Progress in Musculoskeletal Proteomics: Another Look at Rheumatoid and Osteoarthritis

Bryan Krastins, BS, Reuben Gobezie, MD, David A. Sarracino, PhD, Thomas S. Thornhill, MD, Peter J. Millett, MD, MSc

Brigham and Women's Hospital

INTRODUCTION

Arthritis is one of the most common chronic debilitating diseases encountered by physicians throughout the world. Arthritis and chronic pain have become a major public health problem resulting in significant economic, social and human costs (1). Rheumatoid arthritis (RA) and osteoarthritis (OA) are two of the most common types of arthritis affecting the US population. Osteoarthritis currently affects 21 million people in the United States and, because the US population is growing older, the prevalence of osteoarthritis is increasing as well (2). By the year 2020 more than 50 million people in the United States will be affected by arthritis (3,4). Therefore, this will increase the direct and indirect dollars spent on healthcare and arthritis by more than 25% (2). Also, osteoarthritis is the most common reason for total hip and total knee replacements.

Rheumatoid arthritis is less common than osteoarthritis but still affects 440 million people worldwide (5,6). Rheumatoid arthritis is a complex autoimmune disease that results in chronic joint inflammation and destruction of the articular tissue. In the United States, the average life expectancy of affected patients is reduced by between 3-18 years, 80% of patients are disabled after 20 years, and each RA patient requires an average of $5,919 per year in healthcare costs (7,8). However, to date, in both RA and OA, the complex molecular mechanisms explaining the onset and continuation of the inflammation and the broad clinical heterogeneity, including different disease progression and therapy responses, are not fully understood.

Modern strategies designed to obtain deeper insights into the molecular complexity related to the pathology of OA and RA

Mr. Krastins is a research assistant at the Harvard Partners Center for Genomics and Genetics.
Dr. Gobezie is a resident in the Harvard Combined Orthopaedic Residency Program
Dr. Sarracino is Director of Proteomics at the Harvard Center for Genomics and Genetics
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, Harvard Medical School

Address correspondence to:

Brigham & Women's Hospital
75 Francis Street
Boston, MA 02115

make use of the global analyses of both transcriptome patterns determined by RNA microarray technologies and of protein patterns analyzed by two-dimensional gel electrophoresis (2-DE) in combination with mass spectrometry (9). Studies have proven that the correlation between mRNA and protein levels is rather inconsistent in higher organisms like humans and mice (10). Only the proteome identifies post-translational modifications such as phosphoralations and glycosylations, shows regulatory conversion of inactive proforms of enzymes or cytokines into their biological active forms, and displays protein breakdown products (Figure 1).
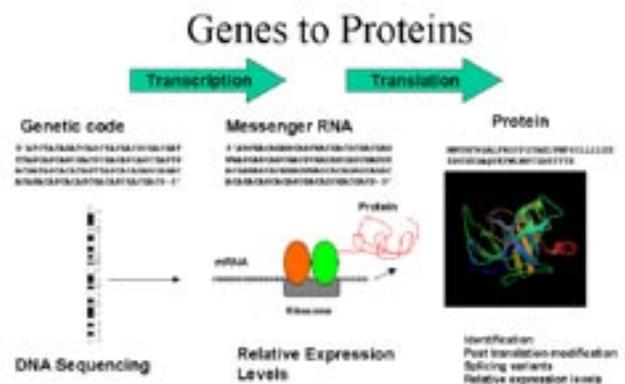

Figure 1. By studying proteins instead of mRNA, both post-translational modifications and splice variants can be identified using a proteomics platform.

Genomics and proteomics are complementary global approaches used to study, interpret and identify the molecular processes that influence a biological system. Thus, measuring mRNA and protein abundances provides more comprehensive information about the macromolecules most closely involved in cellular regulation and operation. These "omic" approaches provide a powerful picture of the transcriptional activity of cells and yield novel data for understanding disorders like arthritis by analyzing tissues throughout the progression of disease (10).

