritis Rheum. 2010;62(9):2680-7.
195. Reesink HL, Watts AE, Mohammed HO, Jay GD, Nixon AJ. Lubricin/proteoglycan 4 increases in both experimental and naturally occurring equine osteoarthritis. Osteoarthritis Cartilage. 2017;25(1):128-37.
196. Ludwig TE, McAllister JR, Lun V, Wiley JP, Schmidt TA. Diminished cartilage-lubricating ability of human osteoarthritic synovial fluid deficient in proteoglycan 4: Restoration through proteoglycan 4 supplementation. Arthritis Rheum. 2012;64(12):3963-71.
197. Hui AY, McCarty WJ, Masuda K, Firestein GS, Sah RL. A systems biology approach to synovial joint lubrication in health, injury, and disease. Wiley Interdiscip Rev Syst Biol Med. 2012;4(1):15-37.
198. Lorenzo P, Bayliss MT, Heinegård D. Altered patterns and synthesis of

extracellular matrix macromolecules in early osteoarthritis. Matrix Biol. 2004;23(6):381-91.

199. Hsueh MF, Önnerfjord P, Kraus VB. Biomarkers and proteomic analysis of osteoarthritis. Matrix Biol. 2014;39:56-66.

200. Popp R, Basik M, Spatz A, Batist G, Zahedi RP, Borchers CH. How iMALDI can improve clinical diagnostics. Analyst. 2018;143(10):2197-203.

VII. Annex

1. Annex I: Supplementary Material of Chapter I.

Annex I. Supplementary data I. HSA depletion results after comparing two strategies: TAC/EtOH chemical depletion and the commercial Thermo HSA depletion kit.

Annex I. Supplementary data II. Proteins commonly and exclusively identified by nanoLC-MALDI-TOF/TOF and nanoLC-ESI-tTOF.

Annex I. Supplementary data III. Intra- and inter-plate protein variability calculated with the triplicate measurements of the RA pool.

Annex I. Supplementary data IV. Statistics for each protein measured by ELISA with or without median normalization of the data and/or outlier removal.

Annex I. Supplementary data V. SAA1, AACT, HPT and A1AG ELISA data normalization including the n factor determination and the data information of each of the groups analyzed.

Annex I. Supplementary data VI. Protein sequences and MRM measured peptides (in orange).

2. Annex II: Supplementary Material of Chapter II.

Annex II. Supplementary data I. Discovery phase results. Number of endogenous peptides and their corresponding unique proteins identified in the secretomes of human articular cartilage.

Annex II. Supplementary data II. Targeted proteomics design. MRM mass spectrometry transitions analyzed in this work, and settings for their analysis.

Annex II. Supplementary data III. Quantification data obtained for the panel of peptides analyzed by MRM mass spectrometry. Results are expressed in peak area ratios of abundance (light/heavy peptides), with a confidence level of $p<0.05$*.

Annex II. Supplementary data IV. Serum and synovial fluid (SF) proteins whose neopeptides were selected for MRM method development.

3. Annex III: Supplementary Material of Chapter III.

Annex III. Supplementary data I. MRM chromatograms showing the presence of CILP1-IVG and CILP1-TFL in a cartilage digest sample (at 24.9min and 32.8 min respectively), confirmed with a high confidence (greater than 98%) by peptide identification search using the ProteinPilot software, whereas they are not detected and/or identified in the depleted and/or non-depleted serum digests.

Annex III. Supplementary data II. MRM chromatograms showing the presence of PRG4-GGS and PRG4-GFG in a cartilage digest sample (at 14.1min and 20.2min respectively), confirmed with a high confidence (99%) by peptide identification search using the ProteinPilot software, whereas they are hardly detected and no identified in the non-depleted serum digests. In the case of depleted serum digests, the peptides and analogous proteins are detected by MRM, however this was not sufficient to achieve protein and/or peptide identification when MS/MS spectra were searched in the ProteinPilot software.

Annex III. Supplementary data III. First iMALDI PRG4-GFG response curve with 8 points made in triplicates from 0 to 2500 fmol, used for assay validation. Light (L/NAT) and heavy (H/SIS) peak intensities and their corresponding L/H ratios, including the average and coefficient of variation (CV) data of each triplicate measurements.

Annex III. Supplementary data IV. iMALDI PRG4-GFG assay intra-day and inter-day variability and precision measurements of low (117.5 fmol), medium (470 fmol) and high (2350 fmol) concentrations made with 5 replicates.

Annex III. Supplementary data V. iMALDI PRG4-GFG accuracy (% nominal) measurements of low (117.5 fmol), medium (470 fmol) and high (2350 fmol) concentrations made with 5 replicates.

