Proteomics: Applications to the Study of Rheumatoid and Osteoarthritis

Reuben Gobezie, MD, David S. Sarracino, PhD, Christopher Evans, PhD, Thomas S. Thornhill, MD, Peter J. Millett, MD, MSc

Brigham and Women's Hospital

INTRODUCTION

Rheumatoid arthritis (RA) and osteoarthritis (OA) comprise two of the most common chronic musculoskeletal disorders encountered by physicians throughout the world. At the start of this millennium, the United Nations declared the years between 2000 and 2010 the "Bone and Joint Decade" in an attempt to highlight the growing impact orthopaedic conditions will have on world health as life expectancy increases and the potential for realizing a cure for these problems is attained through future research advances.[1,2] Indeed, the impact on health from musculoskeletal disorders is tremendous. A survey conducted by the American Academy of Orthopaedic Surgeons reported that 7.3 million orthopaedic procedures were performed in US hospitals in 1995.[1] Of these, osteoarthritis and back pain were the most commonly treated problems. Musculoskeletal disorders account for $215 billion each year in health care costs and loss of economic productivity.[1]

Rheumatoid arthritis, although less common than OA, still affects 1% of the world population.[3,4] The long-term prognosis for RA is poor: the average life-expectancy of affected patients is reduced by between 3-18 years and 80% of patients are disabled after 20 years.[5,6] In the United States, each patient with RA requires an average of $5,919 per year in healthcare costs.[6] Contemporary drugs for RA are slow-acting with limited efficacy and many side-effects. Despite the many advances in our understanding of the pathophysiology of both RA and OA, the etiology of these disorders continues to elude us.

However, we are in the midst of a revolution in research design, techniques and capabilities. Proteomics, defined as the large-scale analysis of proteins, is emerging as a powerful field with large promise for un-locking many of the pathophysiologic mechanisms of disease. As a whole, proteomics encompasses many technical disciplines including light and electron microscopy, array and chip experiments, genetic read-out experiments, and mass spectroscopy (MS). However, of these various disciplines, MS-based proteomics is proving to be the technique of choice for high throughput analysis of complex protein samples.

The explosive development in MS-based proteomics has been made possible by several recent advances in the biomedical sciences. In the 1990s, biological mass spectroscopy evolved as a tool for rapid and powerful large-scale protein analysis and enabled scientists to overcome the limitations of protein analysis imposed by two-dimensional gel electrophoresis.[7] This rapidly evolving technology combined with the completion of the Human Genome Project in July 2000 and public access to the entire human genome have defined the beginning of this new era in biomedical research.

Still, proteomics, MS-based proteomics included, has many significant technical challenges to overcome. Mass spectroscopy of individual proteins has enabled us to develop the ability to identify almost any protein, analyze the protein for the presence of post-translational modifications (PTM's), characterize its protein-protein interactions and provide structural information about the specific protein in gas-phase experiments. However, mass spectroscopy of individual proteins does not equate to MS-based proteomics. The potential of proteomics promises a high-throughput simultaneous analysis of many proteins in a specific physiologic state. As of yet, the advances in proteomics have translated into very few clinically useful applications.

Nevertheless, each technological breakthrough that permits a new type of measurement or improves the quality of data analysis expands the range of potential applications for this very promising field. Our group is using MS-based proteomics and a novel experimental design to explore the potential of this technology for analysis of the complex mixture of proteins in synovial fluid from patients with early and end-stage RA and OA. We hope to identify specific biomarkers and potentially new etiologic factors in these diseases.

An Overview of MS-based Proteomics

As alluded to earlier, MS-based proteomics is a burgeoning field most of whose borders have yet to be explored. Mass spectrometric analysis occurs in the gas phase on ionized analytes. The two most commonly used methods for MS are electrospray

Dr. Gobezie is a resident in the Harvard Combined Orthopaedic Residency Program

Dr. Sarracino is the Director of Proteomics at the Harvard Center for Genomics and Genetics

Dr. Evans is Professor in the Department of Orthopaedic Surgery at the Brigham and Women's Hospital, Harvard Medical School

Dr. Thornhill is Professor of Orthopaedic Surgery and Chairman of the Department of Orthopaedic Surgery at Brigham and Women's Hospital, Harvard Medical School