If one assumes that key proteins diagnostic to a disease are higher in concentration at the site of the disease than in blood and urine where the concentration may be less, then discovery efforts will be more successful by studying biologic fluids closer to the site of the disease, like synovial fluid from OA and RA patients. Following this assumption, our group has decided to utilize mass spectrometry-based (MS) proteomics and a novel experimental design to explore 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.

MS-BASED PROTEOMICS – AN OVERVIEW

Proteomics involves the study of proteins expressed by the genome in a biological system. Ideologically, proteomics proposes to identify all proteins expressed under different physiologic and pathophysiologic conditions, characterize protein modifications, analyze protein-protein interactions, and protein structure (11). Such knowledge can provide a basis for understanding genetic variants, gene function, and the mechanisms of action needed to develop the means to diagnose, treat and prevent against disease.
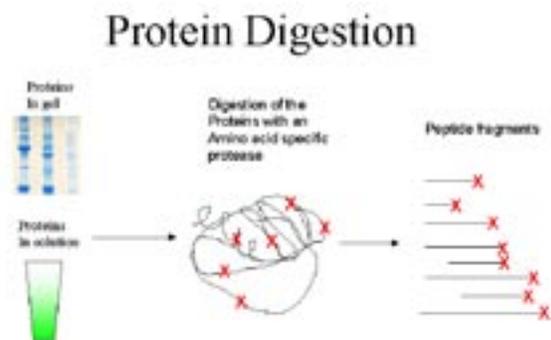

Figure 2. Two different proteomic approaches used to identify the proteome of a sample.

Proteomic measurement techniques are generally divided into two different approaches (Figure 2). Gel-based approaches, such as 2-DE, typically utilize matrix-assisted laser desorption/ionization (MALDI), mass spectrometry, and peptide mass finger printing for the identification of gel separated proteins. The second technique is a solution-based approach. Solution-based approaches generally utilize liquid chromatography (LC) separations with electrospray ionization (ESI) tandem mass spectrometry for the identification of peptides obtained by tryptic digestion of complex protein mixtures such as whole cell lysates or serum. Both of these approaches are invaluable to identifying proteins, however, they still suffer from significant limitations; specifically, the lack of high-throughput necessary to analyze the complex mixtures of proteins that compose most biological tissues (11,12).

During the last several years, the technology for MS-based proteomics has evolved so that the high resolution and mass measurement accuracy of Fourier transform ion cyclotron resonance (FT-ICR) is combined with the high efficiency capillary LC separation, and amino acid sequence information to enable comprehensive high-throughput measurements. This accurate mass and time-tag approach provides high quality quantitative data on protein identity and abundance at high sensitivity without having to perform multiple tandem MS steps, thus dramatically improving the throughput of proteomic measurements.

MS-BASED PROTEOMICS – NEW TECHNOLOGY

Of the different mass spectrometers available, the FT-ICR holds the greatest potential for researchers because of the ultra-high resolution and mass accuracy that can be achieved (12).

Such performance allows unequivocal mass assignment and resolution of species that could not otherwise have been distinguished by other mass spectrometers (13). In FT-ICR mass spectrometry, the ions are generated externally in a separate ion source and then injected into a container known as the cell, which is usually cubical or cylindrical in form. The FT-ICR analyzer cell is located within a strong magnetic field, generated by using a superconducting magnet (13,14).

Electrospray ionization is the ionization technique most frequently coupled with the FT-ICR mass spectrometry. The ESI technique is advantageous because it results in the vaporization and ionization of a sample while minimizing fragmentation
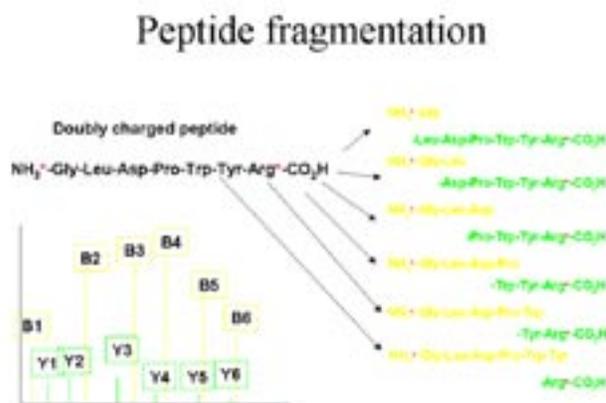

Figure 3. An example of the fragmentation pattern of a doubly charged peptide.