Annex III. Supplementary data VI. Second iMALDI PRG4-GFG response curve with 8 points made in triplicates from 0 to 2500 fmol, used for sample measurement. Light (L/NAT) and heavy (H/SIS) peak intensities and their corresponding L/H

ratios, including the average and coefficient of variation (CV) data of each triplicate measurements.

Annex III. Supplementary data VII. Second iMALDI PRG4-GFG response curve representation with Qualis-SIS software of 8 points made in triplicates from 0 to 2500 fmol, used for sample measurement.

Annex III. Supplementary data VIII. SPM and SCM measurements of the 38 serum samples analyzed using the validated iMALDI PRG4-GFG assay in fmol, fmol/µL serum, µg/mL serum. Difference (%) between the cited SPM and SCM measurements.

Annex III. Supplementary data IX. SISCAPA-MRM CILP1-IVG response curve chromatogram representation and values of 6 points ranging from 0 to 500 fmol, made in duplicates. Light (NAT) and heavy (SIS) peak areas and their corresponding NAT/SIS ratios, including the average and coefficient of variation (CV) data of each duplicate measurements.

Annex III. Supplementary data X. SISCAPA-MRM CILP1-TFL response curve chromatogram representation and values of 6 points ranging from 0 to 500 fmol, made in duplicates. Light (NAT) and heavy (SIS) peak areas and their corresponding NAT/SIS ratios, including the average and coefficient of variation (CV) data of each duplicate measurements.

Annex III. Supplementary data XI. Concentration measurements of 12 serum samples analyzed using the validated SISCAPA-MRM CILP1-IVG and CILP1-TFL multiplex assay in fmol, fmol/µL serum, µg/mL serum and the corresponding average values.

4. Annex IV: articles in which I participated during my stage as a predoctoral student.

- Patricia Fernandez-Puente*, **Lucía González-Rodríguez***, Valentina Calamia, Florencia Picchi, Lucía Lourido, María Camacho-Encina, Natividad Oreiro, Beatriz Rocha, Rocío Paz, Anabel Marina, Carlos García, Francisco J. Blanco and Cristina Ruiz-Romero. *Analysis of endogenous peptides released from osteoarthritic cartilage unravels novel pathogenic markers*. Molecular & Cellular Proteomics July 27, 2019, mcp.RA119.001554. https://doi.org/10.1074/mcp.RA119.001554.
  *These authors contributed equally to this work.

- F.J. Blanco, M. Camacho-Encina, **L. González-Rodríguez**, I. Rego-Pérez, J. Mateos, P. Fernández-Puente, L. Lourido, B. Rocha, F. Picchi, M.T. Silva-Díaz, M. Herrero, H. Martínez, J. Verges, C. Ruiz-Romero and V. Calamia. *Predictive modeling of therapeutic response to chondroitin sulfate/glucosamine hydrochloride in knee osteoarthritis.* Therapeutic Advances in Chronic Disease, July 15, 2019, Manuscript ID TAJ-19-01-OA-004.R2.

- Fernández-Puente P, Calamia V, **González-Rodríguez L**, Lourido L, Camacho-Encina M, Oreiro N, Ruiz-Romero C, Blanco FJ. *Multiplexed mass spectrometry monitoring of biomarker candidates for osteoarthritis*. Volume 152, 30 January 2017, pages 216-225. https://doi.org/10.1016/j.jprot.2016.11.012

5. Annex V: Resumen de la tesis en español.

La articulación se define como el órgano que une varios huesos, cuya función principal es la de permitir su correspondiente movimiento. Las articulaciones sinoviales son las más comunes. Este tipo de articulaciones están esencialmente formadas por una cavidad articular, la cual está rodeada por una membrana sinovial cuyas células secretan líquido sinovial, responsable de la lubricación y nutrición del cartílago articular. El cartílago articular es de tipo hialino y se considera uno de los componentes más importantes de las articulaciones, ya que previene la fricción ósea y amortigua las fuerzas ejercidas sobre el hueso. Está compuesto primordialmente por matriz extracelular (ECM en inglés) y condrocitos (2%), definidos como las células especializadas responsables del desarrollo de la ECM del cartílago. A su vez, la ECM está compuesta principalmente por agua, colágenos y proteoglicanos, así como otras proteínas no-colágenas y glicoproteínas presentes en menor proporción, cuya prevalencia varía en función de las capas del cartílago (superficial, media, profunda). Es importante conocer las características de la articulación para poder entender las enfermedades que afectan a este particular órgano, definidas como enfermedades reumáticas y músculo-esqueléticas (RMD en inglés).