Dr. Millett is Assistant Professor of Orthopaedic Surgery and member of the Harvard Shoulder Service, at Brigham and Women's Hospital and Massachusetts General Hospital, Harvard Medical School

Please address correspondence to:

Dr. Peter Millett
Department of Orthopaedic Surgery
Brigham & Women's Hospital
75 Francis Street
Boston, MA 02115

ionization (ESI), which ionizes the analytes out of a solution, or matrix-assisted laser desorption/ionization (MALDI), which sublimates and ionizes the analytes from a crystalline matrix using laser pulses.[8] ESI-MS is preferred for the analysis of complex mixtures of proteins whereas MALDI is commonly used for simpler protein mixtures because of its simplicity, excellent mass accuracy, high resolution and sensitivity.

Protein identification using MALDI is achieved with peptide-mass fingerprinting,[9] a technique whereby experimental peptide masses are matched against a calculated list of all peptide masses in a protein database. This method requires a purified target protein and is therefore combined with some prior method of protein fractionation such as 1D or 2D gel electrophoresis.

ESI is usually used with ion trap analyzers, an instrument that 'traps' ions for a given time interval prior to subjecting them to MS or MS/MS analysis.[10] The first generation 3D ion traps had relatively low mass accuracies; however, newer 2D ion traps have high sensitivities, mass accuracies, resolution and dynamic ranges. ESI coupled to ion traps is used to construct collision induced spectra (CID). [11] A peptide CID spectra generated from MS analysis can be compared against a comprehensive protein sequence database using various algorithms. Generally, three methods are used to identify proteins from CID spectra.[8] In one method, peptide sequence tags (short peptide sequences specific for a particular protein that are derived from a spectra's peak pattern) can be used with the mass information to determine the 'parent' protein. A second method, the 'cross-correlation' method, compares the spectra obtained from the experimental sample with theoretical spectra derived from protein databases to yield a 'matched' spectra and the likely identity of the protein. With a third method, termed 'probability based matching', the calculated fragments from peptide sequences in the database are compared with observed peaks and a score is generated that represents the statistical probability that a given spectra matches a peptide from the database. Hence, with MS-based proteomics, identification of proteins is limited to species whose proteome has been extensively characterized into protein databases.

Pushing the Envelope

New technology and techniques for combining mass spectroscopy, or tandem mass spectroscopy, allow us to achieve unprecedented sensitivity and specificity for identifying individual proteins within complex protein mixtures like synovial fluid. Hence, the goal of determining the proteome (a profile of all proteins expressed in the extracellular and/or intracellular environment) of body tissue in specific disease states is becoming a reality.

The development of LC-MS/MS is the foundation on which MS-based proteomics is built.[8,12,13]Theoretically, this method of protein analysis can detect very low abundance proteins in a complex mixture of peptides, although significant quantities of protein sample are required and the technique can be tedious. The basic techniques behind LC-MS/MS were pioneered by Hunt and colleagues during their study of MHC class I-associated peptides.[12] Generally, complex protein mixtures are digested with trypsin, usually after pre-separation by 1DE. The peptides are loaded on two-dimensional (strong cation exchange/reverse phase) or three-dimensional (strong cation exchange/avidin/reversed phase) chromatography columns and the eluants analyzed by MS or MS/MS. MS is a relatively poor instrument for quantification of proteins, due to the poorly understood relationship between the measured signal intensity and the quantity of analyte present. Hence, quantitative techniques have been developed for use with LC-MS/MS-the most popular is stable isotope dilution.[14,15] In this method, analytes with the same identity but different stable isotope composition are easily distinguished by MS due to their mass difference. Quantification is achieved using the ratio of signal intensities from the isotopic pairs.

How has this technology been used to study protein profiles in disease states so far? One of the most exciting studies sought to characterize the proteomes of four stages in the *Plasmodium falciparum* life-cycle using LC-MS/MS.[16] This study was able to identify 2,415 parasite proteins and, of these, 51% were hypothetical proteins confirming that the hypothetical ORFs predicted by gene modeling algorithms were functioning coding regions. These proteins are of particular interest as they represent potential targets for new anti-malarial therapies. Mass spectroscopy is also being applied to identify protein profiles in other diseases including prostate, ovarian and lung cancers, and various disorders relating to pre-term pregnancy such as eclampsia. To be sure, MS holds much promise to identify many of the pathophysiologic factors involved in virtually every disease.