(Figure 3). The minimization of fragmentation is important for biological molecules and complex mixtures because fragments can interfere with the mass spectrum when trying to determine the original constituents (15,16).

FT-ICR mass spectrometers are best known for their excellent mass accuracy (measured in ppm) and resolution (11). Mass accuracy is a measurement of how well the observed ratio of mass to charge (m/z) correlates with the 'real value' and is extremely important in situations when multiple ion fragmentation patterns could be assigned to a single m/z value. Resolution is extremely important for separating closely spaced signals from a complex mixture of multiply-charged ions generated by ESI. While most mass spectrometers have the capability to resolve ions in the 10,000 to 100,000 ppm range, the FT-ICR mass spectrometer has the ability to routinely reach resolutions of hundreds of thousands to a few million ppm (14,15). The ability to reach such high resolutions can sometimes make the difference between identifying and not identifying a species of interest.

The FT-ICR mass spectrometer has the ability to comprehensively identify, characterize, and quantify proteins in complex biological systems and is considered essential for revealing the response of an organism to stimuli such as injury or disease (11,13,15). By utilizing the FT-ICR's properties such as mass accuracy and high resolution, we believe that our proteomics research probing for etiological mechanisms

of OA and RA in a complex biological fluid such as synovial fluid, will benefit greatly because the FT-ICR technology allows for protein identifications to be made at greater depths, while quantifications can be made with better accuracy, precision and greater confidence.
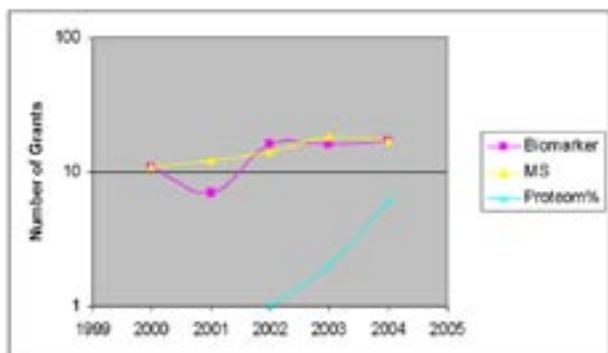

Figure 4. A graph depicting the number of funded grants by the NIH over the past 5 years.

CURRENT BIOMARKER DISCOVER RESEARCH

The need to identify biological markers for musculoskeletal disorder is arguably one of the most significant milestones challenging the scientific community today. Protein biomarkers, in particular, serve as the most likely 'candidate proteins' involved in disease pathogenesis. In addition, these biomarkers are important since they, theoretically, would allow us to develop diagnostics and potential therapeutic agents against disease as well as better characterize disease progression (Figure 4). Furthermore, the biomarkers that accurately substitute for a clinical endpoint can become accepted as an alternate endpoint over time and be used as a diagnostic and prognostic measure in the clinic.

There are several factors that are involved when attempting to identify the potential success of a biomarker candidate. Factors critical to the success of biomarker discovery strategies aimed at providing measurements in biological fluids include the following: 1) carefully identified samples from patients with a condition suitable for analysis; 2) a "global" analytic tool with a wide dynamic range to interrogate a large number of proteins in a reasonable time frame; 3) a candidate selection strategy that is unbiased; and 4) a validation strategy that can be executed quickly, using samples from patients with a diverse range of diseases including the disease of interest (10,17,18).

Achieving all of these aims is not simple and each element poses a different problem. One of the most important aspects of biomarker discovery is the analytic method used.