Existen más de 200 RMD, aunque esta tesis se centra en dos de las más comunes: artrosis (OA en inglés) y artritis reumatoide (RA en inglés). Se estima que, en conjunto, estas enfermedades afectan a alrededor de 2 billones de personas a nivel mundial, fundamentalmente a mujeres, siendo además la segunda causa de incapacitación. A pesar de las diferencias que existen entre ellas (la RA se caracteriza por la presencia de inflamación articular debido a procesos autoinmunes, mientras que en la OA se produce el deterioro de la articulación por desgaste, sin necesidad de ocasionar inflamación articular), ambas producen síntomas similares (dolor y rigidez articular y/o pérdida de funcionalidad). El diagnóstico de estas enfermedades se realiza mediante análisis de marcadores no específicos, los cuales monitorizan otros procesos fisiopatológicos como la

inflamación o la respuesta inmune, así como mediante diversas técnicas de imagen (radiografía, resonancia magnética), cuyo uso frecuente es dañino para el paciente. En el caso de la OA, no existe cura para esta enfermedad, por lo que los tratamientos actuales están dirigidos a paliar el dolor y mejorar la funcionalidad de las articulaciones en las fases tempranas, mientras que como último recurso puede llegar a ser necesario el remplazo articular. En la RA, los avances obtenidos mediante el tratamiento con fármacos modificadores de la enfermedad (DMARDs) y los nuevos agentes biológicos ha supuesto una importante mejora en el manejo de los pacientes, pero aún las estrategias terapéuticas se establecen mediante ensayo-error. Debido a estas limitaciones en cuanto al diagnóstico y tratamiento, existe una gran necesidad de encontrar nuevos biomarcadores que faciliten el entendimiento de los procesos patológicos que tienen lugar en estos pacientes con el fin de definir mejor el diagnóstico y tratamiento efectivo para el manejo de estas enfermedades.

La proteómica se define como la disciplina encargada del estudio del proteoma (conjunto de proteínas que expresa una célula u organismo), así como de la estructura y función de las proteínas. Al contrario que el genoma (el cual es el mismo para todas las células de un mismo organismo), el proteoma es altamente complejo y dinámico. Esto es debido a que sus componentes varían en función de la célula que los expresa, así como de otros procesos relacionados con la expresión proteica (*splicing* alternativo, modificaciones postraduccionales) o con factores externos (condiciones patológicas, situaciones de estrés, exposición a ciertos fármacos). De todas las aplicaciones proteómicas, la búsqueda de biomarcadores proteicos es la más utilizada en clínica, ya que permite determinar qué proteínas se encuentran alteradas en procesos patológicos específicos, con el fin de caracterizar los mecanismos moleculares y biológicos específicos de la enfermedad, posibilitar la mejora del diagnóstico y la identificación de dianas farmacológicas efectivas. La búsqueda de biomarcadores proteicos es generalmente llevada a cabo mediante el desarrollo de varias fases preclínicas antes de la evaluación clínica final, en las que el número de biomarcadores

candidatos disminuye y el número de muestras analizadas aumenta con el transcurso de la búsqueda. Estas fases preclínicas son: 1) descubrimiento o identificación de candidatos biomarcadores; 2) cualificación de la abundancia diferencial de los candidatos identificados; 3) verificación de los mismos en un número más grande de muestras; 4) validación de los biomarcadores verificados mediante inmunoensayos dirigidos en cohortes de cientos de muestras.

Para ello, se pueden llevar a cabo diferentes estrategias proteómicas, que generalmente comprenden 3 fases diferenciadas: obtención de la muestra (cartílago, secretoma de cartílago, líquido sinovial, suero o plasma, utilizadas en este caso para la búsqueda de biomarcadores de OA/RA), aislamiento y purificación de proteínas (mediante técnicas de depleción o enriquecimiento) y finalmente análisis de proteínas mediante espectrometría de masas y/o inmunoensayos. El aislamiento y purificación de proteínas es uno de los pasos cruciales, especialmente en las primeras fases de la búsqueda, donde el procesamiento de la muestra tiende a ser moderado-extenso con el fin de identificar y cuantificar el máximo número de biomarcadores candidatos. Además, la complejidad de la muestra es también un factor importante en la elección del procesamiento de la misma. Particularmente, las 12 proteínas sanguíneas más abundantes representan el 95% del contenido proteico total en plasma o suero. El resto de proteínas se encuentran en un amplio rango dinámico de concentración, en el que las proteínas de interés para la búsqueda de biomarcadores (que suelen ser proteínas secretadas por tejidos o de señalización) se hallan en concentraciones bajas, del orden de ng/mL o pg/mL. Por esta razón, numerosas técnicas de depleción de proteínas mayoritarias y de enriquecimiento de proteínas minoritarias (mediante técnicas no selectivas o dirigidas de precipitación físico-químicas, de inmunoafinidad o kits comerciales) han sido desarrolladas y utilizadas individualmente y en combinación dependiendo de la muestra y análisis posterior.