How would our research add to what we know so far about OA and RA?

First of all, three reasons highlight why research into the etiologic mechanisms of OA and RA would benefit from the application of proteomics, and LC-MS/MS in particular. First, despite the advanced state of current knowledge on pathophysiologic mechanisms of these two diseases, we still do not know the etiological factor(s) that result in OA or RA. Second, the techniques that have been applied to discover what we do know about the pathophysiology of these diseases have been derived almost exclusively from the pre-proteomics era. Lastly, as a result of limits imposed by pre-proteomics era techniques for protein analysis, namely gel electrophoresis, strategies for identifying potential etiologic factors as well as determining their protein interactions have focused on hypothesis-driven research. This approach builds incrementally on what is already known about a specific diseases or mechanism and logically investigates plausibly important candidate genes or proteins. However, the ability to analyze complex mixtures of proteins with high-throughput techniques that permit simultaneous analysis of thousands of proteins has encouraged the development of a different approach to problems in research, a "discovery-based" approach.[17] As of yet, this "discovery-based" approach to investigating disease pathogenesis using high-throughput analysis of complex protein mixtures like synovial

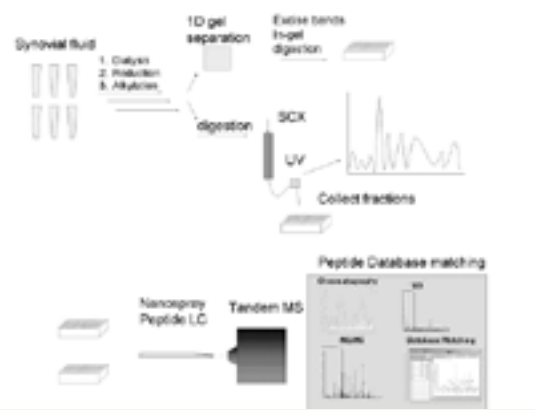

Fig. 1. We have successfully used two different methods to separate proteins prior to determining their identity using tandem mass spectroscopy: 1D gel electropheresis and liquid chromatography.

fluid in specific disease states has not been harnessed in the study of OA or RA.

The work from our group is innovative in that it will use LC-MS/MS with a "discovery-based" approach in an attempt to gain further insight into the disease pathogenesis of RA and OA. (Fig. 1) Our hope is to (1) identify new candidate proteins for further study as potential etiologic agents in OA and RA, using some of the conventional techniques outlined above; and (2) determine an "expression signature" for both RA and OA in order to identify potential biomarkers that could be used to develop improved screening tools for these diseases.

At present, RA is diagnosed primarily by criterion from clinical disease manifestations and the presence of rheumatoid factor (IgM-RF) in the serum of these patients. Rheumatoid factor is suboptimal because its relatively low specificity and sensitivity limit its diagnostic usefulness in the early phases of disease. Although other auto-antigens are being studied, including RA33, Sa, p68, calpastatin, perinuclear factor and antiperinuclear factor (APF) none of these antigens have demonstrated the kind of specificity and sensitivity for RA that translate into a reliable tool for early detection of this disease.[18-21] The need for a reliable biomarker for detection of RA early in the disease is particularly pressing since most of the contemporary antirheumatic therapies can best address the disease in its early phases.

Radiographic and clinical criteria are used as lagging indicators to diagnose OA, as they are usually sensitive only after the destruction of articular cartilage is well advanced; no biochemical markers for early diagnosis of OA have been developed. Again, this situation represents an unmet need for earlier diagnostic tools, as most novel therapeutic interventions such as cytokine receptor antagonists aim to stop progression of OA in its early stages.

The determination of distinct protein profiles for OA and RA, as well as the identification of candidate proteins involved in the pathogenesis of these diseases may represent two ideological outcomes from one result. That is, the protein profiles determined from an attempt at the complete characterization of the proteome of synovial fluid from patients at various stages of OA and RA may yield multiple proteins that can both serve as a potential biomarkers and plausible candidate proteins for further study. In fact, this study is a critical step in a multi-step process to both determine the etiologic factors behind OA and RA as well as identify new biomarkers that can be reliably used to screen for these diseases very early in their progression.