To achieve a greater in depth analysis of complex biological fluids such as synovial fluid, a protein fractionation step of each sample is required (Figure 5). There are several techniques investigators can use to separate samples containing a complex mixture of proteins. These techniques include depletion of abundant proteins, size-fractionation using both liquid chromatography and SDS-PAGE analysis, and two-dimensional gel electrophoresis. Using these techniques with the aid of mass spectrometry technology has shown to be a very useful platform for detecting potential biomarkers in both serum and synovial fluid from patients with OA and RA (19,20).

Drynda et al. have used 2-DE in combination with mass spectrometry to visualize and to identify proteins in synovial fluid and plasma samples from patients with RA and OA. By using the 2-DE fractionation technique, the small calcium binding protein S100A9 was identified as a discriminatory marker protein in synovial fluid by using a global proteomic analysis (21). Another study by Sinz et al. demonstrated that the use of 2-DE combined with MALDI-TOF-MS for the analysis of proteins present in plasma and synovial fluid of various patients demonstrated changes in protein patterns in body fluids of patients suffering from various forms of arthritis. By means of mass spectrometry, they also were able to identify molecular parameters and determine differential protein structure modifications that are thought to play a role for specifically determining a defined pathological state of diseased joints (22).
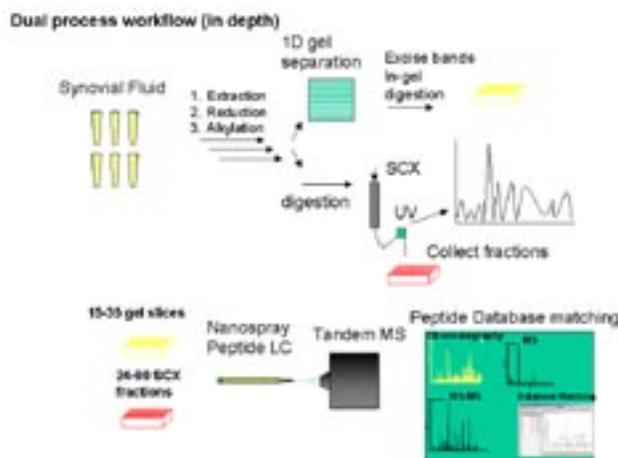

Figure 5. Two different methods to separate proteins prior to determining their identity using tandem mass spectroscopy: 1D gel electrophoresis and liquid chromatography.

Another group of researchers have taken a two-step proteomic approach toward the detection of potential protein biomarkers in serum and synovial fluid. In their first step, two-dimensional liquid chromatography followed by tandem mass spectrometry was used to generate a complete protein profile of synovial fluid from patients with erosive RA and nonerosive RA. In the next step, selected candidate markers from synovial fluids were verified in serum using quantitative mass spectrometry methods (23). By using this two-step approach to biomarker discovery, several biomarker candidates for RA were verified and a subset of these could possibly be used as an indicator for RA disease severity. The clinical application of such biomarkers, although requiring several levels of validation, could aid clinicians with a tool for selecting patients whose conditions merit high-cost and high-risk treatment early enough in the disease course to prevent unnecessary joint destruction and/or disability.

Given the promise and power of proteomics, the industry and technology present must adapt to produce informative biomarkers for diagnosis and prognosis. There are several factors that will play an important role for the future of biomarker discovery. First, the clinician must provide well-characterized and stored samples from various disease states and conditions. Second, samples must be collected carefully and longitudinally obtained samples must be collected in order to further the understanding of individual variability over time. Thirdly, methodology improvements must allow for greater in-depth analysis of samples with a higher throughput. Finally, powered studies to validate markers across multiple disease setting to avoid identification of biomarkers that pertain to ill health (24-26).