Así, por un lado, la espectrometría de masas (MS en inglés) es extensamente usada con fines de descubrimiento y verificación de biomarcadores candidatos, debido a su capacidad para detectar numerosas proteínas a la vez de manera específica. El típico análisis mediante MS incluye tres fases: 1) ionización de la muestra (en este caso péptidos o proteínas) mediante fuentes de ionización como MALDI (Matrix Assisted Laser Desorption/Ionization) o ESI (Electrospray ionization); 2) separación de los iones formados según su relación m/z mediante diversos analizadores de masa como TOF (Time of Flight), Q (Quadrupole), LIT (Linear Ion Trap), Orbitrap; y 3) detección y registro de los iones con valores específicos de m/z. Primordialmente, mediante MS se obtienen las masas totales de proteínas/péptidos analizados (MS) y sus patrones de fragmentación o iones producto (espectrometría de masas en tándem o MS/MS). En MS/MS se usan dos etapas consecutivas de análisis de masa: 1) aislamiento de un ion precursor de la proteína o péptido de interés por su m/z (MS); 2) análisis de la m/z de los iones del producto formados por fragmentación espontánea o inducida del ion precursor seleccionado (MS/MS) mediante diferentes técnicas como CID (Collision-induced dissociation). Así, los espectros de MS/MS contienen fragmentos de los precursores que están relacionados con la secuencia de los péptidos correspondientes. Esta información, junto con los espectros de MS, ayuda a dilucidar la identificación de la secuencia de estas proteínas y/o péptidos.

Existen dos tipos principales de estrategias para identificar proteínas: la técnica *bottom-up* y la *top-down*. En el enfoque *bottom-up* se detectan péptidos de proteínas digeridos, mientras que en el enfoque *top-down* se analizan directamente las proteínas intactas por ESI-MS. La aplicación de la estrategia *top-down* es limitada debido a la baja sensibilidad y a la dificultad para ionizar muchas proteínas de gran peso molecular, mientras que la estrategia *bottom-up* es más común, ya que la solubilización y el fraccionamiento de la muestra antes del análisis por MS es más fácil para los péptidos que para las proteínas. En cuanto a la cuantificación de los péptidos o proteínas identificadas en MS, se utilizan normalmente dos tipos de técnicas: técnicas de cuantificación sin marcaje y

técnicas basadas en marcajes metabólicos o químicos utilizando distintos tipos de isótopos estables que se unen a las proteínas o péptidos, como el iTRAQ (Isobaric tags for relative and absolute quantitation) o el uso de péptidos marcados (heavy, SIS peptides en inglés). Los métodos sin marcaje no son realmente precisos, aunque son técnicas más simples y menos costosas que proporcionan una mayor profundidad analítica en una gran cantidad de muestras. Por su parte, las técnicas de marcaje analizan diferentes análogos de proteínas/péptidos marcados, los cuales producen eficiencias de ionización y señales de respuesta de MS similares en comparación con las proteínas/péptidos no marcados. Generalmente se utilizan dos estrategias proteómicas de MS para el descubrimiento y la verificación de biomarcadores: proteómica "*shotgun*", que usa técnicas de cuantificación relativa para el cribado inicial a gran escala de potenciales biomarcadores candidatos expresados diferencialmente, y proteómica dirigida o "*targeted*", como la MRM (Multiple Reaction Monitoring), basada en la cuantificación absoluta y verificación reproducible, sensible y precisa de los biomarcadores candidatos previamente seleccionados.

Por otro lado, la técnica ELISA (Enzyme-Linked ImmunoSorbent Assay) es una de las mejores opciones con fines de validación de biomarcadores, ya que ofrece una gran sensibilidad a la hora de analizar un gran número de proteínas de forma individual. Sin embargo, existen otras limitaciones a la hora de aplicar esta técnica, como por ejemplo la baja capacidad de analizar múltiples analitos a la vez, o el uso de ELISAs comerciales cuyos anticuerpos posean una baja especificidad. Los métodos de análisis proteómicos pueden abordar estas limitaciones con el objetivo de poder cuantificar grandes paneles de proteínas biomarcadoras. En este sentido, las técnicas híbridas de inmunoafinidad acopladas a espectrometría de masas (IA-MS) brindan importantes ventajas frente a las estrategias proteómicas individuales, ya que combinan la sensibilidad característica de los inmunoensayos como ELISA con la especificidad y capacidad de la MS para analizar simultáneamente varios analitos. Además, estas técnicas híbridas se consideran una de las mejores opciones para el análisis de proteínas minoritarias,

así como de variantes proteicas, en muestras complejas como el suero o plasma, incluso con fines de validación clínica.