First, the identification of reliable biomarkers is a multi-step process. (Fig. 2) Our approach to this problem involves utilizing LC-MS/MS to determine a quantitative analysis of each

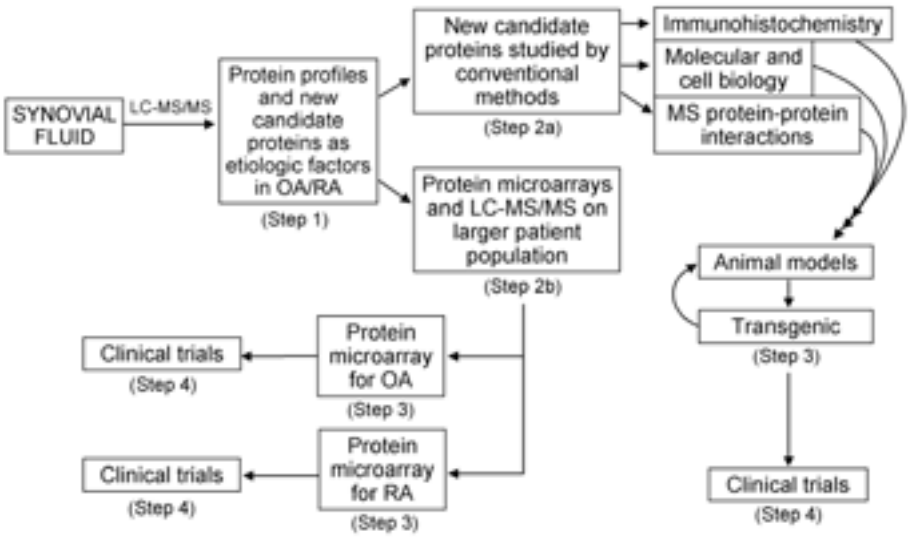

Fig. 2 This project represents a critical first step in a multi-step research effort.

protein in the protein profile of synovial fluid for both early and late OA and RA. Although a daunting task, our group has been able to make significant progress towards achieving this first objective. A comparative analysis of the proteomes for synovial fluid in each of these disease states should yield valuable insight into potential candidate proteins involved in the pathogenesis of both OA and RA. Next, customized protein microarrays would be constructed through our collaboration with the Harvard Partners Center for Genomics and Genetics in order to verify the data from our pilot study protein profiles. Once all of this data has been acquired and experimentally verified, clinical trials using this technology could commence.

Second, the identification of plausible candidate proteins from our analysis of the synovial fluid proteomes for early and late OA and RA would require further study using more conven-

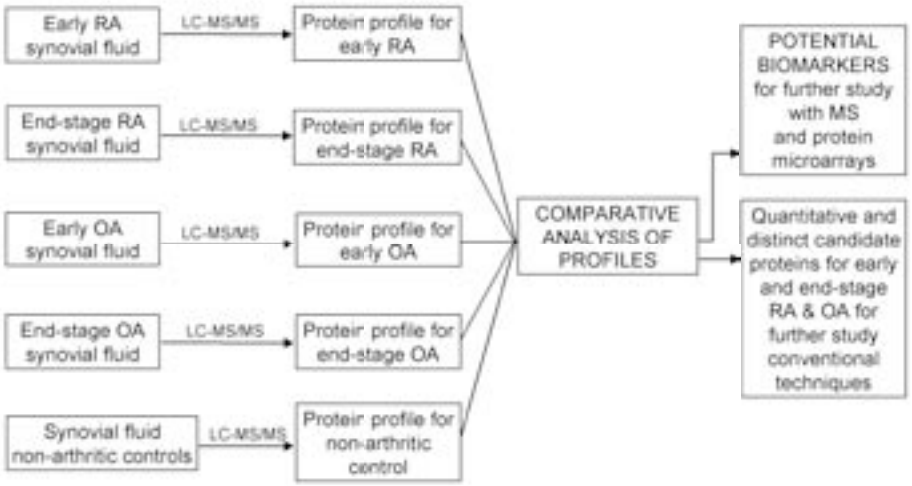

Fig. 3 Schematic design of our experimental design for using proteomics to study OA and RA.

tional techniques including immunohistochemistry, molecular and cellular biology. (Fig. 3) Our collaboration with the laboratory of Dr. Christopher Evans will enable us to better characterize the protein-protein interactions using hypothesis-driven experimental models, as well as to develop animal models using his expertise with gene transfer and molecular biology to study the role of these candidate proteins in disease pathogenesis and potential therapy in both RA and OA. Furthermore, cutting-edge techniques in mass spectroscopy would also be implemented in an attempt identify protein-protein interactions with these candidate proteins using proteomics techniques.