OUR BIOMARKER RESEARCH – ADVANCEMENTS FOR OA AND RA

There are several reasons why research into the etiologic mechanisms of OA and RA will benefit from using a proteomics platform. First, although potential biomarkers for both OA and RA have been currently detected, we currently do not know the etiological factors that result in OA and RA. Second, the techniques that have been applied to discover what we do know about the pathophysiology of these diseases have been low throughput and the number of samples analyzed has been statistically insignificant, thus directly affecting the patient population used to identify potential protein biomarkers. Finally, as a result of limits imposed by pre-proteomics era techniques for protein analysis, the strategies for identifying potential etiologic factors as well as determining protein interactions, protein modifications and protein breakdown products have focused on hypothesis driven research. However, the ability to analyze complex mixtures of proteins using a high-throughput approach which allows for the simultaneous identification of thousands of proteins at a time has encouraged the development of a biomarker discover based research approach.

The work from our group is innovative in that it we used liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) with a high-throughput protein biomarker discover based approach in an attempt to gain further insight into the disease pathogenesis of OA and RA. Our hope is to identify new candidate proteins for further study as potential etiologic factors in OA and RA. We also hope to determine an expression signature for both OA and RA in order to identify potential biomarkers that could be used to develop improved early detection and screening tools used for these diseases.

Currently, 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 as an early detection marker of the disease. Although other autoantigens are being studied as potential early staged biomarkers, thus far none of these antigens have demonstrated the kind of specificity and sensitivity for the use as an early detection tool for RA (27-30). The need for a reliable biomarker for the detection of early RA is particularly pressing since most of the contemporary anti-rheumatic therapies can best address the disease in its early phases.

Presently, radiographic and clinical criteria are used as lagging indicators to diagnose OA. These techniques are sensitive enough to diagnosis OA only after the destruction of articular cartilage is well advanced. To date, there are no biochemical markers for detection leading to early diagnosis of OA. This problem faced by physicians represents an unmet need for earlier diagnostic detection tools for OA.

The determination of 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. The protein profiles determined from the complete characterization of the proteome of synovial fluid from patients at various disease stages of OA and RA may yield multiple proteins that can both serve as potential biomarkers for earlier diagnostic detection tools as well as lead to further studies.

Our approach to identify reliable protein biomarkers involves utilizing LC-MS/MS to determine the quantitative analysis of each protein in the protein profile of synovial fluid for normal patients and both early and late OA and RA. A comparative analysis of the synovial fluid proteomes in each of these diseased and non-diseased states should yield valuable insight into identifying potential candidate proteins involved with the pathogenesis of both OA and RA.

Next, the identification of plausible candidate proteins from our analysis of the synovial fluid proteomes from early and late OA and RA disease states would require further analysis using more conventional techniques including immunohistochemistry, molecular and cellular biology. Also, we would like to better characterize the protein-protein interactions using hypothesis driven experimental models by studying the potential protein biomarkers we identify from the pilot study. We also would like to use the pilot data to develop animal models using gene transfer and molecular biology to study the role of these candidate proteins in disease pathogenesis therapy in both OA and RA. Again, using the data obtained from the pilot study, we plan to utilize the excellent mass accuracy and high resolution of the FT-ICR mass spectrometer to identify protein-protein interactions, protein modifications and also calculate the absolute quantification of our potential protein biomarkers from both early and late OA and RA synovial fluid samples.

In summary, we hope the implementation of proteomics technology will allow us to identify protein profiles and potentially new etiologic proteins involved in the pathogenesis of early and late OA and RA. Ultimately, the insights we gain from this study will results in the development of sensitive and specific biomarkers for both OA and RA that will help to improve our ability to detect these diseases early in a patient's progression. In addition, the novel candidate proteins that we identify using these proteomic techniques may yield potentially valuable therapeutic targets for new drug development.