Entre las estrategias proteómicas más emergentes, la peptidómica actualmente recibe gran consideración para el hallazgo de biomarcadores proteicos, ya que está dirigida a fragmentos de proteínas producidos endógenamente. Éstos pueden indicar la presencia y/o el estado de condiciones patológicas particulares, ya que los péptidos endógenos o neopéptidos pueden generarse mediante síntesis activa o procesamiento proteolítico aberrante de proteínas precursoras más grandes. Debido a que la presencia y función anormal de las proteasas envueltas en el proceso de la OA es uno de los mayores factores implicados en el deterioro del cartílago y su ECM, la consecución de un estudio peptidómico sería una atractiva opción para la detección de fragmentos proteicos o neopéptidos con valor biomarcador del proceso de esta enfermedad.

Particularmente, en esta tesis, con el fin de encontrar proteínas con papel biomarcador en RA y OA y así mejorar el diagnóstico y la selección de terapias efectivas en estas enfermedades, se han llevado a cabo los siguientes trabajos y obtenido las correspondientes conclusiones:

1) Búsqueda de biomarcadores que permitan monitorizar la actividad de la enfermedad en pacientes con RA utilizando el procedimiento convencional anteriormente descrito para ello.

Inicialmente, se ha llevado a cabo la primera fase de descubrimiento de biomarcadores candidatos aplicando la técnica de marcaje iTRAQ en 8 muestras resultantes de la mezcla de plasma de 80 pacientes diagnosticados con RA (4 muestras combinando pacientes con RA de alta actividad de enfermedad y otras 4 con pacientes con RA de baja actividad de enfermedad). Así, estas 8 muestras se han deplecionado individualmente con un método de precipitación química con TCA y etanol, previamente estandarizado en el laboratorio, con el fin de descartar la albúmina (proteína mayoritaria en sangre). Una vez deplecionadas,

las 8 muestras fueron individualmente digeridas con tripsina con el fin de obtener péptidos trípticos, los cuales se marcaron con los correspondientes reactivos iTRAQ y se combinaron para ser analizados conjuntamente mediante nanoLC-MS/MS, con MALDI-TOF/TOF y ESI-tTOF. De la consecución de esta fase, 186 proteínas fueron identificadas, de las cuales 11 se encontraron significativamente alteradas en pacientes con alta actividad de RA frente a los de baja actividad, una vez realizada la cuantificación relativa mediante comparación de los marcajes iTRAQ.

Una vez identificados los 11 candidatos biomarcadores para la monitorización de pacientes con diferente actividad de enfermedad en RA, se ha diseñado un método de MRM para la monitorización de estas proteínas, el cual incluye 26 péptidos trípticos correspondientes a estas mismas. Así, con la ayuda de este método y mediante nanoLC-MS/MS, con ESI-QTRAP, se han analizado de forma independiente las mismas 80 muestras de plasma utilizadas en la fase de descubrimiento. De este análisis, 5 péptidos, correspondientes a las proteínas HPT, AACT, SAA1, A1AG1 y A1AG2, se han visto aumentados de forma significativa (p-valor < 0.05) en plasmas de pacientes de RA con alta actividad, verificando así el papel biomarcador de estas 5 proteínas que se había encontrado en la anterior fase de descubrimiento.

Por último, las proteínas HPT (Haptoglobin), AACT (Alpha 1 Antichymotrypsin, SAA1 (Serum Amyloid A1) y A1AG1 (Alpha-1-acid glycoprotein 1) fueron validadas mediante ELISA en 420 muestras de plasma, entre las que se incluyeron 170 muestras adicionales de pacientes con RA, 80 muestras de pacientes con lupus eritematoso sistémico (SLE en inglés), 80 muestras de pacientes con artritis psoriática (PsA en inglés) y 90 muestras de donantes sanos (HD en inglés). Las muestras de pacientes con SLE y PsA se incluyeron en el estudio como controles de enfermedad frente a la RA, con el fin de evaluar el papel clasificador de los marcadores propuestos en el diagnóstico de enfermedades reumáticas. Además, en cada una de estas tres patologías reumáticas se

incluyeron también pacientes con alta y baja actividad de la enfermedad, con el fin de a su vez detectar su papel como marcador de actividad en estas patologías. Una vez realizada la puesta a punto individual de los ELISAs (estimación de la dilución de las muestras de plasma y determinación del rango de la recta patrón), las 420 muestras de plasma fueron analizadas simultáneamente para cada biomarcador candidato. Estadísticamente, las 4 proteínas evaluadas fueron confirmadas como marcadores de actividad de RA, así como también se pudo validar su papel como clasificadoras de otras enfermedades reumáticas inflamatorias autoinmunes, ya que pueden distinguir también entre pacientes con RA, SLE y PsA. Al mismo tiempo, AACT también logra distinguir entre pacientes con SLE que tengan diferentes actividades de enfermedad, mientras que HPT y SAA1 consiguen distinguir entre pacientes con PsA con diferentes actividades de enfermedad.