In summary, we hope the implementation of proteomics technology will permit us to identify protein profiles and potential new etiologic proteins involved in the pathogenesis of late OA and RA. Ultimately, the insights we gain from this study will result in the development of sensitive and specific biomarkers for both OA and RA that would improve our ability to detect these diseases early in their progression. In addition, the novel candidate proteins that we identify using these proteomics techniques will yield valuable therapeutic targets for new drug development.

References

1. **Boskey AL.** Musculoskeletal disorders and orthopaedic conditions. J American Medical Assoc 2011; 285: 619-623.
2. **Pramer A.** Musculoskeletal conditions in the United States. Rosemont, Ill: American Academy of Orthopaedic Surgeons; 1999.
3. **Schellekens GA, Visser H, de Jong BA, *et al.*** The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis Rheum 2000; 43: 155-169.
4. **Scott DL, Symmons DP, Coulton BL, *et al.*** Long-term outcome of treating rheumatoid arthritis: results after 20 years. Lancet 1987; 1: 1108-11.
5. **Pincus T, Callahan LF.** Taking mortality in rheumatoid arthritis seriously-predictive markers, socioeconomic status and comorbidity. J Rheumatol 1986; 13: 841-5.
6. **Yelin E, Wankle LA.** An assessment of the annula and long-term direct costs: the impact of poor function and functional decline. Arthritis Rheum 1999; 42: 1209-18.
7. **Goldring MB.** The role of the chondrocyte in osteoarthritis. Arthritis Rheum 2000; 43: 1916-1926.
8. **Aebersold R, Mann M.** Mass spectrometry-based proteomics. Nature 2003; 422: 198-207.
9. **Mann M, Wilm MS.** Error tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem 1994; 66: 4390-4399.
10. **Fenn JB, Mann M, Meng CK, *et al.*** Electrospray ionization for the mass spectrometry of large biomolecules. Science 1989; 246: 64-71.
11. **Aebersold R, Goolett DR.** Mass spectrometry in proteomics. Chem Rev 2001; 101: 269-295.
12. **Hunt DF.** Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 1992; 255: 1261-1263.
13. **Link AJ.** Direct analysis of protein complexes using mass spectrometry. Nature Biotechnol 1999; 17: 676-682.
14. **Han DK, Eng J, Zhou H, *et al.*** Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nature Biotechnol. 2001; 19: 946-951.
15. **Gygi SP, Rist B, Griffin TJ, *et al.*** Proteome analysis of low abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J Proteome Res 2002; 1: 47-54.
16. **Florens L, Washburn MP, Raine JD, *et al.*** A proteomic view of the Plasmodium falciparum life cycle. Nature 2002; 419: 520-506.
17. **Hassfeld W, Steiner G, Graninger W, *et al.*** Autoantibody to the nuclear antigen RA33; a marker for early rheumatoid arthritis. BR J Rheumatol 1993; 32: 199-203.
18. **Depres N, Boire G, Lopez-Longo FJ, *et al.*** The SA system: a novel antigen-antibody system specific for rheumatoid arthritis. J Rheumatol 1994; 21: 1027-33.
19. **Blass S, Haferkamp C, Specker C, *et al.*** Rheumatoid arthritis: autorective T cells recognizing a novel 68K autoantigen. Ann Rheum Dis 1997; 56: 317-322.
20. **Lackner KJ, Schlosser U, Lang B, *et al.*** Autoantibodies against human calpastatin in rheumatoid arthritis: epitope mapping and analysis of patient sera. Br J Rheumatol 1998; 37: 1164-71.
21. **Nienhuis RLF, Mandema EA.** A new serum factor in patients with rheumatoid arthritis: the antiperinuclear factor. Ann Rheum Dis 1964; 23: 302-5.