References

1. **Nour K:** Appreciating the predicament of housebound older adults with arthritis: portrait of a population. *Can J Aging* 2005 Spring; 24(1): 57-69.
2. **Rubin BR:** Osteoarthritis. *Journal of the American Association* 2001; 101(4): S2-S5.
3. **Pramer A:** Musculoskeletal conditions in the United States. *Rosemont, Ill: American Academy of Orthopaedic Surgeons*; 1999.
4. **Boskey AL:** Musculoskeletal disorders and orthopaedic conditions. *J American Medical Assoc* 2011; 285: 619-623.
5. **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.
6. **Scott DL, Symmons DP, Coulton BL, et al:** Long-term outcome of treating rheumatoid arthritis: results after 20 years. *Lancet* 1987; 1: 1108-11.
7. **Pincus T, Callahan LF:** Taking mortality in rheumatoid arthritis seriously-predictive markers, socioeconomic status and comorbidity. *J Rheumatol* 1986; 13: 841-5.
8. **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.
9. **Amin AR:** A need for a 'whole-istic functional genomics' approach in complex human diseases: arthritis. *Arthritis Res Ther* 2003; 5(2): 76-9.
10. **Patterson SD:** Proteomics: beginning to realize its promise? *Arthritis Rheum* 2004 Dec; 50(12): 3741-4.
11. **Olsen JV:** Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A* 2004 Sep 14; 101(37): 13417-22.
12. **Pasa-Tolic L:**Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* 2004 Oct; 37(4): 621-4, 626-33, 636.
13. **Qian WJ:** Quantitative Proteome Analysis of Human Plasma following in Vivo Lipopolysaccharide Administration Using 16O/18O Labeling and the Accurate Mass and Time Tag Approach. *Mol Cell Proteomics* 2005 May; 4(5): 700-709.
14. **Barrow MP:** Principles of Fourier transform ion cyclotron resonance mass spectrometry and its application in structural biology. *Analyst* 2005 Jan; 130(1): 18-28.
15. **Strupat K:** High-Throughput Analysis in Proteomics: Parallel Mass Detection using Finnigan LTQ FT. *Thermo Electron Corporation* 2004 Application Note: 30046.
16. **Shen Y:** Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal Chem* 2004 Feb 15; 76(4): 1134-44.
17. **Kantor AB:** Biomarker discovery by comprehensive phenotyping for autoimmune diseases. *Clin Immunol* 2004 May; 111(2): 186-95.
18. **Robinson WH:** Millennium award recipient contribution. Proteomics for the Development of DNA Tolerizing Vaccines to Treat Autoimmune Disease. *Clin Immunol* 2002 April; 103(1): 7-12.
19. **Gianazza E:** Monitoring the effects of drug treatment in rat models of disease by serum protein analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002 May 5; 771(1-2): 107-30. Review.
20. **Lorenz P:** Proteome analysis of diseased joints from mice suffering from collagen-induced arthritis. *Clin Chem Lab Med* 2003 Dec; 41(12): 1622-32.
21. **Drynda S:** Proteome analysis reveals disease-associated marker proteins to differentiate RA patients from other inflammatory joint diseases with the potential to monitor anti-TNFalpha therapy. *Pathol Res Pract* 2004; 200(2): 165-71.
22. **Sinz A:** Mass spectrometric proteome analyses of synovial fluids and plasmas from patients suffering from rheumatoid arthritis and comparison to reactive arthritis or osteoarthritis. *Electrophoresis* 2002 Sep; 23(19): 3445-56.
23. **Liao H:** Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. *Arthritis Rheum*. 2004 Dec; 50(12): 3792-803.
24. **Dotzlaw H:** A pattern of protein expression in peripheral blood mononuclear cells distinguishes rheumatoid arthritis patients from healthy individuals. *Biochim Biophys Acta* 2004 Jan 14; 1696(1):121-9.
25. **Xiang Y:** Proteomic surveillance of autoimmunity in osteoarthritis: identification of triosephosphate isomerase as an autoantigen in patients with osteoarthritis. *Arthritis Rheum* 2004 May; 50(5): 1511-21.
26. **Moseley TA:** Interleukin-17 family and IL-17 receptors. *Cytokine Growth Factor Rev* 2003 Apr; 14(2): 155-74. Review.
27. **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.
28. **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.
29. **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.
30. **Nienhuis RLF, Mandema EA.** A new serum factor in patients with rheumatoid arthritis: the antiperinuclear factor. Ann Rheum Dis 1964; 23: 302-5.