En resumen, en esta búsqueda de biomarcadores de actividad, las proteínas HPT, SAA1, AACT y A1AG1 se han validado como marcadores de actividad en RA, SLE y PsA, así como confirmado como clasificadores de enfermedades reumáticas inflamatorias autoinmunes similares. Este panel de proteínas se ha validado en un total de 420 muestras de plasma de pacientes con RA, SLE, PsA y HD, y puede llegar a ser de gran ayuda para la monitorización clínica del estado de estos pacientes a nivel molecular, así como para la consiguiente selección de terapias más específicas y efectivas para ellos, reduciendo así la carga social, económica y material que supone el seguimiento y tratamiento de esta enfermedad, así como de patologías reumáticas semejantes.

2) Estudio peptidómico en muestras de secretoma de cartílago, líquido sinovial y suero de pacientes con OA y donantes sanos.

Como ya se ha descrito, la peptidómica puede ser de gran interés para la búsqueda de biomarcadores en OA, ya que el deterioro propio del cartílago en esta enfermedad está ligado a diversos procesos moleculares, entre ellos a la

acción de las proteasas. Éstas pueden llegar a producir fragmentos proteicos (o neopéptidos) durante el inicio o transcurso de la enfermedad, pudiendo así valorarse como marcadores candidatos de diagnóstico o monitorización de la OA.

Con el fin de obtener un primer panel de neopéptidos en secretoma de cartílago, explantes (*punch*) de cartílago de 3 pacientes con OA y 2 donantes sanos fueron cultivados con el fin de obtener las correspondientes fracciones de proteínas secretadas (secretomas) de cartílago. Seguidamente, los secretomas fueron procesados mediante ultrafiltración con ayuda de filtros de tipo Amicon de 10 kDa, para el aislamiento de neopéptidos presentes en estos secretomas. Los neopéptidos aislados fueron analizados mediante LC-MS/MS con la ayuda de un equipo con analizador Orbitrap, identificándose así 1175 neopéptidos correspondientes a 101 proteínas. De ellas, se seleccionaron 54 péptidos pertenecientes a 17 proteínas con el fin de desarrollar un método MRM y verificar estos posibles biomarcadores candidatos en un mayor número de muestras. Estos 54 péptidos fueron escogidos en base a su presencia en el mayor número de secretomas analizados, así como debido a la correspondiente puntuación obtenida en su identificación y el papel de las proteínas a la que pertenecen en relación con la ECM del cartílago. Así, una vez evaluada la presencia de este panel de péptidos mediante MRM en nuevas muestras de secretomas, se definió un método final de MRM, en el que se monitorizaron 23 neopéptidos pertenecientes a 9 proteínas diferentes en secretomas de un total de 40 pacientes con OA de rodilla o cadera y donantes sanos. Finalmente, 8 neopéptidos pertenecientes a las proteínas PRELP (Prolargin), CLUS (Clusterin), CILP1 (Cartilage intermediate layer protein 1), COMP (Cartilage Oligomeric Matrix Protein) y MGP (Matrix Gla Protein) se encontraron significativamente alterados en los secretomas de pacientes con OA. A su vez, es interesante destacar la diferencia que existe en cuanto a la presencia de estos péptidos según procedan de secretomas de cartílago de rodilla o cadera. Así, los neopéptidos pertenecientes a CLUS y COMP fueron diferencialmente detectados en cadera, mientras que los correspondientes a CILP1 y PRELP se vieron significativamente alterados en rodilla, lo cual sugiere una

secreción de neopéptidos dependiente de la articulación y asociada al progreso de la OA. En particular, uno de los neopéptidos de PRELP (DSNKIETIPN) ofrece el mejor valor como biomarcador de OA, ya que al comparar secretomas derivados de tejidos sanos con artrósicos, se ha estimado un AUC (Area Under the Curve) de 0.891 al realizar estadísticas de curva ROC (Receiver Operating Characteristic).

Una vez descrito este panel inicial de neopéptidos en secretoma de cartílago, el siguiente objetivo se centró en detectarlo también en muestras más accesibles clínicamente como el líquido sinovial o suero. Aunque se probaron diferentes técnicas de aislamiento de péptidos libres en estas matrices, así como combinaciones de ellas, ninguna de las estrategias facilitó el análisis de los 8 neopéptidos encontrados en el secretoma. Las técnicas de IA-MS anteriormente descritas podrían ayudar y facilitar su análisis en estas matrices, pero a su vez conllevan una larga y tediosa puesta a punto al tratarse de neopéptidos de proteínas de la ECM del cartílago.

Por todo ello, se llevó a cabo una nueva búsqueda independiente de un panel de neopéptidos en suero y líquido sinovial. Con este fin, se evaluaron diferentes técnicas de precipitación química y de ultrafiltración con filtros de diversos pesos moleculares, así como combinaciones entre ambas. La combinación de una precipitación química inicial mediante acetonitrilo y una posterior ultrafiltración con filtros Amicon de 10 kDa fue la estrategia más efectiva para llevar a cabo el aislamiento del mayor número de neopéptidos en suero, y por ello fue consiguientemente aplicada en la búsqueda de neopéptidos en líquido sinovial y suero de pacientes con OA y donantes sanos. En esta búsqueda, llevada a cabo en muestras de suero y líquido sinovial mediante LC-MS/MS con la ayuda de un Orbitrap como en el caso de la búsqueda en secretoma, se identificaron más de 1000 neopéptidos. 21 de ellos, correspondientes a 16 proteínas diferentes, fueron detectados mediante MRM con la ayuda de péptidos marcados. De este análisis dirigido, 6 péptidos pertenecientes a APOA4 (Apolipoprotein A-IV), ITIH4 (Inter-alpha-trypsin inhibitor heavy chain H4), CO3

(Complement component C3) y KNG1 (Kininogen 1) se vieron alterados entre muestras de pacientes OA y donantes sanos, aunque en este caso sólo se podría hablar de tendencia debido a que el número reducido de muestras analizadas no permitió realizar un análisis estadístico exhaustivo para confirmar el papel biomarcador de estos neopéptidos.

En resumen, en este estudio peptidómico se ha detectado un panel de 8 neopéptidos con posible valor biomarcador en la OA en secretomas de cartílago. Todos estos neopéptidos pertenecen a proteínas relacionadas con la ECM del cartílago (como la PRELP, CLUS, CILP1, COMP o MGP), aunque su presencia se ve alterada dependiendo de la articulación evaluada. La identificación de estos péptidos en matrices más accesibles clínicamente como líquido sinovial o suero no ha podido llevarse a cabo al utilizar técnicas convencionales de aislamiento de neopéptidos. Por ello, se ha llevado a cabo una nueva búsqueda independiente de neopéptidos en suero y líquido sinovial, en la que 6 neopéptidos pertenecientes a APOA4, ITIH4, CO3 y KNG1 se han visto alterados en muestras de suero de pacientes con OA y donantes sanos. Al tratarse de sólo una tendencia observada en un número reducido de muestras, es necesario el análisis en un mayor número de muestras independientes con el fin de obtener resultados con valor estadístico que confirmen el posible papel biomarcador de estos 6 neopéptidos candidatos para el diagnóstico de la OA. Aunque es necesario un análisis más profundo de este panel de neopéptidos, los resultados obtenidos del actual estudio peptidómico podrían llegar a facilitar el diagnóstico y monitorización de pacientes con OA, ya que la formación de estos fragmentos peptídicos se ha visto alterada en muestras de pacientes artrósicos frente a donantes sanos.

3) Desarrollo de técnicas híbridas de inmunoafinidad acopladas a espectrometría de masas con el fin de detectar proteínas pertenecientes a la matriz extracelular del cartílago en muestras de suero.

Muchas de las proteínas con papel biomarcador proceden de tejidos específicos al sistema u órgano afectado(s) por una patología en particular. Debido al gran rango dinámico de muestras clínicamente accesibles como plasma o suero, estos potenciales biomarcadores se encuentran en concentraciones muy bajas, ya que las proteínas que se secretan de tejidos específicos se encuentran notablemente diluidas en este tipo de fluidos biológicos. Esta baja abundancia hace que estas proteínas sean prácticamente indetectables mediante técnicas convencionales basadas en MS. En el caso de los inmunoensayos, la determinación de la abundancia de estas proteínas se puede ver alterada por la especificidad de los anticuerpos empleados. Por ello, el uso conjunto de ambas técnicas proporciona numerosas ventajas vinculadas a cada una de ellas, lo que ha generado un gran interés en el desarrollo de métodos de IA-MS. Así, en este trabajo se ha llevado a cabo el desarrollo de dos técnicas de IA-MS, Immuno-MALDI (iMALDI) y Stable isotope standards and capture by anti-peptide antibodies (SISCAPA), con el fin de detectar dos proteínas de capas superficiales e intermedias de cartílago (PRG4 o Proteoglycan 4 y CILP1 respectivamente) en muestras de suero de pacientes artrósicos y donantes sanos. Ambos métodos utilizan péptidos sintéticos marcados y anticuerpos anti-péptido, con el fin de enriquecer y cuantificar de forma absoluta el péptido endógeno en suero con ayuda del análogo marcado.

Por un lado, con el fin de desarrollar un ensayo de iMALDI que permitiese la detección conjunta de 4 péptidos trípticos pertenecientes a CILP1 y PRG4 (así como sus correspondientes péptidos marcados), se probaron diferentes parámetros en cada una de las fases del desarrollo del ensayo. Inicialmente, se establecieron las condiciones de conjugación de anticuerpos a partículas magnéticas, después de fijar la relación entre la concentración de anticuerpo y la

cantidad de partículas magnéticas necesarias para su saturación y valorar el uso de un *crosslinker* para la fijación de ambos. A continuación, se determinó el protocolo de incubación, lavado y elución de la muestra. En último lugar, se definió un protocolo final de iMALDI, utilizando los parámetros escogidos para efectuar un análisis óptimo de CILP1 y PRG4 en suero. Después de analizar diferentes muestras de suero, sólo el ensayo del péptido PRG4-GFG cumplió con los requisitos de la presente metodología, incluyendo la detección óptima de ambos péptidos (péptido endógeno del suero y péptido análogo marcado), la ausencia de fondo interferente en el espectro de masas, así como aceptables espectros e intensidades de señal. Por ello, se configuró y validó un ensayo iMALDI para medir PRG4-GFG en muestras de suero, cuya curva patrón ofreció valores de regresión satisfactorios (R^2: 0,992) y brindó a su vez estimaciones de variabilidad total inferiores al 15%.

Una vez validada analíticamente la respuesta y variabilidad del ensayo iMALDI para la detección de PRG4 en suero, se analizaron 38 muestras de suero de pacientes con OA y donantes sanos, aunque no se observaron diferencias significativas al analizar los dos grupos. Con el fin de comprobar la utilidad de la técnica de iMALDI establecida para la cuantificación absoluta de PRG4, se analizaron las mismas 38 muestras de suero mediante el uso de un ELISA comercial. Nuevamente, no se encontraron diferencias significativas entre los dos grupos, aunque sí se detectó una baja correlación entre las mediciones de PRG4 mediante iMALDI y ELISA (r: 0,36). Particularmente, los niveles séricos de PRG4 determinados mediante ELISAfueron de alrededor de 2.5 µg/mL, valores de medición aproximadamente 4 veces más bajos que los obtenidos utilizando el iMALDI desarrollado (10 µg/mL). Probablemente, esta diferencia entre métodos es debida al hecho de que en iMALDI sólo se utiliza un anticuerpo para capturar el contenido total de PRG4, mientras que en el ELISA se utilizan dos anticuerpos, los cuales deben unirse conjuntamente para poder detectar PRG4, así como de la especificidad de los anticuerpos utilizados en ambas técnicas. Adicionalmente, otras 34 muestras de suero de pacientes con OA y donantes sanos se analizaron

mediante ELISA y en este caso se ha visto una tendencia más clara de esta proteína a estar aumentada en pacientes con OA, aunque sigue sin ser significativa (p-valor: 0.191).

Por otro lado, como el iMALDI no era lo suficientemente sensible o adecuado para detectar ninguno de los dos péptidos trípticos seleccionados de CILP1, se desarrolló un ensayo SISCAPA-MRM con el fin de cuantificar conjuntamente estos péptidos y la proteína correspondiente. Las rectas patrón para ambos péptidos detectados en suero ofrecieron valores de regresión satisfactorios (R^2: 0.97 para CILP1-IVg y R^2: 0.99 para CILP1-TFL). Así, sueros de 6 pacientes con OA y 6 donantes sanos se analizaron utilizando el ensayo SISCAPA multiplex desarrollado, aunque no se encontraron diferencias significativas entre los dos grupos al cuantificar ninguno de los dos péptidos. Concretamente, los niveles séricos de CILP1-IVG fueron de alrededor de 1.3 µg/mL en ambos grupos mientras que los niveles de CILP1-TFL aumentaron ligeramente en OA en comparación con los controles sanos (8.76 vs 6.19 µg/mL, p-valor: 0.132).

En resumen, se ha desarrollado un ensayo iMALDI para la cuantificación de PRG4 en suero y se ha comparado y confirmado su utilidad con la ayuda de un ELISA comercial frente a la misma proteína. Los valores de PRG4 obtenidos mediante iMALDI son 4 veces superiores que los detectados mediante ELISA, lo cual sugiere que iMALDI es capaz de detectar mayor abundancia de PRG4 en suero. Aunque el análisis de estas muestras de suero de pacientes con OA y donantes sanos no supuso diferencias significativas entre grupos, un análisis posterior en un mayor número de muestras mostró una tendencia más clara a estar aumentada en pacientes con OA. Además, ya que CILP1 no pudo ser cuantificada mediante iMALDI, se desarrolló un ensayo SISCAPA-MRM en formato multiplex para su cuantificación absoluta que se empleó sobre 12 muestras de suero de pacientes con OA y donantes sanos. Los resultados mostraron una tendencia de aumento de CILP1 en pacientes con OA, aunque de nuevo no estadísticamente significativa.